Stochastic Dominance

Investment Decision Making under Uncertainty

2nd Edition

Haim Levy

Springer

STOCHASTIC DOMINANCE

Investment Decision Making under Uncertainty

Second Edition

Stochastic Dominance, Second Edition

by Haim Levy
Myles Robinson Professor of Finance
The Hebrew University of Jerusalem

Previously published books in the series:

Viscusi, W. Kip: STUDIES IN RISK AND UNCERTAINTY

Luken, R.: ENVIRONMENTAL REGULATION: TECHNOLOGY, AMBIENT AND BENEFITS-BASED APPROACHES

Shubik, M.: RISK, ORGANIZATIONS AND SOCIETY

Edwards, W.: UTILITY THEORIES: MEASUREMENTS AND APPLICATIONS

Martin, W.: ENVIRONMENTAL ECONOMICS AND THE MINING INDUSTRY

Kunreuther, H. and Easterling, D.: THE DILEMMA OF A SITING OF HIGH-LEVEL NUCLEAR WASTE REPOSITORY

Christe, N.G.S. and Soguel, N.C.: CONTINGENT VALUATION, TRANSPORT SAFETY

Battigalli, P., Montesano, A. and Panunzi, F.: DECISIONS, GAMES AND MARKETS

Freeman, P. and Kunreuther, H.: MANAGING ENVIRONMENTAL RISK THROUGH INSURANCE

Kopp, R.J., Pommerehne, W. W. and Schwartz, N.: DETERMINING THE VALUE OF NON-MARKETED GOODS

Bishop, R.C. and Romano, D.: ENVIRONMENTAL RESOURCE VALUATION: APPLICATIONS OF THE CONTINGENT VALUATION METHOD IN ITALY

Jeanrenaud, C.; Soguel, N.C.: VALUING THE COST OF SMOKING

Eeckhoudt, L.: RISK AND MEDICAL DECISION MAKING

Viscusi, W. Kip: THE RISKS OF TERRORISM

STOCHASTIC DOMINANCE

Investment Decision Making under Uncertainty

Second Edition

by

Haim Levy
Myles Robinson Professor of Finance
The Hebrew University of Jerusalem

Springer

Library of Congress Control Number: 2005933466
ISBN-10: 0-387-29302-7 e-ISBN 0-387-29311-6
ISBN-13: 978-0387-29302-8

Printed on acid-free paper.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

springeronline.com

To Tal, Shira, Tamar, Neta, Romi, Maya, and Alon

TABLE OF CONTENTS

PREFACE

This second edition of *Stochastic Dominance* is devoted to investment decision-making under uncertainty. The book covers four basic approaches to this process:

a) The stochastic dominance (SD) approach, developed on the foundation of von-Neumann and Morgenstern[1] expected utility paradigm.

b) The mean-variance approach developed by Markowitz[2] on the foundation of von-Neumann and Morgenstern's expected utility or simply on the assumption of a utility function based on mean and variance.

c) The "almost" stochastic dominance (ASD) rules and the "almost" mean-variance rule (AMV). No matter whether one employs objective or subjective probabilities, the common stochastic dominance criteria and the mean variance rule may lead to paradoxes: they are unable to rank prospect A which yields $1 with a probability of 0.01 and a million dollars with probability of 0.99, and prospect B which yields $2 with certainty. This is an absurdity as in any sample of subjects one takes, 100% of subjects choose A. The "almost" stochastic dominance criteria and "almost" mean variance rule, which have been recently been developed by Leshno and Levy in 2002[3], suggest a remedy to such paradoxes.

d) The non-expected utility approach, focusing on prospect theory (PT) and its modified version, cumulative prospect theory (CPT). This theory is based on an experimental finding showing that subjects participating in laboratory experiments often violate expected utility maximization: they tend to use subjective probability beliefs that differ systematically from the objective probabilities and to base their decisions on changes in wealth rather than on total wealth.

The above approaches are discussed and compared in this book. We also discuss cases in which stochastic dominance rules coincide with the mean-variance rule and cases in which contradictions between these two approaches may occur. We then discuss the relationship between stochastic dominance rules and prospect theory, and establish a new investment decision rule which combines the two and which we call prospect stochastic dominance (PSD). Though it seems that prospect theory and the mean-variance rule and, in particular, the equilibrium capital asset pricing model (CAPM) cannot coexist, we show in this book that under weak assumption there is no contradiction between these seemingly remote

---

[1] von Neumann, J., and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, N.J., 1953.

[2] Markowitz, H.M., "Portfolio Selection," *Journal of Finance*, 1952, pp. 77-91.

[3] Leshno, M. and H. Levy, "Preferred by All and Preferred by Most Decision Makers: Almost Stochastic Dominance," *Management Science*, August 2002, 48, 8, pp. 1074-1085.

paradigms. Finally, in contradiction to the S-shape preferences, Markowitz suggested as early as 1952 that preferences are typically *reverse* S-shaped. This book suggests a new stochastic dominance rule called Markowitz's[4] stochastic dominance (MSD) rule corresponding to all reverse-S-shape preferences. We present experimental tests, which examine the validity of the various hypotheses regarding preferences.

Concepts similar to stochastic dominance have been known for many years but the two papers published by Hadar and Russell, and Hanoch and Levy in 1969, and the paper published by Rochschild and Stiglitz in 1970[5] paved the way for a new paradigm called stochastic dominance, with hundreds of studies following in their tracks. These studies, deal with theoretical as well as empirical issues in various areas of economics, finance, accounting, statistics, agriculture, and medicine.

The need to develop the stochastic dominance rules, at least in the view of the author of this book, stems from paradoxes that are sometimes revealed by the commonly used mean-variance rule. To be more specific, there are cases in which a clear-cut choice between two risky assets exists, yet the mean-variance rule is unable to rank the two alternate investments. When I was a second-year MBA student, only the mean-variance investment rule was taught. I presented my teacher with the case of two alternative investments: x providing $1 or $2 with equal probability and y providing $2 or $4 with equal probability, with an identical initial investment of, say, $1.1. A simple calculation shows that both the mean and the variance of y are greater than the corresponding parameters of x; hence the mean-variance rule remains silent regarding the choice between x and y. Yet, any rational investor would (and should) select y, because the lowest return on y is equal to the largest return on x. Well, this is a trivial case in which the mean-variance rule fails to show the superiority of one investment over another. However, there are many more such cases in which the mean-variance rule is unable to rank two investments. These cases are sometimes quite complex and the superiority of one investment over the other cannot be detected by the naked eye: hence my motivation to develop general decision rules, well-known nowadays as stochastic dominance rules. Later on, in 2002, Leshno and Levy have found paradoxes also within the stochastic dominance paradigm, even after the mean-variance paradoxes have been resolved. Therefore, they suggested "almost" stochastic dominance (ASD) rules to solve the remaining paradoxes.

---

[4] Markowitz, H.M., "The Utility of Wealth," *Journal of Political Economics*, 1952, 60, pp. 151-156.

[5] Hadar, J. and W.R. Russell, "Rules for Ordering Uncertain Prospects," *American Economic Review*,1969, pp. 25-34; Hanoch, G. and H. Levy, "The Efficient Analysis of Choice Involving Risk," *Review of Economic Studies*, 1969, pp. 335-346 and Rothschild, M. and J. Stiglitz, "Increasing Risk. I. A Definition," *Journal of Economic Theory*, 1970, pp. 225-243.

In spite of its common use, the mean-variance rule coincides precisely[6] with the expected utility paradigm in the following well-known two cases: normal distribution of returns in the face of risk aversion, and quadratic utility function. The quadratic utility function, apart from the disadvantage of restricting the analysis to one type of preference, has other drawbacks, too. Hence, most researchers focus on the normal case. However, in the normal case, we face a severe problem: the returns range is from minus infinity to plus infinity but actual rates of returns are bounded by −100%, that is, the asset price can drop to zero and no further. Assuming lognormal distribution provides a rescue from this limited liability issue because lognormal distribution is defined only for non-negative asset prices, thus conforming with the fact that, in practice, asset prices can not be negative. However, although the assumption of lognormal distribution overcomes the limited liability issue, it gives rise to another issue: if each asset is lognormally distributed, linear combination of these assets (a portfolio) will not be lognormally distributed. The continuous time model suggested by Merton,[7] overcomes this difficulty of lognormal distribution because, for any finite terminal date, each selected portfolio will be lognormally distributed. However, the lognormal case is saved at the cost of assuming a continuous time model in which the investor revises the portfolio weights continuously. This implies that even a tiny transaction cost will induce a negative rate of return on the terminal date; hence, existing transaction costs ruin the model.

The stochastic dominance framework does not suffer from the above deficiencies. However, this paradigm has its own deficiencies: a method to construct all the efficient portfolios and a separation theorem (as in the mean-variance framework) have yet to be developed. In 2003, Post[8] has made an important step in this direction. He introduced a technique to find whether the market portfolio is second degree efficient relative to all diversified portfolios composed from a given set of assets. We do not have yet a stochastic dominance equilbrium, but Post's work is an important step in this direction. Therefore, we suggest that stochastic dominance rules do not substitute for the mean-variance rule but rather offer an alternative approach, complimenting rather than replacing it. However, in cases that are not related to portfolio construction (e.g., applications in agriculture, medicine, statistics, and some applications in economics and finance), employment of the stochastic dominance paradigm is superior because no assumptions are needed regarding the distribution of returns.

The book starts with various commonly used measures of risk (Chapter 1) leading up to the expected utility paradigm (Chapter 2) which shows that the only relevant measure of risk is the risk premium. As the risk premium varies from one investor

[6] One can use the mean-variance rule as an approximation to the expected utility. For more details, see Levy, H., and H. Markowitz, "Approximating Expected Utility by a Function of Mean and Variance," *American Economic Review,* 1979, pp. 308-317.

[7] Merton, R.E., "An Intertemporal Capital Pricing Model," *Econometrica*, September 1973, pp. 867-887.

[8] Post, T., "Empirical test for stochastic dominance efficiency," *Journal of Finance*, 2003, 58, p. 1905-1931.

to another, we conclude that, in general, no one single objective index has the capacity to rank investments by their risk. Thus, the whole distribution of returns rather than one measure of profitability and one measure of risk has to be considered. Chapter 3 constitutes the heart of the book. In this chapter, we develop and discuss first, second, and third degree stochastic dominance rules (FSD, SSD and TSD, respectively). In Chapter 4, we extend the stochastic dominance rules to include riskless assets. In order to do this, we first reformulate the stochastic dominance rules in terms of distribution quantiles rather than cumulative distributions (this can be done for FSD and SSD but not TSD). Algorithms for all these stochastic dominance rules are provided in Chapter 5. Having the general stochastic dominance rules with no constraints on the distribution returns under our belt, we proceed in Chapter 6, to stochastic dominance for specific distributions including normal, lognormal and other truncated distributions. Then, in Chapter 7, we provide empirical evidence regarding the effectiveness of the stochastic dominance rules as well as the mean-variance rule.

In Chapter 8 we present a few of the many applications of stochastic dominance rules in various fields of research. We discuss these applications briefly and provide references to these studies at the end of the book. Chapter 9 is devoted to the definition of situations in which one asset is identified as "more risky" than another asset, and the extension of this definition to DARA utility functions as well as to the case where the riskless asset exists. Chapter 10 is devoted to stochastic dominance and diversification and, in particular, to the effect of changes in the cumulative distributions on diversification. Chapter 11 analyzes the effect of changes in the assumed investment horizon on the efficient set in the frameworks of mean-variance and stochastic dominance.

The capital asset pricing model (CAPM) of Sharpe and Lintner[9] is undoubtedly one of the main cornerstones of modern finance. However, the CAPM holds only under a set of confining assumptions, one of them being that all investors have the same investment horizon. In Chapter 12, based on the work of Levy and Samuelson[10], we use stochastic dominance rules to show that the CAPM holds under a much wider set of assumptions, even if investors do not have the same horizon.

Chapter 13 is a new chapter covering the new decision rules called "almost" stochastic dominance (ASD) rules. This chapter reveals that the existing SD rules and the MV rule alike may be unable to rank two prospects, where it is obvious that in any sample of subjects one takes, all would choose one specific prospect. ASD rules correct for this deficiency of SD and MV rules.

---

[9] Sharpe, W.F., "Capital Asset Prices: A Theory of Market Equillibrium Under Conditions of Risk," *Journal of Finance*, September, 1964, pp: 425–442, and Lintner, J., "Security Prices, Risk, and Maximal Gains from Diversification," *Journal of Finance*, December 1965, pp. 587–615.

[10] Levy, H., and P.A., Samuelson, "The Capital Asset Pricing Model with Diverse Holding Periods," *Management Science*, November 1992, pp. 1529–1542.

Chapter 14 is devoted to non-expected utility theory with Prospect Theory suggested by Kahaneman and Tversky[11] as the main competing theory. We suggest a reconciliation between the two competing theories by focusing on short- and long-term investment decisions.

Chapter 15 is a new chapter. In this chapter we develop prospect stochastic dominance (PSD) and Markowitz's stochastic dominance (MSD) rules. We present experimental tests which examine the validity of risk averse preferences, cumulative prospect theory (CPT) preferences and Markowitz's preferences.

Chapter 16 concludes the book with suggestions for further research and presentation of unsolved problems in the area of investment decision making, with emphasis on stochastic dominance. Readers interested in this field are welcome to pursue these research ideas.

CHANGES IN THE SECOND EDITION

1. Ch. 1: Value at Risk (VaR) as a measure of risk is added.

2. Ch. 2: A discussion of utility of wealth and utility of change of wealth is added.

3. Ch. 3: Risk-Seeking Stochastic Dominance (RSD) is added.

4. Ch. 4: The quantile formulation of third-degree stochastic dominance, as stated in the first edition was wrong, hence is omitted from the second edition.

5. Ch. 5: Most of the published TSD algorithms are wrong, including the one published in the first edition of the book. In the second edition, we provide a new and correct TSD algorithm, accompanied with an empirical study testing the effectiveness of TSD with the new algorithm.

6. Ch. 7: Discussion and proofs of convex stochastic dominance (CSD) are added.

7. A new chapter 13, analyzing "almost" stochastic dominance is added.

8. A new chapter 15, analyzed prospect stochastic dominance, and Markowitz's stochastic dominance, accompanied with the experimental findings is added.

---

[11] Kahaneman, D., and A. Tversky, "Prospect Theory of Decisions Under Risk," *Econometrica*, 1979, pp. 263–291.

9. Since the publication of the first edition of this book, several of the research ideas in the area of stochastic dominance presented in the first edition have been conducted. In the second edition, we suggest some additional research ideas.

AUDIENCE

This book is intended mainly for Ph.D. students, advanced MBA students specializing in finance, and advanced MA economics students interested in the economics of uncertainty. The book can be used also as a supplementary book in post-graduate courses on portfolio selection and investment decision making under uncertainty.

ACKNOWLEDGEMENT

I would like to thank Moshe Levy, and Zvi Wiener who read several chapters of this book and to Moshe Leshno and Boaz Leibovitz who read and commented on the whole manuscript. I derived a great deal of benefit from their comments on the earlier drafts of this book. I also thank Moshe Levy for providing me the proofs of convex stochastic dominance (see Chapter 7). I would also like to thank Yoel Hecht for preparing all the figures and to him and Allon Cohen for checking and improving the algorithms for the stochastic dominance rules. Thierry Post inspired me with both enthusiasm and his new approach to analyze market efficiency with SD criteria. I would like to thank Hyla Berkowitz for typing the many drafts of this book and Maya Landau for her editorial contribution. Finally, I would like to thank the editor, Sean Lorre, for his great help. Without his encouragement and effort, the book would not be published on time.

1

ON THE MEASUREMENT OF RISK

1.1 WHAT IS RISK?

As we go about our daily business, we inevitably overhear snippets of financial wisdom. "The investment is too risky"; "The risk involved in the investment is relatively low"; "By diversifying your investment portfolio you can reduce the risk"; "Putting all your eggs in one basket is too risky". Claims such as these ignore the notion of riskiness but the exact definition of risk and, and in particular, how to measure it, remain vague. People may have a "feel" as to what risk means but, if asked how to measure it, or to rank a number of investment prospects by their risk, there would be little consensus. We would probably be offered diverse intuitive explanation, some quite colorful. Few would furnish a quantitative answer. Webster's dictionary is hardly illuminating in this respect; among its definitions of risk we find:[1]

> "Exposure to the chance of injury of loss"
>
> "A hazard or dangerous chance"
>
> "The hazard or chance of loss"
>
> "The degree of probability of such loss"
>
> "The amount which the insurance company may lose"

The ambiguity surrounding the notion of risk should not surprise us. Indeed, the definition and quantification of risk is neither simple nor straightforward. On the other hand, risky and riskless positions and risky and riskless assets are far less problematic. Let us, therefore, start by describing risky and riskless positions and then discuss a few risk indexes suggested in the literature. We will start with the definition of riskless position. A riskless position is a situation in which a given financial outcome will be realized with certainty, that is with a probability of 1. For example, if we buy U.S. Treasury bills which mature in one year at \$1,000, the market price today is \$950, and no further coupon is paid, we have a riskless financial position and the rate of return is certain and equal to $\frac{\$1,000}{\$950}-1\cong 5.26\%$. Why is this financial position riskless? It is riskless because the United States government cannot go bankrupt. Even if the government has no cash to pay the \$1,000 per bond at the end of the year, it can raise more taxes or even print money and use this money to pay its bondholders; hence, the \$1,000

---

[1] See Webster's Encyclopedic unabridged dictionary, Gramercy Books, New York, 1989.

payment per bond is guaranteed. (We assume zero probability of political revolution and a new regime failing to honor the government debt).

Thus, if the holding period is one year, the investor who buys U.S. short-term government bonds are in a riskless position;[2] there is only one value, a rate of return of 5.26% (or a cash flow of $1,000) which is obtained with a probability of 1. In such a scenario, we say that the investor has a riskless position or that he/she is investing in a riskless asset. Indeed, short-term government bonds are commonly used as a proxy for the riskless asset. We emphasize that it is a proxy for riskless assets: It does not exactly correspond to the riskless asset because there is always the possibility of an increase in the interest rate which may induce a capital loss. However, any such loss is generally very small in the case of short-term bonds.

Formally, the definition of riskless asset is that the future outcome (return, rate of return, or cash flow) has only one value x where, in our example, x = $1,000 or a 5.26% rate of return, and $p(x) = 1$.

From this definition, it is easy to deduce the definition of risky position or risky asset. A risky position is a situation in which there is more than one financial outcome, say $x_1, x_2, \dots, x_n$ and, for at least one value x_i, $0 < p(x_i) < 1$, where p denotes a probability of x_i occurring. Note that if there is one value such that $0 < p(x_i) < 1$, there must be at least one more observation, x_j, with $0 < p(x_j) < 1$. The total probability must be equal to 1; $\Sigma\, p_i = 1$. By this definition, the future value of a risky asset may have more than one value x_i, with $0 < p(x_i) < 1$. For example, IBM stock held for a year represents a risky position because many future values for the stock price (hence for the rates of return), are possible.

Let us denote the future possible monetary outcomes (or rates of return) and their corresponding probabilities by the pair (x,p(x)).[3] Each investment will be characterized by such a pair. If there is only one value x with $p(x) = 1$, it will be a riskless asset. If there is more than one value x with $0 < p(x) < 1$ for all values x, then it will be a risky asset or risky investment.

Frank Knight distinguished between *risk* and *uncertainty*.[4] Risk is defined by Frank Knight as a pair of values (x, p(x)) (with at least one value x_i for which ($0 < p(x_i) < 1$)) such that both x and p(x) are *known*. Uncertainty is a pair (x,p(x)) such that the possible values of x are known but p(x) is *unknown*. For example, if you

---

[2] For simplicity, we assume no inflation. If inflation does exist, riskless position is defined as investment in government bonds linked to the consumer price index. Such an investment guarantees a riskless position in real terms.

[3] If the distribution of returns is continuous, then the density function f(x) characterizes the investment rather than the probability p(x).

[4] Frank Knight, *Risk, Uncertainty and Profit*, Boston and New York, Houghton Mifflin Company, 1921.

roll a balanced die and the prize is equal to the number shown on top in thousand dollars and you pay $4,000 to play this game, you have the following risk:

An Illustration of Risk

Outcome of the die (x):	**1**	**2**	**3**	**4**	**5**	**6**
Cash flow (x • $1,000–$4,000)	–$3000	–$2,000	–$1,000	0	+$1,000	+$2,000
Probability	1/6	1/6	1/6	1/6	1/6	1/6

If, on the other hand, you are not sure that the die is balanced (maybe there is a layer of lead under the face of number 1!), then we have *uncertainty* – the possible outcomes are known but the probabilities p_i, i = 1, 2, …, 6 are unknown.

An Illustration of Uncertainty

Outcome of the die:	**1**	**2**	**3**	**4**	**5**	**6**
Cash flow (x • $1,000–$4,000)	–$3000	–$2,000	–$1,000	0	+$1,000	+$2,000
Probability	p_1	p_2	p_3	p_4	p_5	P_6

In practice, outside the casino or laboratory and except for some national lotteries, the probabilities are unknown. When we invest in the stock of General Motors , Xerox, or any other firm, we have no option but to estimate both x and p(x), subjectively. Thus, whenever p(x) is unknown, we simply substitute it with subjective probabilities, shifting from an uncertain situation to a (subjective) risky situation. Therefore, throughout this book, we use the words risk and uncertainty, interchangeably.

So far, we have defined risky and riskless positions and risky and riskless assets. Let us now turn to the quantification of risk. Can risk be quantified? Is it possible to create an index of risk? Can financial positions or assets be valued by their risk? As we shall see, it is far more difficult to quantify risk than to define risky situation or risky asset. Take the example of the balanced die. As there are six possible results (i.e., more than one) with a probability of $0 < p(x) = 1/6 < 1$ for each possible value, we have a risky position. Can you tell what the risk is? Is it –$3,000 , –$2,000, or –$1,000 or is it maybe the average loss? To show how difficult it is to quantify risk and to compare the relative risk of various investments, assume that there is another investment with an equal chance of either a $3,000 gain or a loss of $2,000. Which cash flow is more risky, the one corresponding to the die or this one? It is very hard to tell. Indeed, it is difficult to measure the risk involved in risky positions and to rank investments by their risk.

We turn next to several alternative suggestions which have appeared in the financial and economic literature on how to measure risk.

1.2 MEASURES OF RISK

a) Domar and Musgrave risk indexes

As we noted in Section 1.1, risk is sometimes defined in terms of loss. In line with this intuition, Domar and Musgrave (D&M) formulated a quantitative index of risk that takes into account all possible negative or relatively low outcomes.[5] They state:

> "Of all possible questions which the investor may ask, the most important one , it appears to us, is concerned with the possibility of actual yield being less than zero, that is with a probability of loss. This is the essence of risk."

Accordingly, they proposed the following risk index (RI):

$$\mathbf{RI}=-\sum_{x_i\le 0}p_i x_i \tag{1.1}$$

Note that because $x_i \le 0$, RI is a positive number; thus, the higher RI, the more risky the investment. In this case, RI is the truncated mean and x=0 is the truncation point. Namely, we calculate the mean return of the negative numbers, only.

If the random variable is continuous, the risk index (RI) will be:

$$\mathrm{RI}=-\int_{-\infty}^{0} f(x)x\,dx \tag{1.1'}$$

All the risk indexes given in this chapter are defined for \$1 of investment; hence x is the rate of return in percent and the risk index is given in percent, too. However, if an investment of \$I rather than \$1 is involved, and you want to measure the risk corresponding to investment of \$I , simply multiply x by I to obtain dollar amount figures. The risk index measured in dollars will then correspond to the \$I investment.

Realizing that many investors feel they have failed in their investment if they earn less than the riskless interest rate, D&M suggested the following modified version of their risk index:

$$\mathrm{RI}=-\sum_{x_i\le r}p_i(x_i)(x_i-r) \tag{1.2}$$

---

[5] E. Domar and R.A. Musgrave, "Proportional income taxation and risk taking," *Quarterly Journal of Economics*, LVII, May, 1944.

Here, all the deviations $(x_i - r)$ (for $x_i < r$ only) are multiplied by the probability p_i to obtain x_i. For the continuous random variable, RI is

$$RI = \int_{-\infty}^{r} f(x)(x-r)dx \qquad (1.2')$$

where r is the riskless interest rate.

Example:

Suppose that we have the following investment:

x:	−50%	−10%	5%	50%	100%
p(x)	1/5	1/5	1/5	1/5	1/5

and r = 6%

Then, by eq. (1.1), the risk index will be:

$$RI = -[\,1/5\,(-50\%) + 1/5\,(-10\%)] = +\,12\%,$$

and by eq. (1.2), the risk index will be,

$$\begin{aligned} RI &= -[\,1/5\,(-50\% - 6\%) + 1/5\,(-10\% - 6\%) + 1/5\,(5\% - 6\%)] \\ &= [1/5\,(-56\%) + 1/5\,(-16\%) + 1/5\,(-1\%)] = 14.6\%. \end{aligned}$$

Finally, if \$10,000 is invested, then the risks involved in dollar terms (rather than percentage terms) will be \$1,200 and \$1,460, respectively. To obtain this result, simply multiply x_i and r in eq. (1.1) and eq. (1.2) by \$10,000.

Thus, using either eq. (1.1) or eq. (1.2) it is possible to calculate a risk index for all alternative investments and then to rank them by their risk. Note that the higher the riskless interest rate, the higher the risk given by eq.(1.2).

D&M's measures of risk are very appealing . Indeed, they conform with our intuition. However, they do contain some drawbacks and not all investors would agree with the resultant risk ranking that they produce. To illustrate this, consider the following two investments with only one negative outcome for each: (the positive outcomes are irrelevant for the risk index given by eq. (1.1), hence not presented here).

	Rates of Return	Probability
Investment A	–50%	0.1
Investment B	–10%	0.5

Using the risk index given by eq. (1.1), both investments have the same risk because RI = 0.1 (-50%) = 0.5 (-10%) = 5%.

Would investors consider the above two investments as equal in terms of their risk? Probably not. For many investors a 10% loss might not spell disaster but a 50% loss might mean bankruptcy and total catastrophe. Therefore, some would consider the –50% rate of return to be much more risky. Thus, the main disadvantage of the D&M indexes of risk is that they do not take into account the differential damage of the various negative monetary returns.

b) Roy's Safety First Rule

According to A.D. Roy, investors are mainly concerned with avoiding the possibility of "disaster". Based on this premise, he proposed his principle of "Safety First" as a guideline in selecting the investments.[6] He rejects the modern utility theory (discussed in detail in Chapter 2) asserting that:

> " A man who seeks advice about his actions will not be grateful for the suggestion that he maximize expected utility."

Focusing on the investment's safety, Roy proposed that risk is measured in terms of the probability that the future income will be lower than d, where d is the disaster level as perceived by the investor. Roy's risk index (RI) is defined as follows:

$$RI = p(x \leq d) \qquad (1.3)$$

where p stands for probability. To elaborate on this index we need first to define the mean and variance of the distribution of outcomes.

[6] See A.D. Roy, "Safety First and the Holding of Assets" *Econometrica*, July, 1952.

The distribution mean, μ, and its variance σ^2 are given by:

$$\mu = \int_{-\infty}^{\infty} x f(x) dx ;$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

where x denotes future income (a random variable) with mean μ and standard deviation σ. Using Chebycheff's inequality, we have:

$$p\{|x-\mu| > k\sigma\} \leq 1/k^2$$

where $k \geq 0$. Select $k\sigma = (\mu - d)$ to obtain:

$$p\{|x-\mu| > (\mu-d)\} \leq \frac{\sigma^2}{(\mu-d)^2}$$

and, *a fortiori:*

$$p\{(\mu-x) \geq \mu-d\} = P(x \leq d) \leq \frac{\sigma^2}{(\mu-d)^2}$$

where μ is the expected value and σ is the standard deviation (σ is the square root of σ^2) of the distribution of returns as defined above.

Thus, $\sigma^2/(\mu\text{-}d)^2$ is the upper bound of the probability of a disaster and $p\ (x \leq d)$ is the risk index. If the whole distribution is known, this probability can be calculated precisely and a risk index can be assigned to each investment. However, if only σ, μ are known (rather than the whole distribution), Roy suggests selecting the investment that minimizes $\sigma^2/(\mu\text{-}d)^2$, which, in turn, will minimize the upper bound of the probability of disaster as estimated by Chebycheff's inequality.

Example:

We show in this example how investments are ranked by Roy's risk index and the pitfalls of this risk index.

Suppose that you have the following two investments:

Investment A		Investment B	
Rate of Return (%)	**Probability**	**Rate of Return (%)**	**Probability**
–50	1/100	–1	2/100
5	9/100	5	3/100
30	90/100	40	95/100

Suppose that d=0 (i.e., a rate of return below zero is considered to be disastrous). Then, using Roy's index, the risk of A = 1/100 < the risk of B = 2/100.

Hence, investment B is more risky. Yet, most investors would probably rank investment A as more risky because of the possibility of a 50% loss. Thus, Roy's risk index takes into account the probability of an outcome below d but not the size of the loss. Secondly, Roy's risk index is subjective. If for one investor, d = 0, investment B will be riskier than investment A. If, for another investor, any outcome below the market interest rate, which is assumed to be r = 6% , would be considered as a disaster, the risk of A will be 10% and the risk of B will be 5%. Thus, with d=6% investment A will be riskier than investment B. Because d is determined by the individual investor, the risk of investments cannot be ranked objectively.

c) Dispersion as a Risk Index: Variance and Standard Deviation

Because risk occurs when there is more than one possible outcome, it would seem natural to measure it by one of the common dispersion measures, such as, the range of outcomes (i.e., maximum return minus minimum return). However, the most common measure of risk is the variance (σ^2) or standard deviation (σ) of the distribution of returns which is given by:

$$\sigma_x^2 = \sum P(x_i)(x_i - Ex)^2$$

for a discrete distribution and,

$$\sigma_x^2 = \int f(x)(x - Ex)^2 \, dx$$

for a continuous distribution (as defined above see Ex = μ). The square root of σ^2 gives the standard deviation , σ.

Investors are interested in the investment's profitability as best estimated by the expected value of the returns. The standard deviation indicates possible deviations of the realized returns from their expected value, hence a high standard deviation is intuitively identified with high risk. Because of its simplicity and intuitive grasp as a risk measure, this index of risk is widely accepted among professional investors as well as academics. Based on this risk index, Markowitz (1952) developed the mean-variance analysis which was the base for the Capital Asset Pricing Model developed by Sharpe (1964) and Lintner (1965). [7, 8, 9] (Both Markowitz and Sharpe won the 1990 Nobel Prize in Economics mainly for these two important contributions).

Although this risk measure is widely accepted, it too has its drawbacks. The main objection is that it takes both the "good" and the "bad" deviations from the mean into account: However, only the "bad" deviations (to the left of the mean) imply losses whereas the "good" deviations (to the right of the mean) imply gains . To see this drawback of σ^2 as a measure of risk, consider the following two investments:

	Investment A		**Investment B**	
	x (%)	p (x)	x (%)	p (x)
	–1	1/3	–1	2/6
	2	1/3	2	3/6
	5	1/3	8	1/6
Mean (E)	2		2	
Variance (σ^2)	6		9	

where

$$\sigma_A^2 = 1/3(-1-2)^2 + 1/3(2-2)^2 + 1/3(5-2)^2 = 6$$
$$\sigma_B^2 = 2/6(-1-2)^2 + 3/6(2-2)^2 + 1/6(8-2)^2 = 9$$

Hence, $\sigma_B^2 > \sigma_A^2$

[7] Markowitz, H.M., "Portfolio Selection," *Journal of Finance*, 7(1952), 77-91.
[8] Sharpe, William F., "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance,* September 1964, pp. 428-442.
[9] Lintner, J., "Security Prices, Risk and Maximal Gains from Diversification," *Journal of Finance,* 20(1965), 587-616.

Both investment A and investment B have a mean return of 2% and the probability of the risk of a relatively low return of −1% is identical in both investments. However, by the variance (or standard deviation), B is riskier than A. What induces σ_B^2 to be larger than σ_A^2 ? It is the +8% deviation to the right of the mean – a windfall by all accounts ! This example shows that the variance, which takes into account both positive and negative deviations from the mean, may be misleading. It shows investment B to be riskier due to its high possible return of +8%. Hence, investment B seems to be penalized due to its most attractive feature – its possible high rate of return. In order to overcome this drawback of the variance measure an index based on semi-variance was introduced.[10] The Semi-Variance (SV) index takes into account only deviations to the left of certain critical values, which are generally selected to be equal to the mean. This index is defined next.

d) Semi-Variance (SV) as an Index of Risk.

Semi-variance as a measure of risk is formally defined as:

$$SV = \sum_{x_i \le A} P(x_i)(x_i - A)^2 \text{ for discrete distributions}$$

and

$$SV = \int_{-\infty}^{A} f(x)(x-A)^2\,dx \text{ for continuous distributions} \qquad (1.4)$$

where A is some constant such that earning less than A would be considered as a failure. Generally, A is selected to be equal to E(x), hence the name semi-variance; it takes only the negative deviations from the mean into account. However, this method also has its critics: Not all investors would agree that it ranks risks properly and, in addition, not all would agree with the selected return level A. We demonstrate in the following example:

[10] The semi-variance has been suggested by Markowitz, see H.M. Markowitz, *Portfolio Selection*, New York, Wiley, 1959.

	Investment A		Investment B	
	x	p(x)	x	p(x)
	1	1/5	−1	1/20
	2	1/5	3	18/20
	3	1/5	7	1/20
	4	1/5		
	5	1/5		
Mean	3		3	
SV	1		4/5	

The semi-variance (with A=E) is calculated as follows:

The semi-variance (with A=E) is calculated as follows:

$$SV_A = 1/5\,(1\text{-}3)^2 + 1/5\,(2\text{-}3)^2 = 5/5 = 1$$

$$SV_B = 1/20\,(\text{-}1\text{-}3)^2 = 16/20 = 4/5$$

Hence, by the semi-variance risk index, investment A is riskier than investment B. However, suppose that a negative income implies bankruptcy. In such a case, many investors, contrary to the semi-variance ranking, would consider investment B to be riskier because it is possible to obtain a negative income in investment B, but not in investment A.

Thus, semi-variance overcomes some of the difficulties of variance as a measure of risk, but it does not provide a universally acceptable, unequivocal, objective measure of risk.

e) Baumol's Risk Measure

William Baumol agrees that variability is a source of risk.[11] However, he claims that the risk is due to the possibility of earning less than some critical level, or "floor". Baumol argued that the standard deviation *per se* is not a good measure of risk:

> "an investment with relatively high standard deviation (σ) will be relatively safe if its expected value (E) is sufficiently high."

Let us illustrate Baumol's approach with an example. Consider the following two investments:

	Investment A	**Investment B**
E	2	20
σ	1	2

According to the variance risk index, investment B would be ranked as riskier because the standard deviation of investment B is *larger* than that of investment A. However, as the probability of an income below E-kσ is bounded by $1/k^2$ (by Chebycheff's inequality, see Roy's risk index), Baumol claims that investments such as B are most likely to end up with a higher realized return than investments such as A; hence investment B is safer rather than riskier than investment A. To illustrate this claim, suppose that the return on investment B deviates, say 5 standard deviations to the *left* of the mean (a very pessimistic outcome), and the rate of return on investment A deviates, say 5 standard deviation to the *right* of the mean (a very optimistic outcome). Even in this extreme case in favor of investment A, we will still have:

Rate of Return of A $= 2 + 5 \cdot 1 = 7 <$ Rate of Return of B $= 20 - 5 \cdot 2 = 10$.

In this example, the realized return on investment A is lower than the realized return on investment B, hence investment A is riskier than investment B!
Note that by Chebycheff's inequality, the probability of deviating k= 5 standard deviations to the left is smaller than $1/k^2 = (1/5)^2 = 1/25 = 4\%$. Therefore, Baumol claims that investments such as B are probably safer than investments such as A.

[11] See W.J. Baumol, "An Expected Gain in Confidence Limit Criterion for Portfolio Selection," *Management Science*, October 1963, 10, pp. 174-182.

He therefore proposes the following risk index:

$$RI = E - k\sigma \tag{1.5}$$

where k is some constant selected by the investor representing his/her safety requirement such that the return is unlikely to fall below it. Accordingly, the higher the floor E-kσ (or lower bound), the safer the investment. In the above example, if k=3, we have:

$$RI_A = 2 - (3 \cdot 1) = -1$$

$$RI_B = 20 - (3 \cdot 2) = 14$$

and, because 14% > -1% , by Baumol's criterion, investment B is safer than investment A (or investment A is riskier than investment B).

Note that a return below E-kσ is also possible but, because the probability of such an event is relatively small (less than $1/k^2$), it is ignored. However, lack of consensus is likely to prevail regarding the negligibility of a given probability; that is, not all investors would agree on the selected value of k.

Thus, although Baumol's risk (or rather safety) index is intuitively appealing, it has two drawbacks: it ignores the probability (no matter how small it is) of a return falling below the floor, and it is subjective: k=2 standard deviations from the mean may be appropriate for one investor but for another more risk averse investor, k=5 standard deviations are required. Thus, the ranking of an investment risk may vary due to variations in the value assigned to k.

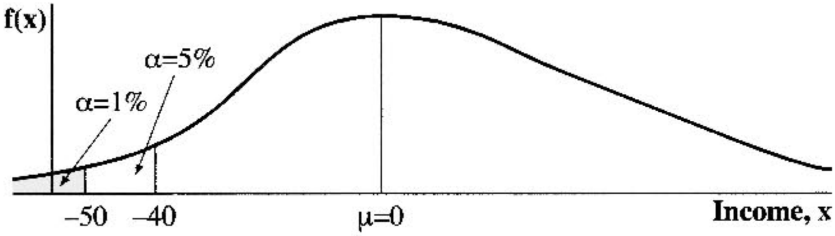
f) Value at risk-VaR(α)[12]

Recently a very common risk index used by practitioners and in particular by financial institutions, e.g., banks, is called VaR(α) – which stands for Value at Risk. VaR(α), indicates the maximum possible loss, when α percent of the left tail distribution is ignored. In a way, the idea of VaR is related to the risk measures suggested by Baumol and Roy. In particular, by Baumol's rule, the extreme left tail of the distribution is ignored, i.e., returns falling in a given left range are considered very unlikely. For example, suppose that the bank's profit on its investment in assets traded in the market is given by Figure 1.1.

---

[12]For an excellent discussion and analyses of VaR and other risk measures, see Philippe Jorion , Value at Risk, McGraw-Hill, New York 1997.

Figure 1.1: The VaR(α) as a risk measure



In this case the VaR ($\alpha = 1\%$) = –50, i.e., when the left tail corresponding to α = 1% is ignored, the maximum loss is –50. The regulator who is in charge of the stability of the banking system may require the bank to hold safe assets of about n x \$50 million where n = 3–4. The VaR can be used by the bank's management also for internal risk management. Finally, one can choose another level of α: For example, VaR ($\alpha = 5\%$) = –40. In the general case,

$$\text{VaR}(\alpha) = \mu - L \tag{1.6}$$

where μ is the mean of the distribution and L is the value such that $\Pr(x \leq L) = \alpha$. Thus, the risk is measured as the maximum deviation from the mean when the left tail of the distribution is ignored. For short time intervals (e.g., one day) $\mu \cong 0$, hence the risk is measured as deviation from zero as done in the above graphical example.

g) Shortfall VaR

This is another risk index with a focus on the left tail of the distribution of returns. It is the expected loss when the expectation is calculated only over the left tail domain.

Generally, for a continuous distribution the α shortfall risk index is given by

$$SF(\alpha) = \int_{-\infty}^{x_0} x f(x) dx \quad \text{where} \quad \int_{-\infty}^{x_0} f(x) dx = \alpha \tag{1.7}$$

h) Loss as an Alternative Cost: The Minimax Regret

Leonard Savage proposed the minimax regret criterion for selecting among risky actions or risky investments. [13] The main thrust of this rule is that investors should choose the investment that offers the minimum risk of possible losses due to a

[13] See Leonard Savage, "The Theory of Statistical Decision," *Journal of American Statistical Association*, 46, 1951, 55-67.

wrong choice; hence the regret measures the risk of making a wrong investment choice. According to this rule, losses are due to the alternative costs, or wrong investment choices. To illustrate, suppose that there are three stocks with the following rates of return:

Rates of Return (in percent)

	State of Nature		
Stock	**State 1 (S_1): Inflation**	**State 2 (S_2): Stagflation**	**State 3(S_3): Recession**
1	50	30	1
2	4	15	10
3	8	6	7

The *minimax regret* criterion determines that the investor calculates the maximum possible regret for each stock and that the stock with the minimum of these maximum regrets be chosen. The stock with the minimax regret is the one with the lowest risk. Suppose that the investor selects stock i and state of nature S_j occurs. If stock i gives the maximum rate of return in state S_j, there will be no regret; the investor will have chosen wisely. If, however, other stock promises the maximum rate of return in this state, the investor will *not* have chosen wisely; his regret is measured by the difference between the maximum rate of return in state S_j and the rate of return realized by having chosen stock i. Thus, the risk is due to selecting stock i when more could be earned by selecting stock j. Regret measures the risk of losses (alternative costs) due to a wrong stock selection. A regret payoff table looks like this:

Regret Table (in %)

	State of Nature		
Stock	**S_1: Inflation**	**S_2: Stagflation**	**S_3: Recession**
1	0	0	9
2	46	15	0
3	42	24	3

For example, if S_1 occurs and the investor selects stock 2, the actual return will be 4%. If he/she had selected wisely (stock 1), 50% could have been earned; hence the regret (or alternative loss) involved with selecting stock 1 will be 46%. With this method, 46% is considered as a loss because this is the additional rate of return that could have been earned if the investor had selected the stock wisely. All values in the regret table are calculated in a similar way.

Savage proposed that the maximum regret of each stock be compared and that the stock with a minimum of such maximums be selected; hence the name minimax regret. In our example, stock 1 would be selected because 9% is lower than 42% and lower than 46%. The minimax regret measures the risk of choosing the wrong stock; it minimizes the risk measured by the maximum possible (alternative) loss from choosing the wrong stock.

Thus, the risk of each stock is measured by the return relative to the rate of return on other alternative investments and not by its own returns. Although the notion of alternative costs is intuitively very appealing, this measure of risk has two major drawbacks. First, adding one more stock may change the relative risk of the stock itself even if the additional stock is irrelevant because it is not chosen. For instance, in our example, the addition of a fourth stock (which the investor will not select) may change the risk ranking. To illustrate, if we add a fourth stock yielding 0% at S_1, 10% at S_2 and 50% at S_3, the regret loss function will be:

	Rates of Return (in %)			**Regret Table (in %)**		
	State of Nature			**State of Nature**		
Stock	S_1	S_2	S_3	S_1	S_2	S_3
1	50	30	1	0	0	49
2	4	15	10	46	15	40
3	8	6	7	42	24	43
4	0	10	50	50	20	0

According to the minimax regret rule, stock 3 is now the least risky, whereas before we introduced the fourth stock, stock 1 was the least risky. Thus, adding a fourth stock which is not selected by the minimax regret criterion, changes the risk ranking of the other stocks. This is called the "irrelevant alternative" effect: The introduction of a fourth stock changes the ranking of our decision even though the fourth stock is irrelevant because it is not selected.

The second major drawback of the minimax regret is that it does not take the probability of the various states into account. Let us go back to our original example (before the fourth stock was introduced) and assume that the probability of S_1, S_2, and S_3 is 1%, 1%, and 98%, respectively. Thus, by selecting stock 1, we have a 98% probability of earning a very low income of 1%. However, by selecting stock 2, we have only a 1% probability of earning 4% (which is higher than the 1% return of stock 1) and a 98% probability of earning 10% or more. In short, the regret function measures risk due to wrong choice but it does not take into account the probability of the various states of nature and, therefore, it does not fully measure the risk of each stock. There would have to be an elaboration on a rule which takes the probabilities of the various states of nature into account. However, this would be a new rule and not the minimum regret criterion.

1.3 SUMMARY

In general, investors do not like risk. However, there is wide lack of agreement as to the index that should be employed to measure risk .

We have discussed a number of measures of risk. Each has its pros and cons. None are free of serious drawbacks. There have been other attempts at quantifying risk but the above examples suffice to demonstrate that risk is very hard to quantify. Indeed, there is little consensus on the acceptability of any given risk index. The acceptability of a risk index tends to be a subjective matter, hence notions of personal utility function and risk preference have to be introduced and incorporated in the measurement of risk.

The most natural risk index is the maximum amount of money that a person is willing to pay to an insurance firm to offset a given risk. The (average) amount paid to transfer the risk to another party is called the *risk premium.* However, different investors have different notions as to the magnitude of the risk premiums that they would be willing to pay to rid themselves of a given risk. We therefore have to draw the unavoidable conclusion that, in the most general case, risk cannot be objectively quantified. As revealed in the above discussion, there is no one objective measure of risk. One investor may rank investment A as riskier than investment B and another investor may rank investment B as riskier than investment A. Indeed, only under certain specific conditions (e.g., normal distributions, lognormal distributions) can risk be quantified such that all investors within a given class (based on their utility function and risk preference) will agree on the risk ranking of a given range of investments (as will be discussed later on in the book).

In order to measure the risk premium, we first need to introduce the concepts of utility function and expected utility. These will be discussed in Chapter 2.

KEY TERMS

Risk

Risk Index

Uncertainty

Semi-Variance (SV) Index

Baumol's Risk Measure

Roy's Risk Index

The Minimax Regret Criterion

Risk Premium

Value at Risk (VaR)

Shortfall VaR

2

EXPECTED UTILITY THEORY

2.1 INTRODUCTION

Let us now turn to what, in the eyes of the investor, is probably the main *raison d'être* of investment, namely, profitability. In focusing on risk in our first chapter, by no means do we belittle this all-important function of investment. Our discussion of risk simply serves to emphasize that, in arriving at an investment decision, the risk of the investment has to be weighed against its profitability. Thus, both profitability and risk have to be incorporated in the decision making process. We devote this chapter to the expected utility criterion that takes into account the whole distribution of returns (risk and return).

Investors face many alternative investment choices . In order to compare the risk and return of alternative investments decision criteria are needed. Most of this book is devoted to investment decision rules that rely on the expected utility paradigm. The expected utility framework does not analyze risk and return separately; it considers the whole distribution of returns simultaneously. Moreover, in this framework, there is no need to define risk.

This chapter deals with the foundations of expected utility theory. We will first discuss a number of investment criteria and then we will analyze how these decision criteria are related to the expected utility framework.

2.2 INVESTMENT CRITERIA

a) The Maximum Return Criterion (MRC).

The Maximum Return Criterion (MRC) is employed when there is no risk at all. According to this rule, we simply choose the investment with the highest rate of return. By making the right choice, we ensure maximum return on the invested wealth at the end of the investment period. Textbooks on economics or price theory are replete with models aimed at maximizing profits, or maximizing return. Let us illustrate the MRC with an example of an optimal production decision by a firm. Let P be the price of the product per unit, Q the quantity of units produced by the firm, and C(Q) the production costs. The firm's objective is to decide on the optimal quantity, Q^*, to be produced such that the profit, $\pi(Q)$, is maximized.

Thus, the objective of the firm is:

$$\text{Max } \pi(Q) = P \cdot Q - C(Q).$$

Taking the first derivative and equating it to zero, we obtain the well-known result that at the optimum production level, the following will hold:

$$P = C'(Q^*)$$

which implies that the marginal revenue, P, will be equal to the marginal cost $C'(Q)$. The value Q^* is the optimal number of units to be produced because, by selecting Q^*, the firm maximizes its return $\pi(Q)$.

Can we apply this MRC rule to selection among uncertain investments and, in particular, to selection of a portfolio of securities that have uncertain returns? As we shall see, this rule is applicable only when the returns are certain (as in the case of selecting the optimum production); it is not applicable in the case of uncertain returns. Indeed, when MRC is recommended in most economics textbooks, it is assumed (implicitly or explicitly) that the price of the product P and the costs C(Q) are certain.

To demonstrate that MRC is applicable only in the case of certain cash flows, let us first explain what we mean by an applicable decision rule. A decision rule is said to be applicable if it can be employed in a non-arbitrary manner. It is not applicable if it can be employed in more than one way. For instance, it is not applicable if investment A is shown to be better than investment B when the rule is used in one way, and an opposite ranking is obtained if it is employed in a different way. Let us explain this notion via a numerical example. Suppose that you want to rank the following four investments in order to arrive at an investment decision:

Investment A		Investment B		Investment C		Investment D	
x	p(x)	x	p(x)	X	p(x)	x	p(x)
+4	1	+5	1	−5	1/4	−10	1/5
				0	1/2	+10	1/5
				+40	1/4	+20	2/5
						+30	1/5

where x is the return (in $s or percentages) and p(x) is the probability of obtaining x. The MRC rule tells us that investment B dominates investment A because it has a higher return. However, it is ambiguous regarding the other pairs of investments and, therefore, it is not applicable to these investments. For example, if we pick the −5 return of investment C, then investment B is better than investment C. However, if we compare the +40 return of investment C with investment B, the opposite ranking is obtained. With uncertain investments, we do not obtain by the MRC a clear-cut unique ranking of investments because the ranking is a function of the arbitrary pairs of returns chosen for comparison. Therefore, MRC is not applicable in the case of uncertainty. It is not a "bad" or a "good" rule for uncertain situations, it is simply not applicable. A modified version of the MRC whereby the highest possible of all returns for each investment is identified and the investment with the highest maximum is then selected, technically helps overcome this problem to some extent. In this case, the rule is applicable and, in our example, investment C with the highest return of 40% is selected. The modified MRC is applicable to uncertainty because its result is not a function of the way it is employed. However, it can be misleading . For instance, let us reduce the probability of the +40% return of investment C to 1/1000, increase the probability of the −5% return to, say, 999/1000, and reduce the probability of x=0 to zero. By the modified MRC, investment C is still the most desirable investment. This is an obvious drawback because very few investors would consider investment C with these new probabilities to be the best investment.

Finally, it should be emphasized that when MRC is employed in finance and economics (especially in price theory), it is assumed that certainty prevails, that is, that there is only one possible return on the investment under consideration. In such cases, the MRC has no drawbacks and it is applicable. However, the certainty assumption regarding future returns is very unrealistic. We, therefore, need to search for other investment criteria.

b) The Maximum Expected Return Criterion (MERC)

The Maximum Expected Return Criterion (MERC) identifies the investment with the highest *expected* return and thereby overcomes the problem of non-unique ranking.

To employ this rule we first calculate the expected return of each possible investment. For investment A, it is 4, for B it is 5, and for C and D it is as follows:

$$E_c(x) = 1/4\,(-5) + 1/2\,(0) + 1/4\,(40) = 8.75;$$

$$E_d(x) = 1/5\,(-10) + 1/5\,(10) + 2/5\,(20) + 1/5\,(30) = 14.$$

Thus, the MERC provides a clear and an unambiguous ranking: In our example, investment D has the highest expected return. Thus, by MERC, investment D is ranked as the best investment.

The fact that the MERC provides an unambiguous ranking of risky investments does not imply that this rule should be employed in all instances. We are merely stating that, technically, the MERC is applicable to certainty and to uncertainty: Its theoretical justification has yet to be shown; hence, it is not necessarily the optimal rule. Actually, as we shall see below, this rule is not optimal and may lead to paradoxes (or irrational decisions) such as the famous St. Petersburg Paradox.

The St. Petersburg Paradox

The St. Petersburg Paradox first came to light in the 18th century and its solution paved the way to modern utility theory.

To illustrate this paradox, consider a game which requires a coin to be tossed until the first head shows up. The prize is $\$2^{x-1}$ where x is the number of tosses until the first head shows up. The game is over when the first head shows up. Theoretically, the game can be infinite. How much would you pay to participate in such a game? Or, specifically, what certain amount would you be willing to accept to be indifferent between playing the game for free or receiving this certain sum? This certain amount is called the *certainty equivalent* of the game. Note that the game can be seen as a risky investment. For example, if you pay, say \$100 for the game and the first head shows up after the first toss, you win in the game 2^{1-1} = \$1 and you lose \$99. How much would you be ready to pay for this risky investment? Experiments with this question reveal that most subjects report a very small certainty equivalent amount (\$2-\$3 in most cases). However, by the MERC, the certainty equivalent of this game is infinite; hence the paradox: Investors are ready to pay only a very small amount for an investment whose expected value is infinite. To see this, let us calculate the expected prize of this game. It is:

$$\sum_{x=1}^{\infty} \frac{1}{2^x} 2^{x-1} = \infty$$

where $\frac{1}{2^x}$ is the probability of an event $\underbrace{T,T,T,...T}_{x-1}H$ which denotes the first head showing up on the x^{th} toss (T repeats x-1 times and then H occurs, x = 1, to 2,...where T stands for "tail shows" and H for "head shows up"). For example, if T appears 3 times in a row and then H appears, we obtain $2^{4-1} = 2^3$ =\$8 and the probability of such an event is $(1/2)^4 = 1/16$. As x can take on any number, we have a summation from x=1 up to infinity.

This paradox reveals the drawback of the MERC: it can lead to results that would be unacceptable to most investors because no investor would require an infinite amount, nor even a large amount, as the certainty equivalent. Indeed, way back in the 18th century, Nikolaus Bernoulli and Gabriel Cramer[1] suggested that investors, in making their decisions, aim at maximizing the expected utility of money like E(log (w)) or $E(w^{1/2})$ where log (w) or $w^{1/2}$ are possible utility functions, w stands for wealth, and E stands for expected value. Thus, according to Bernoulli and Cramer, what is
important to investors is the utility derived from the money received rather than the money itself. With a log(w) function, for example, we have log (10) =1, log(100) = 2 etc. Hence, the utility derived from the first \$10 is equivalent to the utility derived from the next \$90, showing a decreasing marginal utility of money (more details on the meaning of utility function will be provided later on in the chapter.)

---

[1] For more details on the solution of Bernoulli and Cramer, see H.L. Levy and M. Sarnat, *Portfolio and Investment Selection: Theory and Practice*, Prentice Hall International, 1984.

Indeed, by calculating the expected utility, these two utility functions produce a reasonable solution. With the log (w) function (substituting the prize 2^{x-1}, for w) we obtain: [2]

$$E(\log w)=\sum_{x=1}^{\infty}\frac{1}{2^x}\log 2^{x-1}=\log 2\sum_{x=1}^{\infty}\frac{x-1}{2^x}=\log(2).$$

Thus, w=2 is the *certainty equivalent.* [3]

(Note that log (2) can also be considered to be the expected utility of the certainty equivalent because $2 is received with a probability of 1.) In other words, the investor will be indifferent between receiving $2 for sure or playing the St. Petersburg game because $2 also yields the utility of log (2). If you offer the investor a higher sum, say, $3 for sure or, alternatively, the chance to play the game for free by the expected utility criterion , the investor should choose the $3 for sure because log (3) > E(log(w)) = log 2.

Similarly, with the function $U(w) = w^{1/2}$ suggested by Cramer, we obtain:

$$E(w^{\frac{1}{2}})=\sum_{x=1}^{\infty}\frac{1}{2^x}(2^{x-1})^{1/2}=1/2+\sqrt{2}/4+\sqrt{(2)^2}/8+\sqrt{(2)^3}/16+\cdots=1/2\cdot 1/(1-\sqrt{2}/2)$$

$$=1/(2-\sqrt{2})\cong\$1.707$$

hence, the investor will be indifferent between receiving $(\frac{1}{2-\sqrt{2}})^2=\$2.914$ for sure or, alternatively, playing the St. Petersburg game for free, because both alternatives yield the same expected utility of $U(w) = w^{1/2} = (2.914)^{1/2} =$ \$1.707.

The St. Petersburg paradox demonstrates why the MERC may be misleading and unacceptable. We have also seen above that by assuming that investors make

---

[2] In the calculation of the expected utility we employ the following:

$$\sum_{x=1}^{\infty}\frac{x-1}{2^x}=1/4+2/8+3/16\cdots=(1/4+1/8+1/16\cdots)+(1/8+1/16+1/32\cdots)+(1/16+1/32+1/64)\cdots=1$$

[3] Note that if we ask how much the player is ready to pay to play this game, the formula will be a little different. In such a case, we have to solve for the following equation, U(w) = EU(w+y-p) where w is the initial wealth, y is the prize received from the game, and p the price the player is willing to pay to participate in such a game.

investment decisions by the expected utility EU(w), and not by the expected return, E(w), we are able to solve the St. Petersburg Paradox. But is this solution to the paradox sufficient for the claim that investors should always select among the various investments according to the expected utility criterion (i.e., select the investment with the highest expected utility)?

Although, solving a paradox indicates a good property of the maximum expected utility criterion (MEUC), it cannot serve as justification for employing the MEUC in all cases. Yet, as we shall prove below, the maximum expected utility criterion (MEUC) is the correct rule as long as certain axioms are fulfilled. We will show below that if certain axioms are accepted, then the MEUC is the optimal decision rule. We will first discuss the axioms and then prove that the MEUC is the optimal rule given these axioms. Finally, we will discuss the relationship between MEUC, MERC and MRC, and analyze a few properties of the expected utility criterion.

2.3 THE AXIOMS AND PROOF OF THE MAXIMUM EXPECTED UTILITY CRITERION (MEUC).

Although Bernoulli and Cramer succeeded in solving the St. Petersburg Paradox, they did not provide a theoretical foundation for their solution. This came only in the 20th century when Ramsey and later on, von-Neumann and Morgenstern developed the theory of expected utility which determines that alternative investments should be ranked by their expected utility.[4, 5] The expected utility proof can be formulated in various ways. Here, we adopt six axioms from which the maximum expected utility criterion easily follows. We first discuss the axioms and then provide the proof.

a) The Payoff of the Investments

Suppose that you have to make a choice between two investments called also lotteries, which are denoted, by L_1 and L_2. These two investments can be written as:

---

[4] See F.P. Ramsey, "Truth and Probability," in *The Foundations of Mathematics and Other Logical Essays*, London: K. Paul, Trench, Trusner and Co., 1931. See also, J.M. Keynes, *Essays in Biography*, London: Rupert Hart-Davis, 1951.

[5] See J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton, N.J.: Princeton University Press, 3rd ed., 1953.

$$L_1 = \{p_1\, A_1,\, p_2\, A_2, \ldots, p_n\, A_n\}$$

$$L_2 = \{q_1\, A_1\,,\, q_2\, A_2, \ldots, q_n\, A_n\}$$

where A_i are the possible outcomes with probability p_i and q_i , respectively, and the outcomes are ranked from the smallest (A_1) to the largest (A_n) . Thus, under L_1 we have probability p_1 to get A_1, probability p_2 to get A_2, etc. Similarly, under L_2 we have probability q_1 to get A_1, probability q_2 to get A_2, etc. These are mutually exclusive and comprehensive events, that is, only one outcome can be realized under each investment and $\Sigma p_i = \Sigma q_i = 1$. In practice, it is rare for the two investments to have an identical series of outcomes $A_1, A_2, \ldots A_n$, but this fact imposes no constraints on our analysis: Simply write down all the outcomes of the two options under consideration and assign probabilities of p_i or q_i equal to zero when relevant. For example, if $L_1 = \{\frac{1}{4}4, \frac{3}{4}5\}$ (which should be read as probability of 1/4 to get 4 and a probability of 3/4 to get 5), $L_2 = \{\frac{1}{2}1, \frac{1}{2}10\}$, then we can write these two investments as $L_1 = \{\frac{0}{1}1, \frac{1}{4}4, \frac{3}{4}5, \frac{0}{1}\ 10\}$ and $L_2 = \{\frac{1}{2}1, \frac{0}{1}4, \frac{0}{1}5, \frac{1}{2}\ 10\}$.

b) The Axioms

Axiom 1: Comparability. By this axiom, when faced by two monetary outcomes, say A_i and A_j , the investor must say whether he/she prefers A_i to A_j ($A_i \succ A_j$) (where the sign $\succ$ means "prefers " as distinct from the sign $>$meaning "greater than"), or A_j to A_i ($A_j \succ A_i$) or whether he/she is indifferent between the two ($A_i \sim A_j$) (where the sign $\sim$ means "indifferent"). By this axiom, the answer "I do not know which monetary outcome I prefer" is simply not accepted.

Axiom 2: Continuity. If A_3 is preferred to A_2 and A_2 is preferred to A_1 then there must be a probability $U(A_2)$ ($0 \le U(A_2) \le 1$) such that,

$$L = \{\, (1\text{-}U(A_2))\, A_1,\ (U(A_2))\, A_3\} \sim A_2\,.$$

Thus, the investor will be indifferent between two choices: to receive A_2 with certainty or to receive either A_1 with probability $1\text{-}U(A_2)$ or A_3 with probability $U(A_2)$. For a given A_1 and A_3 , these probabilities are a function of A_2; hence, the notation $U(A_2)$. Why is this axiom called the *continuity* axiom? Simply choose $U(A_2) = 1$ to obtain $L = A_3 \succ A_2$ (because by assumption $A_3 \succ A_2$). Then choose $U(A_2) = 0$ to obtain $L = A_1 \prec A_2$ (because by assumption $A_1 \prec A_2$). Thus, if you increase *continuously* $U(A_2)$ from zero to 1, you will hit a value $U(A_2)$ such that $L \sim A_2$.

Previously, we used the notation p and q for probabilities. Why do we suddenly switch here to $U(A_i)$ (rather than, say, $p(A_i)$)? The reason is simply because $U(A_i)$ is also the investor's utility function; hence, the new notation. This will be demonstrated as we continue with the proof.

Axiom 3: Interchangeability Suppose that you have a lottery (investment) L_1 given by:

$$L_1 = \{ p_1 A_1, p_2 A_2 , p_3 A_3 \} .$$

Assume, also, that you are indifferent between A_2 and another lottery B, where $B = \{q A_1, (1-q) A_3 \}$.

Then by the *Interchangeability* axiom, you will be indifferent between L_1 and L_2 where $L_2 = \{ p_1 A_1, p_2 B, p_3 A_3\}$.

Axiom 4: Transitivity. Suppose that there are three lotteries, L_1, L_2 and L_3, where $L_1 \succ L_2$, $L_2 \succ L_3$. Then, by the *transitivity* axiom, $L_1 \succ L_3$. Similarly, if $L_1 \sim L_2$ and $L_2 \sim L_3$ then, by this axiom , $L_1 \sim L_3$.

Axiom 5: Decomposability. A *complex* lottery is one in which the prizes are lotteries themselves. A *simple* lottery has monetary values A_1, A_2 etc. as prizes. Suppose that there is a complex lottery L^* such that:

$$L^* = (q L_1, (1-q) L_2)$$

where L_1 and L_2 themselves are (simple) lotteries. L_1 and L_2 are given by:

$$L_1 = \{p_1 A_1, (1-p_1) A_2\}$$
$$L_2 = \{p_2 A_1, (1-p_2) A_2\}$$

Then, by this axiom, the complex lottery L^* can be decomposed into a simple lottery L having only A_1 and A_2 as prizes. To be more specific:

$$L^* \sim L_1 = \{p^* A_1, (1-p^*) A_2\}$$

where $$p^* = qp_1 + (1-q) p_2$$

Axiom 6: Monotonicity
If there is certainty, then the *monotonicity* axiom determines that if $A_2 > A_1$ then $A_2 \succ A_1$. If there is an uncertainty, the *monotonicity* axiom can be formulated in two alternate ways. First:

Let $L_1 = \{p\ A_1, (1\text{-}p)\ A_2\}$,
and $L_2 = \{p\ A_1, (1\text{-}p)\ A_3\}$. If $A_3 > A_2$, hence $A_3 \succ A_1$
then $L_2 \succ L_1$.

Second:
Let $L_1 = \{p\ A_1, (1\text{-}p)\ A_2\}$,
and $L_2 = \{q\ A_1, (1\text{-}q)\ A_2\}$, and $A_2 > A_1$ (hence $A_2 \succ A_1$).
If $p<q$ [or if $(1\text{-}p) > (1\text{-}q)$]

then $L_1 \succ L_2$.

Each of these six axioms can be accepted or rejected. However, if they are accepted, then we can prove that the MEUC should be used to choose among alternative investments. Any other investment criterion will simply be inappropriate and may lead to a wrong investment decision. We turn next to this proof.

c) Proof that the Maximum Expected Utility Criterion (MEUC) is Optimal Decision Rule

Theorem 2.1: The MEUC. The optimum criterion for ranking alternative investments is the expected utility of the various investments.

To prove Theorem 2.1, suppose that you have to make a choice between two investments L_1 and L_2 given by:

$$L_1 = \{p_1\ A_1\ ,\ p_2\ A_2, \ldots, p_n\ A_n\}\ ;$$
$$L_2 = \{q_1 A_1\ ,\ q_2 A_2, \ldots, q_n\ A_n\},$$

and $A_1 < A_2 < \ldots A_n$, where A_i are the various monetary outcomes.

First, note that by the *comparability* axiom, we are able to compare the A_i. Moreover, because of the *monotonicity* axiom we can determine that:

$$A_1 < A_2 < \ldots\ldots A_n \text{ implies } A_1 \prec A_2 \prec \ldots\ldots A_n.$$

Define $A_i^* = \{ (1-U(A_i))\, A_1, U(A_i)\, A_n \}$ where $0 \le U(A_i) \le 1$. By the *continuity* axiom, for every A_i, there is a probability $U(A_i)$ such that:

$$A_i \sim A_i^*$$

Note that for A_1, we have $U(A_1) = 0$; hence $A_1 \sim A_1$ and for A_n, $U(A_n) = 1$, hence $A_n \sim A_n$. For all other values A_i, we have $0 < U(A_i) < 1$ and, due to the *monotonicity* and *transitivity* axioms, $U(A_i)$ increases from zero to 1 as A_i increases from A_1 to A_n.[6]

Substitute A_i by A_i^* in L_1 and, by the *interchangeability* axiom, we obtain:

$$L_1 \sim L_1^* \equiv \{p_1 A_1, p_2 A_2, \ldots p_i A_i^* \ldots\ldots p_n A_n\}$$

where the superscript of L_1 indicates that one element A_i^* has been substituted in L. Then substitute one more element in L_1^* and use the *interchangeability* and *transitivity* axioms to obtain that $L_1 \sim L_1^* \sim L_1^{**}$ where L_1^{**} is the lottery when two elements are substituted. Continue this process and denote the lottery by $\tilde{L}_1$ where all its elements A_i $(i = 1, 2, \ldots, n)$ are substituted by A_i^* to obtain:

$$L_1 \sim \tilde{L} \equiv \{p_1 A_1^*, \ldots p_2 A_2^*, \ldots, p_n A_n^*\}$$

By the *decomposability* and *transitivity* axioms, we have:

$$L_1 \sim \tilde{L}_1 \sim \tilde{\tilde{L}}_1 \equiv \{A_1 \sum p_i (1-U(A_i)), A_n \sum p_i U(A_i)\} \ .$$

We repeat all these steps with lottery L_2 to obtain:

$$L_2 \sim \tilde{L}_2 \sim \tilde{\tilde{L}}_2 = \{A_1 \sum q_i (1-U(A_i)), A_n \sum q_i U(A_i)\} \, .$$

---

[6] To see this, suppose that $6 \sim \{ \frac{1}{4} 1, \frac{3}{4} 10 \} \equiv A_6^*$. Then we claim that for a higher value, say 7, we have $7 \sim \{(1-\alpha\} \ 1, \ \alpha\{10\} \equiv A_7^*$ where $\alpha > 3/4$. Due to the *monotonicity* 7 $\succ$ 6; due to the *transitivity*, $A_7^* \succ A_6^*$, and due to the *monotonicity*, $\alpha > 3/4$, which is exactly what is claimed above.

Recall that $A_n > A_1$. Therefore, by the *monotonicity* axiom, $\tilde{\tilde{L}}_1$ is preferred to $\tilde{\tilde{L}}_2$ if the following condition holds:

$$\Sigma p_i U(A_i) > \Sigma q_i U(A_i) .$$

But, because of the *transitivity* , this also implies the same inequality with the original investments; hence $L_1 \succ L_2$.

How is this result related to expected utility? Assume for a moment that $U(A_i)$ is the utility of A_i. Then, given the above set of axioms, the investment with the highest expected utility is preferred, namely:

$$L_1 \succ L_2 \Leftrightarrow \sum p_i U(A_i) \equiv E_{L_1} U(x) > \sum q_i U(A_i) \equiv E_{L_2} U(x)$$

where x denotes the possible monetary outcomes (the A_i in our proof) and the subscripts L_1 and L_2 denote the expected utility of L_1 and L_2, respectively. We shall see below that the probabilities $U(A_i)$ do, indeed, represent the investor's preference regarding the various outcomes, hence they will also represent the utility corresponding to outcome A_i. Thus, $U(A_i)$ will be shown to be the investor's utility function.

2.4 THE PROPERTIES OF UTILITY FUNCTION

a) Preference and Expected Utility

We proved above that if the expected utility of L_1 is larger than the expected utility of L_2 then L_1 will be preferred to L_2. Actually, preference is a fundamental property reflecting the investor's taste. Therefore, it is more logical to turn the argument around and assert that if L_1 f L_2 then there is a non-decreasing function U_1 such that $E_{L_1} U_1(x) > E_{L_2} U_1(x)$. Note that $L_1 \succ L_2$ is possible for one investor and $L_2 \succ L_1$, is possible for another investor. This implies that there is another non-decreasing function U_2 reflecting the second investor's preference such that $E_{L_2} U(x) > E_{L_1} U(x)$. This non-decreasing function is called *utility* function. Why does the function $U(A_i)$ reflect the investor's taste or the investor's utility from money? The reason is that by the *continuity* axiom for any two values

A_1 and A_n (where $A_n > A_1$) and $A_1 < A_i < A_n$, there is a function (probability) $U(A_i)$ such that:

$$\{(1-U(A_i)A_1, U(A_i)A_n\} \equiv A_i^* \sim A_i \ .$$

Not all investors would agree on the specific value of $U(A_i)$ but, for each investor, such a function $U(A_i)$, (with $0 \leq U(A_i) \leq 1$), must exist. Because $U(A_i)$ differs from one investor to another, it reflects the investor's preference; hence, it is called utility function and it reflects the investor's taste or indifference curve. The indifference curve is generally measured by a comparison between uncertain investment and a certain cash flow as we shall see in the next example.

Example:

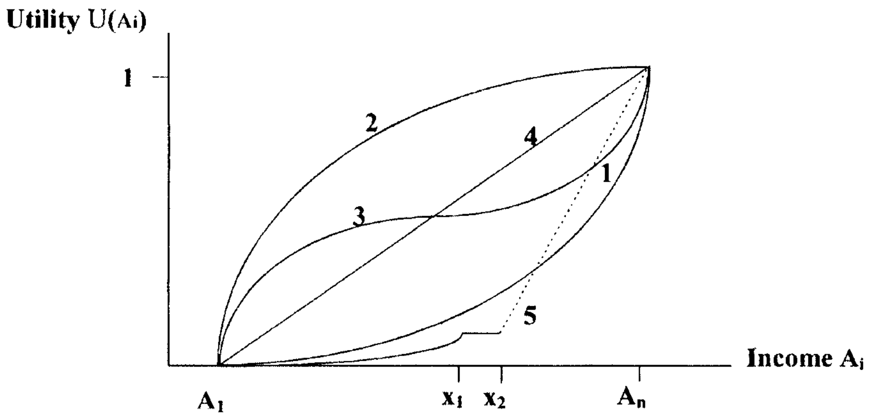
Suppose that $A_1 = \$0$, $A_2 = \$10$ and $A_3 = \$20$.

By the *continuity* axiom, then there is a function $0 < U(A_2) < 1$ such that:

$$L^* \equiv \{ (1\text{-}U(A_2))\ \$0\ ,\ U(A_2)\ \$20 \} \sim \$10\ .$$

If various investors were asked to determine the $U(A_2)$ which would make them indifferent between receiving \$10 for sure or L^*, they would probably assign different values $U(A_2)$. One investor might decide on $U(A_2) = 1/2$. Another investor who dislikes uncertainty might decide on $U(A_2) = 3/4$. Because $U(A_2)$ varies from one investor to another according to his/her taste or preference, it is called the utility function, or the utility assigned to the value A_2. Therefore, $U(A_i)$ is called the utility of A_i , and the investment with the highest expected utility $\Sigma\ P_i\ U(A_i)$ is the optimal investment. The function $U(A_i)$ has only one property: It is $U(A_1) = 0$ for the lowest value A_1 and $U(A_n) = 1$ for the highest value (see Footnote 6) A_n and, due to the *monotonicity* axiom, it increases (in the weak sense) as A_i increases. Thus, the only constraint on $U(A_i)$ is that it is non-decreasing. Figure 2.1 illustrates various possible utility functions: all of them are possible and none of them contradict expected utility theory.

Figure 2.1: Various Utility Functions



Curve 1 is convex, curve 2 is concave, curve 3 has convex as well as concave segments, and curve 4 is linear. Note that $U(A_i)$ do not have to strictly increase through the whole range. We allow also for a function that is constant for some ranges, say range $x_1 \le x \le x_2$ (see curve 5). Thus $U(A_i)$ is a non-decreasing function of A_i. All these functions reflect various preferences; all conform with the *monotonicity* axiom which determines that the higher the monetary outcome A_i, the higher (or equal) the utility $U(A_i)$. Thus, if $A_j > A_i$, then $U(A_j) \ge U(A_i)$. What is the intuitive explanation for the fact that the utility function is non-decreasing in income? The utility function cannot decrease because if $U(A_2) < U(A_1)$ and ($A_2 > A_1$) your utility would increase by simply donating $A_2 - A_1$ to charity; your utility increases (and other people would also be able to enjoy your money!). As we shall see in the next chapter, we will develop stochastic dominance rules for various types of utility functions (e.g., all convex, all concave, etc.).

b) Is U(x) a Probability Function or a Utility Function?

In the proof of MEUC, we assume that U(x) is a probability with $0 \le U(x) \le 1$. However, we also called this function a utility function. Does this mean that the utility of any monetary outcome is bounded between 0 and 1? No, it doesn't: Utility can take on any value, even a negative one. We can start with $0 \le U(x) \le 1$ (as done in the proof of Theorem 2.1) and then conduct a transformation on U(x) and, therefore, it can take on any value, even a negative one, without changing the ranking of the investments. Therefore, we can switch from $0 \leq U(x) \leq 1$ to any other (unbounded) utility function. This is summarized in the following theorem.

Theorem 2.2: A utility function is determined up to a positive linear transformation , where "determined" means that the ranking of the projects by MEUC does not change.

Proof: First, we define a positive linear transformation as $U^*(x) = a+b \cdot U(x)$ where $b > 0$ and $a \gtrless 0$. Suppose that there are two risky investments with returns x and y, respectively. Then by the Theorem claim $EU(x) > EU(y)$, if and only if $EU^*(x) > EU^*(x)$, where:

$$U^* (\cdot) = a + bU (\cdot) \text{ and } b > 0.$$

That is, $U^* (\cdot)$ is a positive linear transformation of $U (\cdot)$.

To see this recall that:

$$EU^*(x) = a + b \cdot EU(x);$$

$$EU^*(y) = a + b \cdot EU(y)$$

and it can be easily seen that for $b > 0$ and $a \lesseqgtr 0$, $EU(x) > EU(y)$ if and only if $EU^* (x) > EU^*(y)$.

Thus, the investments' ranking by U or U^* is identical; if x has a higher expected utility than y with U, it has a higher expected utility with U^* (and vice versa). Therefore, one can shift from U to U^* and vice versa without changing the ranking of the alternative choices under consideration.

Can we use the utility function as a probability function as in the proof of Theorem 2.1? Yes, we can. We use Theorem 2.2 to show this claim.

Suppose that there is a utility function reflecting the investor's preference. Then, one can conduct a linear transformation to obtain another utility function U^* such that U^* will be between zero and one; hence, U^* can be used as a probability in the proof of Theorem 2.1. To demonstrate how such a normalization is carried out, suppose that x reflects all possible values A_i , $x_1 = A_1$ is the lowest monetary value, and $x_n = A_n$ is the highest monetary value. $U(A_1)$ and $U(A_n)$ are unrestricted utilities which can even be negative. We can then conduct a linear positive transformation such that

$$U^*(A_1) = a + b\cdot U(A_1) = 0\ ;$$

$$U^*(A_n) = a + b\cdot U(A_n) = 1\ .$$

Thus, we have two equations with two unknowns (a and b), and we can solve for a and b as follows:

Subtract one equation from the others to obtain:

$$b(U(A_n) - U(A_1)) = 1\ ,$$

or

$$b = 1/[U(A_n) - U(A_1)];$$ [7]

and, from the first equation:

$$a = -bU(A_1) = -\ U(A_1)/[U(A_n) - U\ (A_1)]\ .$$

Thus, for any utility function U, we can select a and b such that there will be a new function U^* with $U^*(a) = 0$ and $U^*(b) = 1$. Because such a transformation does not change the project ranking, U^* can be employed both as probability function as in the proof of Theorem 2.1 as well as a utility function for ranking investments.

Example:

Let $U(A_1) = -5$ and $U(A_n) = 95$. Then, by the above solution for a and b, we have:

$$b = 1\ /[95 - (-5)] = 1/100;$$

$$a = -(-5)\ /[95 - (-5)] = 5/100\ .$$

---

[7] Note that $U(A_n) > U(A_1)$, hence $b > 0$ which confirms that a positive linear transformation is employed.

Thus, the function U^* is given by $U^*(x) = 5/100 + 1/100\ U(x)$.

Indeed, for x = a, we have:

$$U^*(a) = 5/100 + 1/100 \cdot (-5) = 0$$

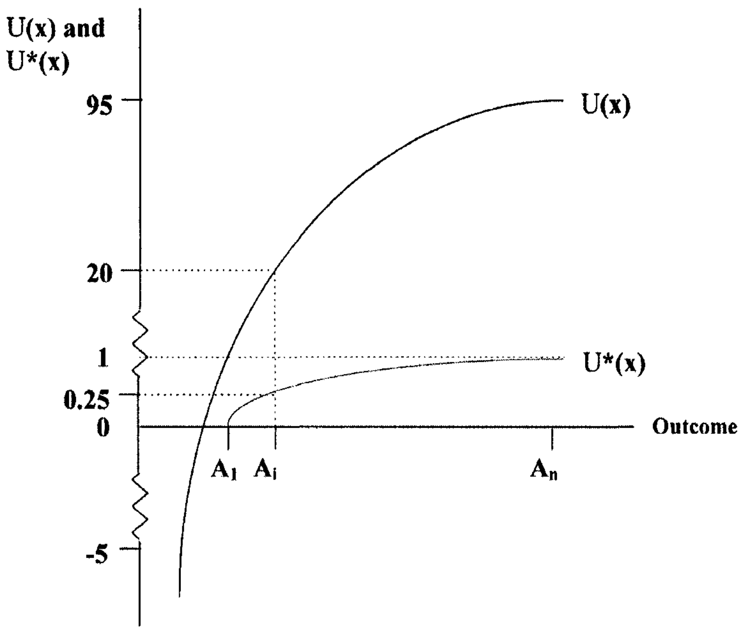
and, for x = b, we have:

$$U^*(b) = 5/100 + 1/100\ (95) = 1\ .$$

Hence $U^*(a) = 0$ and $U^*(b) = 1$ as in the proof of the MEUC.

Figure 2.2 illustrates the utility function U(x) and $U^*(x)$ corresponding to the above example. It should be noted that if there is another value, say A_i, for which $U(A_i) = 20$, then $U^*(A_i) = 5/100 + (1/100) \cdot 20 = 25/100 = 1/4$. All values $U(A_i)$ are determined by the same technique.

Figure 2.2: A linear positive transformation $U^*(x) = a + b \bullet U(x)$



This example illustrates that we can take any utility function U(x) and conduct the linear transformation shifting to $U^*(x)$ without changing the investor's project ranking, and this $U^*(x)$ function can be employed as a probability function, as in the proof of Theorem 2.1.

2.5 THE MEANING OF THE UTILITY UNITS

The utility units, which are called *utiles*, have no meaning: If the utility of investment A is 100 and the utility of investment B is 150, we cannot claim that investment B is 50% better. The reason for this is that we can conduct a positive linear transformation and expand or suppress the difference between the utility of investment A and investment B arbitrarily. In the above example, we have the original utility function for the two values A_1 and A_n.

$$U(A_1) = -5 , \qquad U(A_n) = 95 \text{ , and } U(A_n) - U(A_1) = 100 .$$

After the transformation, we have $U^*(A_1) = 0$ and $U^*(A_n) = 1$; hence the difference between the utility of A_n , and A_1 decreases from 100 to 1.

Because we can shift from U to U^* without changing the ranking of the various investments, we can say that the only important thing is the ranking of the investment by the expected utility and there is no meaning to the difference in the expected utility of the two projects under consideration. The "utilities" themselves and in particular, the magnitude of the difference of utilities, are meaningless. Moreover, a negative utility does not imply that the investment is unattractive. We demonstrate this in the following example:

Example:

Suppose that you are offered one of the following cash flows (denoted by x) for the same amount of money or free of charge corresponding to two distinct investments denoted by A and B:

Investment A		Investment B	
X	**p(x)**	**x**	**p(x)**
5	1/2	8	1
10	1/2		

Suppose that your preference is given by $U(x) = x^2$. Which cash flow would you select? By the MEUC, we have to select the one with the highest expected utility.

Simple calculation reveals:

$$E_AU(x) = (1/2)\,5^2 + (1/2)\,10^2 = 62.50\ ;$$

$$E_BU(x) = (1)\ 8^2 = 64.$$

Hence, B is preferred. Now use the utility function, $U^*(x) = 100\,x^2$, which is a positive linear transformation of U(x), with a = 0 and b = 100. With this new function we have:

$$E_A\,U_1^*(x) = 100 \cdot 62.5 = 6,250;$$

$$E_B\,U_1^*(x) = 100 \cdot 64 = 6{,}400\,.$$

Hence, $EU_B(x) > E_AU(x)$.

The difference, which was 64-62.5 = 1.5 between $E_BU(x)$ and $E_AU(x)$, increases to 6,400–6,250 = 150 with $E_B\,U_1^*(x)$ and $E_A\,U_1^*(x)$. This does not imply that B becomes much better because the only factor that is relevant is that $E_BU(x) > E_A$

$U(x)$ (i.e., B is ranked above A), and the magnitude of the difference $E_AU(x)$ - $E_BU(x)$ is meaningless.

Now consider the following utility function $U_2^*(x) = -100 + x^2$ which is again a positive linear transformation of U(x) with $a = -100$ and $b = 1$.

With this function we have :

$E_A\ U_2^*(x) = -100 + 62.5 = -37.5;$

$E_B\ U_2^*(x) = -100 + 64 = -36.$

We obtain a negative expected utility. Does this mean that the investor should reject both cash flows? No, it doesn't. With no cash flow (x=0), the utility is:

$U_2^*(x) = -100 + 0^2 = -100$

and, because $-36 > -100$, the investor is better off selecting investment B. Thus, utility and expected utility can be negative. We cannot infer that the investment with a negative expected utility should be rejected. We simply rank all investments by their expected utility and select the one with the highest expected utility. Note, however, that the option not to invest at all may have the highest expected utility in which case all of the projects will be rejected.

One might be tempted to believe that any monotonic increasing transformation, not necessarily linear, also maintains the ranking of the project. This is not so. To see this, assume that we have U(x) for which $U(x) \geq 0$ (e.g., $U(x) = x^2$), the function employed earlier. Consider the following increasing monotonic transformation:

$$U^*(x) = [U(x)]^2 \text{ (for } U(x) > 0).$$

Because, by assumption, $U(x) = x^2$ (all $x > 0$), $U^*(x) = x^4$. This is an increasing monotonic transformation of U(x).

Recall that with the above example and with $U(x) = x^2$, we have $E_BU(x)= 64 > E_AU(x) = 62.5$. Let us show that the ranking of these two investments is reversed with $U^*(x)$. We have:

$E_AU^*(x) = (1/2)\ 5^4 + (1/2)\ 10^4 = (1/2)\ 625 + (1/2)\ 10{,}000 = 5{,}312.5;$

$E_BU^*(x) = (1)\ 8^4 = 4.096\ .$

Hence, $E_AU^*(x) > E_B\ U^*(x)$, which differs from the ranking by U(x) obtained before. Thus, a positive *linear* transformation is allowed (does not change the ranking), and a positive *monotonic* transformation is not allowed because it may change the ranking of the project under consideration.

To sum up, a linear positive utility transformation does not affect the project's ranking and, as this is what is important for the investor, such a transformation is allowed. However, a monotonic transformation which is not linear may, or may not, keep the project's ranking unchanged and, therefore, it is not permissible.

Secondly, we can choose a transformation, which makes the utility function, intercept any two points a_1 and a_2. To do this, select $U(x_1) = a_1$ and $U(x_2) = a_2$. In our specific case, we selected $a_1 = 0$ and $a_2 = 1$ such that the utility can be used as a probability function. However, any pair (a_1, a_2), which we want the utility function to intercept, can be selected.

2.6 MRC, MERC AS SPECIAL CASES OF MEUC

Proving that MEUC is optimal does not imply that MRC and MERC should never be employed. On the contrary, these two criteria are special cases of MEUC.

Let us first show that if the utility function is linear of the type,

$$U(x) = a + bx \qquad (b > 0)$$

then MEUC and MERC coincide.

To see this, let us compare two investments denoted by x and y. By the MEUC we have that $x \succ y$ if and only if $EU(x) > EU(y)$. But with linear utility function we have:

$$EU(x)>EU(y) \Leftrightarrow a + bEx > a + b\,Ey \quad (\text{for } b > 0);\ \text{hence,}\ Ex > Ey.$$

Thus, the project ranking by EU(•) is the same as the ranking by the expected value; hence, for linear utility function, the MERC coincides with the MEUC.

Finally, if returns are certain, MRC is obtained as a special case of MEUC. To see this, consider two certain investments (1,x) and (1,y) where 1 means that the probability of obtaining x and y, respectively, is equal to one. Then, by the MRC, $x \succ y$ if $x > y$. But, because of the *monotonicity* of the utility function, $x > y \Rightarrow U(x) > U(y)$ and $1 \cdot U(x) > 1 \cdot U(y)$. However, the last inequity implies that $EU(x) > EU(y)$ for the degenerated case where the probability is 1.

2.7 UTILITY, WEALTH AND CHANGE OF WEALTH

Denote by w the investor's initial wealth and by x the change of wealth due to an investment under consideration. While w is constant, x is a random variable. The utility function is defined on total wealth w+x. The inclusion of w in the utility is crucial as the additional utility due to the possession of a risky asset, e.g., a stock or a bond, depends on w. For
example, a poor person with w = \$10,000 may appreciate an addition of x = \$1,000 (in terms of utility) more than a millionaire with w = \$10 million.

Despite the importance of w in the expected utility paradigm, there is evidence that investors in the decision making processes tend to ignore w and focus on change of wealth. Thus, instead of looking at U(w+x) the investors make decisions based on U(x). To the best of our knowledge, the first one to suggest that in practice investors make decisions based on change of wealth was Markowitz as early as 1952.[8] However, only in 1979[9] when Kahneman and Tversky published their famous Prospect Theory study, has this issue received more attention by economists. Indeed, one of the important components of Prospect Theory is that in practice (based on experimental findings) decisions are made based on change of wealth, x rather than total wealth w + x.

Making decisions based on change of wealth contradicts expected utility paradigm, but it does not contradict Stochastic Dominance (SD) framework despite the fact that SD is derived within expected utility paradigm. As we shall see in this book (see next chapter), if distribution F dominates distribution G for all U(w+x) in a given set of preferences (e.g., risk-aversion), the same dominance is intact also for all U(x) in the same set of preferences. In other words, the *partial ordering* of SD which determines the efficient and inefficient sets of investments

[8]Markowitz, H., "The utility of wealth," *Journal of Political Economy*, *60*, 1952, pp. 151-156.
[9]Kahaneman, D.K., and Tversky, A., "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, *47*, 1979, pp. 263–291.

(as well as dominance by Markowitz's Mean-Variance rule) is invariant to w, while the selection of the optimal choice from the efficient set does depend on w. As we discuss in this book, in employing SD rules which provide the various efficient sets (i.e., partial ordering), we can safely ignore w.

2.8 SUMMARY

If a certain set of axioms is accepted , then the optimum investment criterion is the maximum expected utility criterion (MEUC) and all investments should be ranked by their expected utility. The ranking of the project is important but the magnitude of the difference of expected utility of two investments under consideration is meaningless because one can expand or shrink this difference by conducting a positive linear transformation. Thus, although utility function is cardinal (each project is not only ranked but also assigned a number), actually it is ordinal because what really matters in investment decisions is only the ranking of the projects.

The maximum return criterion (MRC) and the maximum expected return criterion (MERC) are special cases of MEUC where we have certainty or linear utility function, respectively.

In the proof of MEUC, nothing is assumed regarding the shape of the utility function apart from monotonicity (i.e., it is non-decreasing). In the next chapter we develop the stochastic dominance rules, which are optimal decision rules for various possible utility functions (i.e., subsets of all possible utility functions).

Key Terms

The Maximum Return Criterion (MRC).

The Maximum Expected Return Criterion (MERC).

The St. Petersburg Paradox.

Certainty Equivalent.

Comparability Axiom.

Continuity Axiom.

Interchangeability Axiom.

Transitivity Axiom.

Decomposability Axiom.

Monotonicity Axiom.

Optimal Decision Rule.

Utility Function.

Utiles.

Positive Linear Transformation.

Positive Monotonic Transformation.

Partial Ordering

Efficient Set

Inefficient Set

3

STOCHASTIC DOMINANCE DECISION RULES

3.1 PARTIAL ORDERING: EFFICIENT AND INEFFICIENT SETS

We have seen that the MEUC is the optimal investment criterion. If there is full information on preferences (e.g., U(w) = log (w)), we simply calculate EU(w) of all the competing investments and choose the one with the highest expected utility. In such a case, we arrive at a *complete ordering* of the investments under consideration: there will be one investment which is better than (or equal to) all of the other available investments. Moreover, with a complete ordering, we can order the investments from best to worst. Generally, however, we have only partial information on preferences (e.g., risk aversion) and, therefore, we arrive only at a *partial ordering* of the available investments. Stochastic dominance rules as well as other investment rules (e.g., the mean-variance rule) employ partial information on the investor's preferences or the random variables (returns) and, therefore, they produce only partial ordering.

Let us illustrate the notion of partial ordering and complete ordering. Suppose that all that we know is that the utility function is non-decreasing with $U' \geq 0$, namely, investors always prefer more money than less money.

Thus, we have partial information on U and its precise shape is unknown. Later on in the chapter, we will develop an investment decision rule corresponding to this partial information. This rule, called First Degree Stochastic Dominance (FSD), is appropriate for all investors with $U' \geq 0$ (with a strict inequality at some range). Let us use this decision rule (any other rule corresponding to other partial information can be used in a similar way) to introduce some definitions, all of which are commonly used in the financial literature and which are needed for the explanation of partial and complete ordering.

The *feasible set* (FS) is defined as the set of all available investments under consideration. Suppose that the feasible set, denoted by FS, is represented by the area FS (see Figure 3.1). There are no other investments and therefore these five investments represent all available investments. Then, using an investment rule such as the FSD, we divide the whole feasible set, FS, into two sets: the efficient set (denoted by ES) and the Inefficient Set (denoted by IS). These two sets are mutually exclusive and comprehensive; namely:

(ES) U (IS) = FS (where U here denotes union).

Figure 3.1 demonstrates this division of the feasible set, FS into the two sets, IS and ES. In this example, the feasible set includes five investments A, B, C, D and E. Each investment included in the feasible set must be either in the ES or in the IS.

To facilitate the definition and discussion, suppose that the only information we have is that $U' \geq 0$. Thus, $U \in \mathbf{U_1}$ if $U' \geq 0$ where $\mathbf{U_1}$ is the set of all non-decreasing utility functions. We demonstrate below the concept of the efficient set and the inefficient set and the relationship between the two sets for this type of information (namely $U \in \mathbf{U_1}$). Similar analyses hold for other sets of information on preferences regarding the distribution of returns. Before we define the ES and IS formally, we need the following definitions.

Dominance in U_1: We say that investment I dominates investment II in $\mathbf{U_1}$ if for *all* utility functions such that $U \in \mathbf{U_1}$, $E_I U(x) \geq E_{II} U(x)$ and for at least one utility function $U_0 \in \mathbf{U_1}$, there is a strict inequality.

Efficient set in U_1: An investment is included in the efficient set if there is no other investment that dominates it. The efficient set includes all undominated investments. Referring to Figure 3.1, we can say that investments A and B are efficient. Neither A nor B dominates the other. Namely, there is a utility function $U_1 \in \mathbf{U_1}$ such that

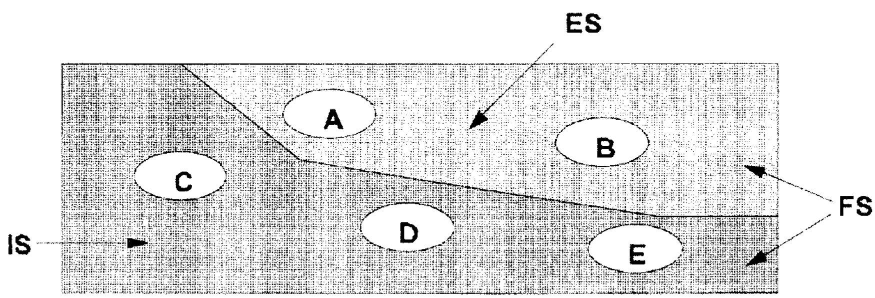
$$E_A U_1(x) > E_B U_1(x).$$

and there is another utility function, $U_2 \in \mathbf{U_1}$ such that

$$E_B U_2(x) > E_A U_2(x).$$

Thus, neither A nor B is the "best" for all investors included in the group $U \in \mathbf{U_1}$. Some investors may prefer A and some may prefer B, and there is no dominance between A and B.

Figure 3.1: The Feasible, Efficient, and Inefficient Sets



Inefficient set in U_1: The inefficient set, IS, includes all inefficient investments. The definition of an inefficient investment is that there is at least one investment in the efficient set that dominates it. Figure 3.1 shows that investments C, D and E are inefficient. For example, we may have the following relationships:

$$E_A\, U(x) > E_C\, U(x)$$

$$E_A\, U(x) > E_D\, U(x)$$

$$E_B\, U(x) > E_E\, U(x)$$

$$\text{for all } U \in \mathbf{U_1}.$$

Thus, the efficient investment A dominates investment C and D, and the efficient investment B dominates investment E. There is no need for an inefficient investment to be dominated by *all* efficient investments. One dominance is enough to relegate an investment to the inefficient set. To be more specific in the above example, A does not dominate E. However, if we also had:

$$E_A\, U(x) > E_E\, U(x) \text{ for all } U \in \mathbf{U_1}$$

it would not add anything to the partial ordering because investment E is already dominated by investment B and hence it is inefficient and no investor will select it.

The partition of the feasible set, FS, to the efficient set (ES) and inefficient set (IS) depends on the information available. In the above example, we assume the information, $U \in \mathbf{U_1}$. If, for example, in addition to $U' > 0$, we assume also that $U'' < 0$ or any other relevant information, we will get another partition of the FS to IS and ES reflecting this additional information. However, in principle, the definition of the ES and IS, and the dominance relationship are as defined and illustrated in the above case (i.e., for $U \in \mathbf{U_1}$). The only difference would be that we would have to change $\mathbf{U_1}$ to $\mathbf{U_i}$ where $\mathbf{U_i}$ is the set of utilities corresponding to the assumed information.

Generally speaking, for any given piece of information, the smaller the efficient set relative to the feasible set, the better off the investors. To demonstrate this, suppose that there are 100 mutual funds (investments) and an investment consultant wishes to advise his clients which funds to buy. Assume that the only information known is that $U \in \mathbf{U_1}$. Suppose that the consultant has an investment decision rule corresponding to this information. Employing this rule, the FS is divided into the ES and the IS. An ES that includes, say, 10 mutual funds, is much better than an ES that includes, say, 80 mutual funds. In the former case, the consultant can suggest that his clients choose their investment from among the 10 mutual funds according to their specific preference (which are unknown to the investment consultant) whereas in the latter case, he would advise them to choose from among 80 mutual funds. Clearly, in the second case, the consultant is not being very helpful to his clients because he eliminated only 20 mutual firms from the possible choices. In the extreme case, if we are lucky and the efficient set includes only one mutual fund, the investment consultant will have provided very sharp, clear advice to his clients. As we shall see in this and the next chapter, the more the assumptions or information on preferences or on the distribution of returns, the more restricted the efficient set.

In investment choice with partial information (hence with partial ordering) there are two decision stages, the first involving the investment consultant and the second, the individual investor. The two stages are as follows:

a) The objective decision:

In the first stage, the initial screening of investments is accomplished by partitioning the FS into the ES and IS. Because the ES generally includes more than one investment and we cannot tell which one is the best, this stage provides only partial ordering. If we possess full information (e.g., $U(x) = \log x$), then the efficient set will include only one investment (or several with the same expected utility, but the investor can pick one of them arbitrarily) and we arrive at a complete ordering of the investments.

b) The subjective decision:

The optimum investment choice by the investor from the ES. This optimal choice maximizes the investor's expected utility. This is a subjective decision because it depends on the investor's preferences.

The decision rules developed in this book are used for the initial screening of investments (the first decision stage) in which some of the investment is relegated to the inefficient set based on partial information (e.g., $U' \geq 0$). All investors will agree on this partition into the efficient and inefficient sets. In the second stage, each investor will select the *optimal portfolio* from the efficient set according to his/her preferences. In this stage, there will be little or no agreement between investors, each will select his/her optimal portfolio according to his/her specific preferences. It is possible, although very unlikely, that the efficient set based on partial information will include only one portfolio. In such a case, all investors will also agree on the optimal portfolio and the two stages will converge.

3.2 FIRST DEGREE STOCHASTIC DOMINANCE (FSD)

In this section, we prove and discuss the FSD in detail. However, because stochastic dominance rules rely on distribution functions, some discussion of probability function is called for before turning to the FSD rule.

a) Probability function, density function and cumulative probability function

Let us first define probability function, density function and cumulative probability function (distribution function).

The pair (x,p) where x is the outcome and p(x) is its corresponding probability is called a probability function. If the random variable x is continuous, then the probability function is replaced by the density function f(x). The cumulative probability function denoted by F(x) is given by:

$$F(x) = P(X \le x) = \sum_{X \le x} P(x) \text{ for a discrete distribution}$$

and

$$F(x) = \int_{-\infty}^{x} f(t)dt \text{ for a continuous random variable}$$

where X denotes a random variable and x a particular value.

We have $F(-\infty) = 0$ and $F(+\infty) = 1$.

Example:

1. Discrete random variable

Suppose that the x and p are given as follows:

X	**p(x)**
–5%	¼
0%	1/8
5%	1/8
20%	½

Then F(x) is given by

$$F(x) = \begin{cases} 0 & x < -5\% \\ 1/4 & -5\% \le x < 0 \\ 3/8 & 0 \le x < 5\% \\ 1/2 & 5\% \le x < 20\% \\ 1 & x \ge 20\% \end{cases}$$

Figure 3.2a and 3.2b demonstrate the probability and the cumulative probability functions, respectively. As we can see, even though the probability function is discrete, the cumulative probability function is continuous from the right-hand side. If \$1 is invested, then the terminal wealth is $w = 1\cdot(1+x)$ and F(w) is similar to F(x) but shifted \$1 to the right.

2. Continuous random variable

When we have a continuous random variable, the density function replaces the probability. To see this, suppose that the outcome has a uniform distribution (i.e., the same density everywhere) with range $5 \le x \le 15$. As the total area under the density function must be equal to 1, the density function must be 1/10. Figure 3.3 demonstrates the density function. The cumulative distribution function in the range $5 \le x \le 15$ is given by:

$$F(x) = \int_5^x f(t)dt = \int_5^x (1/10)dt = \tfrac{x-5}{10}$$

Fig. 3.2a: Probability Function

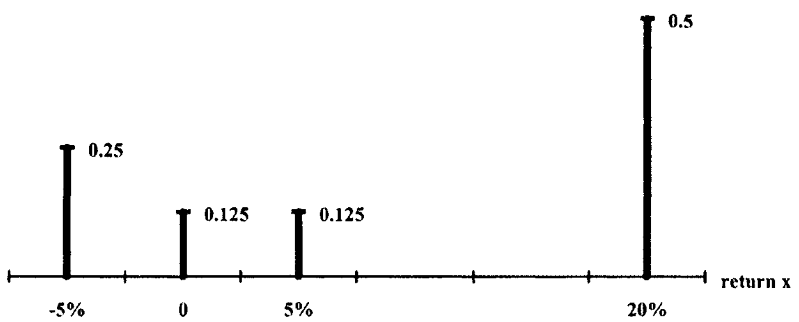
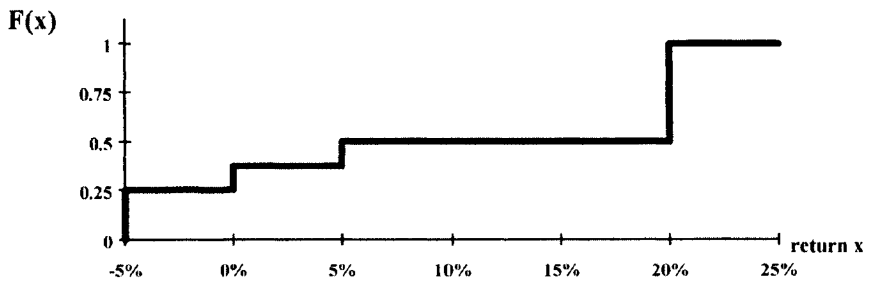


Fig. 3.2b: Cumulative Probability Function



for the range $5 \le x \le 15$, and $F(x) = 0$ for $x < 5$ and $F(x) = 1$ for $x \ge 15$.

b) The FSD rule

Suppose that the investor wishes to rank two investments whose cumulative distributions are F and G. We denote these two investments by F and G, respectively. The FSD rule is a criterion that tells us whether one investment dominates another investment where the only available information is that $U \in U_1$, namely that $U' \ge 0$ and, to avoid the trivial case of U′ coinciding with the horizontal axis, there is a range where $U' > 0$. Actually, this is the weakest assumption on preference because we assume only that investors like more money rather than less money, which conforms to the monotonicity axiom. For most of the proofs in this chapter, we assume that U is a continuous non-decreasing function which implies that it is differentiable apart from a set of points whose measure is zero. We also deal below with continuous random variables and then extend all the stochastic dominance results to discrete random variables.

Theorem 3.1: Let F and G be the cumulative distributions of two distinct investments. Then F dominates G by FSD (which we denote by FD_1G, where D_1 denotes dominance by the first degree and the subscript 1 indicates that we assume only one

Figure 3.3a: Density Function

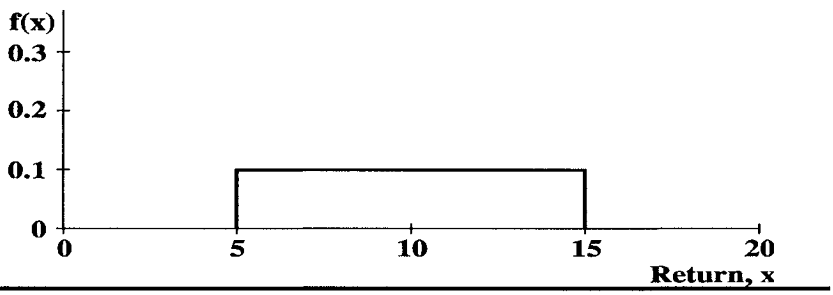
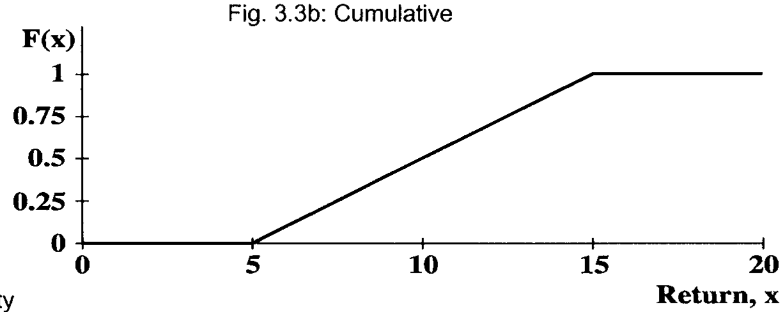


Fig. 3.3b: Cumulative



Probability

piece of information on U, namely that U is nondecreasing) for all $U \in \mathbf{U_1}$ if and only if $F(x) \le G(x)$ for all values x, and there is at least some x_0 for which a strong inequality holds. As FSD relates to $U \in \mathbf{U_1}$, it can be summarized as follows:

F(x) ≤ G(x)	⇔	**$E_FU(x) \ge E_GU(x)$**
for all x with a strong inequality for at least one x_0.		for all $U \in \mathbf{U_1}$ with a strong inquality for at least one $U_0 \in \mathbf{U_1}$

Defining $G(x) - F(x) = I_1(x)$ (again the subscript 1 reminds us that we dealing with first degree stochastic dominance), then the condition for FSD of F over G is that $I_1(x) \ge 0$ for all x and $I_1(x_0) > 0$ for some x_0.

Proof: In all the proofs in this chapter, for simplicity and without loss of generality, we assume that x is bounded from below and from above, namely, $a \le x \le b$ which implies that for $x \le a$ $F(x) = G(x) = 0$ and for $x \ge b$ $F(x) = G(x) = 1$. However, all the results are intact also for $-\infty < x < \infty$. The extension to unbounded random variables is given in Hanoch and Levy [1969][1] and Tesfatsion [1976][2].

Sufficiency: It is given that $F(x) \le G(x)$ for all values x and we have to prove that $E_FU(x) \ge E_GU(x)$ for all $U \in \mathbf{U_1}$. By the definition of expected utility we have:

$$\Delta \equiv E_F U(x) - E_G U(x) = \int_a^b f(x)U(x)dx - \int_a^b g(x)U(x)dx$$

where a is the lower bound and b the upper bound; namely, $a \le x \le b$, and f(x) and g(x) are the density functions of the returns on the two investments F and G, respectively.

The difference in expected utilities denoted by Δ can be rewritten as:

$$\Delta \equiv \int_a^b [f(x) - g(x)]U(x)dx\ .$$

[1] Hanoch, G. and H. Levy, "The Efficiency Analysis of Choices Involving Risk," *Review of Economic Studies*, 36, 1969, pp. 335–346.

[2] Tesfatsion, L., "Stochastic Dominance and the Maximization of Expected Utility," *Review of Economic Studies*, 43, 1976, pp. 301–15.

Integrating by parts and recalling that the integral of the density functions f and g are the corresponding cumulative distribution functions F and G, respectively (namely $F(x) = \int_a^x f(x)dt$ and G(x) is defined similarly) we obtain:

$$\Delta = [F(x) - G(x)]U(x)\Big|_a^b - \int_a^b [F(x) - G(x)]U'(x)dx .$$

The first term on the right-hand side is equal to zero because for x=b, we have [F(b) -G(b)] = 1 –1 = 0 (recall b is the upper bound and $P(X \leq b) = 1$) and for x =a, we have F(a) = G(a) = 0. Thus, we are left with:

$$\Delta = E_F U(x) - E_G U(x) = \int_a^b [G(x) - F(x)]U'(x)dx = \int_a^b I_1(x)U'(x)dx \qquad (3.1)$$

where $I_1(x) \equiv G(x) - F(x)$.

By the theorem assumption, we have that $I_1(x) \geq 0$ for all values x. As for $U \in \mathbf{U_1}$, we have $U'(x) \geq 0$; the integral is non-negative because an integral of a non-negative number is non-negative. Thus, we conclude that the left-hand side of eq. (3.1) is non-negative or:

$$E_F U(x) \geq E_G U(x) \text{ for all } U(x) \in \mathbf{U_1}.$$

To assure a strict dominance of F over G (namely, to avoid the case that $\Delta = 0$ for all $U \in \mathbf{U_1}$), we need to find at least one $U_0 \in \mathbf{U_1}$ such that $E_F U_0(x) > E_G U_0(x)$. To see that such $U_0(x)$ exists in $\mathbf{U_1}$, recall that by the FSD condition, there is at least one value x_0 for which $F(x_0) < G(x_0)$ and, because F and G are continuous from the right (this is a property of all cumulative distributions), there is $\varepsilon > 0$ such that $F(x) < G(x)$ in the range $x_0 < x \leq x_0 + \varepsilon$. Take the utility function $U_0 = x$ with $U_0'(x) = 1$ (of course $U_0 \in U_1$) to obtain:

$$E_F U_0(x) - E_G U_0(x) = \int_a^b [I_1(x)U_0'(x)]dx$$

$$\geq \int_{x_0}^{x_0+\varepsilon} [I_1(x)]dx > 0$$

where the last inequality holds because $I_1(x) \geq 0$ and $U'(x) \geq 0$ and, in the range $x_0 < x \leq x_0 + \varepsilon$, both are strictly positive.

Note that if $E_FU(x) = E_GU(x)$ for all $U \in \mathbf{U_1}$, then F will not dominate G. However, it is shown above that the condition $F(x_0) < G(x_0)$ guarantees that there will be at least one utility function $U_0 \in \mathbf{U_1}$ such that $E_FU_0(x) > E_GU_0(x)$. Because for all other $U \in \mathbf{U_1}$, $E_FU(x) \geq E_GU(x)$ and, for U_0, a strong inequality holds, we conclude that F D_1G. Thus, we have proved that if $F(x) \leq G(x)$ for all values x and $F(x_0) < G(x_0)$ for some x_0, then $E_FU(x) \geq E_GU(x)$ for all $U \in \mathbf{U_1}$ and there is at least one $U_0 \in \mathbf{U_1}$ such that $E_FU_0(x) > E_GU_0(x)$, hence F D_1G.

Necessity:

We have to prove that:

$E_FU(x) - E_GU(x) \geq 0$ for all $U \in \mathbf{U_1} \Rightarrow F(x) \leq G(x)$ for all x. We prove this claim by a contradiction. Assume that indeed $F(x) \leq G(x)$ for all x but for one value x_1 this condition is violated, namely, $F(x_1) > G(x_1)$. Due to the continuity from the right of the probability distribution function, there must be a range $x_1 \leq x \leq x_1 + \varepsilon$ such that $F(x) > G(x)$. We will show that there is a utility function $U_0 \in \mathbf{U_1}$ such that:

$$E_F\, U_0(x) < E_G\, U_0(x)$$

in contradiction to the assumption that $E_F\, U(x) \geq E_GU(x)$ for all $U \in \mathbf{U_1}$. From this we will conclude that if $E_FU(x) \geq E_GU(x)$ holds for all $U \in \mathbf{U_1}$, such a violation is impossible, namely $F(x_1) > G(x_1)$ is impossible and $F(x) \leq G(x)$ must hold in the whole range.

To prove the necessity, suppose that a violation $F(x_1) > G(x_1)$ does exist for x_1; hence it also exists in the range $x_1 \leq x \leq x_1 + \varepsilon$. Choose the following utility function:

$$U_0(x) = \begin{cases} x_1 & x < x_1 \\ x & x_1 \leq x \leq x_1 + \varepsilon \\ x_1 + \varepsilon & x > x_1 + \varepsilon \end{cases}$$

This utility function is illustrated in Figure 3.4. Because this utility function is monotonic nondecreasing, it belongs to U_1. This utility function is differentiable almost everywhere. Because $U'(x) = 0$ for $x < x_1$ and for $x > x_1 + \varepsilon$ we have (see eq.(3.1)):

$$E_F U_0(x) - E_G U_0(x) = \int_a^b [G(x) - F(x)]\, U_0'(x) dx$$

$$= \int_a^{x_1} [G(x) - F(x)] \cdot 0 dx + \int_{x_1}^{x_1+\varepsilon} [G(x) - (x)] dx + \int_{x_1+\varepsilon}^{b} [G(x) - F(x)] \cdot 0 dx = \int_{x_1}^{x_1+\varepsilon} [G(x) - F(x)] dx < 0.$$

The above inequality holds because $U'(x)$ is equal to 1 in the range $x_1 \le x \le x_1 + \varepsilon$ and, by assumption, $F(x) > G(x)$ in this range. Thus, if there is such a violation in the FSD condition (namely $F(x_1) > G(x_1)$), we have found $U_0 \in \mathbf{U_1}$ such that $E_F U_0(x) < E_G U_0(x)$, or G is preferred to F by this specific preference. Therefore, if it is given that $E_F U(x) \ge E_G U(x)$ for *all* $U \in \mathbf{U_1}$ such a violation is impossible, which proves by the indirect method that: $E_F U(x) - E_G U(x) \ge 0$ for all $U \in \mathbf{U_1} \Rightarrow F(x) \le G(x)$ for all x

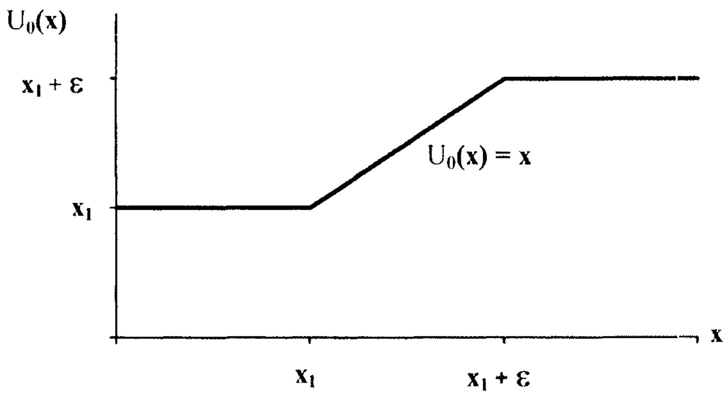
So far we have discussed the conditions under which F dominates G. For reasons of symmetry we claim that G dominates F by FSD (or GD_1F) if $G(x) \le F(x)$ for all x and there is at least one value x_0 where $G(x_0) < F(x_0)$.

c) Graphical exposition of the FSD rule

In Figure 3.5 we see five cumulative distributions representing the feasible set, namely all possible investments. It is easy to show that the FSD efficient set contains F_3 and F_4 and the FSD inefficient set contains F_1, F_2 and F_5. Several conclusions can be drawn from Figure 3.5:

1) FSD dominance requires that the two distributions being compared not cross but they can tangent each other. For example, F_3 dominates F_2 in spite of the fact that there is a range where $F_2(x) = F_3(x)$. Note that F_3 dominates F_2 because the following holds:

Figure 3.4: The Utility Function U



$F_3(x) \leq F_2(x)$ for all values and there is at least one value x_0 for which $F_3(x_0) < F_2(x_0)$ (see x_0 in Figure 3.5).

2) An inefficient investment should not be dominated by all efficient investments. Dominance by one investment is enough. In our example, F_4 does not dominate F_1, F_2 and F_5 (because they intersect) but F_3 dominates all these three investments. Thus, in order to be relegated into the inefficient set, it is sufficient to have one investment that dominates the inefficient investment.

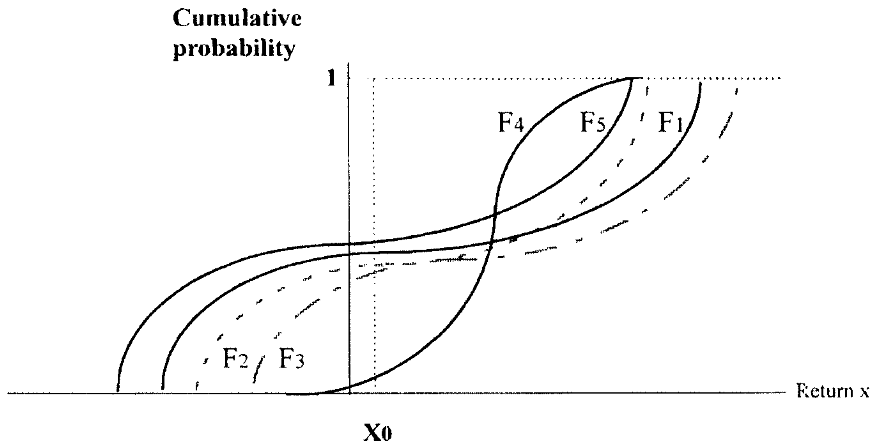
3) In the inefficient set, one investment may or may not dominate another investment in the inefficient set. For example, $F_1 D_1 F_5$ but $F_1 \not{D}_1 F_2$ and $F_2 \not{D}_1 F_1$ (where the slash on D denotes "does *not* dominate"). However, dominance or no dominance within the inefficient set is irrelevant because all investments included in this set are inferior; no investor with preference $U \in U_1$ will select an investment from the inefficient set.

4) An investment within the inefficient set cannot dominate an investment portfolio within the efficient set because if such dominance were to exist then the latter would not be included in the efficient set. For example, if F_2 were to dominate F_3, then F_3 would not be an efficient investment.

5) Finally, all investments within the FSD efficient set must intercept. In our example, F_3 and F_4 intercept. Without such an interception, one distribution would dominate the other and neither would be efficient.

The interception of F_3 and F_4 implies that there is a $U_1 \in \mathbf{U_1}$ such that:

Figure 3.5: The FSD Efficient and Inefficient Sets



$$E_{F_3}U_1(x) > E_{F_4}U_1(x)$$

and there is another utility function $U_2 \in \mathbf{U_1}$ such that:

$$E_{F_4}U_2(x) > E_{F_3}U_2(x)$$

Thus, all investors in the class $U \in \mathbf{U_1}$ will agree on the content of the FSD efficient and inefficient sets; none of them will select their optimum choice from the inefficient set. However, they may disagree on the selection of the optimal investment from the efficient set; one may choose F_3 and another may choose F_4.

d) FSD: A numerical example of FSD

We shall see later on that the FSD investment rule can be extended to discrete random variables; hence we illustrate FSD here with discrete returns.

Let us assume that we have the following three investments denoted by F_1, F_2 and F_3. where x is the given rate of return in percent.[3]

[3] Utility is defined in terms of terminal wealth. Hence, for \$1 investment, a rate of return of –5% implies a terminal wealth of \$0.95. In most examples we will use the rates of return rather than terminal wealth without affecting the analysis. However, in some cases, we need to adhere to terminal wealth. For example, if the utility function is U(x) = log x then it is not

For simplicity of presentation and without loss of generality, we assume that these three investments are the only available investments.

Investment F_1		**Investment F_2**		**Investment F_3**	
x	**p(x)**	**X**	**p(x)**	**x**	**p(x)**
−10%	1/2	−5%	¼	−5%	1/5
30%	1/2	0%	¼	2%	1/5
		10%	¼	15%	1/5
		40%	¼	40%	2/5
		Expected value			
10		$\frac{45}{4}=11\frac{1}{4}$		$\frac{92}{5}=18\frac{2}{5}$	

Which investment dominates which by the FSD rule? In order to answer this question, let us first calculate the cumulative probability of each of these investments:

$$F_1(x)=\begin{cases}0 & x<-10\% \\ 1/2 & -10\le x<30\% \\ 1 & x\ge 30\%\end{cases}$$

$$F_2(x)=\begin{cases}0 & x<-5 \\ 1/4 & -5\le x<0 \\ 1/2 & 0\le x<10 \\ 3/4 & 10\le x<40 \\ 1 & x\ge 40\end{cases}$$

$$F_3(x)=\begin{cases}0 & x<-5 \\ 1/5 & -5\le x<2 \\ 2/5 & 2\le x<15 \\ 3/5 & 15\le x<40 \\ 1 & x\ge 40\end{cases}$$

To find the efficient set and the inefficient set, we need to perform pairwise comparisons. Examination of each pair of distributions shows that the following holds:

defined for x = −5% but it is defined for x = 0.95. We will elaborate on this issue as we proceed.

a) F_1 does not dominate F_2 because for x = –10, $F_1(-10) = 1/2 > F_2(-10) = 0$. Similarly, F_1 does not dominate F_3 because $F_1(-10) = 1/2 > F_3(-10) = 0$.

b) F_2 does not dominate F_1 because for x = 10 we have:

$$F_2(10) = 3/4 > F_1(10) = 1/2.$$

c) F_3 does not dominate F_1 because for x = 15,

$$F_3\,(15) = 3/5 > F_1\,(15) = 1/2.$$

d) F_2 does not dominate F_3 because for the value x = 0,

$$F_2(0) = 1/2 > F_3(0) = 1/5.$$

e) F_1 does not dominate F_3 because for x = 10 we have:

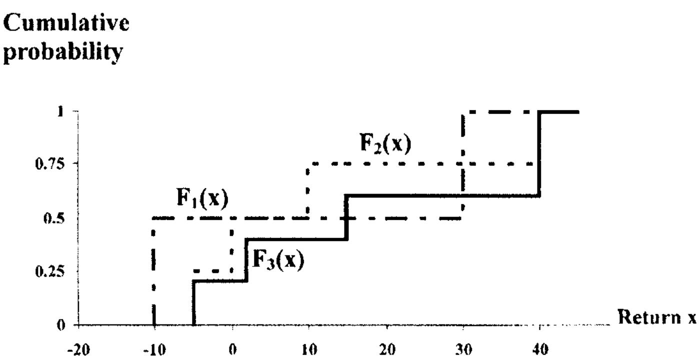
$$F_1(10) = 1/2 > F_3(10) = 2/5.$$

f) Finally, $F_3\ D_1\ F_2$ because for all values x, $F_3(x) \le F_2(x)$ and there is at least one x_0 for which $F_3(x_0) < F_2(x_0)$ (e.g., for $x_0 = 0$, $F_3(0) = 1/5 < F_2(0) = 1/2$).

Thus, with these three possible investments, we have six pairwise comparisons. In this specific case, we obtain only one dominance; $F_3\ D_1\ F_2$. Thus, F_1 and F_3 are included in the efficient set and F_2 is the only investment to be included in the inefficient set.

Figure 3.6 demonstrates that all distributions intercept except for F_2 and F_3. This conforms with our conclusions regarding the partition of the feasible set into the efficient and inefficient sets.

Figure 3.6: The Cumulative Distributions of Three Hypothetical Investments



e) The intuitive explanation of FSD

First-degree stochastic dominance rule implies that if FD_1G then F must be below G (in the weak sense) for the whole range of x. Why is the distribution *below* its competing distribution preferred? To see the intuition of the FSD dominance rule, let us first rewrite the FSD rule as follows:

The condition $F(x) \le G(x)$ for all x can be rewritten as:

$$1 - F(x) \ge 1 - G(x) \text{ for all x.}$$

But, because $F(x) = p(X \le x)$, $1 - F(x) = p(X > x)$. If FD_1G, then for all values x, the probability of obtaining x or a value *higher* than x is larger under F than under G. Such a probability, which would be desired by every investor, explains the dominance of F over G by the FSD criterion. Let us illustrate this "higher than" probability property of FSD. Suppose that under both, F and G, the following outcomes are possible:

X (in %)	**–5**	**0**	**5**	**10**
1–F(x)	1	1	.8	.2
1–G(x)	1	.8	.4	.1

Accordingly, the probability of obtaining (–5%) or more will be the same under both distributions. However, for the other possible outcomes, the probability of obtaining x or more will be higher under F than under G. For example, $P(x \ge 5\%) = .8$ under distribution F and only .4 under distribution G. This probability property of FSD dominance of the event "higher than" would be desired by all investors (recall the monotonicity axiom, see chapter 2), and, therefore, intuitively explains the dominance of F over G for all $U \in \mathbf{U_1}$.

3.3 OPTIMAL RULE, SUFFICIENT RULES AND NECESSARY RULES FOR FSD

An *optimal* decision rule is defined as a decision rule, which is *necessary* and *sufficient* for dominance. The FSD rule is the optimal rule for $U \in \mathbf{U_1}$ because, as proved above, it is a sufficient and a necessary condition for FSD. Mathematically, an optimal rule for the set $U \in \mathbf{U_1}$ is defined as follows:

$$E_FU(x) \ge E_GU(x) \text{ for all } U \in \mathbf{U_1} \Leftrightarrow FD_1G.$$

Namely, FD_1G implies that for every $U \in \mathbf{U_1}$, F is preferred over G by MEUC and the converse also holds: if it is known that for every $U \in \mathbf{U_1}$, F is preferred over G, then $F(x) \leq G(x)$ holds for all values x with a strict inequality for some $x = x_0$.

An optimal rule is the best available rule for a given set of information. Suppose that we know that $U \in \mathbf{U_1}$ but there is no information on the precise slope of U. This means that there is no better rule than the FSD for the information (or assumptions) asserting that $U \in \mathbf{U_1}$ which, in turn, implies that there is no other investment rule that provides a smaller efficient set than the FSD efficient set, and which conforms to MEUC. Thus, an optimal decision rule for all U such that $U \in \mathbf{U_1}$ provides the smallest efficient set for the given information on preferences.

Of course, an arbitrary rule such as F dominates G for all $U \in \mathbf{U_1}$ if $E_F(x) \geq E_G(x)$ can be employed. According to this rule the efficient set includes only one portfolio (we assume, for the sake of simplicity, that no two portfolios have an identical mean), hence it provides an efficient set which would be probably smaller than the FSD efficient set. However, this rule is an arbitrary rule and contradicts the MEUC, hence it should not be used to relegate investments to the inefficient set. An optimal decision rule yields the smallest efficient set provided it does not contradict the MEUC. This is demonstrated in the following example.

Example:

Figure 3.7 demonstrates a feasible set with three hypothetical cumulative distributions F_1, F_2 and F_3. The FSD efficient set includes investments F_2 and F_3. If we employ the maximum expected return rule to rank the available investments, the efficient set will be smaller and consist of only F_3. (It can be seen graphically that $E_{F_3}(x) > E_{F_2}(x)$). The fact that this rule provides a smaller efficient set than the FSD efficient set does not make it a good rule because it contradicts the MEUC by mistakenly relegating portfolio F_2 to the inefficient set. Because there may be $U_0 \in \mathbf{U_1}$ such that:

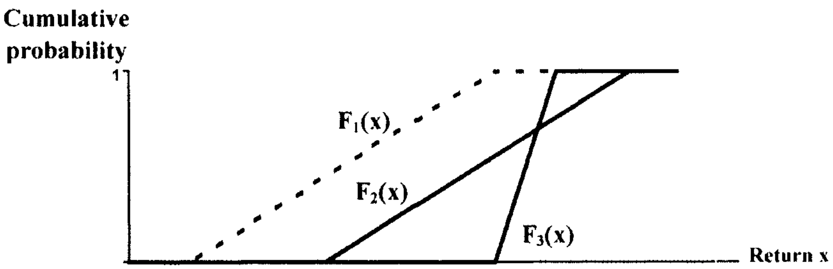
$$E_{F_2}U_0(x) > E_{F_3}U_0(x)$$

Namely, F_2 may be the optimal investment for some legitimate investor with $U \in \mathbf{U_1}$. Thus, $E_F(x) \geq E_G(x)$ does *not* imply that $E_F\,U(x) > E_G\,U(x)$ for all $U \in \mathbf{U_1}$; hence this rule is not a sufficient rule, and hence not an optimal investment rule. As defined above, an optimal rule for a given set of information (e.g., $U \in \mathbf{U_1}$) is a necessary and sufficient rule which, in turn, provides the smallest efficient set without contradicting the MEUC. Let us elaborate on what we mean by "not contradicting the MEUC." In employing FSD, we can safely assume that for any investment relegated to the inefficient set, there is at least one superior investment in the efficient set. Therefore, no investor with $U \in \mathbf{U_1}$ will choose his/her optimal investment from the inefficient set. In this sense, FSD division of the feasible set into the efficient and inefficient set does not contradict the MEUC. By the same logic, when we say that the maximum mean rule contradicts the MEUC, we mean that it may relegate an investment that is optimal for some investors $U \in \mathbf{U_1}$ to the inefficient set. This cannot occur with FSD. Therefore, FSD does not contradict the MEUC, whereas the rule which relies on the means may contradict the MEUC.

a) Sufficient rules

Suppose that there is a sufficient rule for $\mathbf{U_1}$ which we denote by s. If F dominates G

Figure 3.7: Three Hypothetical Cumulative Distributions



by this sufficient rule denoted by FD_sG (where the subscript s denotes a sufficient rule), then $E_FU(x) \geq E_GU(x)$ for all $U \in \mathbf{U_1}$. Formally:

$$FD_sG \Rightarrow E_FU(x) > E_GU(x) \text{ for all } U \in \mathbf{U_1}.$$

Any decision rule with the above property is defined as a sufficient investment rule. Sufficient investment rules do not contradict the MEUC because if F dominates G by a sufficient rule, G will, indeed be inferior to F for all $U \in \mathbf{U_1}$ and, therefore, it should be relegated to the inefficient set. Thus, the results obtained by employing a sufficient rule do not contradict the MEUC. If this is the case, why not employ sufficient rules for the partition of the feasible set into the efficient and inefficient set? The reason is that a sufficient rule may yield an overly large efficient set. In other words, a sufficient rule may not be powerful enough; hence, it may not distinguish between F and G even though F is preferred over G by all $U \in \mathbf{U_1}$. Let us demonstrate a few sufficient rules for $U \in \mathbf{U_1}$.

Sufficient rule 1: F dominates G if $\text{Min}_F(x) \geq \text{Max}_G(x)$.

This is a sufficient rule because whenever it holds, FD_1G (namely $FD_sG \Rightarrow FD_1G$) which, in turn, implies that $E_FU(x) \geq E_GU(x)$ for all $U \in \mathbf{U_1}$.

Example:

Assume the following three investments F, G and H:

F		G		H	
X	p(x)	x	p(x)	x	p(x)
5	1/2	2	3/4	2	¾
10	1/2	4	1/4	6	¼

$Min_F(x) = 5 \geq Max_G(x) = 4$; hence by this sufficient rule, F dominates G. Indeed, $FD_sG \Rightarrow FD_1G$. To see this, note that in this example (regardless of the probabilities), $F(x) \leq G(x)$ which implies that $E_FU(x) \geq E_GU(x)$ for all $U \in \mathbf{U_1}$.

In this case, both the sufficient rule and the optimal rule reveal that F is preferred to G. However, suppose that instead of 4 with a probability 1/4 we have 6 with a probability of 1/4, as given in investment H. In this case we obtain:

$$Min_F(x) = 5 < Max_H(x) = 6$$

and, by this rule $F\cancel{D}_sH$. Yet FD_1H because $F(x) \leq H(x)$ for all values x (see Figure 3.8a and 3.8b); hence, this sufficient rule and the FSD rule do not provide the same dominance relationship (i.e., the same efficient set). Because the FSD rule is optimal for $U \in \mathbf{U_1}$ we can safely conclude that the above sufficient rule is not powerful enough to reveal the preference of F over H.

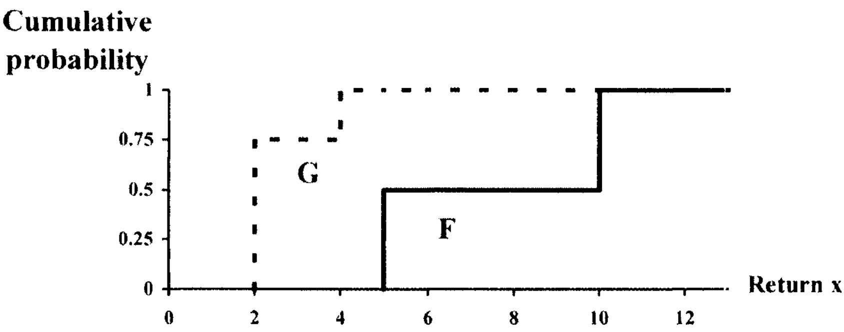
Sufficient rule 2: F dominates G if $F(x) \leq G(x)$ for all x

and there is at least one value, x_0, such that:

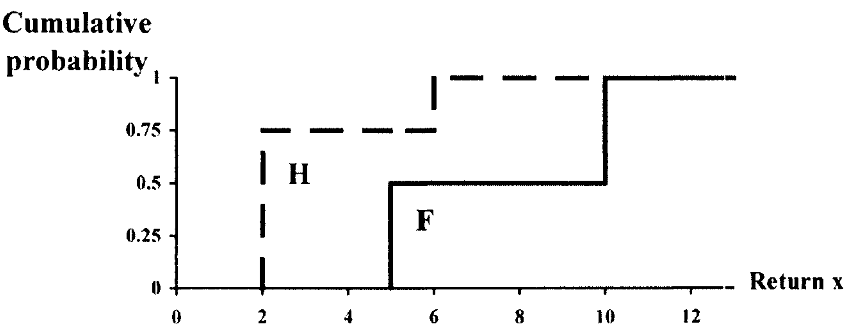
$$F(x_0) + a \leq G(x_0)$$

where a is some fixed positive number. Suppose that we have a specific sufficient rule with a = 4/5. Obviously, if the sufficient rule holds, it implies that $F(x) \leq G(x)$ for all x; hence FD_1G. Therefore, by the definition of a sufficient rule, this is a sufficient rule for $U \in \mathbf{U_1}$. However, if FDsG by this sufficient rule, it is still possible that FD_1G and the sufficient rule will not be powerful enough to unveil this dominance. In the two examples demonstrated in Figure 3.8, there is no value x such that $F(x) + 4/5 \leq G(x)$; hence by this sufficient rule, neither F nor G dominates the

Figures 3.8: The Rule $Min_F(x) \geq Max_G(x)$ is a Sufficient Rule for All $U \in \mathbf{U_1}$



3.8a: $Min_F(x)=5>Max_G(x)=4$



3.8b: $Min_F(x)=5<Max_H(x)=6$ Yet FD_1G

other. Yet, FD_1G (and FD_1H) and the sufficient rule is not strong enough to reveal this dominance. If we select another value a, say, a = 1/4, both the sufficient rule and FSD reveals that F dominates G because for say x = 9, we have

$$F(9) + 1/4 = 1/2 + 1/4 < G(9) = 1 \text{ (and similarly, } F(9) + 1/4 = 3/4 < H(9) = 1).$$

Thus, if we have an optimal rule, it should always be used in investment screening. However, in some cases we do not have optimal rules. In such cases we have to use sufficient rules for investment screening. Fortunately, for the set $U \in \mathbf{U_1}$, we have an optimal rule, the FSD rule, and there is no need to use sufficient rules such as those cited above.

b) Necessary rules

Suppose that $E_F U(x) \geq E_G U(x)$ for all $U \in \mathbf{U_1}$, implies that some condition must hold (e.g., $E_F(x) \geq E_G(x)$). Then we call this condition a *necessary* rule for dominance. We discuss below three necessary conditions (or rules) for dominance in $\mathbf{U_1}$. There are many more necessary rules but these three are the most important ones.

Necessary rule 1:

The Means.

If FD_1G, then the expected value (or the mean return) of F must be greater than the expected value of G. Hence, $E_F(x) > E_G(x)$ is a necessary condition for FSD. Formally:

$$FD_1G \Rightarrow E_F(x) > E_G(x).$$

Proof:

The difference in the mean returns is given by:

$$E_F(x) - E_G(x) = \int_a^b [f(x) - g(x)]x dx.$$

Integrating by parts yields:

$$E_F(x) - E_G(x) = [F(x) - G(x)]x\Big|_a^b - \int_a^b [F(x) - G(x)]dx.$$

The first term on the right-hand side is zero (because $F(b) - G(b) = 1 - 1 = 0$ and $F(a) = 0$ and $G(a) = 0$).

Hence:

$$E_F(x) - E_G(x) = \int_a^b [G(x) - F(x)]dx\ .$$

It is given that FD_1G; therefore, $F(x) \leq G(x)$ with at least one strict inequality. Hence, $G(x) - F(x) \geq 0$ with a strong inequality for at least one value x. Because cumulative distributions are continuous from the right, there is a range for which there is a strict inequality; hence the integral on the right-hand side is positive which implies that $E_F(x) > E_G(x)$. Thus, the superior investment by FSD must have a larger mean than that of the inferior investment. Note that the difference in expected value is equal to the difference in expected utility for the specific linear utility function $U_0(x) = x$.

Necessary rule 2:

Geometric Means:

If FD_1G, then the geometric mean of F must be larger than the geometric mean of G. Formally:

$FD_1G \Rightarrow \bar{x}_{geo.}(F) > \bar{x}_{geo.}(G)$, where geo. stands for geometric and the bar over x denotes mean value.

Proof:

The geometric mean is defined only for positive numbers. Suppose that we have a risky project given by the distribution (x_i, p_i) where i = 1, 2,..,n. The geometric mean denoted by $\bar{X}_{geo.}$ will be defined as follows:

$$\bar{x}_{geo.} = x_1^{p_1} x_2^{p_2} \ldots x_n^{p_n} \text{ or } \bar{x}_{geo.} = \prod_{i=1}^{n} x_i^{p_i} \text{ and } x_i \geq 0 \text{ for all i}$$

where $x_i = (1+R_i)$ and R_i is the ith rate of return. Thus, R_i can be negative or positive, but $x_i \geq 0$. Thus, x_i is the terminal wealth of \$1 invested (see footnote 3). If we take the logarithms from both sides, we have:

$$\log(\bar{x}_{geo.}) = \Sigma p_i \log(x_i) = E\,(\log(x)).$$

We prove below that if FD_1G, then $\bar{x}_{geo.}$ (F) ≥ $\bar{x}_{geo.}$ (G). To see this, recall that, by eq. (3.1), for every utility function $U \in \mathbf{U_1}$ we have:

$$E_F U(x) - E_G U(x) = \int_a^b [G(x) - F(x)]U'(x)dx.$$

For the utility function U(x) = log(x), this formula shows the following specific relationship:

$$E_F \log(x) - E_G \log(x) = \log_F(\bar{x}_{geo}) - \log_G(\bar{x}_{geo}) = \int_a^b [G(x) - F(x)](\partial \log x / \partial x)dx$$

where $\partial \log(x)/\partial x = U'(x)$.

It is given that $F(x) \leq G(x)$ and that there is at least one value x_0 for which $F(x_0) < G(x_0)$. Because with $U(x) = \log x$ we have $U'(x) = \partial \log(x) / \partial x = 1/x > 0$ (because $x = 1+R \geq 0$), and $F(x_0) < G(x_0)$ implies that $F(x) < G(x)$ for some range $x_0 \leq x \leq x_0 + \varepsilon$, the right-hand side will be positive; hence the left-hand side will be positive. However, because the left-hand side is the difference in the logarithms of the geometric means, and the logarithm is a monotonic increasing function of the geometric means, we can conclude that:

$$FD_1G \Rightarrow \bar{x}_{gco.}(F) > \bar{x}_{gco.}(G)$$

In the above two examples, we use the *linear utility* function and the *logarithmic utility* function and employ the following logic to determine the necessary conditions for dominance: F is preferred to G by some decision rule for these specific functions (e.g., $E_F(x) > E_G(x) \Leftrightarrow E_FU_0(x) > E_GU_0(x)$) and, because $U_0 \in \mathbf{U_1}$, then for any pairwise comparison, if FD_1G then also $E_FU_0(x) > E_GU_0(x)$ must hold. Therefore, the necessary condition (e.g., $E_F(x) \geq E_G(x)$) must hold.

The same logic can be applied in analyzing other specific utility functions $U_0(x) \in \mathbf{U_1}$ to establish more necessary rules for FSD. We will elaborate on more necessary rules later on when we examine undominated portfolios (diversification) of assets.

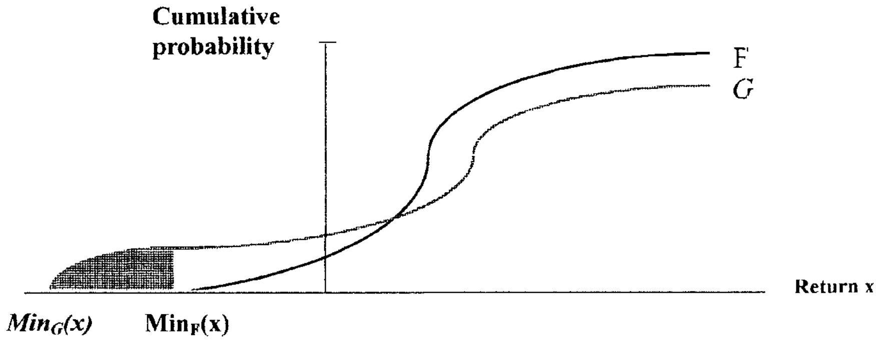
Necessary rule 3: the "Left Tail" condition.

If FD_1G, it is necessary that:

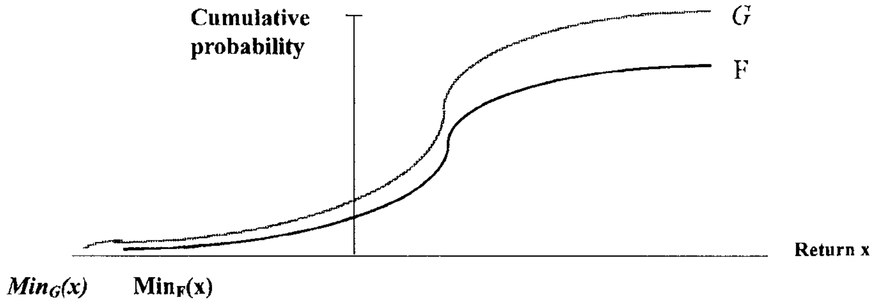
$$\text{Min}_F(x) \geq \text{Min}_G(x).$$

This means that distribution G starts to accumulate area (or probability) before distribution F. This is called the "left tail" condition because the cumulative distributions imply that G has a thicker (in the weak sense) left tail. Note that if the necessary condition does not hold and $\text{Min}_F(x) < \text{Min}_G(x)$

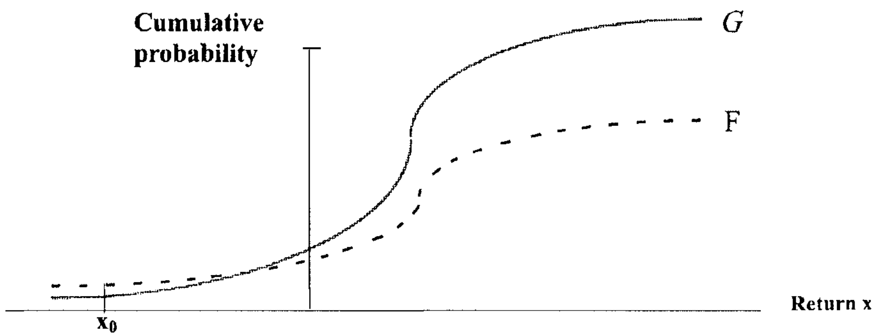
Figure 3.9. The "Left Tail" Necessary Condition for FSD



3.9a: The Necessary Condition Holds But There Is No FSD



3.9b: The Necessary Condition Holds and There Is FSD



3.9c: The Necessary Condition Does Not Hold Hence There Is No FSD

$\equiv$ k, there will be a value $x_0 = 0$ such that $F(x_0) > G(x_0) = 0$ and, therefore, F cannot dominate G. Thus, if FD_1G, it is necessary to have that $Min_F(x) \geq Min_G(x)$. Figure 3.9 demonstrates this necessary condition. In Figure 3.9a $Min_F(x) > Min_G(x)$; hence the necessary condition for FSD is intact. Note that the "left tail" condition does not guarantee dominance by FSD. If, for example, F intercepts G later on (for larger values x), then FSD would not hold (see F and G in Figure 3.9a) even though the necessary condition holds. Thus, the left-tail condition is a necessary but not a sufficient condition for dominance. In Figure 3.9b, $Min_F(x) > Min_G(x)$; hence G has a "thicker tail" and this necessary condition for the dominance of F over G holds. Also, F and G do not intercept later on; hence FD_1G. In Figure 3.9.c, F has a thicker tail than G. Thus, the necessary condition for the dominance of F over G does not hold and, therefore, $F \not{D}_1 G$. Thus, even though $F(x) < G(x)$ for most of the range x, because of its thicker left tail, we have that for, say, $x = x_0$ (see Figure 3.9c), $F(x_0) > G(x_0)$; hence $F \not{D}_1 G$ and there is no FSD of F over G.

Let us go back to the previous example focusing on F and G in Figure 3.8a. In this case, FD_1G. Let us verify that the three necessary conditions hold:

First G has a thicker left tail (or $Min_F(x) = 5 > Min_G(x) = 2$). Secondly, the mean return necessary condition for dominance also holds because:

$$E_F(x) = 1/2 \cdot 5 + 1/2 \cdot 10 = 7.5 > 3/4 \cdot 2 + 1/4 \cdot 4 = 2.5 = E_G(x).$$

Finally, the condition corresponding to the geometric means also holds because:

$$5^{1/2} \cdot 10^{1/2} = 7.07 > 2^{3/4} \cdot 4^{1/4} \cong 2.38$$

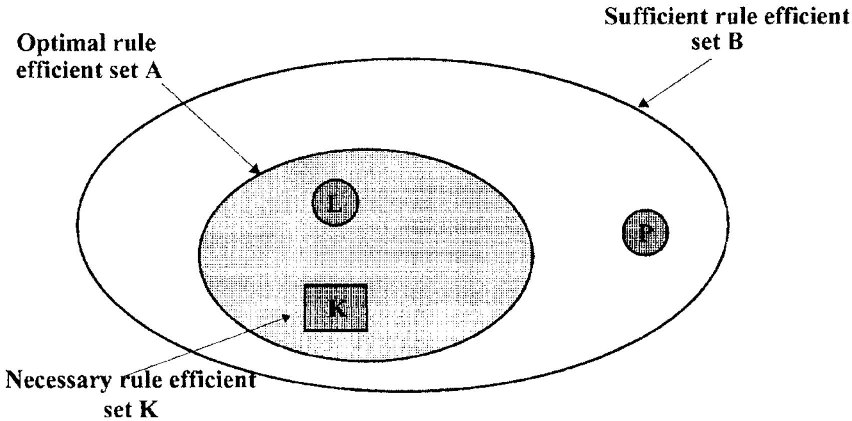
(where here outcome stands for 1 + rate of return).

3.4 TYPE I AND TYPE II ERRORS WHEN SUFFICIENT RULES OR NECESSARY RULES ARE NOT OPTIMAL FOR INVESTMENT SCREENING.

Suppose that we do not have an optimal investment rule and, therefore, we employ a sufficient investment decision rule. Because we are not employing an optimal rule, we may commit an error which we call here a Type I error. If we employ a necessary rule for investment screening, we may commit another error, which we call a Type II error. Figure 3.10 demonstrates these two possible errors. The area given by circle A represents the FSD efficient set when the optimal FSD rule is employed. The efficient set induced by the sufficient rule is given by the area inside circle B. We see that the efficient set includes investments, which are in B but not in A, namely inefficient investments, such as investment P. This is not a serious error. Recall that we employ decision rules for the first stage screening, namely for eliminating some investments from the feasible set. In the second stage each investor selects his/her optimum portfolio (according to personal preference) from the efficient set. The Type I error is not a serious error because if the investor selects wisely from among the investments in area B, he/she will chose one from inside area A; hence in the final stage, no one will invest in an investment such as P and no harm will have been done. This is why Type I error is not considered to be serious. The main drawback of employing a sufficient rule, which is not optimal, is that the partial ordering of the investments may result in an ineffective result, namely a relatively large efficient set. A decision rule, which induces a relatively large efficient set, is an *ineffective decision rule*. In the extreme case, the sufficient rule may be very ineffective, yielding an efficient set, which is equal to the feasible set. In such cases, none of the investments are relegated to the inefficient set and the initial screening stage will have been worthless.

A necessary rule (which is not a sufficient rule and hence not an optimal rule) for investment screening may result in a Type II error: We may relegate an investment which maximizes the expected utility of some legitimate utility function $U \in U_1$ to the inefficient set. The efficient set derived by employing a necessary rule for FSD (e.g., one portfolio dominates the other if it has a higher geometric mean) is represented by the area given by a subset of the area A, say, by the K inside the circle A (see Figure 3.10). Here, for example, investment L is relegated to the inefficient set; hence it is not presented to the investors by the investment consultant for consideration even though it is possible that for some utility function, investment L may maximize the expected utility. The investment consultant presents the investors only with those investments that pass the first screening. Therefore, a Type II error is considered to be serious: Investors who are not presented with investment L, will not choose it even though it may be the best investment for some of them. In conclusion, Type I errors are not necessarily serious and, therefore, there is no harm in using sufficient rules. However, necessary rules (which are not sufficient) should not be employed for the first stage of investment screening because the consequences of Type II errors can be serious.

Figure 3.10: Type I and Type II errors induced by employing sufficient or necessary rules which are not optimal



The question that arises, therefore, is when should necessary rules which are not sufficient be employed? A necessary rule can be employed to facilitate the pairwise comparisons of potential investments. To illustrate, suppose that we have two distributions F and G. By the FSD rule, we need to examine whether F dominates G and, if no dominance is found, we also have to check whether G dominates F. However, suppose it is given that $E_F(x) > E_G(x)$. In such a case, we compare only whether F dominates G, and there is no need to examine whether G dominates F; G cannot dominate F because it has a lower expected return. Thus, by examining the necessary rule of the means, we reduce the number of comparisons from 2 to 1. The FSD is a fairly simple rule. Therefore, when FSD is employed, the benefit gained by using necessary rules to reduce the number of comparisons is limited. However, as will be shown, when more complicated rules are employed, and when we have a large number of investments in the feasible set, the benefit of reducing the number of pairwise comparisons by using necessary rules in the screening process becomes highly apparent.

3.5 SECOND DEGREE STOCHASTIC DOMINANCE (SSD)

a) Risk aversion

So far, the only assumption that we have made is that $U \in \mathbf{U}_1$, namely, $U' \geq 0$. There is much evidence that most, if not all, investors are probably risk averters. Therefore, let us develop a decision rule appropriate for all risk averters. In all the discussions below, we deal only with non-decreasing utility functions, $U \in \mathbf{U}_1$, and we add the assumption of risk aversion. Let us first define risk aversion.

Risk aversion can be defined in the following alternative ways:

1. The utility function U has a non-negative first derivative and a non-positive second derivative ($U' \geq 0$ and $U'' \leq 0$) and there is at least one point at which

$$U' > 0 \text{ and one point at which } U'' < 0.$$

2. If we take any two points on the utility function and connect them by a chord, then the chord must be located either below, or on, the utility function and there must be at least one chord which is located strictly below the utility function. Figure 3.11 demonstrates a risk averse utility function: the chords ab and cd are located on the utility function, and the chord be is located below it.

3. The expected utility is smaller or equal to the utility of the expected return. To be more specific, suppose that we have an investment which yields x_1 with probability p and x_2 with a probability 1–p: Then:

$$U(E(x)) = U(p\,x_1 + (1-p)\,x_2) \geq pU(x_1) + (1-p)\,U(x_2) = E\,U(x)$$

   and there will be at least one possible hypothetical investment for which this inequality is strong. This property of concave functions (namely $U' \geq 0$ and $U'' \leq 0$) is called *Jensen's Inequality:* accordingly, for any concave function, the following will hold:

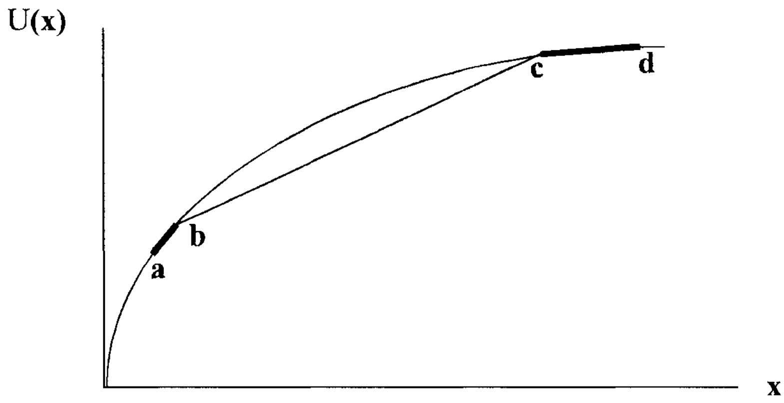
$$U\,(Ex) \geq EU(x).$$

   The chord definition of risk is appropriate for a random variable which can take only two values. In this sense, *Jensen's Inequality* is more general than the chord definition of risk aversion because it holds also for random variables (returns) which can take more than two values.

4. A risk averter will not play a *fair game.* A fair game is defined as a game in which the price of a ticket to play the game is equal to the expected prize. For example, if we roll a die and the number that appears on the top corresponds to the prize in dollars and the ticket to play the game costs \$3.50, it will be a fair game because $E(x) = \sum_{x=1}^{6} \frac{\$x}{6} = \$3.5$. Risk averters will never play a fair game.

5. Risk averters will be ready to pay a positive risk premium to insure their wealth (e.g., purchase fire insurance for their house). Let us elaborate. Define by w the investor's certain wealth except for the house (say cash in the bank) and by x, the house which is exposed to the possible risk of fire. If insurance is not bought, then the value of the house will be a random variable, x and the expected utility

Figure 3.11: Risk Averse Utility Function



will be EU (w+x). By *Jensen's Inequality*, for any concave U we obtain:

$$EU\ (w+x) \leq U\ (w + Ex).$$

Therefore, there is a value $\pi \geq 0$ such that:

$$EU\ (w+x) = U\ (w+Ex - \pi).$$

π which solves this equation is called the *risk premium* or the maximum amount, by which *on average*, an individual is willing to reduce his expected wealth in order to rid himself of the risk. We emphasize that this is the average or expected payment. For example, if an individual insures his home and fire does not break out, he pays the insurance company. If a fire does break out, the insurance firm pays the homeowner. Thus, π represents the expected payment of the insured to the insurance firm given the probability of a fire breaking out. The risk premium is also the average gross profit of the insurance firm, namely, the profit before taxes and other expenses apart from payments to the insured. Thus, the assertion that $\pi \geq 0$ for any risky investment is identical to the assertion that U is concave; hence a non-negative π can also be used as a definition of risk aversion.

All these definitions are consistent except for definition 1 which may not hold for a non-differential utility function such as the one given in Figure 3.4. The other definitions hold for both differential and non-differential utility function; hence they are more general.

We will use all these risk-aversion definitions interchangeably. We define the set of all concave utility functions corresponding to risk aversions by $\mathbf{U_2}$. Of course, $\mathbf{U_2} \subseteq \mathbf{U_1}$ when $\mathbf{U_1}$ corresponds to FSD. Before turning to the risk averters' investment rule, let us be reminded that although all investors would agree that $U \in \mathbf{U_1}$, not all would agree that $U \in \mathbf{U_2}$. Nevertheless, there is much evidence that for virtually all investors, $U \in \mathbf{U_2}$. The fact that cost of capital of most firms is generally higher than the riskless interest rate indicates that stockholders are risk averse and require a risk premium. Similarly, the long-run average rate of return on common stock in the U.S. for the period 1926–1997 is about 12% per year whereas the annual rate of return on Treasury Bills is about 3%. The 9% difference represents the risk premium again reflecting the fact that most stockholders are risk averters.[4] Thus, although the consensus regarding the assumption $U \in \mathbf{U_2}$ is incomplete, it is generally accepted that most investors are risk averters and, therefore, it is worthwhile to develop a decision rule for all $U \in \mathbf{U_2}$.

In the next theorem, we provide a decision rule for all $U \in \mathbf{U_2}$. Once again, we first assume continuous random variables and the results are extended to discrete random variables afterwards.

b) The SSD investment decision rule

Theorem 3.2 provides the SSD investment rule.

Theorem 3.2:

Let F and G be two investments whose density functions are f(x) and g(x), respectively. Then F dominates G by second degree stochastic dominance (SSD) denoted by FD_2G for all risk averters if and only if:

$$I_2(x) \equiv \int_a^x [G(t) - F(t)]dt \geq 0$$

for all $x \in [a,b]$ and there is at least one x_0 for which there is a strict inequality. This Theorem can also be stated as follows:

$$\int_a^x [G(t) - F(t)]dt \geq 0 \quad \Leftrightarrow \quad E_F\, U(x) - E_G\, U(x) \geq 0$$

for all x with at least one strict inequality for some x_0		for all $U \in \mathbf{U_2}$ with at least one $U_0 \in \mathbf{U_2}$ for which there is a strict inequality

[4] See R. Ibbotson and Associate, *Stocks, Bonds, Bills and Inflation*. (Chicago, IL.; Ibbotson Associate various yearbooks).

We will first prove the sufficiency of this theorem and then, the necessity. This will be followed by an intuitive explanation and graphical demonstration of the SSD.

Sufficiency

We have to prove that

$$I_2(x) \equiv \int_a^x [G(t) - F(t)]dt \geq 0 \Rightarrow E_F\, U(x) - E_G\, U(x) \geq 0 \qquad \text{for all } U \in \mathbf{U_2}$$

By equation (3.1) we have seen that:

$$E_F\, U(x) - E_G\, U(x) = \int_a^b [G(x) - F(x)]U'(x)\, dx.$$

Integrating the right-hand side by parts yields:

$$E_F U(x) - E_G U(x) = U'(x) \int_a^x [G(t) - F(t)]\, dt \Big|_a^b - \int_a^b U''(x)(\int_a^b [G(t) - F(t)]dt)dx.$$

Because G(a) = F(a) = 0, this can be simplified as follows:

$$E_F U(x) - E_G U(x) = U'(b) \int_a^x [G(t) - F(t)]dt - \int_a^b U''(x)(\int_a^b [G(t) - F(t)]dt)dx. \quad (3.2)$$

By the sufficiency assumption, we have $\int_a^x [G(t) - F(t)]dt \geq 0$ for all x, and particularly for x=b and, because $U'(b) \geq 0$, the first term on the right-hand side is non-negative. The second term is also non-negative because, by the risk aversion assumption, U is concave; hence $U'' \leq 0$, and $\int_a^x [G(t) - F(t)]dt \geq 0$ by the assumption.

Thus, if the integral condition $I_2(x) \geq 0$ holds, then $E_F U(x) \geq E_G U(x)$ for all $U \in \mathbf{U_2}$. Finally, we need to show that there is some $U_0 \in \mathbf{U_2}$ such that $E_F U_0(x) > E_G U_0(x)$; namely, that a strict inequality holds for $U_0(x)$ (otherwise it is possible that $E_F U(x) = E_G U(x)$ for all $U \in \mathbf{U_2}$ and neither F nor G dominates the other). To show this, recall that Theorem 3.2 requires at least one strict inequity for some x_0; hence there is x_0 such that $\int_a^{x_0} [G(t) - F(t)]dt > 0$.

Because of the continuity of $\int_a^x [G(t)-F(t)]dt$, there is $\varepsilon > 0$ such that $\int_a^x [G(t)-F(t)]dt > 0$ for all x, $x_0 - \varepsilon \le x \le x_0 + \varepsilon$. Choose a utility function such that $U' > 0$ and $U'' < 0$ in the range of $x_0 - \varepsilon \le x \le x_0 + \varepsilon$, say $U_0(x) = -e^{-x}$. For this utility function we have:

$$E_F(U_0) - E_G(U_0) = U_0'(b)\int_a^b [G(t)-F(t)]dt + \int_a^x -U_0''(x)\int_a^x [G(t)-F(t)]dtdx$$

$$\ge \int_a^{x_0-\varepsilon} -U_0''(x)\int_a^x [G(t)-F(t)]dtdx + \int_{x_0-\varepsilon}^{x_0+\varepsilon} -U_0''(x)\int_a^x [G(t)-F(t)]dtdx$$

$$+ \int_{x_0+\varepsilon}^{b} -U_0''(x)\int_a^x [G(t)-F(t)]dtdx +$$

$$\ge \int_{x_0-\varepsilon}^{x_0+\varepsilon} -U_0''(x)\int_a^x [G(t)-F(t)]dtdx = \int_{x_0-\varepsilon}^{x_0+\varepsilon} -U_0''(x)I_2(x)dx > 0.$$

Note that because SSD holds, all terms are non-negative and, due to the fact that $I_2(x) > 0$ and $U''(x) < 0$ in the range $x_0 - \varepsilon \le x \le x_0 + \varepsilon$, the last term is strictly positive. This guarantees that there is at least one $U_0 \in \mathbf{U_2}$ such that $E_F U_0(x) > E_G U_0(x)$; hence we can conclude that F dominates G for all $U \in \mathbf{U_2}$. Note that the strict inequality is needed to avoid the trivial case where G and F are identical.

Necessity

We have to prove that:

$$E_F U(x) - E_G U(x) \ge 0 \text{ for all } U \in \mathbf{U_2} \text{ implies that } \int_a^x [G(t)-F(t)]dt \ge 0$$

for all $x \in [a,b]$.

Once again, we employ the indirect method. The logic of this proof is as follows: Suppose that the integral condition is violated for some value x_0. Then we can show that there is $U_0 \in \mathbf{U_2}$ for which $E_F\, U_0(x) < E_G\, U_0(x)$ which is in contradiction to the assumption of the Theorem asserting that $E_F U(x) \ge E_G U(x)$ for all $U \in \mathbf{U_2}$. Hence, if we wish the inequality $E_F U(x) > E_G\, U(x)$ to hold for all $U \in \mathbf{U_2}$ (including the U_0), then the violation

$I_2(x_0) = \int_a^{x_0} [G(t) - F(t)]dt < 0$ cannot hold. To see this, suppose that

$I_2(x_0) < 0$, and select the function U_0 as follows:[5]

[5] The function $U_0(x)$ is not differentiable at $x = x_0$. However, for this specific function, we can always obtain a differentiable utility function that is arbitrarily close to U_0. To see this, let x_0 be such that

$$\int_a^{x_0} [G(x) - F(x)]dx = \delta < 0.$$

Choose $\varepsilon > 0$ such that:

(a) $$\int_{x_0-\varepsilon}^{x_0} |G(x) - F(x)|dx < \frac{|\delta|}{4}$$

(b) $$\int_a^{x_0-\varepsilon} [G(x) - F(x)]dx < \frac{\delta}{2}.$$

Define U_0 as follows:

$$U_0(x) = \begin{cases} (x - x_0) + \frac{\varepsilon}{2} & \text{for} \le x_0 - \varepsilon \\ -\frac{(x - x_0)^2}{2\varepsilon} & \text{for } x_0 - \varepsilon \le x \le x_0 \\ 0 & \text{for } x > x_0 \end{cases}$$

$U'_0(x)$ is a continuous function such that $U'_0(x) > 0$ for $x < x_0$ and $U'(x) = 0$ for $x \ge x_0$. Obviously, $U_0 \in \mathbf{U_2}$ and U'_0 is a continuous function. Then, for this specific $U_0(x)$, by eq. (3.1), we have:

$$E_F(U_0(x)) - E_G(U_0(x)) = \int_a^b [G(x) - F(x)]U'(x)dx =$$

$$= \int_a^{x_0-\varepsilon} [G(x) - F(x)]dx + \int_{x_0-\varepsilon}^{x_0} [G(x) - F(x)](-\tfrac{(x-x_0)}{\varepsilon})dx + \int_{x_0}^b [G(x) - F(x)] \cdot 0dx$$

But, by assumption $\int_a^{x_0-\varepsilon} [G(x)-F(x)]dx < \frac{\delta}{2}$, and $\left| \int_{x_0-\varepsilon}^{x_0} [G(x)-F(x)](-\frac{(x-x_0)}{\varepsilon})dx \right| < \frac{|\delta|}{4}$

(Note that $\frac{x_0 - x}{\varepsilon} < 1$ in this range).

Therefore, we have that:

$$E_F(U_0(x)) - E_G(U_0(x)) < \tfrac{\delta}{4} < 0.$$

$$U_0(x) = \begin{cases} x & x \le x_0 \\ x_0 & x > x_0 \end{cases}$$

Obviously, $U_0 \in \mathbf{U_2}$ (see the above definition of risk aversion). Then, for this specific $U_0(x)$, by eq. (3.1), we have:

$$E_F U_0(x) - E_G U_0(x) = \int_a^b [G(x) - F(x)]U'(x)dx = \int_a^{x_0} [G(x) - F(x)]dx + \int_{x_0}^b [G(x) - F(x)] \cdot 0dx.$$

(Note that $U'(x) = 1$ in the range $x \le x_0$ and $U'(x) = 0$ for $x > x_0$). However, because by assumption $\int_a^{x_0} [G(x) - F(x)]dx < 0$, $E_F U_0(x) < E_G U_0(x)$ which is in contradiction to the assumption that FD_2G. Thus, if $E_F U(x) \ge E_G U(x)$ for *all* $U \in \mathbf{U_2}$, then $I_2(x) \ge 0$ for *all* x (no violation is allowed!), which completes the proof.

c) Graphical exposition of SSD

The SSD integral condition ($I_2(x) \ge 0$) for dominance implies that the area enclosed between the two distributions under consideration should be non- negative up to every point x. When we examine whether F dominates G, whenever F is below G, we denote the area enclosed between the two distributions by "+"area, and whenever G is below F, we denote the area enclosed between the two distributions by "–" area. When we examine whether G dominates F, the opposite area signs are used.

Figures 3.12a and 3.12b illustrate two cumulative probability distributions F and G. SSD dominance may exist irrespective of the number of times that the two distributions intersect. Let us look first at Figure 3.12a. Can G dominate F by SSD? The answer is negative: Up to x_1, the integral condition does not hold:

$$I_2(x_1) = \int_a^{x_1} [F(x) - G(x)]dx < 0$$

Can F dominate G? As we can see from Figure 3.12a, for any value x up to x_1, the integral $\int_a^x [G(t) - F(t)]dt > 0$; hence such a dominance is possible. However, to have SSD dominance of F over G, $I_2(x)$ has to be non-negative for every value x, and this does not occur in Figure 3.12a. For example, in Figure 3.12a, up to x_2 we have:

$$\int_a^{x_1} [G(x) - F(x)]dx + \int_{x_1}^{x_2} [G(x) - F(x)]dx < 0$$

where the first area is positive and the second is negative. F dominates G by SSD only if the "+" area is greater than the "–" area. The graph presented in Figure 3.12a clearly reveals that in this specific example, this does not occur; hence $\int_a^{x_2}[G(x)-F(x)]dtx<0$ and, therefore, in our example, F does not dominate G by SSD either. Thus, neither of the two distributions given in Figure 3.12a dominates the other by SSD. Figure 3.12b demonstrates F dominance over G by SSD: Here the integral $\int_a^{x}[G(t)-F(t)]dt>0$ for all values x, and there is at least one strong inequality, say at x_1. From these two graphs we can conclude that F dominates G if for any negative area (for example the area between x_2 and x_3, see Figure 3.12b), there is a positive area located to the left of x_2 which is equal or larger than the negative area.

In Figure 3.12 there are only a few intersections of F and G. Let us generalize the SSD condition for a larger number of intersections between F and G. By the integral condition for any negative area (i.e., a range $x_2 \le x \le x_3$, See Figure 3.12) there must be positive areas located earlier such that the sum of the positive areas is larger than the sum of all negative areas accumulated up to x_3 (see Figure 3.12). Let us elaborate. Denote by S^- and S^+ the negative and positive areas (in absolute values), respectively. We employ the absolute values of the areas for all area comparisons (for the sake of brevity, this will not be mentioned again). Suppose that F and G intersect n times, n=1, 2,... (If they do not intersect, then there is FSD which implies SSD). We order all the areas enclosed between F and G from the lowest intersection points of F and G to the highest intersection points of F and G as follows:

$$S_1,S_2,S_3,\ldots\ldots\ldots.S_n$$

where S_i can be a positive area or a negative area. Suppose that S_i^- is the first negative area. Then, by the SSD rule, for the first negative area S_i^-, we must have that:

$$S_i^- \le \sum_{j=1}^{i-1} S_j^+ .$$

Now suppose that i is the first negative area and m is the second negative area. Then, by the SSD, we must have that:

$$S_i^- + S_m^- \le \sum_{j=1}^{m-1} S_j^+ .$$

In general, for the k^{th} negative area, we must have that:

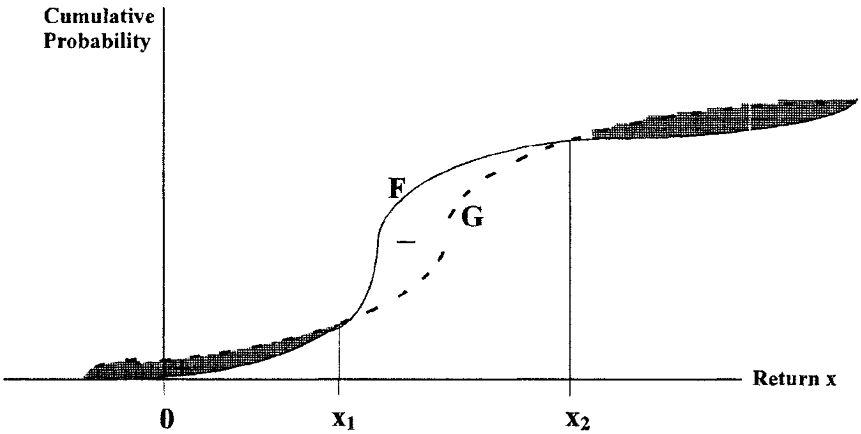
$$\sum_{i=1}^{k} S_i^- \leq \sum_{j=1}^{l} S_i^+ .$$

where there are l positive areas before the k^{th} negative area.

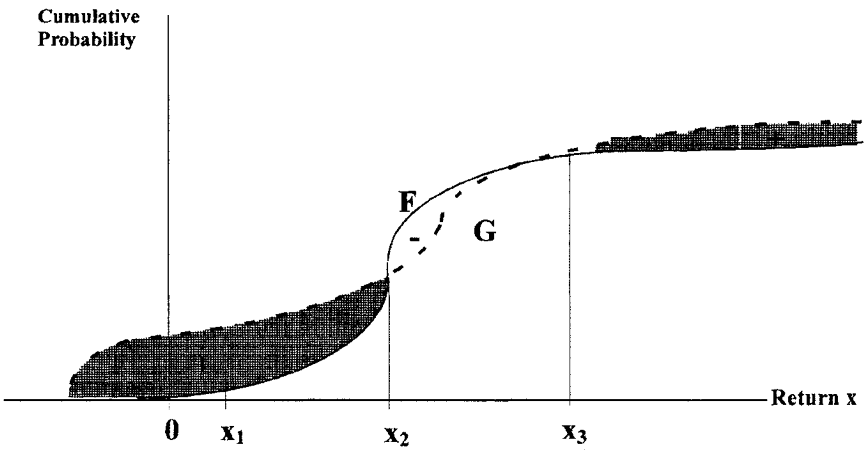
Namely, up to any point corresponding to a negative area, the sum of the positive areas must be larger than the sum of the negative areas.

Finally, note that in order to check whether $I_2(x) \geq 0$ for all values x, it is sufficient to check $I_2(x)$ for the intersection points of F and G. For example, if $I_2(x_3) \geq 0$ (see Figure 3.12b) where x_3 is an intersection point, it will clearly be positive for $x < x_3$ because, by moving to lower values x, we decrease the negative areas that make up $I_2(x)$. Thus, $I_2(x)$ should be calculated only for the intersection points of F and G.

Figure 3.12: The area enclosed between the two distributions F and G



3.12a: Neither F nor G Dominates the Other by SSD



3.12b: F Dominates G by SSD

Example:

As we shall see later on in the chapter, the SSD rule holds also for discrete random variables; hence without loss of generality, we can use the following discrete example:

Suppose that the outcomes and the probabilities of two investments, F and G are given as follows:[6]

Investment F			**Investment G**		
x	**p(x)**	**F(x)**	**X**	**p(x)**	**G(x)**
–5	1/5	1/5	-10	1/10	1/10
0	1/5	2/5	0	7/10	8/10
6	1/5	3/5	10	2/10	1
8	1/5	4/5			
12	1/5	1			

Can we tell whether either of these investments dominates the other by SSD?

Figure 3.13 illustrates the cumulative distributions of investments F and G. As can be seen from this figure, investment G does not dominate F by SSD because:

$$\int_{-10}^{-5} [F(t) - G(t)]dt = -5 \cdot \tfrac{1}{10} = -\tfrac{1}{2} < 0.$$

Let us examine whether investment F dominates investment G. As explained above, there is no need to calculate the area between the two integrals at every point x; it is sufficient to examine the intersection points of the two distributions. The reason, once again, is that if, for example, the integral is positive up to the intersection point x=0 (see Figure 3.13), it will be positive up to any value $x < 0$ because at $x = 0$, the negative area reaches its maximum. Up to $x = -5$, this area is positive:

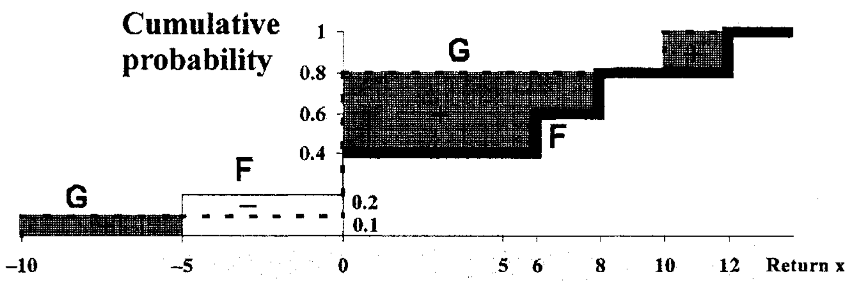
$$\int_{-10}^{-5} [G(t) - F(t)]dt = 5\tfrac{1}{10} = \tfrac{1}{2}.$$

For $-5 \leq x \leq 0$, we have a negative area. The negative area reaches its lowest value (or highest absolute number) at the value $x = 0$:

$$\int_{-5}^{0} [G(t) - F(t)]dt = -5\tfrac{1}{10} = -\tfrac{1}{2}.$$

[6] We can shift to terminal wealth by adding \$1 to all figures of x without affecting the result of this example. See also footnote 3.

Figure 3.13: A numerical example: F dominates G by SSD



Therefore, for any x < 0, the integral is positive because the negative area is smaller than –½. For example, for x = –2 we have:

$$\int_{-10}^{-2}[G(t)-F(t)]dt = \tfrac{1}{2} - 3\tfrac{1}{10} = \tfrac{2}{10}.$$

Finally, for x = 0 we have:

$$\int_{-10}^{0}[G(t)-F(t)]dt = +\tfrac{1}{2} - \tfrac{1}{2} = 0.$$

Thus, for all values x up to x = 0, we have:

$I_2(x) \geq 0$.

For the range $0 \leq x \leq 6$, the positive area increases by $(8/10 - 4/10) \cdot 6 = 4/10 \cdot 6 = 24/10$. Then, in the range $6 < x < 8$, the positive area further increases by $2 \cdot (8/10 - 6/10) = 4/10$. Finally, in the range $10 < x < 12$, we have an additional positive area of $2 \cdot (1 - 8/10) = 2 \cdot 2/10 = 4/10$.

These results can be summarized as follows:

x	$I_2(x) = \int_a^x [G(t)-F(t)]dt$	
x <–10	0	(=0)
x =–5	1/2	(=0 + 1/2)
x =0	0	(= 1/2 – 1/2)
x =6	24/10	(=0 + 24/10)

x =8	28/10	(=24/10 + 4/10)
x =10	28/10	(=28/10 + 0)
x =12	32/10	(=28/10 + 4/10)
x>12	32/10	(=32/10 + 0)

As we can see from the Figure 3.13 as well as from the detailed calculation, $I_2(x) \geq 0$ for all x, and there is at least one value for which a strong inequality holds; hence, investment F dominates investment G by SSD.

d) An Intuitive Explanation of SSD

Recall that if F dominates G by SSD, then for any negative area (G<F) there will be a positive area (F<G) which will be greater or equal to the negative area and which will be located before the negative area (or for smaller value x). For simplicity, assume that there is only one negative area S_2^- and one positive area S_1^+ and that the negative area is smaller in magnitude than the positive area. By equation (3.1) we have:

$$E_F U(x) - E_G U(x) = \int_a^b [G(x) - F(x)] U'(x) dx.$$

However, U′ is a declining function of x (by the assumption of risk aversion $U'' < 0$), hence the positive area (which corresponds to a lower value x and hence a higher value U′(x)) is multiplied by a larger number U′(x) than the negative area which comes later on and, therefore, the total integral is non-negative. This implies that for any $U \in \mathbf{U_2}$, $E_F U(x) \geq E_G U(x)$. To further illustrate this intuitive explanation, suppose without loss of generality, that U′ is constant over each range when either F is above G or G is above F. Suppose that we have the following hypothetical figures with four intersections:

Intersections (i)	**Area (S_i)**	**Marginal Utility (U′)**	$S_i \cdot U'_i$	$\sum_{i=1}^{j} S_i U'_i$
1	S_1 positive = +2	10	20	20
2	S_2 negative = –2	9	–18	2
3	S_3 positive = +10	8	+80	82
4	S_4 negative = –8	7	–56	26

The SSD rule requires that for any negative area, the sum of all areas (positive and negative) located before the negative area must be greater than the negative area under consideration. For example, if $S_2 = -2$, we have $S_1 = +2$; hence the area

S_2 will be smaller or equal than the area S_1. Similarly, for $S_4 = -8$, we have that $|-8| \le 10 - |-2| + 2 = 10$. This can be rewritten as $|-8| + |-2| < 10 + 2$. Thus, the condition $\sum_{i=1}^{k} S_i^- \le \sum_{i=1}^{l} S_i^+$ holds for k=2 and for k=4 in our example (when l stands for the number of the positive areas before the k^{th} negative area). In the above example, this requirement holds and $I_2(x) > 0$. Because $I_2(x) \ge 0$, we obtain that $E_FU(x) - E_GU(x) \ge 0$. Note that to obtain that $E_FU(x) > E_GU(x)$, it is crucial to assume that $U'(x)$ decreases as x increases (risk aversion). If this does not hold and U' in the range S_4 is, say, +15 rather than +7, F dominance over G is no longer guaranteed in spite of the fact that $I_2(x) \ge 0$.

We will demonstrate the interrelationship between differences of the positive and negative areas and utility function and, in particular, risk aversion, in the following example.

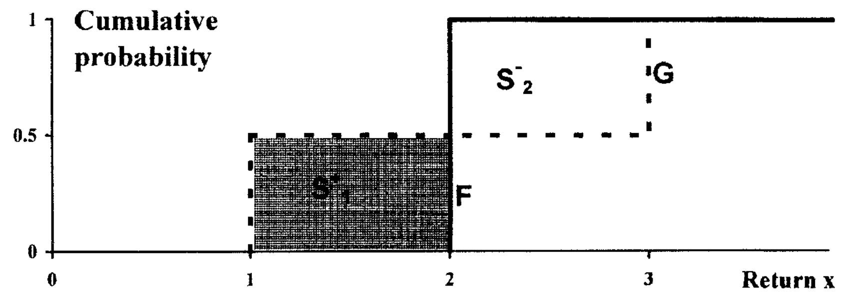
Example:

Suppose that a risk averter has to make a choice between the following two investments:

Investment G		**Investment F**	
x(in $)	**p(x)**	**x(in $)**	**p(x)**
1	1/2	2	1
3	1/2		

Note that F is riskless because 2 is obtained with a probability 1. Figure 3.14 illustrates these two distributions. It is not difficult to see that F dominates G by SSD (or FD_2G) because $\int_a^x [G(t) - F(t)]dt \ge 0$ for all values x, and there is at

Figure 3.14: A Comparison of a Riskless Asset and a Risky Asset: F Dominates G by SSD



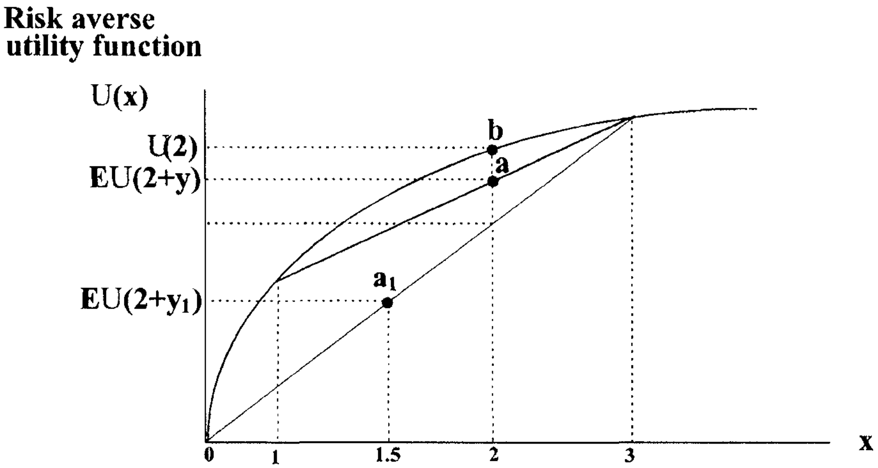
least one strict inequality. F has a +\$1 relative to G represented by the "+" square (S_1^+) or an advantage of a monetary value of \$1 · probability of 1/2. G has +\$1 more than F at a higher value of x represented by the next negative square (S_2^-), or an advantage of a monetary value of \$1 · probability of 1/2. As F "receives" so to speak, the "+" square for a lower wealth and as U′ is declining, the value of the "+" square monetary value (S_1^+) *in utility terms* is larger than the value in utility terms of the negative square, S_2^- that F loses in comparison to G; hence F will be preferred over G by all risk averters.

So far, the explanation is very similar to the explanation given in the previous example. Let us now introduce an even more intuitive explanation using the definitions of risk aversion given earlier in this chapter. Let us show that F, which dominates G by SSD, has a higher expected utility for all risk averters using the chord definition of risk aversion. To see this, we rewrite investment G as F plus return y such that y can have values −1 or +1 with equal probability. Namely:

$$\text{G} = \text{F} + \text{y when } y \equiv \begin{cases} -\$1 \; probability \frac{1}{2} \\ \\ +\$1 \; probability \frac{1}{2} \end{cases}$$

Suppose that a risk averter holds investment F. Will he/she be willing to receive, free of charge, the cash flow of y? If the answer is positive (for every risk averter), then G will dominate F or GD_2F. If the answer is negative, then FD_2G. In fact, the answer is negative. To see this, consider an investor whose wealth is (\$2) and who is offered the cash flow of y, namely ±\$1 with an equal probability. Figure 3.15 demonstrates that a risk averter will prefer not to receive y free of charge, hence FD_2G. The reason why y is rejected is that EU(2+y) (see point a on the chord) is smaller than U(2) (see point b). If risk aversion is assumed, the chord will always fall below the utility function

Figure 3.15: Risk Averse Utility Function and SSD Preference



and therefore y will be rejected. Because the above analysis is independent of the precise shape of the utility function (as long as $U'' < 0$), it holds for all risk averters; hence we can safely conclude that F D_2 G.

This provides an intuitive explanation for why, if the integral $I_2(x)$ is positive up to any value x, every risk averse investor will prefer F over G. In our example, the "+" area is exactly equal to the "–" area; hence we are comparing two distributions with equal means. If the plus area is bigger than the negative area then, *a fortiori,* there will be a preference for F. To see this, assume that we have the following two investments.

Investment G		Investment F	
x	**p(x)**	**X**	**p(x)**
0	1/2	2	1
3	1/2		

In this example, we have $E_G(x) = 1.5$ and $E_F(x) = 2$. Thus, we reduce the mean of G relative to the previous example. It is easy to verify that F dominates G by SSD. To explain this preference, once again, rewrite G as follows:

$$G = F + y_1 \text{ when } \qquad y_1 = \begin{Bmatrix} -2 \text{ } probability \tfrac{1}{2} \\ +1 \text{ } probability \tfrac{1}{2} \end{Bmatrix}$$

A risk averter will not be willing to shift from F to G, because y_1 is an unfair game, hence decreases the expected utility. Figure 3.15 demonstrates this case: Point a_1

corresponds to $EU(2+y_1)$ and it is situated below point b corresponding to U(2). Because $U(2) > EU(2+y_1)$ for any concave function, y_1 is rejected and hence F_2DG.

3.6 SUFFICIENT RULES AND NECESSARY RULES FOR SSD

a) Sufficient rules

As in the case of FSD, there exist many sufficient rules for risk aversion which imply SSD. We will consider here three of such sufficient rules.

Sufficient rule 1: The FSD rule is a sufficient rule for SSD.

To see this recall that if FD_1G then $F(x) \le G(x)$ for all x. Therefore $G(x) - F(x) \ge 0$ for all x and, because the integral of non-negative numbers is non-negative, we have:

$$F(x) \le G(x) \Rightarrow \int_a^x [G(t) - F(t)]dt \ge 0 \Rightarrow E_F U(x) \ge E_G U(x) \text{ for all } \mathbf{U} \in \mathbf{U}_2.$$

Thus, if risk aversion is assumed, the FSD rule can be employed and any investment relegated to the inefficient set with FSD will also be relegated to the inefficient set with SSD. However, this sufficient rule may result in a relatively large efficient set. For example, FSD would not discriminate between F and G in Figure 3.14 yet FD_2G (or F dominates G by SSD).

Sufficient rule 2: $Min_F(x) > Max_G(x)$ is a sufficient rule for SSD.

Like the FSD rule, this rule implies that $F(x) \le G(x)$ for all values x and, because the latter implies SSD, we can state that:

$$Min_F(x) \ge Max_G(x) \Rightarrow FSD \Rightarrow SSD = E_F U(x) \ge E_G U(x) \text{ for all } U \in U_2.$$

Indeed, any rule, which is sufficient for FSD, must be also sufficient for SSD because FSD dominance implies SSD dominance.

Sufficient rule 3: Let us now look at a sufficiency rule for SSD which bears no relationship to FSD dominance; namely, F dominates G if:

$$\int_a^x [G(t) - F(t)]dt \ge k, \text{ for all values x where } k > 0.$$

This “k rule” is sufficient for SSD because if it holds for all x then $I_2(x) \ge 0$ and $I_2(x_0) > 0$ for some x_0. In the example portrayed in Figure 3.14, F dominates G by

SSD. However, neither F nor G dominates the other by the "k rule" because for x=3, $I_2(x) = 0$ and $I_2(x) = 0 < k$. This, once again, confirms that a sufficient rule results in a relatively large efficient set.

b) Necessary rules for SSD

Necessary rule 1:

The means. $E_F(x) \geq E_G(x)$ is a necessary condition for dominance of F over G in U_2. Note that unlike FSD, here a strong inequality $E_F(x) > E_G(x)$ is not a necessary condition for SSD.

To see this recall that:

$$E_F(x) - E_G(x) = \int_a^b [G(x) - F(x)]dx = I_2(b)$$

and SSD implies that the integral $I_2(x)$ is non-negative for any value x and, therefore, it should hold in particular for x=b. Hence, by the SSD requirement, the right-hand side is non-negative which implies that $E_F(x) \geq E_G(x)$. Thus, we prove that if FD_2G, then the expected return of F must be greater or equal to the expected return of G.

Necessary rule 2:

Geometric means. $\overline{x}_{geo.}(F) \geq \overline{x}_{geo.}(G)$ is a necessary condition for dominance of F over G by SSD. To see this, recall that if FD_2G, then $E_FU(x) \geq E_GU(x)$ for all $U \in \mathbf{U_2}$. However, because $U(x) = \log(x) \in \mathbf{U_2}$ we must have also that:

$$E_F \log(x) = \log_F(\overline{x}_{geo}) \geq E_G \log(x) = \log_G(\overline{x}_{geo.}).$$

(See definition of geometric mean in section 3.3 above.)

Therefore, $\log\ (\overline{x}_{geo.}(F)) \geq \log(\overline{x}_{geo.}(G))$ (See FSD necessary rules discussion) but, because a log function is an increasing monotonic function, we can safely conclude that:

$$F\,D_2G \Rightarrow \overline{x}_{geo}(F) \geq \overline{x}_{geo.}(G).$$

Hence, the geometric means rule as defined above is a necessary rule for SSD.

It is interesting that for FSD dominance, we obtain similar necessary rules with the distinction that for FSD, the expected value and the geometric mean of the superior investment must be strictly larger than their counterparts of the inferior distribution and for SSD they can be greater *or equal* to their counterparts.

Necessary rule 3:

The "left tail" rule. A necessary rule for FD_2G is that $Min_F(x) \geq Min_G(x)$, namely the left tail of G must be "thicker". The proof is simple and similar to that used for FSD. If the necessary rule does not hold, namely,

$Min_F(x) < Min_G(x)$ then denote $Min_F(x)$ by x_k to obtain:

$$\int_a^{x_k}[G(t) - F(t)]dt = \int_a^{x_k}[0 - F(t)]dt < 0$$

(because at x_k, G will still be zero but F will be positive) and, therefore, $F\not{D}_2G$.

If FD_2G, $Min_F(x) \geq Min_G(x)$; hence, it is a necessary rule for dominance.

Other sufficient rules and necessary rules for SSD do exist, but the ones described above (in particular the necessary rules) are the most important.

We turn now to another rule called Third Degree Stochastic Dominance (TSD).

3.7 THIRD DEGREE STOCHASTIC DOMINANCE (TSD)

a) A preference for positive skewness as a motivation for TSD

So far we have assumed either that $U \in \mathbf{U}_1$ $(U' \geq 0)$ from which we derived the corresponding FSD rule or, alternatively, that $U \in \mathbf{U}_2$ $(U' \geq 0$ and $U'' \leq 0)$ from which we derived the corresponding SSD rule. In this section we derive a decision rule called Third Degree Stochastic Dominance (TSD) corresponding to the set of utility functions $U \in \mathbf{U}_3$ where $U' \geq 0$, $U'' \leq 0$ and $U''' \geq 0$.

However, before we turn to this rule, let us first discuss the economic rationale for the additional assumption asserting that $U''' \geq 0$. The assumptions $U' \geq 0$ and

$U'' \leq 0$ are easier to grasp: $U' \geq 0$ simply assumes that the investor prefers more money to less money (which stems from the monotonicity axiom), and $U'' \leq 0$ assumes risk aversion: other things being equal, investors dislike uncertainty or risk. But what is the meaning of the assumption $U''' \geq 0$? What is the economic justification for such an assumption? As we shall see, U‴ is related to the distribution skewness.

b) The definition of skewness

Skewness of a distribution of rate of return or the distribution's third central moment, denoted by μ_3, is defined as follows:

$$\mu_3 = \sum_{i=1}^{n} p_i (x_i - Ex)^3 \qquad \text{for discrete distributions}$$

(where n is the number of observations and (p_i, x_i) is the probability function),

and
$$\mu_3 = \int_{-\infty}^{\infty} f(x)(x - Ex)^3 dx \quad \text{for continuous distribution.}$$

The prizes of a lottery game are generally positively skewed because the small probability of winning a very large prize. Similarly, the value of an uninsured house is negatively skewed because of the small probability of a heavy loss due to a fire or burglary. Finally, for symmetrical distributions, the negative and positive deviations cancel each other out and the skewness is zero. The following illustrates three distributions, one with a positive skewness, one with a negative skewness, and one symmetrical distribution with zero skewness.

Example 1: The prizes of a lottery

Suppose that the prize, x, and the probability of x occurring p(x) is given by:

X	**p(x)**
$0	0.99
$1,000	0.01

There is a 0.99 probability of a zero monetary outcome and 0.01 probability of a $1,000 monetary outcome. The mean is:

$$E(x) = (0.99) \cdot \$0 + (0.01) \cdot \$1{,}000 = \$10$$

and the skewness is:

$$\mu_3 = 0.99(0-10)^3 + 0.01(1{,}000-10)^3 = 0.99(1000) + (0.01)(990)^3 = 9{,}702{,}000$$

which, as we can see, is very large and positive. Indeed, the skewness of virtually all real lotteries is positive and very large. As we shall see, a possible skewness may provide the incentive to participate in a lottery if $U''' > 0$.

Example 2: The Value of An Uninsured House

Suppose that you have a house valued at $100,000 and it is uninsured. If a fire breaks out, we assume (for simplicity only) a total loss; hence, the value of the house will be zero. We have the following information regarding x (the value of the house) and the probability, p(x):

X	p(x)
0	0.01
$1,000,000	0.99

The expected value of the house is:

$$E(x) = (0.01) \cdot 0 + 0.99 \cdot (\$100,000) = \$99,000.$$

The skewness is large and negative:

$$\mu_3 = 0.01 \cdot (0 - 99,000)^3 + 0.99 \cdot (100,000 - 99,000)^3 = -(9.702) \cdot 10^{12}$$

Example 3: Symmetrical Distribution

Finally, suppose that we toss a coin. If "heads" shows up, you get $10 and if "tails" shows up, you pay $8. The expected value is $1:

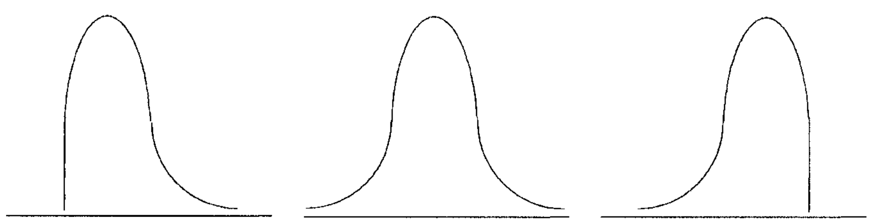
$$\tfrac{1}{2} \cdot (\$10) + \tfrac{1}{2} \cdot (-\$8) = \$1$$

and because there is an equal probability of deviating from the mean, the payoff distribution is symmetrical and the skewness is equal to zero:

$$\mu_3 = \tfrac{1}{2} \cdot (10 - 1)^3 + \tfrac{1}{2} \cdot (-8 - 1)^3 = \tfrac{1}{2} \cdot 9^3 + \tfrac{1}{2} \cdot (-9^3) = 0.$$

Figure 3.16 illustrates three distributions: a – positively skewed, b – symmetrical with zero skewness and c – negatively skewed.

Figure 3.16: The Density Function f(x) of positively Skewed, Symmetrical, and Negatively Skewed Distributions



3:16a: Positive Skewness $\mu_3>0$ | 3.16b: Symmetrical Distribution $\mu_3=0$ | 3.16c: Negative Skewness $\mu_3<0$

c) Lottery, insurance and preference for positive skewness

Let us turn now to the interrelationship between skewness and U'''. Friedman and Savage (F&S) and Kahneman and Tversky (K&T) suggest a positive approach toward understanding investor preference.[7,8] The main thrust of this approach is as follows: By observing how investors behave, we can draw conclusions regarding their preferences. In particular, F&S claim that investors buy lottery tickets and insure their homes. By insuring their homes, they reduce the variance of future value as well as the negative skewness of an uninsured house to zero. The observed behavior of purchasing home insurance can be interpreted in two ways: People insure their homes because they dislike variance, or (and) that they dislike negative skewness.

Home insurance ensures a certain income; hence, the insurance company is actually selling the negative skewness (as well as variance).

Similarly, when people buy a lottery ticket, variance and positive skewness are created. Thus, buying a lottery ticket can be explained by asserting that the investor likes variance (which is very unlikely) or the investor likes positive skewness because with a lottery ticket both variance and skewness increase.

To see the relationship between U'' and skewness more precisely, let us expand the utility function into a Taylor series about the point (w+Ex) where the utility function is U(w+x), w denotes the initial certain wealth and x is a random variable (which can be the uninsured house or the lottery ticket):

---

[7] Milton Friedman and Leonard J. Savage, "The Utility Analysis of Choices Involving Risk," *The Journal of Political Economy*, LVI, No. 4, August 1948.

[8] See Kahneman, Daniel and Tversky, Amos, "Prospect Theory of Decision Under Risk," *Econometrica*, Vol. 105, 1990.

$$U(w+x) = U(w+Ex) + U'(w+Ex)(x-Ex) + \frac{U''(w+Ex)}{2!}(x-Ex)^2$$

$$+\frac{U'''(w+Ex)}{3\%}(x-Ex)^3 + \ldots.$$

Taking the expected value from both sides and using the fact that E(x–Ex) = 0 yields:

$$EU(w+x) = U(w+Ex) + \frac{U''(w+Ex)}{2!}\sigma_x^2 + \frac{U'''(w+Ex)}{3!}\mu_3 + \ldots.$$

If other factors are held constant, then the higher $\sigma^2{}_x$, the lower the expected utility of a risk averter (because $U'' < 0$), and the higher the skewness, the higher the expected utility as long as $U''' > 0$. Therefore, if $U'' < 0$, the investor will dislike the variance (other factors being held constant!) and if $U''' > 0$, the investor will dislike negative skewness and like positive skewness.

By insuring the house, both σ^2 and μ_3 (which, with no insurance, are positive and negative, respectively) become zero; hence if $U'' < 0$, the reduction of the variance to zero by itself, will increase the expected utility, and if $U''' > 0$, the negative skewness will be replaced by zero skewness which, once again, will increase the expected utility. Insurance firms charge a risk premium; hence the expected wealth decreases which, in turn, decreases the expected utility. Taking all these factors into account, an insurance policy is worthwhile only if the expected utility increases.

Similarly, participation in an unfair lottery increases both the variance and the skewness (both of which, without the lottery ticket, are zero). If risk aversion is assumed, the large positive skewness mitigates the negative effect of the increase in the variance of those participating in lotteries.

The observation that people buy home insurance and participate in lotteries does not constitute conclusive evidence that $U''' > 0$ (because variance and expected return also change). However, these behaviors conform with the hypothesis that investors like positive skewness or dislike negative skewness which, in turn, provides support for (but not proof of) the hypothesis that $U''' > 0$.

d) Empirical studies and positive skewness preference (or $U''' > 0$).

Let us now seek out stronger evidence for the hypothesis that $U''' > 0$.

The rate of return on stocks are generally positively skewed. The intuitive reason for this is that, at most, a stock price can drop to zero (–100% rate of return).

However, the stock price is unbounded from above; hence the distribution of rates of return will have a long right tail which, in turn, may induce a positive skewness (see Figure 3.16a).

In the case of lotteries and home insurance it was difficult to separate the effect of changes in the variance and changes in the skewness; hence we could not definitively conclude that $U''' > 0$ from the fact that individuals buy insurance and lottery tickets. Stock market rates of return can be used to ascertain whether U''' is indeed positive: The effect of the variance can be separated from the effect of the skewness by conducting multiple regression analysis. To be more specific, the following cross-section regression can be performed:

$$\bar{R}_i = a_1 + a_2\sigma_i^2 + a_3\mu_{i3} + a_4\mu_{i4} \ldots\ldots + a_k\mu_{ik}$$

where μ_{ik} is the k^{th} central moment of the i^{th} mutual fund (the first k moments are included in the regression), $\sigma^2_i = \mu_{i2}$ is the variance, and $\bar{R}_i$ is the i^{th} stock average rate of return. The regression coefficients (if significant) determine how the various moments of the distribution affect the expected rate of return $\bar{R}_i$. For example, if $a_2 > 0$ this means, other things being held constant, that on average investors dislike variance. Using the Taylor series expansion, this implies that $U'' < 0$ because if $a_2 > 0$, the higher σ^2, the higher the required average rate of return $\bar{R}_i$. This means that investors do not like variance and, therefore, they will require compensation on investments with relatively large variance. Similarly, if $a_3 < 0$, we can conclude that investors like positive skewness and dislike negative skewness because an asset i with a high positive skewness (with $a_3 < 0$) implies that $\bar{R}_i$ is relatively small; investors consider positive skewness as a good feature; hence they will be willing to receive a relatively low average return.

The market dynamic for price determination of risky assets is as follows: Suppose that a firm takes an action such that the skewness of the returns on the stock increases. Then, if investors like positive skewness, the demand for the stock will increase the stock's price and, therefore, for a given future profitability, the average rate of return with the new high price will be lower.

Table 3.1 reports such a regression result using the rates of return of mutual funds. First, 25 annual rates of return are used to calculate $\bar{R}_i, \sigma_i^2, \mu_{3i,}\mu_{3i}, \mu_{4i}$, etc. for the ith mutual fund. Then, using these time series estimates, the above regression was run to estimate a_1, a_2, and other regression coefficients. Table 3.1 reveals a number of interesting points:

1. Only the first three coefficients, a_1, a_2, and a_3, are significant. All other coefficients are insignificant.

2. The R^2 is quite high (86%).

3. Investors (on average) do not like variance because $a_2 > 0$. Thus, we can conclude that for most investors, $U'' < 0$.

Table 3.1: Analysis of Regression Results for the 25-year Period

Source of Variance	Sum of squares	Degrees of Freedom	Mean Squares	F''	Critical Value	R^2
Due to Regression	544.3	2	272.2	162.3	$F_{0.99}$=5.01	0.86
Deviation from the Regression	92.2	55	1.7			

**Due to tolerance limit only two moments remain in the regression; that is, the contribution of the other 18 moments to the F value is negligible.

Variable	Coefficient	t Value	Critical Value (99%)
Constant (a_1)	7.205	16.2	$t_{0.99} = 2.40$
Variance (a_2)	0.019	9.9	
Skewness (a_3)	–8.4	–0.000064	

4. Investors, on average, like positive skewness and dislike negative skewness (because $a_3 < 0$); hence, for most (not necessarily for all) investors, $U''' > 0$.

In this analysis, a_2 measures the variance effect and a_3 the skewness effect on average rates of return. Unlike the lottery and insurance example, here each effect is estimated separately. Market data, as revealed in Table 3.1, shed light on most investors (because market prices are determined by them), but we cannot draw conclusions for all investors. Thus, the results presented in Table 3.1 support the hypothesis that for *most* investors, $U''' > 0$.

There is considerable empirical evidence supporting our hypothesis regarding the sign of U'' and U''': As far as we know, the first researcher to empirically discover this preference for positive skewness was F.D. Arditti who performed a similar regression with individual stocks (rather than mutual funds) and found that investors like positive skewness and dislike negative skewness; hence, for most investors,

$$U'' < 0 \text{ and } U''' > 0.$$[9]

[9] See F.D. Arditti, "Rate and the Required Return on Equity," *Journal of Finance*, March 1967. pp. 19–36.

e) Decreasing Absolute Risk Aversion (DARA), and Positive Skewness Preferences (or U''' > 0)

Another rationale for the assumption that U''' > 0 relies on the observation that the higher the investor's wealth, the smaller the risk premium that he/she will be willing to pay to insure a given loss. Arrow and Pratt determined that the risk premium is given by $\pi(w)$ where, $\pi(w) = -\frac{\sigma^2}{2} \cdot \frac{U''(w)}{U'(w)}$ [10, 11]

(This is Pratt's formulation but Arrow's formulation is very similar). It has been observed that the larger the wealth w, the smaller the average amount $\pi(w)$ that the investor will be willing to give up in return for getting rid of the risk (rich people do not need insurance!). Formally, this claim is that $\partial\pi/\partial w < 0$. Using the above definition of $\pi(w)$, this means that the following should hold:

$$\frac{\partial\pi(w)}{\partial w} = -\frac{\sigma^2}{2} \cdot \frac{U'(w)U'''(w) - [U''(w)]^2}{[U'(w)]^2} < 0$$

and this can hold only if $U'''(w) > 0$. Thus, we conclude from the observation that $\partial\pi(w)/\partial w < 0$ that $U'''(w) > 0$ (Note that the converse does not hold; $U'''(w) > 0$ does not imply $\partial\pi/\partial w < 0$).

To sum up, participation in a lottery and buying insurance provides some evidence that $U'''(w) > 0$. The empirical studies and the observation that $\partial\pi(w)/\partial w > 0$ provide much stronger evidence for the preference for positive skewness (and aversion to negative skewness) which, in turn, strongly support the hypothesis that $U'''(w) > 0$. This evidence is strong enough to make it worthwhile to establish an investment decision rule for $U \in \mathbf{U}_3$ where $U' \geq 0$, $U'' \leq 0$ and $U'''(w) \geq 0$.

Once again, we will first prove the investment decision rule for continuous random variables and then extend it to the discrete case.

f) The optimal investment rule for $U \in \mathbf{U}_3$: TSD

The optimal investment rule for $U \in \mathbf{U}_3$ is given in the following Theorem.

Theorem 3.3

Let F(x) and G(x) be the cumulative distributions of two investments under consideration whose density functions are f(x) and g(x), respectively. Then F

---

[10] See K.J. Arrow, *Aspects of the Theory of Risk: Bearings*, Helsenki, Yrjö Jahnssonin Säätiö, 1965.
[11] See J.W.Pratt, "Risk Aversion in the Small and in the Large," *Econometrica*, January-April, 1964.

dominates G by *Third Degree Stochastic Dominance* (TSD) if and only if the following two conditions hold:

a) $$I_3(x) = \int_a^x \int_a^z [G(t) - F(t)]dtdz \geq 0 \qquad \text{for all x}$$

(for the sake of brevity, we denote the double integral by $I_3(x)$; hence we require that $I_3(x) \geq 0$).

b) $$E_F(x) \geq E_G(x) (\text{or } I_2(b) \geq 0)$$

and there is at least one strict inequality, namely:

$$I_3(x) \geq 0 \text{ and } I_2(b) > 0 \Leftrightarrow E_F U(x) \geq E_G U(x) \text{ for all } U \in \mathbf{U}_3 \tag{3.3}$$

To have a dominance we require that either $I_3(x_0) > 0$ for some x, or $I_2(b) > 0$ which guarantees a strong inequality holds for some $U \in \mathbf{U}_3$. (recall that $U \in \mathbf{U}_3$ if $U' \geq 0$, $U'' \leq 0$ and $U''' \geq 0$).

We call such a dominance third-degree because assumptions of the third order are made on U (i.e. , $U' > 0$, $U'' \leq 0$ and $U''' \geq 0$). If F dominates G by TSD we write it as FD_3G where the subscript 3 indicates a third order stochastic dominance. We prove below first the sufficiency, and then the necessity, of TSD. This will be followed by an example, a graphical exposition, and an intuitive explanation.

Sufficiency:

We have to prove that if the left-hand side of (3.3) holds, then the right-hand side will also hold. Namely, if the above two conditions hold, every risk averter with $U'''(w) \geq 0$ will prefer F over G.

We have already seen in SSD proof (see eq. 3.2) that the following holds:

$$E_F U(x) - E_G U(x) = U'(b) \int_a^b [G(x) - F(x)]dx + \int_a^b -U''(x)(\int_a^x [G(t) - F(t)]dt)dx.$$

Integrating by part the right-hand side second term yields:

$$\int_a^b -U''(x)(\int_a^x [G(t) - F(t)]dt)dx = -U''(x)(\int_a^x \int_a^z [G(t) - F(t)]dtdz)\Big|_a^b$$

$$+\int_a^b U'''(\int_a^x\int_a^z [G(t)-F(t)]dtdz)dx = -U''(x)I_3(x)\Big|_a^b + \int_a^b U'''(x)I_3(x)dx$$

$$= -U(b)I_3(b) + \int_a^b U'''(x)\, I_3(x)\, dx$$

where
$$I_3(x) = \int_a^x\int_a^z [G(t)-F(t)]dtdz.$$

(Note that for convenience we change the order of the integrals. However, because x and z are independent and for $x = z$ we have $I_3(x) = I_3(z)$, we have $\int_a^x\int_a^z(\bullet) = \int_a^z\int_a^x(\bullet)$ and changing the order of the integrals does not change the value of the integral). Collecting all these results, we obtain:

$$E_F U(x) - E_G U(x) = U'(b)I_2(b) - U''(b)I_3(b) + \int_a^b U'''(x)I_3(x)dx \qquad (3.4)$$

By the sufficiency condition $I_3(x) \geq 0$ for all values x and hence also for x=b, we have $I_3(b) \geq 0$. By the assumption that $U \in \mathbf{U}_3$, we have $-U''(b) \geq 0$ and $U'''(x) \geq 0$. Therefore, the second and third terms on the right-hand side of eq. (3.4) are non-negative. We now have to show that the first term on the right-hand side of eq. (3.4) is also non-negative. However, we already have shown that:

$$E_F(x) - E_G(x) = \int_a^b [G(x)-F(x)]dx \equiv I_2(b).$$

By the sufficient condition of TSD, we require that $I_2(b) \geq 0$ and, by the monotonicity of U, we require that $U'(b) \geq 0$. Therefore, the first term on the right-hand side is also non-negative. Because all three terms on the right-hand side of eq. (3.4) are non-negative, we conclude that:

$$E_F U(x) \geq E_G U(x) \qquad \text{for all } U \in \mathbf{U}_3.$$

Finally, for strict dominance, we need to show that there is at least one $U_0 \in \mathbf{U}_3$ such that a strict preference exists, namely:

$$E_F U_0(x) > E_G U_0(x).$$

To see this, recall that the dominance condition requires at least one strict inequality (either $I_2(b) > 0$ or $I_3(x_0) > 0$ for some x_0). Let us assume first that:

$$I_2(b) \equiv E_F(x) - E_G(x) > 0.$$

Choose $U_0 = \log(x) \in \mathbf{U}_3$ (for terminal wealth $x > 0$) then $U'(x)$ at $x = b$ is given by $1/b > 0$ and, therefore:

$$U'(b)\int_a^b [G(x)-F(x)]dx = U'(b)[E_F(x) - E_G(x)] > 0$$

which implies that $E_F U_0(x) > E_G U_0(x)$ (recall the other terms cannot be negative).

Let us discuss the other possibility where $E_F(x) = E_G(x)$ and there is at least one value x_0 for which there is a strict inequality, $I_3(x_0) > 0$. Due to the continuity of $I_3(x)$, there exists $\varepsilon > 0$ such that for $x_0 - \varepsilon \le x \le x_0$, $I_3(x) > 0$.

Select the following $U_0 \in \mathbf{U}_3$:

$$U_0(x) = \begin{cases} -e^{-(x_0-\varepsilon)} & a \le x \le x_0 - \varepsilon \\ -e^{-x} & x_0 - \varepsilon \le x \le x_0 \\ -e^{-x_0} & x_0 \le x \le b \end{cases}$$

Then, for this specific utility function, $U'(b) = U''(b) = 0$ and we have (see eq. 3.4):

$$E_F U_0(x) - E_G U_0(x) =$$

$$= \int_a^b U_0'''(x)\left(\int_a^x\int_a^z [G(t)-F(t)]dtdz\right)dx = \int_a^{x_0-\varepsilon} U_0'''(x)\left(\int_a^x\int_a^z [G(t)-F(t)]dtdz\right)dx +$$

$$+ \int_{x_0-\varepsilon}^{x_0} U_0'''(x)\left(\int_a^x\int_a^z [G(t)-F(t)]dtdz\right)dx +$$

$$+ \int_{x_0}^{b} U_0'''(x)\left(\int_a^x\int_a^z [G(t)-F(t)]dtdz\right)dx.$$

Because $U'''(x)=0$ for $x \le x_0 - \varepsilon$ and for $x > x_0$, the first and the third terms vanish and we are left with the second term on the right-hand side. The second term is equal to $\int_{x_0-\varepsilon}^{x_0} U_0'''(x) I_3(x)dx = \int_{x_0-\varepsilon}^{x_0} e^{-x} I_3(x)dx > 0$, because $e^{-x} > 0$ and $I_3(x) > 0$ in the range $x_0 - \varepsilon \le x \le x_0$. Thus, we have proved that if $I_3(x_0) > 0$, then there is $U_0 \in \mathbf{U}_3$ such that $E_F U_0(x) > E_G U_0(x)$. To sum up, if the two conditions hold with at least one strict inequality, we have proved that $E_F U(x) \ge E_G U(x)$ for all $U \in \mathbf{U}_3$ and that there is at least one $U_0 \in \mathbf{U}_3$ such that $E_F U_0(x) > E_G U_0(x)$, Hence, F dominates G for all $U \in \mathbf{U}_3$ or FD_3G.

Necessity:

We have to prove that:

$$E_F U(x) \geq E_G U(x) \text{ for all}$$

$$U \in \mathbf{U}_3 \Rightarrow I_3(x) \equiv \int_a^x \int_a^z [G(t) - F(t)]dtdz \geq 0 \text{ and } E_F(x) \geq E_G(x).$$

Once again, we prove this claim by contradiction. Suppose that there is x_0 such that $I_3(x_0) < 0$. Then, due to the continuity of $I_3(x)$, there will be an $\varepsilon > 0$ such that $I_3(x) < 0$ for all $x_0 - \varepsilon \leq x \leq x_0 + \varepsilon$. Define:

$$U_0(x) = \begin{cases} -e^{-(x_0-\varepsilon)} & \text{for } x \leq x_0 - \varepsilon \\ -e^{-x} & \text{for } x_0 - \varepsilon \leq x \leq x_0 + \varepsilon \\ -e^{-(x_0+\varepsilon)} & \text{for } x \geq x_0 \end{cases}$$

$U_0 \in \mathbf{U}_3$ because $U' \geq 0$ $U'' \leq 0$ and $U''' \geq 0$. For $x_0 - \varepsilon < x < x_0 + \varepsilon$ there are strict inequalities in all these derivatives. Also $U'(b) = U''(b) = 0$. Therefore, the first two terms of the right-hand side of eq. (3.4) vanish and we are left with:

$$E_F U_0(x) - E_G U_0(x) = \int_a^b U'''(x) \left(\int_a^x \int_a^z [G(t) - F(t)]dtdz \right) dx$$

$$= \int_a^{x_0-\varepsilon} U'''(x) I_3(x) dx + \int_{x_0-\varepsilon}^{x_0+\varepsilon} U'''(x) I_3(x) dx + \int_{x_0+\varepsilon}^{b} U'''(x) I_3(x) dx.$$

However, $U'''(x) = 0$ for $x \leq x_0 - \varepsilon$ and for $x \geq x_0 + \varepsilon$. Thus, we have:

$$\Delta_0 \equiv E_F U_0(x) - E_G U_0(x) = \int_{x_0-\varepsilon}^{x_0+\varepsilon} U'''(x) \left(\int_a^x \int_a^z [G(t) - F(t)]dtdz \right) dx.$$

However, because by assumption, $\int_a^x \int_a^z [G(t) - F(t)]dtdz < 0$ for all $x_0 - \varepsilon \leq x \leq x_0 + \varepsilon$, and $U'''(x) > 0$, the right-hand side is negative; hence $E_F(U_0(x)) < E_G(U_0(x))$.

Thus, if $I_3(x_0) < 0$ for some, x_0, $\Delta_0 < 0$, then $F \not{D}_3 G$. Therefore, if FD_3G, namely for every $U \in \mathbf{U}_3$, $E_F U(x) \geq E_G U(x)$, it is impossible to have a value x_0

such that $\int_a^{x_0}\int_a^{z}[G(t)-F(t)]dtdz<0$. Thus, a dominance in $\mathbf{U}_3$ implies that the above integral must be non-negative up to every value x.

Now let us turn to the other condition. Once again, we employ the indirect method. We will show that if $E_F(x) < E_G(x)$ then there is $U_0 \in \mathbf{U}_3$ such that $E_FU_0(x) < E_GU_0(x)$; hence if FD_3G (namely for all $U \in \mathbf{U}_3$), it is impossible to have $E_F(x) < E_G(x)$.

Suppose that $E_F(x) < E_G(x)$. Choose $U_0(x) = -e^{-kx}$ with $k > 0$. For this function $U'(x) = ke^{-kx} > 0, U''(x) = -k^2e^{-kx} < 0$, and $U'''(x) = k^3e^{-kx} > 0$; hence, $U_0(x) \in \mathbf{U}_3$. Expand $U_0(x)$ to a Taylor series about x=0 to obtain:

$$U_0(x) = -1 + kx \frac{k^2x^2}{2!} + \frac{k^3x^3}{3!} + \ldots$$

Take a value $k \to 0$ (still $U_0(x) \in \mathbf{U}_3$) to obtain:

$$EU_0(x) = -1 + kEx + o(k).$$

For this utility function with $k \to 0$ for the two investments under consideration, we have:

$$E_FU_0(x) = -1 + k\,E_F(x) + o(k)$$

$$E_GU(x) = -1 + k\,E_G(x) + o(k).$$

Therefore, for a sufficiently small positive value k:

$$E_F\,U_0(x) - E_GU_0(x) = k\,(E_F(x) - E_G(x)) + o(k).$$

Therefore, if $E_F(x) < E_G(x)$, then for this utility function (recall that $U_0 \in \mathbf{U}_3$), we have $E_FU_0(x) < E_GU_0(x)$ (because $o(k)/k \to 0$ as $k \to \infty$, hence choose $k \to \infty$ to obtain this result). Therefore, if we assume that FD_3G, namely F dominates G for all $U \in \mathbf{U}_3$, the inequality $E_F(x) < E_G(x)$ is impossible which completes the necessity proof.[12]

---

[12] Another $U_0(x)$ that can be employed in the necessity proof is a linear utility function in most of the range with a small range $x_0 \leq x \leq x_0 + \varepsilon$ at which $U''' > 0$. Thus, this function is close to the linear function (for sufficiently small ε) which can be used to prove that $E_F(x) \geq E_G(x)$ is a necessary condition for TSD of F over G.

g) Graphical exposition of TSD

A preference of one investment over another by TSD may be due to the preferred investment having a higher mean, a lower variance, or a higher positive skewness. In the first example given below, we compare two distributions with equal means and equal variances but different skewness. In the second example, we compare two investments with equal means and equal skewness but different variance.

a) Example 1: FD_3G with equal means and equal variances

	Investment G		**Investment F**	
	x	**p(x)**	**X**	**p(x)**
	0	1/4	1	3/4
	2	3/4	3	1/4
Expected value:	1.5		1.5	
Variance:	0.75		0.75	

Figure 3.17a provides the cumulative distributions corresponding to these two investments. As we can see, the two distributions intercept, hence, there is no FSD,

$(F\not{D}_1G, G\not{D}_1F)$. Let us check whether there is SSD. $G\not{D}_2F$ because $Min_G(x) = 0 < Min_F(x) = 1$; hence, the 'left tail' necessary condition for dominance of G over F does not hold. Does F dominate G by SSD? To answer this question let us draw $I_2(x)$. Figure 3.17b draws the integral:

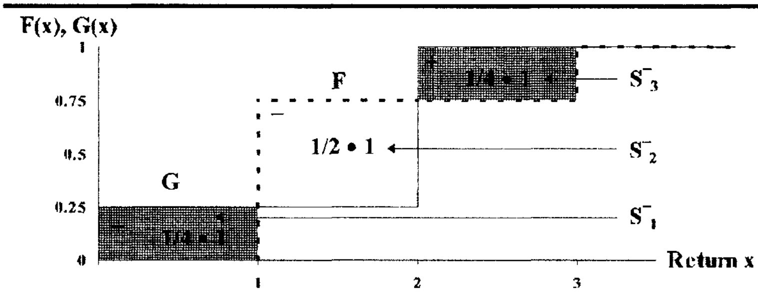
$$I_2(x) = \int_a^x [G(t) - F(t)]dt$$

for all values x.

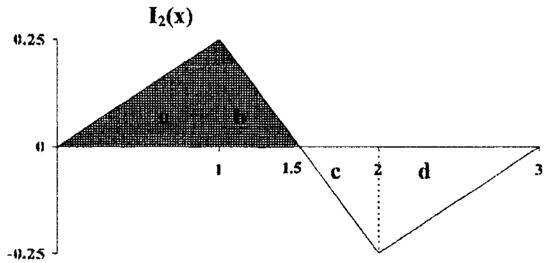
In the range $0 \le x \le 1$, $I_2(x)$ increases linearly with x reaching its maximum at x=1 when $I_2(x) = 1/4 \cdot 1 = 1/4$ which is equal to the first "+" area in Figure 3.17a. Then $I_2(x)$ decreases linearly as x increases because the second area of Figure 3.17a is negative. As the second negative area is equal to $-(\frac{3}{4} - \frac{1}{4}) \cdot 1 = -\frac{1}{2}$, $I_2(x)$ reaches its minimum at x=2 where $I_2(x) = \frac{1}{4} - \frac{1}{2} = -\frac{1}{4}$. As the line corresponding to $I_2(x)$ is equal to $+\frac{1}{4}$ at x = 1 and $-\frac{1}{4}$ at x = 2, it must intercept the horizontal axis at x = 1.5. Then for the range $2 < x \le 3$, once again we have a positive area equal to $(1 - \frac{3}{4}) \cdot (3 - 2) = \frac{1}{4}$. Therefore, at this range, $I_2(x)$ increases linearly and at x=3, it is equal to zero. From Figure 3.17b it is also clear that $F\not{D}_2G$ because $I_2(x) < 0$ in the range $1.5 < x < 3$; hence, $I_2(x)$ is not positive everywhere and

FD_2G. Actually, this can be seen directly from Figure 3.17a because the negative area (S_2^-) is greater than the previous positive area (S_1^+). Therefore:

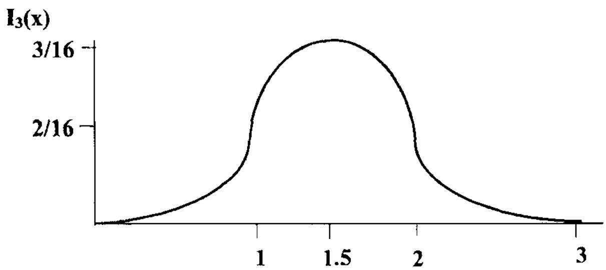
Figure 3.17: A Graphical Illustration of TSD



3.17a: The cumulative distributions F and G



3.17b: The area $I_2(x)=\int_a^x [G(t)-F(t)]dt$



3.17c: The area $I_3(x)=\int_a^x \int_a^z [G(t)-F(t)]dtdz$

$$I_2(2) = \int_a^2 [G(t) - F(t)]dt = ¼ - ½ = -¼ < 0.$$

To sum up, neither F nor G dominates the other by FSD or by SSD.

Let us use Figure 3.17b to show that there is TSD of F over G. To see this we draw the curve $I_3(x)$ in Figure 3.17c which is nothing but the area under the curve given in Figure 3.17b up to any point x. It is a little complex to see that FD_3G; hence, we add more details to our calculation. Up to point x=1, (see the curve in Figure 3.17b), the accumulated area under $I_2(x)$ grows at an accelerated rate as x increases from 0 to 1. Therefore, in this range, the curve in Figure 3.17c is convex reaching the value (¼ · 1)/2 = 2/16 which is the area of triangle a. Then comes the area of triangle b corresponding to the range $1 < x \leq 1.5$; in this range $I_3(x)$ is still increasing but at a diminishing rate; hence, in the range $1 < x \leq 1.5$, $I_3(x)$ is concave. At x = 1.5 we have:

$$I_3(x) = 2/16 + [(1.5 - 1) \cdot (1/4)]/2 = 3/16.$$

From the value x = 1.5, $I_3(x)$ starts declining because $I_2(x)$ is negative. First $I_3(x)$ declines at an increasing rate corresponding to triangle c. Because triangle c and b are symmetrical, $I_3(x)$ is reduced to 2/16, at x = 2. Then, $I_3(x)$ decreases at a diminishing rate corresponding to triangle d. Because the area of triangle d is equal to ((3 – 2) · ¼)/2 = 2/16 $I_3(x)$ is equal to zero for all $x \geq 3$ (see Figure 3.17c).

Because $I_3(x) \geq 0$ for all x, and $I_3(x_0) > 0$ for some x, and $E_F(x) = E_G(x)$ we can safely conclude that FD_3G. This example illustrates a case in which there is TSD but no FSD or SSD.

It is interesting to note that in this example we have:

$$E_F(x) = E_G(x)$$
$$\sigma_F^2(x) = \sigma_G^2(x) \text{ and } FD_3G.$$

However, it is reasonably simple to explain why FD_3G: F has a positive skewness whereas G has a negative skewness and we know that $U \in \mathbf{U}_3$ implies that $U''' > 0$; hence, there is a preference for positive skewness. Thus, the preference of F over G by TSD can be explained by the skewness difference of these two investments. Let us elaborate. From Taylor's expression we have:

$$EU(x) = U(w + Ex) + \frac{U''(\cdot)}{2!}\sigma_x^2 + \frac{U'''(\cdot)}{3!}\mu_3(x) + \ldots.$$

and $U \in \mathbf{U}_3$ implies that $U'''>0$.

For the distributions with the same mean and variance, the one with the highest skewness, μ_3, may have an advantage over the others because $U''' > 0$.

Let us first calculate the third central moment to show that F is, indeed, positively skewed and that G is negatively skewed:

$$\mu_3(G) = \tfrac{1}{4} \cdot (0 - 1.5)^3 + \tfrac{3}{4} \cdot (2 - 1.5)^3 = \tfrac{1}{4} \cdot (-3.375) + \tfrac{3}{4}(-0.125) + \cong -0.750$$

and

$$\mu_3(F) = \tfrac{3}{4} \cdot (0 - 1.5)^3 + \tfrac{1}{4} \cdot (3 - 1.5)^3 = \tfrac{3}{4} \cdot (-0.125) + \tfrac{1}{4} \cdot (3.375)$$

$$\cong -0.094 + 0.844 = +0.750.$$

From the above example, it is tempting to believe that for TSD dominance, the dominating investment has to have a larger skewness. Of course, this is not generally true; it is possible for F to dominate G by FSD (hence by SSD and TSD) but F may have a larger or smaller variance as well as a larger or smaller skewness relative to G as long as F has a higher mean return than G. The more interesting case is to analyze the role of skewness when two distributions with equal means are compared where there is TSD but no FSD or SSD. Can we say that in such a case the dominating investment necessarily has a higher skewness? The answer here is, once again, negative. In the next example we compare two symmetrical distributions with equal means: namely two distributions with $\mu_3(F) = \mu_3(G) = 0$ and $E_F(x) = E_G(x)$.

Example 2: FD_3G with equal means and equal skewness

	Investment F		**Investment G**	
	X	**p(x)**	**x**	**p(x)**
	1.49	½	1	¼
	3.51	½	2	¼
			3	¼
			4	¼
Expected value:	2.5		2.5	
Variance:	1.02		1.25	
Skewness μ_3:	0		0	

Suppose that you consider the above two investments, F and G:

Figure 3.18 illustrates the cumulative distributions of these two investments. Because F and G intercept, we can conclude that $F \not{D}_1 G$ and $G \not{D}_1 F$, and there is no FSD. Let us now check whether there is SSD: $G \not{D}_2 F$ because the left tail of G is "thicker" than the left tail of F, and $F \not{D}_2 G$ because up to x=2, we have:

$$\int_1^2 [G(x) - F(x)]dx = \tfrac{1}{4} \cdot (0.49) + \tfrac{1}{4}(-0.51) = \tfrac{1}{4} \cdot (-0.02) < 0$$

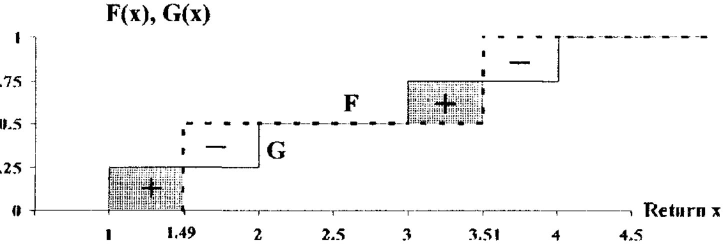
Thus, there is neither FSD nor SSD. Let us now check whether there is TSD. Calculating the expected return we find:

$$E_F(x) = \tfrac{1}{2} \cdot (1.49) + \tfrac{1}{2} \cdot (3.51) = 2.5$$

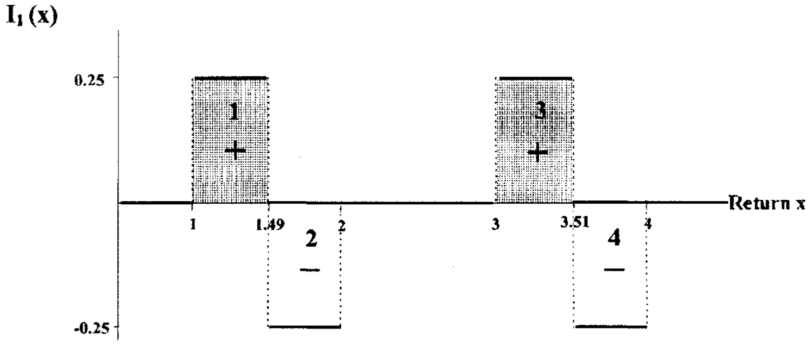
$$E_F(x) = \tfrac{1}{4} \cdot 1 + \tfrac{1}{4} \cdot 2 + \tfrac{1}{4} \cdot 3 + \tfrac{1}{4} \cdot 4 = 2.5$$

Because $E_F(x) = E_G(x)$, each distribution can, potentially, dominate the other. However, because of the required second condition of TSD, G cannot dominate F. To see this, note that in the range $1 \le x < 1.49$, $G > 0$ and $F = 0$; hence, $\int_1^{1.49} \int_1^z [F(t) - G(t)]dtdz < 0$ and, therefore, $G \not{D}_3 F$. In what follows we show that FD_3G. Figure 3.18a shows the cumulative distributions F and G. Figure 3.18b presents the difference $I_1(x) = G(x) - F(x)$, and Figure 3.18c depicts the integral $I_2(x) = \int_a^x [G(t) - F(t)]dt$. As we can see, the integral $I_2(x)$ is negative in some range which confirms the previous conclusion that F does not dominate G by SSD. Figure 3.18d provides the integral of the curve given in Figure 3.18c; namely, $I_3(x) = \int_a^x \int_a^z [G(t) - F(t)]dtdz$. (The precise calculation of this area is given in Table 3.2 and will be discussed below). Note that up to $x = 1.49$, the accumulation of area takes place at an accelerated pace and for $x > 1.49$, we continue to accumulate positive area (up to a given point) but at a slower pace. This is induced by the fact that we accumulated the area under the first triangle in Figure 3.18c. Then there is a range $1.98 \le x \le 3.02$, where the integral $I_3(x)$ is diminishing and from $x = 3.02$, it increases once again in the same way as it increased for the first triangle. The calculation in Table 3.2 confirms that the curve in Figure 3.18d never crosses the horizontal axis which implies that for all x, $I_3(x) \ge 0$ and there is at least one x for which $I_3(x_0) > 0$, hence FD_3G.

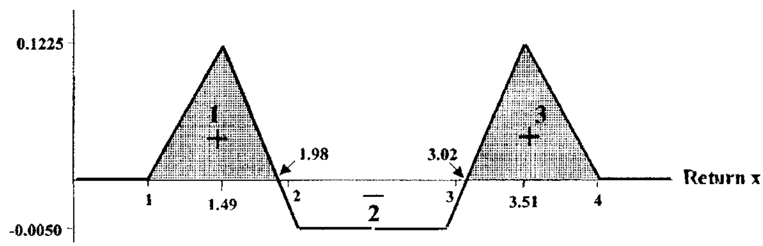
Figure 3.18: TSD with Equal Means and Zero Skewness



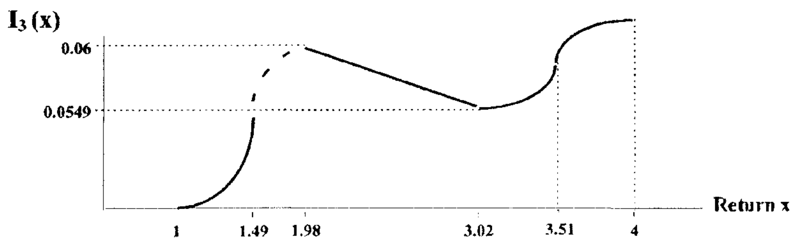
3.18a: F and G



3.18b: $I_1(x)$



3.18c: $I_2(x)$



3.18d: $I_3(x)$

Let us now turn to the detailed calculation of $I_3(x)$. Table 3.2a on page 104 provides the difference $I_1(x) = G(x) - F(x)$ for various points x. For example, up to point x = 1.49, F(x) = 0.5, G(x) = 0.25; hence, $I_1(x) = G(x) - F(x) = -0.25$ (see Figure 3.18a).

Table 3.2b provides the cumulative area $I_2(x) = I_2(x) = \int_a^x [G(t) - F(t)]dt$ up to any value x.

For example, up to x = 1.49, the area $I_2(x) = 0.49 \cdot 0.25 = +0.1225$. Up to x = 2, we have $0.1225 - (0.51) \cdot 0.25 = 0.1225 - 0.1275 = -0.0050$. The accumulation of the area $I_2(x)$ up to any value x is illustrated in Figure 3.18c. Note that for symmetry, the base of the triangle 1 is $0.49 \cdot 2 = 0.98$; hence, the declining line crosses the horizontal axis at x = 1.98. Table 3.2c uses Figure 3.18c to calculate $I_3(x)$ up to any point x. Note that $I_3(1.98) = 0.060$ and $I_3(3.02) = 0.060 - 0.0549 > 0$. Therefore, up to x = 3.02, $I_3(x) \geq 0$ for all x; hence, for all x up to x = 3.02, we have $I_3(x) \geq 0$. For higher values, x the function $I_3(x)$ increases even further, therefore, FD_3G.

This, together with the fact that $E_F(x) = E_G(x) = 2.5$, is sufficient to conclude that FD_3G.

Thus, from the above two examples of TSD, we see that positive skewness which plays a central role in TSD, does not tell the whole story; in the above example FD_3G in spite of the fact that $E_F(x) = E_G(x)$, there is no FSD and no SSD and the two distributions are symmetrical! This makes the intuitive explanation of TSD quite difficult, as we shall see in the following.

h) The intuitive explanation of TSD

We provide here an intuitive explanation for TSD. The intuitive explanation, as mentioned above, is not as simple as that of FSD or SSD. In addition, we use a discrete distribution; hence, the explanation now is heuristic and not precise mathematically.

Assume that we have the following utility function U(x), from which the various derivatives can be calculated,

X	U(x)	U′(x)	U″(x)	U‴(x)
1	100	—	—	—
2	110	10	—	—
3	116	6	–4	—
4	120(117)	4(1)	–2(–5)	2(–1)
5	123	3	–1	1
6	125	2	–1	0
7	126.5	1.50	–0.5	0.5

Table 3.2: A Numerical Example of TSD

Table 3.2a: $I_1(x)$

X	F(x)	G(x)	$I_1(x) = [G(x)–F(x)]$
<1	0	0	0
1	0	0.25	+0.25
1.49	0.50	0.25	–0.25
2	0.50	0.50	0.00
3	0.50	0.75	–0.25
3.51	1.00	0.75	–0.25
4	1.00	1.00	0.00

Table 3.2b: $I_2(x)$

X	$I_1(x) = [G(x)–F(x)]$	Non-cumulative area of G–F for each interval		Cumulative area $I_2(x) = \int_{-\infty}^{x} ([G(t) - F(t)]dt$
<1	0	—		0
1	+0.25	—		0
1.49	–0.25	0.49 x 0.25	= +0.1225	+0.1225
2	0.00	–0.51 x 0.25	= –0.1275	–0.0050
3	–0.25	1 x 0.00	= 0.00	–0.0050
3.51	–0.25	–0.25 x 0.51	= +0.1275	+0.1225
4	0.00	0.49 x (–0.25)	= –0.1225	0.0000

Table 3.2c: $I_3(x)$ (based on Figure 3.18c)

	Triangle 1	Trapezoid 2	Triangle 3
Return x	1–1.98	1.98–3.02	3.02–4
Midpoint	1.49	—	3.51
Base	L=(1.49–1) x 2=0.98	L_1 = 3–2 = 1	L = (4–3.51) x 2=0.98
		L_2 = 3.02–1.98 = 1.04	
Height	+0.1225	–0.0050	0.1225
Area	+0.1225 x 0.98/2	–0.005 x (1+1.04)/2	+0.1225 x 0.98/2
	=0.0600	= –0.0051	=0.0600
Cumulative area $I_3(x)$	+0.0600	+0.0549	+0.1149

Because this is a discrete distribution, we calculate U' as follows:

$U'(x_{i+1}) = (U(x_{i+1}) - U(x_i))/(x_{i+1} - x_i)$ where in our example $x_{i+1} - x_i = 1$ for all values i. $U''(x)$ and $U'''(x)$ are calculated in a similar way. For example, $U'''(x_{i+1}) = (U''(x_{i+1}) - U''(x_i))$. In this example, because $U'(x) \geq 0$, $U''(x) \leq 0$ and $U'''(x) \geq 0$ this function is included in the set U_3. Figure 3.19 illustrates this function and its derivatives. Note that $U'''(x) \geq 0$ but $U'''(x)$ is not necessarily a declining function. In our example, it decreases from 2 to zero and then increases back to 0.5. However, the fact that $U''' \geq 0$ implies that $U''(x)$ is a non-decreasing function of x. However, it is not necessarily strictly concave (see Figure 3.19b). If U'' were not non-decreasing as x increases, then the condition $U'''(x) \geq 0$ would be violated. To see this, let us look at the numbers in the brackets corresponding to x = 4. For this hypothetical utility function, $U(4) = 117$, $U'(4) = 1$ and $U''(4) = -5$ and $U'''(4) = -1 < 0$. Because with these figures $U'' \leq 0$ is a decreasing function of x in some range, we obtain that $U'''(x) < 0$.

$U''(x) \geq 0$ also implies that $U'(x)$ must be a declining function of x, as shown in Figure 3.19a, namely $U'(x)$ is a convex function. To see this, consider once again the numbers in brackets corresponding to x=4. If $U(4) = 117$, then $U'(4) = 1$ and U' decreases at an accelerated rate (namely it is not convex because it declines from 10 to 6 and then from 6 to 1). We still have $U''(4) = -5 < 0$ but $U'''(4)$ will be -1 which contradicts the requirement that $U'''(x) \geq 0$ for all x. Thus, $U'(x)$ must decline at a slower rate as x increases, otherwise $U'''(x) < 0$. From the above example, we see that the assumption that $U'''(x) \geq 0$ implies that $U'(x)$ is a declining convex function. Note that if $U'(x)$ were to decline at a constant pace, say from 10 to 6 and then from 6 to 2, then U'' would be -4 in both cases and $U''' = 0$; namely, the requirement $U'''(x) \geq 0$ would still hold. Thus, we allow U' to decline at the same pace at some range, and because of the required condition $U'''(x) > 0$ for some x, it must be strictly convex at some range.

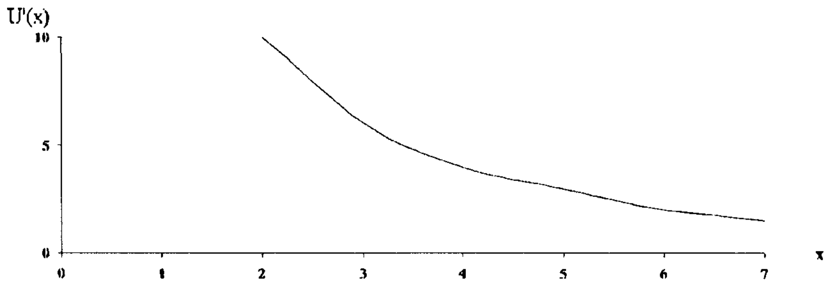
Let us now show how the fact that $U'(x)$ is convex declining function of x explains why we may have TSD but no SSD even for symmetrical distributions.

We have:

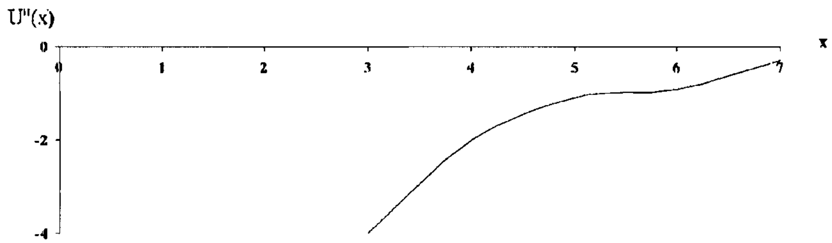
$$E_F U(x) - E_G U(x) = \int_a^b [G(x) - F(x)] U'(x) dx.$$

Because U′(x) is a declining convex function, we allow a first positive area to be followed by a second larger negative area such that SSD does not hold but TSD may hold. To see how this is possible, let us first assume, for simplicity only, that U′(x) is strictly declining. Therefore, the weight of the negative area in utility terms is smaller than the weight of the positive area in utility terms. To introduce this idea, consider the previous example (see Figure 3.18) where we have the following areas of $I_2(x)$:

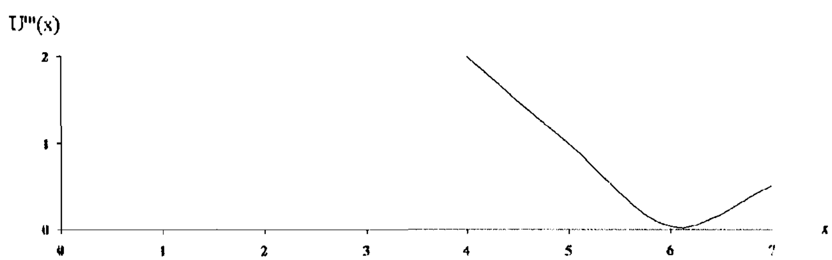
Figures 3.19: The Derivatives of A Function $U \in U_3$



3.19a: U′ (x)



3.19b: U″ (x)



3.19c: U‴ (x)

Area (j)	Size of Area	Hypothetical	Area • U′	Total expected Utility
(see Fig. 3.18)	S_j	U'_j	$(S_j U'_j)$	$\sum_{i=1}^{j} S_i \cdot U'_i$
1	+0.25·0.49	4	+0.4900	+0.4900
2	−0.25·0.51	3	−0.3825	+0.1075(=0.49 − 0.3825)
3	...	...	...	...
4	...	...	...	...

Thus, although the second negative area is larger than the previous positive area, if U′ is strictly declining (in our example from 4 to 3), then in utility terms for $U \in \mathbf{U}_3$, the positive area in the above example is worth more than the negative area; hence, FD_3G is possible.

Of course, the larger negative area that is allowed is a function of the positive area that precedes it as well as the relative location of these two areas on the horizontal axis. We elaborate below on the relationship between the convexity of U′, the location of the various areas and TSD.

The above example is overly simplistic, serving merely to introduce the relationship between the convexity of U′(x) and TSD. In the example below, we discuss the importance of the location of the various areas for the existence of TSD. We also show that to have TSD it is required to have a convex U′(x) but U′(x) = 0 in some range is possible.

We illustrate this by means of Figure 3.20. In Figure 3.20a there is TSD but no SSD (see also Figure 3.17 corresponding to this example). Note that $E_F(x) = E_G(x)$; hence, the total negative area (enclosed section F and G) in absolute terms is equal to the total positive area. We show below, and intuitively explain, the difference between F and G and the relationship to TSD criterion.

First note that by construction for the various blocks (or areas between F and G), we have a = b = c = d. Because U′ is declining, there is a utility gain from blocks a + b and a utility loss from blocks c + d. However, because U′ is convex, the gain from the difference between the two areas "a" and "b" in utility terms is larger than the loss from the blocks c and d in utility terms. To see this, suppose that U′, on average, is U′ = 4 in block a, U′ = 2 in block b (and c) and U′ = 1 in block d. Because 4 − 2 > 2 − 1 and the difference in the location of blocks is equal to 1, we can safely conclude that U′ is a convex function. Thus, we have:

$$\Delta \equiv E_F U(x) - E_G U(x) = (a \cdot 4 - b \cdot 2) - (c \cdot 2 - d \cdot 1)$$

where a, b, c and d represent the area of the various blocks in absolute terms.

Because a = b and c = d we have:

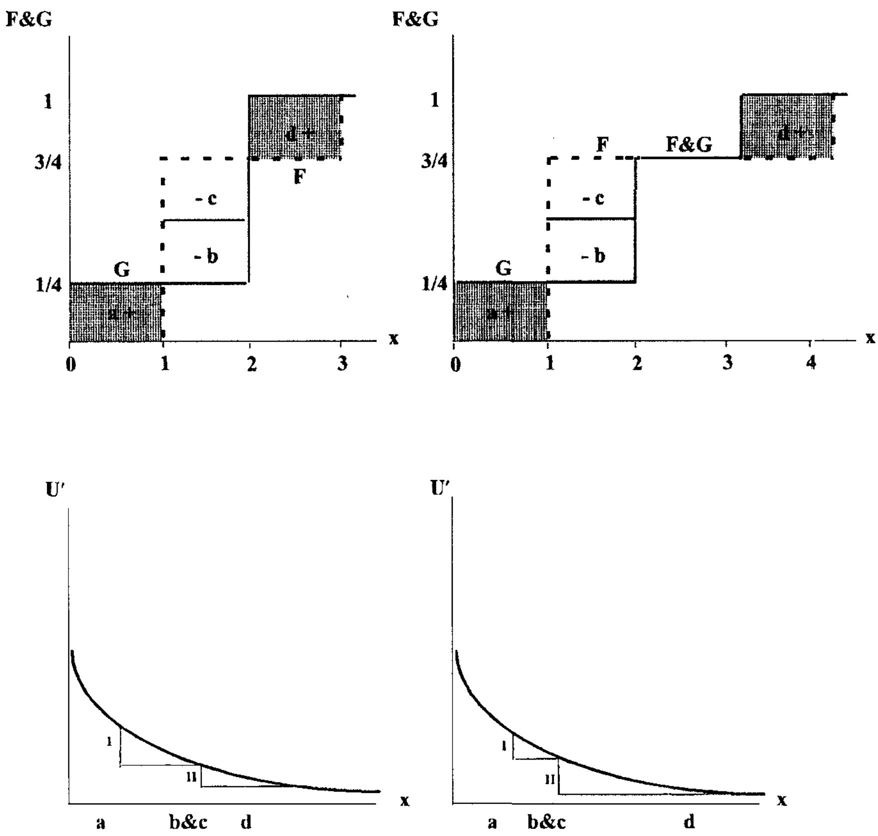
$$\Delta \equiv a\,(4-2) - c\,(2-1)$$

and because also c = a, we simply have:

$$\Delta = a\,(2-1) = a > 0$$

Note that if U′ were not convex Δ, could be negative. However, $U'''(x) \geq 0$ implies that U′(x) is convex; hence, Δ > 0 and there is TSD. This explanation allows for U′(x) to be zero and not strictly declining as long as the function U′(x) is even

Figure 3.20: The role of a convex U′(x) and the location of the "blocks areas" on the existence of TSD



weakly convex. For example, the explanation given above remains intact if $U' = 4$ in block a and $U' = 0$ in all other blocks. In such a case, U' will be convex albeit only weakly so in some segments.

A convex $U'(x)$ and the location of the various blocks is critical for the existence of such dominance. Figure 3.20b illustrates a case where block d is located far away to the right of block c. In such a case, even though U' is convex, we do not have TSD: $U'(a) - U'(b)$ may be smaller than $U'(c) - U'(d)$ because the distance corresponding to the various blocks is not the same as in Figure 3.20a. Thus, block d should be not far away from block c to guarantee the TSD of F over G. The convexity of $U'(x)$ and the location of the various blocks are crucial for the existence of TSD.

Finally, in the above examples, we have $E_F(x) = E_G(x)$ which implies that the total positive areas enclosed between F and G are equal to the total negative areas. If, on the other hand, $E_F(x) > E_G(x)$ then, in Figure 3.20a, either areas a or d (or both) increase, or areas b or c (or both), decrease and, *a fortiori*, we have TSD of F over G. Also, if $E_F(x) > E_G(x)$, even if block d shifted somewhat to the right of areas b+c, TSD may still exist. However, with such a shift, dominance is not guaranteed and $I_3(x)$ should be calculated to confirm the existence of TSD.

3.8 SUFFICIENT RULES AND NECESSARY RULES FOR $U \in U_3$

a) Sufficient rules

TSD is a necessary and sufficient decision rule for all $U \in \mathbf{U_3}$. However, here too, we can establish various sufficient rules and necessary rules for $U \in \mathbf{U_3}$ dominance. Here are a few examples:

Sufficient rule 1: FSD is a sufficient rule for TSD

If FD_1G, then $F(x) \leq G(x)$ for all x with at least one strong inequality. This implies that: $E_F(x) > E_G(x)$ and $I_3(x) \geq 0$ because FSD implies that the superior investment has a higher mean and that $I_1(x) = [G(x) - F(x)]$ is non-negative. However, because the integral of $I_1(x)$ is $I_2(x)$; $I_2(x) > 0$ and $I_3(x)$, which is the integral of $I_2(x)$ is also non-negative.

Sufficient Rule 2: SSD is a sufficient rule for TSD

If FD_2G, then:

$$I_2(x) = \int_a^x [G(t) - F(t)]dt \geq 0 \text{ for all x.}$$

Then:

$$I_3(x) = \int_a^x I_2(t)dt \geq 0 \text{ because } I_2(x) \geq 0.$$

$I_2(x) \geq 0$ for all x implies that it holds also for x = b; hence, $E_F(x) \geq E_G(x)$.

Thus, FD_2G implies that the two conditions required for TSD dominance hold; hence, FD_3G. To add one more explanation for these sufficiency rules recall that:

$$FD_1\, G \Rightarrow E_F\, U(x) \geq E_G\, U(x) \text{ for all } U \in \mathbf{U_1}$$

and because $\mathbf{U_1} \supseteq \mathbf{U_3}$, it is obvious that $E_FU(x) \geq E_GU(x)$ for all $U \in \mathbf{U_3}$. A similar explanation holds for the sufficiency of SSD because $\mathbf{U_2} \supseteq \mathbf{U_3}$.

Of course, many more sufficient rules are possible (e.g., $I_3(x) \geq a$ where $a > 0$ and $E_F(x) \geq E_G(x)$). However, the most important sufficient rules for $U \in \mathbf{U_3}$ are the FSD and SSD conditions.

b) Necessary rules for TSD

Necessary rule 1: The Means

Unlike FSD and SSD, TSD explicitly requires that $E_F(x) \geq E_G(x)$ in order to have FD_3G. This condition on the expected values is a necessary condition for dominance in $\mathbf{U_3}$. Note that for FSD and SSD we had to prove that this condition was necessary for dominance but for TSD, there is nothing to prove because it is explicitly required by the dominance condition.

Necessary rule 2: The Geometric means:

Suppose that FD_3G. Then:

$$E_F(\log(x)) \geq E_G(\log(x)) \text{ because } U_0(x) = \log(x) \in \mathbf{U_3}.$$

However, we have seen before that $E_F(\log(x)) \geq E_G(\log(x))$ implies that the geometric mean of F must be greater or equal to the geometric mean of G. Hence, it is a necessary condition for dominance in $\mathbf{U_3}$.

Necessary rule 3: The "Left Tail" condition

Like FSD and SSD, for FD_3G, the left tail of the cumulative distribution of G must be "thicker" than the left tail of F. In other words, $Min_F(x) \geq Min_G(x)$ is a necessary condition for FD_3G. To see this, suppose that the necessary condition does not hold. Namely, $Min_F(x) \equiv x_0 < Min_G(x) \equiv x_1$.

Then, for $x_1 > x > x_0$ we have:

$$I_3(x) = \int_a^x \int_a^z [G(t) - F(t)]dtdz < 0$$

because $F(x) = 0$ and $G(x) > 0$ in this range and, therefore, FD_3 G. Thus, if it is given that FD_3G, $Min_F(x) \geq Min_G(x)$; hence, it is a necessary condition for TSD.

3.9 DECREASING ABSOLUTE RISK AVERSION (DARA) STOCHASTIC DOMINANCE (DSD)

DSD stand for DARA Stochastic Dominance. Let us first discuss DARA utility functions and then turn to the discussion of DSD.

a) DARA utility functions.

Arrow and Pratt defined the absolute risk aversion measure "in the small" as the risk premium $\pi(w)$ given by:

$$\pi(w) = -\frac{U''(w)}{U'(w)}$$

where w is the investor's wealth. (Actually it is also multiplied by some positive constant ($\sigma^2/2$ or h/4) but this constant does not change the following analysis). It is claimed that investor's behavior reveals that $\frac{\partial \pi(w)}{\partial w} < 0$. Namely, the more wealth the investor has, the less, on average, will he/she be willing to pay for insuring against a given risk. This property is called decreasing absolute risk aversion (DARA). In the previous section, we saw that $\partial\pi(w)/\partial w < 0$ implies that $U'''(w) > 0$.

We also developed the TSD criterion corresponding to the set $U \in \mathbf{U}_3$. However, note that $\mathbf{U}_3$ is a set of utility functions which is wider than the set of all functions with decreasing absolute risk aversion (DARA) because $U'''(w)>0$ does not imply that $\partial\pi(w)/\partial w < 0$. If we denote the set of all DARA utility functions by $\mathbf{U}_d$, we will have the following relationship: $\mathbf{U}_d \subseteq \mathbf{U}_3$. In this section, we discuss a stochastic dominance for all DARA functions.

It is obvious from the above definitions of $\mathbf{U}_d$ and $\mathbf{U}_3$ that $FD_3G \Rightarrow FD_dG$ because FD_3G implies that $E_FU(x) \geq E_GU(x)$ for all $U \in \mathbf{U}_3$ and, because $\mathbf{U}_3 \supseteq \mathbf{U}_d$, $E_FU(x) \geq E_GU(x)$ also for all $U \in \mathbf{U}_d$. Because the converse is not true, we conclude that the DARA efficient set must be smaller or equal to the TSD efficient set.

FSD, SSD and TSD are relatively easy to employ. The criterion for all $U \in \mathbf{U}_d$ called Decreasing Absolute Risk Aversion Stochastic Dominance, or for short DARA Stochastic Dominance denoted by DSD, is much harder to employ. A natural way to analyze DSD is to write the utility function in terms of $\pi(x)$. Indeed, it is possible to express U(x) in terms of $\pi(x)$. To see this, first note that if the absolute risk aversion index $\pi(x) = -U''(x)/U'(x)$ is known at any point x (we use x here for the terminal value x = (1+R) where R is the rate of return), then there is full information on U(x). To see this, note that:

$$\pi(x) = \frac{-U''(x)}{U'(x)} = -\frac{\partial \log U'(x)}{\partial x}.$$

Hence:

$$-\int_a^x \pi(t)dt = \log U'(x) + c_1.$$

Thus:

$$U'(x) = e^{-\int_a^x \pi(t)dt + c_1}$$

and

$$U(x) = \int_a^x \left(e^{-\int_a^z \pi(t)dt + c_1}\right) dz + c_2, \text{ where } c_1, c_2 \text{ are constants.}$$

Therefore, if $\pi(x)$ is known for all wealth levels, then U(x) will be fully known (up to multiplicative and additive constants). Indeed, Hammond used this formulation to reach conclusions regarding preference of one investment over an other under various restrictions on $\pi(x)$.[13]

To find a dominance criterion for all $U \in \mathbf{U}_d$, we need a decision rule such that, if it holds, $E_FU(x) \geq E_GU(x)$ for all U with $\partial\pi(x)/\partial(x) < 0$. A natural way to achieve

---

[13] See J.S. Hammond III, "Simplifying the Choice Between Uncertain Prospects where Preference is Nonlinear, *Management Science, 20*, 1974;

this (as in the case of the other SD rules), is to carry out integration by parts of U(x). However, in this case, integration by parts such that $\pi'(x)$ will appear in the expression of U(x) does not lead to a clear rule as it did with the three stochastic dominance rules; hence, the DSD is very difficult to analyze.

Although Hammond reaches some important conclusions for a restricted group of investors, unfortunately, it is impossible to give a simple criterion for DSD stated only in terms of F and G as in the case of FSD, SSD and TSD. Rather we need to employ a relatively complicated algorithm procedure to prove or disprove DSD in specific given cases. These algorithms can be found in Vickson who provides several necessary conditions and several sufficient conditions for DSD.[14] Vickson also provides an algorithm for detecting DSD under various restrictions on the number of intersections between the two cumulative distributions under consideration. He also provides an example in which TSD does not hold and DSD holds, namely he shows that the DSD efficient set may be strictly smaller than the TSD efficient set. The detailed discussion can be found in Vickson; here we prove the equal means case, $E_F(x) = E_G(x)$, a case where DSD and TSD coincide.

b) DSD with equal mean distributions

Theorem 3.4: Let F and G be two cumulative distributions corresponding to two continuous random variables, with $E_F(x) = E_G(x)$. Then TSD and DSD are equivalent.

Proof:

Suppose that FD_3G. Then, because $\mathbf{U}_3 \supseteq \mathbf{U}_d$, it is obvious that FD_dG. Thus, $FD_3G \Rightarrow FD_dG$.

Now we have to show is that the opposite also holds. Namely, for equal mean distributions, $FD_dG \Rightarrow FD_3G$. To prove this claim, let $U_{k,x_0}(x), (k > 0, x_0 > a)$ be the utility function whose derivative is:

$$U'_{k,x_0}(x) = \begin{cases} 1/k \cdot e^{k(x_0 - x)} & \text{for } x \le x_0 \\ 1/k & \text{for } x > x_0. \end{cases}$$

The absolute risk aversion of $U_{k,x_0}(x)$ is a nonincreasing step function:

[14] See R.G. Vickson, "Stochastic Dominance Tests for Decreasing Absolute Risk Aversion in Discrete Random Variables," *Management Science*, 21, 1975 and "Stochastic Dominance Tests for Decreasing Absolute Risk Aversion II: General Random Variables," *Management Science*, 23, 1977.

$$\pi(x) \equiv -\frac{U''_{k,x_0}(x)}{U'_{k,x_0}(x)} = \begin{cases} k & \text{for } x \le x_0 \\ 0 & \text{for } x > x_0. \end{cases}$$

$\pi(x)$ is not differentiable everywhere; therefore, this utility function is not in $\mathbf{U}_d$. However, $U_{k,x_0}(x)$ can be closely approximated arbitrarily by a utility function in $\mathbf{U}_d$. The difference in expected utility can be written as:

$$E_F(U_{k,x_0}(x)) - E_G(U_{k,x_0}(x)) = \int_a^b [G(x) - F(x)]U'_{k,x_0}(x)dx$$

which, in our specific case of preference, reduces to:

$$E_F(U_{k,x_0}(x)) - E_G(U_{k,x_0}(x)) = \int_a^{x_0} [G(x) - F(x)]\frac{e^{k(x_0-x)}}{k}dx + \int_{x_0}^b [G(x) - F(x)]\tfrac{1}{k}dx.$$

By the Taylor expansion, for $x \in [a, x_0]$:

$$e^{k(x_0-x)} = 1 + k(x_0 - x) + \frac{k^2e^k(x_0 - \theta(x))}{2!}.$$

where $\theta(x)$ is the Cauchy residual, $a \le \theta(x) \le x_0$.

Therefore, for $x \in [a, x_0]$:

$$U'_{k,x_0}(x) = \frac{e^{k(x_0-x)}}{k} = \frac{1}{k} + (x_0 - x) + \frac{ke^k(x_0 - \theta(x))}{2!}.$$

Then using the fact that $1/k$ is a common factor of $U'_{k,x_0}(x)$ in the whole range of x, we obtain:

$$E_F(U_{k,x_0}(x)) - E_G(U_{k,x_0}(x)) = \int_a^b [G(x) - F(x)]\frac{1}{k}dx + \int_a^{x_0} [G(x) - F(x)](x_0 - x)dx$$
$$+ \int_a^{x_0} [G(x) - F(x)]\frac{ke^{k(x_0-\theta(x))}}{2!}dx.$$

But, because $\int_a^b [G(x) - F(x)]dx = E_F(x) - E_G(x)$, we obtain:

$$E_F(U_{k,x_0}(x)) - E_G(U_{k,x_0}(x)) = \frac{E_F(x) - E_G(x)}{k} + \int_a^{x_0} [G(x) - F(x)](x_0 - x)dx$$
$$+ \int_a^{x_0} [G(x) - F(x)] \frac{ke^{k(x_0 - \theta(x))}}{2!} dx.$$

Because, by assumption, $FD_dG, E_F(U_{k,x_0}(x)) \geq E_G(U_{k,x_0}(x))$ for all $k > 0$ and $x_0 > a$, the right-hand side must also be non-negative. Because by the assumption of the Theorem $E_F(x) = E_G(x)$ and letting $k \to 0$, the first term and the third term on the right-hand side are equal to zero; hence, the second term is non-negative. Thus, we now have to show that $\int_a^{x_0} (x_0 - x)[G(x) - F(x)]dx \geq 0$ (which does not depend on k) implies that $I_3(x) \geq 0$ for all values x; namely, for the equal means, distribution FD_dG also implies FD_3G.

Thus, we need to show that:

$$\int_a^{x_0} (x_0 - x)[G(x) - F(x)]dx \equiv \int_a^{x_0} (x_0 - x)I_1(x)dx = I_3(x_0).$$

To see this, write $I_3(x_0) = \int_a^{x_0} I_2(x)dx$. Integrating by parts yields:

$$\int_a^{x_0} (x_0 - x)I_1(x)dx = (x_0 - x)I_2(x)\Big|_a^{x_0} + \int_a^{x_0} I_2(x)dx = I_3(x_0).$$

Because x_0 can be chosen arbitrarily, $I_3(x) \geq 0$ for all x which completes the proof.

Note that the is emphasis here is on the two distributions having equal means. If $E_F(x) > E_G(x)$, then .1/k $[E_F(x) - E_G(x)] > 0$ and it is possible that FD_dG even though $I_3(x_0) < 0$. The preference of F over G, namely $E_FU_0(x) > E_GU_0(x)$ in this specific example is due to the difference in the means and $I_3(x_0)$ can be negative.

3.10 RISK-SEEKING STOCHASTIC DOMINANCE (RSSD)

We first present the risk-seeking SD rule and then provide a graphical and intuitive explanation.

a) The Risk-Seeking Stochastic Dominance (RSSD) Rule

In the financial and economic literature it is very uncommon to claim that risk-seeking prevails in the whole domain of outcomes. For example, Friedman and Savage[15] and Markowitz[16] claim that the observed behavior of people indicates that risk seeking prevails in some (but not in all) domains of outcomes. Kahneman and Tversky[17] advocate that risk-seeking prevails in the negative domain $x < 0$. Therefore, the risk-seeking criterion developed below is important as it will be used later on in the book for Prospect Theory value functions, and for Markowitz's preferences, which are not concave everywhere (see Chapter 15). Let us turn first to the RSSD criterion.

Suppose that the returns on the options under consideration fall in the risk-seeking domain. What is the appropriate SD rule? A utility function belongs to the set of risk-seeking utility function denoted by $\bar{U}_2$, if $U' \geq 0$ and $U'' \geq 0$ (and to avoid trivial cases there is at least one utility function with strict inequalities). Theorem 3.5 below provides the $\overline{\text{SSD}}$ rule which is the stochastic dominance rule corresponding to $\bar{U}_2$, i.e., for all risk seekers. Note that while SSD and U_2 correspond to risk-aversion $\overline{\text{SSD}}$ and $\bar{U}_2$ correspond to risk-seeking.

Theorem 3.5: Let F and G be two investments whose density functions are f(x) and g(x), respectively. Then F dominates G by $\overline{\text{SSD}}$ *denoted by* $\overline{\text{FD}_2\text{G}}$ *for all risk seekers if and only if,*

$$I_2^*(x) = \int_x^b [G(t) - F(t)]dt \geq 0$$

for all $x \in [a,b]$ *and there is at least one* x_0 *for which there is a strict inequality. This theorem can also be stated as follows:*

$$\int_x^b [G(t) - F(t)]dt \geq 0 \quad \Leftrightarrow \quad E_F U(x) \geq E_G U(x)$$

for all x with at least one strict inequality for some x_0	*for all* $U \in \bar{U}_2$ *with at least one* $U_0 \in \bar{U}_2$ *for which there is a strict inequality*

---

[15] Friedman, M. and L.J. Savage, "The utility of analysis of choices involving risk," *The Journal of Political Economics, 56,* 1948, pp. 279–304.

[16] Markowitz, H.M., "The Utility of Wealth," *The Journal of Political Economy, 60,* 1952b, pp. 151-156.

[17] Kahneman, D. and A. Tversky, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica 47*, 1979, pp. 263–291.

We will first prove the sufficiency of this theorem and then, the necessity. This will be followed by an intuitive explanation and graphical explanations.

Sufficiency

By eq. (3.2) we have,

$$E_F U(x) - E_G U(x) = U'(b)\int_a^b [G(t)-F(t)]dt - \int_a^b U''(x)(\int_a^x [G(t)-F(t)]dt\,dx$$

(3.5)

The second term can be rewritten as

$$-\int_a^b U''(x)(\int_a^x [G(t)-F(t)])dt\,dx = -\int_a^b U''(x)(\int_a^b [G(t)-F(t)]dt\;dx$$

$$+\int_a^b U''(x)\int_x^b ([G(t)-F(t)]dt\,dx$$

However, as we have

$$-\int_a^b [G(t)-F(t)]dt\int_a^b U''(x)dx = -\int_a^b [G(t)-F(t)]dt[U'(b)-U'(a)] \quad (3.5')$$

we can rewrite (3.5) as follows,

$$E_F U(x) - E_G U(x) = U'(a)\int_a^b [G(t)-F(t)]dt + \int_a^b U''(x)(\int_x^b [G(t)-F(t)]dt)dx \quad (3.6)$$

As by the $\overline{\text{SSD}}$ criterion $\int_x^b [G(t)-F(t)]dt \geq 0$ for all x and as $U''(x) \geq 0$ the second term on the right-hand side of eq. (3.6) is non-negative. The first term is non-negative as $\overline{\text{SSD}}$ criterion holds for all x and in particular for x = a and $U'(a) \geq 0$. Thus, if the $\overline{\text{SSD}}$ condition holds it implies that $E_F U(x) \geq E_G U(x)$.

Finally, in a proof similar to SSD, it can be easily shown that if there is a strict inequality in the $\overline{\text{SSD}}$ rule for some x, then there is some $U_0 \in \bar{U}_2$ such that $E_F U_0(x) > E_G U_0(x)$, which completes the sufficiency side of the proof.

Necessity

As in the proof of SSD, we also employ here the indirect method. Suppose that for some value x_0

$$\bar{I}_2(x) = \int_{x_0}^{b}[G(t)-F(t)]dt<0$$

Choose a risk-seeking utility function $U \in \bar{U}_2$ as follows:[18]

$$U_0(x) = \begin{cases} x_0 & x \le x_0 \\ x & x > x_0 \end{cases}$$

Plugging this function in eq. (3.6) the first term on the right-hand side vanishes (as $U'(a) = 0$) and what's left is

$$E_F U(x_0) - E_G U(x_0) = \int_{x_0}^{b} U_0''(x)\left(\int_{x_0}^{b}[G(t)-F(t)]dt\right)dx < 0$$

(for $x < x_0$, $U''=0$, hence the lower bound of the integral is x_0.)

Thus, in order to have $E_F U(x) \ge E_G U(x)$ for *all* $U \in \bar{U}_2$ we cannot have a violation of $\overline{SSD}$ rule even for one value x. Therefore, it is necessary for $\overline{SSD}$ that the condition of Theorem 3.5 holds.

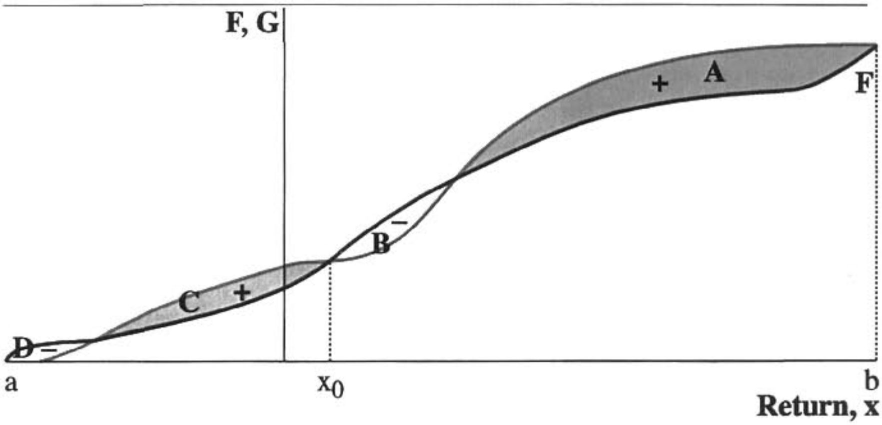
b. Graphical exposition of $\overline{SSD}$

As in SSD, also with $\overline{SSD}$ we calculate the area enclosed between the two cumulative distributions. However, this time the area accumulation is not done from the lower bound, a, to x but rather from the upper bound, b, to x, i.e., from the end point b up to any value x.

Figure 3.21 demonstrates two distributions F and G where F dominates G by $\overline{SSD}$. Let us elaborate. As can be shown, the last area denoted by A > 0, hence there is a chance that FDG by $\overline{SSD}$. Then A + B > 0 as area $|B|$ by construction is smaller than A. Of course, A + B +C > 0 and A + B + C + D > 0. Thus, the integral condition of Theorem 3.5 holds for all values x and F dominates G by $\overline{SSD}$.

[18]One can also establish a similar utility function which is differential in all points (see footnote 5).

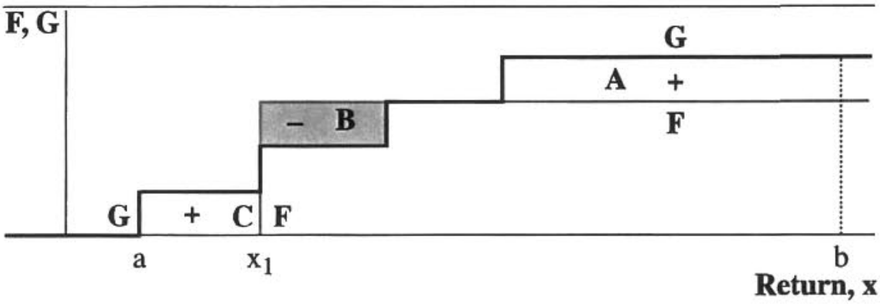
Figure 3.21: F dominates G by $\overline{SSD}$ (risk-seeking criterion)



c. *The Relationship Between SSD and* $\overline{SSD}$

One is tempted to believe that if F dominates G by $\overline{SSD}$ then G dominates F by SSD. This is not true and one counter example is sufficient to show this claim: in Figure 3.22 option F dominates G by $\overline{SSD}$, yet G does not dominate F by SSD. To see that G does not dominate F by SSD, note that integral $I_2(x) = \int_a^{x_1}[F(t)-G(t)]\,dt < 0$, hence G does not dominate F by SSD. F dominates F by $\overline{SSD}$, as A > 0 and A + B > 0, and of course A + B + C > 0, see Figure 3.22.

Figure 3.22: F dominates G by SSD and $\overline{SSD}$ but not by FSD



We have demonstrated that if one option dominates the other by $\overline{SSD}$ it is not necessary to have an opposite dominance by SSD. Yet such opposite dominance is possible in some specific cases. For example, Figure 3.14 reveals that dominates G by SSD and it is easy to verify that in this specific case, G dominates F by $\overline{\text{SSD}}$, as we have in this specific case,

$\int_x^b [F(t) - G(t)]dt \geq 0$, for all x,[19] i.e., we obtain $\overline{\text{SSD}}$ of G over F.

The intuitive explanation of $\overline{\text{SSD}}$ is similar to the intuitive explanation of SSD but this time $U'(x)$ is increasing with x rather than decreasing with x. To provide the intuitive explanation recall that

$$E_F U(x) - E_G U(x) = \int_a^b [G(x) - F(x)] U'(x)dx$$

By $\overline{\text{SSD}}$ criterion for each negative area enclosed between F and G there must be a larger positive area *located to the right* of it, (e.g., for area $B < 0$ there must be area $A > 0$ located to the right of it and $A > |B|$, see Figure 3.21). As each area is multiplied by $U'(x)$ and $U'(x)$ is increasing with x, the addition, in the utility terms, of the positive area to the term $\Delta \equiv E_F U(x) - E_G U(x)$ is larger than the deduction from Δ due to the negative area which proceeds it. Hence, when we sum up all positive and negative contributions to $\Delta \equiv E_F U(x) - E_G U(x)$, it adds up to a positive term, hence $E_F U(x) \geq E_G U(x)$.

d. The Relationship Between FSD, SSD and $\overline{\textbf{SSD}}$

If FSD holds, then $F(x) \leq G(x)$ which implies that both $\int_a^x [G(t) - F(t)]dt \geq 0$ and $\int_x^b [G(t) - F(t)]dt > 0$, hence SSD and $\overline{\text{SSD}}$ also hold. This makes sense as F dominates G for all $U \in \mathbf{U}_1$ and $\mathbf{U}_2$ and $\bar{\mathbf{U}}_2$ are subsets of $\mathbf{U}_1$.

So far, we have seen that FDG by FSD implies a dominance of F over G by SSD and by $\overline{\text{SSD}}$. Is it possible to have that one option dominates the other by SSD and $\overline{\text{SSD}}$ yet not by FSD? The answer is positive and one example is sufficient to show this claim. Consider, once again, F and G, as drawn in Figure 3.12b. As it can be easily shown that F dominates G by SSD and by $\overline{\text{SSD}}$, yet the two distributions cross and therefore, there is no FSD. Thus, it is possible that F dominates G for all $U \in \mathbf{U}_2$ and for all $U \in \bar{\mathbf{U}}_2$, but not for all $U \in \mathbf{U}_1$, which includes functions which neither belong to $\mathbf{U}_2$ nor to $\mathbf{U}_2^*$, e.g., functions with both concave and convex segments.

---

[19]It can be shown that the situation FDG by SSD and GDF by $\overline{\text{SSD}}$, is possible only if $E_F(x) = E_G(x)$, as the linear utility function is a borderline between risk-seeking and risk averse utility functions.

3.11 N^{TH} ORDER STOCHASTIC DOMINANCE

So far we have discussed First, Second, and Third Degree Stochastic Dominance (as well as DSD) where FSD assumes $U'>0$, SSD assumes $U'>0$ and $U'' < 0$ and TSD assumes $U'>0$, $U''<0$ and $U''' >0$. Now suppose we know that $U''''<0$. We can then integrate the last term of eq. (3.4) by parts to obtain:

$$E_F U(x) - E_G U(x) = U'(b)I_2(b) - U''(b)I_3(b) + U'''(b)I_4(b) - \int_a^b U''''(x)I_4(x)dx$$

where:

$$I_4(x) = \int_a^x \int_a^v \int_a^z [G(t) - F(t)]dtdzdv$$

Thus, if $I_2(b) \geq 0$, $I_3(b) \geq 0$ and $I_4(x) \geq 0$ for all x, then $E_F U(x) \geq E_G U(x)$ for all U $\in$ $\mathbf{U_4}$ where U $\in$ $\mathbf{U_4}$ if $U' \geq 0$, $U'' \leq 0$, $U''' \geq 0$ and $U'''' \leq 0$. In such a case, we say that F dominates G by the fourth order stochastic dominance of FD_4G.

Let us now generalize the results by defining n^{th} order stochastic dominance.

Theorem 3.5: Suppose that we have information on the first n derivatives. Define U $\in$ $\mathbf{U_n}$ as the set of all utility functions such that all odd derivatives are positive and all even derivatives are negative. Then FD_nG (or F dominates G by n^{th} order stochastic dominance) if and only if:

$$I_j(b) \geq 0 \qquad j = 1, 2, \ldots, n-1$$
$$I_n(x) \geq 0 \qquad \text{for all } x$$

and there is at least one strict inequality.

The proof is a trivial extension of the previous discussion of $\mathbf{U_4}$. We simply integrate the last term by parts again and again until we obtain:

$$E_F U(x) - E_G U(x) = \sum_{j=1}^{n-1} (-1)^{j+1} U^j(b) I_{j+1}(b) + \int_a^b U^n(x) I_n(x)dx$$

if n is an odd number

and:

$$E_F U(x) - E_G U(x) = \sum_{j=1}^{n-1} (-1)^{j+1} U^j(b) I_{j+1}(b) - \int_a^b U^n(x) I_n(x)dx$$

if n is an even number.

Because, by assumption, all odd derivatives are positive and all even derivatives are negative, then if $I_j(b) \geq 0$ for all $j \geq 1, 2, .., n-1$ and $I_n(x) \geq 0$, for all x,

$$E_F U(x) \geq E_G U(x) \text{ for all } U \in \mathbf{U_n}$$

Note that, unlike the transparent intuition of FSD, SSD and even the intuition of TSD, the economic intuition of high order stochastic dominance is somewhat vague. However, there are some important utility functions with positive odd derivatives and negative even derivatives. For example the change in signs of the derivatives holds for $U(x) = \log(x)$, for X^{α}/α where $\alpha < 1$, and for $-e^{-\alpha x}$ $(\alpha > 0)$.

3.12 STOCHASTIC DOMINANCE RULES: EXTENSION TO DISCRETE DISTRIBUTIONS

In this section, we extend all the previous stochastic dominance results to the case of discrete random variables.[20] Suppose that we have a discrete distribution with jumps at $x_0 = a, x_1, x_2, \ldots, x_n$ and $x_{n+1} = b$, namely $a \leq x \leq b$. Then the probability of x_i occurring will be:

$$p(x_i) = p\,(x = x_i) = F(x_i) - F(x_{i-1}),\ i = 1,2\ldots,n+1$$

where $F(x_i)$ is the cumulative probability. The expected utility of investment F with a discrete distribution is given by:

$$\begin{aligned} E_F U(x) &= \sum_{i=0}^{n+1} p(x_i)U(x_i) = F(x_0)U(x_0) + \sum_{i=1}^{n}[F(x_i) - F(x_{i-1})]U(x_i) + (1 - F(x_n))U(b) \\ &= F(x_0)U(x_0) - F(x_n)U(x_{n+1}) + U(b) + \sum_{i=1}^{n} F(x_i)U(x_i) - \sum_{i=1}^{n} F(x_{i-1})U(x_i) \end{aligned}$$

(Note that $1-F(x_n) = P(X = x_{n+1})$ and $X_{n+1} = b$)

Using the following relationship:

$$\begin{aligned} &F(x_0)U(x_0) - F(x_n)U(x_{n+1}) + \sum_{i=1}^{n} F(x_i)U(x_i) \\ &= \sum_{i=1}^{n} F(x_{i-1})U(x_{i-1}) - F(x_n)[U(x_{n+1}) - U(x_n)] \end{aligned}$$

[20] Actually, this follows from the properties of Riemann - Stieltjes integral. However, we provide here the detailed proof.

the expected utility, $E_FU(x)$ can be rewritten as:

$$F_FU(x) = -\sum_{i=1}^{n+1} F(x_{i-1})[U(x_i) - U(x_{i-1})] + U(b).$$

Note that the last term in the summation for i = n + I is simply

$F(x_n)(U(x_{n+1}) - U(x_n))$.

Using the following relationship:

$$U(x_i) - U(x_{i-1}) = \int_{x_{i-1}}^{x_i} U'(x)dx,$$

we obtain:

$$E_FU(x) = U(b) - \sum_{i=1}^{n+1} F(x_{i-1}) \int_{x_{i-1}}^{x_i} U'(x)dx,$$

and, because F(x) is constant at each interval of x the expected utility can be rewritten as:

$$E_FU(x) = U(b) - \sum_{i=1}^{n+1} \int_{x_{i-1}}^{x_i} F(x_{i-1})U'(x)dx.$$

Furthermore, because:

for	$i = 1$	$x_{i-1} = x_0 = a$
and for	$i = n + 1$	$x_{n+1} = b$,

we finally obtain:

$$E_FU(x) = U(b) - \int_a^b F(x)U'(x)\,dx$$

Applying the same technique to distribution G, we obtain:[21]

---

[21] Note that G may not start at $x_0 = 0$ or may not end at $x_{n+1} = b$. In such a case, simply add these two values with zero probability to obtain the same formulation as for F. In such a case, we may have $p(x_0) = 0$ and $p(x_n) = 0$ but this does not change the generality of the above proof.

$$E_G U(x) = U(b) - \int_a^b G(x)U'(x)\,dx$$

and hence:

$$E_F U(x) - E_G U(x) = \int_a^b [G(x) - F(x)]U'(x)dx$$

as obtained in eq. (3.1) for the continuous random variables.

Using this formula for discrete random variables, integration by parts can be carried out to prove the SD rules in exactly the same manner as with the continuous random variables. Therefore, we can conclude that all the results regarding FSD, SSD, TSD, DSD, etc., hold for continuous and discrete random variables alike. In the following example, we conduct a direct calculation of expected utility and show that the same results are obtained by employing eq. (3.1) in spite of the fact that we have a discrete random variables.

Example:

Suppose that we have the following two discrete random variables:

G		**F**	
X	**p(x)**	**X**	**p(x)**
1	½	0	½
2	½	4	½

and assume that $U(x) = x^2$ (for $x > 0$) with $U'(x) = 2x > 0$.

Note that eq. (3.1) holds for all utility functions; hence there is no need to assume risk aversion. Therefore, $U(x) = x^2$ (for $x > 0$) can be safely employed to check whether eq. (3.1) is intact for discrete random variables. The expected utility of F and G can be calculated directly as follows:

$E_G U(x) = \frac{1}{2}\,1^2 + \frac{1}{2}\,2^2 = 2.5$

$E_F U(x) = \frac{1}{2}\,0^2 + \frac{1}{2}\,4^2 = 8$

Hence, $\Delta \equiv E_F U(x) - E_G U(x) = 8 - 2.5 = 5.5$

Let us check whether we get the same answer by employing eq. (3.1).

Figure 3.21 provides the distributions F(x) and G(x) corresponding to our example. As can be seen, in the range $0 \le x \le 1$, $G(x) - F(x) = -½$, in the range $1 < x \le 2$, $G(x) - F(x) = 0$, and in the range $2 \le x \le 4$, $G(x) - F(x) = ½$ and for $x < 0$ and $x > 4$, $G(x) - F(x) = 0$. Also $U'(x) = 2x$.

Thus, using eq. (3.1) we obtain:

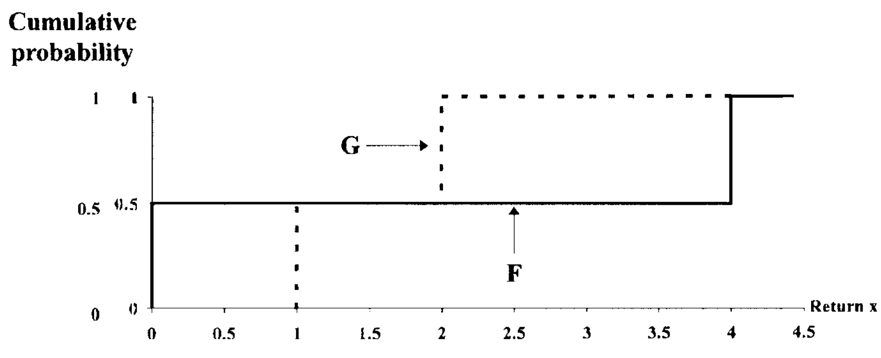
$$E_F U(x) - E_G U(x) = \int_0^4 [G(x) - F(x)]U'(x)dx = \int_0^1 (-1/2)\cdot 2xdx + \int_1^2 0\cdot 2xdx + \int_2^4 (1/2)\cdot 2xdx$$

$$= -\frac{x^2}{2}\Big|_0^1 + \frac{x^2}{2}\Big|_2^4 = -[1/2 - 0] + [8 - 2] = 5½$$

exactly as obtained with the direct calculation.

In the proof of eq. (3.1) we assume that x is bounded by a and b, namely $x \in [a,b]$. For instance, an investor in the stock market may choose $a = -100\%$ and b = a very large number, say 10 billion percent, which practically covers all possible returns. Moreover, we assume that for both F and G, $X_0 = a$ and $x_{n+1} = b$. What is the effect on the proof of eq. (3.1) for discrete random variables if the range of one distribution is wider than the range of the other distribution? To illustrate, suppose that alternatively we select $a_1 < a$

Figure 3.23: The Cumulative Distributions



and $b_1 > b$. Does the extension of the bound from [a,b] to $[a_1,b_1]$ affect the value given in eq. (3.1)? Generally, for the calculation of EU(x), we take a to be the lower value and b the higher value. However, if we were to take a lower value than a as a lower bound or a higher value than b as an upper bound, the results of the expected utility calculation would remain unchanged. Take, for example, $a_1 < a$, then:

$$E_FU(x) = U(b) - \int_{a_1}^{b} F(x)U'(x)dx =$$
$$= U(b) - \int_{a_1}^{a} F(x)U'(x)dx - \int_{a_1}^{b} F(x)U'(x)dx$$
$$= U(b) - \int_{a}^{b} F(x)U'(x)dx$$

because $F(x) = 0$ for $x < a$.

Now take $b_1 > b$ as an upper bound.

Here we get:

$$E_FU(x) = U(b_1) - \int_{a}^{b} F(x)U'(x)dx = U(b_1) - \int_{a}^{b} F(x)U'(x)dx - \int_{b}^{b_1} F(x)U'(x)dx$$

But, because $F(x) = 1$ for $x > b$, we obtain:

$$E_FU(x) = U(b_1) - \int_{a}^{b} F(x)U'(x)dx - \int_{a}^{b_1} U'(x)dx$$
$$= U(b_1) - \int_{a}^{b} F(x)U'(x)dx - [U(b_1) - U(b)]$$
$$= U(b) - \int_{a}^{b} F(x)U'(x)dx$$

exactly as obtained before when b rather than b_1 was the upper bound. Thus, the lower and upper bounds [a,b] can be selected arbitrarily without affecting the results, and without affecting eq. (3.1), as long as:

$a \leq$ minimum (x_F, x_G)
$b \geq$ maximum (x_F, x_G).

In the above example, we selected $a = 0$, $b = 4$ which complies with this requirement. However, selecting say $a = -10$ and $b = +20$ will not change the value of $E_FU(x)$, $E_GU(x)$, or the difference between the expected values.

3.13 THE ROLE OF THE MEAN AND VARIANCE IN STOCHASTIC DOMINANCE RULES

In all the SD rules discussed above, $E_F(x) \geq E_G(x)$ is a necessary condition for dominance. It is natural to ask whether there is a condition on the variances which is also a necessary condition for dominance. The answer is generally negative. To see this, take the following example:

	Investment G		Investment F	
	X	p(x)	X	p(x)
	1	½	2	½
	2	½	4	½
Expected rate of return:	1.5		3	
Variance:	¼		1	

Thus, $\sigma_F^2 > \sigma_G^2$, yet F dominates G by FSD, hence also by SSD, TSD and DSD as well as higher order stochastic dominance. Therefore, we can conclude that the superior investment does not necessarily have a lower variance. However, if the two random distributions under consideration have equal means ($E_F(x) = E_G(X)$), then a necessary condition for the dominance of F over G by SSD (FD_2G) is that $\sigma_F^2 \le \sigma_G^2$. To see this, recall that $FD_2G \Rightarrow E_FU(x) \ge E_GU(x)$ for all $U \in \mathbf{U_2}$. Take the quadratic utility function $U_0(x) = x + b\,x^2$ when $b < 0$ (it is defined only for the range $U'(x) \ge 0$), hence $U_0 \in \mathbf{U_2}$.

$$EU_0(x) = Ex + b(Ex)^2 + b\,\sigma_x^2 \text{ (because } \sigma_x^2 = Ex^2 - (Ex)^2\text{)}.$$

Suppose that x corresponds to distribution F and y to distribution G. Therefore:

$$\Delta \equiv E_FU_0(x) - E_GU_0(y) = (Ex - Ey) + b[(Ex)^2 - (Ey)^2] + b(\sigma_x^2 - \sigma_y^2).$$

Because $U_0 \in \mathbf{U_2}$ and FD_2G (by assumption), we know that $\Delta \ge 0$. By assumption $Ex = Ey$, therefore:

$$\Delta \ge 0 \Rightarrow b(\sigma_x^2 - \sigma_y^2) \ge 0$$

and because $b < 0$, we must have $\sigma_x^2 \le \sigma_y^2$. Thus, for equal mean distributions, $\sigma_x^2 \le \sigma_y^2$ is a necessary condition for SSD dominance of x over y.

The quadratic function $U_0(x)$ cannot be used to show that $\sigma_F^2 \le \sigma_G^2$ is also a necessary condition for TSD dominance of F and G because the quadratic utility function is not included in $\mathbf{U_3}$. However, we can choose:

$$U_k(x) = -(x - b)^2 + 1 - e^{-kx}$$

where k > 0 and b is the upper bound of x. It can be easily verified that $U_k(x) \in U_3$. For $k \to 0, 1 - e^{kx} \to 0$ and TSD dominance also implies that for this specific function, $EU(x) \geq EU(y)$, namely:

$$-E(x-b)^2 \geq -E(y-b)^2$$

or:

$$Ex^2 - 2Ex\,b + b^2 \leq Ey^2 - 2Ey\,b + b^2$$

where x and y correspond to F and G, respectively. For Ex = Ey , it implies that a necessary condition for TSD dominance of x over y is:

$$Ex^2 \leq Ey^2$$

or:

$$Ex^2 - (Ex)^2 = \sigma_x^2 \leq Ey^2 - (Ey)^2 = \sigma_y^2$$

(recall that Ex = Ey by assumption) which completes the proof.

Finally, if $E_F(x) > E_G(x)$, the condition $\sigma_F^2 \leq \sigma_G^2$ is neither sufficient nor necessary for SSD and TSD dominance. The fact that this condition is not necessary for dominance is shown in the above example when FD_2G (hence also FD_3G) and $\sigma_F^2 > \sigma_G^2$. To show that the condition is not sufficient for SSD and TSD dominance take the following example provided by Hanoch and Levy (1969)[22]:

	Investment F			Investment G	
	X	**p(x)**		**X**	**p(x)**
	1	0.80		10	0.99
	100	0.20		1000	0.01
Expected rate of return:	20.8		>	19.9	
Variance:	1468		<	9703	

F has a higher mean and a lower variance. Does this mean that for every $U \in \mathbf{U_2}$, (and $U \in \mathbf{U_3}$), F will be preferred over G? Not really. Take the utility function

[22] Hanoch, G. and H. Levy, "The Efficiency Analysis of Choices Involving Risk," *Review of Economic Studies*, 36, pp. 335–346, 1969.

$U_0 = \log_{10}(x)$ where $U_0 \in \mathbf{U_2}$ (and also $U_0 \in \mathbf{U_3}$).

To show this claim, we conduct a simple calculation with $U_0(x) = \log_{10} x$ and the above example reveals that:

$$E_F U(x) = 0.4 < E_G U(x) = 1.02$$

and the distribution with the lowest mean and highest variance is preferred by this specific risk averse preference! Thus, $Ex \geq Ey$ and $\sigma_x \leq \sigma_y$ is not sufficient for SSD (and TSD) dominance. This might seem surprising but is it really? Recall that G has a much higher positive skewness and as with log function $U''' > 0$, the preference of G over F may be due to the large positive skewness of G. Hence the preferred investment by this specific risk averse investor may have a lower mean and a higher variance.

3.14 SUMMARY

In this chapter we discussed stochastic dominance rules for the partial ordering of uncertain projects. The most important rules are FSD (for $U \in \mathbf{U_1}$), SSD for $U \in \mathbf{U_2}$ and TSD for $U \in \mathbf{U_3}$. We also discussed the DSD criterion for $U \in \mathbf{U_d}$, $\overline{\text{SSD}}$ for $U \in \bar{U}_2$ (risk seeking) as well as n^{th} order stochastic dominance. U_1 is defined as the set of all preferences with $U' \geq 0$, $\mathbf{U_2}$ is the set of all preferences with $U' \geq 0$ and $U'' \leq 0$, $\mathbf{U_3}$ is the set of preferences with $U' \geq 0$, $U'' \leq 0$ and $U''' \geq 0$, $\mathbf{U_d}$ is the set of all preferences with $U' \geq 0$, $U'' \leq 0$, $U''' \geq 0$, and $\partial\pi(w)/\partial(w) \leq 0$ where $\pi(w)$ is the risk premium is defined as $-U''(w)/U'(w)$. The set $\bar{U}_2$ is the set of all preferences with $U' \geq 0$ and $U'' \geq 0$.

We have the following relationships:

a) $\mathbf{U_1} \supseteq \mathbf{U_2} \supseteq \mathbf{U_3} \supseteq \mathbf{U_d}$

b) $\mathbf{U_1} \supseteq \bar{U}_2$

c) FSD $\Rightarrow$ SSD $\Rightarrow$ TSD $\Rightarrow$ DSD

d) FSD $\Rightarrow$ $\overline{\text{SSD}}$

Figure 3.24: The Utility Classes and the Resulting Efficiency Sets

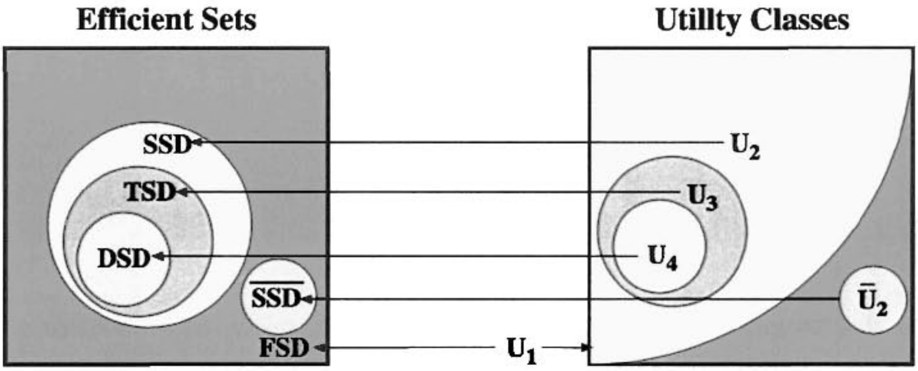


Figure 3.24 summarizes the relationship between the various groups of utilities and the resulting efficient sets. As we can see, the more information we assume on U, the smaller the class of preferences and the smaller the size of the resulting efficient set. Yet, the $\bar{U}_2$ set is only a subset of U_1 and hence $\overline{SSD}$ efficient set is a subset of FSD and as there is no relationship between $\bar{U}_2$ and other sets of preferences there is no relationship between the $\overline{SSD}$ efficient set and the other efficient set. The inefficient set corresponding to each SD rule is the feasible set less the corresponding efficient set.

Key Terms

Complete Ordering

Partial Ordering

Feasible Set (FS)

Efficient Set (ES)

Inefficient Set (IS)

Objective Decision

Subjective Decision

Optimal Portfolio

First Degree Stochastic Dominance (FSD)

Probability Function

Density Function

Cumulative Probability Function

Necessary Rules

Sufficient Rules

The Means Rule

The Geometric Means Rule

Linear Utility Function

Logarithmic Utility Function

The "Left Tail" Condition

Investment Screening

Ineffective Decision Rule

Risk Aversion

Jensen's Inequality

Fair Game

Second Degree Stochastic Dominance (SSD)

Skewness

Positive Skewness

Decreasing Absolute Risk Aversion (DARA)

Third Degree Stochastic Dominance (TSD)

Symmetrical Distributions

Convex Function

Concave Function

DARA Stochastic Dominance (DSD)

Risk Seeking

Risk Seeking Stochastic Dominance ($\overline{SSD}$)

n^{th} Order Stochastic Dominance

4

STOCHASTIC DOMINANCE: THE QUANTILE

In Chapter 3 the various stochastic dominance rules were stated in terms of cumulative distributions denoted by F and G. In this chapter FSD and SSD stochastic dominance are restated in terms of distribution *quantiles*. Both methods yield the same partition of the feasible set into efficient and inefficient sets. The formulas and the stochastic dominance rules based on distribution quantiles are more difficult to grasp intuitively but, as will be shown in this chapter, they are more easily extended to the case of diversification between risky asset and riskless assets. They are also more easily extended to the analysis of stochastic dominance among specific distributions of rates of return (e.g., lognormal distributions). Such extensions are quite difficult in the cumulative distribution framework.

In Chapter 3, we showed that the SD based on cumulative distributions is optimal. In this chapter, we will show that the rules based on quantiles lead to the same dominance relationship as those based on cumulative distributions and, therefore they are also optimal. However, note that while the shift from cumulative distributions to quantiles is legitimate with FSD and SSD, this is not the case with TSD and higher stochastic dominance rules, where only the cumulative formulas are correct (see Chapter 5).

4.1 THE DISTRIBUTION QUANTILE

Let us first introduce some of the definitions and notations used in this chapter. The P^{th} quantile of a distribution ($0 < P \le 1$) is defined as the *smallest* possible value Q(P) for the following to hold:

$$p\,(X \le Q(P)) \ge P \tag{4.1}$$

where p denotes probability and P denotes cumulative probability. For convenience, the quantile for P = 0 is defined as the minimum value of X (the random variable) if it exists.[1] Hence, the whole range $0 \le P \le 1$ is covered: Q(P) is the smallest possible value of the random variable where the probability of obtaining this value or a lower value is at least P. For a continuous density function, there will always be an equality on the right-hand side of eq. (4.1). Because the random variable is generally denoted by X, and a specific value by x, the accumulated value of probability P up to this value is denoted by x_p. The value, x_p thus defined is equal to Q(P); hence it is also the P^{th} quantile. Therefore:

---

[1] For example, in rolling a die, Q(0)=1. This definition is appropriate only in the case if the quantile exists. For example, the quantile is not defined in the normal distribution where P = 0 and P = 1. Thus, without loss of generality, we define Q(P) for all $0 \le P \le 1$, bearing in mind that in some cases, we need to confine ourselves to the range $0 < P < 1$.

$$p\,(X \le x_P) = p\,(X \le Q(P)) = P(x_P) = F(X_p).$$

Thus, $X_P = Q(P)$ stands for the P^{th} quantile and $P(x_p) = F(x_p)$ stands for the cumulative probability, and either of these notations can be employed. For a strictly increasing cumulative distribution denoted by F, the quantile is defined as the inverse function:

$$Q(P) = x_P = F^{-1}(P)$$

where x_p is defined as above. In the rest of this chapter we use either x_P or Q(P) for the P^{th} quantile and either F(x) or P(x) for the cumulative distribution. Generally, F denotes the cumulative probability. However, if and when we wish to compare two distributions, we reserve the notations F and G for the risky projects and, therefore, use the notation P (rather than F or G) for the cumulative probability.

Example:

Let x have a uniform distribution in the range (a, b) where a = 1, b = 3. Because b −a = 3−1 = 2, we have f(x) = ½ where f(x) is the density function of x. The cumulative distribution of a uniform distribution is given by:

$$F(x)=\begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x \ge b \end{cases}$$

And in our specific case:

$$F(x)=\begin{cases} 0 & x < 1 \\ \dfrac{x-1}{2} & 1 \le x \le 3 \\ 1 & x \ge 3 \end{cases}$$

In the range $1 \le x \le 3$, the distribution F(x) is given by $F(x) = \dfrac{x-1}{2}$. The inverse function, F^{-1} is in the range, $x = 2F(x) + 1$. However, because $F(x) = P(x) = p(X \le x)$, by definition, x is the P^{th} quantile, which we denote by Q(P). Therefore, we have the following relationship between Q(P) and F(x):

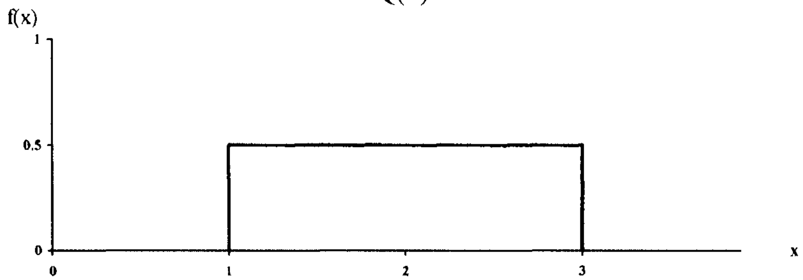
$$Q(P) = 2P(x) + 1 = 2F(x) + 1.$$

Recall F(x) = P(x); hence the relationship can be rewritten as Q(F(x))= 2F(x) + 1. For example, if x = 1, F(x) = 0 (or P =0) and $Q(0) = 2 \cdot 0 + 1 = 1$.

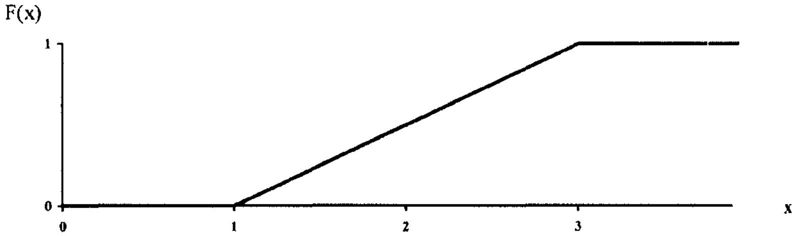
Similarly, if F(x) = ½, then Q(½) = 2 · ½ + 1 = 2 and, finally, if F(x) = 1. then Q(1) = 2 · 1 + 1 = 3. This example illustrates that the P^{th} quantile is the value of x such that up to this value, the accumulated probability is exactly P.

Figure 4.1 demonstrates the density function f(x), the cumulative distribution F(x) and the quantile (inverse) function Q(P), corresponding to our specific example. First, note that because f(x) is continuous and f(x) > 0 in the range 1 ≤ x ≤ 3, F(x) is a monotonically increasing function in the range 1 ≤ x ≤ 3.

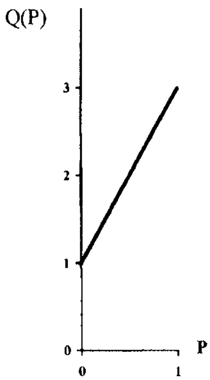
Figure 4.1: The density function f(x), the cumulative distribution F(x), and the quantile function
Q(P)



4.1a: The density function



4.1b: The cumulative distribution



4.1c: The quantile function

The slope of F(x) = ½ and the slope of the quantile function is 2. The function F(x) represents the quantile function because for every value P, it gives us the corresponding value x_P (or Q(P)). To see this, draw a horizontal line (not shown in Figure 4.1) from the vertical axis at a specific value P, until it intersects F, and then a vertical line to the horizontal axis until it crosses the horizontal axis at the p^{th} quantile x_p. If we were to plot the quantile on the vertical axis and P on the horizontal axis, the graph would change and the *inverse* function would provide the distribution quantile (see Figure 4.1b). Although the graph changes when we change the role of the axes, the basic relationship between P (or F) and Q(P) remains the same. However, as our aim here is to demonstrate the quantile approach in comparison to the cumulative distribution approach presented in Chapter 3, it is more appropriate not to reverse the role of the axes and to leave the quantile on the horizontal axis.

For continuous density functions with strictly increasing cumulative distributions, Q(P) is the value such that up to Q(P), the cumulative probability is *exactly* P. However, for discrete distributions (or for continuous functions with strictly increasing non-cumulative distributions), there are many values x corresponding to a given value P, and, therefore, the quantile definition will be modified by eq. (4.1). Thus, for a discrete distribution, Q(P) is defined as the smallest value in each step of the cumulative distribution (see eq. (4.1)), bearing in mind that for monotonically increasing cumulative distributions, we have equality on the right-hand side of eq. (4.1). Let us demonstrate the relationship between P(x) = F(x) and Q(P) for a discrete distribution.

Example:

Suppose that we roll a balance die. Hence, we have:

$$F(x)=\begin{cases} 0 & \text{for } x<1 \\ 1/6 & \text{for } 1 \le x < 2 \\ 2/6 & \text{for } 2 \le x < 3 \\ 3/6 & \text{for } 3 \le x < 4 \\ 4/6 & \text{for } 4 \le x < 5 \\ 5/6 & \text{for } 5 \le x < 6 \\ 1 & \text{for } x \ge 6 \end{cases}$$

The quantile function Q(p) as defined above is given by:

$$Q(P)=\begin{cases} 1 & \text{for } 0 \le P \le \frac{1}{6} \\ 2 & \text{for } \frac{1}{6} < P \le \frac{2}{6} \\ 3 & \text{for } \frac{2}{6} < P \le \frac{3}{6} \\ 4 & \text{for } \frac{3}{6} < P \le \frac{4}{6} \\ 5 & \text{for } \frac{4}{6} < P \le \frac{5}{6} \\ 6 & \text{for } \frac{5}{6} < P \le 1 \end{cases}$$

Thus, if F(x) is a step function, the quantile Q(P) will also be a step function.

Note, however, that Q(P) is continuous from the left, whereas F(x) is continuous from the right.

Figure 4.2 illustrates the probability function p(x), the cumulative distribution P(x) = F(x), and the quantile function Q(P) corresponding to our discrete example. Note that because p(x) = 1/6 for x = 1,2,...,6, we obtain jumps in F(x) at these values. In each range (e.g., in the range $3 < x < 4$), F(x) is constant because there is zero probability p(x) for the values x where $3 < x < 4$.

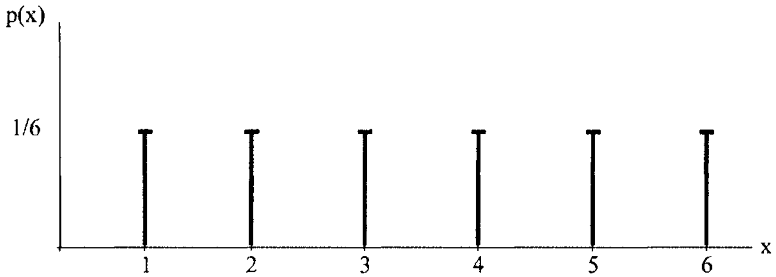
Let us now examine the graph of the quantile function. We defined the P^{th} quantile in this case as the minimum value in each range (see eq. (4.1)). In our specific example, this minimum value is Q(P) = 1 for $0 < P \le 1/6$. Similarly, in the range $1/6 < P \le 2/6$, Q(P) is the smallest value (i.e., equal to 2), and so on until finally, for $5/6 < P \le 1$, we have Q(1) = 6 (see Figure 4.2c). Finally, note that by eq. (4.1), for P = 0, the quantile is not defined (it is $-\infty$). However, we defined it as +1, the minimum value that has a positive probability, if it exists (see footnote 1).

For simplicity of the presentation, most of the proofs and illustrations of the stochastic dominance criteria deal with continuous density functions with monotonically increasing cumulative distributions in the relevant range of x. However, the results remain intact for discrete distributions. Later on in the book, when we discuss the empirical evidence of stochastic dominance studies, we will return to the discrete definition of the quantile because empirical cumulative distributions are, by construction, step functions.

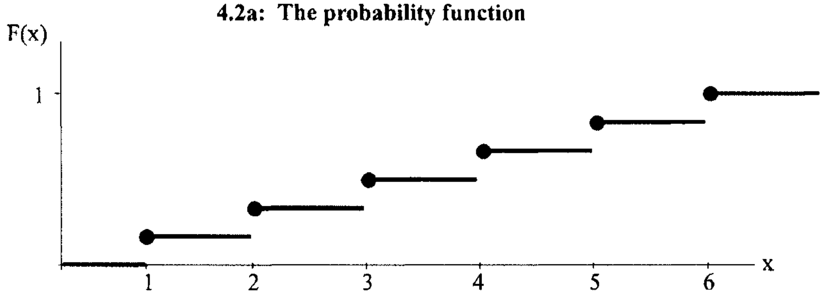
4.2 STOCHASTIC DOMINANCE RULES STATED IN TERMS OF DISTRIBUTION QUANTILES

For some specific distributions (e.g, normal distributions), the quantile does not exist for P = 0 and P = 1 (see footnote 1). In such cases, we confine ourselves to the range, $0 < P < 1$. However, in the rest of the chapter we relate to the range $0 \le P \le 1$, bearing in mind that the various statements relate to each P in the range $0 \leq P \leq 1$ only if it exists.

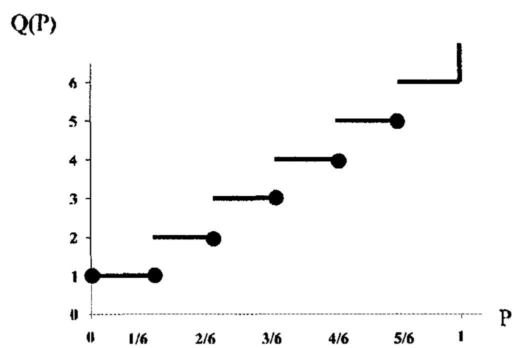
Figure 4.2: Probability function p(x), the cumulative distribution F(x), and the quantile function Q(P)



4.2a: The probability function



4.2b: The cumulative distribution



4.2c: The quantile function

a) The FSD Rule with Quantiles

The following Theorem (Theorem 4.1) formulates FSD in terms of the distribution quantile.

Theorem 4.1:

Let F and G be the cumulative distributions of the return on two investments. Then FD_1G if and only if:

$$Q_F(P) \geq Q_G(P) \text{ for all } 0 \leq P \leq 1$$

and there is at least one value P_0 for which a strict inequality holds.

Proof:

We first assume monotonic increasing cumulative distributions (in the relevant range) and show that the condition $Q_F(P) \geq Q_G(P)$ for all P, implies that F is never above G and, conversely, if F is below G (in the weak sense) everywhere, $Q_F(P)$ will also be greater than $Q_G(P)$ for all P. Specifically, we need to prove that the following holds:

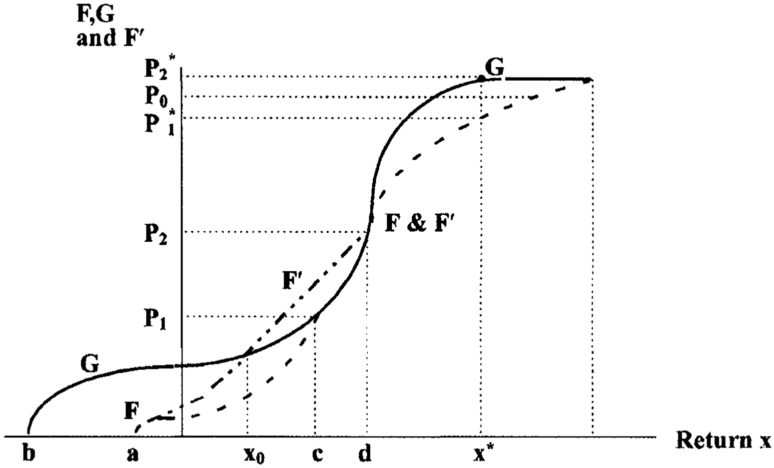
$$Q_F(P) \geq Q_G(P) \text{ for all P } \Leftrightarrow F(x) \leq G(x) \text{ for all x} \tag{4.2}$$

and, if a strict inequality holds on the left-hand side for some P_0, then a strict inequality must hold on the right-hand side for some x_0. Because $F(x) \leq G(x)$ is an optimal decision rule for all $U \in U_1$, if eq. (4.2) holds, $Q_F(P) \geq Q_G(P)$ will also be an optimal decision rule for all $U \in U_1$.

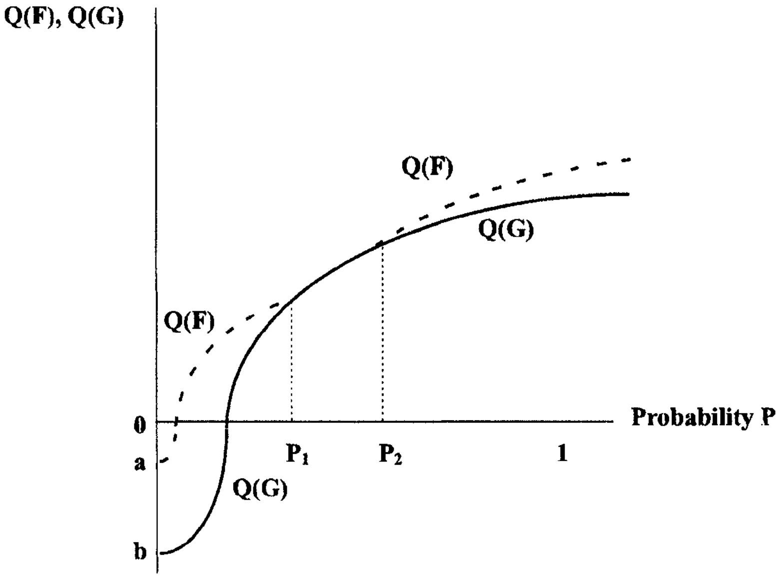
To prove this Theorem, assume first that $F(x) \leq G(x)$. Because FD_1G and

$F(x) \leq G(x)$ for all x, for any arbitrary x^*, we have $F(x^*) = P_1^* \leq G(x^*) = P_2^*$. Then, x^* will be both the P_1^{*th} quantile of distribution F and P_2^{*th} quantile of distribution G (see Figure 4.3a). Because, by assumption, F and G are monotonic increasing functions of x (in the relevant range of x), the quantile functions are also monotonic increasing functions of P in the range $0 \leq P \leq 1$. Because $P_1^* \leq P_2^*$, due to the monotonicity of the quantile function, we have $Q_G(P_1^*) \leq Q_G(P_2^*)$. Because $Q_F(P_1^*) = Q_G(P_2^*) = x^*$, we have $Q_G(P_1^*) \leq Q_F(P_1^*)$. Because this argument holds for any value x^* chosen arbitrarily, by covering all possible values of x^*, we also cover all values of $0 \leq P \leq 1$. Hence, we conclude that the condition $F(x) \leq G(x)$ for every value x, implies that $Q_F(P) \geq Q_G(P)$ for every value P. The same logic can be employed to show that the condition $Q_F(P) \geq Q_G(P)$, implies that $F(x) \leq G(x)$ for all x.

Figure 4.3: The cumulative distribution functions and the corresponding quantile functions



4.3a: The cumulative probability functions



4.3b: The quantile functions

Let us elaborate on the graphical proof. Consider the two distributions F and G given in Figure 4.3. First note that in this example $F(x) \le G(x)$ for all x and there is a strict inequality for say $x = x_0$; hence FD_1G. With cumulative distributions, for each value x we check whether F is below G (a vertical comparison, see for example x^*). With quantile distributions, we check whether for a given value P, $Q_F(P) \ge Q_G(P)$ (a horizontal comparison, see for example P_0). However, because $Q_F(P) \ge Q_G(P)$ if and only if F is below G (in the weak sense), the quantile statement of FSD is identical to the FSD rule formulated in terms of F and G.

In Figure 4.3, in the range $c \le x \le d$, $F(x) = G(x)$ (see Figure 4.3a). This implies that in the range $P_1 \le P \le P_2$, also $Q_G(P) = Q_F(P)$ (see Figure 4.3b). Also, note that in our example, G starts to the left of F ($b < a$). Hence, for the left tail of the distribution, $Q_G(P) < Q_F(P)$, and G cannot dominate F.

We have shown that dominance in the quantile framework is identical to dominance in the cumulative distribution framework. Now we need to show that if $F(x) > G(x)$ for some value x (e.g., $F \not\!\!D_1 G$ in the cumulative distribution framework) such a dominance will not hold with the quantile approach either (and vice versa). To see this, replace F with F′ which is above F in the range $x < d$ and for $x \ge d$ it coincides with F (see dashed line in Figure 4.3a). Because at point c we have $F'(c) > G(c)$, F′ does not dominate G by FSD. Similarly, there is a value P (e.g., $P = P_1$) for which $Q_{F'}(P_1)$. $< Q_G(P_1)$ and there is no dominance of F′ over G in the quantile framework either. Thus, the cumulative distributions and the quantile formulations of FSD yield the same dominance relationship and can be used interchangeably. We can, therefore safely switch from the definition of FSD in terms of F(x) and G(x) to the definition of FSD in terms of $Q_F(P)$ and $Q_G(P)$, and vice versa.

The above proof of Theorem 4.1 is not appropriate for strictly non-increasing cumulative distributions, and in particular, for discrete distributions, because if F and G are step functions, then at each step, there will be many values x corresponding to a given P. In other words, for discrete distributions F and G, Q_F and Q_G are not strictly monotonic increasing functions as required by the above proof. However, the Theorem's claim remains intact for discrete distributions. Let us demonstrate this claim graphically with an example of a discrete distribution. Suppose that under F, one of the values 1, 2 or 3 is obtained with an equal probability of 1/3, and under G, the value 1 with a probability of 2/3 or the value 3 with a probability of 1/3 is obtained. Figure 4.4a plots the cumulative distribution of these two prospects. We see that because $F(x) \le G(x)$, FD_1G. Figure 4.4b depicts the quantile functions $Q_F(P)$ and $Q_G(P)$ using the definition of the quantiles given by eq.(4.1). As can be seen, even for a discrete distribution, $F(x) \le G(x)$ and $Q_F(P) \ge Q_G(P)$ are equivalent because the quantile function is nothing but a mirror image of the cumulative distribution: If F(x) is below

G(x), then $Q_F(P)$ must be above $Q_G(P)$.[2] Thus, Theorem 4.1 remains intact for continuous and discrete random variables alike.

Figure 4.4: Cumulative distributions and the corresponding quantile functions

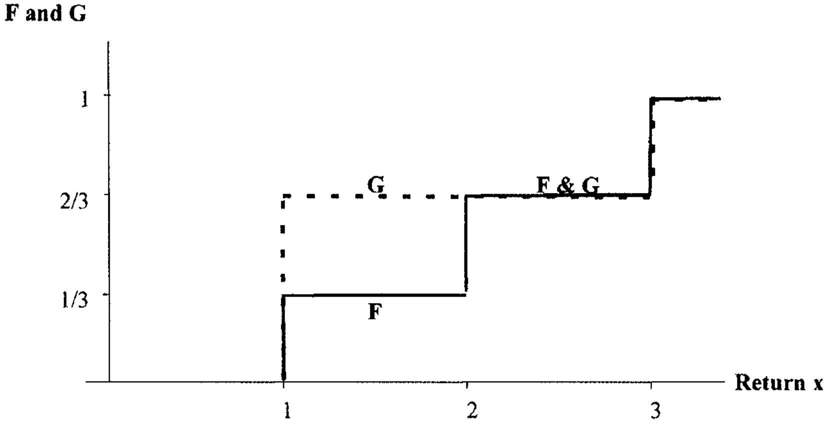
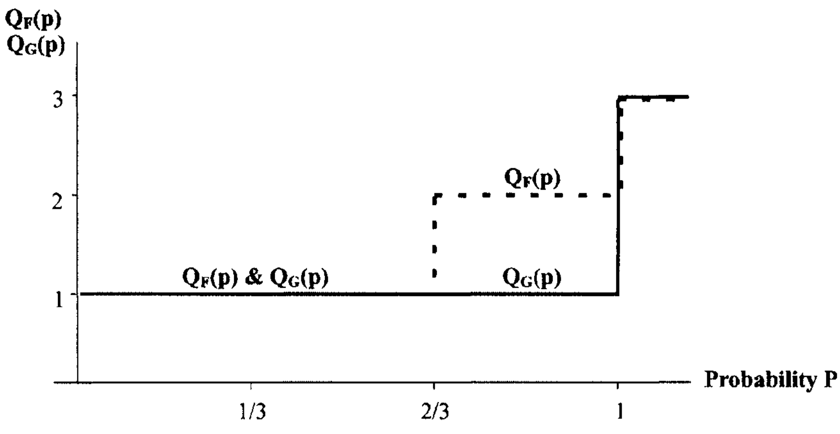


Figure 4.4a: Cumulative distributions



[2] The formal proof for the discrete distribution is different from that given in Theorem 4.1 because Q(P) is not a monotonically increasing function of P. We prove here that the sufficiency side and the necessity side of the proof is very similar. Given that $F(x) \leq G(x)$, we have to show that this implies that $Q_F(P) \geq Q_G(P)$ even if the random variable is discrete. Let P^* be selected arbitrarily where $0 \leq P^* \leq 1$. Denote $Q_G(P^*) = x_2$ and $Q_F(P^*) = x_1$. We need to show that $x_1 \geq x_2$. The proof is by contradiction. Assume $x_1 < x_2$. Because x_2 is the smallest value for eq. (4.1) to hold, x_1 and x_2 cannot be located on the same "step" of G(x). Hence, $G(x_1) < G(x_2)$. Thus, we have $G(x_1) < G(x_2) = P^*$ and $P^* = F(x_1) \leq F(x_2)$. Hence, $G(x_1) < F(x_1)$, which contradicts the assumption that $F(x) \leq G(x)$ for all x.

b) The SSD Rule with Quantiles

In the next Theorem (Theorem 4.2), we formulate SSD in terms of the distribution quantile

Theorem 4.2:

Let F and G be the two distributions under consideration with quantiles $Q_F(P)$ and $Q_G(P)$, respectively. Then FD_2G, if and only if, $\int_0^P [Q_F(t) - Q_G(t)]dt \geq 0$ for all P $(0 \leq P \leq 1)$ and there is a strict inequality for at least one P_0

Proof:

First note that if F dominates G by SSD (FD_2G), then $\int_a^x [G(t) - F(t)]dt \geq 0$ for every value x, and there is a strict inequality for some value x_0 (see Chapter 3, Theorem 3.2). We have seen that if the above integral is non-negative at all intersection points of F and G, then it will be non-negative for all values x (for simplicity of the proof, assume a finite number of intersections of F and G). Thus, it is enough to examine the integral at the intersection points of F and G. However, at all intersection points of F and G we have:

$$\int_0^P [Q_F(t) - Q_G(t)]dt = \int_a^{x_i} [G(t) - F(t)]dt \tag{4.3}$$

where (x_i, P_i) are the horizontal and vertical coordinates of all intersection points, respectively. Therefore, if comparison of F and G reveals that FD_2G (namely, the integral is positive at all intersection points x_i), then the criterion stated in terms of the quantiles will also reveal this dominance (namely, the integral will be positive at all points P_i corresponding to the intersection points x_i of F and G). Similarly, if there is no dominance, say up to some intersection point x_1, we obtain:

$$\int_a^{x_1} [G(t) - F(t)]dt < 0$$

then also up to P_1 corresponding to x_1, we obtain:

$$\int_0^{P_1} [Q_F(t) - Q_G(t)]dt < 0$$

and there will be no dominance of F over G in either framework.

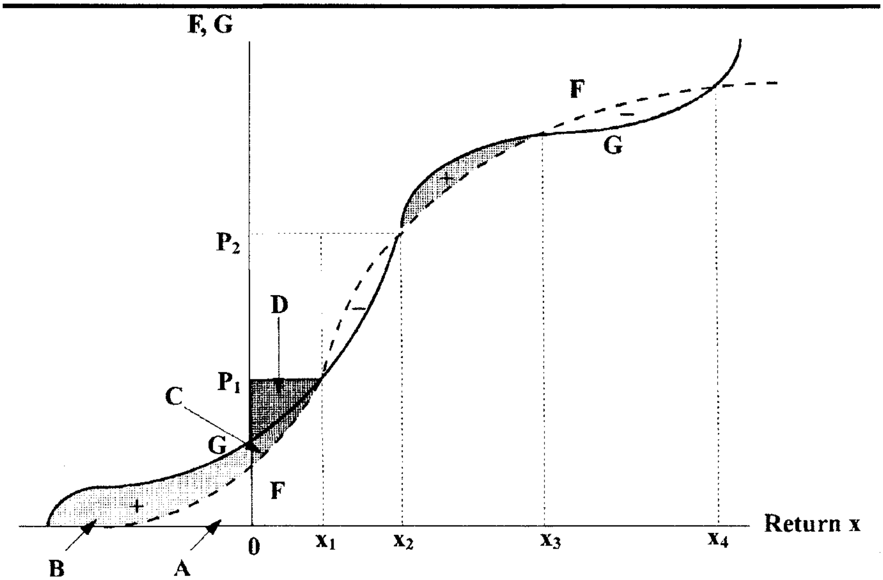
To further illustrate that eq. (4.3) holds for all intersection points of F and G, consider the example given in Figure 4.5.

Let us first consider the area enclosed between F and G in the range $x_1 \le x \le x_2$. The area enclosed between F and is equal to the area below G minus the area below F. Because the curves F and G represent the quantiles of distributions F and G, respectively, the area $\int_{P_1}^{P_2}[Q_F(t) - Q_G(t)]dt$ will correspond to the area enclosed between G and F in this range (x_1, x_2); It is formed by the area left of the curve F up to the vertical axis (which is the quantile of distribution F) minus the area left to curve G (which is the quantile of distribution G) in the range $P_1 \le P \le P_2$. Note that in the cumulative distribution framework, we sum up the area below the curves up to the horizontal axis whereas in the quantile framework, we sum up the area left of F and left of G (up to the vertical axis). By the same token, it can be shown that between any two intersection points x_i and x_{i+1}, the following holds:

$$\int_{x_i}^{x_{i+1}}[G(t) - F(t)]dt = \int_{P_i}^{P_{i+1}}[Q_F(t) - Q_G(t)]dt$$

where $F(x_i) = G(x_i) = P_i$ and $F(x_{i+1}) = G(x_{i+1}) = P_{i+1}$ and (x_i, P_i) and (x_{i+1}, P_{i+1}) correspond to the two intersection points. Now, let us consider the first area left of x_1

Figure 4.5: Second degree stochastic dominance: the cumulative distribution framework and the quantile framework



where x takes negative as well as positive values. For x < 0, the location of the area is left of the vertical line; hence it is negative. Let us show that even for the negative area, eq. (4.3) remains intact. First note that:

$$\int_0^{P_i} Q_F(t)dt = \int_0^{F(0)} Q_F(t)\,dt + \int_{F(0)}^{P_i} Q_F(t)dt$$

and:

$$\int_0^{P_i} Q_G(t)dt = \int_0^{G(0)} Q_G(t)\,dt + \int_{G(0)}^{P_i} Q_F(t)dt$$

Therefore:

$$\int_0^{P_i} [Q_F(t) - Q_G(t)dt = \int_0^{F(0)} Q_F(t)\;\;dt - \int_0^{G(0)} Q_G(t)dt + \int_{F(0)}^{P_i} Q_F(t)dt - \int_{G(0)}^{P_i} Q_G(t)dt$$

From Figure 4.5, we see that:

$$\int_0^{P_i} [Q_F(t) - Q_G(t)]dt = -A - [-(B+A)] + (C+D) - D) = B + C$$

which is equal to:

$$\int_0^{x_i} [G(t) - F(t)dt.$$

Finally, because eq. (4.3) holds for any interval between two intersection points, it also holds for any interval $x \leq x_i$ where x_i is an intersection point; hence, eq. (4.3) holds for all intersection points. If at all intersection points, the right-hand term of eq. (4.3) is non-negative and there is at least one intersection point where it is strictly positive, the same will hold for the left-hand side of eq. (4.3), and both frameworks will reveal a dominance of F over G by SSD. Similarly, if for at least one intersection point, the right-hand side of eq. (4.3) is negative, there will also be a value P_i such that the left-hand side of eq. (4.3) will be negative and neither framework will reveal dominance of F over G. Finally, note that this proof holds for continuous and discrete distributions alike because eq. (4.3) holds also for discrete distributions.

4.3 STOCHASTIC DOMINANCE RULES WITH A RISKLESS ASSET: A PERFECT CAPITAL MARKET

The decision rule most commonly employed in the choice among risky prospects is Markowitz's mean-variance rule.[3] According to this rule, x will be preferred over y if $E(x) \geq Ey$ and $\sigma_x^2 \leq \sigma_y^2$ and there is at least one strict inequality. However, with empirical data as well as with *ex-ante* estimates, $E(x) > E(y)$ and $\sigma_x^2 > \sigma_y^2$ are common. In such cases, the mean-variance rule will be unable to distinguish between x and y. Sharpe (1964)[4] and Lintner (1965)[5] have shown that if a riskless interest rate at which one can borrow or lend exists, a sharper rule can be developed whereby x will be preferred over y by the mean-variance rule if and only if:

$$\frac{Ex - r}{\sigma_x} > \frac{Ey - r}{\sigma_y}$$

where r stands for the riskless interest rate. Thus, both x and y may be located in the mean-variance efficient set with no riskless asset, and the addition of a riskless asset may relegate y to the inefficient set. We employ a similar notion below, and show that stochastic dominance between F and G is revealed by allowing diversification between a risky asset and the riskless asset. To develop stochastic dominance rules with a riskless asset, we rely on the stochastic dominance rules stated in terms of the distribution quantiles presented in the previous section.

a) FSD with a Riskless Asset: The FSDR Rule.

For simplicity of notation and without loss of generality, assume that all random variables under consideration have monotonic increasing cumulative distributions on the relevant range of x. Denote the return on a portfolio composed of a riskless asset and a risky asset by x_α where $x_\alpha = (1 - \alpha) r + \alpha x$ (where $\alpha > 0$) and by F_α the cumulative distribution of X_α.[6] By the quantile definition we have (for a positive α):

$$P = p_F(x \leq Q_F(P)) = p_F((1 - \alpha)r + \alpha x \leq \alpha Q_F(P) + (1 - \alpha)r)$$
$$= p_F(x_\alpha \leq \alpha Q_F(P) + (1 - \alpha)r).$$

---

[3] See H.M. Markowitz, "Portfolio Selection", *Journal of Finance*, 1952.

[4] See W.F. Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk", *Journal of Finance*, 1964.

[5] See J. Lintner, "Security Prices, Risk and Maximal Gains From Diversification", *Journal of Finance*, 1965.

[6] $\alpha > 0$ implies no short sales of the risky asset. Similar rules (though much more complicated) exist for the case where short sales of x is allowed.

However, because by definition, $P_{F_\alpha}(x_\alpha \le Q_{F_\alpha}(P)) = P$, we find that the P^{th} quantile of distribution Fα is given by:

$$Q_{F_\alpha}(P) = \alpha Q_F(P) + (1 - \alpha)\, r.$$

Similarly, we can show that for distribution G, we have:

$$Q_{G_\alpha}(P) = \alpha Q_G(P) + (1 - \alpha) r.$$

From the relationship between $Q_{F_\alpha}(P)$ and $Q_F(P)$, we see that by mixing x with a riskless asset, we rotate the cumulative distribution, F, about the point (r, F(r)). If $Q_F(P^*) = r$, then $P^* = F(r)$. For this particular value we have $Q_{F_\alpha}(P^*) = Q_F(P^*) = r$. The same relationship holds with respect to G. We will elaborate on this rotation of the distribution when we discuss Figure 4.6 below.

Using this quantile's relationship, we will show that even if $F \not\!D G$ and $G \not\!D F$, by FSD, if certain conditions hold, then for every mix of G with the riskless asset, there exists a mix of F with the riskless asset which dominates it by FSD. Let us first define this form of dominance, known as *First Degree Stochastic Dominance with a Riskless Asset (FSDR).*

Definition:

Let $\{F_\alpha\}$ denote the set of all possible mixes of F with the riskless asset and let $\{G_\beta\}$ denote the set of all possible mixes of G with the riskless asset ($\alpha, \beta > 0$), then F will dominate G by FSDR, if for every G_β $\{G_\beta\}$, there is $F_\alpha \in \{F_\alpha\}$ such that F_α dominates G_β by FSD. In such a case, we say that $\{F_\alpha\}\ D_1 \{G_\beta\}$ which implies that an investor with preference $U \in U_1$ will always be better off mixing F with the riskless asset rather than G with the riskless asset, even though it may be that neither F nor G dominates the other by FSD. This dominance is also denoted by $FD_{r1}G$, signifying that F dominates G by FSD with a riskless asset, where the subscript r_1 denotes two things: r for the riskless asset and 1 for first-degree dominance. Thus, F dominates G by FSDR denoted by $FD_{r1}G$ or by $\{F_\alpha\}\ D_1 \{G_\beta\}$. These alternative notations will be used interchangeably. It might seem that an infinite number of comparisons is needed to establish a dominance of $\{F_\alpha\}$ over $\{G_\beta\}$ because the sets $\{F_\alpha\}$ and $\{G_\beta\}$ are infinite. However, we shall see in the next Theorem (Theorem 4.4) that if a positive value α can be found such that $F_\alpha D_1 G$, then we can safely conclude that $\{F_\alpha\} D_1 \{G_\beta\}$.

Theorem 4.4:

If a value α can be found such that $F_\alpha D_1 G$, then $\{F_\alpha\}\ D_1\ \{G_\beta\}$.

Proof:

Suppose that such an α exists. Thus, it is given that $F_\alpha D_1 G$. This implies that

$$Q_{F_\alpha}(P) \equiv \alpha\, Q_F(P) + (1-\alpha)\, r \geq Q_G(P) \text{ for all } 0 \leq P \leq 1 \qquad (4.4)$$

and there is a strict inequality for some value of P. Take a distribution $G_\beta \in \{G_\beta\}$. We need to prove that for every $\beta > 0$, there is $\gamma > 0$ such that:

$$Q_{F_\gamma}(P) \equiv \gamma\, Q_F(P) + (1-\gamma)\, r \geq \beta\, Q_G(P) + (1-\beta)\, r \equiv Q_{G_\beta}(P) \qquad (4.5)$$

and there is a strict inequality for some value P_0, namely $\{F_\alpha\}\ D_1\ \{G_\beta\}$.

To see that (4.4) implies (4.5), simply multiply (4.4) by $\beta > 0$ and add $(1-\beta)r$ to both sides to obtain:

$$\alpha\beta\, Q_F(P) + \beta(1-\alpha)\, r + (1-\beta)\, r \geq \beta\, Q_G(P) + (1-\beta)\, r$$

for all P and a strict inequality holds for some P_0 or:

$$\alpha\beta\, Q_F(P) + (1-\alpha\beta)\, r \geq \beta\, Q_G(P) + (1-\beta)\, r$$

for all P with a strict inequality for some P_0. Choose $\gamma = \alpha\beta > 0$ to complete the proof.

Thus, if we find a value $\alpha > 0$ such that $F_\alpha D_1 G$, we can safely conclude that $\{F_\alpha\}$ D_1 $\{G_\beta\}$ or that F dominates G by FSDR ($FD_{r1}G$). One way to find whether such an α exists is to try all possible combinations x_α. Of course, this could well be an endless task. In the next theorem (Theorem 4.5) we establish the conditions for the existence of such a value α. We show that, in practice, only one comparison of F and G is needed to verify whether such an α exists.

Theorem 4.5 (FSDR):

Let F and G be the cumulative distributions of two distinct risky assets and r be the riskless interest rate. Then F dominates G by FSDR if and only if:

$$\beta_0 \equiv \underset{0 \leq P < F(r)}{\text{INF}} \frac{Q_G(P) - r}{Q_F(P) - r} \geq \underset{F(r) < P \leq 1}{\text{SUP}} \frac{Q_G(P) - r}{Q_F(P) - r} \equiv \beta_1 \qquad (4.6)$$

where $F(r) = p_F(X \leq r)$.

Proof:

Using the results of Theorem 4.4, we need to find only one positive value α such that $Q_{F_\alpha}(P) \geq Q_G(P)$ for all P. Thus, we have to show that (4.6) holds if and only if there is $\alpha > 0$ such that:

$$Q_{F_\alpha}(P) = \alpha Q_F(P) + (1-\alpha)\, r \geq Q_G(P) \qquad \text{for all } 0 \leq P \leq 1$$

or:

$$\alpha(Q_F(P) - r) \geq Q_G(P) - r \text{ for all } 0 \leq P \leq 1 \tag{4.7}$$

Let us divide the whole range $0 \leq P \leq 1$ into two ranges: the range $0 \leq P < F(r)$ and the range $F(r) < P \leq 1$. By definition of the quantile, in the range $0 \leq P < F(r)$ also $Q_F(P) < r$, or $Q_F(P) - r < 0$. Divide (4.7) by the negative number $Q_F(P) - r$ to obtain:

$$\alpha \leq \frac{Q_G(P) - r}{Q_F(P) - r},$$

and, because this inequality should hold for all $P < F(r)$, we need to find a value α such that:

$$\alpha \leq \underset{0 \leq P < F(r)}{\mathrm{INF}} \frac{Q_G(P) - r}{Q_F(P) - r} \equiv \beta_0.$$

because if $\alpha < \beta_0$, then α is not larger than any other value $\frac{Q_G(P) - r}{Q_F(P) - r}$ in the range $0 \leq P < F(r)$.[7]

Similarly, in the range $F(r) < P \leq 1$, $Q_F(P) > r$, we need to find α such that:

$$\alpha \geq \underset{F(r) < P \leq 1}{\mathrm{INF}} \frac{Q_G(P) - r}{Q_F(P) - r} \equiv \beta_1.$$

If $\alpha \geq \beta_1$ and also $\alpha \leq \beta_0$, then $Q_{F_\alpha}(P) \geq Q_G(P)$ in the whole range $0 \leq P \leq 1$. However, if condition (4.6) of Theorem 4.5 holds, a value α can be found such that $\beta_1 \leq \alpha \leq \beta_0$,

[7] Note that at the point $P=F(r)$, $Q_F(P) = r$ and $Q_G(P) < r$ (see below); hence we have in the range $0 \leq P \leq F(r)$ $\lim_{P \uparrow F(r)} \beta_0 = \infty$. Therefore, we know for sure that the β_0 is not located at this point. Hence, the point $F(r) = P$ can be added to eq. (4.6) without lose of any generality. In such a case, we have the range $0 \leq P \leq F(r)$, rather than $0 \leq P < F(r)$ and the proof remains the same.

and hence condition (4.6) will guarantee that F dominates G by FSDR.[8] Finally, note that because there is a value P such that $Q_F(P) > r$, β_1 in this range is positive; hence $\alpha > 0$. (see also footnote 8). The necessity of the condition of Theorem 4.5 is straightforward: if condition (4.6) does not hold, an α will not be found such that $Q_{F_\alpha}(P) \geq Q_G(P)$ for all P; hence there is no $\alpha > 0$ such that F_α will dominate G.

Because $G \in \{G_\beta\}$, F will not dominate G by FSDR or $\{F\alpha\} \not{D} \{G_\beta\}$.

Finally, if FD_1G, then for $\alpha = 1$, F will dominate G by FSD or $FD_{r1}G$. Thus, FSD implies FSDR. Therefore, the FSDR efficient set will be smaller than or equal to the FSD efficient set.

b) Graphical Illustration of the FSDR Rule

Figure 4.6 illustrates a hypothetical distribution F and two distributions F_{α_1} and F_{α_2} corresponding to x_{α_1} and x_{α_2}. First, note that $\min_F(x) < r < \max_F(x)$, otherwise we have arbitrage positions (see footnote 8). Secondly, by definition, $F(r) = P\,(x \leq r)$. Because F is a monotonic non-decreasing function of x, then for $P < F(r)$, $Q_F(P) < r$ and for $P > F(r)$, $Q_F(P) > r$. Now, let us examine the distribution F_α.

By definition of $Q_{F\alpha}(P)$ we have:

$$Q_{F_\alpha}(P) = \alpha Q_F(P) + (1-\alpha)r = Q_F(P) - (1-\alpha)Q_F(P) + (1-\alpha)r$$

$$= Q_F(P) + (1-\alpha)(r - Q_F(P)).$$

From this relationship, we can draw the following conclusions:

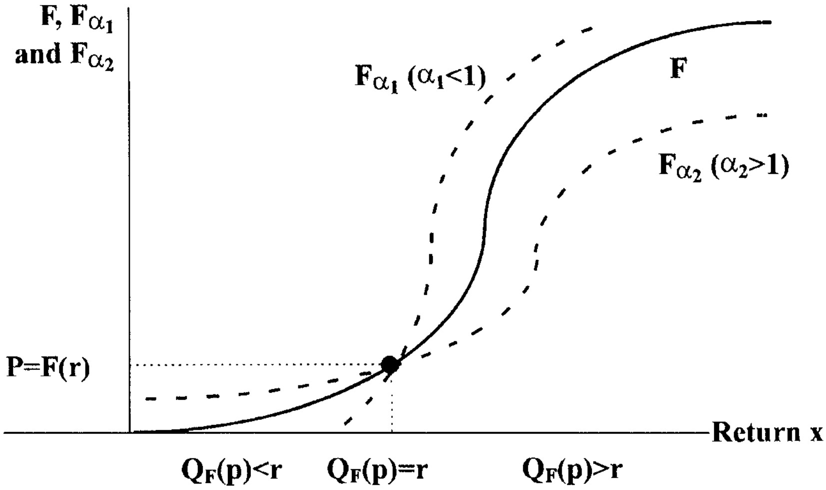
1 For the point $P = F(r)$, we have $Q_F(P) = r$ (see Figure 4.6). Thus, no matter what α we employ, at this point $Q_{F_\alpha}(P) = Q_F(P)$. This means that F_α and F intercept at the point (F(r), r).

2 If $\alpha < 1$, then for $Q_F(P) < r$, we have $Q_{F_\alpha}(P) > Q_F(P)$ and for, $Q_F(P) > r$, we have $Q_{F_\alpha}(P) < Q_F(P)$ (see F_{α_1} in Figure 4.6).

---

[8] To avoid possible arbitrage, we assume that there is a value P for which $Q_F(P) < r$ and there is a value P for which $Q_F(P) > r$; in other words, distribution F crosses the degenerated distribution of r. A similar relationship holds with distribution G. Hence, SUP $(\cdot)$ in the range $F(r) < P \leq 1$ must be positive, and therefore, if inequality (4.6) holds, there will be $\alpha > 0$ such that $F_\alpha D_1 G$.

Figure 4.6: The effect of lending (F_{α_1}) and borrowing (F_{α_2}) on the cumulative distribution of returns

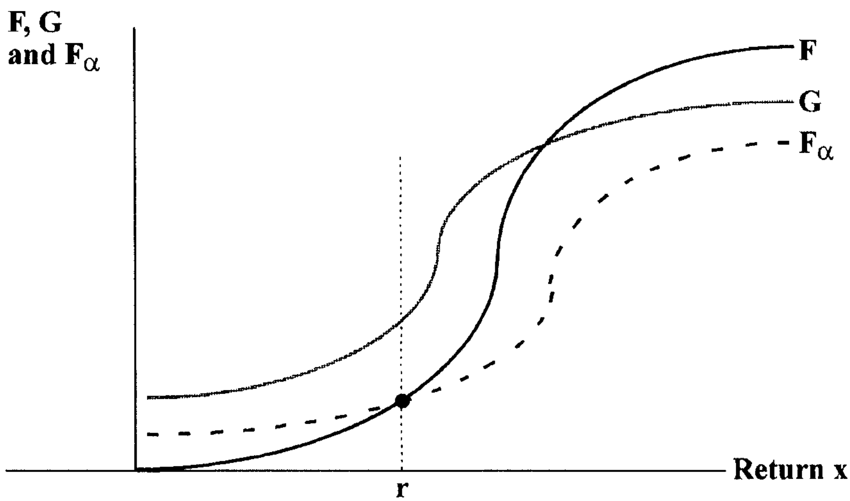


3 If $\alpha > 1$, then for $Q_F(P) < r$, $Q_{F_\alpha}(P) < Q_F(P)$ and for $Q_F(P) > r$, $Q_{F_\alpha}(P) > Q_F(P)$ (recall $1 - \alpha < 0$). The case $\alpha > 1$ is given by F_{α_2} in Figure 4.6.

Thus, all distributions F_α intercept at the same point (F(r), r) and, therefore, by mixing a random variable with the riskless asset, the cumulative distribution will always rotate about the point (F(r), r). If $\alpha > 1$, we have a leverage effect; the distribution F_α will be flatter than distribution F. If $\alpha < 1$, the risk of the investment will be reduced due to investment in riskless bonds and F_α will become more condensed relative to F around the point $x = r$. In the extreme case where $\alpha = 0$, F_α will be a vertical line rising from point r; all of the investment will be in the riskless asset. Now let us compare two distributions F and G as illustrated in Figure 4.7. We see that neither F nor G dominates the other by FSD. However, it is possible to rotate F and to obtain F_α which dominates G by FSD; hence F dominates G by FSDR.

Finally, note that as all distributions F_α intercept at point (F(r), r), in order to have some α for which $F_\alpha D_1 G$, we must have $F(r) \leq G(r)$ as plotted in Figure 4.7. Thus, $F(r) < G(r)$ is a *necessary* condition for FSDR of F over G. If there is a α such that $F_\alpha D_1 G$, then for every G_β there will be F_γ (where $\gamma = \alpha\beta$) such that $F_\gamma D_1 G_\beta$. The

Figure 4.7: F and G intercept but F_α dominates G by FSD



intuition of this claim is as follows: Suppose that we want to rotate G such that G_β will intercept F_α. We simply rotate F_α by mimicking the G rotation to obtain a new distribution F_γ such that $F_\gamma D_1 G_\beta$ (not shown on Figure 4.7).

Example:

Consider the distribution of returns on the following two investments:

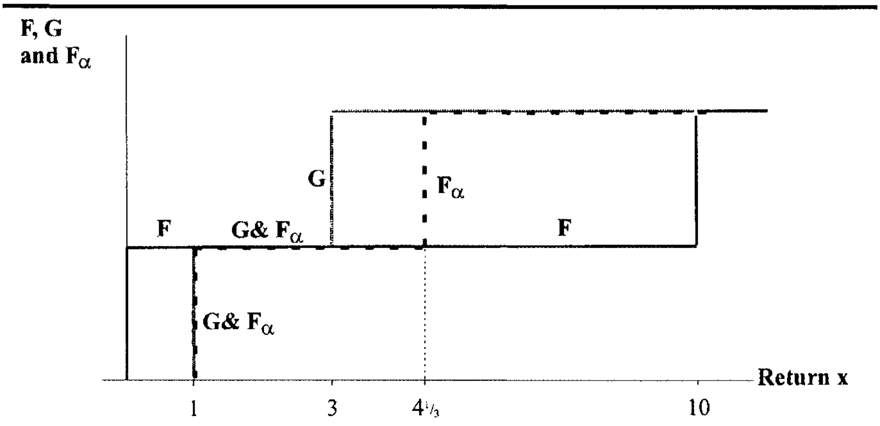
Investment F		**Investment G**	
x	**P(x)**	**x**	**p(x)**
0	½	1	½
10	½	3	½

Because F and G intersect, neither F nor G dominates the other by FSD. Assume that r = 1.5. Choose α = 1/3 to obtain the following returns for F_α:

$$1/3 \cdot 0 + (1 - 1/3) \cdot 1.5 = 1 \text{ with a probability of } ½$$
$$1/3 \cdot 10 + (1 - 1/3) \cdot 1.5 = 4⅓ \text{ with a probability of } ½$$

Thus, we find α = 1/3 such that $F_\alpha D_1 G$; hence F dominates G by FSDR. Figure 4.8 demonstrates F, G and F_α corresponding to the above example. Note that F and G intercept but F_α is located below (in the weak sense) G in the whole range.

Figure 4.8: The distributions corresponding to the numerical example: F and G intercept but Fα dominates G by FSD



c) SSD with a Riskless Asset: The SSDR Rule

We turn now to second-degree stochastic dominance with a riskless asset (SSDR). As with FSDR, it is enough to find one positive value α such that $F_\alpha D_2 G$ in order to conclude that F dominates G by SSDR, or more specifically, that

$\{F_\alpha\}$ D_2 $\{G_\beta\}$ or $FD_{r2}G$ where the subscript r2 composed from r and 2, where 2 denotes second-degree stochastic dominance and r denotes the existence of a riskless asset.

To see this, recall that F_α dominates G by SSD if:

$$\int_0^P (\alpha Q_F(t) + (1-\alpha)r)dt \geq \int_0^P Q_G(t)dt \text{ for all } 0 \leq P \leq 1$$

with a strict inequality for some P_0. This can be rewritten as:

$$\int_0^P (\alpha Q_F(t) - r)\, dt \geq \int_0^P (Q_G(t) - r)dt \text{ for all } 0 \leq P \leq 1.$$

Multiply both sides by β and add r to obtain:

$$\int_0^P (\alpha\beta Q_F(t) + (1-\alpha\beta)r]dt \geq \int_0^P [\beta Q_G(t) + (1-\beta)r]dt.$$

Thus, if F_α D_2G, then for any mix G_β there will be a mix F_γ (where $\gamma = \alpha\beta$), such that $F_\alpha D_2 G$. Therefore, we conclude that if there is a positive α such that $F_\alpha D_2 G$, then F will dominate G by SSDR (or $\{F_\alpha\}$ D_2 $\{G_\beta\}$). We turn now to the condition which quarantees that such $\alpha > 0$ exists.

Theorem 4.6 (SSDR):

Let F and G be the cumulative distributions of two distinct investments with quantiles $Q_F(P)$ and $Q_G(P)$, respectively. Then F will dominate G by SSDR (or $\{F_\alpha\}$ will dominate $\{G_\beta\}$ for every risk-averter) if and only if:

$$\delta_0 \equiv \underset{0 \le P < P_0}{\text{INF}} \frac{\int_0^P [Q_G(t) - r]dt}{\int_0^P [Q_G(t) - r]dt} \ge \underset{P_0 < P \le 1}{\text{SUP}} \frac{\int_0^P [Q_G(t) - r]dt}{\int_0^P [Q_F(t) - r]dt} \equiv \delta_1 \quad (4.8)$$

where r is the riskless interest rate and P_0 is the value which solves the following equation:

$$rP_0 = \int_0^{P_0} Q_F(t)\, dt. \quad (4.9)$$

Proof:

First note that because F and G are risky investments, we can assume, without loss of generality, that $E_F(x) > r$ and $E_G(x) > r$; otherwise all investors would invest in the riskless asset rather than in the risky asset (recall that F and G are the only two investments under consideration and no diversification is allowed).

Because: $\int_0^1 Q_F(t)dt = \int_a^b xf(x)dx = E_F(x)$ [9], there is $P_0 < 1$, such that (4.9) holds.

As shown above, F dominates G by SSDR if and only if there is a positive α such that:

[9] Note that $\int_a^b xf(x)dx = xF(x)\big|_a^b - \int_a^b F(x)dx = b - \int_a^b F(x)dt$ but this is exactly the area left of F, namely $\int_0^1 Q_F(t)dt$. Because $\int_0^P Q_F(t)dt$ is a continuous increasing function of P, the solution to eq. (4.9) is unique.

$$\int_0^P Q_{F_\alpha}(t)dt \equiv \int_0^P [(1-\alpha)r+\alpha Q_F(t)]dt \geq \int_0^P Q_G(t)dt \tag{4.10}$$

for every value P ($0 \leq P \leq 1$) with at least one strict inequality.

Rearrange (4.10) to obtain:

$$\alpha \int_0^P [Q_F(t)-r]dt \geq \int_0^P [Q_G(t)-r]dt \tag{4.10'}$$

for every value P (with at least one strict inequality)

From (4.9) we conclude that $\int_0^P [Q_F(t)-r]dt \gtreqless 0$ if and only if $P \gtreqless P_0$, and hence a positive α has to be found such that[10]

$$\alpha \leq \frac{\int_0^P [Q_G(t)-r]dt}{\int_0^P [Q_F(t)-r]dt} \quad \text{for all } 0 \leq P < P_0. \tag{4.11}$$

Because (4.11) has to hold for all values $0 \leq P < P_0$, choose α such that:[11]

$$\alpha \leq \underset{0 \leq P < P_0}{\text{INF}} \frac{\int_0^P [Q_G(t)-r]dt}{\int_0^P [Q_F(t)-r]dt} \equiv \delta_1 \tag{4.12}$$

[10] Note that for $P < P_0$, the denominator of (4.11) is negative and hence the inequality sign of (4.11) is reversed. Also, in order to find a positive α such that (4.11) holds, the numerator must be also negative because the denominator is negative. This implies that the value P_1 which solves the equation $rP_1 = rP_1 = \int_0^{P_1} Q_G(t)dt$ must be larger than the value P_0 which solves equation (4.9). This property guarantees that the denominator of (4.11) in the range $0 \leq P < P_0$ will also be negative. Indeed, $P_1 > P_0$ is a necessary condition for FSDR.

[11] Note that for $P \uparrow P_0$ by (4.9) and (4.10'), the numerator of δ_0 is negative and the denominator approaches zero from below. Hence, in the range $0 \leq P \leq P_0$, $\lim_{P \uparrow P_0} = \infty$. Therefore, the INF is not achieved at $P = P_0$ and the range $0 \leq P \leq P_0$ can be written in (4.12) rather $0 \leq P < P_0$, with no change in the proof.

Then for this α, (4.12) will hold over the entire range $0 \le P < P_0$. Similarly (4.10) will hold for all $1 \ge P > P_0$ if a positive value α can be found such that:

$$\alpha \ge \underset{P_0 < P \le 1}{SUP} \frac{\int_0^P [Q_G(t) - r]dt}{\int_0^P [Q_F(t) - r]dt} \equiv \delta_1 \tag{4.13}$$

Equation (4.10) and (4.10′) hold in the entire range of $0 \le P \le 1$; hence F will dominate G in the SSDR sense, if and only if (4.12) and (4.13) hold simultaneously. Thus, we need to find α such that:

$$\delta_1 \le \alpha_0 \le \delta_0 \tag{4.14}$$

However, if the condition of Theorem 4.6 holds, such an α will exist.

Finally, because $E_G(X) > r$ and $E_F(X) > r$, $\delta_1 \ge \frac{E_F(x) - r}{E_G(x) - r} > 0$. Thus, if eq. (4.8) holds,[12] a *positive* α can always be found such that (4.10) will hold over the whole range of P. Note that if (4.8) does not hold, then for any positive value α that is chosen, there will be some value P such that (4.10) will not hold for this particular value. Thus, the conditions of Theorem 4.6 are necessary and sufficient for dominance of F over G by SSDR. The interesting feature of the SSDR rule (like the FSDR rule) is that there is no need for an infinite number of comparisons in order to determine whether $\{F_\alpha\}$ dominates $\{G_\beta\}$. Only one comparison of F and G is needed. If (4.8) holds with respect to F and G, we conclude that $\{F_\alpha\}$ dominates $\{G_\beta\}$ by SSD without having to compare all the elements in $\{F_\alpha\}$ with all the elements in $\{G_\beta\}$.

Example:

To facilitate the understanding of the suggested rule, let us illustrate how it can be applied by considering the following simple numerical example:

Investment F		**Investment G**	
X	**p(x)**	**x**	**p(x)**
0	1/2	1	½
14	1/2	10	½

[12] The requirement of at least one strict inequality in (4.10) (or 4.10′) implies that we exclude the possibility of $\int_0^P [Q_G(t) - r)dt / \int_0^P [Q_F(t) - r]dt$ being constant for all values P. In other words, we preclude the trivial case of identical F and G.

where F and G denote two risky investments and x is the additional return in dollars per \$100 investment. The cumulative distributions of these two investments is given in Figure 4.9. It is easy to verify that neither F nor G dominates the other by the SSD:

$$\int_0^1 [G(t)-F(t)]dt=-1/2<0; \quad \text{hence, F does not dominate G by SSD.}$$

Similarly:

$$\int_0^{14} [F(t)-G(t)]dt=-1\tfrac{1}{2}<0; \quad \text{hence G does not dominate F by SSD.}$$

Assume now that investors can borrow and lend money at a riskless interest rate of 4% (or \$4 per \$100). Does F dominate G by SSDR? The first step in the application of the SSDR criterion is to find the value P_0 for which $\int_0^{P_0} Q_F(t)dt=rP_0$.

Figure 4.9 shows that P_0 is given by $14\,(P_0-1/2)=4P_0$; hence, P_0=0.7.

In the range $0 \le P \le 0.7$ we obtain:

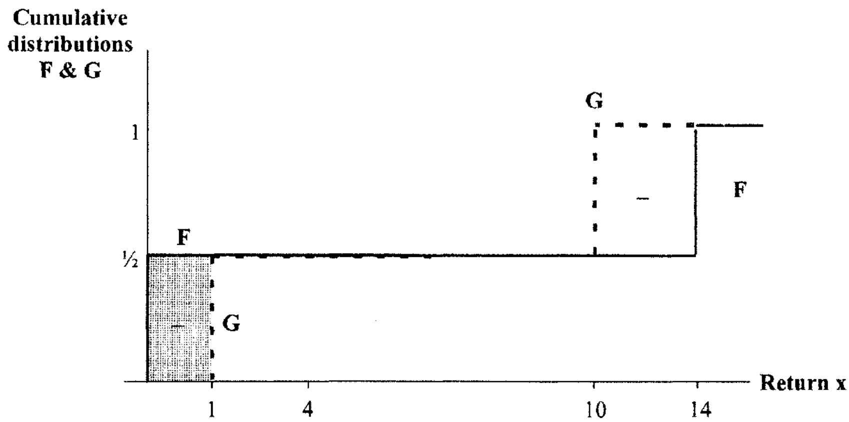
$$\underset{0 \le P \le 0.7}{\text{INF}} \frac{\int_0^P [Q_G(t)-r]dt}{\int_0^P [Q_F(t)-r]dt} = 0.75$$

and this value is obtained for all $0 < P \le 0.7$ (Note that because $P \to 0.7$ from below, this ratio approaches ∞ as $P \to 0.7$). In the range $0.7 < P \le 1$ we have,

$$\underset{0.7 \le P \le 1}{\text{SUP}} \frac{\int_0^P [Q_G(t)-r]dt}{\int_0^P [Q_F(t)-r]dt} = 0.50$$

and this value is obtained for P = 1. Thus eq. (4.8) holds and F dominates G by SSDR. To confirm the above result, construct an investment distribution F_α with 50% invested in F and 50% invested in riskless bonds namely, $\alpha = ½$. In this case, we obtain \$2 (given by $1/2 \cdot 0 + 1/2 \cdot 4$) with a probability of 1/2 and \$9 (given by $1/2 \cdot 14 + 1/2 \cdot 4$) with a probability of 1/2. It is easy to verify that

Figure 4.9: F does not dominate G by SSD but F dominates G by SSDR.



$\int_0^x [G(t) - F_{0.5}(t)]dt \geq 0$ for all values x with at least one strict inequality (e.g., for x=2) and, therefore, $F_{0.5}$ dominate G by the SSD. Because a value $\alpha > 0$ was found such that F_α dominates G by SSD, F_α dominate G by SSDR (or $\{F_\alpha\}$ D_2 $\{G_\beta\}$). Note that in order to show that there is an α_0 such that F_α dominates G, any value α can be picked from the range $0.50 \leq \alpha \leq 0.75$. This example illustrates how the SSDR criterion is applied and shows that F may dominate G by SSDR even though F does not dominate G by the SSD rule.

d) The SD and SDR Efficient Sets

The greater the number of assumptions regarding preference or the availability of the riskless asset, the smaller the efficient set. To be more specific, if there is FSD of F over G, then for $\alpha = 1$, F_α will dominate G which implies that F dominates G by FSDR ($FD_{r1}G$). Similarly, if F dominates G by SSD, then for $\alpha = 1$ $F_\alpha D_2 G$ and F will dominate G by SSDR ($FD_{r2}G$). Finally, if FD_3G, $FD_{r3}G$ will hold.[13] These stochastic dominance relationships can be summarized as follows:

FSD	$\Rightarrow$	SSD	$\Rightarrow$	TSD
$\Downarrow$		$\Downarrow$		$\Downarrow$
FSDR	$\Rightarrow$	SSDR	$\Rightarrow$	TSDR

Because all of the rules are transitive, FSD implies SSDR as well as TSDR, and SSD implies TSDR.

[13] However, TSD cannot be stated in terms of the quantiles and one should employ the cumulative distributions.

4.4 STOCHASTIC DOMINANCE RULES WITH A RISKLESS ASSET: AN IMPERFECT CAPITAL MARKET

In the previous section we assumed a perfect capital market in which investors can borrow and lend at the same interest rate, r. In practice, the capital market is imperfect and the borrowing rate denoted by r_b is higher than the lending rate denoted by r_l. In this section we establish necessary and sufficient rules for dominance where $r_b > r_l$.

To develop SDR rules corresponding to an imperfect market we need the following Theorem (Theorem 4.9).

Theorem 4.9:

If x dominates y by FSD, SSD or TSD then x^* will also dominate y^* by FSD, SSD and TSD, respectively, where:

$$x^* \equiv ax + b, \; y^* \equiv ay + b, \qquad \text{and } a > 0, b \lesseqgtr 0 .$$

Proof:

Denote the distribution of x by F, and of y by G, and the distributions of x^* and y^* by F^* and G^*, respectively. By definition:

$$P = p_F\,(x \le Q_F\,(p)) = p_F\,(ax + b \le aQ_F(p) + b) = p_{F^*}\,(x^* \le Q_{F^*}(P))$$

(and a similar relationship holds with y). Therefore, the quantile of the various distributions are related as follows:

$$Q_{F^*}(p) = aQ_F(P) + b$$

$$Q_{G^*}(p) = aQ_G(P) + b$$

Using the quantile formulation of FSD, SSD and TSD, the proof of the Theorem is immediate (because b cancels out and a is a positive constant). In an imperfect market, it is not trivial that if there is $\alpha > 0$ such that F_α dominates G, then $\{F_\alpha\}$ will dominate $\{G_\beta\}$ because $\gamma = \alpha\,\beta > 1$ (borrowing) and $\beta < 1$ (lending) are possible. Hence, we cannot employ the same interest rate as in the proof for the case of a perfect capital market. However, the next Theorem (Theorem 4.10) shows that in spite of this difficulty, if an α is found such that $F_\alpha D_i G$, then $\{F_\alpha\}\ D_i\ \{G_\beta\}$ (i = 1, 2, 3).

Theorem 4.10:

Let x and y be two random variables with cumulative distributions F and G, respectively, and $r_b > r_l$. Then, $\{F_\alpha\}$ will dominate $\{G_\beta\}$ by FSD, SSD, or TSD, respectively, if and only if there is at least one value α (either $0 < \alpha \leq 1$, or $\alpha > 1$) such that $F_\alpha DG$ (i.e., F_α dominates G) by FSD, SSD, or TSD, respectively.

Proof:

The necessity is straightforward: If there is no α such that F_α dominates G, then $\{F_\alpha\}$ will not dominate $\{G_\beta\}$ because $G \in \{G_\beta\}$. In order to prove the sufficient condition, we have to show that if there is a positive α such that $F_\alpha D_i G$, i = 1, 2, 3 (or x_α dominates y by FSD, SSD, or TSD), then for every positive β, there will be some positive γ such that $F_\gamma D G_\beta$ by the appropriate criterion. We show that if $\gamma = \alpha\,\beta$ is chosen, then $F_\gamma D_i G_\beta$ (i = 1, 2, 3). Thus, given that $F_\alpha D_i G$ (i = 1, 2, 3), we have to prove that $F_\gamma D_i\, G_\beta$ (i = 1, 2, 3) for all $\beta > 0$. To prove this, let us first look at an artificial random variable $(x_\alpha)_\beta = \beta\, x_\alpha + (1 - \beta)r$, where $\beta > 0$. According to Theorem 4.9, a positive linear transformation ($\beta > 0$) does not change the stochastic dominance relationship between variables. Thus, if $x_\alpha\, D_i y$ (i = 1, 2, 3) also $(x_\alpha)_\beta D_i\, y_\beta$ (i = 1, 2, 3). Because FSD, SSD, and TSD are transitive, all that we need to show is that $x_{\alpha\beta}\, D_i\, (x_\alpha)_\beta$ (i = 1, 2, 3). The transitivity of these rules implies that $x_{\alpha\beta} D_i\, y_\beta$ (i = 1, 2, 3). The expressions of $(x_\alpha)_\beta$ and $x_{\alpha\beta}$ are given by:

$$(x_\alpha)_\beta = \begin{cases} \beta[\alpha x + (1-\alpha)r_l] + (1-\beta)r_l & \text{if } \alpha < 1 \text{ and } \beta < 1 \\ \beta[\alpha x + (1-\alpha)r_b] + (1-\beta)r_l & \text{if } \alpha > 1 \text{ and } \beta < 1 \\ \beta[\alpha x + (1-\alpha)r_b] + (1-\beta)r_b & \text{if } \alpha > 1 \text{ and } \beta > 1 \\ \beta[\alpha x + (1-\alpha)r_l] + (1-\beta)r_b & \text{if } \alpha < 1 \text{ and } \beta > 1 \end{cases} \tag{4.15}$$

$$x_{\alpha\beta} = \begin{cases} \alpha\beta x + (1-\alpha\beta)r_l & \text{if } \alpha\beta < 1 \\ \alpha\beta x + (1-\alpha\beta)r_b & \text{if } \alpha\beta > 1 \end{cases} \tag{4.16}$$

By rearranging the terms in (4.15) for the two cases where both α and β are smaller (or larger) than 1, we obtain $(x_\alpha)_\beta = x_{\alpha\beta}$ for these two cases, and because $(x_\alpha)_\beta$ dominates x_β, $x_{\alpha\beta}$ also dominates y_β. It is easy to verify that in the other two cases (note that each case splits into two cases i.e., $\alpha\beta < 1$ and $\alpha\beta > 1$), $x_{\alpha\beta}$ is equal to the variable $(x_\alpha)_\beta$ plus a positive term. Thus, $x_{\alpha\beta}$ dominates $(x_\alpha)_\beta$ by FSD which also implies dominance by SSD and TSD. Thus, the fact that $(x_\alpha)_\beta$ dominates y_β by FSD, SSD, and TSD, implies that $x_{\alpha\beta}$ dominates y_β, by these rules (i.e., F_γ (or x_γ) which dominates G_β (or y_β) was found where $\gamma = \alpha\beta$). Note that the variable $(x_\alpha)_\beta$ is an "artificial" variable in the sense that such an investment does not exist. We use this variable only to prove the sufficiency. To sum up, Theorem 4.10 guarantees that even though $r_b > r_l$, if some α is found such that $F_\alpha\, D_i G$ (i = 1, 2, 3), there is no need for additional comparisons because it implies that: $\{F_\alpha\}\, D_i\, \{G_\beta\}$ (i = 1, 2, 3).

In the next three theorems we establish FSDR, SSDR and TSDR where $r_b > r_l$, namely, conditions for the existence of $\alpha > 0$ such that $F_\alpha D_i G$, i = 1,2,3. The proofs are very similar to the perfect market case; hence they are omitted.[14] Note, however, that in the Theorems given below for $F_\alpha DG$ we have two possible conditions: If $\alpha < 1$, we use r_l (the lending rate) and if $\alpha > 1$, we use r_b (the borrowing rate),

Theorem 4.11 (FSDR):

Let F and G be the cumulative distributions of two risky assets. Assume that the borrowing interest rate r_b is higher than the lending rate r_l. Then, a necessary and sufficient condition for the dominance of $\{F_\alpha\}$ and $\{G_\beta\}$ for all $U \in \mathbf{U}_1$ will be either:

$$\underset{0 \le P < F(r_l)}{\text{INF}} \frac{Q_G(p) - r_l}{Q_F(p) - r_l} \ge \underset{F(r_l) < P \le 1}{\text{SUP}} \frac{Q_G(p) - r_l}{Q_F(p) - r_l} \tag{4.17}$$

or:

$$1 < \underset{0 \le P < F(r_b)}{\text{INF}} \frac{Q_G(p) - r_b}{Q_F(p) - r_b} \ge \underset{F(r_b) < P \le 1}{\text{SUP}} \frac{Q_G(p) - r_b}{Q_F(p) - r_b} \tag{4.18}$$

In order to guarantee that such an α will be smaller than 1 (because we are lending rather than borrowing money), we require the SUP in (4.17) to be smaller than 1. Similarly, when borrowing takes place, we need the INF in (4.18) to be greater than 1.

Theorem 4.12 (SSDR):

Let F and G be the cumulative distributions of two risky investments, and let r_b and r_l be the riskless borrowing and lending interest rates, respectively. Then, a necessary and sufficient condition for the dominance of $\{F_\alpha\}$ over $\{G_\beta\}$ for all $U \in \mathbf{U}_2$ (i.e., for all risk-averters) will be that at *least* one of the following conditions holds:

$$\underset{0 \le P < P_l}{\text{INF}} \frac{\int_0^P (Q_G(t) - r_l)dt}{\int_0^P (Q_F t) - r_l)dt} \ge \underset{P_l < P \le 1}{\text{SUP}} \frac{\int_0^P (Q_G(t) - r_l)dt}{\int_0^P (Q_F(t) - r_l)dt} \le 1 \tag{4.19}$$

or:

---

[14] For detailed proofs, see Levy, H., and Kroll, Y., "Stochastic Dominance with a Riskless Asset: An Imperfect Market," *Journal of Financial and Quantitative Analysis*, 1980.

$$1 < \underset{0 \le P < P_b}{\text{INF}} \frac{\int_0^P (Q_G(t) - r_b)dt}{\int_0^P (Q_F t) - r_b)dt} \ge \underset{P_b < P \le 1}{\text{SUP}} \frac{\int_0^P (Q_G(t) - r_b)dt}{\int_0^P (Q_F(t) - r_b)dt} \qquad (4.20)$$

where P_l is the value which solves the equation $\int_0^{P_l} (Q_F(t) - r_l)dt = 0$, and P_b is the value which solves the equation $\int_0^{P_b} (Q_F(t) - r_b)dt = 0$.

4.5 SUMMARY

In this chapter, the FSD and SSD stochastic dominance rules were formulated in terms of the distribution quantiles. The advantage of the stochastic dominance rules based on the quantile formulation is that they can be extended relatively easily to the case where lending and borrowing at the riskless interest rate is allowed. A mix of the riskless asset with the risky asset, x_α, is given by $x_\alpha = (1-\alpha)$ $r+\alpha$ x whose cumulative distribution is F_α. If there is an $\alpha > 0$ such that F_α dominates G by FSD, SSD or TSD, then we can safely conclude that for any mix of G with the riskless asset there will be a mix of F with the riskless asset which dominates it by these rules, respectively. Thus, if $F_\alpha DG$ for one positive value α, then the set $\{F_\alpha\}$ will dominate the set $\{G_\beta\}$. Such dominance is called Stochastic Dominance with a Riskless Asset (SDR). For $U \in \mathbf{U}_1$ we developed FSD with a riskless asset denoted by FSDR. Similarly, for $U \in \mathbf{U}_2$ we presented SSDR. Finally, we established stochastic dominance rules with a riskless asset in the more realistic case where the borrowing rate r_b is higher than the lending rate r_l. The quantile framework is also more convenient for developing an algorithm for FSD and SSD (as will be seen in the next chapter), but not allowed to be employed instead of the cumulative distributions for TSD and higher degree SD rules.

Key Terms

Quantile Distribution

The FSD Rule with Quantiles

The SSD Rule with Quantiles

First-degree Stochastic Dominance with a Riskless Asset: The FSDR Rule

Second-degree Stochastic Dominance with a Riskless Asset: The SSDR Rule

ALGORITHMS FOR STOCHASTIC DOMINANCE

To obtain the efficient sets corresponding to the various stochastic dominance (SD) rules, we need to know the precise shape of the various distributions under comparison. For example, for SSD, in order to check whether dominance exists or not, we need to calculate the area enclosed between the two distributions under consideration, F and G; and in order to be able to carry such a calculation, we need to know the precise distribution of the rates of return. Thus, we need to know the distributions in applying the various dominance rules. In practice, stochastic dominance rules are commonly applied to *empirical distributions* (e.g., *ex-post rates of returns* of mutual funds or of other available portfolios). In such cases, if we have n observations, say, n annual rates of return, or n monthly rates of return, then each observation is commonly assigned an equal probability of 1/n. Thus, we have empirical distributions which, for investment decision making, generally serve also as estimates of future distributions. Relatively simple SD and SDR algorithms exist for this framework of uniform discrete distributions. The first algorithm for FSD and SSD was developed by Levy and Hanoch in 1969,[1] and the algorithms for TSD as well as FSDR, SSDR and TSDR were developed by Levy and Kroll.[2] Porter, Wart and Ferguson also developed algorithms that employ necessary rules to help cut down the number of necessary comparisons. This is particularly beneficial when a large number of portfolios are compared. This chapter is devoted to the various SD and SDR algorithms.[3]

5.1 USING THE NECESSARY CONDITIONS AND TRANSITIVITY TO REDUCE THE NUMBER OF COMPARISONS

Suppose that there are N distributions (e.g., portfolios, or mutual funds) in the feasible set. Then, in principle, we have to carry $P_2^N = \frac{N!}{(N-2)!}$ comparisons where P denotes permutations. We emphasize permutations rather than combinations, because for each pair of distributions, F and G, we have to examine whether F dominates G, and, if such dominance does not exist we also have to check whether G dominates F. Thus, if, for example, there are 100 investments in the feasible

[1] See, Levy, H., and Hanoch, G., "Relative Effectiveness of Efficiency Criteria for Portfolio Selection," *Journal of Financial and Quantitative Analysis*, 1970.

[2] See, Levy, H., and Kroll, Y., "Efficiency Analysis with Borrowing and Lending: Criteria and their Effectiveness," *Review of Economics and Statistics*, 1979.

[3] Porter, R.B., Wart, J.R., and Ferguson, D.L., "Efficient Algorithms for Conducting Stochastic Dominance Tests on a Large Number of Portfolios," *Journal of Financial and Quantitative Analysis*, 1973.

set, N = 100 and, therefore, the maximum number of comparisons that we need to conduct will be $\frac{100!}{98!} = 100 \cdot 99 = 9900$. However, the number of comparisons can be drastically reduced by employing some necessary rules for dominance and the transitivity property of the SD and SDR rules. For example, the mean rate of return can be calculated and all 100 investments corresponding to the above example can be ordered in descending order:

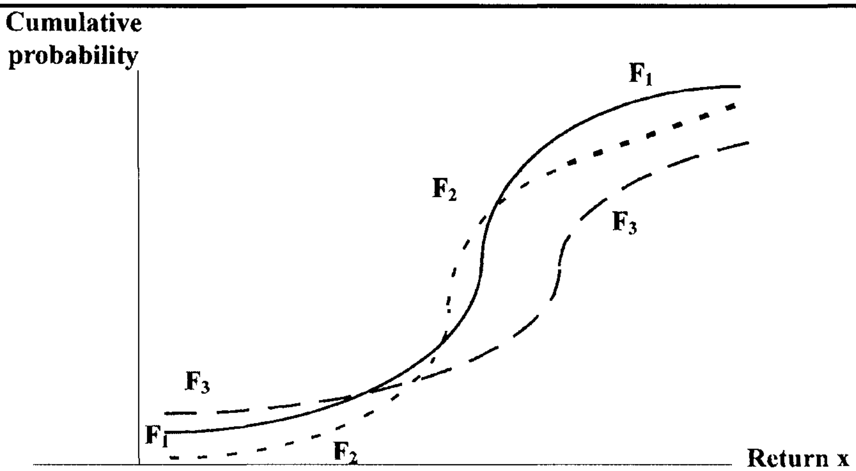
$$\overline{X}_{100} > \overline{X}_{99} > \overline{X}_{98} > \ldots > \overline{X}_1 \quad .$$

A necessary condition for SD dominance (by all SD rules) is that the superior investment must have a higher (or equal) mean. Therefore, the number of necessary comparisons is reduced to C_2^N (where C stands for combinations). For example, if we wish to compare whether a distribution ranked 100 dominates a distribution ranked 99, there is no need to check whether distribution 99 dominates distribution 100; it is impossible because $\overline{X}_{100} > \overline{X}_{99}$ and dominance of 99 over 100 violates the necessary condition regarding the means for dominance. Using this necessary condition for dominance reduces, in our specific example, the number of comparisons to:

$C_2^{100} = \frac{100!}{(100-2)!2!} = 4{,}950$. By the same token, we can use the other necessary rules, in particular, the "left tail" necessary condition (which is easy to check), to further reduce the number of comparisons.

The necessary rules should be checked before any dominance calculations are conducted. To illustrate further how these necessary rules can be used to reduce the number of necessary comparisons, consider Figure 5.1 which presents three distributions F_1, F_2 and F_3. There are $P_2^3 = \frac{3!}{(3-2)!} = 6$ permutations. However, for FSD, SSD as well as TSD, it is suggested that the mean of each investment be calculated first and that this information be used on the means to reduce the number of comparisons to $C_2^3 = \frac{3!}{(3-1)!2!} = 3$.

Figure 5.1: Using necessary conditions to reduce the number of comparisons



If it is given that $\bar{x}_{F_3} > \bar{x}_{F_2} > \bar{x}_{F_1}$, we will be left with only the following three comparisons: Does F_3 dominate F_2? Does F_3 dominate F_1? Does F_2 dominate F_1? All other comparisons (i.e., F_2DF_3?, F_1DF_3?, F_1DF_2?) will not be necessary because of the required condition of the means. Using the "left tail" necessary condition, we can also safely conclude that F_3 cannot dominate F_2 and F_1 because it has a thicker left tail. So, using these two necessary conditions (the mean and the left tail), we are left with only one comparison; namely, does F_2 dominate F_1? Such a dominance is possible because F_2 has a higher mean than F_1, and F_1 has a thicker left tail than F_2. Thus, in our specific example, these two necessary conditions reduce the number of necessary comparisons from six to one. Of course, we could also calculate the geometric means of all available distributions and use this necessary condition (for FSD, SSD and TSD) to reduce the number of necessary comparisons.

Knowing that FSD dominance implies SSD dominance and SSD dominance implies TSD dominance, it is suggested that FSD be conducted first. Then SSD will be conducted only on the FSD efficient set and, finally TSD will be conducted only on the SSD efficient set. For example, suppose that out of 100 investments under consideration, 20 are in the SSD efficient set: Then, because the TSD efficient set is a subset of the SSD efficient set, the maximum number of comparisons for TSD will be only (using also the necessary condition of the means) $C_2^{20} = \frac{20!}{18!2!} = 190$ rather than $C_2^{100} = \frac{100!}{98!2!} = 4950$.

Finally, as we will see below, FSD, SSD, and TSD investment rules (as well as the corresponding SDR rules) are *transitive rules*. The transitivity property implies that for three distributions F, G and H, if FDG and GDH , then also

FDH. To see why this transitivity property reduces the number of necessary comparisons, once again, suppose that we have, say, 100 investments. Then, after ranking these investments by their means, we will have a maximum of 4,950 comparisons to conduct. Suppose that we find that F_{100} dominates, say, 40 out of the available 100 investments. Then these 40 investments will be relegated into the inefficient set. Suppose that F_i is relegated into the inefficient set. Should we keep F_i in the meantime in order to check whether F_iDF_j? (where j is some other investment $i \neq j \neq 100$). There is no need because, if F_iDF_j, due to the transitivity, also $F_{100}DF_j$. Thus, it is guaranteed that the inferiority of F_j will be discovered by F_{100}, hence F_i can be safely relegated into the inefficient set. This means that after comparing F_{100} to all other investments (99 comparisons), if 40 investments are dominated by F_{100}, the maximum number of comparisons left will be $\frac{59!}{2!(59-2)!} = \frac{59 \cdot 58}{2} = 1{,}711$, where 59 are all 100 mutual funds less the ones which were relegated into the inefficient set by F_{100} less F_{100} itself for which comparison with the others has already been conducted. Using the transitivity property of the decision rule under consideration, the total number of comparisons in the above hypothetical example is reduced from 4,950 to 1,711 + 99 = 1,810 (where 99 is the number of comparisons with the others).

By the same token, the number of comparisons will be further reduced when more investments are relegated to the inefficient set by investments other than F_{100}, say by F_{99}, F_{98}, etc.

In the following theorem, it is claimed that all optimal rules are transitive. Because FSD, SSD, TSD and all the corresponding rules with a riskless asset (FSDR, SSDR and TSDR) are optimal, the investment decisions will also be transitive.

Theorem 5.1:

An optimal investment decision rule is a transitive rule.

Proof: Let F, G, and H denote three investments and assume that FD_iG and GD_iH by some decision rule D_i. We need to prove that if D_i is an optimal rule, then FD_iH is implied.

Suppose that we employ an optimal rule D_i corresponding to utility class $\mathbf{U}_i$. Then , because an optimal rule is also a sufficient rule, we have:

FD_iG implies $E_FU(x) \geq E_GU(x)$ for all $U \in \mathbf{U}_i$

GD_iH implies $E_GU(x) \geq E_HU(x)$ for all $U \in \mathbf{U}_i$

Because real numbers are transitive, we have $E_F U(x) \geq E_G U(x) \geq E_H U(x)$ for all $U \in \mathbf{U}_i$. However, because an optimal rule is also a necessary rule, $E_F U(x) \geq E_G U(x)$ for all $U \in \mathbf{U}_i$ implies that FD_iH, which completes the proof.

If a decision rule is sufficient but not necessary, $E_F U(x) \geq E_H U(x)$ will not imply that F dominates H by this rule. As FSD, SSD, and TSD as well as FSDR, SSDR, and TSDR are optimal rules, by employing those rules we can use the transitivity property in the algorithm presented below; namely, any dominated investment can safely be relegated into the inefficient set.

5.2 THE FSD ALGORITHM

Suppose that there are n observations (e.g., n annual rates of return). Denote the rates of return of F and G by x and y, respectively. Reorder the observations x and y from the lowest to the highest value such that:

$$x_1 \leq x_2 \leq \ldots \qquad \ldots \leq x_n$$
$$y_1 \leq y_2 \leq \qquad \ldots \leq y_n .$$

Assigning a probability of 1/n to each observation (If there are two identical observations, write them one after the other and assign a 1/n probability to each one) the FSD algorithm will be stated as follows:

The FSD algorithm: F (or x) dominates G (or y) by FSD if and only if $x_i \geq y_i$ for all i = 1,2,...,n and there is at least one strict inequality.

This claim implies that if $x_i \geq y_i$ for all i, and $x_i > y_i$ for some i, then with equal probability assigned to each observation, F must be located below G in the whole range, and at some range, it will be strictly below G. To see this, recall that $F(x_i) = i/n$ (i = 1,2,...,n) and $G(y_i) = i/n$ (i = 1,2,...,n).[4] If $x_i \geq y_i$ for all i, then for all P ($P = \frac{1}{n}, \frac{2}{n} .. \frac{i}{n} .. \frac{n}{n}$), $Q_F(P) \geq Q_G(P)$ and, for some value P, $Q_F(P) > Q_G(P)$, which implies that FD_1G (or F is below G).

Similarly, if $x_i \geq y_i$ (and for some i a strict inequality holds) for all $i \neq j$ but for one observation j we have $x_j < y_j$, then F and G must intercept; hence there is no FSD. This property is demonstrated in the next example.

Example:

Suppose that we have the following five annual rates of return for three assets:

[4] This definition holds only if $x_i \neq x_j$ for all (i,j). Otherwise, one needs to modify $F(x_i)$ and $G(x_i)$ to include cases where $x_i = x_j$. For simplicity only, and without loss of generality, we assume in this chapter that $x_i \neq x_j$ for all (i,j).

Rates of return (in%)			
Year	x (or F)	y (or G)	z (or H)
1	–4	–15	–15
2	0	0	0
3	10	9	12
4	15	15	15
5	–10	–5	–5

Is there FSD? First let us reorder the observations from the lowest to the highest;

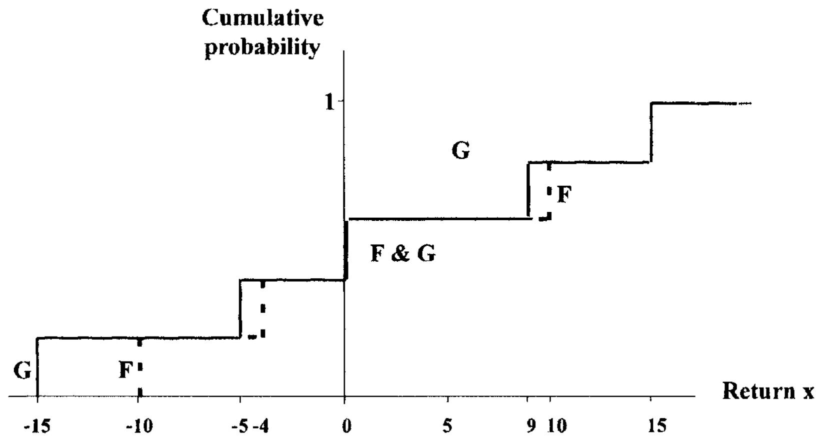
x_i(or F)	y_i(or G)	z_i(or H)
–10	–15	–15
–4	–5	–5
0	0	0
10	9	12
15	15	15

Does G dominate F by FSD? No, it doesn't because the left-tail necessary distribution does not hold: $\text{Min}_F(x) = -10 > \text{Min}_G(y) = -15$, hence G cannot dominate F by FSD. Does F dominate G by FSD? Using the FSD algorithm we have:

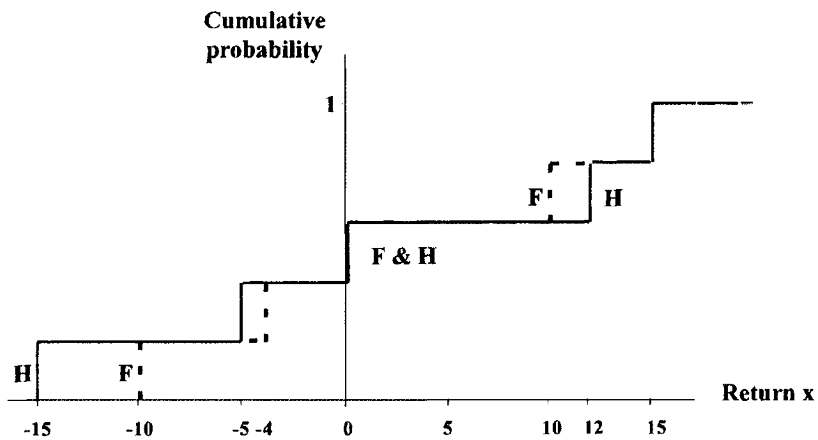
$$\begin{aligned} x_1 &= -10 > y_1 = -15 \\ x_2 &= -4 > y_2 = -5 \\ x_3 &= 0 = y_3 = 0 \\ x_4 &= 10 > y_4 = 9 \\ x_5 &= 15 = y_5 = 15 \end{aligned}$$

Thus, $x_i \geq y_i$ for all i and for, say, i=1, $x_1 > y_1$; hence, FD_1G. If, instead of G we had distribution H (or rates of return z), then we would have no dominance because for i = 4, we would have $x_4 = 10 < z_4 = 12$; hence, the condition required by the algorithm does not hold and F does not dominate H by FSD. Figure 5.2 demonstrates these two comparisons of F and G, and F and H, with a probability 1/5 assigned to each observation.

Figure 5.2: A demonstration of FSD algorithm



5.2a: F dominated G by FSD



5.2b: F does not dominate H by FSD

As can be seen from Figure 5.2, $x_i \geq y_i$ for all i implies that FD_1G (see Figure 5.2a). However, in the comparison of F and H, for $i = 4$, we have $x_4 = 10 < z_4 = 12$; hence, H is below F in the range $10 < x < 12$. The algorithm condition $x_i \geq z_i$ for all i does not hold and F does not dominate H as shown in Figure 5.2b.

5.3 THE SSD ALGORITHM

As with FSD, we rank all observations from the lowest to the highest and assign a probability of 1/n to each observation. Then we define $X'_i (i=1,2,...,n)$, as follows:

$$X'_1 = x_1$$
$$X'_2 = x_1 + x_2$$
$$\vdots$$
$$X'_i = \sum_{j=1}^{i} x_j$$
$$\vdots$$
$$X'_n = \sum_{j=1}^{n} x_j$$

and Y'_i is defined similarly.

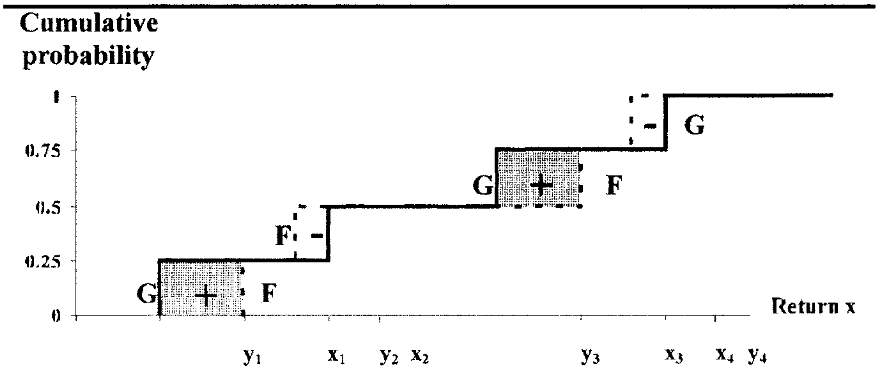
The SSD Algorithm: F (or x) dominates G (or y) by SSD if and only if:

$$X'_i \ge Y'_i \text{ for all } i = 1, 2,...,n$$

and there is at least one strict inequality.

Let us explain this algorithm via Figure 5.3. G does not dominate F by SSD because $\text{Min}_G(x) = y_1 < \text{Min}_F(x) = x_1$. Does F dominate G by SSD?

Figure 5.3: A demonstration of SSD algorithm



Using the SSD quantile framework, F dominates G by SSD if and only if $I_2(P) \equiv \int_0^P [Q_F(t) - Q_G(t)]dt \geq 0$ for all $0 \leq P \leq 1$ (and there is at least one strict inequality). As can be seen from Figure 5.3, it is sufficient to check this inequality only at suspect points P = ¼, ½, ¾ and 1 because, if $I_2(¼) > 0$, then for any P < ¼, $I_2(P) \geq 0$. Similarly, if in addition to $I_2(¼) > 0$, also $I_2(½) \geq 0$, then for any P ≤ ½, $I_2(P) \geq 0$, and if also $I_2(¾) \geq 0$, then for any P ≤ ¾, $I_2(P) \geq 0$. Finally, if also $I_2(1)$, then for any P ≤ 1, $I_2(P) \geq 0$.

Let us now calculate $I_2(P)$ at these four suspected points:

At P = ¼ we have:
$I_2(¼) = ¼\,(x_1 - y_1)$ and, if $x_1 \geq y_1$ (or $X'_1 \geq Y'_1$), then $I_2(¼) \geq 0$.

Similarly, for P = ½ we have:

$$I_2(½) = (x_1\,¼ + x_2\,¼) - (y_1\,¼ + y_2\,¼)$$
$$= ¼\,(x_1 + x_2 - y_1 - y_2)$$

and if:

$$X'_2 \equiv x_1 + x_2 \geq y_1 + y_2 \equiv Y'_2,$$

then, $I_2(½) \geq 0$.
Similarly, at P = ¾ we have:

$$I_2(¾) = (¼\,x_1 + ¼\,x_2 + ¼\,x_3) - (¼\,y_1 + ¼\,y_2 + ¼\,y_3)$$
$$= ¼\,(x_1 + x_2 + x_3 - y_1 - y_2 - y_3) = ¼\,(x'_3 - y'_3).$$

Hence, if:

$$X'_3 \geq Y'_3,\ I_2(¾) \geq 0.$$

Finally, at P = 1, it can be shown in a similar way that:

$$I_2(P) = ¼\,(X'_4 - Y'_4) \text{ and if } X'_4 \geq Y'_4 \text{ also } I_2(1) \geq 0.$$

When there are n observations, it can be shown in a similar way that if $X'_i \geq Y'_i$ for all i (and to avoid a trivial case that F = G everywhere we require at least one strict inequality), then FD_2G. Thus, if the inequalities required by the SSD algorithm hold, then $I_2(P) \geq 0$ for all $0 \leq P \leq 1$; hence FD_2G.

Example:

In this example, we employ FSD and SSD algorithms, and show that there is no FSD but there is SSD of F over G.

Let F and G have the following rates of return:

Year	**x_i(or F)**	**y_i(or G)**
1	5	9
2	10	10
3	2	–4

Is there FSD? SSD? Let us first reorder the rates of return in increasing order:

i	**x_i(F)**	**y_i(G)**
1	2	–4
2	5	9
3	10	10

Using these figures, we next calculate X_i' and Y_i' :

i	X_i' (G)	Y_i' (G)
1	2	–4
2	7	5
3	17	15

With these figures, the condition $x_i \geq y_i$ does not hold for all i (see i = 2); hence, there is no FSD. However, $X_1' \geq Y_1'$ for i = 1, 2, 3 (and there is at least one strict inequality); hence FD_2G.

5.4 THE TSD ALGORITHM

One is tempted to employ the same technique employed in FSD and SSD to establish the TSD algorithm. However, we cannot use in the TSD case the same technique for two reasons:

a) Unlike FSD and SSD, with TSD we cannot compare the two distributions under consideration only at the points of jumps in probabilities. The reason is that the difference in the relevant FSD and SSD integrals is linear but the difference corresponding to the TSD integral is not linear, hence a suspected point, which may violate TSD may be interior. The first researchers to claim this point are Fishburn and Vickson.[5]

b) Using FSD and SSD, one can use an algorithm, which is based either on the distributions' quantiles, or on the cumulative distributions. The switch between cumulative distributions and quantiles is completely wrong with TSD and, in this case, one should adhere to the comparison of the cumulative distributions.

Porter, Wart and Ferguson[6] compare the two distributions only at jump points with jumps in probabilities, hence their TSD algorithm is wrong. Levy & Kroll[7] and Levy[8] employ the quantile approach to establish the TSD algorithm and hence it is also wrong. Ng[9], has shown by a simple numerical example that the switch from cumulative distributions to quantiles is not allowed in the TSD case while it is fully legitimate in the FSD and SSD cases.

Thus, in the derivation of TSD algorithm, we employ cumulative distribution and check the integral condition also in interior points, which correct for the errors in the previous published algorithms.[10]

In the derivation of the TSD algorithm we employ the following definitions corresponding to the cumulative distribution:

---

[5] P.C. Fishburn and R.G. Vickson, "Theoretical foundation of stochastic dominance," in G.A. Whitmore and M.C. Findlay, editors, *Stochastic Dominance: An Approach to Decision-Making Under Risk*. Lexington, Toronto, 1978.

[6] Porter, R.B., J.R. Wart, and D.L. Ferguson, "Efficient algorithms for conducting stochastic dominance tests on a large number of portfolios," *Journal of Financial and Quantitative Analysis,"* 1973, 8:71–82.

[7] Levy, H., and Y. Kroll, "Efficiency analysis with borrowing and lending criteria and their effectiveness," *Review of Economics and Statistics*, 1979, 61:25–30.

[8] Levy, H., *Stochastic Dominance*, Kluwer Academic Press, Boston, 1998.

[9] Ng, M.C., "A remark on third stochastic dominance," *Management Science*, 46, 2000, pp. 870–873.

[10] The TSD algorithm presented in this section is based on Levy, H., Leshno, M., and Hechet, Y., "Third Degree Stochastic Dominance: An Algorithm and Empirical Study," 2004, working paper, Hebrew University, Jerusalem, Israel.

2. Basic Definitions

Define,

$$F_2(x) = \int_{-\infty}^{x} F(t)dt \tag{5.1}$$

and

$$F_3(x) = \int_{-\infty}^{x} F_2(t)dt \tag{5.2}$$

where F is the cumulative distribution and the subscripts 2 and 3, as we shall see later on correspond to second and third degree SD, respectively. For two assets with distribution functions F and G we say that F dominates G by the TSD criterion if for all $-\infty < x < \infty$, $F_3(x) \leq G_3(x)$ with at least one strict inequality and $E_F(X) \geq E_G(X)$.

Assume that the empirical returns of the two assets denoted by F and G respectively are: $x_1 \leq x_2 \leq \cdots \leq x_{n-1} \leq x_n$ and $y_1 \leq y_2 \leq \cdots \leq y_n$. Let z_1, $z_2, \ldots, z_{2n}$, be the common grid of x_1 and y_j, i.e. $z_k = x_j$ for some i or $z_k = y_j$ for some j. Thus we have,

$$F(x) = \begin{cases} 0 & x < x_1 \\ \frac{1}{n} & x_1 \leq x < x_2 \\ \vdots & \\ \vdots & \\ \frac{n-1}{n} & x_{n-1} \leq x < x_n \\ 1 & x_n \leq x \end{cases} \tag{5.3}$$

Therefore, an area under the step function $F(x)$ is given by:

$$F_2(x)=\int_{-\infty}^{x}F(t)dt=\begin{cases}0 & x\le x_1\\ \frac{1}{n}(x-x_1) & x_1\le x\le x_2\\ \frac{2}{n}x-\frac{1}{n}(x_1+x_2) & x_2\le x\le x_3\\ \vdots & \\ \frac{k}{n}x-\frac{1}{n}\left(\sum_{i=1}^{k}x_i\right) & x_k\le x\le x_{k+1}\\ \vdots & \\ x-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right) & x_n\le x\end{cases} \tag{5.4}$$

and

$$F_3(x)=\int_{-\infty}^{x}F_2(t)dt=\begin{cases}0 \quad \text{for} x\le x_1\\ \frac{1}{2n}(x-x_1)^2 \quad \text{for } x_1\le x\le x_2 \\ F_3(x_k)+\int_{x_k}^{x}F_2(t)dt=F_3(x_k)+\frac{k}{2n}(x^2-x_k^2)-\frac{1}{n}\left(\sum_{i=1}^{k}x_i\right)(x-x_k)\\ \text{for } \quad x_k\le x\le x_{k+1} \ k=2,\dots,n-1\\ F_3(x_n)+\int_{x_n}^{x}F_2(t)dt=F_3(x_n)+\frac{1}{2}(x^2-x_n^2)-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)(x-x_n)\\ \text{for } x_n\le x\end{cases} \tag{5.5}$$

In a similar way we define G, G_2 and G_3 for the other distribution under comparison. We will use these definitions in the TSD algorithm given below. Note that for every k the intervals $[z_k,z_{k+1}]$ satisfy $[z_k,z_{k+1}]\subseteq[x_i,x_{i+1}]$ and $[z_k,z_{k+1}]\subseteq[y_j,y_{j+1}]$ for some i and j. Thus one can define $F_1(z)$ by the following way: for each $z\in[z_k,z_{k+1}]$ define $F_1(z)$ by the projection of $F_1(x)$ from $[x_i,x_{i+1}]$ on $[z_k,z_{k+1}]$ where $[z_k,z_{k+1}]\subseteq[x_i,x_{i+1}]$. Similarly we define $F_2(z)$ and $F_3(z)$.

Having these definitions and relationships between F, F_2 and F_3 we turn now to the TSD algorithm. With no loss of generality assume that G does not dominates F with the TSD criterion (otherwise we change their names respectively), and we

need to verify whether F dominates G by the TSD criterion. Thus we need to check whether $F_3(x) \le G_3(x)$ for all x with at least one strict inequality, and $E_F(X) \ge E_G(X)$. Define H(x) as the difference
$H(x) = G_3(x) - F_3(x)$. As $F_3(x)$ and $G_3(x)$ are parabolas in each interval $[z_k, z_{k+1}]$ (see equation 5.5), H(x) is also a parabola or a segment of a parabola in each interval $[z_k, z_{k+1}]$. Therefore, for any z, $z_k \le z \le z_{k+1}$, H(z) is a polynomial function of order 2 (a parabola). Therefore, for any z, $z_k \le z \le z_{k+1}$ we have that $0 \le H(z)$ if and only if $0 \le H(z_k)$, $0 \le H(z_{k+1})$ and

i. $0 \le H'(z_k)$ or

ii $0 \ge H'(z_k)$ and $0 \ge H'(z_{k+1})$ or

iii $0 > H'(z_k)$ and $0 \le H'(z_{k+1})$. And because H(z) is a parabola in $[z_k, z_{k+1}]$ it can be written as $H(z) = az^2 + bz + c$, we require also that $H\left(\frac{-b}{2a}\right) \ge 0$. We can calculate the coefficients of parabola in terms of (5.5). To show this, let $[z_k, z_{k+1}] \subseteq [x_i, x_{i+1}]$ and $[z_k, z_{k+1}] \subseteq [y_j, y_{j+1}]$ therefore we have the following explicit expressions for a, b, and c:

$$c = G_3(y_j) - \frac{j}{2n} y_j^2 + \frac{1}{n}\left(\sum_{t=1}^{j} y_t\right) y_j - \left[F_3(x_i) - \frac{i}{2n} x_i^2 + \frac{1}{n}\left(\sum_{t=1}^{i} x_t\right) x_i\right]$$

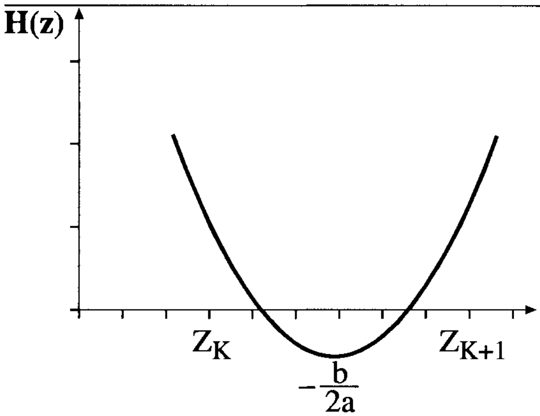
and

$$b = \frac{1}{n}\left(\sum_{t=1}^{i} x_t\right) - \frac{1}{n}\left(\sum_{t=1}^{j} y_t\right)$$

$$a = \frac{j=i}{2n}$$

To be more specific, any polynomial of order 2 (a parabola) has one extremum point (minimum or maximum). If H(z) is an increasing function at the end point z_k (i.e. $0 \le H'(z_k)$, this is case 1 above) we can conclude that for all $z \in [z_k, z_{k+1}]$, $H(z) \ge 0$. When H(z) is a decreasing function at the end points z_k, z_{k+1} (i.e. $0 \ge H'(z_k)$ and $0 \ge H'(z_{k+1})$) then, $0 \le H(z_k), 0 \le H(z_{k+1})$, once again, guarantees that H(z)≥0 for all $z \in [z_k, z_{k+1}]$ (this is case 2 above). When H(z) is increasing at z_k and decreasing at z_{k+1} then H(z) has a maximum point and therefore, $0 \leq H(z_k), 0 \leq H(z_{k+1})$ guarantees that H(z) $\geq$ 0 for all $z \in [z_k, z_{k+1}]$. We turn now to the most interesting case: if H(z) decreases at z_k and increases at z_{k+1}, the conditions $0 \leq H(z_k), 0 \leq H(z_{k+1})$ do not guarantee that H(z) is non negative for all $z \in [z_k, z_{k+1}]$, because its minimum is in the open interval $z \in (z_k, z_{k+1})$ and therefore we need to check that its minimum value of H(z) is non negative (see Figure 5.4). This case corresponds to case (iii) above, namely we check the value of H(z) in its minimum point. In the existing algorithms this possibility is overlooked because H(z) is checked only at points with probabilities jumps i.e. at z_k, z_{k+1} but not at any value within its interval.

Figure 5.4: H(z) is positive at z_k and z_{k+1}, however, H(z) is not non-negative for all points in the interval $[z_k, Z_{k+1}]$, ($H(z) = az^2 + bz + c$).



Note that if H(z) $\geq$ 0 at the minimum point then H(z) $\geq$ 0 at the whole interval $[z_k, z_{k+1}]$. Let us elaborate on this minimum point. The minimum point is given for z where $H'(z) = 0$, i.e. at the point $(\frac{-b}{2a})$ where $H(z) = az^2 + bz + c$. We need to verify that $H(\frac{-b}{2a}) \geq 0$. If $H(\frac{-b}{2a}) < 0$ then there is no TSD despite the fact that at any probability jump H(z) $\geq$ 0 [11]. Therefore to check for TSD dominance we need to go back to the distributions of F_3, F_2 H(z) and $H'(z)$.

[11]Note that if H(z) would be linear rather than a parabola, checking H(z) at probability jumps would be sufficient. Indeed this is the case with FSD and SSD algorithms.

Note that,

$$\frac{d}{dx}F_3(x) = F_2(x) = \begin{cases} 0 & x \le x_1 \\ \frac{1}{n}(x - x_1) & x_1 \le x \le x_2 \\ \frac{2}{n}x - \frac{1}{n}(x_1 + x_2) & x_2 \le x \le x_3 \\ \vdots & \\ \frac{k}{n}x - \frac{1}{n}\left(\sum_{i=1}^{k} x_i\right) & x_k \le x \le x_{k+1} \\ \vdots & \\ x - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) & x_n \le x \end{cases} \tag{5.6}$$

(and $\frac{d}{dx}G_3(x)$ is defined in a similar way).

Therefore in order to verify whether *F* dominates *G* with the TSD criterion it is enough to verify that

1. $E_F(X) \ge E_G(X)$

2. $F_3(z) \le G_3(z)$ for $z \in \{x_i, y_i : i = 1,2,...n\}$ (which is the algorithm which checks the integral at the probability jump points. However because the minimum point is not checked this is not a sufficient condition for dominance). We need also to check the interior point:

3. If for some k, $0 \ge H'(z_k) \equiv [G_2(z_k) - F_2(z_k)]$
and $0 \le H'(z_{k+1}) \equiv [G_2(z_{k+1}) - F_2(z_{k+1})]$ we need to check also
whether $H(\frac{-b}{2a}) \ge 0$ before a dominance by the TSD is declared.

To be more specific, assume that for some *k*, $0 \ge H'(z_k) = [G_2(z_k) - F_2(z_k)$ and $0 \le H'(z_{k+1}) = G_2(z_{k+1}) - F_2(z_{k+1})$. Let i and j be such $z_k \le y_j \le z_{k+1}$ and $z_k \le x_i \le z_{k+1}$. In the equation H(z) = $az^2 + bz + c$, a is the difference between the coefficients of z^2 in G_3 and F_3. We have that $a = \frac{j=i}{2n}$ and $b = \frac{1}{n}\left(\sum_{t=1}^{i} x_t\right) - \frac{1}{n}\left(\sum_{t=1}^{j} y_t\right)$. Another way to calculate $\frac{-b}{2a}$ is given by

$\frac{-b}{2a} = z_k - \frac{H'(z_k)(z_{k+1} - z_k)}{H'(z_{k+1}) - H'(z_k)}$. Checking the value of $H(\frac{-b}{2a})$ is made by computing $G_3(-\frac{b}{2a}) - F_3(-\frac{b}{2a})$ using eq. (5.5).

The following example illustrates that the existing TSD algorithm is wrong because it may conclude that there is a TSD dominance where actually such dominance does not exist.

5.5. A NUMERICAL EXAMPLE SHOWING THE FLAW IN EXISTING TSD ALGORITHM:

We illustrate the flaw of the existing TSD algorithm with the following numerical example. Suppose that the empirical distribution of F and G is given by the 6 annual rate of return presented in the following table:

	Period 1	Period 2	Period 3	Period 4	Period 5
Distribution F	5%	10%	10%	20%	30%
Distribution G	5%	10%	20%	20%	20%

The functions F_2,G_2 and F_3,G_3, given in equations 4 and 5, at the probability jump points are presented in the following table:

	5%	10%	20%	30%
F_2	0	0.0083	0.075	0.1583
G_2	0	0.0167	0.0667	0.1667
$G_2 - F_2$	0	0.0084	-0.0083	0.0084
F_3	0	0.000208	0.004375	0.016041
G_3	0	0.000416	0.004583	0.01625
$G_3 - F_2$	0	0.0002083	0.0002083	0.0002083

Because $G_2 - F_2$ is positive at 10% and negative at 20% then F does not dominate G by the SSD rule. Similarly, it is easy to see that G does not dominate F by SSD. Is there a TSD? According to the existing algorithm which checks the integral condition only at end points, F dominates G by the TSD rule, because $G_3 - F_3$ is positive at all probability jump points 10%, 20% and 30%. We claim that this does not guarantee that F dominates G by the TSD because interior point should also be examined. Indeed $G_3 - F_3$ is negative at the point 25%, and therefore F does not dominate G by the TSD. To be more specific, the corresponding number for 25% is F_2 (25%) = G_2(25%), = 0.1166, G_2(25%) − F_2(25%) = 0, and F_2(25%) = $9.1666667*10^{-3}$, G_3(25%) = $9.1663333*10^{-3}$, hence G_3(25%) − F_3(25%) = $3.3*10^{-7} < 0$.

Using the existing algorithm one does not check the value of $G_3 - F_3$ at 25% because the point 25% is not an observation in the sample, hence it is not a probability jump point.

We suggest the following steps in checking TSD dominance:

1. Sort the empirical rates of return of F and G.

2. If min (F) < min(G) then rename F to G and G to F, (min (F) < min(G) means that after excluding all the values in F and G that are equal the minimum value of F is less than the minimum value of G).

3. Use equations (5.4) and (5.5) to calculate the value of $F_2(x_k)$ and $G_2(y_k)$, $F_3(x_k)$ and $G_3(y_k)$, $k = 1,\ldots,n$.

4. Take all the values z_i, where each z_i is equal to some x_j or y_j , $i = 1,2,\ldots, 2n$ and sort the values z_i, i.e., $z_1 \le z_2 \le \cdots \le z_{2n}$.

5. If $H(z_k) = G_3(z_k) - F_3(z_k) \ge 0$, $k = 1,2,\ldots,2n$ and

6. If for some $k \le 2n - 1$, $0 > (G_2(z_k) - F_2(z_k))$ and $0 \le (G_2(z_{k+1}) - F_2(z_{k+1}))$ then let i and j be such $z_k \le y_j \le z_{k+1}$ and $z_k \le x_i \le z_{k+1}$. Check if $H(\frac{-b}{2a}) \ge 0$ where $a = \frac{j-i}{2n}$; and $b = \frac{1}{n}\left(\sum_{t=1}^{i} x_t\right) - \frac{1}{n}\left(\sum_{t=1}^{j} y_t\right)$; If $H(z_k) \ge 0$ for $k = 1,\ldots,2n$ and $H(\frac{-b}{2a}) \ge 0$, and there is at least one strict inequality then F dominates G by the TSD rule.

5.6 THE EMPIRICAL RESULTS

In Chapter 7 we provide the efficient sets corresponding to various decision rules as published in the literature. However, in this section we provide a comparison of the empirical efficiency of the various rules, with an emphasize on the difference between the wrong existing TSD algorithm and the correct TSD algorithm, developed in this chapter. The data set employed in the empirical study covers weekly, monthly and quarterly rates of return of a sample of mutual funds corresponding to the period April 1997- July 2002. We construct the efficient sets using M-V criterion, the SSD and the TSD rules. In building the TSD efficient sets we used the correct version of the algorithm as well as the wrong algorithm, which compares to integral only at end points of probability jumps.
The data set corresponding to 37 mutual funds. We are mainly interested in examining the following aspects of SD and M-V efficiency sets:

1. The relative size of the wrong and correct TSD efficient sets.
2. The relative effectiveness of SSD and TSD rules in reducing the size of the efficient sets.
3. The relative magnitudes of the SD and M-V efficient sets.

4. A comparison of the efficient sets induced by switching from weekly to monthly and quarterly data.

Table 5.1 provides the results corresponding to the 37 mutual funds.

Table 5.1: The percentage of Mutual Funds in the Efficient Set

	M-V	SSD	TSD	TSD (wrong algorithm)
Weekly	19%	30%	24%	3%
Monthly	24%	27%	22%	5%
Quarterly	30%	35%	22%	8%

A few conclusions can be drawn from this table. First note that the wrong TSD algorithm always yield a smaller efficient set than the correct TSD algorithm. For example, with weekly data the efficient set is 24% of the population with the correct TSD algorithm and only 3% with the wrong TSD algorithm. The fact that with the wrong TSD algorithm a smaller efficient set is obtained is not surprising in light of the fact that the correct TSD algorithm imposes one more check before TSD dominance is established. The difference between these two sets is quite large, and it is large also for the other two investment horizons: 5% versus 22% monthly data and 8% versus 22% with quarterly data. Thus, the correction in the TSD algorithm has a substantial effect on the size of the efficient set.

As explained above the size of the TSD efficient set is always smaller or equal to the SSD efficient set. One wonders whether to assume only risk aversion or to add the assumption $u''' > 0$. We found that with quarterly data the additional assumption $u''' > 0$ is most beneficial as it reduces the efficient set from 35% to 22% of the population. For the other two horizons the benefit from the additional assumption $u''' > 0$ is also beneficial but to a lower degree.

The last analysis relates to the comparison between the M-V and SD efficient sets. If the distributions are normal the SSD and M-V efficient set must be identical. However, with the empirical non-normal distributions there is no predicted relationship between the M-V and SSD efficient sets and the content and the size of the M-V and SSD efficient sets are generally different. This is indeed the case in our study, see Table 5.1. Recall, however, that if distributions are not normal the M-V rule is not consistent with expected utility theory and one should rely on SSD rather than M-V.

Finally, note that the efficient set is affected by the assumed investment horizon. It is possible that for an investor for "one week" horizon, a certain portfolio may be inefficient but for a longer horizon, e.g., one month is efficient.

5.7. THE SDR ALGORITHM

a) FSDR Algorithm

Let us demonstrate the FSDR algorithm in detail. The logic of SSDR is very similar but for the TSDR one needs to use the cumulative distributions rather than the quantile's approach. We focus here only on FSDR and SSDR.

By eq. (4.6) in Chapter 4, F dominates G by FSDR if and only if:

$$\underset{0 \le P < F(r)}{\text{INF}} \frac{Q_G(P)-r}{Q_F(P)-r} \ge \underset{F(r) \le P \le 1}{\text{SUP}} \frac{Q_G(P)-r}{Q_F(P)-r}. \tag{5.7}$$

However, because $(Q_G - r)/(Q_F - r)$ is constant in each step (where a step is defined as a range of P where the quantiles of both F and G are constant), this ratio needs to be computed only once at each step. In the previous section, x and y denoted two random variables. Here x_1 and x_2 denote the two random variables and y_1 and y_2 denote these random variables after subtraction of the riskless interest rate, r. To be more specific, we define $y_{i,j} = x_{i,j} - r$ where $x_{i,j}$ is the j^{th} ranked observation of distribution i (i.e., $x_{i,1}$ is the lowest observation of distribution i) and r is the riskless interest rate. Suppose that we wish to compare two distributions 1 and 2, namely $y_{1,j}$ and $y_{2,j}$ corresponding to $x_{1,j}$ and $x_{2,j}$, where F is the cumulative distribution of x_1 and G of x_2. We carry out the following steps:

Efficiency step:

(i) If there is at least one distribution k such that $y_{k,1} > 0$, then distribution k will dominate all other distributions i with $y_{i,1} < 0$ because $y_{k,1} > 0$ implies that $x_{k,1} > r$ (i.e., the lowest obervation will be larger than r and we will have an arbitrage position; borrowing an infinite amount and investing in k). If there is more than one distribution where the minimum value is greater than r, all of them will be in the efficient set.

(ii) Any distribution i with $y_{i,n} < 0$ (where n corresponds to the largest observation) will be inefficient, because r will dominate the distribution i by FSD. It should be emphasized that the occurrence of conditions (i) and (ii) is very unlikely.

(iii) $x_1 D_1 x_2 \Rightarrow x_1 D_{r1} x_2$.

If none of these efficiency steps holds, we carry out the following steps:

a) Find an index S such that $y_{1,S} \le 0$ and $y_{1,S+1} > 0$.
b) Find an index T such that $y_{2,T} \le 0$ and $y_{2,T+1} > 0$.

If T < S then $x_1 \not{D}_{r1} x_2$ (because it implies that F(r) < G(r) which violates the FSDR necessary condition for dominance of F over G). If T ≥ S, go to step c.

c) Compute $\underset{j \le S}{MIN}(y_{2,j}/y_{1,j}) = \underline{M}$. Compute $\underset{j>S}{MAX}(y_{2,j}/y_{1,j}) = \overline{M}$.

Then $x_1 D_{r1} x_2$ if and only if $\underline{M} \ge \overline{M}$. Note that $\underline{M} \ge \overline{M}$ implies that eq. (5.5) holds which guarantees that F (or x_1) dominates G (or x_2).

b) SSDR Algorithm

Define $Y'_{ij} = \sum_{t=1}^{j} y_{i,t}$ where $y_{i,j}$ is defined as above.

Efficiency Steps

(i) If $y_{k,1} > 0$ for some distribution k, then distribution k will dominate all other distributions i with $y_{i,1} < 0$ (see FSDR explanation).

(ii) Any distribution with $y_{i,n} < 0$ will be inefficient (see FSDR explanation).

(iii) If $Y'_{i,n} \le 0$, then distribution i will be inefficient (because its mean will be lower than r and as r will dominate x_i by SSD, such dominance will exist also with SSDR because SSD ⇒ SSDR).

(iv) $x_1 D_{r1} x_2$, or $x_1 D_2 x_2 \Rightarrow x_1 D_{r2} x_2$.

If the efficiency steps do not hold, we conduct the following steps:

a) Find an index S such that $Y'_{1,S} \le 0$ and $Y'_{1S+1} > 0$.
b) Find an index T such that $Y'_{1,T} \le 0$ and $Y'_{1,T+1} > 0$.

If T < S, then the necessary condition for dominance of F and G (or x_1 over x_2) by SSDR will not hold; hence, $x_1 \not{D}_{r2} x_2$. If T = S, proceed to step c. If T > S, proceed to step d.

c) Compute P_0 and P_1:[12]

[12] Actually we are looking for the values q_0 and q_1 which fulfill the two requirements, $\int_0^{q_0}(Q_F(t)-r)dt = 0$ and $\int_0^{q_1}(Q_G(t)-r)dt = 0$, respectively and a necessary condition for dominance of F over G is that $q_0 \le q_1$.

$$P_0 = -Y'_{1,S} \,/\, Y'_{1,S+1}\;,\; P_1 = Y'_{2,T} \,/\, y_{2T+1}\,.$$

If $P_0 > P_1$, then $x_1 \not{D}_{r2} x_2$. If $P_0 \le P_1$, then the necessary condition for SSDR dominance will hold, and we proceed to step d.

d) Compute $\underline{M}' = \underset{j \le S}{\text{MIN}}\,(Y'_{2,j} \,/\, Y'_{1,j})$ and

$\overline{M}' \equiv \underset{j>S}{\text{MAX}}(Y'_{2,S} \,/\, Y'_{1,S})$. Then, $x_1 D_{r2} x_2$ by SSDR if and only if $\underline{M}'$
$\ge \overline{M}'$ [13]

For discrete distributions with n observations, q_0 and q_1 are given by:

$$Y'_{1,S} + (q_0 - \tfrac{S}{n})\, y_{1,S+1} = 0$$

and:

$$Y'_{2,T} + (q_1 - \tfrac{T}{n})\, y_{2,T+1} = 0$$

Therefore, a necessary condition for dominance of F over G is that,

$$q_0 = \frac{Y'_{1,S}}{y_{1,S+1}} + \frac{S}{n} \le \frac{Y'_{2,T}}{y_{2,T+1}} + \frac{T}{n} = q_1$$

However, as in this step of the algorithm we require that T=S, we obtain $P_0 \gtrless P_1 \Leftrightarrow q_0 \gtrless q_1$.

Hence, we can switch from q_0 and q_1 to P_0 and P_1, respectively.

[13] The SSDR relationship is determined by the ratio $\gamma(p) = \int_0^p (Q_G(t) - r)dt \,/\, \int_0^p (Q_F(t) - r)dt$ in the range $0 \le p < p_0$ and in the range $p_0 < p \le 1$, respectively (see Chapter 4). However, for a step function, it is easy to show that within each step (apart from the step where the denominator of $\gamma(p)$ is equal to zero, i.e., where $p = p_0$) $\gamma(p)$ is continuous and monotonic, hence the extremum points at each step are obtained at the end points. Therefore, it is sufficient to examine $\gamma(p)$ at these end points. At the neighborhood of $P=P_0$, $\gamma(p)$ is not continuous and not monotonic. However, the necessary condition $p_0 < p_1$ guarantees that at this point $\lim_{p \to p_0^-} \gamma(P) = \infty$ and $\lim_{p \to p_0^+} \gamma(p) = -\infty$, hence if there is SSDR this discontinuity point does not violate it, because at this point $\text{INF}(\cdot) > \text{SUP}(\cdot)$. However, if $P_0 > P_1$, $\lim_{p \to p_0^-} \gamma(P) = -\infty$ and $\lim_{p \to p_0^+} \gamma(p) = +\infty$ and therefore $\underset{0 \le P < P_0}{\text{INF}}\gamma(p) < \underset{P_0 < P \le 1}{\text{SUP}}\gamma(p)$ and there is no SSDR dominance. Therefore, it is very important to check whether $P_0 < P_1$ before we proceed with the algorithm.

5.8 SUMMARY

Stochastic dominance rules are derived with very weak set assumptions – a clear benefit in comparison to other decision rules. The cost is that these rules may not be strong enough ending up with a relatively large efficient set. To measure this cost, one needs to conduct empirical studies, which reveal the size of the efficient set in comparison to the feasible set. To conduct such studies, one needs efficient algorithms, to which this chapter is devoted.

We first discuss how one can reduce the number of pairwise comparisons by employing necessary conditions for dominance. Then FSD, SSD and TSD algorithms are provided. While FSD and SSD algorithms are based on jump points of the cumulative distributions stated in terms of either cumulative distributions or their quantiles, the TSD must be stated in terms of cumulative distributions and interior points should also be examined; hence its relative complexity. Most existing published TSD algorithms in the literature are wrong; hence, we provide here the correct TSD algorithm. Finally, by adding the possibility to lend and borrow money at the riskless interest rate, one can drastically reduce the relative size of the efficient set. For this purpose we develop stochastic dominance rules with a riskless asset called SDR rules. We provide in this chapter algorithms for FSDR and SSDR.

Key Terms

Empirical distributions

Ex-post rates of return

The FSD algorithm

The SSD algorithm

The TSD algorithm

Investment horizon

Transitive rules

FSDR algorithm

SSDR algorithm

Efficiency step

STOCHASTIC DOMINANCE WITH SPECIFIC DISTRIBUTIONS

In the derivation of the SD and SDR rules presented in the previous chapters (see Chapters 3 and 4), assumptions on preference, U_i are made but no assumptions are made on the shape of the distributions of rates of return. In that sense, stochastic dominance rules are *distribution-free decision rules*. However, assumptions on the shape of the distributions of rates of return can be added and, in some cases, *parametric investment decision rules* can be derived because the rules will be stated in terms of the distribution's parameters (e.g., mean and variance).

Such parametric rules yield the same partition of the feasible set into the efficient and inefficient sets as the partition obtained by the distribution-free stochastic dominance (SD) rules. Therefore, under the added assumptions regarding the distributions of rates of return, the rules presented in this chapter can safely substitute the SD rules. The advantage of stating the dominance relationship in terms of the distribution's parameters (rather than the cumulative distributions) is that it allows us to reach a number of important results (e.g., optimal diversification among risky assets and equilibrium prices of risky assets – the well known *Capital Asset Pricing Model (CAPM)*). It is important to emphasize that these results cannot be reached without additional assumptions to those assumed so far. In this chapter, we make assumptions on the distributions of returns as well as on preferences, and obtain the well-known Markowitz *Mean-Variance (M-V) Rule*[1] as well as other rules (e.g., the *mean-coefficient of variation rule (M-C)*). These rules, under the added assumption, yield the same results as the SD rules.

6.1 NORMAL DISTRIBUTIONS

a) Properties of the Normal Distribution

Suppose that the rate of return x is normally distributed. Then, the density function f(x) is given by:

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for all } -\infty < x < \infty$$

where μ and σ are the expected rate of return and the standard deviation of x, respectively. The density function f(x) is symmetrical (i.e., all odd moments including the skewness, μ_3, are equal to zero). Normal distributions possess a number of other

[1] Markowitz, H.M., "Portfolio Selection," *Journal of Finance*, 7, March 1952, pp. 77–91, *Portfolio Selection*, New York, Wiley 1959, and *Mean-Variance Analysis in Portfolio Choice and Capital Markets*, Basil Blackwell, New York, 1987.

important properties which will be used in this chapter. Let us review them briefly before using them in the various proofs.

Property a:

Let x_1 and x_2 be two random variables with normal distributions whose cumulative distributions are F and G, respectively. It is given that $x_1 \sim N(\mu_1, \sigma_1)$ and $x_2 \sim N(\mu_2, \sigma_2)$. Each of these two random variables can be standardized as follows:

$$z=\frac{x_1-\mu_1}{\sigma_1} \text{ and } z=\frac{x_2-\mu_2}{\sigma_2}$$

such that the standardized variable z has a mean of $E(z) = 0$ and a standard deviation of $\sigma_z = 1$. The two parameters of z (hence also the density function of z), are the same regardless of the parameters (μ_1, σ_1) or (μ_2, σ_2). Therefore, for any selected value, z_0, the following holds:

$$P(z \le z_0)=P_F\left(\frac{x_1-\mu_1}{\sigma_1} \le z_0\right) = P_G\left(\frac{x_2-\mu_2}{\sigma_2} \le z_0\right)$$
$$= P_F(x_1 \le \mu_1+z_0\sigma_1) = P_G(x_2 \le \mu_2+z_0\sigma_2)$$

where P_F and P_G are probabilities corresponding to distributions F and G, respectively.

Suppose that there is a value x_0 such that:

$$\frac{x_0-\mu_1}{\sigma_1}=\frac{x_0-\mu_2}{\sigma_2} \equiv z_0 \qquad (6.1)$$

Then, for this particular value, $x_0 = \mu_1 + z_0\sigma_1 = \mu_2 + z_0\sigma_2$ and $F(x_0) = G(x_0) = \phi(z_0)$, where $\phi(z)$ denotes the cumulative distribution of z.

Are there any other values like x_0 which fulfill eq. (6.1) and for which F and G reach the same cumulative probability? Because eq. (6.1) is linear in x_0, there is, *at most*, one value x_0. To see this recall that x_0 is given by the equation:

$$\frac{x_0-\mu_1}{\sigma_1}=\frac{x_0-\mu_2}{\sigma_2}.$$

Solve for x_0 to obtain:

$$x_0=\frac{\mu_1\sigma_2-\mu_2\sigma_1}{\sigma_2-\sigma_1}.$$

From this we can conclude that two cumulative normal distributions intersect, at most, once. Moreover, if $\sigma_1 = \sigma_2$, they will never intercept.

Property b:

If $\sigma_2 > \sigma_1$, then g(x) has "thicker tails" than f(x), where f(x) and g(x) stand for the normal density function with σ_1 and σ_2, respectively. This implies that G(x) has a thicker left tail than F(x). To see this, write the ratio:

$$\frac{g(x)}{f(x)} = \frac{\frac{1}{\sqrt{2\pi\sigma_2}} e^{-1/2(\frac{x-\mu_2}{\sigma_2})}}{\frac{1}{\sqrt{2\pi\sigma_1}} e^{-1/2(\frac{x-\mu_1}{\sigma_1})}}$$

or:

$$\frac{g(x)}{f(x)} = \frac{\sigma_1}{\sigma_2} e^{+\frac{1}{2}[\ (\frac{x-\mu_1}{\sigma_1})^2 - (\frac{x-\mu_2}{\sigma_2})^2]}$$

Because $\sigma_2 > \sigma_1$ by assumption, as $x \to \infty$ or $x \to -\infty$, the exponential term approaches infinity and $g(x)/f(x) \to \infty$. Thus, there are values x_1 and x_2 such that for $x < x_1$ and $x > x_2$ we have $g(x)/f(x) > 1$ (namely, g(x) has "thicker" tails).

Property c:

Assume that $\sigma_2 > \sigma_1$; hence, g(x) has thicker tails than f(x). Therefore, F(x) and G(x) intercept exactly once and F(x) intercepts G(x) from below. If $\sigma_1 = \sigma_2$, F(x) and G(x) will not intercept. Thus, F and G intercept once at most, and the one with the lower variance σ intercepts the other distribution "from below".

Property d:

If x is normally distributed, $x \sim N(\mu, \sigma)$, then a linear combination such as $x_\alpha = \alpha x + (1-\alpha) r$ (where r is constant and $\alpha > 0$) will also be normally distributed:

$$x_\alpha \sim N(\alpha\mu + (1-\alpha)r, \alpha\sigma).$$

We will now use these properties, in comparing the SD rules with the M-V rule.

b) Dominance Without a Riskless Asset

We first assume that, in the absence of the riskless asset, investors have to choose between x and y. We will use these well-known properties of normal distributions in the next two theorems.

Theorem 6.1:

Let x and y denote the return on two distinct investments whose cumulative distributions are F and G, respectively. Assume that x and y, the two random variables, are normally distributed with the following parameters:

$$x \sim N(\mu_1, \sigma_1)$$

$$y \sim N(\mu_2, \sigma_2).$$

Then F will dominate G by FSD (FD1G), if and only if the following holds:

a. $\mu_1 > \mu_2$

b. $\sigma_1 = \sigma_2$

Proof:

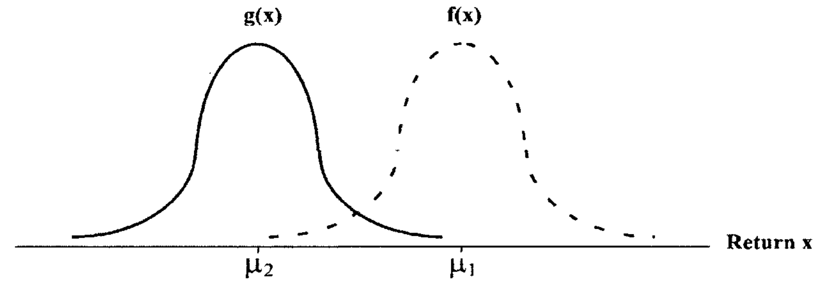
Figure 6.1 plots two density functions of normal random variables and two corresponding cumulative distributions with $\mu_1 > \mu_2$ and $\sigma_1 = \sigma_2$. Because $\sigma_1 = \sigma_2$, F and G do not intercept. The condition $\mu_1 > \mu_2$ implies that $F(x) < G(x)$ for all x; hence, FD_1G. The other side of the proof is very similar: If FD_1G, then F and G will not intercept. Because of the normality assumption, two normal distributions will not intercept if and only if, $\sigma_1 = \sigma_2$; and because $F(x) < G(x)$, $\mu_1 > \mu_2$, which completes the proof.

In the next theorem we show that if normality and risk aversion are assumed, then the mean variance rule will coincide with SSD. Indeed, this result constitutes the justification of the M-V rule and the CAPM, where the latter is based on the M-V rule in the expected utility paradigm.

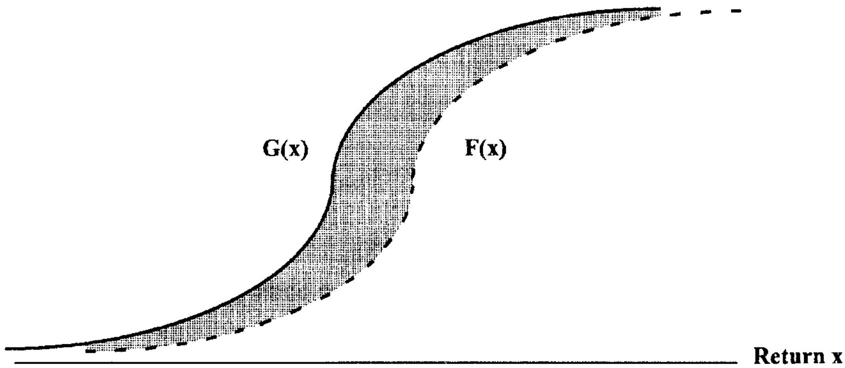
Theorem 6.2:

Denote, as before, by x and y, the return on two investments whose cumulative distributions are F and G, respectively with $x \sim N(\mu_1, \sigma_1)$, $y \sim N(\mu_2, \sigma_2)$. Then, F will dominate G by SSD if and only if F dominates G by the mean-variance rule. Namely:

Fig. 6.1: Two density functions and the corresponding cumulative normal distributions with $\mu_1 > \mu_2$ and $\sigma_1 = \sigma_2$



6.1a: The density functions



6.1b: The cumulative probability functions

a. $\mu_1 \geq \mu_2$

b. $\sigma_1 \leq \sigma_2$

with at least one strong inequality (i.e., M-V dominance) if and only if F dominates G by SSD.

Proof:

We need to analyze several situations. Suppose first that $\mu_1 > \mu_2$ and $\sigma_1 = \sigma_2$. Then F dominates G by M-V rule. But, as we saw in Theorem 6.1, in this case FD_1G and, because FSD implies SSD, also FD_2G.

Now let us turn to the second case: Assume that $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$. In this case, F and G intersect exactly once and F intersects G from below. This is illustrated in Figure 6.2. G does not dominate F by SSD because the "left tail" necessary condition does not hold. Does F dominate G by SSD? Yes, it does. To see this recall that:

$$\mu_1 - \mu_2 = \int_{-\infty}^{+\infty} [G(x) - F(x)]dx \qquad \text{(see Chapter 3, eq. 3.1)}$$

and, because $\mu_1 = \mu_2$, the total positive area in Figure 6.2 is equal to the total negative area. Therefore:

$$\int_{-\infty}^{x} [G(t) - F(t)]dt \geq 0 \quad \text{for all x}$$

(and a strict inequality holds for some x); hence, F dominates G by SSD.

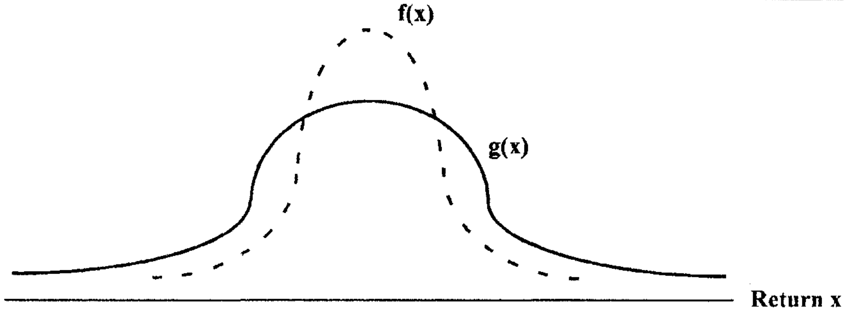
The last scenario occurs when $\mu_1 > \mu_2$ and $\sigma_1 < \sigma_2$. This is similar to the case presented in Figure 6.2 but f(x) and F(x) are shifted to the right relative to the case $\mu_1 = \mu_2$ and $\sigma_2 < \sigma_2$; hence, *a fortiori* FD_2G.

The necessity side of the proof is similar and is obtained by simply reversing the arguments: If FD_2G, then $\mu_1 \geq \mu_2$ because of the necessary condition on the means for dominance; and FD_2G implies that $\sigma_1 \leq \sigma_2$ because of the left tail necessary condition for dominance. Hence, FD_2G implies that F dominates G by the M-V rule.

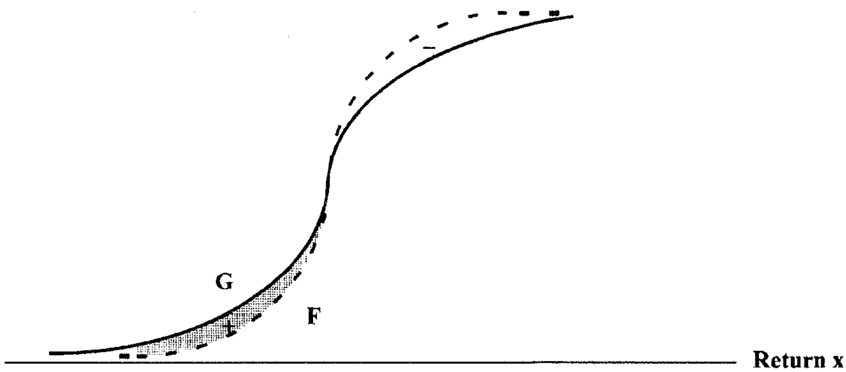
Finally, at least one strict inequality is required (either in a or in b given in Theorem 6.2) otherwise the two normal distributions will be identical and F cannot dominate G. This completes the proof.

Thus, the M-V rule and SSD yield the same partition of the feasible set into the efficient and inefficient sets as long as the returns are normally distributed. Because SSD is an optimal rule for $U \in U_2$, M-V is also an optimal rule for $U \in U_2$ when normality of returns is assumed. Thus, Markowitz's well-known M-V rule coincides with SSD when normal distribution is assumed. This is the main justification for using the M-V investment decision rule in the expected utility paradigm.

Fig. 6.2: F intercepts G from below and the "+" area equal to the "–" area



6.2a: The Density Functions with $\mu_F = \mu_G$ and $\sigma_F < \sigma_G$



6.2b: The Cumulative Distributions F and G with $\mu_F = \mu_G$ and $\sigma_F < \sigma_G$

We will now analyze decision rules for normal distributions when borrowing and lending is allowed. This is the framework for the deviation of Sharpe-Lintner's CAPM.

c) Dominance With a Riskless Asset

Sharpe[2] and Lintner[3] have shown that the asset (or portfolio) that maximizes the slope of the line originating from the riskless interest rate, r, on the vertical axis, also maximizes the investor's expected utility (or reaches the highest indifference curve) as long as risk-aversion is assumed. The line corresponding to the maximum slope is called *Capital Market Line (CML)*. In this section we will

[2] Sharpe, W.F., "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, 19, September 1964, pp. 425–442.

[3] Lintner, J., "Security Prices, Risk and Maximal Gain from Diversification," *Journal of Finance*, Dec. 1965, pp. 587–615.

show that for normal distributions, the portfolio that maximizes the slope of the CML dominates all other portfolios by FSD, as well as by SSD.

Theorem 6.3:

Suppose that the return on the two assets, x and y, are normally distributed: x ~ N (μ_x, σ_x) and y ~ N (μ_y, σ_y) and $\mu_x > r$, $\mu_y > r$, where r stands for the interest rate on the riskless asset. Then x will dominate y by FSDR if and only if:

$$\frac{\mu_x - r}{\sigma_x} > \frac{\mu_y - r}{\sigma_y} \tag{6.2}$$

Note that we have here dominance by FSDR, not only by SSDR.

Proof:

Denote the distribution of x by F, and of y, by G. In Chapter 4, we saw that $\{F_\alpha\}$ D_1 $\{G_\beta\}$ (or F dominates G by FSDR) if and only if there is $\alpha \geq 0$ such that $F_\alpha D_1 G$. Thus, we need to show that inequality (6.2) holds, if and only if such an α exists. We distinguish between two cases:

Case a:

$\mu_x \geq \mu_y$, $\sigma_x \leq \sigma_y$ (with at least one strict inequality).

This case implies that $(\mu_x - r)/\sigma_x > (\mu_y - r)/\sigma_y$ must hold. The mix of x with the riskless asset is given by x_α where $x_\alpha = \alpha x + (1 - \alpha)r$. Let us select a levered strategy, namely $\alpha > 1$. The leverage increases the expected rate of return and standard deviation of the portfolio. We have the following relationship between the standard deviation of x and x_α: $\sigma_{x_\alpha} = \alpha\sigma_x$, and for $\alpha > 1$, the standard deviation increases with leverage. The mean of x_α and x are related as follows:

$$\mu_{x_\alpha} = \alpha\mu_x + (1-\alpha)r$$

$$= \mu_x + (\alpha - 1)\mu_x + (1 - \alpha)r$$

$$= \mu_x + (\alpha - 1)(\mu_x - r).$$

Hence, with $\alpha > 1$, $\mu_{x_\alpha} > \mu_x$ as long as $\mu_x > r$. Thus, leverage ($\alpha > 1$) increases both the mean and the standard deviation. Because, by assumption, $\sigma_x < \sigma_y$, we can choose $\alpha > 1$ such that $\alpha\sigma_x = \sigma_y$. As $\mu_{x_\alpha} > \mu_x > \mu_y$, the levered portfolio, x_α, has, by construction, the same variance as y and a higher mean than y. Therefore, by Theorem 6.1, $F_\alpha D_1 G$, or F dominates G by FSDR, which completes the-proof for case a.

Case b:

Assume that $\mu_x > \mu_y$ and also $\sigma_x > \sigma_y$ (or, alternatively, $\mu_x < \mu_y$ and $\sigma_x < \sigma_y$). Hence, with no riskless asset, there is no dominance by the M-V rule. However, by eq. (6.2):

$$\frac{\mu_x - r}{\sigma_x} > \frac{\mu_y - r}{\sigma_y} \qquad (6.2')$$

We have to show that (6.2′) implies that $\{F_\alpha\}$ D_1 $\{G_\beta\}$, or, that there exists $\alpha > 0$ such that $F_\alpha D_1 G$. Multiply and divide the left-hand side of (6.2′) by α and then add and subtract r in the numerator to obtain:

$$\frac{\mu_x - r}{\sigma_x} = \frac{\alpha(\mu_x - r)}{\alpha\sigma_x} = \frac{\alpha\mu_x + (1-\alpha)r - r}{\alpha\sigma_x} > \frac{\mu_y - r}{\sigma_y} \qquad (6.3)$$

Because of (6.2′), (6.3) holds for any selected α. Choose $0 < \alpha < 1$ such that $\alpha\sigma_x = \sigma_y$ (note that as $\sigma_x > \sigma_y$, by assumption, we must have $\alpha < 1$). Inequality (6.3) plus $\alpha\sigma_x = \sigma_y$ implies that $\alpha\mu_x + (1-\alpha)\, r = \mu_{x_\alpha} > \mu_y$. We then obtain a new distribution $x_\alpha = \alpha x + (1-\alpha)\, r \sim N\,(\alpha\mu_x + (1-\alpha)\, r,\ \alpha\sigma_x)$. ($x_\alpha$ is distributed normally because a combination of a normal random variable and a constant is distributed normally). Because $\mu_{x_\alpha} > \mu_y$, $\sigma_{x_\alpha} = \alpha\sigma_x = \sigma_y$, and F_α is normally distributed, we conclude that F_α dominates G by FSD (see Theorem 6.1). Thus, we have found a value $0 < \alpha < 1$ such that F dominates G by FSD; hence, FDr_1G, which completes the proof.

The necessity side of the proof is straightforward: If F dominates G by FSDR, then there is ($\alpha > 0$) such that $F_\alpha D_1 G$. Because of the normality assumption, for this α we must have $\mu_{x_\alpha} \geq \mu_y$ and $\sigma_{x_\alpha} = \sigma_y$ with at least one strict inequality. This implies that for this α inequality (6.3) holds; hence, eq. (6.2) holds, which completes the necessity side of the proof.

6.2 LOGNORMAL DISTRIBUTIONS

Normal distributions are easy to handle mathematically. However, the lognormal distribution probably has more economic justification. First, stock prices cannot be negative: The normal distribution with a range of $-\infty < x < \infty$ implies that this is possible. Secondly, most distributions of rates of return observed in the market are positively skewed: This contradicts the possibility of rates of return being normally distributed because normal distributions are symmetrical. The existence of positive skewness in the stock market conforms with the assumption that returns are lognormally distributed because the lognormal distribution is positively skewed. Finally, when portfolio revisions are allowed on a continuous basis, the portfolio return (by the Central Limit Theorem) will be lognormally distributed at the end of each finite period.[4] Thus, the lognormal distribution has several important supportive arguments. Hence, it is worthwhile to establish a decision rule for it.

a) Properties of the Lognormal Distribution

Definition of Lognormal Distribution

Let us first define the lognormal distribution and show its relationship to the normal distribution. Suppose x is the return on an investment. We then define y such that y = log x. If y is normally distributed:

$$y \sim N(\mu, \sigma)$$

Where $E(\log x) = \mu$ and $\sigma(\log x) = \sigma$, then x itself will be lognormally distributed:

$$x \sim \Lambda\ (\mu, \sigma)$$

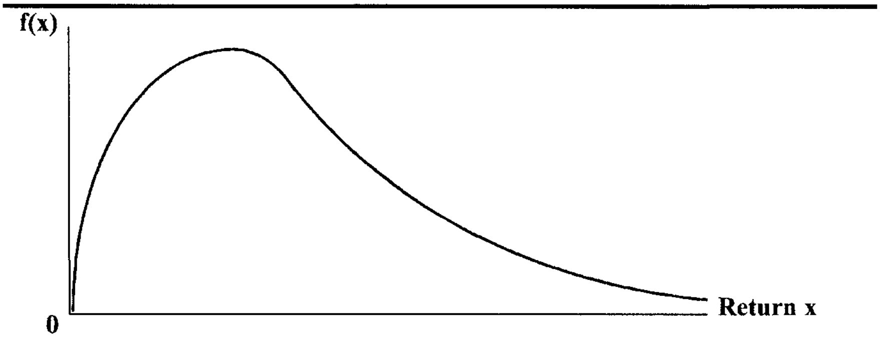
where Λ stands for lognormal distribution. Note that both the normal distribution and the lognormal distribution are fully determined by the same two parameters, μ and σ. However, this does not mean that the two distributions are identical. The density function f(x) of the lognormal random variable is given by:

$$f(x) = \begin{cases} \dfrac{1}{x\sqrt{2\pi\sigma}} e^{-1/2\frac{(\log x-\mu)^2}{\sigma^2}} & \text{if } x>0 \\ 0 & \text{otherwise} \end{cases}$$

The density function is defined only on the positive range (x > 0) and unlike the normal distribution it is positively skewed, as illustrated in Figure 6.3. The first two moments of a lognormal distribution are given by:

[4] Merton, R.C., "An Intertemporal Capital Asset Pricing Model," *Econometrica*, September, 1973.

Figure 6.3: The density function of a lognormal random variable



$$E(x) = e^{\mu + 1/2\sigma^2} \tag{6.4}$$

$$\sigma_x^2 = e^{2\mu+\sigma^2}[e^{\sigma^2} - 1] \tag{6.5}$$

where x is the return as distinct from log x. (For comparison, note that the first two moments of the normal distribution are μ and σ.)

The P^{th} quantile of the lognormal distribution is given by:

$$Q_\wedge (p) = e^{\mu + Q_z(p)\sigma} \tag{6.6}$$

where $Q_z(P)$ is the P^{th} quantile of the normal standardized distribution, and μ and σ are the expected value and the standard deviation of y = log x, respectively.[5]

Two lognormal distributions intersect, at most, once. The intersection point is given by the value P_0 which solves the equation, $Q_{\wedge F}(P_0) = Q_{\wedge G}(P_0)$, or $\mu_F + Q_z(P_0)\,\sigma_F = \mu_G + Q_z(P_0)\,\sigma_G$ (see eq. 6.6). Thus, the intersection point is at point $Q_z(P_0)$ such that:

$$Q_z(P_0) = \frac{\mu_F - \mu_G}{\sigma_G - \sigma_F}.$$

If $\sigma_G = \sigma_F$, there will be no intersection point. If $\sigma_F \neq \sigma_G$, there will be exactly one intersection point of F and G. If $\sigma_F < \sigma_G$, then for relatively low values of P(where

[5] For more details on the lognormal distribution, see Aitchison, J., and J.A.C. Brown, *The Lognormal Distribution*, Cambridge: Cambridge University Press, 1963.

$Q_z(P)$ is negative), $Q_{\Lambda\,F}(P) > Q_{\Lambda\,G}(P)$ and, for relatively high values of P (where $Q_z(P)$ is positive), $Q_{\Lambda\,F}(P) < Q_{\Lambda\,G}(P)$ (see eq. 6.6); hence F will intercept G from below. Similarly, if $\sigma_G < \sigma_F$, G will intercept F from below.

In the analyses below, as with the normal distribution, we distinguish between two cases, one with no riskless asset, and one with a riskless asset.

b) Dominance Without a Riskless Asset

Using the aforementioned properties of lognormal distributions in the next two theorems, we establish the conditions for FSD and SSD for two lognormal distributions.

Theorem 6.4:

Let F and G be two distinct lognormal distributions such that:

$$x_F \sim \Lambda\ (\mu_F, \sigma_F)$$

$$x_G \sim \Lambda\ (\mu_G, \sigma_G)$$

where μ and σ are the mean and standard deviations of log (x), respectively. Then FD_1G if and only if:

a. $\mu_F > \sigma_G$

b. $\sigma_F = \sigma_G$

Proof:

Sufficiency:

Because $\mu_F > \mu_G$, and $\sigma_F = \sigma_G$, we have:

$$Q_{\Lambda\,F}(P) = e^{\mu_F + Q_z(p)\sigma_F} > Q_{\Lambda\,G}(P) = e^{\mu_G + Q_z(p)\sigma_G}$$

for all P; hence, F dominates G by FSD.

Necessity:

Suppose that $\mu_F < \mu_G$ (but $\sigma_F = \sigma_G$). Then for $U_0 = \log x \in U_1$, $E_F U_0(x) = \mu_F < E_G U_0(x) = \mu_G$; hence, F does not dominate G by FSD. Therefore, $\mu_F > \mu_G$ is a necessary condition for FSD.

Now suppose that $\sigma_F \neq \sigma_G$. If $\sigma_F > \sigma_G$ then for $Q_z(P) \to -\infty$, $Q_{\wedge F}(P) < Q_{\wedge G}(P)$, and for $Q_z(P) \to +\infty$, $Q_{\wedge F}(P) > Q_{\wedge G}(P)$; hence, F and G intercept and there is no FSD. If $\sigma_F < \sigma_G$, the same holds but the inequalities are reversed. Thus, if $\sigma_F \neq \sigma_G$, F does not dominate G by FSD. Therefore, $\sigma_F = \sigma_G$ is a necessary condition for FSD dominance. We turn now to SSD dominance.

Theorem 6.5:

Let F and G be two distinct lognormal distributions as in Theorem 6.4. F dominates G by SSD if and only if:

a. $E_F(x) \geq E_G(x)$

b. $\sigma_F(x)/E_F(x) \equiv C_F(x) \leq \sigma_G(x)/E_G(x) \equiv C_G(x)$

with at least one strict inequality. (Note that the conditions for dominance are stated in terms of the parameters of x and not in terms of the parameters of y = log x). Conditions a and b, above, provide the Mean-Coefficient of variation rule (M-C) which is similar to the well-known Mean-Variance rule with one distinction: The coefficient of variation, C, substitutes for the standard deviation as a measure for risk.

Proof:

First note that $C^2 = (\sigma/E)^2$. By eq. 6.4 and eq. 6.5 we have:

$$C^2 = e^{2\mu+\sigma^2}[e^{\sigma^2}-1]/(e^{\mu+1/2\sigma^2})^2 = (e^{\sigma^2}-1)$$

Hence, $C_F(x) \leq C_G(x)$ is the same condition as $\sigma_F \leq \sigma_G$.

Hence, condition b of Theorem 6.5 is identical to the condition $\sigma_F \leq \sigma_G$. We use this result in the proof below. We also use the property that two lognormal cumulative distributions intercept once, at most, and the one with the lower σ intercepts the other "from below". The quantiles of the two distributions are given by:

$$Q_{\wedge F}(P) = e^{\mu_F + Q_z(P)\sigma_F}$$

$$Q_{\wedge G}(P) = e^{\mu_G + Q_z(P)\sigma_G}$$

Let us first discuss the case where $C_F = C_G$ (or $\sigma_F = \sigma_G$) and $E_F(x) > E_G(x)$. By eq. 6.4, these conditions imply that $\mu_F > \mu_G$. However, if $\mu_F > \mu_G$ and $\sigma_F = \sigma_G$ (or $C_F = C_G$), then FD_1G (see Theorem 6.4), which implies FD_2G.

Now let us turn to the case where $C_F < C_G$ and $E_F(x) \geq E_G(x)$. Because $C_F < C_G$, also $\sigma_F < \sigma_G$; hence, F and G intercept once and F intercepts G from below. Also:

$$E_F(x) - E_G(x) = \int_0^\infty [G(x) - F(x)]dx \geq 0 \text{ (see Chapter 3, eq: 3.1)}$$

(Note that as x is lognormally distributed, $F(x) = G(x) = 0$ for $x < 0$).

The integral is non-negative due to the assumption, $E_F(x) \geq E_G(x)$. Because F crosses G from below, we can safely state that $\int_0^x [G(x) - F(x)]dx \geq 0$ for all x (and there is at least one strict inequity); hence, FD_2G, which completes the sufficiency side of the proof. The necessity side of the proof is straightforward: If $C_F > C_G$, also $\sigma_F > \sigma_G$ and, therefore, G intercepts F from below; hence, F has a "thicker" left tail and cannot dominate G. Similarly, if $E_F(x) < E_G(x)$, F cannot dominate G by SSD because $E_F(x) \geq E_G(x)$ is a necessary condition for dominance of F over G. Hence, conditions a and b of Theorem 6.5 are necessary and sufficient conditions for SSD dominance of F over G.

c) Dominance With a Riskless Asset

If $y = \log x$ is normally distributed, then, by definition, x will be lognormally distributed, namely, $x \sim \Lambda(\mu, \sigma)$.

When borrowing and lending is allowed, then the random variable is:

$$x_\alpha = \alpha x + (1 - \alpha) r$$

where x stands for the return of the risky asset.

First note that if we define $y_\alpha = \log(\alpha x)$ $(\alpha > 0)$, then, if $y = \log x$ is normally distributed, $y_\alpha = \log \alpha + \log x$ will also be normally distributed with a mean of $\log \alpha + \mu$ and standard deviation of σ. Thus, αx will be lognormally distributed with these two parameters.

However, this is not the case for x_α because of the additional constant $(1-\alpha)r$. Let us elaborate: The random variable x_α is similar to the random variable x but it is shifted by the constant $(1-\alpha)r$. This shift adds one more parameter to the distribution function and, therefore, x_α is distributed lognormally with three parameters: the constant $(1-\alpha)r$ and the mean and standard deviation of $y_\alpha = \log(\alpha x)$. However, because $E(y_\alpha) = \log \alpha + E \log x = \log \alpha + \mu$, and $\sigma(y_\alpha) = \sigma(\log x) = \sigma$, we have that the three parameters are $(1-\alpha)r$, $\log \alpha + \mu$, and σ. Therefore, we denote this three-parameter distribution as follows:[6]

$$x_\alpha \sim \Lambda((1-\alpha)r, \log \alpha + \mu, \sigma).$$

Thus, in order to find conditions for dominance of F over G by FSDR or by SSDR, we need to compare three-parameter lognormal distributions. Because the proofs of FSDR and SSDR are rather long, we give below two conditions (Theorem 6.6 and Theorem 6.7) for dominance without a proof.

Theorem 6.6:

Let F and G be the cumulative distributions of two options with two-parameter lognormal distributions given by:

$$x_F \sim \Lambda\ (\mu_F, \sigma_F),\ x_G \sim \Lambda\ (\mu_G, \sigma_G)$$

where r is the riskless asset interest rate. Then the necessary and sufficient conditions for $\{F_\alpha\}\ D_1\ \{G_\beta\}$ (or F dominates G by FSDR) are:

I. $F(r) < G(r)$

II. $\sigma_F \geq \sigma_G$.

It can be shown that condition (I) is equivalent to the following condition:

$$\frac{\mu_F - \log r}{\sigma_F} > \frac{\mu_G - \log r}{\sigma_G}$$

For proof, see Levy and Kroll (1976)[7]

Theorem 6.7:

Let F, G and r be as in Theorem 6.6. Then a necessary and sufficient condition for dominance of F over G by SSDR is that *either* one of the following conditions holds:[8]

---

[6] See footnote 5.

[7] Levy H., and Kroll, Y., "Stochastic Dominance with Riskless Assets," *Journal of Financial and Quantitative Analyses*, 11, December 1976, pp. 743–773.

[8] The proof of this theorem is very long and cumbersome; hence, for the sake of brevity, it is not provided in the book. It appears in Kroll, Y., "*Preferences Among Combinations of Risky Assets and a Riskless Asset: Criteria and Implication*," Ph.D. dissertation, Hebrew University, Israel, 1977.

I. F dominates G by SSD

II. $\sigma_F \geq \sigma_G$ and $P_0 \leq P_1$

where P_0 and P_1 are given by the following equations:

$$\int_0^{P_0} e^{\mu_F + z_n(P)\sigma_F} dP = rP_0$$

$$\int_0^{P_1} e^{\mu_G + z_n(P)\sigma_G} dP = rP_1$$

6.3 TRUNCATED NORMAL DISTRIBUTIONS

The normal distribution has the following two drawbacks when applied to a choice among risky assets:

1. Its range is $-\infty < x < \infty$. This is inconsistent with the fact that the price of a risky asset (stock or bond) cannot be negative. Also, such a wide price range (range of x) is also unrealistic. Thus, stock returns will not conform to the normal distribution precisely.

2. Suppose that $x \sim N(\mu_x, \sigma_x)$ and $y \sim N(\mu_y, \sigma_y)$. Furthermore, assume that we have the following parameter: μ_x is very large, say, $\mu_x = 10^6$, $\mu_y = 1$, $\sigma_x = 1.01$ and $\sigma_y = 1$. Then there will be no dominance between x and y (because x has a thicker left tail). However, probably risk averters and risk seekers alike would choose x due to its very large mean relative to the mean of y. Thus, the M-V rule fails to show the intuitive preference of x over y.

In order to overcome these two difficulties of the normal distribution, it is suggested that the normal distribution be truncated so that the returns will be bounded between two values, say $L < x < M$. For example, if $x \sim N(10\%, 20\%)$ (i.e., $\mu = 10\%$ and $\sigma = 20\%$), we may want to truncate the normal distribution from below such that the density function for any value x, $x \leq \mu - 5.5\sigma = 10\% - 5.5 \cdot 20\% = -100\%$ is zero. By doing so, we disallow negative prices; the rate of return cannot be lower than -100% (i.e., the stock price drops to zero). We turn now to the decision rule for the choice among truncated normal distributions.

a) Symmetrical truncation

Suppose that there are two normal distributions with cumulative distributions F and G, and density functions f(x) and g(x), respectively. To obtain the truncated normal distributions with symmetrical truncation, the density functions f(x) and g(x) are defined as zero for all values that deviate more than some fixed number of standard deviations from the mean.

Moreover, assume that the truncation points are determined such that $\alpha/2$ of the area under f and g is located in each tail of the distribution. Then, to obtain the truncated normal distribution, an area of α is shifted from the tails to the center of the distribution; hence, the total area under the truncated density function remains 1. Under such truncation we have:

$$\Phi(\frac{A_1-\mu_1}{\sigma_1}) = \Phi(\frac{A_2-\mu_2}{\sigma_2}) = \alpha/2$$

$$1-\Phi(\frac{B_1-\mu_1}{\sigma_1}) = 1-\Phi(\frac{B_2-\mu_2}{\sigma_2}) = \alpha/2$$

where Φ stands for the cumulative area under the standardized normal distribution, A_i (i=1, 2) are the lower truncation points of the two distributions, f(x) and g(x), respectively, and B_i(i=1, 2) are the upper truncation points. Suppose that the area of $\alpha/2$ corresponds to δ standard deviations from the mean up to each truncation point. Then we have the following relationships:

$$A_1 = \mu_1 - \delta\sigma_1 \text{ and } A_2 = \mu_2 - \delta\sigma_2$$

$$B_1 = \mu_1 - \delta\sigma_1 \text{ and } B_2 = \mu_2 - \delta\sigma_2$$

Recall that with a normal distribution, we need to deviate the same number of standard deviations from the mean corresponding to a probability of $\alpha/2$ in the tail, regardless of μ and σ.

The truncated cumulative distributions $F^*(x)$ and $G^*(x)$ as a function of the non-truncated normal distributions F(x) and G(x) are as follows:[9]

$$F^*(x) = \begin{cases} 0 & x<A_1 \\ \frac{F(x)-\alpha/2}{1-\alpha}) & A_1 \le x \le B_1 \\ 1 & x > B_1 \end{cases}$$

[9] For the density function and other properties of truncated normal distribution, see Johnson, N., and S. Kotz *Continuous Univariate Distributions*, Boston: Houghton Mifflin, 1970.

$$G^*(x) = \begin{cases} 0 & x<A_2 \\ \dfrac{G(x)-\alpha/2}{1-\alpha} & A_2 \le x \le B_2 \\ 1 & x > B_2 \end{cases}$$

Note that $F(B_1) = 1-\alpha/2$; hence, $F^*(B_1)=\dfrac{1-\alpha/2-\alpha/2}{1-\alpha} = 0$. Also, $F(A_1) = \alpha/2$;

hence, $$F^*(A_1)=\frac{\alpha/2-\alpha/2}{1-\alpha}=0$$

The same holds for $G^*(x)$; namely $G^*(A_2) = 0$ and $G^*(B_2) = 1$.

In the following theorems, we establish conditions for dominance for truncated normal distributions.

Theorem 6.8:

Let F and G be the cumulative distributions of two normal distributions with parameters (μ_1, σ_1) and (μ_2, σ_2), respectively, and F^* and G^* are the corresponding truncated cumulative probability distributions. Let δ be defined as above, namely, a deviation of δ standard deviations from the mean corresponding to areas of $\alpha/2$ in the tail of the normal distribution. Then:

a) If $\mu_1 > \mu_2$, $\sigma_1 > \sigma_2$, then F^* will dominate G^* by FSD if and only if:

$$(\mu_1 - \mu_2) / (\sigma_1 - \sigma_2) > \delta$$

b) If $\mu_1 > \mu_2$, $\sigma_1 < \sigma_2$, then F^* will dominate G^* by FSD if and only if:

$(\mu_1 - \mu_2) / (\sigma_2 - \sigma_1) > \delta$

where F^* and G^* are the truncated distributions of F and G, respectively.

Proof:

Case (a):

This is the most interesting case because the M-V rule fails to reveal a preference, whereas with the truncation, there is FSD. First note that $B_1 = \mu_1 + \delta\sigma_1$ and $B_2 = \mu_2 + \delta\sigma_2$ and, because $\mu_1 > \mu_2$ and $\sigma_1 > \sigma_2$, always $B_1 > B_2$; hence, for a relatively large value of x, F is below G. Let us investigate the lower bounds. First note that F and G intersect once and, because $\sigma_1 > \sigma_2$, F has a thicker left tail. The intersection of the two normal distributions, F and G, at x_0 is given by:

$$\frac{x_0 - \mu_1}{\sigma_1} = \frac{x_0 - \mu_2}{\sigma_2}$$

or:

$$x_0 = \frac{\mu_1\sigma_2 - \mu_2\sigma_1}{\sigma_2 - \sigma}.$$

The lower bound is given by:

$A_1 = \mu_1 - \delta\sigma_1$, $A_2 = \mu_2 - \delta\sigma_2$; hence, $A_1 \gtrless A_2$ if and only if, $(\mu_1 - \mu_2) / (\sigma_1 - \sigma_2) \gtrless \delta$. This case is illustrated in Figure 6.4a. Note that because F intersects G from below, the condition $A_1 > A_2$ implies that $x_0 < A_2 < A_1$. Figure 6.4b illustrates a case where the condition of the theorem does not hold; hence $A_1 > A_2 > x_0$.

Let us first show that if $A_1 > A_2$ (i.e., the condition of the theorem holds) also, $x_0 < A_2$, as shown in Figure 6.4a. Recall that:

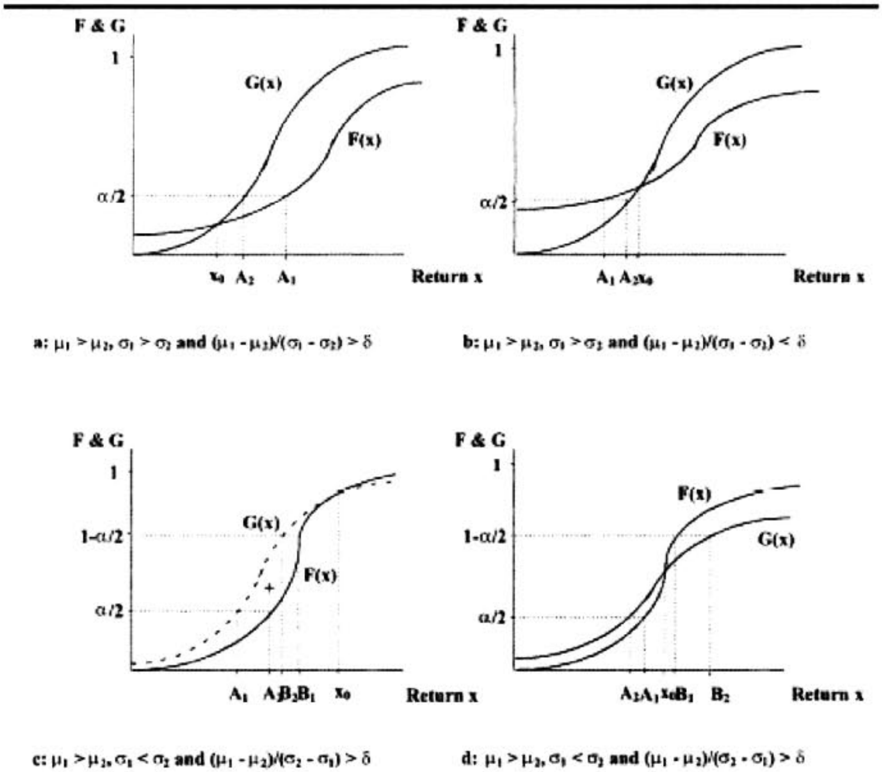
$$\frac{A_1 - \mu_1}{\sigma_1} = \frac{A_2 - \mu_2}{\sigma_2} = -\delta$$

As $A_1 > A_2$, we have:

$$\frac{A_2 - \mu_1}{\sigma_1} < \frac{A_2 - \mu_2}{\sigma_2}$$

But from this we can conclude that $A_2 > \frac{\mu_1\sigma_2 - \mu_2\sigma_1}{\sigma_2 - \sigma_1} = x_0$ (note that the equality is reversed because we divide by a negative number: $\sigma_2 - \sigma_1$). Hence, $A_2 < A_1$ implies that $A_1 > A_2 > x_0$ as drawn in Figure 6.4a. Thus, if the condition of the theorem holds, $A_1 > A_2 > x_0$ and $B_1 > B_2$, and we have to show is that F^* dominates G^* by FSD. To prove the sufficiency side of the theorem, note that if the condition of the theorem holds, then $A_1 > A_2$. For all values $x < A_2$, we have $G^*(x) = 0$ and $F^*(x) = 0$. For values $A_2 < x < A_1$, we have $G^*(x) > 0$ and $F^*(x) = 0$. For $A_1 < x < B_2$, we have:

Figure 6.4: Various normal distributions and their truncation points



$$F(x)<G(x) \Rightarrow \frac{F(x)-\alpha/2}{1-\alpha} < \frac{G(x)-\alpha/2}{1-\alpha} \Rightarrow F^*(x) \leq G^*(x).$$

Finally, for the range $B_2 < x < B_1$, $G^*(x) = 1$ and $F^*(x) < 1$. Therefore $F^*(x) < G^*(x)$. Thus, $F^*(x) \leq G^*(x)$ in the whole range (and a strict inequality holds at the same range), namely FD_1G, which completes the sufficiency side of the proof.

The necessity side follows: If the condition of the theorem does not hold, then $A_1 < A_2 < x_0$, as described in Figure 6.4b. Hence, in the range $A_1 < x < A_2$, $F^*(x) = \dfrac{F^*(x)-\alpha/2}{1-\alpha} > 0$ and $G^*(x) = 0$; namely, $F^*(x) > G^*(x)$. Therefore, F^* does not dominate G^* by FSD.

Case (b):

In this case it is assumed that F dominates G by the M-V rule. Let us show that with the truncation, there is preference of F over G by FSD: The condition $\mu_1 > \mu_2$, $\sigma_1 < \sigma_2$ implies that $A_1 = \mu_1 - \delta\sigma_1 > A_2 = \mu_2 - \delta\sigma_2$.

However, for the upper bound, we have $B_1 = \mu_1 + \delta\sigma_1$ and $B_2 = \mu_2 + \delta\sigma_2$, and we have:

$$B_1 \gtrless B_2 \text{ if, } (\mu_1 - \mu_2)/(\sigma_2 - \sigma_1) \gtrless \delta.$$

Figure 6.4c corresponds to this case. First note that because $\sigma_1 < \sigma_2$ (by assumption), F intersects G from below. If the condition of the theorem holds, $B_1 > B_2$. But because $\frac{B_1 - \mu_1}{\sigma_1} = \frac{B_2 - \mu_2}{\sigma_2}$, we have $\frac{B_2 - \mu_1}{\sigma_1} < \frac{B_2 - \mu_2}{\sigma_2}$ and, therefore, $B_2 > \frac{\mu_1\sigma_2 - \mu_2\sigma_1}{\sigma_1 - \sigma_2} = x_0$.

Thus, we conclude that $B_1 > B_2 > x_0$. Because also $A_1 > A_2$ (see above) $F^*(x) \leq G^*(x)$ in the whole range (and there is at least one strict inequality), which completes the sufficiency side of the proof. The necessity side of the theorem, once again, follows: If the condition of the theorem does not hold, then $x_0 < B_1 < B_2$ and, in the range $x > x_0$, $F^*(x) \geq G^{x_0}$ (with some strict inequality) and $F^*(x)$, does not dominate $G^*(x)$ by FSD (see Figure 6.4d).

Discussion:

With truncated normal distributions, some paradoxes of the M-V rule are solved. For example, assume that $\mu_1 = 10^6$, $\mu_2 = 1$, $\sigma_1 = 1.01$, $\sigma_2 = 1$ (see our previous example). The M-V rule cannot distinguish between these two options. However, with truncated distributions, because μ_1 is much larger then μ_2, we have $\frac{\mu_1 - \mu_2}{\sigma_1 - \sigma_2} > \delta$ for any reasonably selected δ, which implies that investment 1 dominates investment 2 by FSD, which conforms with intuition. The M-V rule does not distinguish between these two distributions but once a truncation is assumed, the dominance is revealed.

Baumol[10] also claims that if $\mu_1 > \mu_2$ and $\sigma_1 < \sigma_2$, investment 1 will dominate investment 2 by the M-V rule, which conforms with intuition. However, if $\mu_1 > \mu_2$ and $\sigma_1 > \sigma_2$, even though the M-V rule cannot distinguish between the two distributions, in many cases there will be an obvious dominance of investment 1 over investment 2. To overcome this difficulty, Baumol suggests an investment criterion asserting that if $\mu_1 > \mu_2$ and $\sigma_1 > \sigma_2$,

[10] See Baumol, W.J., "An Expected Gain Confidence Limit Criterion for Portfolio Selection," *Management Science*, October, 10, 1963, pp. 174–182.

investment 1 will dominate investment 2 if and only if $\mu_1 > \mu_2$ and, in addition, the following inequality holds: $\mu_1 - k\sigma_1 \geq \mu_2 - k\sigma_2$ where k is a positive number (see Chapter 1, eq. (1.5)).

Baumol's rule (which is not based on expected utility), coincides with FSD if the distributions are truncated normal distribution. To see this, like Baumol, assume that $\mu_1 > \mu_2$ and $\sigma_1 > \sigma_2$. Rewrite Baumol's second condition as $(\mu_1 - \mu_2)/(\sigma_1 - \sigma_2) \geq k$. Thus, for $\delta \equiv k$, preference of investment 1 or 2 by Baumol's criterion will coincide with the preference by FSD, hence conforming with intuition as well as with expected utility paradigm. This is not a surprising result because both Baumol's rule and the rule suggested in the above two theorems disregard the tails of the distributions.

b) Non-Symmetrical Truncation

Suppose that area α_1 of the left tail and area α_2 of the right tail of the normal distribution are moved to the center. This means that the truncation points are δ_1 and δ_2 standard deviations from the mean, respectively. Then:

$$\Phi\left(\frac{A_1 - \mu_1}{\sigma_1}\right) = \Phi\left(\frac{A_2 - \mu_2}{\sigma_2}\right) = \alpha_1$$

$$1 - \Phi\left(\frac{B_1 - \mu_1}{\sigma_1}\right) = 1 - \Phi\left(\frac{B_2 - \mu_2}{\sigma_2}\right) = \alpha_2$$

and the A_1, A_2, B_1, and B_2 are the bounds of the two distributions as defined in section a. With such non-symmetrical truncation, we have:

$$F^*(x) = \begin{cases} 0 & x < A_1 \\ \dfrac{F(x) - \alpha_1}{1 - \alpha_1 - \alpha_2} & A_1 \leq x \leq B_1 \\ 1 & x > B_1 \end{cases}$$

and:

$$G^*(x) = \begin{cases} 0 & x < A_2 \\ \dfrac{G(x) - \alpha_1}{1 - \alpha_1 - \alpha_2} & A_2 \leq x \leq B_2 \\ 1 & x > B_2 \end{cases}$$

If α_1 corresponds to δ_1 standard deviations from the mean, and α_2 corresponds to δ_2 standard deviations from the mean, we have the following relationships between the bounds and δ_i:

$$A_1 = \mu_1 - \delta_1\sigma_1,\ A_2 = \mu_2 - \delta_1\sigma_2$$

$$B_1 = \mu_1 + \delta_2\sigma_1,\ B_2 = \mu_2 + \delta_2\sigma_2$$

The dominance rule for non-symmetrical truncation is formulated in the next theorem. The proofs are very similar to the proof corresponding to symmetrical truncation; hence, it will be given here in brief.

Theorem 6.9:

Let F and G be two normal distributions with parameters (μ_1, σ_1) and (μ_2, σ_2), respectively, and F^* and G^* are the corresponding truncated normal distributions, with truncation determined by δ_1 and δ_2, as defined above. Then:

Case (a):

If $\mu_1 > \mu_2$, $\sigma_1 > \sigma_2$, then F^* will dominate G^* by FSD if and only if $(\mu_1 - \mu_2)/\ (\sigma_1 - \sigma_2) > \delta_1$.

Case (b):

If $\mu_1 > \mu_2$ and $\sigma_1 < \sigma_2$, then F^* will dominate G^* by FSD if and only if $(\mu_1 - \mu_2)/(\sigma_2 - \sigma_1) > \delta_2$.

Proof:

Case (a):

G intercepts F from "below" because $\sigma_1 > \sigma_2$. Like in the analysis of symmetrical truncation, $A_1 > A_2 > x_0$ (where x_0 is the intersection point of F and G) only if $A_1 = \mu_1 - \delta_1\sigma_1 > A_2 = \mu_2 - \delta_1\sigma_2$ or $(\mu_1 - \mu_2)/(\sigma_1 - \sigma_2) > \delta_1$. $B_1 > B_2$ because $\mu_1 + \delta_2\ \sigma_1 > \mu_2 + \delta\sigma_2$. Thus, as in the case of symmetrical truncation, if $(\mu_1 - \mu_2)/(\sigma_1 - \sigma_2) > \delta_1$, we have FSD dominance of F^* over G^*.

Case (b):

Here F intercepts G from below. $A_1 = \mu_1 - \delta_1\sigma_1 > A_2 = \mu_2 - \delta_1\sigma_2$ because $\mu_1 > \mu_2$ and $\sigma_1 < \sigma_2$. However, for FSD dominance, we need to have $B_1 = \mu_1 + \delta_2\sigma_1 > B_2 = \mu_2 + \delta_2\sigma_2 > x_0$. However, this occurs only if $(\mu_1 - \mu_2)/(\sigma_2 - \sigma_2) > \delta_2$, which completes the proof.

6.4 DISTRIBUTIONS THAT INTERCEPT ONCE

Many distributions intercept at *most* once (e.g., normal distributions, lognormal distributions, and uniform distributions). Making assumptions on the shape of the distribution, we can find a criterion for FSD or SSD stated in terms of the distribution parameters. However, in the case of one intersection at most of F and G, the general principle for SSD dominance is as follows:

a. Ascertain whether there is FSD. All investments dominated by FSD should be eliminated.

b. All remaining cumulative distributions intercept exactly once. For these distributions, the criterion for dominance of F over G by SSD is:

   1. F must cross G from below, and

   2. $E_F(x) \geq E_G(x)$.

Because two stages are involved, the above procedure is called a *two-stage criterion*. (Actually, this criterion is equivalent to the requirement that conditions 1 and 2 hold if F and G intercept and 2 holds if they do not intercept).

Let us assume first that one intersection occurs, and F intersects G from below because:

$$E_F(x) - E_G(x) = \int_{-\infty}^{\infty} [G(x) - F(x)]dx.$$

Then, if $E_F(x) \geq E_G(x)$, it implies that $\int_{-\infty}^{\infty} [G(x) - F(x)]dt \geq 0$ up to any value x; hence, FD_2G.

We illustrate the two-stage criterion with uniform distributions with a density function f(x) given by:

$$f(x) = \begin{cases} 0 & x < \alpha \\ 1/(\beta - \alpha) & \alpha \le x \le \beta \\ 0 & x < \beta \end{cases}$$

where (α, β) are the parameters of the uniform distribution.

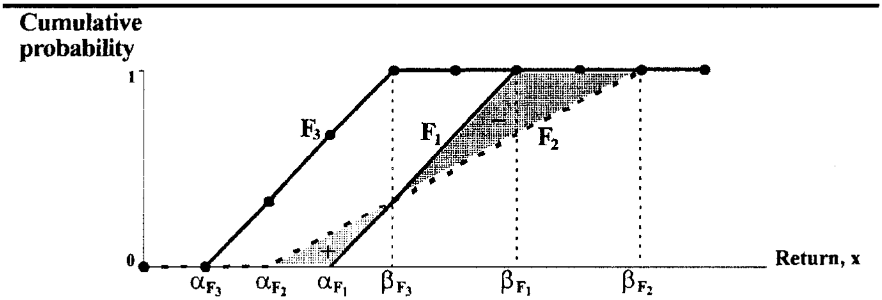
Figure 6.5 illustrates three cumulative uniform distributions. At stage a, F_3 is eliminated because $F_1D_1F_3$. F_1 intercepts F_2 from below. If the "+" area is larger or equal to the "–" area, then:

$$\int_{-\infty}^{x} [F_2(t) - F_1(t)]dt \ge 0 \text{ for all x; hence } F_1D_2F_2.$$

In our specific example (see Figure 6.5) the "+" area is smaller than the "–" area; hence, both F_1 and F_2 are located in the SSD efficient set.

The two conditions, (a) F intercepting G from below, and (b) $E_F(x) \ge E_G(x)$, can sometimes be stated in terms of the distribution parameters. For example, in the normal distribution case, for SSD of F over G, we require that $\sigma_F < \sigma_G$ (i.e., F intercepts from below), and $E_F(x) \ge E_G(x)$ which guarantees that the "+" area will be greater than the "–" area. With a lognormal distribution for SSD of F over G, we require, once again, that $E_F(x) \ge E_G(x)$ and, in addition, that $\sigma_F < \sigma_G$. However, the condition $\sigma_F < \sigma_G$ also implies that F intercepts G from below. In the uniform case, F dominates G by SSD if $\alpha_F > \alpha_G$. (F intercepts G from below and $(\alpha_F + \beta_F)/2 \ge (\alpha_G + \beta_G)/2$, or $E_F(x) \ge E_G(x)$.) Thus, the conditions $E_F(x) \ge E_G(x)$ and F intercepting G from below, in all previous proofs are kept, but we establish these principles in terms of the distribution parameters (i.e., μ, σ) or even the coefficient of variations.

Figure 6.5: Three uniform distributions



6.5 SUMMARY

In this chapter, we analyze stochastic dominance rules with some additional information (or assumptions) regarding the distribution of the rates of return. We find that the SSD rule coincides with Markowitz's well known M-V rule when the distributions of rates of return are normal. Thus, for normal distributions, the M-V rule conforms with the expected utility paradigm. When borrowing and lending is allowed, the condition for dominance of x over y by FSDR, as well as SSDR is $(\mu_x - r)/\sigma_x > (\mu_y - r)/\sigma_y$. However, as the market portfolio provides the highest slope in the $\mu - \sigma$ space, it dominates all other portfolios by FSDR. Thus, for normal distributions, the CAPM, which relies on the selection of the market portfolio by all investors, conforms with the expected utility paradigm.

In practice, the possibility that rates of return are normally distributed is questionable because prices of risky assets cannot be negative. The lognormal distribution is confined to positive values only; hence, it seems to be more suitable for investment choices. For lognormal distributions the mean-coefficient of variation rule (M-C) coincides with SSD; hence, the coefficient of variation, σ/E, is the appropriate measure of risk.

Another way to overcome the difficulty of non-negative prices is to truncate the normal distributions. We found conditions for FSD dominance of one truncated distribution over another, thereby revealing obvious dominance undetectable by the M-V criterion.

Key Terms

Distribution-free decision rules

Parametric investment decision rules

Mean-coefficient of variation rule (M-C)

Normal distributions

Capital Market line (CML)

Lognormal distributions

Truncated normal distributions

Baumol Rule

Two-stage criterion

M-V rule

CAPM

7

THE EMPIRICAL STUDIES

In judging the quality of an investment decision making rule, two factors have to be taken into account: a) its underlying assumptions; b) its *effectiveness* in terms of the relative size of the resultant efficient set. Based solely on the first factor, the FSD is the best rule because the only assumption needed for its derivation is that $U \in \mathbf{U}_1$ or $U' \geq 0$. However, the FSD rule is likely to be ineffective in that the resultant efficient set may not be much smaller than the feasible set. Generally, the larger the number of assumptions (e.g., risk aversion, decreasing absolute risk aversion), the smaller the induced efficient set.

Figure 3.24 in Chapter 3 illustrates the relationship between the various stochastic dominance rules and the relative size of their corresponding efficient sets. The assumption of the FSD rule that $U'>0$ produces the largest efficient set. If, in addition, $U'' < 0$ (the SSD rule) is also assumed, the corresponding efficient set will be smaller, and if $U'''>0$ (the TSD rule) is also assumed, the efficient set obtained will be even smaller. In addition, because FSD $\Rightarrow$ SSD $\Rightarrow$ TSD, we have that the SSD efficient set is not only smaller than the FSD efficient set, but also that it forms a subset of the FSD efficient set. Similarly, the TSD efficient set constitutes a subset of the SSD efficient set. By adding the riskless asset, we obtain the FSDR, SSDR and TSDR rules where the TSDR efficient set is a subset of the TSD, the SSDR efficient set is a subset of the SSD efficient set, and the FSDR efficient set is a subset of the FSD efficient set. Also, we have FSDR $\Rightarrow$ SSDR $\Rightarrow$ TSDR; hence relations among the efficient sets with the SDR decision rules are similar to the relations of SD rules. In this Chapter we apply the various investment decision rules discussed in the previous Chapter to a series of empirical studies. So far we have discussed SD criteria for pair-wise comparisons. In this chapter we introduce the concept of convex stochastic dominance (CSD) corresponding to a comparison of 3 or more distributions simultaneously, which allows a further reduction in the various efficient sets with no additional assumption made neither on preferences nor on the distributions of returns.

Empirical studies of the SD decision rules generally focus on three main issues: the effectiveness of the various SD rules, the performance of mutual funds relative to an unmanaged portfolio (e.g., the Standard and Poor's index), and inherent statistical errors. The first two topics; effectiveness and performance, although two separate issues, are generally reported jointly; hence we shall do likewise. In view of the abundance of empirical studies, our presentation will be limited to those that, in our view, best exemplify the aforementioned three main issues.

Finally, we have shown in Chapter 5 that the existing TSD (and TSDR) algorithms are wrong. Yet we report here the efficient sets as published in the literature. Thus, we have to remember that the correct TSD and TSDR efficient sets may be different than what is reported here. In particular, the TSD and TSDR efficient sets reported by Porter (see below) are two small and using the correct algorithm may increase the efficient set corresponding to TSD and TSDR.

7.1 THE EFFECTIVENESS OF THE VARIOUS DECISION RULES: A PERFECT MARKET

The first empirical study of SD was conducted by Levy and Hanoch (L&H) in 1970[1], just one year after the publication of the stochastic dominance rule by Hadar & Russell [2] and Hanoch & Levy. [3] L&H developed FSD and SSD algorithms and applied them as well as the M-V rule to quarterly rates of return over the period 1958–1968 in the Israeli stock market. Due to the relatively small number of shares traded in the Israeli stock market in 1958, subperiods 1962–1968 and 1965–1968 were also examined. The shorter the period under consideration, the larger the number of available shares (the feasible set) but the smaller the number of quarterly rates of return per share. Table 7.1 summarizes the results of this empirical study.

Table 7.1. The Efficient Sets Obtained on Israeli Stock Market Shares

Period	Number of shares included (the feasible set)	No. of observations	Size of the efficient set		
			FSD	SSD	M-V
1958–1968	16	41	15	4	4
1962–1968	37	25	32	10	6
1965–1968	138	13	66	7	9

Source: Levy & Hanoch, 1970.

The FSD rule, which is virtually assumption-free, is seen to be very ineffective, especially in the first two periods where the number of observations is relatively large. As will be shown later on, this result characterizes other empirical studies, too. The SSD rule is very effective, hence indicating that the addition of the assumption of risk aversion substantially reduces the size of the efficient set. The mean-variance (M-V) efficient set is very similar in size to the SSD efficient set, but these two sets are not identical and, for the shortest subperiod, the M-V efficient set is larger than the SSD efficient set.

[1]Levy, H., and Hanoch, G., "Relative Effectiveness of Efficiency Criteria for Portfolio Selection," *Journal of Financial and Quantitative Analysis,"* Vol. 5, 1970, pp. 63–76.

[2]Hadar, J. and Russell, W.R., "Rules for Ordering Uncertain Prospects," *American Economic Review*, Vol. 59, 1969, pp. 25–34.

[3]Hanoch, G. and Levy, H., "The Efficiency Analysis of Choices Involving Risk," *Review of Economic Studies*, Vol. 36, 1969, pp. 335–46.

Levy and Sarnat (L&S) (1970)[4] applied the FSD, SSD and M-V rule to annual rates of return of American mutual funds. Table 7.2 reports the results of their study:

Table 7.2: The FSD, SSD and M-V Efficient Sets: American Mutual Funds with Annual Rates of Return

Period	Number of funds	Size of the efficient set FSD	SSD	M-V
1946–1967	58	50	17	12
1958-1967	87	62	16	14
1958-1967	149	89	18	21

Source: Levy & Sarnat, 1970.

Application of the decision rules to portfolios (mutual funds) rather than individual stocks does not change the basic results observed in Table 7.1: The FSD is relatively ineffective, especially for the longer periods where more rates of returns are available. Once again, addition of the risk aversion (SSD) assumption drastically reduces the efficient sets; from 50 to 17 for the period 1946–1967, from 62 to 16 for the period 1958–1967, and from 89 to 18 for the period 1958–1967. The M-V efficient set is similar in magnitude to the SSD efficient set and, for one subperiod, it is even larger.

Porter and Gaumnitz (P&G)[5] (1972) who also tested the effectiveness of SD rules, added the TSD rule.[6] Their data consist of 72 monthly rates of return on 893 portfolios generated randomly from 925 stocks chosen from the Chicago Price Relative tape for the period 1960–1965. The results are in line with the results of L&H and L&S. The P&G study also found that TSD is significantly more effective than SSD and M-V (see footnote 6), and that most of the options included in the M-V efficient set, but not in the SSD efficient set tend to be those with low mean and low variance.

Porter (1973a)[7] conducted an extensive empirical comparison of stochastic dominance and mean-variance efficiency using the same data as P&G. The content and size of the various efficient sets were examined for monthly, quarterly, semi-annual, and annual rates of return. The results of this analysis are shown in Table 7.3.

---

[4]Levy, H., and Sarnat, M., "Alternative Efficiency Criteria: An Empirical Analysis," *Journal of Finance*, Vol. 25, 1970, pp. 1153–58.

[5]Porter, R.B., and Gaumnitz, J.E., "Stochastic Dominance vs. Mean Variance Portfolio Analysis: An Empirical Evaluation," *American Economic Review*, Vol. 62, 1972, pp. 438–46.

[6]As mentioned in the introduction, the TSD algorithm that is in the literature is inaccurate; hence the TSD and TSDR results published in the various empirical studies may contain some errors.

[7]Porter, R.B., "An Empirical Comparison of Stochastic Dominance and Mean-Variance Choice Criteria," *Journal of Financial and Quantitative Analysis*, Vol. 8, 1973, pp. 587–608.

Table 7.3. Size of Efficient Sets (total population: 893 portfolios)

	Type of Data			
	Monthly	**Quarterly**	**Semi-annual**	**Annual**
Number of Observations	72	24	12	6
	Number of portfolios in efficient set			
FSD	893	676	404	101
SSD	216	127	82	32
TSD	146	69	44	12
M-V	67	65	56	41
	Percent of total 893 portfolios			
FSD	100.0	75.7	45.2	11.3
SSD	24.2	14.2	9.2	3.6
TSD	16.3	7.7	4.9	1.3
M-V	7.5	7.3	6.3	4.6

Source: Porter, 1973a.

Again, the main results of the previous studies are confirmed. The most striking results are the ineffectiveness of FSD with monthly data, and the relatively low sensitivity of the size of M-V efficient sets to type of data. Another interesting finding is the similarity of the SSD, TSD, and M-V efficient sets, a similarity that becomes even more marked the longer the period under consideration. An additional finding (not reported in Table 7.3) is that options that are efficient according to M-V but not according to SSD or TSD, generally have a low mean and low variance.

In another study, Porter (1974)[8] compares SSD with two criteria based on the mean and the lower semi-variance (instead of the variance) of the distributions. In the first version, the *semi-variance* S_E is defined as:

$$S_E = \begin{cases} E[R-E(R)]^2 & R \le E(R) \\ 0 & R > E(R) \end{cases}$$

where E(R) is the expected value of the rates of return.

[8]Porter, R.B., "Semi-variance and Stochastic Dominance: A Comparison," *American Economic Review*, Vol. 64, 1974, pp. 200–204.

In the second version, the semi-variance S_h is measured in terms of a reference point, h, as follows:

$$S_h = \begin{cases} E[R-E(R)]^2 & R \leq h \\ 0 & R > h \end{cases}$$

Theoretically, E-S_h (but not E-S_E, see Porter [1974]) is a subset of the SSD efficient set, and the M-V efficient set is not such a subset. Therefore, Porter postulated that there would be more consistency between the SSD and E-S_h efficient sets than between the SSD and M-V efficient sets. Porter's empirical results are summarized in Table 7.4. As can be seen, there are no E-S_h efficient options that are not SSD efficient, but there are some M-V efficient options that are not SSD efficient.

Table 7.4. Size of Efficient Sets: Correspondence between Criteria

	Type of Data	
	Monthly	**Quarterly**
SSD	216	127
M-V	69	65
E–S_E	70	62
E–S_h (h = 0)	50	33
E–S_h (h = 0.1)	49	33
SSD and		
M-V	61	45
E–S_E	70	55
E–S_h (h = 0)	50	33
E–S_h (h = 0.1)	49	33

Source: Porter, 1974.

This result was expected because the efficient set produced by E-S_h is defined as a subset of the SSD efficient set, whereas the M-V efficient set is not. The E–S_h efficient set includes a relatively small number of SSD efficient options; for example, only 33 of the 127 SSD efficient options are efficient with the E-S_h efficient rule (see quarterly data in Table 7.4). The relatively small E-S_h efficient set is a drawback rather than advantage of the E–S_h rule because it can lead to elimination of an optimum option from the efficient set. In short, the E-S_h rule relegates many options to the inefficient set, some of which may constitute the optimum choice for some risk averters.

Joy and Porter (J&P) (1974)[9] employed FSD, SSD, and TSD to test the performance of mutual funds relative to market performance (as measured by the Dow-Jones Industrial Average, DJIA). J&P used the data on 34 mutual funds, employed in earlier studies by Sharpe (1966) and Arditti (1971). They found that by FSD, none of the 34 mutual funds dominates or is dominated by the DJIA, by SSD, none of the mutual funds dominates the DJIA but 6 are dominated by it, and by TSD, none of the mutual funds dominates the DJIA, but 9 are dominated by it. These results confirm Sharpe's empirical M-V performance results.

Vickson and Altman (V&A) (1977)[10] investigated the relative effectiveness of the Decreasing Absolute Risk Aversion SD (DSD) criterion, which is a stochastic dominance criterion for decreasing absolute risk aversion (DARA) utilities. Using a data base consisting of 100 portfolios generated from 20 stocks listed on the Toronto Stock Exchange, they found that DSD does not significantly improve effectiveness relative to the TSD criterion.

From the aforementioned results, we see that the addition of assumptions such as $U'''>0$ or DARA does not significantly reduce the size of the efficient set. Hence, let us now turn to assumptions regarding the capital market, and in particular, to the existence of a riskless asset, bearing in mind that such an assumption plays a key role in the M-V analysis: It reduces the number of M-V efficient *levered* portfolios from infinity to only one efficient *unlevered* portfolio. Thus, adding the riskless asset may also lead to a substantial reduction in the size of the SD efficient sets.

Levy and Kroll (1979a)[11] tested Stochastic Dominance with a Riskless asset (SDR) criterion. The data for this study were the annual rates of return of 204 mutual funds in 1965–1974, 73 mutual funds in 1953–1974, and 27 mutual funds in 1943–1974. In all cases, the assumption of borrowing and lending at a riskless interest rate reduced the size of the efficient set significantly, sometimes quite dramatically.

The effect of including a riskless asset on the relative effectiveness of the decision criteria is illustrated in Table 7.5. As can be seen (part A of the table), the assumption of borrowing and lending at a riskless interest rate leads to an impressive reduction of the size of the efficient sets for risk averters. However, FSDR remains relatively ineffective. In the study, the SD and SDR criteria were also employed to compare the performance of mutual funds relative to the Fisher

---

[9] Joy, O.M., and Porter, R.B., "Stochastic Dominance and Mutual Fund Performance," *Journal of Financial and Quantitative Analysis*, Vol. 9, 1974, pp. 25–31.

[10] Vickson, R.G., and Altman, M., "On the Relative Effectiveness of Stochastic Dominance Rules: Extension to Decreasingly Risk-Averse Utility Functions," *Journal of Financial and Quantitative Analysis*, Vol. 12, 1977, pp. 73–84.

[11] Levy, H., and Kroll, Y., "Efficiency Analysis with Borrowing and Lending: Criteria and their Effectiveness," *Review of Economics and Statistics*, February 1979.

Arithmetic Average Index, which serves as a proxy for an unmanaged portfolio (Table 7.5, panel B). The most impressive result of this comparison is that with M-VR, SSDR, or TSDR, we can, in most cases, identify dominance relations between the funds and the unmanaged portfolio. However, no conclusions can be reached from these investment criteria without a riskless asset. A second important finding is that in most cases, the mutual funds are inferior to the unmanaged portfolio. This inferiority increases when the riskless asset is available.

Table 7.5. Mutual Fund: Efficient Sets and Peformance, 1953–1974 (total population 73 mutual funds)

Criterion and dominance condition	Without riskless asset	With Risk-Free Asset at Rate 2 percent	4 percent	6 percent
Part A: Size of the Efficient Sets				
M-V, M-VR	11	1	1	1
FSD, FSDR	68	51	57	56
SSD, SSDR	16	9	6	5
TSD, TSDR	15	5	3	2
Part B: Performance of Funds Compared to Fisher Index				
M-V, M - VR				
Superior funds	6	40	28	18
Inferior funds	2	33	45	55
No dominance	65	—	—	—
FSD, FSDR				
Superior funds	—	1	—	—
Inferior funds	—	—	—	2
No dominance	73	72	73	71
SSD, SSDR				
Superior funds	4	18	13	10
Inferior funds	2	22	34	32
No dominance	67	33	26	11
TSD, TSDR				
Superior funds	3	22	16	13
Inferior funds	3	29	43	33
No dominance	67	22	14	3

Source: Levy and Kroll, 1979a.

7.2 THE EFFECTIVENESS OF THE VARIOUS DECISION RULES: AN IMPERFECT MARKET

Although the SDR results described in Table 7.5 are encouraging, they are based on the unrealistic assumption of a perfect market in which $r_b = r_l$. Kroll and Levy (K&L) (1979)[12] tested the effectiveness of the various SDR criteria under the realistic assumption of $r_b > r_l$ which pertains in the market. The main results for an imperfect market are presented in Table 7.6. The imperfection imposed does not significantly change the size of the efficient sets. Note that under this imperfection assumption, the *separation property* of the M-VR criterion no longer holds, and the M-VR efficient set sometimes includes more than one risky option. Another result of K&L (1979) (not shown in Table 7.6) is that in some cases corresponding to an imperfect market, the TSDR decision rule is more effective than the M-VR, although in most cases their effectiveness is similar.

Table 7.6. Size of Efficient Sets in Imperfect and Perfect Markets (total population 73 mutual funds)

	FSDR	SSDR	TSDR	M-VR
Imperfect market				
$r_b = 4\%, r_l = 2\%$	57	10	5	1
$r_b = 5\%, r_l = 2\%$	58	10	5	3
Perfect market				
$r_b = r_l = 2\%$	51	9	5	1
$r_b = r_l = 4\%$	57	6	3	1
$r_b = r_l = 6\%$	56	5	2	1

Source: Kroll & Levy, 1979b

7.3 THE PERFORMANCE OF MUTUAL FUNDS WITH TRANSACTION COSTS

In part B of Table 7.5, we compare the performance of mutual funds and the Fisher Index, disregarding possible differences in the transaction costs of individual stocks and mutual funds. However, the cost of buying and selling mutual funds is usually much higher than that of individual stocks. The effect of this difference on the performance of mutual funds was examined by Kroll (1977).[13] The most striking of his results, presented in Table 7.7, is that the effect of transaction costs on the relative performance of mutual funds is not significant

[12]Kroll, Y., and Levy, H., "Stochastic Dominance with a Riskless Asset: An Imperfect Market," *Journal of Financial and Quantitative Analysis*, Vol. 14, June 1979, pp. 179-204.

[13]Kroll, Y., "Preferences Among Combinations of Risky Assets with a Riskless Asset: Criteria and Implications," Ph.D. Thesis, Hebrew University of Jerusalem, 1977.

by SD criteria, but very significant by the SDR criteria. Even with a 3 percent difference in transaction costs by the M-V and SD criteria, the no-dominance situation is the most frequent. However, the effect of transaction costs by the M-VR and SDR criteria is considerable. For example, at 4 percent interest and no difference in transaction costs, 28 funds are superior and 45 inferior by the M-VR criterion. However, at 3 percent difference in transaction costs, all the funds are inferior by M-VR and almost all of them are inferior to the unmanaged portfolio by SSDR and TSDR.

Table 7.7. The Performance of Mutual Funds and the Fisher Index Transaction Costs: 1953–1974

Transaction Costs of 0, 1, 2, and 3 Percent Deducted from Returns:									
	0 Percent			1 Percent			3 Percent		
	Superior Funds	Inferior Funds	No Dominance	Superior Funds	Inferior Funds	No Dominance	Superior Funds	Inferior Funds	No Dominance
Without Riskless Asset									
M-V	6	2	65	1	2	70	—	2	71
FSD		—	73	—	—	73		—	73
SSD	4	2	67	1	4	68	—	7	66
TSD	3	3	67	1	5	67	—	9	64
With Riskless Asset									
Riskless Rate 2 Percent									
M-VR	40	33	—	21	52	—	3	70	—
FSDR	1	—	72	—	—	73	—	9	64
SSDR	18	22	33	12	38	23	—	68	5
TSDR	22	29	22	13	52	8	—	70	3
Riskless Rate 4 Percent									
M-VR	28	45	—	14	59	—	—	73	—
FSDR	—	—	73	—	2	71	—	15	58
SSDR	13	34	26	3	53	12	—	72	1
TSDR	16	43	14	11	57	5	—	73	—
Riskless Rate 6 Percent									
M-VR	18	55	—	8	65	—	—	73	—
FSDR	—	2	71	—	5	68	—	19	54
SSDR	10	52	11	4	61	8	—	73	—
TSDR	13	55	15	5	62	6	—	73	—

Source: Kroll, 1977.

In the face of this strong evidence of the inferior performance of the mutual funds relative to the unmanaged portfolio, how can we explain the popularity of mutual funds among investors? The answer to this question seems to be the inability of most individual investors to hold and manage a well-diversified portfolio, e.g., the Dow Jones Index. Levy and Sarnat (1972)[14] investigated this possibility by comparing the performance of mutual funds with efficient sets constructed from a limited number of popular individual stocks and succeeded in demonstrating that mutual funds may be superior to such portfolios.

Many empirical studies have compared the various SD efficient sets or mutual funds to the unmanaged portfolio. Saunders, Ward and Woodward (1980)[15] tested the performance of U.K. unit trusts relative to the unmanaged portfolio and found that higher order stochastic dominance rules increase the proportion of unit trusts that dominate the unmanaged portfolio. Jean and Helms (1987) applied various necessary conditions to the distributions moments of SD dominance in order to establish efficient sets. These efficient sets were then compared to the true efficient sets derived from the optimal SD rules rather than the necessary rules only. They found that the application of simple necessary rules yields a set of portfolios similar to the correct SD efficient set. For more results of empirical SD studies and comparison to efficient sets derived by other rules, see Tehranian (1980)[16] and Okuney (1988)[17].

7.4 FURTHER REDUCTION IN THE EFFICIENT SETS: CONVEX STOCHASTIC DOMINANCE (CSD)[18]

With pair-wise comparisons we obtained in previous chapters the FSD and SSD efficient sets. As FSD and SSD are optimal rules for $U \in \mathbf{U}_1$ and $U \in \mathbf{U}_2$, respectively, no further reduction in the efficient set is possible with pair-wise comparisons. However, if one does not confine herself to pair-wise comparisons, the efficient set may be further reduced. The technique of a comparison of three or more distributions is called convex stochastic dominance (CSD). The intuition of CSD is as follows: suppose that three distributions F, G and H are, say in the FSD efficient set. Suppose that for some subset of $\mathbf{U}_1$, e.g., $\mathbf{U}_s \subset \mathbf{U}_1$, G is preferred over F and therefore G is not eliminated from the FSD efficient set by F. But it is

[14] Levy, H., and Sarnat, M., "Investment Performance in an Imperfect Securities Market," *Financial Analyst Journal*, Vol. 28, 1972, pp. 78–81.

[15] Saunders, A., Ward, C., and Woodward, R., "Stochastic and the Performance of UK Unit Trusts," *Journal of Financial and Quantitative Analysis*, June 1980.

[16] Tehranian, H., "Empirical Studies in Portfolio Performance Using Higher Degrees of Stochastic Dominance," *Journal of Finance*, 35, March 1980, 159-171.

[17] Okuney, J., "A Comparative Study of the Gini's Mean Difference and Mean Variance in Portfolio Analysis," *Accounting and Finance*, 28, 1988, pp. 1-15.

[18] The convex FSD and SSD proofs presented in this Chapter are taken from Levy, M., "Is Stochastic Dominance Efficient Set Really Efficient? A Joint Stochastic Dominance Analysis," Hebrew University, Jan. 2004, working paper. Yet, the most general proof with many applications is given by Fishburn, P.S., "Convex Stochastic Dominance with Continuous Distributions," *Journal of Economic Theory*, 1974, 7, pp. 143–158.

possible that for U_S, H is preferred over G, therefore G can be eliminated from the efficient set. We demonstrate the CSD with FSD and SSD.

FSD, CSD with Three Assets in the Efficient Set (N=3)

Suppose that there are three options, F, G, and H in the FSD efficient set. The set of all non-decreasing utility functions $\mathbf{U}_1$ (where $U \in \mathbf{U}_1$ if $U' \geq 0$) can be partitioned to $\mathbf{U}_1 = U_1^F \cup U_1^G$, where U_1^F is the set of all preferences that prefer F over G, and U_1^G is the set of all preferences that prefer G over F. Though neither F nor H directly dominate G by FSD, we provide below a condition such that for all $U \in U_1^G$, option H provides a higher expected utility than option G. If this condition holds, G can be eliminated from the efficient set, because all investors with $U \in U_1^F$ prefer F over G, and all investors with $U \in U_1^G$ prefer H over G. Thus, though F does not dominate G and H does not dominate G, *jointly* F and H dominate G. Let us state the condition for the elimination of an option from the FSD efficient set by a joint dominance.

Theorem 7.1: Suppose that H, F and G are in the traditional FSD efficient set. Any investment G can be eliminated from the FSD efficient set if there exist two other investments, F and H, such that:

$$F(x) - G(x) \leq G(x) - H(x) \quad \textit{for all } x, \tag{7.1}$$

where G(x), F(x), and H(x), are the cumulative distributions of the three respective investments.

Proof:

Given (7.1), we show below that any investor who prefers G over F will also prefer H over G (and symmetrically any investor who prefers G over H will also prefer F over G). As by assumption F, G, and H are in the traditional FSD efficient set, we have some preferences $U \in U_1^G$, i.e. preferences revealing a higher expected value under G than under F. Let us focus on these preferences. A preference of G over F implies:

$$EU_G - EU_F = \int_a^b [F(t) - G(t)]U'(t)dt \quad > 0 \qquad \text{for all } U \in U_1^G \tag{7.2}$$

But (7.2) also implies the preference of H over G for all $U \in U_1^G$ because:

$$EU_H - EU_G = \int_a^b [G(t) - H(t)]U'(t)dt \quad \geq \int_a^b [F(t) - G(t)]U'(t)dt \quad > 0 \tag{7.3}$$

where the first inequality follows from eq. (7.1) and the non-negativity of U'. Thus, condition (7.1) implies that though for all $U \in U_1^G$ G is preferred over F, no investor will choose G, and it can be eliminated from the FSD efficient set. Q.E.D.

Extension to N Assets In the FSD Efficient Set

Theorem 7.2 Consider n assets that are in the traditional FSD efficient set. Any asset G can be eliminated from the FSD efficient set if there exist N-1 other assets, $F_1, F_2, \ldots, F_{N-2}$ *and H, and a vector of positive weights* $\alpha_1, \alpha_2, \ldots \alpha_{N-2}$, *such that:*

$$\sum_{i=1}^{N-2} \alpha_i F_i(x) - G(x) \le G(x) - H(x) \quad \text{for all x,} \tag{7.4}$$

where G(x), $F_i(x)$, and H(x), are the cumulative distributions of the respective investments, and $\sum_{i=1}^{N-2} \alpha_i = 1$.

Proof

The convex FSD criterion with N assets, as stated in eq. (7.4) implies that any investor who prefers G over all the F's will prefer H over G. This, of course, means that G is FSD inefficient, because no investor will choose G: she will either prefer one of the F's over G, or if she prefers G over all the F's, she will prefer H over G. To see that the condition of eq. (7.4) indeed implies that any investor who prefers G over all the F's also prefers H over G, note that the preference of G over all the F's for an investor with preference U_0 can be stated as:

$$EU_G - EU_{F_1} = \int_a^b (F_1(t) - G(t))\, U_0'(t)dt > 0$$

$$EU_G - EU_{F_2} = \int_a^b (F_2(t) - G(t))\, U_0'(t)dt > 0$$

.

.

.

$$EU_G - EU_{F_{N-2}} = \int_a^b (F_{N-2}(t) - G(t))\, U_0'(t)dt > 0$$

Multiplying the first equation by α_1, the second by α_2, and so forth, and adding all the equations we obtain:

$$\int_a^b \left(\left(\sum_{i=1}^{N-2} \alpha_i F_i(t)\right) - G(t)\right) U_0'(t)\, dt > 0$$

(recall that all the α's are positive with $\sum_{i=1}^{N-2} \alpha_i = 1$). If condition (7.4) holds, i.e.,

$$\sum_{i=1}^{N-2} \alpha_i F_i(x) - G(x) \leq G(x) - H(x) \text{ for all x,}$$

then,

$$EU_H - EU_G = \int_a^b [(G(t) - H(t)] U_0'(t) dt > \int_a^b \left(\left(\sum_{i=1}^{N-2} \alpha_i F_i(t)\right) - G(t)\right) U_0'(t) dt > 0$$

i.e., H is preferred over G for this investor. Q.E.D.

Theorem 7.3 states the condition for convex SSD.

Theorem 7.3: Suppose that assets F, G and H are in the SSD efficient set. Then, asset G can be eliminated from the SSD efficient set if there exist two other assets, F and H, such that:

$$\int_a^x [(F(t) - G(t)] dt \leq \int_a^x [G(t) - H(t)] dt \quad \textit{for all } x. \tag{7.5}$$

Proof:

The idea is very similar to that of the convex FSD proof, with the difference that eq.(7.2) is modified via integration by parts, and use is made of the information that $U'' \leq 0$. Preference of G over F for all $U \in U_2^G$ implies (by integration by parts of (7.2), see Chapter 3 to:

$$EU_G - EU_F = U'(b) \int_a^b [F(t) - G(t)] dt - \int_a^b U''(x) \left(\int_a^x [F(t) - G(t)] dt \right) dx > 0 \tag{7.6}$$

Using eq. (7.5) and using it once again for the case x = b implies that H is preferred over G for all $U \in U_2^G$, because:

$$EU_H - EU_G = U'(b) \int_a^b [G(t) - H(t)] dt - \int_a^b U''(x) \left(\int_a^x (G(t) - H(t)) dt \right) dx \geq$$

$$U'(b) \int_a^b [F(t) - G(t)] dt - \int_a^b U''(x) \left(\int_a^x (F(t) - G(t)) dt \right) dx > 0, \tag{7.7}$$

where the inequality follows from (7.5) and $U''(x) \leq 0$. Q.E.D.

Thus, all investors with preference $U \in U_2^G$ prefer G over F, but for them H is superior to G. Thus, for $U \in U_2^F$ F is preferred over G, and for $U \in U_2^G$, H is preferred over G. Hence, for all $U \in \mathbf{U_2}$ either F or H is preferred over G, and therefore G can be safely relegated to the SSD inefficient set. It is easy to see that the convex SSD efficient set is a subset of the convex FSD efficient set.

Let us turn now to the N-assets case.

Consider N assets that are in the traditional SSD efficient set. Any investment G can be eliminated from the SSD efficient set if there exist N-1 other assets, $F_1, F_2, \ldots, F_{N-2}$ and H, and a vector of positive weights $\alpha_1, \alpha_2, \ldots \alpha_{N-2}$, such that:

$$\int_a^x \left(\left(\sum_{i=1}^{N-2} \alpha_i F(t) \right) - G(t) \right) dt \le \int_a^x (G(t) - H(t)) dt \quad \text{for all } x, \qquad (7.8)$$

where $\sum_{i=1}^{N-2} \alpha_i = 1$.

The proof is similar to the FSD CSD with N assets.

Levy, M. [2004] (see footnote 18) provides an algorithm to find the convex FSD and convex SSD efficient sets. He found that convex FSD criterion reduces only a little the FSD efficient set, but convex SSD reduces the efficient set by about one third.

Other studies reveal up to 60% reduction in the efficient set by employing CSD. For more details see Meyer (1977[19], 1979[20]) and Bawa, Bondurtha, Rao and Suri (1985)[21].

7.5 SAMPLING ERRORS: STATISTICAL LIMITATIONS OF THE EMPIRICAL STUDIES

In all of the aforementioned empirical studies, the efficient sets are based solely on the empirical distribution. This means that the sample distributions are assumed to be the true distributions, and the significance of the difference between two distributions F and G is not tested. In actual fact, such a test has yet to be developed. However, several tests are available to check whether a given sample distribution F is significantly different from some known distribution F_0 by FSD, or whether F significantly dominates G by FSD where both F and G are unknown. Such statistical tests are more complicated and not available to higher order SD rules. Deshpande and Singh (1985)[22] succeeded in developing an SSD test for large samples, and a different test for small samples for specific distributions, but not a general SSD test for all sample sizes. In both cases, it is assumed that the null hypothesis distribution F_0 is known. Eubank, Schechtman

[19]Meyer, J., "Choices Among Distributions," *Journal of Economic Theory*, 11, 1977, pp. 99–132.
[20]Meyer, J., "Second Degree Stochastic Dominance with Respect to a Function," *International Economic Review*, 18, 1979, pp. 477–487.
[21]Bawa, V.S., J. Bondurtha, M.R., Rao and H.L. Suri, "On Determination of Stochastic Dominance Optimal Set," *Journal of Finance*, 40, 1985, pp. 417–431.
[22]Deshpande, J.V. and H. Singh, "Testing for Second Order Stochastic Dominance," *Comm. Statist., Part A: Theory and Methods*, 14, 1985, pp. 887–897.

and Yitzhaki (1993)[23] extended the SSD test to the case where both distributions F and G are unknown. Their test, however, is also confined to large samples. Thus, we have statistical tests for FSD, but only limited tests for SSD, and no tests at all for higher order SD decision rules.
Although no empirical studies have tested for the significance of the SD efficient set, some simulation studies have demonstrated the sampling errors involved in the division of the feasible set into efficient and inefficient sets. Let us elaborate:

The empirical distributions of rates of return in the aforementioned empirical studies are estimates of the (unobserved) true distributions. Even if we are willing to assume that these distributions are stable over time, sampling errors may lead to improper portfolio selection, even if optimum efficiency criteria are used. These errors are not confined to SD analyses; they also characterize other investment decision rules such as the M-V rule.

To demonstrate possible sampling error, recall that in most cases the true distributions of rates of return are unknown. Hence, like any statistical test which relies on a sample of observations, efficient set analyses (in SD as well as M-V frameworks) are also exposed to sampling error. We distinguish between two possible error types of sampling. To demonstrate, let us first assume that portfolio choices rely on two steps: in the first, the investment consultant derives the efficient set, and in the second, the investor selects his/her optimal portfolio from the efficient set. All efficiency analyses focus on the first step.

Type I error: There is a dominance in the population but this dominance is *not* revealed in the sample data. This error is not as severe. It induces a relatively large efficient set from which the optimum portfolio is selected in the second step. For example, suppose that x and y are normally distributed with $\mu_x = 10\%$, $\sigma_x = 10\%$, $\mu_y = 5\%$, $\sigma_y = 20\%$, and that we obtain no dominance in the sample. This constitutes an error but it is not a severe error because both x and y will be situated (mistakenly) in the (sample) efficient set. Thus, both will be presented to the investor by the investment consultant, the investor may select investment x as his/her optimal portfolio, and no error will have occurred.

Type II error: There is *no* dominance in the population but a dominance is revealed in the sample data. This error is serious because portfolios that are optimal for some investors may be relegated to the inefficient set which, in turn, will reduce the investor's expected utility. For example, suppose that x and y are normally distributed with $\mu_x = 10\%$, $\sigma_x = 10\%$, $\mu_y = 5\%$, $\sigma_y = 20\%$. If, in the sample, y dominates x by the M-V rule, x, which is a better investment, will be relegated to the inefficient set. In other words, the efficient set presented to the investor by the investment consultant will not include x and, therefore, x

[23]Eubank, R., Schechtman, E., and Yitzhaki, S., "A Test for Second Order Stochastic Dominance," *Commun. Statist. – Theory Math.*, 22(7), 1993, pp. 1893–1905.

definitely will not be selected by the investor and hence the expected utility of the investor will not be maximized. This constitutes a serious error.

Dickinson (1974)[24] was the first to measure the errors involved in the mean-variance analysis. Johnson and Burgess (1975)[25] used a simulation technique to analyze the sample error of an independent normal distribution and its influence on SSD and M-V efficiency. The assumption of independence is unrealistic because most pairs of portfolios under comparison contain some identical assets and, therefore, they are dependent. In a later study, Levy and Kroll (1980a[26] and 1980c[27]) relaxed the assumption of independence and employed simulations and mathematical analysis to examine type I and type II errors for dependent distributions as well. The purpose of Levy and Kroll's simulations was to measure the effect of sample size and degree of dependence on the dominance relations as well as the magnitude of the sampling errors, by SD and M-V decision criteria. The main results of this study are as follows:

1. Both the size of the sample and the degree of dependence have a strong effect on the relative effectiveness of the various decision criteria which, in turn, implies a strong effect on the sampling errors.
2. The ineffectiveness of FSD is mainly due to sampling error. For instance, in the case of normal distributions, if two distributions have the same standard deviation of 20 percent but their means are 30 and 15 percent, respectively, then the first option will dominate the second by FSD. However, most of the simulations with such pairs of parameters failed to show this dominance in the sample. For example, with a correlation coefficient of 0.5 and a sample of 30 observations, 88 percent of the tests failed (mistakenly) to show FSD dominance. What is even more surprising is that this kind of error becomes more serious as the sample size increases, up to about 30 observations. Thereafter the error levels off, decreasing only slightly as the number of observations becomes very large.
3. The probability of finding dominance in the sample when it exists in the population is much higher with SSD, TSD, or M-V than with FSD. Thus, Type I errors are relatively small with SSD, TSD (and MV). On the other hand, the probability of finding dominance in the sample when it does not exist in the true distribution is also much higher with SSD, TSD, or M-V than with FSD. Thus, Type II errors are relatively large in higher order SD. Therefore, we conclude that criteria that are effective in the

---

[24]Dickinson, J.P., "The Reliability of Estimation Procedures in Portfolio Analysis," *Journal of Financial and Quantitative Analysis*, Vol. 9, 1974, pp. 447–62.

[25]Johnson, K.H., and Burgess, R.C., "The Effects of Sample Sizes on the Accuracy of EV and SD Efficiency Criteria," *Journal of Financial and Quantitative Analysis*, 10, December 1975, pp 813-820.

[26]Levy, H., and Kroll, Y., "Stochastic Dominance: A Review and Some New Evidence," *Research in Finance*, Vol. 2, 1980, pp. 163–227.

[27]Levy, H., and Kroll, Y., "Sampling Errors and Portfolio Efficiency Analysis," *Journal of Financial and Quantitative Analysis*, 15, No. 3, September 1980, pp. 655–688.

sample have a relatively high probability of the more serious error of indicating spurious dominance, and a relatively low probability (in comparison to FSD) of the less serious error of not finding dominance in the sample when it is present in the population.

4. In small samples, the probability of error is sometimes higher by M-V than by SSD or TSD. This result is interesting because, intuitively, one might expect sampling problems to be more serious with SD owing to the fact that the input for SD tests is the whole distribution and the input for M-V tests is only the mean and variance.

7.6 SUMMARY

The stochastic dominance rules as well as the M-V rule are employed to construct the various efficient sets as well as to examine the performance of mutual funds relative to the unmanaged portfolio (e.g., the Standard & Poor's index).

Most of the empirical studies have shown the FSD rule to be ineffective in that it produces a relatively large efficient set. The SSD and M-V are more effective and yield efficient sets that are similar in size (about 10–20 percent of the feasible set) but not necessarily similar in content. In most cases, the TSD efficient set is only slightly smaller than the SSD efficient set; however, some studies have shown the TSD to produce a more substantial reduction in the efficient set. Taking into account that some of the studies employ wrong TSD algorithms, the efficient set is even larger than what is reported in previous studies, and hence becomes very close to the SSD efficient set.

Using the SSDR and the TSDR, we find that the inclusion of the riskless asset (in perfect and imperfect markets, alike) induces a dramatic reduction in the number of portfolios included in the efficient set. In some cases, only 2 out of the 73 mutual funds included in the empirical study end up in the efficient set. The MV-R efficient set, by definition, includes only one investment, the one with the highest Sharpe ratio (see Chapter 11). The common SD efficient set can be further reduced by employing a convex stochastic dominance. Convex Stochastic Dominance (CSD) may eliminate from the efficient set some elements. The CSD may be obtained by a dominance of a linear combination of elements included in the efficient set (which can be considered as a mix strategy) over another element which is also included in the pair-wise efficient set. Finally, the efficient and inefficient sets are exposed to Type I and Type II errors because they rely on sample distributions rather than true population distributions.

In most cases, when the SD and M-V rules are used to test the performance of mutual funds, no dominance relative to the unmanaged portfolio is induced. However, when the riskless asset is added, a high percentage of the mutual funds turn out to be inferior relative to the unmanaged portfolio by the MV-R, SSDR and TSDR.

KEY TERMS

Effectiveness (of the decision making rule)

Feasible set

Mutual fund

Levered portfolio

Unlevered portfolio

Semi-variance

M-VR

Separation property

Fisher Arithmetic Average Index

Transaction cost

Convex stochastic dominance (CSD)

Mix strategy

Type I error

Type II error

8

APPLICATIONS OF STOCHASTIC DOMINANCE RULES

Stochastic dominance rules are applicable in many varied fields and, in particular, finance, economics, insurance, statistics, agriculture, and medicine. Due to space constraints, we will mention only a few such applications in this chapter.

8.1 CAPITAL STRUCTURE AND THE VALUE OF THE FIRM

One of the first implicit applications of FSD in the area of finance was carried out by Modigliani and Miller (M&M) (1969).[1] Let us present here a slightly different analysis from the one suggested by M&M. Applying SD rules, we first show M&M's argument without corporate taxes and then demonstrate the case with corporate taxes.

Suppose that there are two firms identical in all respects apart from their capital structure. Let V_U and V_L stand for the market value of the unlevered and levered firms, respectively. The return to the investor who holds a portion α of the unlevered firms is Y_U (a random variable) given by $Y_U = \alpha X$ where X is the net income of the unlevered firm. Let us discuss two possible alternate cases regarding the value of the firms; $V_U > V_L$, and $V_L > V_U$ and show that these two cases contradict equilibrium and that in equilibrium, $V_L = V_U$ must hold.

a) Suppose that $V_U > V_L$. The investor who holds a portion α of the stock of the unlevered firm invests αV_U and his/her income is αX. M&M suggest that the holder of the unlevered firm should sell αV_U and buy αS_L where S_L is the equity value of the levered firm. Moreover, he/she should lend αB_L where B_L is the debt value of the levered firm. Hence, the total income from the new position will be $\alpha(X - rB_L) + \alpha r B_L = \alpha X$ which is exactly the same as before the transaction. However, because the investor invests αV_U before the transaction and $\alpha(S_L + B_L) = \alpha V_L$ after the transaction, and because by assumption, $V_U > V_L$, the investor is left with wealth $\alpha(V_U - V_L)$ which can be invested at the risk-free asset, r, to give the following end of period cash flow:

$$Y_L = \alpha X + \alpha(V_U - V_L)r = Y_U + \text{a positive constant.}$$

It is easy to verify that Y_L dominates Y_U by FSD. Thus, with the same initial investment as before, a superior distribution of future returns is obtained. Let us explain why $V_U > V_L$ cannot hold in equilibrium.

M&M claim that if $V_U > V_L$, then a cumulative distribution of return $F(Y_L)$ can be created which will dominate $F(Y_U)$ by FSD. Will all investors regardless of their preferences, sell the unlevered firm and invest in the

[1] Modigliani, F. and M.H. Miller, "Reply to Heins and Sprenkle," *American Economic Review*, 59, 1969, pp. 592–595.

levered firm? If so, then the price of the stock of the unlevered firm will drop and the price of the levered firm will increase, and this process will continue as long as $V_U > V_L$. Hence, $V_U > V_L$ is impossible in equilibrium. Before we turn to the case $V_L > V_U$, let us now elaborate on the motives for this financial transaction. We have two alternate explanations why investors should sell the stock of the unlevered firm and buy the stock of the levered firm regardless of prevailing preferences.

i. The investor holds only one risky asset, either the shares of the levered firm or the shares of the unlevered firm. In such a case, dominance of Y_L over Y_U by FSD is a sufficient condition for the investor to shift from the unlevered firm to the levered firm. (Note that *ex-ante*, all investors will be better-off by this shift, yet the realized return of Y_U may be greater than that of Y_L.) However, if another risky asset, say, Z is held with this stock, dominance of Y_L over Y_U will not be sufficient for such a shift. In such a case, $Y_U + Z$ should be compared to $Y_L + Z$. Thus, in a portfolio context, the FSD of Y_L over Y_U shown above is not sufficient for the argument that $V_U > V_L$ is impossible in equilibrium and we need case ii) below.

ii. If, like M&M, one adds the assumption that the returns of the two firms, X, are fully correlated, then Y_L and Y_U will also be fully correlated, in which case, FSD <==> arbitrage. Note if we are dealing with two firms that are identical in all respects except for their leverage, then X will be the same, and there will be a correlation of +1 between the income of the two firms. In this case, even if other risky assets are held, shifting from the unlevered firm will add a positive constant profit regardless of the other assets held. Namely, if Y_L dominates Y_U by FSD, $Y_U + Z$ will also dominate $Y_L + Z$. Moreover, adding the perfect positive correlation assumption guarantees not only that the shift from the unlevered firm to the levered firm is beneficial, *ex ante*, but also that the realized return will increase.

b) So far, we have proved that $V_U > V_L$ is inconsistent with equilibrium. We show here that $V_L > V_U$ is also impossible in equilibrium. Assume now that $V_L > V_U$. The investor who holds a portion α of the shares of the levered firm obtains a return of $Y_L = \alpha\,(X - rB_L)$. By selling the stocks of the levered firm, buying portion α of the shares of the unlevered firm, and borrowing αB_L, his/her income will be $\alpha X - \alpha r B_L = \alpha\,(X - rB_L)$ which is exactly the same as the income obtained before the transaction. However, before the transaction, the investment was αS_L (where S_L is the equity value of the levered firm) and after the transaction, the investment is $\alpha V_U - \alpha B_L = \alpha(V_U - B_L)$ (recall that borrowing reduces the amount invested) but, because by assumption $V_L \equiv S_L + B_L > V_U$, also $S_L > V_U - B_L$. Hence, the investor obtains the same return as before the transaction and can invest the difference $\alpha[S_L - (V_U - B_L)] = \alpha\,[V_L - V_U]$ in the riskless asset.

The return before the transaction was $Y_L = \alpha\,(X - rB_L)$ and after the transaction it is $Y_U = Y_L + \alpha(V_L - V_U)\,r = Y_L$ + a positive constant return. Hence, all investors will shift from the levered firm to the unlevered firm, and this process will continue as long as $V_L > V_U$.

From this analysis, M&M conclude that neither $V_U > V_L$ nor $V_L > V_U$ can hold in equilibrium, hence $V_L = V_U$ and, therefore, with no taxes, capital structure is irrelevant.

Figure 8.1 illustrates case b), where $Y_U = Y_L$ + a positive constant, namely $F(Y_U)$ dominates $F(Y_L)$ by FSD. As mentioned above, if there is FSD and the returns are not fully correlated, the *realized* return on the superior investment can be lower than the realized return on the inferior distributions (compare points X_1 and X_2 in Figure 8.1 corresponding to $F(Y_U)$ and $F(Y_L)$, respectively). However, if the returns are perfectly (and positively) correlated, then the realized return on the superior distribution will always be larger, too. For example, if X_2 occurs with $F(Y_L)$, then with a perfect positive correlation, X_3 must occur, hence, a higher return will be obtained with the superior investment.

Using the same argument, M&M (1969)[2] show that with corporate tax, T, in equilibrium the following must hold: $V_L = V_U + TB_L$, and Arditti, Levy and Sarnat (1977)[3] show that with corporate and personal taxes, the same type of arbitrage can be used to show that, in equilibrium, the following must hold:

$$V_L = V_U + B_L\left[1\frac{(1-T_c)(1-T_g)}{(1-T_p)}\right]$$

where T_c is the corporate tax rate, T_g is the personal capital gain tax, and T_p is the personal income tax (on interest and dividends). Assuming no capital gains tax, Miller (1977)[4] uses another argument to show that the above formula holds, where T_g is replaced with zero.

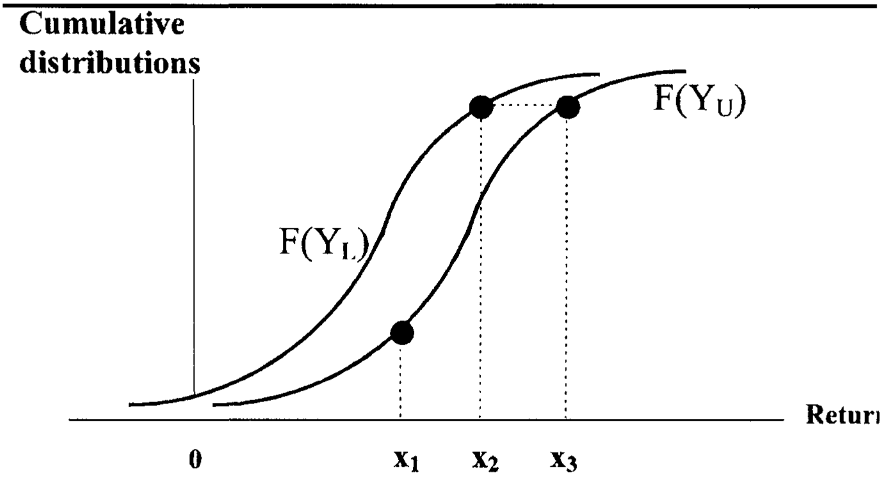
---

[2] See Footnote 1.

[3] Arditti, F., Levy, H. and Sarnat, M., "Taxes, Capital Structure and Cost of Capital: Some Extensions," *The Quarterly Review of Economics and Business*, Summer 1977, pp. 89–95.

[4] Miller, Merton, H., "Debt and Taxes," *Journal of Finance*, May 1977.

Figure 8.1: The cumulative distributions of Y_L and Y_U



8.2 PRODUCTION, SAVING AND DIVERSIFICATION

All investors seek a high mean rate of return but risk averters dislike risk. Thus, it is interesting to analyze, separately, the effect of an increase in the expected rate of return, and the effect of an increase in the risk on the risk averter's optimal investment decision. Rothschild and Stiglitz (R&S) (1971)[5] analyze these effects using stochastic dominance.

Specifically, R&S investigate the following four main issues:

(a) *Investment-Consumption.* An investor allocating his wealth between consumption today and consumption tomorrow, invests the amount not consumed today. In such a scenario, R&S show that, in general, increasing the uncertainty of the return yields ambiguous results. In other words, it is not clear whether a risk-averse investor would respond by increasing or decreasing her/his saving.

(b) *A Portfolio Problem.* Suppose an investor divides his investment between the safe asset and the risky asset. Now, increase the riskiness of the risky asset by adding a mean preserving spread (MPS, see Chapter 10). In this case, the M-V rule (based on the quadratic function) yields the misleading result that the investor will *always* decrease the investment proportion in the risky asset. Instead, R&S prove that the risk-averse investor will only *sometimes* increase his/her holding of

[5] Rothschild, M., and Stiglitz, J.E., "Increasing Risk: II. Its Economic Consequences," *Journal of Economic Theory*, Vol. 3, 1971, pp. 64–84.

the risky asset. (For additional analyses of diversification and stochastic dominance rules, see Chapter 10).

(c) *A Combined Portfolio-Saving Problem.* Suppose that an investor has to decide on how much to consume today and how much to invest. In addition, she/he has to decide on how to diversify between two risky assets. Levhari and Srinivasan (1969)[6] analyze this issue and show that, under certain conditions, increasing the variance of one risky asset and holding the mean constant, induces a reduction in the proportion invested in the risky asset. R&S show that, in general, this conclusion does not hold and ambiguous results are obtained even when the concept of "increase in variance," rather than "increase in risk" is employed as implied by the stochastic dominance rules. Thus, more restrictions on U_2 should be imposed (e.g., $U(W) = \ln W$) to obtain clear-cut results.

(d) *A Firm's Production Problem.* Suppose that Q = P(K, L) where Q is the future uncertain output, K, L stand for the capital and labor inputs, respectively, and P is a production function. R&S analyze the change in the optimum value K as the variability of Q increases. Here, too, unambiguous results are obtained only for a specific utility function. R&S also analyze the optimum output selected by the firm in the face of increase in the uncertainty. As in the previous cases, the results are ambiguous and a function of the characteristics of the absolute and relative risk-aversion measures.

The implication of these theoretical findings is quite discouraging. For example, consider a firm that changes its investment strategy in order to increase the demand for stocks by creating a mean preserving anti-spread, MSPA, or by avoiding MPS (see Chapter 10). The above analyses indicate that the firm is not guaranteed that its market value will increase as a result of this seemingly desirable change. Hence, in a portfolio context, avoiding a mean preserving spread or creating a mean preserving anti-spread is not necessarily desirable.

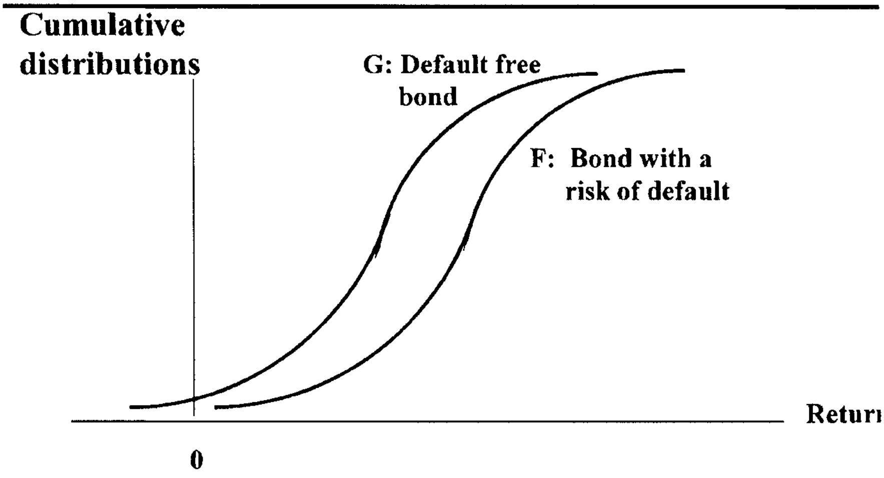
8.3 ESTIMATING THE PROBABILITY OF BANKRUPTCY

Bonds are characterized and rated by their bankruptcy risk. Government bonds have no risk of bankruptcy because the government can always print more money to pay back the bondholders. This is not the case for corporate or municipal bonds which may default on their payments. Suppose that two bonds are identical in all respects except for their risk of bankruptcy. Investors will pay a higher price for the bond with no bankruptcy risk (relative to the bond which may bankrupt), hence the rates of return on such a bond will be lower as long as bankruptcy does not occur. Indeed, when we compare *ex-post* rates of return on two bonds, one with bankruptcy risk and one

---

[6] Levhari, D. and T.N. Srinivasan, "Optimal Saving Under Uncertainty," *Review of Economic Studies*, 36, 1969, pp. 153–164.

Figure 8.2: The cumulative distributions of returns corresponding to two bonds



without, we generally obtain FSD of one distribution over the other. Empirical results reveal a relationship described by Figure 8.2. Distribution F dominates G by FSD where F corresponds to a bond which may default on its payments. Does this imply that every investor should invest in F rather than in G? Not necessarily: There is a probability of bankruptcy of the bond corresponding to distribution F which is not reflected with the *ex-post* data because the data generally relate to existing bonds which, by definition, have not gone bankrupt. There are two ways to incorporate bankruptcy risk into the analyses of such bonds:

a) Comparison of the average rate of return on a sample of various categories of bonds, including bonds that default on their payments with a sample of government bonds.

b) Comparison of distributions such as F and G in Figure 8.2 to estimate their probability of default.

We focus here on approach b) because it addresses itself to the investor's main concern, namely, the probability of bankruptcy.

In adopting approach b), Broske and Levy (1989)[7] employ SD rules to estimate the probability of bankruptcy implied by the market price of bonds. They assume that investors who are risk averters consider investing either in government bonds or in corporate bonds rated Aaa. Let us denote the cumulative distribution of the rates of return on government bonds by $F_G(X)$ and the cumulative distribution of the rates

[7] Broske, Mary S. and H. Levy, "The Stochastic Dominance Estimation of Default Probability," in Thomas B. Fomby and Tae Kun Seo (eds.), *Studies in the Economics of Uncertainty*, in honor of Josef Hadar, Springer Verlag, New York, pp. 91–112, 1989.

of return on investment in Aaa bonds by $F_{Aaa}(X)$. The investor considers putting \$1 either in $F_G(X)$ or in $F_{Aaa}(X)$. For any finite holding period (e.g., one month or one year), the risk involved in each investment consists of two main components: (i) the risk of changes in the rate of interest, and (ii) the risk of default. As Broske and Levy are interested in measuring only default risk, they neutralize the effects of type (i) risk. This is accomplished by comparing the distributions of rates of return of the two types of bonds while holding maturity (or alternatively, duration) and all other relevant factors, except default risk, constant.

Given the above two cumulative distributions, $F_G(X)$ and $F_{Aaa}(X)$, it can be expected that if type (i) risk is held constant, the investor will pay a higher price for $F_G(X)$ because it is default-free. The relatively lower price for investment $F_{Aaa}(X)$ implies that the holding period rates of return on this investment will be higher than the comparable holding period rates of return on government bonds.

However, taking into account default risk, on an *ex-ante* basis, it is expected that neither $F_G(X)$ nor $F_{Aaa}(X)$ will dominate the other by SSD. This is not the case with *ex-post data*. By using *ex-post* rates of return, only bonds of firms that do not default are analyzed, resulting in $F_{Aaa}(X)$ dominating $F_G(X)$ by SSD or even by FSD. In other words, when investigating corporate bonds, two conditions have to be distinguished:

θ_1 – no default, in which the investor obtains an observation drawn at random out of $F_{Aaa}(X)$ as observed in the past.

θ_2 – default, a case in which the investor gets either zero return or some compensation depending on the severity of the default.

With *ex-post* data, only firms that do not default during the period covered in the study are examined. Because *ex-post* data include only firms that do not default, this data apply only to state θ_1 and state θ_2 is not explicitly represented in the data. Broske and Levy take the *ex-post* data and incorporate state θ_2 in the following manner: As explained above, with *ex-post* data it is expected that $F_{Aaa}(X)$ will dominate $F_G(X)$. By assigning a probability to condition θ_2, Broske and Levy derive a new distribution $F^*_{Aaa}(X)$ from $F_{Aaa}(X)$. Then this probability is changed until neither $F_G(X)$ nor $F^*_{Aaa}(X)$ dominates by SSD. Assuming that the market is efficient and is in equilibrium, and that investors are risk averters, neither $F_G(X)$ nor $F^*_{Aaa}(X)$ dominates by SSD. The probability derived from the no-dominance condition is the upper limit of the risk of default of the Aaa bond as assessed by the market.

Applying the theoretical model to market data, we find that the probability of bankruptcy of Baa bonds is almost twice that of the corresponding probability of Aaa bonds. For a graphical illustration of the relationship between $F_G(X)$, $F_{Aaa}(x)$ and

$F^*_{Aaa}(X)$, and a more detailed analysis, see Broske and Levy (1984).[8] For another study using SSD in the bond market see Chaing (1987).[9]

8.4 OPTION EVALUATION, INSURANCE PREMIUM AND PORTFOLIO INSURANCE

The option valuation model developed by Black & Scholes (B&S) (1973)[10] is undoubtedly one of the most important contributions to modern finance. However, according to this model, there can be only one equilibrium price for the option on which all will agree, resulting in no trade in the option at all. For example, if the observed call price falls below the B&S value, all investors will want to buy the option and no one will want to sell it; hence, the price will go up (if allowed to go up with no transaction) with no actual transaction. This situation is, of course, not realistic. Moreover, in order to obtain the B&S option valuation model, it must be assumed that investors trade continuously, that there are no transaction costs, and that the proceeds from short sales can be held by the short seller. When there are transaction costs and investors cannot continuously hedge their portfolios, the B&S valuation model is not intact. An alternative explanation of option pricing and option trading is that trade is conducted in a discrete market (due to transaction costs) and the investor considers whether to buy the option or the underlying asset (or both). In this framework, SD can be used to find upper and lower bounds on the option value and, within this range, trading can take place between investors depending on their preferences. However, if the option price falls outside this range, economic forces will push it back within the bounds. Let us elaborate on these bounds.

Employing SSD, Levy (1985[11] and 1988[12]) obtains bounds on the option values where the investor is allowed to hold either the call option or the underlying asset. In this framework, taxes and transaction costs can be easily incorporated. The bounds obtained are:

Lower bound C_L: $$C_L = S_0 - X/r + 1/rP_c \int_0^{P_1} (X - Q(t)dt,$$

Upper bound C_U: $$C_U = S_0 / E(\tilde{S}_T) \int_X^{\infty} (\tilde{S}_T - X) f(\tilde{S}_T)\, d\tilde{S}_T,$$

[8] See Footnote 7.

[9] Chaing, R., "Some Results on Bond Yield and Default Probability," *Southern Economic Journal*, 53 1987, pp. 1037–1051.

[10] Black, F. and Scholes, M., "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, May/June 1993.

[11] Levy, H., "Upper and Lower Bounds of Put and Call Option Value: Stochastic Dominance Approach," *Journal of Finance*, 40, 1985, pp. 1197–1217.

[12] Levy, H., "Option Valuation Bounds: A Comparative Analysis," *Studies in Banking and Finance*, 5, 1988, pp. 199–220.

where S_0 is the current stock price, X is the exercise price, $\tilde{S}_T$ is the stock price at maturity (a random variable), r is the riskless interest rate, $f\tilde{S}_T)$ is the density function of $\tilde{S}_T$, $P'=P_r(S_T \leq X)$, and P_c is the value that solves the integral:

$$\int_0^{P_c}(Q_c(t)-r)dt=0,$$

where $Q_c(t)$ is the quantile of the distribution of the rate of return on the call option. Under the conditions of this model, in equilibrium, $C_L \leq C_M \leq C_U$, where C_M stands for the market price of the call option. Levy (1985) shows that the value obtained by the B&S model always falls within the bounds (C_L, C_U). For example, for $\sigma = 0.20$, $\mu = 0.10$, $S_0 = \$100$, and $r = 0.03$, the bounds are $C_L = \$7.29$, $C_U = \$14.48$, whereas the B&S value is $C_{B\&S} = \$9.41$ (see Levy, 1988).

Whereas Levy (1985) employs SSDR, searching for all possible combinations of the risky asset with the riskless asset, Perrakis and Ryan (P&R)[13] (1984) derive bounds on the option value by avoiding dominance of three specific portfolios:

(a) one share of the stock at price S_0,

(b) one call at price C_0 and $S_0 - C_0$ is invested in bonds,

(c) S_0/C_0 call options on the stock.

Because not all possible combinations with the riskless asset are considered, the bounds obtained by P&R are wider than those obtained by Levy. Indeed, using the above example, the P&R bounds are $C_L = \$6.38$ and $C_U = \$14.48$. For other studies using various techniques to derive option bounds see, for example, Perrakis (1986,[14] and 1997[15]), Ritchken (1985)[16] and Ritchken and Kuo (1988).[17]

Note that once the market value of the call falls outside the range (C_L, C_U), there is SSDR dominance of one investment over the other. When $C_M < C_L$, the option dominates the stock. Similarly, when $C_M > C_U$, the stock dominates the option. Although this is only a partial equilibrium relationship (because other risky assets are ignored), it can be usefully employed in practice. For example, the put-call

[13] Perrakis, S. and P. Ryan, "Option Pricing Bounds in Discrete Time," *Journal of Finance*, 39, 1984, pp. 519–525.

[14] Perrakis, S., "Option Pricing Bounds in Discrete Time: Extensions and Price of the American Put," *Journal of Business*, 59, 1986, pp. 119–142.

[15] Perrakis, S., "Pricing and Replication of Short-Lived Index Options Under Transaction Costs," Working Paper, University of Ottawa, 1997.

[16] Ritchken, P.H., "On Option Pricing Bounds," *Journal of Finance*, 1985, pp. 1219–1233.

[17] Ritchken, P.H., and S. Kuo, "Option Bounds with Finite Revision Opportunities," *Journal of Finance*, 43, 1988, pp. 301–308.

parity relationship can be employed to derive the condition under which portfolio insurance dominates the strategy of holding an uninsured portfolio.

By the put-call parity, we have: C = S + P – X/r where C is the call price, P is the put price, X is the exercise price, S is the stock price, and r is the riskless interest rate. If $C < C_L$, then the portfolio S + P – X/r will dominate S. But SSDR implies that S + P (plus lending or borrowing) dominates S (plus lending and borrowing). However, S + P is just a portfolio insurance strategy. Thus, the bounds given by Levy (1985) also reveal the condition under which a 100% portfolio insurance is optimal. If S + P dominates S, 100% portfolio insurance will be optimal.

Azriel Levy (1988),[18] extending this result by also analyzing the condition for optimal fractional insurance (i.e., holding X% of a put option for one stock held), found that the bounds given by Levy (1985) remain intact. In fact, Azriel Levy deals with the general condition under which some risk averters will hold the call option (in combination with the stock) in a long position, and some risk averters will hold the call in a short position. Once again, the equilibrium conditions derived from Azriel Levy (1988) are equal to those of Levy (1985). The issue of portfolio insurance has also been analyzed by Clarke (1988)[19] and by Brooks and Levy (1993)[20] employing a simulation technique to investigate whether an insured portfolio dominates an uninsured portfolio. The results show that unless some specific utility functions are assumed, neither the naked portfolio nor the covered portfolio will dominate.

The option bounds can be applied to determine the maximum premium that an investor will agree to pay for insurance. Kroll (1983)[21] uses a similar argument to find the upper and lower bounds on the premium price for insurance of an asset. If the premium falls within the range established by Kroll, no SSDR dominance will exist between the insured and uninsured strategies.

Recently, a series of studies conducted by Constantinides and Zariphopoulou (1997a[22] and 1997b[23]) and by Constantinidis (1998),[24] analyzed the option bonds

---

[18] Levy, A., "Option Equilibrium in an Incomplete Market with Risk Aversion," Working Paper, Bank of Israel and Hebrew University, 1988.

[19] Clarke, R.G., "Stochastic Dominance of Portfolio Insurance Strategies," Working Paper, Brighham Young University, 1988.

[20] Brooks, R. and Levy, H., "Portfolio Insurance: Does it Pay?" *Advances in Future and Option Research*, 1993, JAI Press, pp. 329–353.

[21] Kroll, Y., "Efficiency Analysis of Deductible Insurance Policies," *Insurance, Math. and Economics*, 2, 1983, pp. 119–137.

[22] Constantinides, G. M. and T. Zariphopoulou, "Bounds on Prices of Contingent Claims in Intertemporal Economy with Proportional Transaction Cost and General Preferences," 1994, Working Paper, 1997a, University of Chicago and University of Wisconsin, Madison.

[23] ——————————," Bounds on Option Prices in an Intertemporal Setting with Proportional Transaction Costs and Multiple Securities," Working Paper, 1997b.

[24] Constantinides, G.H., "Transaction Costs and the Volatility Implies by Option Prices," Working Paper, January 1998, University of Chicago.

in a model in which proportional transaction costs and some restrictions on preferences (on the relative risk aversion coefficient) are imposed.

8.5 APPLICATION OF SD RULES IN AGRICULTURAL ECONOMICS

It is well known that one of the main disadvantages of SD analysis (in comparison to M-V analysis) is that an algorithm to find the SD efficient *diversification* strategies has yet to be developed. We can find some efficient diversification strategies but not *all* of the efficient diversification strategies. This disadvantage is virtually irrelevant in applying SD to agricultural economics. A farmer with a given piece of land has a finite number of irrigation methods and he cannot mix them continuously as in portfolio construction. Of course, if he has many farms he may consider applying one irrigation method on farm A and another on farm B, etc. Even in this case, the number of irrigation schemes to be compared is finite (unlike portfolio diversification which is characterized by an infinite number of combinations). In such cases, the application of SD criteria will be superior to any other method because it is distribution-free and makes only minimal assumptions regarding preferences. Similar advantages of the SD framework pertain to the measurement of income inequality (see Atkinson, 1970),[25] the choice of the best advertising strategy by a firm, choosing the best medical treatment (see Stinnett and Mullahy[26]) etc. In this section, we discuss the application of SD to agricultural economics, and in sections 8.6 – 8.8 we discuss the applications of SD to other fields.

Stochastic dominance is widely employed in problem solving in agricultural economics and related areas. It has been used in evaluating alternative stocking rate tactics (see Riechers et al., 1986)[27], in analyzing the efficiency of export earnings (see Gan et al., 1988)[28], in choosing the most efficient crop insurance (Williams, 1988)[29], in evaluating the risk of various agricultural products (Lee et al., 1987)[30] etc. This application of SD to agricultural economics will be illustrated here by one representative study: Harris and Mapp (1986)[31] employ stochastic dominance toanalyzevarious water-conserving irrigation strategies. A computerized plant

---

[25] Atkinson, A. B., "On the Measurement of Inequality," *Journal of Economic Theory*, 2, 1970, pp. 244–263.

[26] Stinnett, A. and Mullahy, J., "Net Health Benefits: A New Framework for the Analysis of Uncertainty in Cost-Effectiveness Analysis," Working Paper, 1997.

[27] Riechers, R.K., J.G. Lee and R.K. Heitschmidt, "Evaluating Alternative Stocking Rate Tactics. A Stochastic Dominance Approach," *American Journal of Agricultural Economics*, 70, 1986.

[28] Gan, C., R.B. Wharton and T.P. Zacharias, "Risk Efficiency Analysis of Export Earnings: An Application of Stochastic Dominance," *American Journal Agricultural Economics*, 70, 1988.

[29] Williams, J.R., "A Stochastic Dominance Analysis of Tillage and Crop Insurance Practices in a Semi-Arid Region," *American Journal of Agricultural Economics*, 70, 1988, 112–120.

[30] Lee, J.B., R.D. Lacewell and J.R. Ellis, "Evaluation of Production and Financial Risk: A Stochastic Dominance Approach," *Canada Journal of Agricultural Economics*, 35, 1987, pp. 109–126.

[31] Harris, Thomas R., and H.P. Mapp, "A Stochastic Dominance: Comparison of Water-Conserving Irrigation Strategies," *American Journal of Agricultural Economics*, 68, 1986.

growth model for sorghum grain using daily weather observations provides the basis for the stochastic dominance analysis. Input data for the plant growth model include daily precipitation, maximum and minimum temperature, and solar radiation for the period May through October. Data corresponding to the gross revenue is obtained by multiplying the yield by the price obtained for the grain. Irrigation costs vary depending on the quantity of water used for irrigation. Each irrigation strategy is replicated 23 times based on the 23 years of available data. The resulting 23 net returns are used for the stochastic dominance comparisons.

Table 8.1 provides the main results. As can be seen, the intensive irrigation method used in practice is dominated by FSD by the seven alternative irrigation methods and one alternative strategy dominates it by SSD. Thus, switching to another irrigation method constitutes a FSD improvement, when the risk is being determined by climate variations across years. Apart from showing that there are irrigation schemes that dominate the current intensive irrigation practice, the authors also employ pairwise FSD, SSD and TSD comparisons to find the efficient set of irrigation schemes, thereby eliminating some of the strategies reported in Table 8.1.

Degree of Stochastic Dominance among Water-Conserving Irrigation Strategies and the Current Practice of Intensive Irrigation

Irrigation strategy	Expected Net return ($/ac)	Standard Deviation Of net return ($/ac)	Mean Yield (cwt/ac)	Mean water application (ac. in./ac)	Degree of stochastic dominance
Intensive irrigation	78.86	24.77	59.20	24.00	
Irrigation by soil water ratio	93.94	21.74	59.04	14.09	FSD over intensive Irrigation
No irrigation in:					
Stage 1[a]	94.23	21.71	59.02	13.89	FSD over intensive Irrigation
Stage 2	92.76	22.00	58.58	13.70	FSD over intensive Irrigation
Stage 3	93.33	21.07	58.64	13.50	FSD over intensive Irrigation
Stage 4	91.29	19.20	57.19	11.15	SSD over intensive Irrigation
Stages 1 and 2	92.66	23.20	58.18	12.91	FSD over intensive Irrigation

Stages 1 and 3	93.55	21.20	58.62	13.30	FSD over intensive Irrigation
Stages 1 and 4	91.39	19.27	57.14	10.96	SSD over intensive Irrigation
Stages 2 and 3	88.12	22.43	56.78	12.13	No dominant strategy
Stages 2 and 4	77.91	30.22	52.74	9.78	FSD by intensive Irrigation
Stages 3 and 4	52.32	40.37	46.23	8.22	FSD by intensive Irrigation
Stages 1, 2, and 3	84.42	22.35	55.57	11.55	No dominant strategy
Stages 2, 3, and 4	25.60	59.79	38.57	4.87	FSD by intensive Irrigation
Stages 1,2,3 and 4	11.43	64.28	34.55	4.70	FSD by intensive Irrigation

[a] Growth stages 1 through 4 refer to emergence to floral initiation, differentiation to the end of leaf growth, end of leaf growth to half-bloom, and half-bloom to physiological maturity, respectively. Sensitivity to soil water stress increases as the plant moves from stage 1 through stages 2, 3, and 4.

Source: Harris and Mapp (1986), see footnote 31.

8.6 APPLICATION OF SD RULES IN MEDICINE

Statistical methods are employed in the selection of alternative health interventions. Data on costs and health effects are commonly used to analyze the cost-effectiveness (CE) ratio of two treatments, say treatment A_1 and A_2, and the confidence intervals of these ratios can then be used to help choose between the two treatments. Recently, Stinnett and Mullahy[32] suggest that FSD and SSD rules be applied to selection between treatments A_1 and A_2. They first define net health benefit (NHB) and then suggest to apply FSD and SSD by comparing the cumulative distributions, F(NHB) to G(NHB) compared, where F and G denote two alternative health interventions, A_1 and A_2. Although many problems have yet to be solved (e.g., how to quantify suffering and death in monetary terms), we believe that we will see many more applications of SD in the medical area and in drug development. For example, suppose that a firm wishes to invest in the research and development (R&D) of a new drug (and for that matter any R&D project): Various strategies will be considered (e.g., independent R&D teams or one large team with a full flow of information) and SD rules will be employed to decide on the most efficient R&D strategy (for more details see Arditti & Levy [1980]).[33] Similarly, in the area of auditing, SD rules may be used to decide on how many teams should audit a firm and the degree of dependency allowed between these teams (see Barlev and Levy, 1996).[34]

[32] See footnote 26.

[33] Arditti, F. and H. Levy, "A Model for Parallel Team Strategy in Product Development," *American Economic Review*, 70, No. 5, December 1980, pp. 1089–1097.

[34] Barlev, B. and Levy, H., "Misuse and Optimum Inspecting Strategy in Agency Problems," *Meiroeconomica*, 1996, pp. 82–104.

8.7 MEASURING INCOME INEQUITY

Economists analyzing income distributions may be interested in questions such as whether the distribution of income is more equal than it was in the past, or whether some countries are characterized by greater income inequality than others, or whether taxes lead to greater equality of income distributions.

Inquiry of this type commonly draws on summary statistics such as the Gini coefficient or the coefficient of variation. Atkinson (1970)[35] suggests that SD rules be employed to analyze such issues. Denoting social welfare by W and the individual utility from income y by U(y), he defines:

$$W=\int_0^{\bar{y}} U(y)f(y)dy$$

where $\bar{y}$ is some upper value that income can be obtained. If there are two policies leading to two income distributions F and F^*, then Atkinson suggests that F will be preferred to F^*, if and only if, F dominates F^* by SSD. Thus, if dominance exists, without specifying U(y), it is possible to determine a preferred income distribution. Atkinson proves that if F dominates F^* by SSD, then the familiar Lorenz curve corresponding to F will always be located above the curve corresponding to F^*. He also analyzes the effect of MPS on income distribution and the Lorenz curve. It would be of interest to extend this line of investigation by analyzing whether there is a relationship between the Lorenz curve and the TSD criterion, and to examine the effect of MPSA on the Lorenz curve. Moreover, if all individuals can borrow and lend money, SDR criteria rather than SD criteria will be more appropriate. Suppose that F does not dominate F^* by SSD and that it does dominate F^* by SSDR. It would be interesting to investigate the relationship between SSDR dominance and the Lorenz curve and the economic interpretation of such dominance in the Lorenz curve framework.

8.8 APPLICATION OF SD RULES IN THE SELECTION OF PARAMETER ESTIMATORS

The expected value of a distribution can be estimated using the mode, the mean or the median. Which of these estimates is best? Selection among these estimators is not always an easy task because each estimator has different characteristics (e.g., unbiasedness, maximum likelihood, etc.).

The statistical literature dealing with the loss function induced by potential error and the selection of estimators is vast. Here we will briefly discuss some issues related to the selection of various estimators that are relevant to stochastic dominance.

---

[35] See footnote 25.

Ben-Horin and Levy (1982)[36] apply stochastic dominance rules to the evaluation of alternative estimators of the variance of a normal distribution. This application is motivated by the notion that any deviation of an estimate from its corresponding population parameters will be accompanied by monetary loss; the higher the deviation, the larger the assumed loss. Therefore, estimators are first screened to determine which of them is efficient according to FSD, SSD and TSD dominance. The estimators of the variance of a normal distribution were chosen to demonstrate how this approach can be applied, in general, to the selection of statistical estimators. Specifically, the following estimators were considered:

$$S_k^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{k}$$

for k = n − 1, n, n+1. The estimators have a χ^2 distribution; hence, the cumulative probability distribution of the error is easily derived. SD rules are then applied to discriminate between estimators. Note that although the distributions are normal, the deviation of the estimate from the true parameter is not distributed normally. Therefore, the mean-variance rule is invalid in this case and stochastic dominance rules should be employed.

For a linear loss function, the mean squared error estimator S_{n+1}^2 is found to be inferior (by SSD) to the maximum likelihood estimator S_n^2. Thus, if the decision maker is a risk averter, and the monetary loss is a linear function of the deviation from the true parameter, then the maximum likelihood estimator will be a better estimator than S_{n+1}^2. However, both the maximum likelihood estimator (S_n^2) and the unbiased estimator (S_{n-1}^2) are found to be undominated by FSD and SSD rules. In other words, using both S_n^2 and S_{n-1}^2 is justified because neither is superior. This result holds for small and large samples alike. In another paper, Ben-Horin and Levy (1984)[37] apply the SD approach to parameter estimation of stable distributions. (For more details of this and similar issues, see Ben-Horin and Levy, 1982 and 1984).

Applying SD and SDR rules in statistics still has a long way to go. It is well known that the selection of significance level, α, in hypothesis testing is quite arbitrary. The monetary loss of Type I errors and Type II errors can be specified and a vector of SD or SDR efficient significant levels, α, calculated. SD and SDR can also be applied to many other widely-used estimates in statistics.

---

[36] Ben Horin, M., and H. Levy, "Evaluating Estimators Using Stochastic Dominance Rules: The Variance of a Normal Distribution," *Comm. Statis.*, 1982.

[37] Ben Horin, M., and H. Levy, "Stochastic Dominance and Parameters Estimation: The Case of Symmetric Stable Distribution," *Insurance Math. and Economics*, 3, 1984, pp. 133–138.

8.9 SUMMARY

Stochastic dominance is most commonly applied in finance and the economics of uncertainty (e.g., capital structure, portfolio diversification, defining risk, estimating bankruptcy risk, and determining option's price bounds). Nevertheless, SD applications to portfolio selection have not exploited their potential because SD algorithms to construct a portfolio of risky assets have yet to be developed. However, in other areas, the portfolio issue does not exist (one action rather than a mix of actions is selected); hence, the application of SD is straightforward. It is applied in areas such as agriculture (choosing the best irrigation system), statistics (finding the efficient estimators), and in medicine (selecting an efficient treatment).

Key Terms

Capital Structure

Levered Firm

Unlevered Firm

Investment-consumption

Ex-post data

Ex-ante data

Bankruptcy

Default

Black & Scholes Model

Lower Bound

Upper Bound

Put-Call Parity

9

STOCHASTIC DOMINANCE AND RISK MEASURES

There are various measures of risk and each of them has its pros and cons. This chapter focuses on the notion of risk in the stochastic dominance framework. We first discuss the concept of *mean preserving spread* (MPS) and risk suggested by Rothschild and Stiglitz (R&S) (1970)[1] and then extend it to the case where riskless asset exists. Finally, we discuss the *mean-preserving spread antispread* (MPSA) which is a risk measure similar to the MPS but which corresponds to DARA risk-averse utility functions. Thus, the MPS corresponds to SSD and the MPSA corresponds to TSD.

In the next section, we review several risk measures, and show that five of them are equivalent and the sixth, which is commonly used as a measure of risk, is not.

9.1 WHEN IS ONE INVESTMENT RISKIER THAN ANOTHER INVESTMENT?

Generally, investors select investments by comparing profitability as well as risk. In order to focus on risk, Rothschild and Stiglitz, in their analysis assume equal means and they also assume, for simplicity, that the random variables are bounded by the range [0,1]. R&S provide five definitions of risk (definitions a, b, c, d and f, below, and we add definition e). To be more specific, R&S do not suggest a quantitative index for risk but rather ask the following question: When is one investment "riskier than" another? Let us first provide their definitions

[a] *y is riskier than x if y is equal to x plus "noise":* If an uncorrelated noise is added to a random variable x such that:

$$y \underset{d}{=} x+z$$

where "d" denotes "the same distribution as" and z has the property $E(z/x) = 0$ for all x, then y will be riskier than x (note that x and y have the same mean). For example, suppose that x yields \$1 or \$5 with equal probability. In addition, suppose that if x = \$1, then z = 0 and if x = \$5, then z = ± \$1 with equal probability. According to R&S, in this example y will be riskier than x.

[b] *Risk Aversion*: y will be riskier than x if both have the same mean and for every $U \in U_2$ (risk aversion) $EU(w+x) \geq EU(w+y)$ for every constant w.

[1] Rothschild, M. and J.E. Stiglitz, 1970, "Increasing Risk: I. A Definition, *Journal of Economic Theory*, 1970, pp. 225–343.

[c] *The "fat tails" criterion or MPS:* y will be riskier than x if the density function of y has more weight in the "tails" than the density function of x. This criterion is not rigorously defined and, therefore, R&S defined mean preserving spread (MPS), a technique to shift the density from the center to the tails, hence increasing the random variable risk. Thus, the "fat tails" criterion is rigorously formalized by the MPS definition. We will discuss MPS in greater detail later on.

[d] *The integral criterion:* y (with cumulative distribution G) will be riskier than x (with cumulative distribution F) if the integral condition is satisfied, namely:

$$\int_0^z [G(t)-F(t)]dt \geq 0 \quad \text{for every z (with at least one strict inequality)}$$

and the expected values of F and G are identical.

[e] *Risk premium:* y will be riskier than x if for all $U \in U_2$, if $\pi_y \geq \pi_x$ where π_x and π_y are the corresponding risk premium solving the equations:

$EU(w+x) = U(w+Ex-\pi_x)$, and $EU(w+y) = U(w+Ey-\pi_y)$, for every constant w, where w is the initial wealth and $U \in U_2$. This criterion follows naturally from the previous five definitions.

[f] *The variance criterion*: By this common definition, y will be riskier than x if both have the same means and y has a larger variance than x.

The equivalence between risk definitions [b] and [d] has already been demonstrated in Chapter 3 by Theorem 3.2 for any two distributions with no constraints on the means, and the proof remains intact also for the special case of equal means. Also, in condition [d] one can eliminate the constant w without affecting the inequality condition. The equivalence between risk definitions [e] and [b] is also straightforward. If y is riskier than x, by definition [b], $EU(w+x) \geq EU(w+y)$ for every w and, therefore, the following will also hold: $U(w+Ex-\pi_x) \geq U(w+Ey-\pi_y)$ for every w. By assumption, we have $E_x = E_y$ and U is a non-decreasing function of x, therefore, $\pi_y \geq \pi_x$. The equivalence between risk definitions [a] and [b] is also straightforward. Recall that y = x+z and E(z/x) = 0 for every fixed x. By definition [b], U is concave.

Taking expectation with respect to z (for a fixed x), due to the concavity of U and the fact that x is fixed with Ex = x, we have:

$$E_xU(x+z) \leq U(E(x) + E(z)) = U(x).$$

where the subscript x denotes that x is fixed.

Taking expectation with respect to x yields $EE_xU(x+z) \le E(U(x))$ or:

$$EU(x+z) = EU(y) \le EU(x)$$

which confirms the equivalence between [a] and [b].

To show that the integral condition [d] is equivalent to the MPS definition [c] (shifting density from center to tails), let us first define mean preserving spread (MPS).

9.2 MEAN PRESERVING SPREAD (MPS)

For simplicity only, let us assume that we are dealing with a continuous random variable. Using the notation used by R&S (1970), let s(x) be a step function defined by:

$$s(x)=\begin{cases} \alpha>0 & \text{for } a<x<a+t \\ -\alpha\le 0 & \text{for } a+d<x<a+d+t \\ -\beta\le 0 & \text{for } b\le x<b+t \\ \beta\ge 0 & \text{for } b+e<x<b+e+t \\ 0 & \text{oterwise} \end{cases} \tag{9.1}$$

where $0 \le a \le a+t \le a+d \le a+d+t \le b \le b+t \le b+e \le b+e+t \le 1$

and α, β, e and d are determined such that the following holds, $\beta e = \alpha d$. The density shift described by eq. (9.1) is called MPS. The shift in density function is the spread (of probability) and, because the mean is preserved, it is called MPS. To be more specific, the condition $\beta e = \alpha d$ guarantees that the density shift will not change the distribution mean. This MPS and its effects on f(x) and F(x) are illustrated in Figures 9.1a,9.1b, 9.1c, and 9.1d.

Figure 9.1a presents the density function f(x), Figure 9.1b, the MPS, Figure 9.1c, the density function g(x) = f(x) + s(x) and, finally, Figure 9.1d, the integral $\int_a^x S(t)dt$ where S is the integral of s given by eq. (9.1). Let us elaborate.

From the definition of s(x) (see eq. (9.1)), we can conclude that:

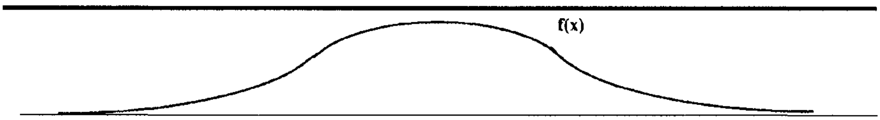
I. $$\int_0^1 s(x)dx=0\equiv S(1)$$ (see Figure 9.1b)

and:

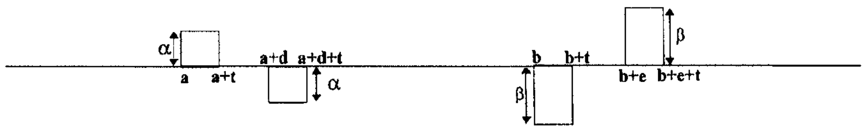
II. $$\int_0^1 xs(x)dx=0.$$

Conclusion I is obtained by the construction of the MPS. Conclusion II follows from the fact that:

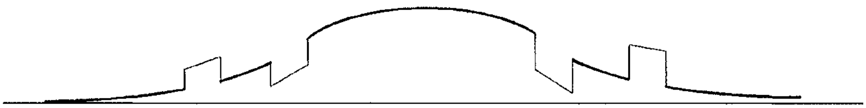
Figure 9.1: MPS and its effects on the density function



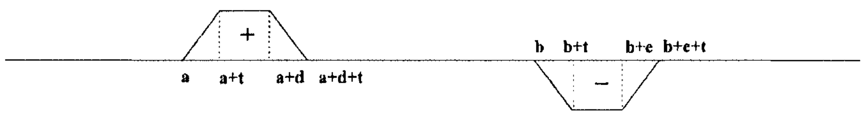
9.1a: The Density Function f(x)



9.1b: The MPS, s(x)



9.1c: g(x) = f(x) + s(x)



9.1d: The Integral $\int_a^x s(t)dt \geq 0$

$$\int_0^1 xs(x)dx = \frac{\alpha}{2}[(a+b)^2 - a^2] - \frac{\alpha}{2}[(a+d+t)^2 - (a+d)^2] - \frac{\beta}{2}[(b+t)^2 - b^2]$$
$$+\frac{\beta}{2}[(b+e+t)^2 - (b+e)^2] = \frac{\alpha}{2}[2at+t^2] - \frac{\alpha}{2}[2(a+d)t+t^2] - \frac{\beta}{2}[2bt+t^2]$$
$$+\frac{\beta}{2}[2(b+e)t+t^2] = (-\alpha d+\beta e)t$$

Because $\alpha d = \beta e$ (by construction of the MPS), $(-\alpha d + be)\, t = 0$ and condition II above is intact.

Thus, if f(x) is a density function, then g(x) = f(x) + s(x) is a density function, too, as long as $g(x) \geq 0$ for all x because $\int_0^1 g(x)dx = \int_0^1 f(x)dx + \int_0^1 s(x)dx = 1 + 0 = 1$. The constraint $g(x) \geq 0$ simply states that s(x) can be shifted only from ranges where $f(x) \geq 0$ and where $f(x) \geq s(x)$.

The addition of a function s(x) to f(x), shifts the probability from the center of f(x) to its tails (or to one of the tails) without affecting the mean. Such a probability shift, as described by eq. 9.1, is called MPS. If the functions f(x) and g(x) are density functions, then we say that g(x) differs from f(x) by a single MPS.

It is easy to verify that if g(x) differs from f(x) by a single MPS, then the integral condition holds; hence, risk definitions [c] and [d] are also equivalent. To see this claim, recall that by the MPS definition, we have g(x) = f(x) + s(x); hence, G(x) = F(x) + S(x), where $S(x) = \int_0^x s(t)dt$. Therfore, $\int_0^x [G(t) - F(t)]dt = \int_0^x S(t)dt \geq 0$ for all values x, and there is a strict inequality for some value x, and for x=1, we get $\int_0^1 S(t)dt = 0$. To see why $\int_0^1 S(t)dt \geq 0$ for all values x, recall that $\int_0^1 [G(x) - F(x)]dx = \int_0^1 S(x)dx = 0$ (because the means are preserved. See Chapter 3, Section 3.3b). Thus, by the equal means constraint of the MPS, at the value x = 1, we have $\int_0^1 S(t)dx = 0$, which implies that the "+" area in Figure 9.1d is equivalent to the "–" area. Therefore, for 0 < x < 1, $\int_0^x [G(t) - F(t)]dt \geq 0$ (see Figure 9.1d).

R&S also prove that if there are two cumulative distributions F and G with equal means, and if G(x) – F(x) satisfy the integral condition (F dominates G by SSD), then sequences F_n and G_n exist such that $F_n \to F$, and $G_n \to G$ and, for each n, G_h could have been obtained from F_n by a finite number of MPS's.

Unfortunately, in their original proof, the MPS used by R&S to shift density from F to G is incorrect because G_n is a declining function of x, namely G_n is not a cumulative probability function, therefore, all the R&S proofs that rely on the non-decreasing property of the cumulative distribution are not intact. Fortunately, Leshno, Levy and Spector (1997)[2] show that by adopting different MPS's to those suggested by R&S, the definition of G_n as a cumulative probability function is guaranteed, and all the results (though not the proofs) of R&S remain intact.

The definition of MPS as a probability shift from the center of the distribution to the tails implies that if F and G have identical means, the addition of MPS's to F will create a new distribution G which will be riskier by the integral condition. Therefore, all four definitions of risk suggested by R&S ([a] to [d]) and the risk premium risk definition ([e]) added in this chapter, are equivalent. If G is riskier than F by one definition, it will also be riskier by all other definitions.

Variance (definition [f] of risk) on the other hand, is generally *not* a measure of risk and, therefore, it is not equivalent to the other five definitions. To see this claim, one counter example will suffice. In Chapter 3 we provide an example with $E_F(x) > E_G(x)$ and $\sigma_F < \sigma_G$; however, risk averters with $U(x) = \ln x$ prefer G to F. This holds *a fortiori* for a lower mean of F such that $E_F(x) = E_G(x)$. Hence, $E_F(x) = E_G(x)$, and $\sigma_F(x) < \sigma_G(x)$ does not imply that $E_FU(x) \geq E_GU(x)$ for all $U \in U_2$. Therefore, ranking investments by their risk on the basis of the variance does not coincide with the other definitions of "riskier than"; hence, it is generally wrong.

9.3 UNEQUAL MEANS AND "RISKIER THAN" WITH THE RISKLESS ASSET

Suppose that we wish to rank two (or more) mutual funds (or any other pair of investments) by their risk. Even if the distributions of their returns are known or estimated with *ex-post* data in general, it will still be impossible to employ R&S's method to rank the mutual funds by their risk because it is highly unlikely that the two investments under consideration will have equal means. As long as $E_F(x) \neq E_G(x)$, R&S's "riskier than" definition is not applicable; it is purely theoretical and not useful in practice.

However, as shown by Levy (1977)[3], when the riskless asset is added, R&S's definition of "riskier than" can be extended to the case of unequal means and can be used to rank mutual funds by their risk. We will focus here on the integral definition of "riskier than" and, because the other four definitions are equivalent to the integral condition, it follows that the extension holds also for the other four definitions of risk.

Suppose that two investments F and G do not have equal means; hence, R&S's method cannot be used to rank F and G by their risk. Let F and G be two risky investments with means $E_F(x)$ and $E_G(x)$, respectively. Without loss of generality, assume that $E_G(x) > E_F(x)$. We now borrow money at the riskless rate r and, from

---

[2] Leshno, M., Levy, H. and Spector, Y., "A Comment on Rothschild and Stiglitz's Increasing Risk: I. A Definition," *Journal of Economic Theory*, 1997, pp. 223–228.

[3] Levy, H., 1977, "The Definition of Risk: An Extension," *Journal of Economic Theory*, 14, 1977, pp. 232–234.

F, create a levered portfolio F_a, where $x\alpha = \alpha x + (1-\alpha)r$ and determine a to fulfill the constraint $E_{F_\alpha}(x)=E_G(x)$. Then

$$\alpha\, E_F(x) + (1-\alpha)r = (E_G(x) \text{ and } \alpha = (E_G(x) - r)/(E_F(x)-r).$$

Clearly, if $E_G(x) > E_F(x)$ (as assumed), $\alpha > 1$. Thus, in this specific case, $F\alpha$ will be a levered portfolio. The next theorem demonstrates that if the means of the two options can be equalized, then R&S's definition of risks can be applied.

Theorem 9.1:

Let F and G be two investments with expected means $E_F(x)$ and $E_G(x)$, respectively. If there is one combination of F with a riskless asset, F_α, where E_{F_α}, $(x)=E_G(x)$, such that G is riskier than F_α, then for any other two combinations, G_β F_δ, with $E_{G_\beta}(x) = E_{F_\delta}(x)$, G_β will be riskier than F_δ. Thus, it is always possible to obtain a less risky position by diversifying F with the riskless asset than by diversifying G with the riskless asset. Hence, F is a less risky investment in spite of the fact that $E_F(x) \neq E_G(x)$.

Proof:

By assumption $E_{G_\beta}(x) = E_{F_\delta}(x)$; hence, $\beta(E_G(x) - r) = \delta(E_F(x) - r)$, or:

$$\delta = \beta[E_G(x) - r] / [E_F(x) - r] = \beta\alpha. \tag{9.2}$$

Given that G is riskier than $F\alpha$, we obtain:

$$\int_{-\infty}^{x} [G(t) - F_\alpha(t)]dt \geq 0, \text{ for all values x.} \tag{9.3}$$

However, F_α can be rewritten as:

$$F_\alpha(x) = P_r(x_\alpha \leq x) = \Pr[(\alpha x + (1-\alpha)r) \leq x] = F[(x - (1-\alpha)r)/\alpha].$$

Substituting $F_\alpha(t)$ in eq. (9.3) we obtain that the following holds:

$$\int_{-\infty}^{x} [G(t) - F\left(\frac{t-(1-\alpha)r}{\alpha}\right)]dt \geq 0 \text{ for all values x.} \tag{9.4}$$

We need to prove that (9.4) implies:

$$\int_{-\infty}^{x} [G_\beta(t) - F_\delta(t)]dt \geq 0, \text{ for all values x.} \tag{9.5}$$

However, because $G_\beta(x) = G[(x - (1 - \beta)r/\beta]$ and

$F_\delta(x) = F[(x - (x - (1 - \delta)r)/\delta]$, eq. (9.5) can be rewritten as:

$$\int_{-\infty}^{x}[G\left(\frac{t-(1-\beta)r}{\beta}\right)-F\left(\frac{t-(1-\delta)r}{\delta}\right)]dt \geq 0, \text{ for all values x.} \qquad (9.6)$$

Recall that by eq. (9.2), $\delta = \beta\alpha$. Substituting $\beta\alpha$ for δ in eq. (9.6) we need to prove that,

$$\int_{-\infty}^{x}[G\left(\frac{t-(1-\beta)r}{\beta}\right)-F\left(\frac{t-(1-\alpha\beta)r}{\alpha\beta}\right)]dt$$
$$= \int_{-\infty}^{x}[G\left(\frac{t-(1-\beta)r}{\beta}\right)-F\left(\frac{[t-(1-\beta)r]/\beta-(1-\alpha)r}{\alpha}\right)]dt \geq 0 \qquad (9.7)$$

for all values of x. Thus, it is sufficient to show that eq. (9.3) implies eq. (9.7). In eq. (9.7), conduct the transformation, $u = [(t - (1 - \beta)r)/\beta]$ with $dt = \beta du$ to obtain

$$\beta\int_{-\infty}^{(x-(1-\beta)r/\beta}[G(u)-F\left(\frac{u-(1-\alpha)r}{\alpha}\right)]du \geq 0. \qquad (9.7')$$

However, $\beta > 0$, therefore, eq. (9.3) implies eq. (9.7′) (or eq. 9.7).

Thus, all that we need to do is to equate the two means of the two random variables at any arbitrary level and then apply the risk measure suggested by R&S. If, on the basis of the chosen arbitrary mean level, one option is found to be riskier than the other, it will be riskier for any other level of selected means. In other words, suppose that by R&S, G is riskier than Fa. If one also wishes to mix G with the riskless asset to create G_β, there is another mix $\delta = \alpha\beta$ such that G_β is riskier than F_δ. Thus, we can conclude that the set $\{G_\beta\}$ (for all β's) is riskier than the set $\{F_\alpha\}$ in the R&S's sense. Namely, for each G_β in the set $\{G_\beta\}$ there is Fa in the set $\{F_\alpha\}$ with the same mean and $F\alpha$ dominates G_β by SSD.

Example:

Let distributions F and G be as follows:

	F		G	
	x	p(x)	y	p(y)
	5%	1/2	4%	½
	9%	1/2	12%	½
Expected value	7%		8%	

The means are not equal and, therefore, we cannot rank F and G by their risk. However, suppose that the interest rate is r = 4%. Borrow:

$$\alpha=(E_G(x)-r)/(E_F(x)-r)=\frac{8-4}{7-4}=\frac{4}{3}=1\tfrac{1}{3}.$$

Therefore, F_α will be:

$1\frac{1}{3}\cdot 5\% - \frac{1}{3}\cdot 4\% = 6\frac{2}{3} - 1\frac{1}{3} = 5\frac{1}{3}\%$ with a probaility of $\frac{1}{2}$,

$1\frac{1}{3}\cdot 9\% - \frac{1}{3}\cdot 4\% = 12 - 1\frac{1}{3} = 10\frac{2}{3}\%$ with a probaility of $\frac{1}{2}$,

It is easy to verify that: $E_{F_\alpha}(x) = E_G(x)$ (= 8%) and that $\int_{-\infty}^{x}[G(t)-F_\alpha(t)]dt \geq 0$ for all values x (and there is at least one strict inequality). Thus, $\{G_\beta\}$ will be riskier than $\{F_\alpha\}$ by Levy's definition (1977).

9.4 "RISKIER THAN" AND DARA UTILITY FUNCTION: MEAN PRESERVING ANTISPREAD

So far, we have defined the "riskier than" concept where $U \in U_2$. We have seen that if a single MPS is added to F, a new distribution G will be created where G = F + MPS and G will be riskier than F in the sense that $E_F U(x) \geq E_G U(x)$ for all $U \in U_2$. Suppose now that we add the assumption that $U \in U_d$ where the subscript d denotes DARA utility functions; In U_2, the two distributions cannot be ranked by their risk, but in U_d one distribution may be riskier than the other. This stems from the fact that dominance can be established in U_d even though there is no dominance in U_2 (recall that $U_2 \supset U_d$).

"Riskier than" for DARA function is equivalent to TSD; before proceeding with this definition recall that when $E_F(x) = E_G(x)$, dominance by TSD and by DARA implies each other (see Chapter 3, Section 3.9b). As we are confining ourselves to equal mean distributions, we will examine TSD dominance rather than DARA dominance. We will show that if one distribution differs from another by a Mean Preserving Spread Antispread (MPSA) and there is TSD, then one distribution will be riskier than the other in U_d even though "riskier than" situation cannot be defined in U_2. Let us first define MPS and MPSA with discrete distributions.

a. Spread and Antispread

Assume that we compare two random variables x and y with a cumulative distribution function (c.d.f.) of F(z) and G(z), respectively. For simplicity and without losing generality, as before, assume that the random variables are bounded r.v.s such that F(0) = G(0) = O and F(1) = G(1) = 1.

Let α and β be non-negative numbers and let x_1, x_2, x_3 and x_4 be any real numbers such that $x_1 \le x_2 \le x_3 \le x_4$ The MPS ('spread') function, S(z) (which is the integral of s(z)) is defined as follows:

$$S(z)=\begin{cases} +\alpha & \text{for } x_1 \le x_2 \\ -\beta & \text{for } x_3 \le x_4 \\ 0 & \text{otherwise} \end{cases} \tag{9.8}$$

where $\alpha(x_2 - x_1) = \beta(x_4 - x_3)$.

Similar to the 'spread' function, the mean preserving antispread function A(z) is defined as follows:

$$A(z)=\begin{cases} -\alpha & \text{for } x_1' \le x_2' \\ +\beta & \text{for } x_3' \le x_4' \\ 0 & \text{otherwise} \end{cases} \tag{9.9}$$

where $\alpha(x'_2 - x'_1) = \beta(x'_4 - x'_3)$.

The antispread shifts the probability mass function in the opposite direction to that of the MPS suggested by R&S.

The "mean preserving spread and antispread" (MPSA) function, SA(z), is defined as follows:

$$SA(z) = S(z) + A(z). \tag{9.10}$$

Given the MPA and MPS definitions, we must have that $x'_i \neq x_i$ for at least one i (i=1, 2, 3, 4). If $x'_i = x_i$ for *all* i (i=1, 2, 3, 4), then S(z) cancels the A(z) and we have $SA(z) \equiv 0$ for all x.

Similar to R&S's MPS, if G(z) = F(z) + SA(z), we say that G(z) differs from F(z) by a single MPSA step. We confine the analysis of risk to one MPS and one mean preserving antispread (MPA). Preserving the means is guaranteed if $\alpha(x_2 - x_1) = \beta(x_4 - x_3)$ and $\alpha(x'_2 - x') = \beta(x'_4 - x'_3)$.

Kroll, Leshno, Levy, and Specter (KLLS) (1995)[4] prove that if no constraints are imposed on the relationship between the MPS and the MPA function, any two random variables with the same mean can be deduced from each other by a sequence of MPSAs.

b. Increasing Risk and DARA

In this section we specify the conditions on the MPSA functions that enable the classification of one random variable as 'more risky' than another random variable for all DARA utility functions.

Let us first restate the TSD (or DARA) dominance for equal mean distributions. Assume that all risk averse investors have DARA utility functions. In the case of E(x) = E(y), a necessary and sufficient condition for all DARA investors to prefer x over y is for x to dominate y by TSD (see Chapter 3, section 3.9b). Thus, if E(x) = E(y), y (whose distribution is G) will be riskier than x (whose distribution is F) for all DARA utility functions if and only if:

$$\int_0^x \int_0^v [G(t) - F(t)]dtdv \geq 0 \tag{9.11}$$

for all x in [0,1] and with a strict inequality for at least one x. This dominance condition is employed in Theorem 9.2.

Definition:

An MPSA function SA = S + A is said to satisfy the TSD criterion if:

$$\int_0^x \int_0^v S(t)dtdv \geq \int_0^x \int_0^v A(t)dtdv \tag{9.12}$$

[4] Kroll, Y., Leshno, M., Levy, H., and Spector, Y., "Increasing Risk, Decreasing Absolute Risk Aversion and Diversification," *Journal of Mathematical Economics*, 24, 1995, pp. 537–556.

for all x in [0,1] and with a strict strong inequality for at least one x.

Note that the MPA 'improves' the distribution under consideration in the sense that it increases expected utility whereas MPS decreases the expected utility. The MPSA that satisfies the TSD criterion generates a distribution G when G = F + MPSA. Distribution G is inferior to F for all DARA utilities, but not necessarily inferior for all risk averters.

Theorem 9.2:

(TSD) Let F and G be the distributions of two equal mean random variables, x and y, respectively, bounded by [0,1]. F will dominate G by TSD if and only if there exists a sequence $\{SA_i\}_{i=0}^{\infty}$ of MPSA satisfying the TSD criterion such that $G = F + \sum_{i=1}^{\infty} SA_i$. The proof can be found in KLLS (1995)[5].

What is the difference between risk ranking by R&S and risk ranking by KLLS? The difference is that TSD dominance may exist – even if there is no SSD dominance: if only $U \in U_2$ is assumed, we may be unable to rank the two options by their risk, but if $U \in U_d$ is assumed, such ranking may be possible. Thus, if we create MPSA such that eq. (9.12) holds, and G = F + MPSA, we can safely conclude that G will be riskier than F in $U \in U_d$. Finally, if $E_F(x) \neq E_G(x)$ and a riskless asset is added, as with SSD, it is sufficient to find one F_α such that G is riskier than F_α in U_d (where F_α and G have equal means) to conclude that $\{G_\beta\}$ will be riskier than $\{F_\alpha\}$ in U_d.

[5] See footnote 4.

9.5 SUMMARY

In Chapter 1, we saw that it is very hard to quantify risk. R&S (1970) establish-several equivalent definitions of risk (and variance is not one of them!) that enable us to define one variable as "more risky" than another for equal means distributions. According to these definitions, risk is not quantified, but ranking investments by their risk is enabled.

Where $E_f(x) \neq E_G(x)$, R&S's definition still holds and it is possible to establish two sets of mixes of the random variables with the riskless asset such that $\{G_\beta\}$ is more risky than $\{F_\alpha\}$. Finally, the addition of a DARA assumption may enable the ranking of investments by their risk in U_d when it is impossible to do so in U_2: While G is riskier than F in U_2 if G = F + MPS, G will be riskier than F in U_d if G = F + MPSA where MPSA denotes mean preserving spread antispread.

Key Terms

Mean Preserving Spread (MPS)

Mean Preserving Antispread (MPA)

Mean Preserving Spread Antispread (MPSA)

The "Fat Tails" Criterion

The Variance Criterion

The Integral Criterion

10

STOCHASTIC DOMINANCE AND DIVERSIFICATION

Stochastic dominance (SD) rules are applicable in selection between mutually exclusive investments but, unlike the mean-variance rule, they cannot identify all possible efficient diversification strategies. Thus, SD rules can tell us whether investment F dominates investment G, or investment G dominates H, but they cannot provide us with the set of combinations of these three assets that dominate all other sets of combinations. Moreover, for two investments F and G, even if it is given that F dominates G, say by SSD, when diversification is considered, one cannot tell unequivocally whether this SSD implies that more than 50% of the wealth should be invested by all risk averters in the superior investment F. Analysis of SD and diversification has been attempted but much still has to be accomplished in this area of research. In this chapter we first discuss some published results obtained in this area, and then we will report some new results.

10.1 ARROW'S CONDITIONS FOR DIVERSIFICATION

a) Diversification Between a Risky and a Riskless Asset

Arrow (1971)[1] studies the diversification policy of risk-averse investors where there are only two assets, one risky asset and one riskless asset which, for simplicity, is assumed to be cash (later on we relax this assumption), with zero rate of return and zero risk. Using Arrow's notations, we have:

X = rate of return on the risky asset

A = initial wealth

a = amount invested in the risky asset

m = A–a, amount invested in the secure asset

Y = final wealth

Thus, the final wealth Y can be written as $Y = m + a\,(1+X) = (m+a) + aX = A+aX$.

The investor's goal is to maximize the expected utility of terminal wealth given by:

$$EU(Y) = E[U(A + aX)] \equiv W(a) \qquad (10.1)$$

where $Y = A + aX$ and a, the decision variable, is restricted between 0 and A.

[1] Arrow, J.K., *Essays in the Theory of Risk Bearing*, Markam Publishing Company, Chicago, 1971.

Taking the first derivative of EU(Y) with respect to a, Arrow analyzes the conditions under which diversification between the two assets is optimal for all risk averse investors (i.e., conditions under which there is a diversified portfolio which dominates the specialized portfolios [i.e., a = 0 or a = A]). To be more specific, Arrow shows that:

$$W'(0) = E[U'(Y)X] \text{ and } W''(a) = E[U''(Y)\,X^2 \qquad (10.2)$$

Because of the risk aversion assumption, $U''(Y) < 0$ for all Y, hence $W''(a) < 0$ for all a. Therefore, $W'(a)$ is decreasing which implies that W(a) must exhibit one of the three possible shapes given in Figure 10.1.

In Figure 10.1a, W(a) has its maximum at a = 0; hence, holding all assets in cash is optimal. This implies that $W'(0) \leq 0$ or that

$W'(0) = U'(A)\,E(X) \leq 0$ (Note that if a=0, Y=A). However, because $U'(A) \geq 0$,

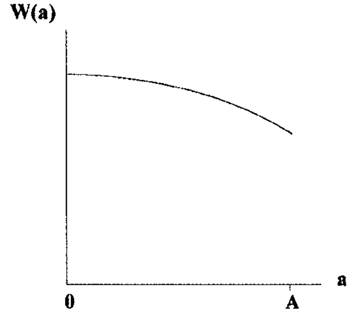
we can conclude that $E(X) \leq 0$. Therefore, the solution a = 0 is optimal if and only if $E(X) \leq 0$. In other words, if $E(X) \leq 0$, holding 100% in cash has a higher (risk averse) expected utility than any other diversification between cash and the risky asset. Therefore, if $E(X) \leq 0$, holding only cash provides a higher expected utility than any mix of cash and the risky asset as long as the utility function $U \in U_2$. Therefore, if $E(X) \leq 0$, we can safely conclude that cash will dominate any mix of cash and the risky asset by SSD (see Chapter 3). Similarly, Arrow shows that specialization in the risky asset (a = A) will be optimal for all risk averse investors if and only if, for all values of X:

$$E\,[U'(A + AX)X] \geq 0 \qquad (10.3)$$

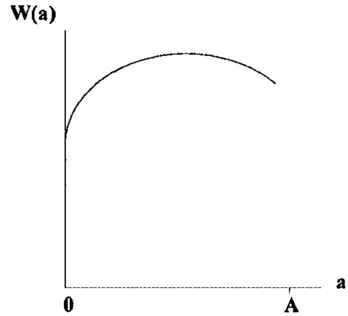
This case corresponds to Figure 10.1c. In other words, the expected utility W(a) increases as we increase a; hence, the solution a = A will be optimal for all risk averters, as long as eq. (10.3) holds. Finally, Figure 10.1b corresponds to the case where an interior diversification policy $0 < a < A$ maximizing expected utility. Of course, this interior optimum is a function of preference, U. We discuss below the dominance condition corresponding to Figure 10.1a and 10.1c and show that it implies SSD, and in some cases, even FSD.

Let us now apply the SD model (i.e., using cumulative distributions) to demonstrate Arrow's conditions for all diversification strategies in cash and the risky asset to be dominated by SSD (and maybe FSD) by the specialized investment (cash or the risky asset) strategy. Let us look first at the case where cash dominates all diversification strategies by Arrow's condition. Because this occurs when $E(X) \leq 0$, X is a random variable which must have at least one negative value and possibly (but not necessarily)

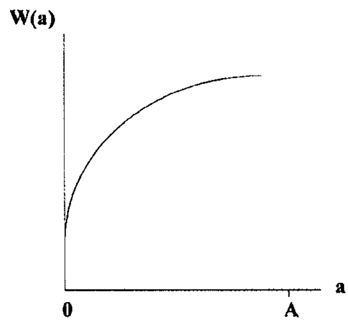
Figure 10.1: Expected utility, W(a) of various diversification strategies in cash and in a risky asset



10.1a: Specialization in cash



10.1b: All diversification strategies are SSD efficient



10.1c: Specialization in the risky asset

some positive values. Therefore, $Y = A + aX < A$ for the negative values of X and $Y = A + aX > A$ for the positive value of X (if they exist). This case is shown in

Figure 10.2a by the two cumulative distributions F(C) and F(X). The cumulative distribution corresponding to 100% cash holding is given by F(C) and the cumulative distribution of any diversification strategy ($0 < a < A$) is given by F(Y). Because cash is certain, there will be one intersection point between F(C) and F(Y) at the most (and no intersection if $X \le 0$ for all values X). We claim that $E(X) \le 0$ implies SSD dominance of cash over the risky assets.

To see this claim, consider the following statements:

1. $E(C) = A$ because all assets are held in cash.

2. $E(Y) = A + aE(X) \le A$ because $E(X) \le 0$ and $0 < a < A$.

3. Therefore, $E(C) \ge E(Y)$.

4. $E(C) \ge E(Y)$ and the one intersection as given in Figure 10.2a where the "+" area is greater or equal to the "–" area (see Chapter 3 and Figure 10.2a), guarantees that C dominates Y by SSD.

5 And the converse also holds, with one intersection described in Figure 10.2a, if C dominates Y then $E(X) \le 0$.

Thus, $E(X) \le 0$ if and only if cash dominates a diversified portfolio by SSD.

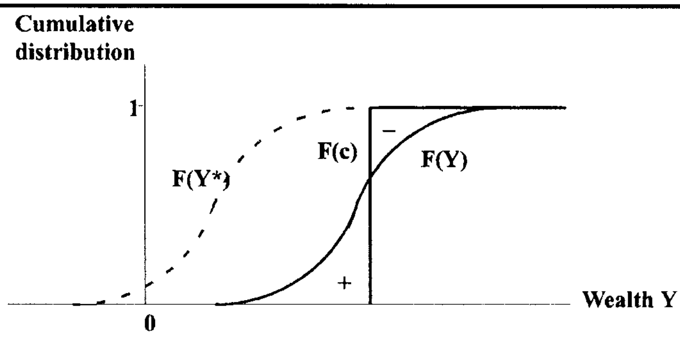
Finally, note that if $X \le 0$ for all X (which is consistent with Arrow's condition $E(X) \le 0$), then $Y \le A$ for all values Y, and F(C) will dominate F(Y) by FSD as well as by SSD. The cumulative distribution corresponding to the case $X \le 0$ for all X is given by $F^*(Y)$ in Figure 10.2a.

Let us now turn to the more realistic case where $E(X) > 0$. We distinguish between two cases: case i) $X < 0$ for some values X; and case ii) $X \ge 0$ for all values X.

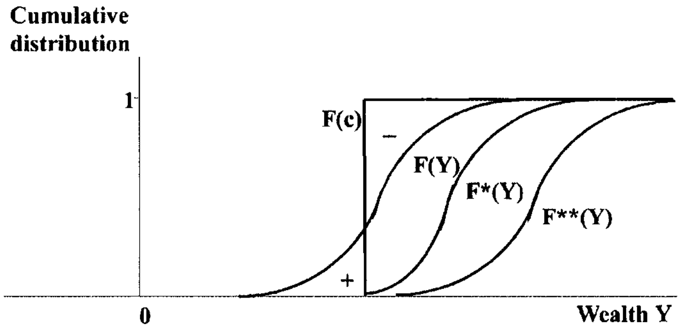
case i): $X < 0$ for some values X

In the case $E(Y) \equiv A + a\,E(X) > A$, for all $0 < a < A$. Hence, the "+" area will be smaller than the "–" area as illustrated in Figure 10.2b. Because $E(Y) > A$, cash cannot dominate Y (see Chapter 3). As long as there is even one value $X < 0$, F(Y) will start to the left of F(C) (because $Y = A + aX$ and X is negative for this particular value $X < 0$); hence, Y will not dominate cash either regardless of the selected

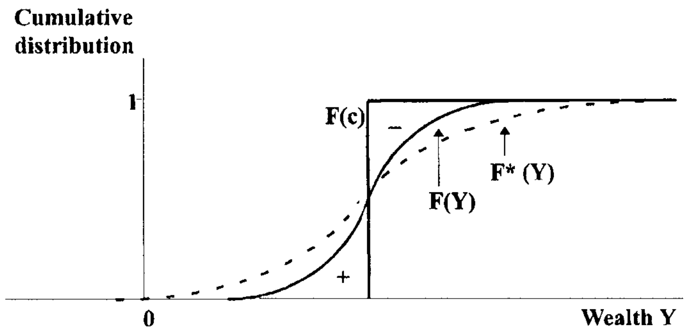
Figure 10.2: A demonstration of ARROW's results in the SD framework



10.2a: E(x)≤0 implies specialization in cash is optimal



10.2b: Min(x)≥0, imply a dominance of the risky asset by FSD



10.2c: E(x)>0, Min(x)<0, imply all strategies 0≤a≤A are efficient SSD

diversification strategy a (as long as $a > 0$). The reason for this is that if there is $X < 0$, then there will be a value Y such that $Y < A$. Thus, $E(X) > 0$ and $X < 0$ for some value implies that neither F(Y) nor F(C) dominates the other.

case ii) $X \geq 0$ for all values X

Let us turn now to the case $E(X) > 0$ and $X \geq 0$ for *all* values X. In this case, $Y = A + aX \geq A$ for all values X (with at least one strict inequality), and diversification with $a > 0$ dominates F(C) by FSD. This is illustrated by $F^*(Y)$ in Figure 10.2b. Thus, in such a case, $F^*(Y)$ dominates the 100% cash investment, not only by SSD but also by FSD. However, if $X \geq 0$ for all values X, we have:

$$Y^{**} = A + AX \geq Y = A + aX$$

for all values a where $A > a > 0$ and X, because $A > a$ and $X \geq 0$. Therefore, by increasing a up to A, we shift the distribution of Y further to the right, and the distribution $F^{**}(Y)$ dominates cash as well as all other possible mixes of cash and the risky asset (e.g., $F^*(Y)$ dominates cash by FSD, as well as by SSD, see Figure 10.2b). The case where specialization in the risky asset is optimal for all risk averse investors corresponds to Figure 10.1c as suggested by Arrow.

To sum up, for the case $E(X) > 0$, if there is even one value $X < 0$, then specializing in the risky asset will not be optimal for all risk averters. There may be some risk averters for whom it is optimal to diversify between cash and the risky asset. However, if $X \geq 0$ for all X, specialization in the risky asset will be optimal for all investors, risk averters and risk lovers, alike, because the strategy $a = A$ will dominate all other diversification policies by FSD.

The last result corresponding to case (ii) above is consistent with Arrow's assertion that $E[U'(A + AX)\, X] \geq 0$ constitutes the condition for dominance of strategy $A = a$ over all diversification strategies (see eq. 10.3). We claim that equation (10.3) is equivalent to the condition $X \geq 0$ given above and, therefore, it is the condition for FSD and not SSD. Let us elaborate on Arrow's condition given by eq. (10.3) and see how it is related to FSD (rather than SSD).

First, if $X \geq 0$ for all X, it is clear that Arrow's condition will hold because $U' \geq 0$. Thus, $X \geq 0$ implies eq. (10.3). Let us now show the converse case: if $X < 0$ even for one value X, then eq. (10.3) cannot hold for all U. To see this claim, suppose that there is some range where $X < 0$. Choose a utility function such that $U' > 0$ for $X < 0$ and $U' = 0$ for $X \geq 0$. Then for this specific utility function (which belongs to U_1 as well as U_2, see Chapter 3), $E[U'(A+AX)\, X] < 0$ and Arrow's condition does not hold. Thus, Arrow's condition for specialization in the risky asset given by eq. (10.3) is equivalent to the condition $X \geq 0$, and the latter implies FSD of the specialized strategy $a = A$ over any other investment strategy.

Note, that the two extreme cases ($E(X) < 0$ and all values $X > 0$) are of limited interest because they rarely hold in practice. The most interesting case is where $E(X) > 0$ and $X < 0$ for some values of X, which corresponds to Figure 10.1b where an interior maximum is possible. In this case, it is possible to have an interior maximum at which $W'(a) = 0$ or $E[U'(Y)\,X] = 0$, where $Y = A + aX$, as defined before. In terms of SD, this means that neither cash nor Y dominates the other by SSD, as shown by F(C) and F(Y) in Figure 10.2c. The reason for the non-dominance is that $E(Y) \geq A$ (because $E(X) > 0$), and Y starts to the left of A (see Figure 10.2c). Moreover, by increasing a, we shift from F(Y) to $F^*(Y)$ where $F^*(Y)$ has a "thicker" left tail as well as a higher expected value than F(Y). $F^*(Y)$ does not dominate F(Y) because it starts to the left of Y. However, F(Y) does not dominate $F^*(Y)$ because it has a lower mean (see Chapter 3). Therefore, $F^*(Y)$ and F(Y) are also located in the SSD efficient set. This means that all possible values $0 \leq a \leq A$ are SSD efficient, and the optimal point a, is a function of the investor preference U. This implies that for the case $E(X) > 0$ and $X < 0$ for some values X, we allow an interior maximum as in Figure 10.1b or a non-interior maximum as shown in Figures 10.1a and 10.1c, and all investment strategies may be optimal depending on preference U.

So far, we assume that the riskless asset is cash, yielding a zero rate of return. If, rather than cash, a riskless bond is considered, then the terminal wealth W(a) is given by:

$$W(a) = (A-a)(1+r) + a(1+X)$$

$$= A(1+r) + a(X-r)$$

$$= A(1+r) + aZ$$

$$\text{where } Z \equiv x-r$$

Therefore, Arrow's results should be modified as follows:

a. Risk averse investors will always invest 100% in the riskless asset if and only if $E(X) \leq r$ (or $E(Z) \leq 0$), where r is the riskless interest rate (SSD).

b. If min $(X) > r$ (or min $Z > 0$), then all investors should invest 100% in the risky asset (FSD).

c. If $E(X) > r$ (or $E(Z) > 0$ and min $(X) < r$ (or min $Z \leq 0$)), then all investment strategies $0 \leq a \leq A$ will be included in the SSD set. The proofs are provided by simple modification of the previous proofs.

b) The Effect of Shifts in Parameters or Diversification

Arrow's analysis of the shift in various parameters on the optimal holding of the two assets (cash and risky asset), produces the following results:

1. A decreasing absolute risk aversion (DARA) implies that $da/dA > 0$. In other words, an increase in wealth A will induce an increase in the demand for the safe asset.

2. Increasing relative risk aversion (IRRA) implies that the wealth elasticity of demand for cash is at least 1. Thus, a 1% increase in wealth A, will induce at least 1% increase in cash held. Therefore, not only does the amount of cash held increase, but the proportion of cash held in the portfolio increases, too.

3. If the risky asset is shifted to the right such that $X(h) = X + h$ ($h > 0$), then, if there is DARA, the demand for the risky asset will increase. Thus, if we have an FSD shift of the above type, and preferences are confined to DARA utility functions, then an increase of X in the portfolio will occur.

Arrow analyzes the effects of other possible shifts in the distribution of X on diversification but he does not analyze the general case where F is shifted to the right (not necessarily by a constant h) such that FSD dominance is created.

If normality is assumed, SSD will be equivalent to the M-V rule and the M-V framework can be used to analyze the effect of changes in the mean return and changes in variance on the optimum portfolio diversification. Tobin[2] (1958) analyzes the demand for cash as a function of changes in the mean return, whereas Levy (1973),[3] employing the multi-asset case, analyzes the effect of changes in the mean, variance and correlations on the demand for each asset. To be more specific, consider the case of n risky assets and one riskless asset: As a result of new information (e.g., a new income statement from IBM reporting unexpectedly large earnings), an investor changes the estimate of the mean return of the firm from μ to μ_1, where $\mu_1 > \mu$. Will the investor invest a higher dollar amount in the stock? A higher proportion in the stock? In general, Levy finds that it is impossible to tell because there are "income" and "substitution" effects. Thus, with no additional restrictions on utility, it is impossible to predict changes in the optimal investment policy even in the M-V framework, let alone in the general case where SD criteria should be employed.

---

[2] Tobin, J., "Liquidity Preferences as Behavior Toward Risk," *Review of Economic Studies*, 25, 1958, pp. 65–86.

[3] Levy, H., "The Demand for Assets Under Conditions of Risk," *Journal of Finance*, 28, March 1973, pp. 79–96.

10.2 EXTENSION OF THE SD ANALYSES IN THE CASE OF TWO RISKY ASSETS

Arrow and Tobin analyze diversification between one risky asset and one riskless asset. In the following, we extend the analysis in several directions, including the case of two risky assets.

Fishburn and Porter (F&P) (1976)[4] analyze diversification between a risky asset and a safe asset by risk averters. An investor who allocates KW of his/her investment capital, W, to the risky asset and (1−K)W to the riskless asset at the riskless interest rate, ρ, where $0 \le K \le 1$, his/her total return X at the end of the period will be given by:

$$X = W[Kr + (1-K)\rho] = W[K(r-\rho) + \rho]$$

where r stands for the rate of return on the risky asset. The investor is assumed to select the diversification strategy K that maximizes the expected utility EU(K: ρ, F): In other words, he/she will choose K which is a function of U and ρ, as well as the random variable with distribution F. The impact of a change in ρ on the optimal diversification between the safe and the risky asset is analyzed first. By taking the first derivative with respect to ρ (and holding F constant) F&P arrive at the following results:

1. Define by $R_A(X)$ the Arrow-Pratt measure of absolute risk aversion. If $R_A(X)$ is constant or an increasing function (e.g., a quadratic utility function), then an increase in ρ will be followed by an increase in the allocation of funds to the safe asset.

2. When $R_A(X)$ is a decreasing function, an increase in ρ may lead to an increase, a decrease or no change in the proportion of W allocated to the safe asset.

3. Define by $R_R(X)$ the relative risk aversion measure. If $R_R(X) < 1$, the investor will increase the proportion of W allocated to the safe asset when ρ increases.

Thus, for the most interesting case of decreasing $R_A(X)$, the results are ambiguous. This is a discouraging finding. Moreover, the financial literature reveals support for the claim that $R_R(X) > 1$, which once again, leads to an ambiguous result.

Let us now change F keeping ρ constant. We can change F in many ways, such as adding a constant to the returns, or adding a_i to each return x_i such that $a_i > 0$ for some i. In these two cases, the new distribution will dominate the first one (F) by

[4] Fishburn, P.C., and R.B. Porter, "Optimal Portfolios with One Safe and One Risky Asset: Effects of Changes in Rates of Return and Risk," *Management Science*, 22, 1976, pp. 1064–1073.

FSD. F&P analyze the effect of such shifts in F on the optimum diversification. It is tempting to believe that such a shift in F to the right will lead to a higher investment proportion in the risky asset. However, F&P show that this is not necessarily the case.

Let us elaborate on Fishburn and Porter's results. Assume a shift from distribution F to distribution G such that G dominates F by FSD. Will the investor allocate a higher proportion of his/her wealth to G in comparison to his/her allocation to F? Fishburn and Porter prove that an increase in the proportion of W allocated to the risky asset, G, will take place only if:

$$R_A(X^*)WK^*(r-\rho) < 1 \text{ for all } r \in (\rho, h).$$

This inequality should hold for any value r in the range (ρ, h) where h is the maximum value, that is G(h) = 1 for all G under consideration, K^* is the optimal allocation with F (before the shift in the distribution), and X^* is the terminal wealth corresponding to K^*.

Kira and Ziemba[5] (1980) extend Fishburn and Porter's analysis to the case of a shift from F to G such that G dominates F by FSD, SSD or, alternatively, by TSD. They first define the following three conditions:

$$\text{(A1)} \quad R_A(W^*)W\lambda^*(X-\rho) \leq 1 \quad \text{for all } X > \rho,$$

$$\text{(A2)} \quad \frac{U'''(W^*)}{U''(W^*)} W\lambda^*(X-\rho) \leq 2 \quad \text{for all } X > \rho,$$

$$\text{(A3)} \quad \frac{-U''''(W^*)}{U'''(W^*)} W\lambda^*(X-\rho) \leq 3 \quad \text{for all } X > \rho,$$

where λ^* is the optimum investment proportion in the risky asset, $(1 - \lambda^*)$ is invested in the safe asset, and X denotes the return on the risky asset. The other notations are as given in Fishburn and Porter. Kira and Ziemba prove the following relationships corresponding to FSD, SSD and TSD, respectively.

a. If G dominates F by FSD and (A1) holds, then $\lambda_G^* \geq \lambda_F^*$; hence, the investor will allocate no less to G than to F when the risky asset improves by FSD.

b. If G dominates F by SSD and (A1) and (A2) hold then $\lambda_G^* \geq \lambda_F^*$.

---

[5] Kira, D. and W.T. Ziemba, "The Demand for Risky Assets," *Management Science*, 26, 1980, pp. 1158–1165.

c. If G dominates F by TSD and (A1) – (A3) holds then $\lambda^*_G \geq \lambda^*_F$.

If a the above conditions are violated, a distribution G can be found that will fulfill the above SD requirements, while maintaining $\lambda^*_G \geq \lambda^*_F$.

Like Fishburn and Porter's condition, conditions (A1) – (A3) are also very restrictive and, in general, an improvement in the risky asset by FSD, SSD or TSD does not guarantee an increase in the investment proportion allocated to the risky asset.

Analysis of SSD diversification with more than two assets is complex and optimal SSD diversification between n risky assets has yet to be developed. However, we can analyze the diversification between two risky assets X and Y, where one differs from the other by a mean preserving spread (MPS) where X and Y are independent. This case was investigated by Hadar and Seo (1988[6], 1990[7]) by analyzing the conditions under which a risk averter will invest more in X than in Y, when X dominates Y either by FSD or alternatively by SSD, and Y is obtained from X by a mean preserving spread. Thus, both X and Y are assumed to be risky assets. Once again, in general, the results are ambiguous. Only when very restrictive conditions are imposed, unambiguous results are obtained. Let us elaborate: Hadar and Seo first assume that X dominates Y by FSD. They prove that when X and Y are *independent,* a risk averter will invest in X at least as much as in Y if and only if the following condition holds:

$$U'(z + b)\, z \text{ is nondecreasing in z for all } 0 < b < 1.$$

This condition can be reformulated to give the following equivalent conditions:

A. $U'(z)\, z$ is nonincreasing in z.

B. $R_R(z + b) - bR_A(z + b) \leq 1$ for all b, $z \geq 0$.

C. $R_R(z) \leq 1$ for all $z \geq 0$.

where R_R and R_A are the Arrow-Pratt relative and absolute risk aversion measures, respectively. Thus, even if X and Y are independent and X dominates Y by FSD, it is not obvious that the investor will invest in X more than in Y.

[6] Hadar, J. and T.K. Seo, "Asset Proportions in Optimal Portfolios," *Review Economic Studies*, 55, 1988, pp. 459–468.

[7] Hadar, J. and T.K. Seo, "The Effects of Shifts in a Return Distribution on Optimal Portfolio," *International Economic Review*, 31, 1990, pp. 721–736.

Hadar and Seo show that when X and Y are *independent* and X dominates Y by SSD such that the difference between them is given by a mean preserving spread (i.e., they have equal means), then the risk-averse investor will not invest less in X than in Y if and only if:

$$U'(z + b)\, z \text{ is concave in } z \text{ for all } 0 < b < 1.$$

This condition can also be rewritten using R_A and R_R. Unfortunately, all these conditions lack economic intuition. They are technical results obtained from solving the first-order condition. Therefore, based on partial information regarding preferences, very little can be said about the investor's response to changes in the distribution of the risky asset. Thus, in general, even if G dominates F by FSD, it is not certain that the investor will invest more than 50 percent of his/her assets in the superior asset.

The implications of these theoretical findings are quite discouraging. For example, consider a firm that changes its investment plan in order to increase the demand for its stocks by creating a mean preserving anti-spread (or by avoiding MPS). The above analyses indicate that the firm is not guaranteed an increase in its market value as a result of this seemingly desirable change. Hence, in a portfolio context, avoiding a mean preserving spread or creating a mean preserving anti-spread is not necessarily desirable.

We see that some restrictive conditions must hold (with respect to preference and distributions) in order to obtain the intuitive result that the more the distribution of the random variable is shifted to the right, the higher the proportion that will be invested in the shifted distribution. Another approach is to impose no restrictions on utility except for $U' > 0$, and require a shift in the distribution of the random variable which is stronger than FSD. Indeed, Lansberger and Meilijson (L&M) (1990)[8] obtain unambiguous results using a strong preference of this type and employing the *Monotone Likelihood Ratio Order* (LR) to derive their results. LR dominance is defined as follows:

Let $X \sim F$ and $X \sim G$: then G will dominate F by the likelihood ratio if $g(X)/f(X)$ is a nonincreasing function of X. The LR condition implies FSD dominance but FSD does not imply LR. Thus, it is possible for $g(X)/f(X)$ not to be monotonic in X, even though G dominates F by FSD. L&M provides the following two propositions:

Proposition 1.

Let X,Y and Z be independent investment returns such that X

---

[8] Landsberger, M. and I. Meilijson, "Demand for Risky Financial Assets: A Portfolio Analysis," *Journal Economic Theory*, 12, 1976, pp. 483–487.

dominates Y by the Likelihood Ratio. Then, for any arbitrary α, α_1, and α_2 such that $\alpha_1 < \alpha_2$, the following will hold:

$$EU(\alpha Z + \alpha_1 X + \alpha_2 Y) \leq EU(\alpha Z + \alpha_2 X + \alpha_1 Y)$$

for all nondecreasing utility functions, U.

Thus, L&M show that if X differs from Y by a strong shift, investors will be better off investing more in X than in Y, which is an unambiguous result.

Proposition 2.

In a portfolio composed of one safe asset and one risky asset, a shift in the distribution of the return on the risky asset in the sense of the Likelihood Ratio will lead to an increase in the demand for this risky asset by all investors with nondecreasing utility functions.

Unike Hadar and Seo, L&M obtain unambiguous results. However, propositions 1 and 2 still refer to independent returns. They claim that their propositions can be employed in practice in constructing a combination of assets to be included in mutual funds, or for that matter, in constructing any managed portfolio. Their claim is theoretically correct. However, it implicitly assumes that risky assets can be found in the market that are dominant by the Likelihood Ratio and that these assets are independent. We believe that such dominance, unfortunately, rarely exists with *ex-post* distributions. However, this is, of course, an empirical question.

To sum up, the SD analysis of investment diversification and the response to changes in the characteristics of returns is not very well developed yet: It yields ambiguous results or it needs very strong restrictions on the distribution of the random variable. We firmly believe that more studies will be devoted to this issue in the future.

10.3 DIVERSIFICATION AND EXPECTED UTILITY: SOME COMMON UTILITY FUNCTIONS

From the previous two sections we see that when an investor faces one risky asset and one riskless asset, and the cumulative distribution of the risky asset is shifted to the right, there is no guarantee that he/she will invest a higher proportion in the risky asset: This is a counter-intuitive result. In this section we examine some of the issues raised in the previous two sections but this time we focus on investors characterized by common preferences (e.g., $W^{1-\alpha}/1-\alpha$, log W, $-e^{-\alpha W}$, etc.) rather than all investors, in general. By changing the risk-aversion parameter α, we are able to analyze how a given level of risk aversion affects the optimum diversification.

Throughout our analysis below, we confine ourselves to two independent assets X and Y, where X is always a risky asset and Y can be either the riskless asset or another risky asset. We use Matlab programs to calculate the optimal investment proportions, Wx and Wy, invested in assets X and Y, respectively, where $Wx + Wy = 1$. For simplicity, we do not allow short sales; hence, $0 \le Wx \le 1$ and $0 \le Wy \le 1$. The investor is assumed to be risk averse with some common representative utility function. (see Levy & Markowitz, [1979][9] and Kroll, Levy & Markowitz, [1984][10]).

For each of these utility functions, $U(\cdot)$, we solve the following equation:

$$\text{Max } EU(Wx + (1-Wx)Y, \ Wx \in [0,1].$$

The various assumed utility functions are given in Table 10.1.

The following five cases were examined:

a) X is a risky investment and Y is the riskless asset with rate of return, r. The optimal proportion Wx is calculated as Y is shifted to the right (increase in the risk-less interest rate, r).

b) X is a risky investment and Y is the riskless asset, r. This time X is shifted to the right until X dominates Y by FSD (see Chapter 3, Section 3.2). We examine how the investment proportions are affected by such a shift.

c) Y is a risky investment and X is constructed from Y by MPS shifts (see Chapter 9, Section 9.2). As we shift more MPS's, X becomes much more risky relative to Y. We examine the effect of MPS on the optimal diversification.

d) Similar to case c) above, but here Y is a risky investment and X is constructed from Y using MPA shifts (see Chapter 9, Section 9.3). Note that as we shift more MPA's, X becomes less risky than Y.

e) The effect of MPSA on the optimum investment proportion.

The following tables and figures give the results for each of these cases.

---

[9] Levy, H., and Markowitz, H.M., "Approximating Expected Utility by a Function of Mean and Variance," *American Economic Review*, 69, No. 3, June 1979, pp. 308–317.

[10] Kroll, Y., Levy, H., and H. Markowitz, "Mean-Variance versus Direct Utility Maximization," *Journal of Finance*, March 1984, pp. 47–61.

a) Shift in r

To solve for the optimal investment proportions, some distribution of returns has to be assumed. Let us assume that the rate of return on the risky investment has the following probability distribution function:

X(%)	P(X)
3	0.2
4	0.2
5	0.2
6	0.2
18	0.2

Note that E(X) = 7.2%. Therefore, we increase r until it is greater than 7.2%.

Table 10.1 provides the investment proportion in X for the various utility functions, as r increases. Figure 10.3 illustrates the change in the proportion invested in the risky asset as it increases.

Table 10.1 The Investment Proportions in X as it increases

	r					
U(W)	6.75%	6.85%	6.95%	7.05%	7.15%	7.25%
Log(W)	1	1	0.9475	0.5536	0.1799	0
$W^{0.9}/0.9$	1	1	1	1	1	0
$W^{0.1}/0.1$	1	1	1	0.6164	0.2000	0
$W-0.05W^2$	1	1	1	1	1	0
$-e^{-0.01W}$	1	1	1	1	1	0
$-e^{-0.1W}$	1	1	1	1	1	0
$-e^{-0.9W}$	1	1	0.9518	0.5634	0.1854	0
$-e^{-0.99w}$	1	1	0.8653	0.5122	0.1685	0
$(W+0.1)^{0.9}$	1	1	1	1	1	0
$(W+0.9)^{0.1}$	1	1	1	1	0.3680	0

Note that Min X < r and E(X) ≥ r for all r ≤ 7.2%. Therefore, neither r nor X dominates the other. Nevertheless, for r ≤ 6.85%, the optimum solution is to invest 100% in the risky asset. Then, as r increases further, diversification takes place and less is invested in X, ending with 100% invested in the riskless asset when r = 7.25. A comparison of various risk-aversion coefficients, as expected, shows that the more risk averse the investor, the faster the shift toward diversification as r increases (compare, for example, $W^{0.1}/0.1$ and $W^{0.9}/0.9$).

b) Shift in X

Let us now assume that r = 7.25% and X is the return on the risky asset. We initially assume that X has the following distribution function:

X%	P(x)
3	0.2
4	0.2
5	0.2
6	0.2
18	0.2

Then we shift X to the right where a shift of 0.1 means that all values of X increase by 0.1% with no change in the probabilities.

Figure. 10.3: The optimal investment proportion (Wx) in the risky asset x, as a function of the riskless interest rate r(%) (various utility functions)

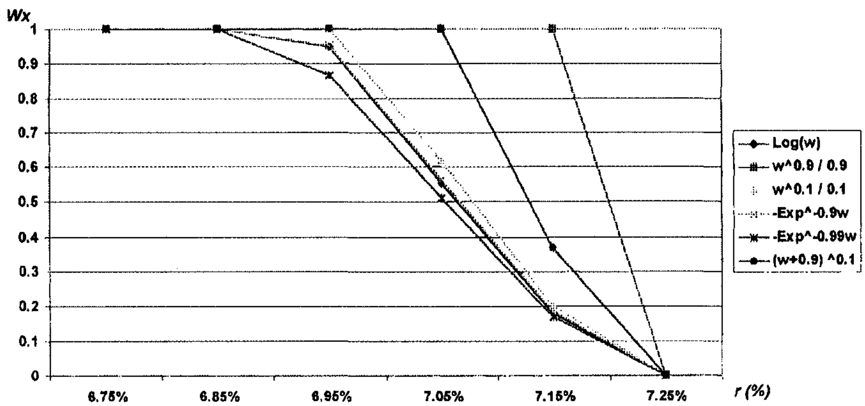


Table 10.2 gives the investment proportion in X for the various utility functions, as X shifts to the right.

Note first that for zero changes and with r = 7.25%, as shown in Table 10.1, investors invest all their wealth in the riskless asset. However, as we add a constant return to X, the investment proportion becomes positive and diversification takes place. With as little a shift as 0.4%, all preferences reveal 100% investment in the risky asset.

Table 10.2. The Investment Proposition in the Risky Asset: Shifts in X

	Shifts in X					
U(W)	0	0.1%	0.2%	0.3%	0.4%	0.5%
Log(W)	0	0.1801	0.5546	0.9501	1	1
$\mathbf{W^{0.9}/0.9}$	0	1	1	1	1	1
$\mathbf{W^{0.1}/0.1}$	0	0.2002	0.6175	1	1	1
$\mathbf{W-0.05W^2}$	0	1	1	1	1	1
$\mathbf{-e^{-0.01W}}$	0	1	1	1	1	1
$\mathbf{-e^{-0.1W}}$	0	1	1	1	1	1
$\mathbf{-e^{-0.9W}}$	0	0.1854	0.5634	0.9518	1	1
$\mathbf{-e^{-0.99W}}$	0	0.1685	0.5122	0.8653	1	1
$\mathbf{(W+0.1)^{0.9}}$	0	1	1	1	1	1
$\mathbf{(W+0.9)^{0.1}}$	0	0.3682	1	1	1	1

c) MPS Shifts.

So far, we have analyzed one risky asset and one riskless asset. We turn now to the case of two risky assets when they differ by one or more MPS's. Let us illustrate with investor Y whose probability distribution function is given by,

Y(%)	P(Y)
3	0.2
5	0.2
8	0.2
11	0.2
30	0.2

Initially X = Y (0 MPS). We then initiate MPS shifts on Y by moving some probability from the center of Y to the left and to the right such that the mean of the new and riskier investment, X, is equal to the mean of Y. To be more specific, we move probabilities from Y = 8 to Y = 5 and Y = 11 in a symmetrical manner. The MPS size is equal to 0.02 (see Table 10.3) and more than one MPS can be employed.

Thus, X is obtained from Y by n MPS shifts where n= 1, 2, 3, 4. For example, with one MPS, we obtain that X = 3, 5, 8, 11, 30 with probabilities 0.2, 0.22, 0.16, 0.22 and 0.2, respectively. When n = 0, X = Y and, as expected, the optimal investment is 50% in each of the two assets. Thus, with MPS shifts, the investment in the more risky asset is less than 50%.

d) MPA Shifts.

By adding to Y an MPS, we obtain an asset X which is more risky than Y. By adding MPA to Y, we obtain an asset X which is less risky than Y. We start with a distribution Y as follows:

Y	**Probability**
1	0.1
5	0.3
8	0.2
10	0.1
15	0.3

Table 10.3. The Investment Proportion in the More Risky Asset (MPS = 0.02)

U(W)	**0 MPS**	**1 MPS**	**2 MPS**	**3 MPS**	**4 MPS**
Log(W)	0.5	0.4989	0.4978	0.4967	0.4956
$W^{0.9}/0.9$	0.5	0.4990	0.4979	0.4969	0.4959
$W^{0.1}/0.1$	0.5	0.4989	0.4978	0.4967	0.4956
$W-0.05W^2$	0.5	0.4990	0.4981	0.4971	0.4962
$-e^{-0.01W}$	0.5	0.4990	0.4981	0.4971	0.4962
$-e^{-0.1W}$	0.5	0.4990	0.4981	0.4971	0.4962
$-e^{-0.9W}$	0.5	0.4990	0.4979	0.4969	0.4959
$-e^{-0.99W}$	0.5	0.4990	0.4979	0.4969	0.4959
$(W+0.1)^{0.9}$	0.5	0.4990	0.4979	0.4969	0.4959
$(W+0.9)^{0.1}$	0.5	0.4990	0.4979	0.4969	0.4959

Then we create a new variable X such that X = Y+ n·MPA where n = 0, 1, 2,…,7. Thus, with one MPA, X obtains the values 1, 5, 8, 10, 15 with the corresponding probabilities of 0.09, 0.3, 0.22, 0.1, and 0.29, respectively.

Note that because the MPA shift creates a new random variable X which is less risky than Y, more than 50% of the wealth is invested in X. Moreover, the more MPA's that are added, the larger the investment proportion in X (see table 10.4).

e) MPSA Shifts

Finally, we analyze the effect of MPSA on the investment proportion in X. We start with (Y, P(Y)), and add to P(Y) MPS shifts and MPA shifts as follows:

Y	P(Y)	P(Y) + MPS	P(Y) + MPS + MPA
1	0.1	0.11	0.11
5	0.3	0.30	0.29
8	0.2	0.18	0.20
11	0.1	0.10	0.09
15	0.3	0.31	0.31

Note that the MPA is "inside" the MPS, hence the MPSA increases risk for all DARA utility functions.

Table 10.4. The Investment Proportion in the Less Risky Asset, X: MPA Shifts (MPA = 0.01)

U(W)	0 MPA	1 MPA	2 MPA	3 MPA	4 MPA	5 MPA	6 MPA	7 MPA
Log(W)	0.5	0.5020	0.5041	0.5061	0.5082	0.5103	0.5124	0.5146
$W^{0.9}/0.9$	0.5	0.5020	0.5041	0.5061	0.5082	0.5103	0.5124	0.5145
$W^{0.1}/0.1$	0.5	0.5020	0.5041	0.5061	0.5082	0.5103	0.5124	0.5146
$W-0.05W^2$	0.5	0.5020	0.5040	0.5061	0.5081	0.5102	0.5123	0.5144
$-e^{-0.01W}$	0.5	0.5020	0.5040	0.5061	0.5081	0.5102	0.5123	0.5144
$-e^{0.1W}$	0.5	0.5020	0.5040	0.5061	0.5081	0.5102	0.5123	0.5144
$-e^{-0.9W}$	0.5	0.5020	0.5040	0.5061	0.5081	0.5102	0.5123	0.5144
$-e^{-0.99W}$	0.5	0.5020	0.5040	0.5061	0.5081	0.5102	0.5123	0.5144
$(W+0.1)^{0.9}$	0.5	0.5020	0.5040	0.5061	0.5081	0.5102	0.5123	0.5144
$(W+0.9)^{0.1}$	0.5	0.5020	0.5040	0.5061	0.5081	0.5102	0.5123	0.5144

Indeed, Table 10.5 reveals that the investment in X is less than 50% and it monotonically diminishes as we add more MPSA's. It is interesting that the results hold also for the quadratic utility even though it is characterized by an increasing absolute risk aversion.

Table 10.5. The Investment Proportions in the Risky Asset: MPSA Shifts MPS = MPA = 0.01

U(W)	0	1	2	3	4	5	6	7	8	9	10
Log(W)	0.5	0.4912	0.4826	0.4744	0.4664	0.4588	0.4513	0.4441	0.4372	0.4304	0.4239
$W^{0.9}/0.9$	0.5	0.4912	0.4827	0.4746	0.4666	0.4590	0.4516	0.4444	0.4375	0.4307	0.4343
$W^{0.1}/0.1$	0.5	0.4912	0.4826	0.4744	0.4665	0.4588	0.4513	0.4441	0-4372	0.4305	0.4239
$W-0.05W^2$	0.5	0.4913	0.4829	0.4747	0.4669	0.4593	0.4519	0.4448	0.4379	0.4312	0.4247
$-e^{0.01W}$	0.5	0.4913	0.4829	0.4747	0.4669	0.4593	0.4519	0.4448	0.4379	0.4312	0.4247
$-e^{0.1W}$	0.5	0.4913	0.4829	0.4747	0.4669	0.4593	0.4519	0.4448	0.4378	0.4311	0.4247
$-e^{0.9W}$	0.5	0.4912	0.4828	0.4746	0.4667	0.4590	0.4516	0.4445	0.4375	0.4308	0.4243
$-e^{0.99W}$	0.5	0.4912	0.4828	0.4746	0.4667	0.4590	0.4516	0.4444	0.4375	0.4308	0.4243
$(W+0.1)^{0.9}$	0.5	0.4912	0.4828	0.4746	0.4667	0.4590	0.4516	0.4445	0.4375	0.4308	0.4243
$(W+0.9)^{0.1}$	0.5	0.4912	0.4828	0.4746	0.4667	0.4590	0.4516	0.4444	0.4375	0.4308	0.4243

10.4 SUMMARY

Tobin and Levy, in the mean-variance framework, and Arrow, in the general expected utility framework, analyze the conditions for the optimality of diversification between a risky asset(s) and a riskless asset. Their main conclusion is that unless the expected return on the risky asset is negative, or min X > r, all mixes, as well as the specialized investment strategies, are efficient.

Arrow (and Tobin and Levy in the M-V framework) also analyzes the effect of changes in the various parameters on the optimal diversification policy. Other researchers extend the above analysis to FSD shifts in the random variable as well as SSD shifts. When the SSD shifts are conducted via the addition of MPS, the main results are inconclusive; a shift in F by FSD or by SSD does not yield a clear-cut result regarding the optimal diversification.

We ran calculations of expected utility for some common utility functions. In all cases, the change in the investment proportion is in the intuitive direction: An increase in the riskless interest rate induces an increase in the investment proportion in the riskless asset. A shift to the right in the cumulative distribution of the risky asset increases the investment proportion in the risky asset. Similarly, when two random variables are considered, MPS and MPA affect the results as follows: The more (less) risky the asset becomes, the smaller (larger) the investment proportion in the asset. Obviously, the simulation results are limited to the utility functions and the numerical examples studied here.

KEY TERMS

Optimal Diversification

MPS Shifts

MPS Shifts

MPSA Shifts

DARA

Utility Function

Monotone Likelihood Ratio Order (LR)

11

DECISION MAKING AND THE INVESTMENT HORIZON

Mean-variance and stochastic dominance efficiency analysis, the stock's beta, and portfolio performance measures are all based on rates of return: historical rates of return in the case of *ex-post* analysis or true rates of return in the case of *ex-ante* analysis. In turn, rates of return are calculated on the basis of the investment horizon. In this chapter we investigate the extent to which the stochastic dominance efficient sets as well as other widely used portfolio analyses depend on the selected investment horizons.

In most of the proofs given in this chapter, it is assumed that the distribution's parameters are known and we show that portfolio analyses are affected by the assumed investment horizon or holding period. In empirical research, we use sample data to estimate the true unknown distribution and here, too, the investment horizon plays an important role. To illustrate, suppose that we have monthly rates of return for the last ten years. We use these rates of return as estimates of the true unknown distributions. We can use the same data set to calculate, say, the quarterly rates of return. Suppose that returns have *identical independent distributions* (i.i.d.) over time. Do we expect any systematic change in the dominance relationship, the portfolio composition, beta and performance measures as we shift from one horizon to another (i.e., from monthly to quarterly rates of return)? Note, that we use the same set of data but we use it differently as we shift from one horizon to another. In other words, does the way we "slice" the data into atomic units affect the portfolio analyses results?

We will show that the mean-variance analysis, the performance measures, betas, co-variance, portfolio composition, as well as stochastic dominance analysis are strongly dependent on the assumed investment horizon even if the rates of return are identical and independent over time (i.i.d.). We show, for example, that investment A may be ranked better than investment B with monthly data and the opposite holds with quarterly data. Such results will be extremely important to investors as well as fund managers in (legally) manipulating the ranking of their funds.

We first focus on M-V analysis and then on SD analysis.

11.1 TOBIN'S M-V MULTI-PERIOD ANALYSIS

If returns are dependent over time, investment efficiency analysis is not invariant to the assumed investment horizon. To see this, consider a stock whose return changes sign every month. Using the bi-monthly rates of return, we may have zero variance, but with monthly rates of return there will be a positive, maybe even very large, variance. Thus, when returns are not i.i.d., the M-V efficiency analysis is affected by the assumed investment horizon. Thus, it is only natural to ask whether the investment horizon affects the efficiency analysis when the returns are characterized by identical (or stationary) independent distribution (i.i.d.) over time, an assumption which conforms with the *random walk* hypothesis. Intution would probably lead us to believe that assuming i.i.d., the investment analysis and, in particular, asset-ranking by performance, will be invariant to the horizon employed. We show in this chapter that this is a misleading intuition.

Tobin (1965) shows that the M-V analysis is affected by the assumed investment horizon even under the i.i.d. assumption. To show Tobin's claim, we need first to derive the formula for the multi-period mean and variance. Denote the one-period rate of return by $x \equiv (1+R)$, where R is the one-period rate of return. Denote the mean of x and its variance by $(1+\mu)$ and σ^2, respectively. The terminal wealth W_n, after n periods, is given by:

$$\tilde{W}_n = \prod_{i=1}^{n}(1+\tilde{R}_i) \tag{11.1}$$

where $\tilde{R}_n$ and $\tilde{W}_n$ are random variables and the initial investment is assumed to be $W_0 = \$1$. The mean of the terminal wealth is given by:[1]

$$E(W_n) = E\left(\prod_{i=1}^{n}(1+\tilde{R}_i)\right) = \prod_{i=1}^{n}(1+\mu_i) = (1+\mu)^n \tag{11.2}$$

where $\mu_i = \mu$ due to the assumption of stationarity over time. The n-period variance is given by:

$$\sigma_n^2 = E\left[\prod_{i=1}^{n}(1+\tilde{R}_i)\right]^2 - (1+\mu)^{2n}. \tag{11.3}$$

Employing the assumption of independence and stationarity, we obtain:

$$\sigma_n^2 = [E(1+\tilde{R})^2]^n - (1+\mu)^{2n} = [\sigma^2 + (1+\mu)^2]^n - (1+\mu)^{2n} \tag{11.3'}$$

or:

$$\sigma_n^2 = \sum_{k=0}^{n-1}\binom{n}{k}(\sigma^2)^{n-k}(1+\mu)^{2k} \tag{11.3''}$$

(For more details on the development of these relationships, see Tobin, 1965.)[2]

---

[1] To derive equation (11.2), we use the stationary and independence assumptions.

From these equations, Tobin derives the following two conclusions:

[a] If x does *not* dominate y by the M-V rule for one-period (n = 1) investors, then such dominance does *not* exist also for multiperiod (n >1) investors either.

[b] If x dominates y by the M-V rule for one-period (n = 1) investors, x may *not* dominate y for the multiperiod (n >1) investors.

Therefore, the size of the M-V efficient set increases (in the weak sense) as the horizon increases.

To see conclusion [a] above, recall that if $\mu_x > \mu_y$ and $\sigma_x^2 > \sigma_y^2$ (and one equality is allowed), then $E((\tilde{W}_n(x)) > E(\tilde{W}_n(y))$ (see eq. 11.2), and $\sigma_n^2(x) > \sigma_n^2(y)$ (see eq. 11.3″), where one equality is allowed. To see conclusion [b] above, assume that $\mu_x > \mu_y$ and $\sigma_x < \sigma_y$; hence, x dominates y for n=1. However, it is possible to have $\sigma_n^2(x) > \sigma_n^2(y)$ (due to the fact that $\mu_x > \mu_y$, see eq. 11.3″); hence, there is no dominance in the multiperiod setting. Thus, the M-V efficient set may increase, but never decreases, as the horizon increases.

11.2 SHARPE'S REWARD-TO-VARIABILITY RATIO AND THE INVESTMENT HORIZON

From the above equations, it is clear that both the mean and the variance of rates of return increase with the assumed horizon. Therefore, it is possible that changes in the investment horizon will not affect Sharpe's (1966)[3] performance index. Unfortunately, this is not the case.

Sharpe's performance index known as the *Reward-to-Variability Ratio* (R/V) is given by:

$$R/V = \frac{\mu - r}{\sigma}$$

where r stands for the riskless interest rate. Using equations (11.2) and (11.3), the multi-period reward-to-variability ratio $(R/V)_n$ is given by:

$$(R/V)_n = \frac{[(1+\mu)^n - 1] - [(1+r)^n - 1)]}{\{[\sigma^2 + (1+\mu)^2]^n - (1+\mu)^{2n}\}^{\frac{1}{2}}}, \qquad (11.4)$$

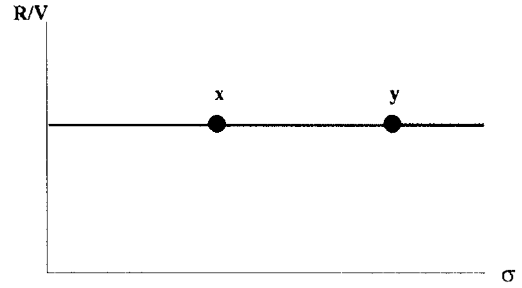
---

[2] Tobin, J., "The Theory of Portfolio Selection," in F. H. Hahn and F.P.R. Brechling, eds., *Theory of Interest Rates*, New York, Macmillan, 1965.

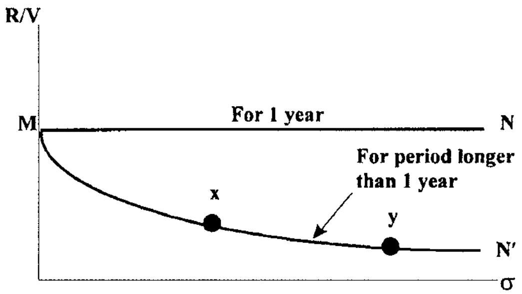
[3] Sharpe, W. F., "Mutual Fund Performance," *Journal of Business*, January 1966.

where a flat yield curve is assumed; hence the multi-period interest rate is $[(1+r)^n - 1]$. Suppose that for n=1, all portfolios are located on a horizontal line A as shown in

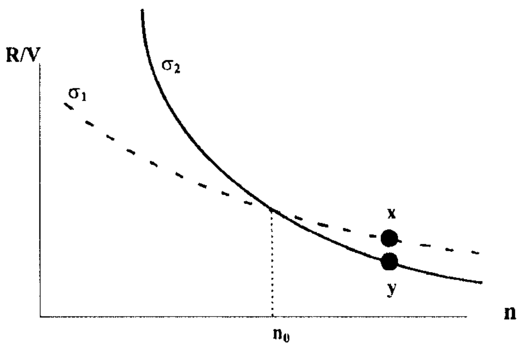
Figure 11.1: The R/V Ratio and the Investment Horizon



11.1a: R/V as a function of σ for a given horizon



11.1b: R/V as a function of σ for various horizons



11.1c: R/V as a function of the horizon for various σ***

Figure 11.1a. Thus, for n=1, no portfolio dominates the other by the R/V ratio. However, using eq. (11.4), Levy (1972)[4] proved that for n > i, $\partial(R/V)_n/\partial\ \sigma < 0$; hence, portfolios with a large one-period σ will be inferior relative to portfolios with low σ (see portfolios y and x in Fig. 11.1a) when multi-period (n > 1) distributions of rates of return are considered. This is illustrated in Figure 11.1b. The implication of this result is that if we were to invest say, for one year, we would be indifferent between two portfolios x and y. However, if we decide to invest for two years, by the R/V criterion, we would shift to the investment with the smaller one-period σ (see x and y in Figures 11.1a, 11.1b and 11.1c). This is somewhat counter-intuitive because the annual rates of return are identical from one year to the next.

Similarly, it can be shown that if the R/V ratio is constant for a given period, say, one year, then a portfolio based on data corresponding to a shorter horizon, say monthly rates of return, with a higher one-year σ will dominate a portfolio with a lower σ by the R/V criterion. (The graph related to this case is not shown.)

The effect of changes in the length of the investment horizon on the R/V ratio is summarized in Figure 11.1c. If the "true" investment horizon is given by n_0, and in the empirical study we use n_0 as the basic unit of time for the estimation of μ and σ, we will not expect any systmatic bias in the one-parameter measures of performance (R/V). Moreover, in equilibrium, all portfolios will have the same R/V ratio, that is, all the curves will intersect at n_0. If, however, we use n_1, where $n_1 < n_0$ as the basic time unit in our empirical research, we will obtain a systematic bias of the one-parameter measure of performance. The portfolio with the highest σ (and hence the highest μ) will tend to have the highest R/V ratio. On the other hand, if the basic time unit is n_2 where $n_2 > n_0$, we can predict that the portfolio with a lower σ will also have a higher R/V ratio.

In most empirical studies related to investment in the stock market, the basic unit of time is usually selected quite arbitrarily as one year or one month. However, by doing this, the researcher ignores the important role of the basic unit of time in the calculation of the rates of return. The above results indicate that more attention should be paid to the selection of the basic time unit. An empirical study based on a yearly rate-of-return data may yield different results from one that uses monthly rate-of-return data. Specifically, the R/V ratio of mutual funds is dependent on the assumed investment horizon because the ranking of mutual funds may change as the horizon changes. The ranking of one mutual fund over another may not reflect performance but simply be induced by the inappropriate division of the period studied into smaller time periods according to which ratios of returns are calculated. It seems that practitioners are not aware of the serious error that may arise due to choosing a wrong horizon. For example, in an article published in *Forbes*, 1997,[5] mutual funds are ranked by R/V ratios based on monthly rates of return. As most investors in

[4] Levy, H., "Portfolio Performance and the Investment Horizon," *Management Science*, 18, 1972, pp. 645–653.

[5] See Mark Hulbert, "Calculated Risk," *Forbes*, Jan. 27, 1997, p. 114.

mutual funds probably have longer horizons, these ratios mistakenly suggest that high risk funds perform better than lower risk funds (see Figure 11.1c).

So far, we have discussed the horizon effect on expected returns, variances and Sharpe's R/V ratio. Other important parameters in portfolio diversification and the evaluation of risk (e.g., beta) are also affected. We now turn to the effect of the assumed horizon on correlations.

11.3 THE EFFECT OF THE INVESTMENT HORIZON ON CORRELATIONS

Mean-variance optimal portfolio diversification among assets relies on the means, variance and the correlations (or covariances) of the rates of return on the various assets. In this section we analyze the horizon effect on correlations. Suppose that returns are i.i.d. What effect will changes in the assumed investment horizon have on the various correlations underlying the M-V portfolio diversification strategy?

Denote by x_i the return (1 + rate of return, which is non-negative) on asset x in period i, say month i, and by y_i the corresponding return on the other asset. Levy (1996)[6] has shown that the n-period correlation (i.e., correlation ρ_n based on a horizon of n time units, say, 12 months) and the one-period correlation, ρ_1 (based on a horizon of, say, one month) are related as follows:

$$\rho_n = \frac{\mathrm{Cov}_n(x,y)}{\sigma_n(x)\sigma_n(y)} = \frac{\sum_{k=0}^{n-1}\binom{n}{k}\rho_1^{n-k}[(1+\mu_i)\,(1+\mu_j)]^k(\sigma_i\sigma_j)^{n-k}}{\left(\sum_{k=0}^{n-1}\binom{n}{k}(\sigma_i^2)^{n-k}(1+\mu_i)^{2k}\right)^{1/2}\left(\sum_{k=0}^{n-1}\binom{n}{k}(\sigma_j^2)^{n-k}(1+\mu_j)^{2k}\right)^{1/2}} \quad (11.5)$$

where μ_i and μ_j are the mean one-period returns of these two assets.

Using an equation similar to (11.5), Schneller (1975)[7] shows that for $|\rho_1| \neq 1$, $\lim_{n\to\infty}\rho_n = 0$. However, Levy (1996) shows that $\lim_{n\to\infty}\rho_n = 0$ as long as $\rho_1 < 1$, including the case where $\rho_1 = -1$. Moreover, there are cases where $\lim_{n\to\infty}\rho_n = 0$ even if $\rho_1 = 1$. The only case where ρ_n is unaffected by the horizon is where $\rho_1 = +1$ and x and y are related by the specific linear relationship $(1+R_x) = b\,(1+R_y)$, i.e., a regression line which goes through the origin where R_x and R_y denote the rates of return on the two investments, respectively (i.e., a regression line that intersects the origin).

[6] Levy, H., "Investment Diversification and Investment Specialization and the Assumed Holding Period." *Applied Mathematical Finance*, 3, 1996, pp. 117–134.

[7] Schneller, I.M., "Regression Analysis for Multiplicative Phenomenon and the Implication for the Measurement of Investment Risk," *Management Science*, 22, 1975, pp. 422–426.

Levy and Schwarz (1997)[8] analyze the relationship between ρ_n^2 and ρ_1^2 for any finite n. They show that ρ_n^2 decreases monotonically as n increases except for the case where $\rho = +1$ and the regression line intersects the origin (in this case ρ_n, as claimed above, remains +1 independent of the selected horizon). The mathematical proofs of these claims can be found in the above mentioned studies. Here, we will demonstrate these findings with some numerical examples.

Example 1:

Consider the case, $(1 + R_x) = b\,(1 + R_y) + c$, where $b > 0$, $c \neq 0$. Choose $b = 0.2$ and $c = 0.9$ in the following example: By having $c \neq 0$, we have $\rho_1 = 1$ but the regression line does not intersect the origin.

One-period return		**Two-period return**[9]	
x	y	X	Y
1.1	1.0	1.21	1.00
1.2	1.5	1.32	1.50
		1.32	1.50
		1.44	2.25

From the above example, it is easy to verify that although $\rho_1 = +1$, $\rho_2 = 0.99 < 1$. Moreover, the larger the number of periods, the lower the correlation.

To show that when $\rho_1 = -1$, ρ_n increases, consider the following example:

Example 2:

One-period return		**Two-period return**	
x	y	X	Y
1.1	1.5	1.21	1.25
1.2	1.0	1.32	1.50
		1.32	1.50
		1.44	1.00

[8] Levy, H., and Schwarz, G. "Correlation and Time Interval Over Which the Variables are Measured," *Journal of Econometrics*, 76, 1997, pp. 341–350.

[9] The two-period returns are obtained by $(1 + R_1)(1+R_2)$. Independence across time is assumed; hence, the return of 1.32 and 1.50 are achieved twice.

The figures of Example 2 stand for the return on the two risky assets. It is easy to verify that $\rho_1 = -1$, $\rho_2 = -0.985 > -1$ and, by continuing this example for more periods, we can easily show that as n increases, ρ_n tends toward zero. Thus, for any correlation ρ_1, including the case $\rho_1 = \pm 1$, as long as it is not of the type $(1 + R_x) = b(1 + R_y)$, ρ_n decreases monotonically as n increases, $\lim_{n\to\infty} \rho_n = 0$, and the correlation matrix converges to the diagonal matrix. This result has strong implications for the optimal diversification strategy of investors with various planned investment horizons.

Finally, though $\lim_{n\to\infty} \rho_n = 0$, the speed of convergence of the correlation matrix to the diagonal matrix is not revealed by eq. (11.5). For various hypothetical one-period parameters, we employ eq. (11.5) to calculate ρ_n. Table 11.1 provides μ and σ of five assets, where these are the one-period parameters.

Part A of Table 11.2 gives the assumed one-period correlation matrix and the other parts give the calculated multi-period correlation matrix (ρ_n for n = 5, 20, and 100) as implied by eq. (11.5). As can be seen, all positive one-period correlations (for $i \neq j$) decrease, and all negative one-period correlations increase as n increases (i.e., ρ_n^2 decreases monotonically toward zero).

The change of the correlation has a direct implication of the portfolio composition,

Table 11.1. The One-Period Assumed Parameters

Asset	**Mean (μ)**	**Standard deviation (σ)**
1	0.18	0.40
2	0.10	0.30
3	0.15	0.25
4	0.20	0.15
5	0.22	0.10

The change of the correlation has a direct implication of the portfolio composition, an issue discussed in the next section.

Table 11.2. The Correlation Matrices (ρij)

	Asset				
Asset	**1**	**2**	**3**	**4**	**5**
	PART A: n = 1 (assumed ρ_{ij})				
1	1.00	−0.25	0.00	0.40	0.15
2		1.00	−0.15	0.25	0.30
3			1.00	0.20	0.17
4				1.00	0.33
5					1.00
	PART B: n = 5 and ρ_{ij} is calculated by eq. (11.5)				
1	1.00	−0.19	0.00	0.35	0.13
2		1.00	−0.13	0.23	0.28
3			1.00	0.18	0.16
4				1.00	0.32
5					1.00
	PART C: n = 20 and ρ_{ij} is calculated by eq. (11.5)				
1	1.00	−0.06	0.00	0.21	0.07
2		1.00	−0.07	0.17	0.21
3			1.00	0.15	0.13
4				1.00	0.30
5					1.00
	PART D: n = 100 and ρ_{ij} is calculated by eq. (11.5)				
1	1.00	0.00	0.00	0.00	0.00
2		1.00	0.00	0.02	0.03
3			1.00	0.03	0.03
4				1.00	0.21
5					1.00

11.4 THE EFFECT OF THE INVESTMENT HORIZON ON THE COMPOSITION OF M-V PORTFOLIOS

Optimum M-V portfolio diversification is a function of the means, variances and co-variances of the rates of return on the various assets under consideration. All these parameters are affected by changes in the assumed investment horizon. Let us examine the effect of the horizon on the optimal asset composition of the Sharpe-Lintner unlevered portfolio. Assuming a set of one-period parameters, we employ the Sharpe-Lintner technique to solve for the optimal unlevered portfolio. We employ the hypothetical one-period parameters given in Table 11.1 and part A of Table 11.2, to calculate the optimal investment diversification. Column 1 of Table 11.3 gives the optimal one-period portfolio composition. Then, we employ eqs.(11.2), (11.3') and (11.5) to calculate the multiperiod parameters which, in turn, are employed to calculate the optimum M-V diversification for various horizons n (n > 1). As can be seen from Table 11.3, the assumed investment horizon has a dramatic effect on the recommended M-V diversification. To illustrate, suppose that we have two fund managers. Both believe in the same one-period parameters. However, one believes his/her investor's horizon is one month (n=1) and the other believes his/her investor should invest, say, for 10 months (n=10). As can be seen from Table 11.3, they would recommend quite different optimal portfolios. These findings have strong theoretical as well as practical implications. Therefore, it is obvious that investment managers should first try to estimate their customer's horizon and only then decide on the optimal portfolio composition. Moreover, more than one portfolio, probably several portfolios should be selected by the fund's manager according to the different horizons of investors.

So far, we have examined the effect of the horizon on optimal diversification with hypothetical data. Let us now turn to examine the horizon effect with actual data.

Table 11.3. The Optimum Investment Proportions xi (%) for Various Horizons (n)

	Number of Periods – n							
Assets	**1**	**2**	**5**	**10**	**15**	**30**	**50**	**100**
1	38.52	41.40	48.29	56.21	61.95	71.35	82.37	95.06
2	31.78	32.76	34.15	33.69	31.83	26.60	17.30	4.94
3	19.59	18.37	14.78	9.89	6.48	2.08	0.33	0.00
4	5.85	4.33	1.57	0.01	–0.27	–0.03	0.00	0.00
5	4.26	3.14	1.21	0.20	0.01	0.00	0.00	0.00
Total	100.0%	100.00%	100.0%	100.0%	100.00%	100.0%	100.0%	100.00%

Using Ibbotson Associate data for the years 1926–1990, Gunthorpe & Levy (1994)[10] solve for the optimal M-V diversification among various types of securities for various horizons. The one-period parameters (means, variances and correlations) are taken from Ibbotson Associates and the multiperiod parameters are calculated by the equations provided in this chapter. The results are presented in Table 11.4.

As can be seen from Table 11.4, also with actual data the investment proportions are strongly affected by the assumed investment horizon. Note that the fact that as the horizon increases, more of the investor's resources should be invested in intermediate government bonds, does not imply that investors should take less risk as the horizon increases: A very risky portfolio may be obtained by borrowing at 4% (the assumed riskless interest rate) and investing in government bonds yielding on average 5.1%. This strategy may provide the investor with a better risk-return profile than that obtained by investing a high proportion of wealth in stock (see also Ferguson and Simman, 1996).[11]

Table 11.4. The Investment Proportions for Various Holding Periods at a Riskless Interest of 4% and with Short Sale Constraints

	Holding Period in Years			
Assets	**1**	**5**	**10**	**20**
Common stocks	17.5	7.3	4.9	2.1
Small stocks	11.9	1.7	0.4	0.0
Long-term corporate bonds	0.0	0.0	0.0	0.0
Long-term government bonds	0.0	0.0	0.0	0.0
Intermediate government bonds	70.6	91.0	94.7	97.9
Total	100.0%	100.0%	100.0%	100.0%
Portfolio mean	7.7	32.9	72.4	184.7
Portfolio standard deviation	8.3	12.9	28.2	66.4

Source: Gunthorpe & Levy (1994).

11.5 THE EFFECT OF THE INVESTMENT HORIZON ON BETA

The CAPM suggests that beta is the risk measure of individual assets as well as portfolios. This measure of risk is also widely used by practitioners. Levhari and Levy

[10] Gunthorpe, D. and Levy, H., "Portfolio Composition and the Investment Horizon," *Financial Analysts Journal*, January-February, 1994, pp. 51–56.

[11] Ferguson, R. and Simman, Y., "Portfolio Composition and the Investment Horizon Revisited," *The Journal of Portfolio Management*, Summer 96, pp. 62–67.

(L&L) (1977)[12], have shown that under the i.i.d. rates of return assumption, the multiperiod beta, β_n, and the one-period beta, β_1, are related as follows:

$$\beta_n = \frac{\sum_{i=0}^{n-1}\binom{n}{i}\beta_1^{n-i}(\sigma_m^2)^{n-i}(\mu_i\mu_m)^i}{\sum_{i=0}^{n-1}\binom{n}{i}(\sigma_m^2)^{n-i}(\mu_m^2)^i} \qquad (11.6)$$

where n denotes the number of periods, μ_i the one-period mean return of the asset under consideration, and μ_m and σ_m^2 are the one-period mean and variance of the market portfolio, respectively (For proof of eq. 11.6 see Levhari & Levy 1977). From eq. (11.6), we see that β_n is generally different from β_1. Assuming that the CAPM holds for n = 1, L&L show that:

a) For neutral stocks with $\beta_1 = 1$, $\beta_n = 1$ (i.e., the assumed horizon does not affect the calculated beta).

b) For one-period aggressive stocks, $\beta_n > \beta_1 > 1$.

c) For one-period defensive stocks, $\beta_n < \beta_1 < 1$.

If the CAPM is assumed to hold for n-periods, we obtain similar results for betas calculated for shorter horizons. To be more specific, for neutral stocks with $\beta_n = 1$, β_1 corresponding to a shorter horizon, will also be equal to unity. For aggressive stocks with $\beta_n > 1$, $\beta_1 < \beta_n$ and for defensive stocks with $\beta_n < 1$, $\beta_1 > \beta_n$. These theoretical findings imply that the beta of aggressive stocks increases the longer the horizon, and the beta of defensive stocks decreases the longer the horizon. Table 11.5 demonstrates these results with a sample of ten defensive stocks (Part A) and ten aggressive stocks (Part B).

Even though the i.i.d. assumption does not necessarily hold with actual data, a very strong result is obtained: In most cases, the betas of defensive stocks decrease as the assumed horizon increases and even becomes negative for very long horizons. The opposite holds for aggressive stocks. It should be emphasized that all these betas are calculated using the same 20-year data set. The only difference is that these 20 years are divided in different ways corresponding to the various investment horizons. These results have very strong implications for the CAPM as well as for practitioners

[12] Levhari, D. and Levy, H., "The Capital Asset Pricing Model and the Investment Horizon," *Review of Economics and Statistics*, 59, 1977, pp. 92–104.

Table 11.5. The Beta of Ten Defensive and Aggressive Stocks Corresponding to Various Horizons

A. Defensive Stocks

Horizon (months)	Idaho Power Corp.	American Can Corp.	National Dairy Products	P. Lorrillard Corp.	American Tobacco	Borden Inc.	Abbott Laboratory	Standard Brands	Greyhound Corp.	Continental Can
1	.4282	.5167	.5281	.6166	.6296	.6372	.6576	.6650	.6752	.7807
2	.4012	.4886	.4655	.5711	.4652	.5912	.5717	.6147	.6651	.7081
3	.3796	.3755	.4475	.3496	.4993	.5684	.5892	.5978	.5773	.6520
4	.3329	.3311	.3400	.4881	.3697	.6142	.5284	.6397	.5340	.6109
5	.1881	.2631	.4428	.2604	.3283	.3449	.6319	.4331	.6709	.5882
6	.3862	.3402	.4119	.4253	.3706	.4330	.3811	.6112	.5294	.6616
8	.4322	.0621	.5309	.4815	.3020	.4627	.2398	.7987	.4907	.3967
10	.2312	.1236	.4777	−.0656	.2438	.4272	.4729	.5325	.4800	.4209
12	.2367	−.0118	.3511	−.4615	.0364	.3390	.4227	.4289	.6188	.2834
15	.1556	.0702	.4544	−1.0612	−.0365	−.0561	.1243	.2008	.1541	.3526
16	.3016	.2049	.5016	−1.0387	.1400	.2723	.1463	.7473	.1719	.4753
20	.1142	−.2563	.3283	−1.1855	−.1060	.2336	.0247	.4002	.2378	.3307
24	.1068	−.2690	.3996	−2.0036	.1657	.0849	.2474	.3771	.7826	.5319
30	.2210	.0101	.2781	−2.8251	.1187	.1360	−.3863	−.0150	−.5545	.3536

A. Aggressive Stocks

Horizon (months)	Evans Products Corp.	Cerro Corp.	Colt Industries	Coinelco Inc.	Anaconda Corp.	Bethlehem Steel	United States Steel	Carpenter Steel	Hooker Chemical	Medusa Portland
1	1.8252	1.5876	1.5353	1.4091	1.2904	1.1664	1.1657	1.1093	1.0650	1.0108
2	1.8002	1.6050	1.7738	1.5725	1.3379	1.1525	1.1778	1.0790	1.1205	1.1431
3	2.1009	1.5389	2.0234	1.6586	1.1890	1.2046	1.1777	1.1384	1.1241	1.1164
4	1.9293	1.6299	1.8991	1.7332	1.3045	.9997	1.0952	1.0051	1.1947	1.3197
5	1.8935	1.6583	2.2407	2.2077	1.1977	1.0340	1.1037	.9627	1.1016	1.2757
6	1.9414	1.6056	2.1994	2.2806	1.2368	1.1495	1.2811	1.1078	1.0907	1.0505
8	2.0450	1.7603	1.7112	2.6780	1.2632	1.1889	1.2164	1.1081	1.2390	1.5629
10	2.0446	1.4159	2.3771	3.2640	1.1683	1.2068	1.1751	1.1476	1.1098	1.4558
12	3.4909	1.6676	1.3451	3.4419	1.4399	1.3672	1.6645	1.6671	1.4250	1.8580
15	3.2813	2.0732	1.5511	2.2901	1.5936	1.9100	1.7893	1.8270	1.3184	2.7472
16	2.2692	1.8526	1.8232	3.4655	1.3207	1.5274	1.4989	1.0993	1.4488	2.0514
20	1.8549	2.0076	4.0385	3.0123	.9667	1.2587	1.0467	.6608	1.1361	1.7524
24	4.1552	3.1044	2.2391	2.6956	1.5806	1.2426	1.3561	1.4828	1.6444	1.2072
30	8.1045	4.0495	2.2277	5.2007	2.1563	3.6042	2.9569	1.7334	1.4476	3.2152

who classify investment risk using beta. For example, suppose that we want to estimate the equity cost of capital of Medusa Portland. Using monthly data, beta is about 1 and using annual data, beta is about 1.85. Hence, the cost of equity will be much higher with annual data than the (annualized) cost of capital obtained with monthly data.

11.6 STOCHASTIC DOMINANCE AND THE INVESTMENT HORIZON

The results of stochastic dominance analyses are also affected by the assumed investment horizon. However, as we shall see in this section, under i.i.d. assumptions, the size of SD efficient sets decreases as the assumed horizon increases. This is in contrast to the results obtained for the M-V efficient set. In this section we provide the results and in Section 11.7 we discuss the results and solve the paradox that they seem to present.

To demonstrate the horizon effect on SD analyses, for simplicity, let us assume two periods only. Denote by x_1^F and x_2^F the returns (1 + rate of return) of option F in period one and two, respectively. Similarly, the returns on option G are x_1^G and x_2^G. The two-period returns are denoted by: $x^F = x_1^F x_2^F$ and $x^G = x_1^G x_2^G$ and their cumulative distribution is denoted by F^2 and G^2, respectively.

Assuming independence over time (random walk) and denoting the two-period return by x, the two-period cumulative distribution F^2 is given by:

$$F^2(x) = \int_0^{\infty} \int_0^{x/x_1} f_1(t_1) f_2(t_2) dt_1 dt_2$$

or:

$$F^2(x) = \int_0^{\infty} F_2(x/t_1) f_1(t_1) dt_1 \qquad (11.7)$$

Similarly, the two-period cumulative distribution of option $G^2(x)$, is given by:

$$G^2(x) = \int_0^{\infty} G_2(x/t_1) g_1(t_1) dt_1 \qquad (11.8)$$

where $g_i(t_i)$, $f_i(t_i)$ (i = 1, 2) are the density functions of the return in period 1 and 2 of the two options, respectively, with cumulative distributions F_i and G_i (i = 1, 2), respectively. In order to examine whether dominance prevails in each period, we need to compare F_1 to G_1 in the first period and F_2 to G_2 in the second period. With a two-period horizon, to establish dominance, we need to compare the more complicated functions $G^2(x)$ to $F^2(x)$ given by eqs. (11.7) and (11.8). With an n-period horizon, the cumulative distributions become much more complex. Levy (1973)[13] uses eqs. (11.7) and (11.8) to establish the following two theorems:

Theorem 11.1:

Let $F^n(x)$ and $G^n(x)$ be the cumulative distributions of two n-period options where n is the number of periods and x is the product of the returns corresponding to each period ($x = x_1, x_2...,x_n$). Then, a sufficient condition for F^n dominance over G^n by FSD is that such dominance exists in each period, namely $F_i(x_i) \leq G_i(x_i)$ for all i, where i denotes the period (i = 1, 2,..,n) and there is at least one strict inequality. Theorem 11.2 extends the results of Theorem 11.1 to the case where the risk aversion is assumed.

Theorem 11.2:

Using the same notation as in Theorem 11.1, a sufficient condition for dominance of F^n over G^n by SSD is that such dominance exists in each period, namely:

$$\int_0^{x_i}[G_i(t_i) - F_i(t_i)]dt_i \geq 0 \qquad \text{for } i = 1, 2,..,n$$

and there is at least one strict inequality.

If, in addition to the independence, we also assume stationarity over time, then we can conclude that for any two horizons n_1 and n_2 where $n_2 > n_1$, the number of options in the n_2 efficient set (FSD as well as SSD) will not be larger than the number of options in the n_1 efficient set. This conclusion stems directly from Theorems 11.1 and 11.2. To see this, recall that under the stationarity assumption, we have $F_1 = F_2... = F_n$ and $G_1 = G_2 =....= G_n$. Thus, if F_iDG_i (by FSD, by SSD, or by TSD, see discussion below) for some period i, by the stationarity assumption, also F_jDG_j for all other periods j. Hence by the above two theorems, F^n dominates G^n, and any option eliminated from the one-period efficient set is also eliminated from the n-period efficient set. Hence, the number of elements in the long-horizon efficient set cannot be larger than the number of elements in the short-horizon efficient set.

It is interesting to note that numerical examples can be found in which the number of elements in the efficient set strictly decreasrd as the horizon increases. Consider the following two-period example, where $F_1 = F_2$ and $G_1 = G_2$.

---

[13] Levy, H., "Stochastic Dominance, Efficiency Criteria, and Efficient Portfolios: The Multi-Period Case," *American Economic Review*, 63, 1973, pp. 986–994.

Example:

One-Period Distribution			
F_1 (or F_2))		G_1 (or G_2	
Outcome	**Probability**	**Outcome**	**Probability**
1	¼	2	1/2
4	¾	10	1/2

Assuming independence over time, the two-period distributions F^2 and G^2 are:

Two-Period Distribution			
F^2		G^2	
Outcome	**Probability**	**Outcome**	**Probability**
1	1/16	4	1/4
4	6/16	20	1/2
16	9/16	100	1/4

Because F_1 and G_1 (or F_2 and G_2) intersect, both of them are included in the one-period FSD efficient set. It is easy to verify that the two one-period distributions are also included in the SSD efficient set. However, $F^2(x) \le G^2(x)$ for every x (with a strict inequality for some value) and, therefore, F^2DG^2 by FSD and, *a fortiori*, by SSD. This numerical example shows that portfolios not eliminated from the one-period SD efficient set might be eliminated from the n-period efficient set.

This example shows that the FSD and SSD efficient set, contrary to the M-V efficient set, decrease (in the weak sense) as the horizon increases.

Levy & Levy (1982)[14] add two more results to the above results:

a) A statement similar to the one contained in Theorems 11.1 and 11.2 holds also for TSD.

b) Extension to the case where a riskless asset is available. In case b) the mix $x_i(\alpha)$ is defined as follows: $x_i(\alpha) = \alpha_i x_i + (1 - \alpha_i) r_i$ where i denotes the period (i = 1, 2,..,n), α_i is the investment proportion in the risky asset with return x_i, and r_i is the riskless interest rate available in period i. Thus, r_i may vary from one period to another. Using this definition, Levy & Levy (1982) prove that

[14] Levy, H. and Levy, A., "Stochastic Dominance and the Investment Horizon with Riskless Asset," *Review of Economic Studies*, 49, 1982, pp. 427–438.

the sufficient condition for multiperiod dominance of F^n over G^n, when mixture with the riskless asset is allowed, is that there exists a non-negative value α_i such that $F^{(\alpha_i)}$ dominates G_i in each period i, that is $F(x_i(\alpha))$ dominates G. This statement is intact for FSD, SSD and TSD.

Thus, with a riskless asset, too, with interest that may vary from one period to another, with stationarity of the random variables across periods, the size of the efficient set is non-increasing with increase in the investment horizon.

11.7 CONTRASTING THE SIZE OF THE M-V AND SD EFFICIENT SET

Tobin shows that if rates of return are i.i.d., then the size of the efficient set will be a non-decreasing function of the investment horizon whereas SD analysis indicates the opposite: the efficient set will be a non-increasing function of the investment horizon. Actually, it is easy to show that by M-V, the size of the efficient set is strictly increasing whereas with say, SSD, it is strictly decreasing. This is illustrated in the next example.

Example:

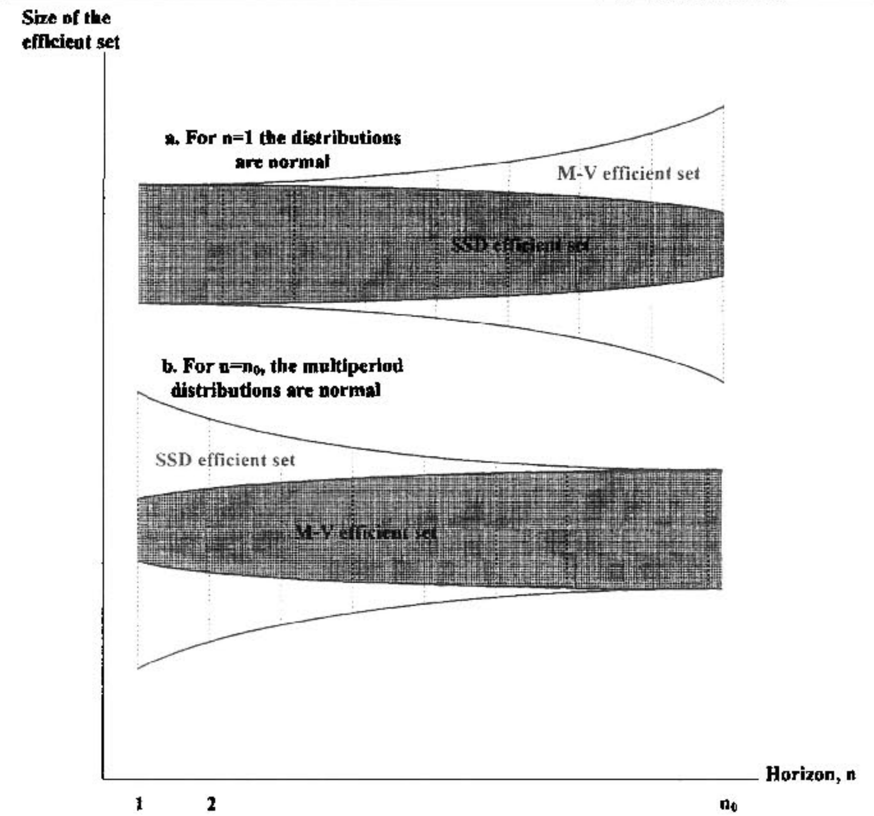
	F_1 (and F_2)		**G_1 (and G_2)**	
	Outcome	**Probability**	**Outcome**	**Probability**
	1	1/4	2	1/2
	3.1	3/4	3.2	1/2
Expected value:	2.575		2.6	
variance:	~ 0.827		0.36	

	F^2		**G^2**	
	Outcome	**Probability**	**Outcome**	**Probability**
	1	1/16	4	1/4
	3.1	6/16	6.4	1/2
	9.61	9/16	10.24	1/4
Expected value:	6.95		6.76	
variance:	7.31		~ 5	

This example reveals that the one-period M-V efficient set includes one element (F) whereas the two-period M-V efficient set includes the two elements exactly as claimed by Tobin. Contrary to the M-V results, the SSD one-period efficient set includes two elements whereas the two-period SSD efficient set includes only one element. Thus, the M-V and SSD analysis reveals conflicting results regarding the effect of the horizon on the size of the efficient set. Generalization of the results of this example is given by Figure 11.2 which compares the size of the M-V and SSD efficient sets as a function of the assumed horizon (under i.i.d.).

How can Tobin's M-V results be reconciled with the contradictory results obtained for SSD? Both are technically correct, but only SSD is theoretically justified. Recall that if normal distributions and risk aversion are assumed, M-V and SSD yield identical efficient sets. Hence, if for n=1 (or for that matter for any other value of n), we assume normal distributions, M-V and SSD will lead to identical efficient sets for that particular horizon, and both rules will be consistent with the expected utility paradigm. However, as n increases, SSD will continue to

Figure 11.2: The size of the M-V efficient set: (a) When the one period distributions are normal, (b) When the multiplied distributions are normal



yield results that are consistent with the expected utility maximization paradigm, but M-V will not because the normality assumption will have been violated. It can be shown that if x_1 and x_2 are normally distributed (hence in each one-period, M-V and SSD efficient set will coincide) the product $x_1 \cdot x_2$ will not be normally distributed anymore. Hence, the increase in the M-V efficient set, though technically correct, is misleading because it includes inefficient options from the risk averse investors point of view. Alternatively, it can be assumed that $x_1 \cdot x_2$ (or $\Pi\ x_i$) is normally distributed; hence the two-period (or n-period) M-V and SSD efficient set will coincide (see lower part of Figure 11.2) but, once again, x_1 and x_2 cannot be normally distributed. Therefore, under such an assumption, in each period, the M-V and SSD efficient sets will not coincide, and only the results produced by SSD will be correct because the SSD criterion does not assume normality and it conforms with the expected utility paradigm.

11.8 SUMMARY

The assumed investment horizon plays a key role in portfolio analysis. This is obvious in the case where distributions are not stationary and are dependent over time. However, we show in this Chapter that M-V efficiency analysis, M-V portfolio diversification, Sharpe's reward-to-variability ratio, beta, and correlations are all dependent on the assumed horizon even if distributions are independent and identical over time (i.i.d.), let alone if the i.i.d. assumption does not hold.

Stochastic dominance efficiency analyses are also dependent on the assumed horizon. The size of the M-V efficient set does not decrease (and may increase) with increase in the investment horizon: The opposite is true with SD efficient set. This apparent contradiction between M-V and SSD is resolved once we recall that if the one-period distributions are assumed to be normal, the multiperiod distributions cannot be normal. Thus, M-V and SSD coincide for $n=1$ (if the one-period distribution is assumed to be normal) but for $n > 1$, the two efficient sets may diverge. SD analysis is distribution-free and, therefore, it is correct and consistent with expected utility paradigm, but M-V is not consistent with this paradigm. Given that the one-period distributions are normal, the M-V multi-period efficient set is incorrect because the distributions are no longer normal and, therefore, it will include portfolios that are SSD inefficient.

Key Terms

Identical Independent Distribution (i.i.d)

Random Walk Hypothesis

Reward to Variability (R/V) Ratio

Defensive Stock

Aggressive Stock

Investment Horizon

Ex-Post

Ex-Ante

12

THE CAPM AND STOCHASTIC DOMINANCE

The M-V and stochastic dominance paradigms represent two distinct branches of expected utility, each implying a different technique for portfolio investment selection. Each paradigm has its pros and cons. The advantage of the M-V approach is that it provides a method for determining the optimal diversification among risky assets which is necessary in establishing the Capital Asset Pricing Model (CAPM) (see Sharpe [1964][1] and Lintner [1965][2]). The main disadvantage of the M-V paradigm is that it relies on the assumption of normal distribution of returns, an assumption not needed for SD rules. The normality assumption is obviously inappropriate for assets traded in the stock market because asset prices cannot drop below zero (−100% rate of return) whereas the normal distribution is unbounded. As the equilibrium risk-return relationship implied by the CAPM has very important results, it has been developed under other frameworks which do not assume normality. Levy (1977)[3] has shown that technically the CAPM holds even if all possible mixes of distributions are log-normal (bounded from below by zero). However, with discrete time models, a new problem emerges: if x and y are lognormally distributed, a portfolio z, where $z = \alpha x + (1 - \alpha) y$, will no longer distribute lognormally. Merton (1973)[4] assumes continuous-time portfolio revisions and shows that under this assumption, the terminal wealth will be lognormally distributed and the CAPM will hold in each single instantaneous period. By employing the continuous portfolio revision, the nagging problem of additivity of lognormal distributions, which characterizes discrete models disappears. However, the disadvantage of the continuous time model is that the CAPM result breaks down if minor transaction costs, no matter how small, are incorporated.

The advantage of stochastic dominance over the M-V approach is that it is distribution-free: There is no need to make assumptions regarding the distributions of rates of return. In addition, transaction costs can be easily incorporated into the stochastic dominance analysis. Its main disadvantage is that, to date, no method has been found to determine SD efficient diversification strategy and it does not provide a simple risk-return relationship such as implied by the CAPM.

---

[1] Sharpe, W.F., "Capital Asset Prices: A Theory of Market Equilibrium," *Journal of Finance*, September 1964, pp. 425–42.

[2] Lintner, J., "Security Prices, Risk and Maximal Gains from Diversification," December 1965, pp. 587–615.

[3] Levy, H., "Multi-Period Stochastic Dominance with One-Period Parameters, Liquidity Preference and Equilibrium in the Log Normal Case," in Alan Blinder and Philip Friedman (eds.) *Natural Resources Uncertainty, Dynamics and Trade: Essays in Honor of Rafael Lusky, Academic Press*, 1977, pp. 91–111.

[4] Merton, R.C., "An Intertemporal Capital Asset Pricing Model," Econometrica, 41, September 1973, pp. 867–887.

In practice, investors do not revise their portfolios continuously, nor do they have identical investment horizons. Therefore, multiperiod discrete models are called for. In this Chapter, which relies heavily on Levy (1973)[5] and in particular on Levy & Samuelson (1992)[6], we present cases where the CAPM holds in a discrete multiperiod setting (i.e, where proportional transaction costs can easily be incorporated). Interestingly, multiperiod SD and M-V arguments are simultaneously employed to prove the CAPM in these cases.

12.1 THE CAPM WITH HETEROGENEOUS INVESTMENT HORIZONS

In the one-period model with normal distribution, the CAPM holds in both the M-V and the SSD framework, because, in this case, these two rules coincide.

In this section we extend the analysis to a multiperiod setting. Investors with various horizons are assumed but portfolio revisions are allowed only at the end of each period. More specifically, the following assumptions are made:

(1) Investors maximize expected utility $EU(W_T)$, where W_T is the terminal wealth at the end of period T. The terminal date T is permitted to vary across investors. However, investors may revise their portfolio at the end of each period, that is, (T −1) times.

(2) Investors are risk averse.

(3) The rates of return on each security i are independent (but not necessarily stationary) over time. Thus:

$$g(R_1, R_2, \ldots, R_T) = f_1(R_1)f_2(R_2) \cdots f_T(R_T)$$

where g is the joint density function and $f_i(R_i)$ (i = 1, 2,...,T) is the density function of a single return period. The distributions of returns can be nonstationary over time.

Four cases in which the CAPM holds, in addition to cases of the one-period Sharpe-Lintner CAPM, Merton continuous time CAPM, and Levy (1973)[7] (1977)[8] lognormal discrete case are given below:

[5] Levy, H., "Stochastic Dominance Among Log-Normal Prospects," *International Economic Review*, 14, 1973, pp. 601–614.

[6] Levy, H. and Samuelson, P., "The Capital Asset Pricing Model with Diverse Holding Periods," *Management Science*, 38, 1992, pp. 1529–1540.

[7] Levy, H., 1973, Ibid.

[8] Levy, H., 1977, Ibid.

a) Quadratic Utility Function

The Sharpe-Lintner CAPM can be derived by assuming either a quadratic utility function or normally distributed returns. In a multiperiod framework, the quadratic utility assumption also leads to the two-fund Separation Theorem, and the CAPM is implied. Investors are shown to continue to make their decisions by the meanvariance rule, but each investor now faces different distributions of terminal wealth because the horizons are heterogeneous. However, at the beginning of each revision period, all investors hold the same portfolio of risky assets and, therefore, the CAPM is implied in each period. These results do not hold if portfolio rebalancing is not allowed, even with quadratic utility functions defined on multiperiod wealth, because the risky portfolio that is optimal for, say, a T = 2 horizon, will generally not be optimal for a T = 1 horizon and the separation property breaks down.
Theorem 12.1 states the CAPM results for the quadratic utility case:

Theorem 12.1:

Suppose that each investor is characterized by a quadratic utility function, but investors have diverse holding periods. If portfolio returns are independent over time, and portfolio rebalancing is allowed, then the CAPM will hold for each single period. For a proof of this Theorem, see Levy & Samuelson (1992)[9].

The quadratic utility function has a few drawbacks and, therefore, we will not elaborate on this case. We elaborate on the other three cases which are economically more interesting and where stochastic dominance and M-V tools are employed simultaneously to prove that the CAPM holds in each single period.

b) Single-Period Normal Distributions

In this section, we assume that the return corresponding to each single period is normal, and that investors maximize $EU(W_T)$ where U is concave, and the terminal wealth W_T is given by:

$$W_T = \prod_{t=1}^{T} \left(\sum_{i=1}^{N+1} x_{it} R_{it} \right) \tag{12.1}$$

where R_{it} is the return on asset i in period t, x_{it} is the investment proportion in the ith asset in period t, N is the number of risky assets, and N + 1 stands for the riskless asset. Because, by assumption, R_{it} is normally distributed, W_T, which is a *product* of normal random deviates, is *not* normally distributed. We now prove that for expected utility maximization, $EU(W_T)$, the Sharpe-Lintner CAPM will hold in each single period regardless of the features of the distribution of W_T, and even when investors

[9] Levy, H. and Samuelson, P., 1992, Ibid.

vary with respect to terminal date T. The results hold for stationary as well as nonstationary distributions. The riskless interest rate may also vary over time.

Theorem 12.2:

Let F_i and G_i (i = 1, 2,...,T) be the one-period cumulative normal distributions, and let F^T and G^T denote the multiperiod distributions of two distinct options F and G. The riskless interest rate is denoted by r. Investors are assumed to be risk-averse and to maximize expected utility of terminal wealth, but they are allowed to revise their portfolios in every period (i.e., T – 1 times). When the portfolio returns are independent over time, the CAPM is intact in each period even when investors have different holding periods T, with no assumptions regarding terminal wealth distribution.

Proof:

Figure 12.1 provides the M-V efficiency frontier where m is the tangency portfolio. A linear combination of portfolio m and the riskless asset r exists which dominates any linear combination of the riskless asset and other risky portfolio by the M-V rule for T = 1 (a single period, see Figure 12.1). However, because by assumption, the one-period returns are normally distributed, such dominance also holds by Second Degree Stochastic Dominance (SSD).

Let F_i (i= 1, 2,...,T) denote the single-period distributions of a linear combination of m and the riskless asset, and let G_i (i=1, 2,...,T) denote a linear combination of the riskless asset, and *any other* risky portfolio. For any G_i, there is at least one F_i which dominates it by SSD for each period i (for example, in Figure 12.1, for selected portfolio G_i given by m_1, portfolio m_2 is selected as F_i), and, therefore, we have that F_i dominates G_i by SSD (for i = 1, 2,...,T). Employing Theorem 11.2 in Chapter 11, we obtain that F^T dominates G^T by SSD. Thus, regardless of the length of the investment horizon T, and regardless of the shape of the multi-period distributions F^T and G^T, in each single period, investors will diversify between m and r, and both the two-fund Separation Theorem and the CAPM hold.

Note that in order to prove dominance of the terminal wealth distribution corresponding to portfolio m (plus the riskless asset) over any other terminal wealth distribution, we first employ the equivalence of SSD and the M-V rule for each single period (normal distributions), and then the relationship between single and multiperiod stochastic dominance.

In the proof, the distributions are not assumed to be stationary over time. Moreover, m and r may vary from one period to another (see Figure 12.1). To see this, simply denote by F_i the linear combinations of r_i and m_i (where m_i is the tangency portfolio corresponding to the ith period). Thus, for any selected strategy. G_i (i = 1, 2,...,T – 1), there is a strategy F_i (i = 1, 2,...,T – 1) which dominates it by SSD where F_i and G_i may vary from one period to another. By Theorem 11.2 in Chapter 11, this implies that F^T will also dominate G^T, where F^T and G^T stand for the distributions of terminal wealth induced by strategies F_i and G_i, respectively. In each period, all investors hold the same combination of risky assets (regardless of the one-period parameters and the one-period riskless interest rate) hence, both the two-fund Separation Theorem and the CAPM hold.

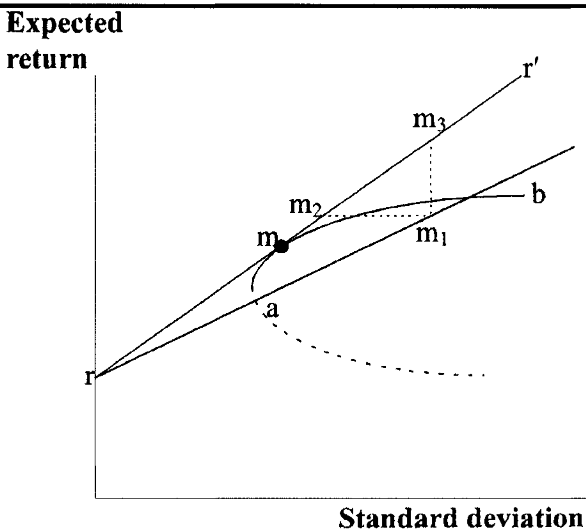
c) Multiperiod Normal Distributions

Let us now show that the two-fund Separation Theorem and the CAPM follow if the distributions of terminal wealth are normal for all W_T (even if for T = 1 the distributions are not normal).

Theorem 12.3:

Suppose that the distributions of terminal wealth F^T and G^T are normal. If (T−1) revisions are allowed and portfolio returns are independent over time, then the

Figure 12.1: The one-period efficient frontier



CAPM will hold for every single period even though the distribution in each single period is not normal.[10]

[10] If the multiperiod distributions are assumed to be normal, some constraints need to be imposed on the one-period distributions. For example, the one-period distributions cannot be normal because a product of normal deviates (terminal wealth) is not normally distributed.

Proof:

To see this, let Figure 12.1 correspond to the rebalancing of the ith single-period portfolio. With the one-period parameters, m_2 dominates m_1 by the M-V rule (but not by expected utility corresponding to the one-period returns, because these returns do not distribute normally). However, such dominance implies dominance by the M-V rule with the multiperiod parameters, too (see equations 11.1 and 11.3 in Chapter 11), irrespective of the investment strategy in all other single periods. Because W_T is assumed to be normal, choosing m_2 in period i will also maximize the expected utility defined on terminal wealth regardless of the distribution characteristics of each single period returns. Recall that for normal distributions SSD and MV rules are equivalent. Because m_2 is a linear combination of portfolio m and the riskless asset, investors will hold portfolio m of risky assets. Because the same logic holds for every period i, a portfolio of risky assets such as m will be selected in each period and the two-fund Separation Theorem will hold for all $i = 1, 2, \ldots, T - 1$ which completes the proof.

Note that W_T can be assumed to be normally distributed for a given $T = T_0$, but normality is violated for $T \neq T_0$. Although, technically, normality of terminal wealth implies the CAPM, in this case diverse holding periods are impossible because they would contradict the normality assumption. (Recall that if x_1 and x_2 are normally distributed, $x_1 \cdot x_2$ will no longer be normally distributed.) Thus, the CAPM also holds when $T - 1$ revisions are allowed and W_T is normally distributed, albeit with the constraint that $T = T_0$ for all investors. This case is analogous to the single-period Sharpe-Lintner framework except that here we allow $T - 1$ portfolio revisions. Finally, if $T_0 = 1$, there will be no revisions, and these cases will collapse into Sharpe-Lintner's single-period model.

d) Lognormal Distributions

(1) Stationary Distributions.

In this section, we continue to employ the assumptions of independence and risk-aversion. However, the distributions of terminal wealth W_T are now assumed to be *lognormal* and diverse holding periods are allowed. First assume stationary return distributions and that portfolio returns are independent over time. This assumption is relaxed in section (2) below.

The assumption of lognormality for W_T may raise some objections.[11] Nevertheless, although it cannot be justified theoretically, there is empirical support that

[11] If the rates of return on the risky assets are independent over time, the terminal wealth cannot be precisely lognormal. To see this, recall that if W_T is lognormally distributed, then log $(W_T) = \Sigma_{t=1}^{T} \log(z_t)$ will be normally distributed where z_t is the return on the portfolio in period t. However, z_t is independent over time (by assumption), and, therefore, each log (z_t)

returns can be approximated quite well by lognormal distributions, and even the sums of lognormal variables approach lognormal distribution. Lintner (1972)[12] who extensively analyzes the shape of the distribution of returns concludes,

> "...on the basis of these simulations that the approximation to log-normally distributed stock is sufficiently good that theoretical models based on these twin premises should be useful in a wide range of applications and empirical investigations" (see also Ohlson and Ziemba [1976][13] and Dexter, Yu and Ziemba [1974][14]).

Let us now prove that the CAPM holds in each single period when W_T is lognormally distributed and diverse holding periods are allowed. Let us first introduce a few notations. Denote the return (i.e., 1 + the rate of return) on investment i in period t by R_{it}, and the corresponding investment proportion by x_{it}. Then the portfolio return in period t is given by:

$$Z_t = \sum_{i=1}^{N+1} x_{it} R_{it} \tag{12.2}$$

where t = 1, 2,...,T, N is the number of risky assets, (N + 1) is the riskless asset, and T is the number of periods. The terminal wealth at the end of T investment periods for one dollar of investment is given by W_T:

$$W_T = \prod_{i=1}^{T} Z_t = \prod_{t=1}^{T} \left(\sum_{i=1}^{N+1} x_{it} R_{it} \right) \tag{12.3}$$

and W_T is assumed to be lognormally distributed for all possible values

$$t = 1, 2, \ldots, T.$$

We use Theorem 12.4 below in the proof of the CAPM.

---

must be normally distributed, or z_t must be lognormally distributed for all t = 1, 2,...T. However, for each single period, if we have, say three portfolios z(1), z(2) and z(3), all these three portfolios must be lognormally distributed. Suppose that z(3) is a linear combination of z(1) and z(2), then z(3) cannot be lognormally distributed which contradicts the assertion that z_3 is lognormally distributed. Thus, the fact that z_1, z_2 and z_3 are not all lognormal implies that W_T cannot be precisely lognormal. Therefore, we may have distributions which are approximately, but not precisely, lognormal. The error in expected utility terms is more relevant than the error in the probability distribution terms (see Markowitz [1991] and Dexter, Yu and Ziemba [1974]).

[12] Lintner, J. "Equilibrium in a Random Walk and Lognormal Securities Market," Discussion Paper, No. 235, Harvard Institute of Economic Research, July 1972.

[13] Ohlson, J.A. and W.T. Ziemba, "Portfolio Selection in a Lognormal Market when the Investor has a Power Utility Function," *Journal of Financial and Quantitative Analysis*, 11, 1976, pp. 51–57.

[14] Dexter, A.S., J.S.W. Yu and W.T. Ziemba, "Portfolio Selection in a Lognormal Market when the Investor has a Power Utility Function: Computational Results," *Stochastic Programming*. Dempster, M.A., (ed.), Academic Press, London, 1980.

Theorem 12.4:

Let E_{F_1}, $\sigma^2_{F_1}$ and E_{G_1}, $\sigma^2_{G_1}$ be the one-period expected return and variance of two alternative prospects F and G, respectively and $E_{F_1} > 0$, $E_{G_1} > 0$. If the portfolio returns are independent over time, the necessary and sufficient conditions for dominance of F_T over G_T for all nondecreasing concave utilities U will be:

$$E_{F_1} \geq E_{G_1} \text{ and } C_{F_1} \equiv \sigma_{F_1} / E_{F_1} \leq \sigma_{G_1} / E_{G_1} \equiv C_{G_1} \tag{12.4}$$

with at least one strict inequality, where C_{F_1} and C_{G_1} are the one-period coefficient of variation of distributions of F_1 and G_1, respectively. F_1 and G_1 denote the one-period return distributions and F_T and G_T denote the corresponding T-period return distributions (namely, the distributions of the terminal wealth, which are assumed to be lognormal given by equation 12.3).

In other words, this theorem claims that instead of looking at the *multiperiod* mean-coefficient of variation rule, the *one-period* mean-coefficient of variation rule should be applied in each portfolio revision, even though the utility is not defined on the one-period distributions.

Proof:

Employing the assumption of the stationarity of returns, the multiperiod expected return and variance are given by $E_{F_T} = \left(E_{F_1}\right)^T$, and $\sigma^2_{F_T} = (\sigma^2_{F_1} + E^2_{F_1})^T - E^{2T}_{F_1}$, respectively (for the relationship between the multiperiod and one-period variances, see equations 11.3, 11.3′ and 11.3″ in Chapter 11); hence, the T-period coefficient of variation C^2, is given by:

$$C^2 = \frac{\sigma^2_{F_T}}{E^2_{F_1}} = \frac{(\sigma^2_{F_1} + E^2_{F_1})^T - E^{2T}_{F_1}}{E^{2T}_{F_1}} = \left(\frac{\sigma^2_{F_1}}{E^2_{F_1}} + 1\right)^T - 1 = (1 + C^2_{F_1})^T - 1$$

and the parameters of distribution G are defined similarly. Thus:

$$E_{F_T} \geq E_{G_T}\text{, if and only if } E_{F_1} \geq E_{G_1}\text{, and :}$$

$$\frac{\sigma_{F_T}}{E_{F_T}} \leq \frac{\sigma_{G_T}}{E_{G_T}} \text{ (or } C_{F_T} \leq C_{G_T}\text{), if and only if } \frac{\sigma_{F_1}}{E_{F_1}} \leq \frac{\sigma_{G_1}}{E_{G_1}} \text{(or } C_{F_1} \leq C_{G_1}\text{),}$$

and, if there is a strict inequality on the left-hand side of each equation, then there will be a strict inequality on the right-hand side, too, and vice-versa. However, the mean-coefficient of variation is an optimal criterion for risk-averse investors confronted by lognormal (multiperiod) distributions (see Chapter 6, Theorem 6.5), therefore, the one-period mean-coefficient of variation is an optimal single-period investment revision criterion for all risk-averters who maximize expected utility $EU(W_T)$ where W_T is the terminal wealth. We use this result to derive the CAPM in each single revised date.

Theorem 12.5:

Assume that the T-period returns are lognormally distributed, the portfolio returns are independent over time, and that investors are allowed to revise their investment portfolios (T – 1) times. Then the two-fund Separation Theorem holds and the Sharpe-Lintner CAPM is intact in each single period even with diverse holding periods.

Proof:

We show here that when investors revise their portfolios (T – 1) times and the terminal wealth (i.e., $W_T = \prod_{t=1}^{T} z_t$) is lognormally distributed, the two-fund Separation Theorem, and hence the CAPM, holds. In the classic CAPM, it is assumed that all investors have the same holding period. In the multiperiod lognormal case, the holding period may vary across investors. Nevertheless, all investors for a given portfolio's expected return, minimize the one-period coefficient of variation and, therefore, they adopt the same diversification strategy in risky assets. However, unlike the classic CAPM, we will show below that the price of a unit of risk varies from one investor to another. Let us now prove the CAPM for this case.

Looking at the *single-period parameters*, each investor who maximizes the expected utility of terminal wealth should minimize the coefficient of variation for a given expected return (see Theorem 12.4). Thus, we need to solve the following Lagrange function:

$$L = \frac{1}{E}\left(\sum_{i=1}^{N} x_i^2 \sigma_i^2 + \sum_{\substack{j=1 \\ j \neq i}}^{N} \sum_{i=1}^{N} x_i x_j \sigma_{ij} \right)^{1/2} + \lambda \left[E - x_i \mu_i - \left(1 - \sum_{i=1}^{N} x_i \right) r \right]$$

where E is the single-period portfolio mean return, (μ_i, σ_i^2), the mean rate of return and variance of the ith security, respectively, and r stands for the one-period riskless interest rate. Minimization of the coefficient of variation for a given mean return E yields the following N + 1 equations with N + 1 unknowns:

$$(12.5)\quad \left.\begin{aligned}
&(1)\frac{1}{E\sigma_E}\left[x_1\sigma_1^2+\sum_{j=2}^{N}x_j\sigma_{1j}\right]=\lambda(\mu_1-r);\\
&(2)\frac{1}{E\sigma_E}\left[x_2\sigma_2^2+\sum_{\substack{j=1\\ j\neq 2}}^{N}x_j\sigma_{2j}\right]=\lambda(\mu_2-r);\\
&\quad\vdots \qquad \vdots \qquad \vdots \quad \vdots\\
&(N)\frac{1}{E\sigma_E}\left[x_n\sigma_N^2+\sum_{j=1}^{N-1}x_j\sigma_{Nj}\right]=\lambda(\mu_N-r),\ \text{and}\\
&(N+1)\ E=\sum_{i=1}^{N}X_i\mu_i+\left(1-\sum_{i=1}^{N}X_i\right)r
\end{aligned}\right\}(N+1)\ \text{equations}$$

where E and σ_E are the portfolio's expected (one period) return and standard deviation, respectively. Multiplying the ith equation (i = 1,...,N) by x_i and summing for all i we obtain:

$$\frac{1}{E\sigma_E}\left[\sum_{i=1}^{N}x_i^2\sigma_i^2+\sum_{\substack{j=1\\ j\neq i}}^{N}\sum_{i=1}^{N}x_ix_j\sigma_{ij}\right]=\lambda\left[\sum_{i=1}^{N}x_i\mu_i-\sum_{i=1}^{N}x_ir\right]$$

which can be rewritten as:

$$\frac{\sigma_E}{E}=\lambda\left[\sum_{i=1}^{N}x_i\mu_i+\left(1-\sum_{i=1}^{N}x_i\right)r-r\right]=\lambda(E-r).$$

Therefore:

$$\frac{1}{\lambda}=\frac{E-r}{\sigma_E/E}=\frac{E-r}{C_E},$$

which is the price of a unit of risk. Note that the price of a unit of risk in the Sharpe-Lintner model is given by $(E-r)/\sigma_E$, whereas in the lognormal case, the price of a unit of risk is stated in terms of the portfolio's coefficient of variation C_E. Moreover, $(E-r)/\sigma_E$ is constant across all investors (the slope of the capital market line), whereas the price of a unit of risk in terms of the coefficient of variation $\frac{E-r}{\sigma_E/E}$ is not constant across investors and depends on the portfolio selected (E, σ_E) by each individual investor.

Hence, we call it the *subjective price of risk.* Rewriting the ith equation of (12.5) yields:

$$\mu_i = r + \frac{1}{\lambda}\cdot\frac{1}{E}\cdot\frac{1}{\sigma_E}\left[x_i\sigma_i^2 + \sum_{\substack{j=1\\ j\neq 1}}^{N} x_j\sigma_{ij}\right]$$

Substituting $\frac{E-r}{\sigma_E/E}$ for $1/\lambda$ (and recalling that the square bracket term is equal to Cov (R_i, R_E)) yields:

$$\mu_i = r + \frac{E-r}{\sigma_E/E}\cdot\frac{1}{E}\cdot\frac{1}{\sigma_E}\text{Cov}(R_i, R_E) \qquad (12.6)$$

where R_E is the return on the portfolio selected by the investor. However, R_E is a linear combination of the *market portfolio* R_m and the riskless asset r where R_m is the return on the market portfolio (see portfolio m in Figure 12.1). Thus, $R_E = \alpha R_m + (1-\alpha)r$ with mean $E = \alpha\mu_m + (1-\alpha)\, r$, standard deviation $\sigma_E = \alpha\mu_m$, and covariance Cov (R_i, R_E) = α Cov (R_i, R_m)..

Substituting these relationships in Eq. (12.6) yields:

$$\mu_i = r + \frac{\alpha\mu_m + (1-\alpha)r - r}{\alpha\sigma_m/E}\cdot\frac{1}{E}\cdot\frac{1}{\alpha\sigma_m}\alpha\text{COV}(R_i, R_m)$$

which reduces to:

$$\mu_i = r + \frac{\alpha(\mu_m - r)}{\alpha\sigma_m}\cdot\frac{\text{Cov}(R_i, R_m)}{\sigma_m}$$

which is the well-known CAPM relationship:

$$\mu_i = r + (\mu_m - r)\,\beta_i.$$

Thus, although the price of a unit of risk is subjective, the CAPM risk-return relation is intact.

Discussion:

Let us look at the economic intuition of this result. Figure 12.2 illustrates the one-period efficient sets in the E–σ and E–σ/E framework (note that E stands for the mean, hence the notations M-V and E–σ are used interchangeably). When the riskless asset is not allowed, only the segment **bc** of Figure 12.2a is efficient in the lognormal case (i.e., by the M-C criterion), even though the E–σ efficient frontier is larger and represented by the segment **ac**. The segment **bc** in the E–σ space is given by the corresponding segment $\mathbf{b}^1\mathbf{c}^1$ in the E–σ/E space in Figure 12.2b. Note that $\mathbf{b}^1$ represents the portfolio with the minimum *coefficient of variation*, whereas point **a** in Figure 12.2a denotes the portfolio with the minimum variance.

When the one-period riskless asset is available, the E-C single-period efficient set is **rr′** (see Figure 12.3a) with corresponding efficient segment **bb′** in terms of terminal wealth (Figure 12.3b). Let us elaborate. The segment **ad** is the E-V efficient frontier of *risky* assets whereas **bd** is the corresponding E-C frontier (see Figure 12.3a). However, when the riskless asset exists, the one-period E-V and E-C frontiers, given by the line **rr′**, coincide. Thus, with lognormal terminal wealth distribution any investor who maximizes the expected utility of terminal wealth should select his/her optimum investment in each single period from the efficient frontier **rr′**. To see this, choose any portfolio, say m_1. Then portfolio m_2 will dominate m_1 by the E-V rule as well as by the E-C rule (see Figure 12.3a). The proof that m_2 dominates m_1 by the E-C rule in the multiperiod case is straightforward: Suppose an investor selects portfolio m_1. By employing the M-C rule in each single period, we see that portfolio m_2 dominates portfolio m_1. Denote the one-period and multiperiod coefficient of variation, by C_1 and C, respectively. Hence, by shifting from m_1 to m_2, all other things being equal, the multiperiod mean increases and the multiperiod coefficient of variation (squared) $C^2 = (1 + C_1^2)^T - 1$ decreases, raising the expected utility of terminal wealth.

Thus, we have a single period two-fund Separation Theorem at each revision date: Regardless of the length of the holding period T, all investors who wish to minimize the coefficient of variation C for a given expected return E, will mix portfolio m with the riskless asset, in a similar way to that prescribed by the E-V framework, and, therefore, the CAPM result holds. Because the E-C is an optimal rule for the multiperiod lognormal distributions, we obtain the Separation Theorem in each single period which implies the CAPM.

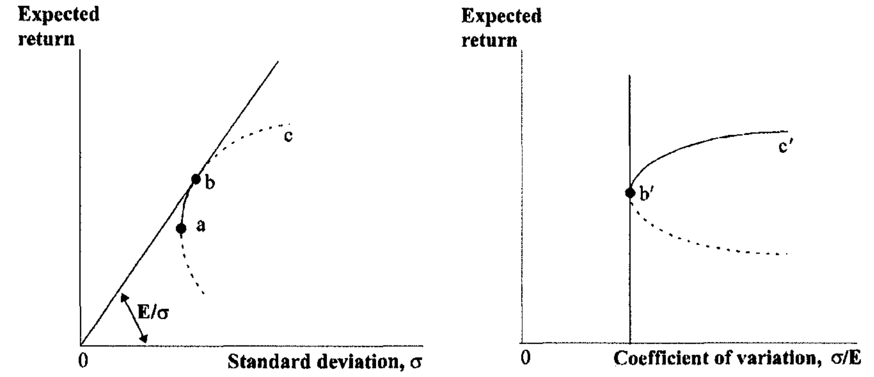
Note that although the return on the riskless asset is certain in every single period, the efficient frontier of the terminal wealth is not linear (Figure 12.3b). To see this, consider date T–1 (i.e., the beginning of the final period). The value of the portfolio is V_{T-1} and the investor invests x in portfolio m and (1 – x) in r. Hence, the terminal wealth of the portfolio at the end of the Tth period is given by V_T:

$$V_T = x\, V_{T-1}\,(1 + R_m) + (1 - x)\, V_{T-1}\,(1 + r)$$

where R_m and r are the rates of return on the risky portfolio and on the riskless asset, respectively. In spite of r being certain at time T – 1, no certain component is observed in this cash flow at any time prior to T – 1, because V_{T-1} is a random variable which depends on the random returns in the previous T – 1 periods. Hence, the multiperiod efficient frontier with a riskless asset is nonlinear. All efficient investment strategies that lie on the line **rr′** (see Figure 12.3a) lead to efficient diversification strategies given by the curve **bb′** in Figure 12.3b, where this figure is drawn in terms of terminal wealth rather than single-period parameters. Investors with preferences U_1 and U_2 have the same investment strategy in risky assets in each single period, but they differ with respect to their use of leverage: U_2 employs more leverage than U_1 and therefore, the mean return and risk are greater.

All of the portfolios located on segment **bb′** (see figure 12.3b) are composed of portfolio m and the riskless asset which, in each single period, provides a Separation Theorem and an equilibrium risk-return relationship such as that obtained in the CAPM.

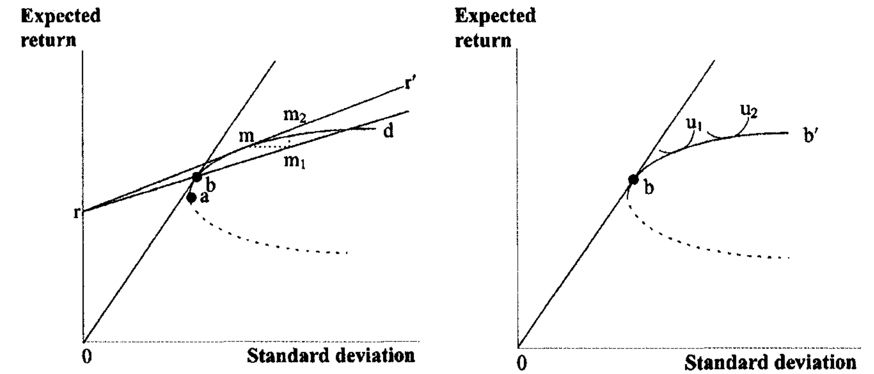
Fig. 12.2: The Lognormal Efficient Frontier without a Riskless Asset



12.2a: Mean - Standard deviation efficient frontier

12.2b: Mean - Coefficient of variation efficient frontier

Figure 12.3: The One-period and Multiplied Frontiers



12.3a: One-period parameters

12.3b: Terminal wealth parameters

(2) Nonstationary Distributions of Returns:

So far, in the lognormal case it was assumed that portfolio returns are stationary over time. Although, the stationarity assumption simplifies the analysis, it also qualifies the results (see Merton [1990][15], Rosenberg and Ohlson [1976][16]). We show below that all the results of the previous section are intact even under nonstationarity, as long as portfolio returns are independent over time. Moreover, by the nonstationarity assumption, the various parameters may change from one period to another, and, therefore, the uniform investment strategy is no longer optimal and the typical investor would be expected to change his/her investment proportions from one period to another. In other words, the asset parameters may change over time, hence the investment strategy will also change. However, the investment strategy does not change as a function of historical realized returns. In spite of the relaxation of the stationarity assumption, we show below that investors still select their portfolios by the mean-coefficient of variation rule in *every single period* which, in turn, implies the CAPM in exactly the same way proved above for the stationary case.

As shown below, the multiperiod variance σ^2 and mean μ are given by:

$$\sigma^2 = \prod_{t=1}^{T}(\sigma_t^2 + \mu_t^2) - \mu^2,$$

$$\mu = \prod_{t=1}^{T}\mu_t$$

The multiperiod coefficient of variation squared is given by $C^2 = (\sigma/\mu)^2$. Using these definitions and the relationship $\mu^2 = \mu_1^2 \cdot \mu_2^2 \cdots \mu_T^2$ yields:

$$C^2 = \prod_{t=1}^{T}\left[\frac{\sigma_t^2 + \mu_t^2}{\mu^2}\right] - \frac{\mu^2}{\mu^2} = \left(\frac{\sigma_1^2 + \mu_1^2}{\mu_1^2}\right)\left(\frac{\sigma_2^2 + \mu_2^2}{\mu_2^2}\right)\cdots\left(\frac{\sigma_T^2 + \mu_T^2}{\mu_T^2}\right) - 1.$$

Hence:

$$C^2 = (C_1^2 + 1)(C_2^2 + 1)\cdots(C_T^2 + 1) - 1 = \prod_{t=1}^{T}(C_t^2 + 1) - 1 \qquad (12.7)$$

where $C_t^2 = \sigma_t^2 / \mu_t^2$ (t = 1, 2,...,T) is the squared coefficient of variation corresponding to period t. Thus, we obtain:

[15] Merton, R.C., *Continuous-Time Finance*, Basil Blackwell, Cambridge, 1990.

[16] Rosenberg, B. and A.G. Ohlson, "The Stationary Distributions of Returns and Portfolio Separation in Capital Markets: A Fundamental Contradiction," *Journal of Quantitative Analysis*, 1976, pp. 393–401.

$$\mu = \prod_{t=1}^{T}\mu_t \text{ and } C^2 = \prod_{t=1}^{T}(1 + C_t^2) - 1 \tag{12.8}$$

where μ and C are the multiperiod portfolio mean and coefficient variation, respectively.

We see from eq. (12.8) that if in period t, for a given μ_t, C_t is minimized, then for a given multiperiod mean μ, the multi-period coefficient of variation C will also be minimized. Therefore, CAPM holds in each single period in the non-stationary case, too.

12.2 SUMMARY

In this Chapter we extend the validity of the single-period CAPM to four multi-period cases where investors are assumed to maximize $EU(W_T)$ where W_T is the terminal wealth. In several of them, stochastic dominance arguments are employed to prove the CAPM (which is based on the M-V paradigm). The four cases are:

(a) Quadratic preferences defined on terminal wealth where (T – 1) revisions are allowed: The CAPM does not hold when revisions are not allowed even when preferences are quadratic.

(b) Normal one-period return distributions with no constraints on the multiperiod distributions: Here, too, a diverse holding period is allowed. This case is an extension of the classic Sharpe-Lintner model, where the single holding period assumption is relaxed and the terminal wealth, W_T, is not normally distributed.

(c) The terminal wealth is assumed to be normal and (T – 1) revisions are allowed. However, the normality assumption is possible only when all investors have the same holding period ($T = T_0$). Thus, in this case diverse holding period is not allowed since the normality property is violated. This is similar to the single period Sharpe-Lintner model with the distinction that (T – 1) revisions are allowed before the portfolio is liquidated.

(d) The terminal wealth, W_T, is lognormally distributed with a diverse holding period. This case is similar to Merton's continuous CAPM model. The distinction, however, is that here we allow a *finite* number of revisions, and transaction costs can therefore be incorporated.

In all four cases, the CAPM risk-return relation holds for both stationary and nonstationary distributions of returns over time. Moreover, the one-period riskless interest rate may also vary from one period to another without affecting the results.

Key Terms

Continuous time Model

Distribution-free

Heterogeneous Investment Horizons

Multiperiod District Models

Revision Period

Two-Fund Separation Theorem

Lognormal Distribution

Stationary Distribution

Coefficient of Variation

Subjective Price of Risk

13

ALMOST STOCHASTIC DOMINANCE (ASD)

The SD criteria as well as the M-V rule may lead to paradoxes in decision making. This chapter is devoted to such paradoxes and to a suggestion how to modify the SD rules and the M-V rule to avoid such paradoxes.

13.1 THE POSSIBLE PARADOXES

In the previous chapters we define the set $\mathbf{U}_1$ as the set of *all* non-decreasing utility functions and the set $\mathbf{U}_2$ as the set of *all* non-decreasing concave utility functions. The sets $\mathbf{U}_1$ and $\mathbf{U}_2$ that we deal with in this book contain all preferences including *extreme* preference which in practice do not conform with observed decision makers' behavior (the same argument can be extended also to U_3, U_4 , etc.). For example, the preference U_0 given by,

$$U_0(x) = \begin{cases} x & x \le x_0 \\ x_0 & x > x_0 \end{cases} \tag{13.1}$$

is included in $\mathbf{U}_1$ as $U_0' \ge 0$; however, it may lead to paradoxical results. To illustrate the possible paradoxical results induced by the employment of U_0, suppose that the investor faces the following two options:

Option F		Option G	
Outcome	Probability	Outcome	Probability
\$1	1/10	\$2	1/10
$\$10^6$	9/10	\$3	9/10

It is easy to verify that the cumulative distributions F and G cross hence there is no FSD. Namely, both F and G are included in the FSD efficient set. What would you choose? There is no doubt that most, if not all, investors would choose option F. Yet, for a preference $U_0 \in \mathbf{U}_1$ given by (13.1), with $x_0 = 2$ prospect G has a higher expected utility.

To see this claim assume, as before, that $a \le x \le b$. The difference in expected utility of G minus F is given by,

$$E_G\, U_0(x) - E_F U_0(x) = \int_a^b [F(x) - G(x)]\, U_0'(x)\, dx =$$

$$= \int_a^2 [F(x) - G(x)]\, U_0'(x)\, dx + \int_2^b [F(x) - G(x)]\, U_0'(x)\, dx = \int_a^2 [F(x) - G(x)]\, dx > 0$$

This result is induced from the fact that $U_0'(x)=0$ for $x > 2$ and $U_0'(x)=1$ for $x \leq 2$. As for $x < 2$, F is above G, we obtain that $E_G U_0(x) > E_F U_0(x)$, in contrast to the fact that virtually all investors would choose F. Thus, there is at least one utility function, U_0, for which G is preferred over F and this utility function belongs to $\mathbf{U}_1$. This paradoxical result holds also if we replace the outcome $\$10^6$ in F by $\$10^n$ where $n \to \infty$, or if we replace the outcome of \$2 in G by a number close to \$1, say \$1.1. What is the source of this paradoxical result? Why in a case where there is no doubt that all investors would choose F, the FSD criterion cannot rank F and G and asserts that both should be included in the FSD efficient set? The explanation is that the set $\mathbf{U}_1$ contains *all* preferences as long as $U' \geq 0$ (and $U' > 0$ in some range). This set may include preferences that do not conform with any investor's behavior. Yet, mathematically, these preferences are included in $\mathbf{U}_1$. For example, U_0 given above (see 13.1) with $x_0 = 2$ reveals that the investor is indifferent whether she gets \$2 or \$2 million as $U_0'(x)=0$ for $x \geq x_0 = 2$. Thus, formally (or mathematically), $U_0 \in \mathbf{U}_1$ but it is reasonable to assume that there is no investor who will choose G because there is no investor with preference like U_0. Namely, the reason for this paradoxical result is that the mathematical set $\mathbf{U}_1$ includes many utility functions $U \in \mathbf{U}_1$ that do not characterize any of the decision makers. Therefore, FSD may lead to such paradoxes.

How can such paradoxes be avoided? To this issue we devote this chapter. To be more specific, we will develop SD dominance rules, which clearly reveal a dominance of F over G for *almost* all preferences. To achieve a dominance rule with no paradoxes we define another set of utility function $\mathbf{U}_1^*$ where $\mathbf{U}_1^* \subset \mathbf{U}_1$, such that preference like U_0 given above (see (13.1)) are excluded. We may call preferences like U_0 as *extreme*, *pathological* or simply *unrealistic* as they, generally, do not conform with human decision making. Thus, we defined a set of preference $\mathbf{U}_1^*$, which does not include extreme preferences. To be more specific, $\mathbf{U}_1^*$ is defined as the set of all monotonic non-decreasing preferences which exist *in practice* (i.e., which conform with investors' choices) and $\mathbf{U}_1$ is the theoretical set of preferences and $\mathbf{U}_1^* \subset \mathbf{U}_1$.

So far we have shown a possible paradox in $\mathbf{U}_1$. Let us see that a similar paradox may exist also in $\mathbf{U}_2$, i.e., with SSD. Suppose you have to choose between F and G given as follows:

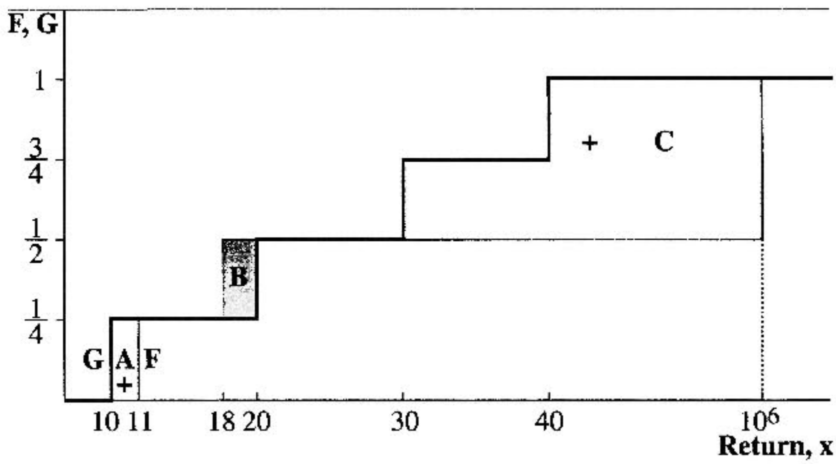
F		G	
Outcome	Probability	Outcome	Probability
11	¼	10	¼
18	¼	20	¼
10^6	½	30	¼
		40	¼

The two cumulative distributions corresponding to F and G are illustrated in Figure 13.1. There is no FSD as F and G cross. There is also no SSD of F over G as $|B| > |A|$. Of course there is also no SSD of G over F as F has a higher mean. Thus, mathematically, both F and G are included in the SSD efficient set because of the area $|B|$ is larger than the area $|A|$. Therefore, no matter how large area C is (see Figure 13.1) there is no SSD dominance. As there is no SSD dominance it means that there is some $U_0 \in U_2$, for which G is preferred over F. For example, for U_0 given by,

$$U_0 : \begin{cases} x & x \leq 20 \\ 20 & x > 20 \end{cases}$$

G is preferred over F, hence F does not dominate G by SSD. Yet, we suspect that in practice 100% of the investors will choose F. Thus, in analogy to FSD here we need to define $\mathbf{U}_2^* \subset \mathbf{U}_2$ such that functions like U_0 given above will not be included in $\mathbf{U}_2^*$, hence the paradox is avoided.

Figure 13.1: No FSD and no SSD



Actually, this possible paradoxical result of SD exists also with the well-known and widely used Mean-Variance (M-V) criterion of Markowitz. To see this, suppose that one faces the following two prospects F and G with the following parameters:

$E_F(x) = \$10^6$ $\qquad$ $E_G(x) = \$10$

$\sigma_F(x) = \$10$ $\qquad$ $\sigma_G(x) = \$9$

By the M-V rule the two prospects cannot be ranked and both should be in the M-V efficient set. What would the investor choose? We doubt whether in a survey of investors we will not obtain that 100% of the choices will be F, hence the paradox also in the M-V framework.

What is the remedy to the paradoxical results of the M-V rule in this case? There are two possibilities here:

i) If normal distributions are assumed, then SSD ~ MV, and we are back in the SD framework. Hence, like in SSD we need to define $\mathbf{U}_2^* \subset \mathbf{U}_2$ which eliminates the paradox and clearly reveals a preference for F.

ii) The M-V rule can be justified also by assuming a quadratic utility function $U_q \subset \mathbf{U}_q$, where $\mathbf{U}_q$ is the set of all quadratic preferences. Note that the quadratic function is given by $U(x) = x - bx^2$ where $b > 0$. Define $\mathbf{U}_q^* \subset \mathbf{U}_q$, where some constraint on the parameter b is imposed to avoid paradoxes, like the one given above. Hence, in a similar way to FSD and SSD one should define $\mathbf{U}_q^* \subset \mathbf{U}_q$ such that in $\mathbf{U}_q^*$ the dominance of F over G in the above example is revealed.

It is interesting that Baumol, as early as in1963, revealed the possible M-V paradoxical results and therefore suggested another rule called "Expected Gain – Confidence Limit Criterion"[1] as a substitute to the M-V rule. Indeed, in some cases the paradoxes are avoided. However, Baumol's rule is a more intuitive rule rather than a mathematical rule which can be justified in expected utility paradigm.

The above paradoxes of FSD, SSD and M-V criteria stem from the fact that the sets, $\mathbf{U}_1$, $\mathbf{U}_2$ and $\mathbf{U}_q$ contain utility functions which are mathematically valid but do not conform in practice with the preferences of virtually all investors. Therefore, to avoid such paradoxes we need to define new decision rules, which are denoted by FSD*, SSD* and M-V* corresponding to the set, $\mathbf{U}_1^* \subset \mathbf{U}_1$, $\mathbf{U}_2^* \subset \mathbf{U}_2$ respectively, where the paradoxical results are avoided. These rules are also called *Almost* FSD (AFSD), *Almost* SSD (ASSD) and *Almost* M-V (AM–V). We call the new criteria Almost Stochastic Dominance (ASD) rules because they relate to "almost" all utility functions in the relevant set. However, if in practice a given ASD criterion is appropriate for all investors, the set $\mathbf{U}_i^*$ $(i = 1,2,3\ldots)$ is indeed the relevant set of

---

[1]See Chapter 1.

preferences and the larger set $\mathbf{U}_i$ (i = 1,2,3...) is simply irrelevant for decision making in practice.

In what follows we first define ASD and mathematically prove the ASD rules, and then we present an experimental study which sheds light on the relationship between $\mathbf{U}_i^*$ and $\mathbf{U}_i$ in practice. We focus in the rest of this chapter on FSD* and SSD* and the same extensions can be conducted with other decision rules.

13.2 FSD* CRITERION CORRESPONDING TO $\mathbf{U}_i^*$.

We provide in this section the intuitive as well as the formal definition of AFSD which we also denote by FSD*. Then we analyze the relationship between FSD* and FSD as well as the relationship between $\mathbf{U}_i^*$ and $\mathbf{U}_1$. We provide also a few examples, which induce paradoxes with FSD but avoided with FSD*. The next section is devoted to SSD*

Suppose that we have two cumulative distributions F and G. Option F has a higher mean return, hence G cannot dominate F by FSD. However, also F does not dominate G as the distributions cross. Figure 13.2 demonstrates the two hypothetical distributions. As we can see, due to the area denoted by B, F does not dominate G by FSD. Moreover, no matter how small the area B is relative to the total positive areas, A+C, still one can always find $U_0 \in \mathbf{U}_1$ such that $F_F U_0(x) < E_G U_0(x)$, which is the reason why F does not dominate G by FSD. The expected utility difference between F and G denoted by Δ can be written as:

$$\Delta \equiv E_F U(x) - E_G U(x) = \int_a^b [G(x) - F(x)] U'(x) dx \qquad (13.2)$$

(see eq. (3.1) in Chapter 3)

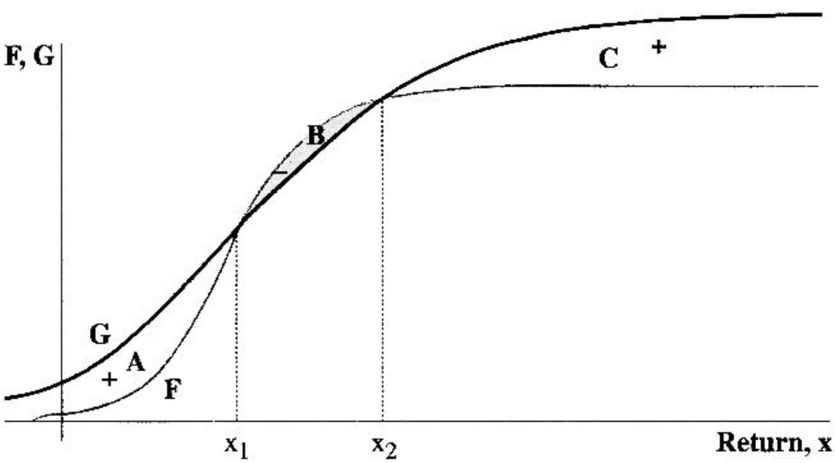
Choose a utility function of the form

$$U_0(x) = \begin{cases} x_1 & x < x_1 \\ x & x_1 \le x \le x_2 \\ x_2 & x > x_2 \end{cases} \qquad (13.3)$$

For this utility function the integral (13.2) in the ranges $x < x_1$ and $x > x_2$ is equal to zero (as $U'(x) = 0$ in the two ranges), and for $x_1 \le x \le x_2$ we have $U_0' = 1$ hence

$\Delta \equiv \int_{x_1}^{x_2}[G(x)-F(x)]\,dx<0$ implying that $E_F U_0(x) < E_G U_0(x)$. Hence for this specific $U_0 \in \mathbf{U}_1$, G is preferred over F, which is the reason why FSD does not exist, even with a relatively little FSD area violation as area B. The utility function U_0 generally does not conform with the observed human decision choices, as it assigns zero additional utility to outcomes in the range $x < x_1$ and in particular it assigns zero additional utility to outcomes $x > x_2$. Thus, even if the area C is indefinitely large, F does not dominate G by FSD, hence the paradox. Let us turn to the definition of $\mathbf{U}_1^*$ and FSD*, which avoids such a paradox.

Figure 13.2: No SSD



(for the values x_1 and x_2, see Figure 13.2)

Suppose that we have two distributions F and G which cross. Let us divide the difference in expected utility integral into two sets. The first set, s_1, is defined over ranges where F > G. The range for which F < G is denoted by its complement $\bar{s}_1$. Formally, s_1 is defined as follows:

$s_1(F,G) = \{t \in [a,b] : G(t) < F(t)\}$, which corresponds to the range (x_1, x_2) in Figure 13.2, and $\bar{s}_1$ is is the complement of s_1.

The difference in expected utility of the two prospects F and G is therefore given by,

$$\Delta \equiv \int_{s_1}[G(x)-F(x)]U'(x)\,dx + \int_{\bar{s}_1}[G(x)-F(x)]U'(x)\,dx \qquad (13.4)$$

The first part (the integral over s_1), is by definition negative, which is the reason for the *no* FSD dominance of F over G, which occurs even if the range s_1 is relatively very small. Using these definitions of s_1 and $\bar{s}_1$ let us turn to the derivation of FSD*.

In deriving FSD* we employ the following technique. We decrease the integral over s_1 to a minimum (i.e., increase it in absolute value) and decrease the integral over s_2, which is positive, *also* to a minimum. Thus, we define $\Delta^* \equiv$ (to distinguish from Δ) as follows:

$$\Delta^* \equiv \text{Sup}[U'(x)] \int_{s_1} [G(x) - F(x)]\,dx + \text{Inf}[U'(x)] \int_{\bar{s}} G(x) - F(x)\,]\,dx \quad (13.5)$$

If for a given utility function $\Delta^*>0$, then for this utility function also $\Delta >0$. The reason is that with Δ the first term as well as the second term of eq. (13.4) are larger relative to the corresponding terms of eq. (13.5).

Thus, for any specific utility function for which $\Delta^* > 0$ also $\Delta > 0$ and for this utility function we assert that F is preferred over G. As we shall see below, we are looking for the set of preferences $\mathbf{U}_1^*$ for which $\Delta^* > 0$, hence $\Delta >0$, and for this set we assert that F dominates G by FSD*, despite the fact that F does not dominate G by FSD. In other words, we are looking for the restriction needed on preferences such that $\Delta^* > 0$. With a little manipulation of eq. (13.5) it is easy to show that $\Delta^* > 0$ if the following holds,

$$\text{Sup}\,(U'(x)) \le \text{Inf}(U'(x)) \frac{-\int_{\bar{S}_1} [G(x) - F(x)\,]\,dx}{\int_{S_1} [G(x) - F(x)\,]\,dx} = \text{Inf}(U'(x)) \frac{\int_{\bar{S}_1} [G(x) - F(x)\,]\,dx}{\int_{S_1} [F(x) - G(x)\,]\,dx} \quad (13.6)$$

Note that the integral over s_1 is negative, hence the inequality sign is reversed and is a division of this term. In addition, in deriving (13.6) we used the relationship,

$$\int_{s_1} [G(x) - F(x)]\,dx = -\int_{s_1} [F(x) - G(x)]\,dx$$

As by definition $U'(x) \le \text{Sup}\,U'(x)$, eq. (13.6) can be rewritten as

$$U'(x) \le \text{Inf}\,U'(x) \frac{\int_{\bar{s}_1} [G(x) - F(x)\,]\,dx}{\int_{s_1} [F(x) - G(x)\,]\,dx} \quad (13.7)$$

Define by ε_1 the ratio of the absolute value of the area over s_1 (which induces the no FSD situation) divided by the total absolute area $\int_a^b |G(x)-F(x)|dx$ enclosed between F and G, to obtain

$$(\frac{1}{\varepsilon}-1)=\frac{\int_{S_1}[F(x)-G(x)]dx+\int_{\bar{S}_1}[F(x)-G(x)]dx}{\int_{S_1}[F(x)-G(x)]dx}-1=\frac{\int_{\bar{S}_1}[G(x)-F(x)]dx}{\int_{S_1}[F(x)-G(x)]dx} \quad (13.8)$$

Hence, eq. (13.7) can be finally rewritten as

$$U'(x)\leq \mathrm{Inf}[U'(x)]\left[\frac{1}{\varepsilon}-1\right] \quad (13.9)$$

Thus, what we have shown here is that if (13.9) holds, for this specific utility function $\Delta^* > 0$, hence also $\Delta > 0$, and F is preferred over G.

Note that if $\varepsilon = 0$, eq. (13.9) always holds. In this case there is no area violation and there is FSD of F over G. Thus, there are no paradoxes and no need to define $\mathbf{U}_1^*$. Alternatively, for $\varepsilon = 0$, $\mathbf{U}_1^* = \mathbf{U}_1$, we have no area violation and therefore FSD* and FSD coincide.

Proposition 13.1 (AFSD or FSD):*[2]

Define by $U_1^*(\varepsilon)=\{U\in \mathbf{U}_1 : U'(x)\leq \mathrm{Inf}\, U'(x)\left[\frac{1}{\varepsilon}-1\right]$ (13.10)

for all x, where $0 < \varepsilon < 0.5$. Then F dominates G by FSD* (almost FSD) if and only if,

$$\int_{S_1}[F(x)-G(x)]dx\leq \varepsilon \int_a^b |F(x)-G(x)|\, dx \quad (13.10')$$

The sufficiency proof follows immediately from the above discussion because if (13.10) holds, $\Delta^*> 0$ and if $\Delta^*> 0$ for a given utility function, it implies that for this utility function also $\Delta > 0$. Hence, if (13.10') holds, then for all preferences $U\in \mathbf{U}_1^*$, $\Delta^*>0$, and also $\Delta >0$ for all $U\in \mathbf{U}_1^*$ and F dominates G by FSD* which completes the sufficiency.

[2]For more details on the proofs and analyses of ASD see Leshno, M. and H. Levy, "Preferred by "All" and Preferred by "Most" Decision Makers: Almost Stochastic Dominance," *Management Science*, 2002, *48*, pp. 1074–1085.

To prove necessity, we have to show that if (13.10') does not hold then there is $U_0 \in \mathbf{U}_1^*$ for which $\Delta < 0$. Such an example of U_0 can be found in Levy and Leshno (see footnote 2).

Before we turn to the example given before, note that it is required that $\varepsilon < 0.5$. Namely, for two distributions we define the one with the less than 50% area violation as a candidate for FSD dominance. In terms of Figure 13.2 it implies that $\varepsilon = \dfrac{|B|}{|A|+|B|+|C|} < 0.5$. In this specific case G cannot dominate F by FSD* as it requires $\varepsilon > 0.50$ correction in area (i.e., $(|A|+|C|)\,/\,(|A|+|B|+|C|) > 0.5$). As by definition (see Chapter 3) we have

$E_F(x) - E_G(x) = A - B + C$ and $A + C > |B|$, the condition $\varepsilon < .05$

implies that a necessary condition for FSD* of F over G is that $E_F(x) > E_G(x)$.

Let us turn back to the paradoxical examples given above. We have shown above that with F = {\$1, 1/10, \$$10^6$, 9/10}, G = {\$2, 1/10, \$3, 9/10} (see example in section 13.1) there is no FSD, a paradoxical result. For example, for the function $U_0 \in \mathbf{U}_1$ which is equal to x for $x \leq \$2$, and 2 for $x \geq 2$ (see 13.1), G is preferred over F, hence there is no FSD. However, with this function $\text{Inf}[U'(x)] = 0$, $\text{Sup}[U'(x)] = 1$, and as $(\frac{1}{\varepsilon} - 1) > 0 (\text{as } \varepsilon < 0.5)$, eq. (13.9) can never be fulfilled, and therefore this specific utility function does not belong to $\mathbf{U}_1^*$. Thus, there is a paradox in $\mathbf{U}_1$ but not in $\mathbf{U}_1^*$. By a similar argument, we exclude from $\mathbf{U}_1^*$ all preferences which do not conform with decision making in practice, hence paradoxes are avoided. Indeed, with FSD*, F dominates G for all $\mathbf{U}_1^*$ which allow a relatively small FSD area violation. This conforms with the intuition that virtually all investors would prefer F over G, given by the above example. In the above example, F dominates G by FSD* i.e., for all $U \in \mathbf{U}_1^*$, but not by FSD, namely not for all $U \in \mathbf{U}_1$.

Note that FSD implies FSD* but not the other way around. To see that, recall that if F dominates G by FSD, then $\varepsilon = 0$ and eq. (13.9) holds, hence there is FSD* of F over G. The reverse argument is, of course, not valid as, if $\varepsilon > 0$ we may have FSD* but we do not have FSD. Finally, it is worth mentioning that as ε decreases eq. (13.9) holds for more utility functions, and therefore $\mathbf{U}_1^*(\varepsilon)$ increases as ε decreases. In the limit as $\varepsilon \to 0$, $\mathbf{U}_1^* = \mathbf{U}_1$ and FSD* and FSD coincide. However, the main purpose of AFSD (or FSD*) is to allow a little FSD violation ($\varepsilon > 0$), hence $\mathbf{U}_1^* \subset \mathbf{U}_1$ and the preferences which generally create paradoxes are excluded from $\mathbf{U}_1$. Of course, what is considered as a paradox and what is a reasonable choice is an empirical or experimental question, which will be addressed later on in this chapter.

To sum up this section we have FSD $\Rightarrow$ FSD*, a necessary condition for dominance of F over G by FSD* is that $E_F(x) \geq E_G(x)$, and that for $\varepsilon = 0$ $U_1^* = U_1$ and FSD and FSD* coincide.

13.3 THE SSD* CRITERION CORRESPONDING TO U_2^*

Let us look first at Figure 13.3. As we can see there is no FSD as the two cumulative distributions cross. Moreover, neither F nor G dominates the other also by SSD. G does not dominate F by SSD because $E_F(x) > E_G(x)$. However, also F does not dominate G by SSD because of the area denoted by B (see Figure 13.3). And no matter how small the area B is relative to the areas A and C, there is no SSD as there is some $U_0 \in U_2$ for which $E_G U_0(x) > E_F U_0(x)$. Take for example the following utility function $U_0 \in U_2$:

$$U_0 = \begin{cases} x & x \leq x_2 \\ x_2 & x > x_2 \end{cases} \quad \text{(for the value } x_2\text{, see Figure 13.3)}$$

For this function we have,

$$E_F U_0(x) - E_G U_0(x) = \int_a^b [G(x) - F(x)] U'(x) dx = \int_a^{x_2} [G(x) - F(x)] dx < 0 \text{ (the}$$

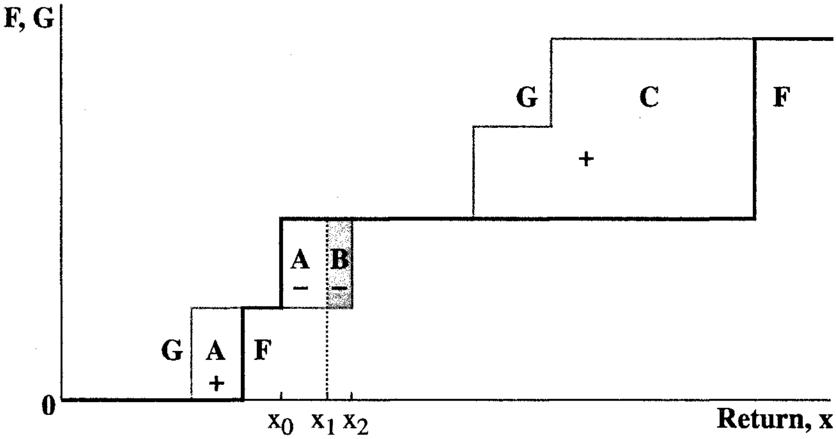
integral over the range $x > x_2$ is equal to zero as $U'(x) = 0$ for $x > x_2$). Thus, for this specific preference, U_0, G is preferred over F. And this assertion is valid regardless of the size of area C which may be very large. Yet, if the area C is very large relative to area B, in practice most risk averse (if not all) investors would choose prospect F which, in turn, constitutes a paradox with SSD. Thus, we need to define $U_2^* \subset U_2$ such that for very large area C for all $U \in U_2^*$, F will dominate G. Such a dominance is called almost SSD and denoted by ASSD or SSD*. Let us turn now to the definition of SSD*.

Suppose that in the above example

$$|B| \,/\, \int_a^b |G(x) - F(x)| \, dx < \varepsilon$$

Namely, the area $|B|$ which induces the no SSD of F over G, divided by the total absolute area enclosed between F and G is smaller than ε (see Figure 13.3).

Figure 13.3: No FSD: No SSD



Formally, the range $s_2(F, G)$ is defined as s_2

$$s_2(F,G) = \left\{ t \in s_1(F,G) : \int_a^t [G(x) - F(x)]dx < 0 \right\}$$

where F and G are the two cumulative distributions and s_1 is the range over which FSD is violated. In terms of Figure (13.3), s_1 is the defined as the range (x_0, x_2) and s_2 is defined as the range (x_1, x_2). Recall that we deal here only with concave functions, i.e., $U \in \mathbf{U}_2$ when $U' \geq 0$ and $U'' < 0$. As before, the difference in expected utility is given by

$$\Delta \equiv E_F U(x) - E_G U(x) = \int_{s_2} [G(x) - F(x)] U'(x)\, dx + \int_{\bar{s}_2} [G(x) - F(x) U'(x)\, dx$$

where s_2 is the range corresponding to the area that if omitted we would have SSD, e.g., range (x_1, x_2) corresponding to area B in Figure 13.3, and $\bar{s}_2$ is the complement of s_2.

Define Δ^*as follows:

$$\Delta^* = \text{Sup}[U'(x)] \int_{s_2} [G(x) - F(x)]dx + \text{Inf}\ [U'(x)] \int_{\bar{s}_2} [G(x) - F(x)]dx \quad (13.11)$$

We claim that if for a specific *concave* utility function $\Delta^* > 0$, it implies that for this *concave* utility function also $\Delta > 0$. The proof is as follows:

$\int_{s_2} [G(x) - F(x)]dx < 0$, hence by substituting $[\ U'(x)$ by

Sup $[U'(x)]$, the contribution of this area to the expected utility difference Δ decreases. In other words this term becomes even more negative. The integral over $\bar{s}_2$ is positive. Though over this range we may have $F < G$ as well as $G < F$, for any negative area when $F > G$ there must be a preceding positive area which is larger (or equal) than the negative area. This assertion stems from the fact that over $\bar{s}_2$, SSD is not violated. In terms of Figure 13.3, the $-A$ area is proceeded by the $+A$ range, see range ox_1. As U' is declining, the positive area contributes to expected utility difference, Δ, over $\bar{s}_2$ is more than the reduction in Δ due to the negative area. By substituting $U'(x)$ by Inf $[U'(x)]$ which is constant for both areas and which is smaller (or equal) than all $U'(x)$ in this range, we diminish the integral over $\bar{s}_2$. Thus, if $\Delta^* > 0$, it follows that $\Delta > 0$.

Let us find the conditions under which $\Delta^* > 0$ and to the definition of $\mathbf{U}_2^*$. From (13.11) we have that $\Delta^* > 0$ if the following holds,

$$\operatorname{Sup}[U'(x)] \le \operatorname{Inf}\,[U'(x)]\;\frac{\int_{\bar{s}_2}[G(x)-F(x)\,dx}{\int_{s_2}[F(x)-G(x)\,dx}$$

(Note that for $\Delta^*>0$ we shift on term of (13.11) to the right, divided by a negative number and reverse the roll of F and G in the denominator as done with FSD* proof).

But as Sup $[U'(x)] \geq U'(x)$ we finally can assert that if

$$U'(x) \le \operatorname{Inf}[U'(x)]\frac{\int_{\bar{s}_2}[G(x)-F(x)]dx}{\int_{s_2}[F(x)-G(x)\,dx} \qquad (13.12)$$

(13.12) holds, $\Delta^* \geq 0$, which implies that $\Delta > 0$. After some algebraic manipulation we obtain that the condition for preference of F over G for a given concave utility function is:

$$U'(x) \le \operatorname{Inf}[U'(x)]\left[\frac{1}{\varepsilon}-1\right] \qquad (13.13)$$

Where

$$\varepsilon \equiv \frac{\int_{S_2}\left|G(x)-F(x)\right|dx}{\int_a^b\left|G(x)-F(x)\right|dx} \qquad (13.14)$$

To illustrate in terms of Figure 13.3 we have, $\varepsilon=|B|/(|A|+|A|+|B|+|C|)$
We summarize these results with the following proposition:

Proposition 13.2: (ASSD or SSD)*

Define by $\mathbf{U}_2^*$ the set of all concave utility functions for which eq. (13.13) holds. Namely

$$\mathbf{U}_2^*(\varepsilon) = \{U\in \mathbf{U}_2 : U'(x)\le \text{Inf}\ [U'(x)[\frac{1}{\varepsilon}-1] \quad (13.15)$$

and ε is defined by eq. (13.14).

Then F dominates G by SSD* (or ASSD) if and only if for any $0 < \varepsilon < 0.5$ (as before, we need $\varepsilon < 0.5$ because we need to conduct less than 50% of area correction to avoid the case that FDG and GDF by SSD*) we have,

$$\int_{S_2}|G(x)-F(x)|\,dx \le \varepsilon \int_a^b |G(x)-F(x)|\,dx \quad (13.16)$$

where s_2 is the range of outcomes which, if eliminated, we would have SSD of F over G.

First note that like the FSD* case, here also a necessary condition for SSD* is that $E_F(x) \geq E_G(x)$. If the condition does not hold, it is possible to create a situation when F dominates G by SSD* and G dominates F by SSD* with the same ε. However, the condition $\varepsilon < 0.5$ guarantees that $E_F(x) \geq E_G(x)$. Let us turn now to the proof of proposition 13.2.

The sufficiency follows immediately from the above discussion: if (13.16) holds, (13.3) holds and Δ^* given by eq. (13.11) is positive, and as for all $U\in \mathbf{U}_2$, we have $\Delta \geq \Delta^*$ also $\Delta \geq 0$. The necessity is similar to the necessity in proposition 13.1 corresponding to FSD*.

A few remarks regarding SSD* are called for:

1) By construction for any two distributions $\varepsilon_{FSD} \geq \varepsilon_{SSD}$ therefore, if F dominates G by FSD* $\Rightarrow$F dominates G by SSD*.

2) We need to add $0 < \varepsilon < 0.5$ to avoid a case where FDG and GDF by SSD*.

3) As $\varepsilon_{FSD} > \varepsilon_{SSD} \Rightarrow \mathbf{U}_1^* \supset \mathbf{U}_2^*$

4) SSD $\Rightarrow$ SSD* because with SSD we have $\varepsilon = 0$ which implies that (13.15) holds.

Finally, let us turn back to the SSD paradox, which has been discussed before and which is illustrated with Figure 13.3. With the function U_0 given above which reveals a paradox we have Sup $[U'(x)] = 1$, Inf $[U'(x)] = 0$, hence for any $\varepsilon > 0$, eq. (13.15) cannot hold. Therefore, the utility U_0 which induces the paradox, indeed is eliminated as it does not belong to $\mathbf{U}_2^*$.

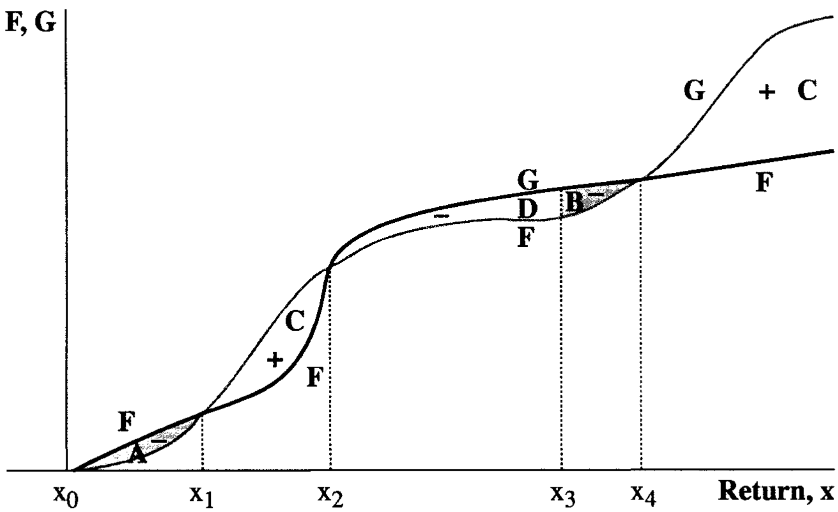
Example

Figure 13.4 demonstrates a case where $E_F(x) > E_G(x)$ but due to two areas marked by A and B there is no SSD of F over G. Yet, if area C is large enough almost all risk investors would choose F. Note that with our notation for FSD*, we have s_1 = {range (x_0, x_1) and range (x_2, x_4)} while for SSD* we have s_2 = {range (x_0, x_1) and range (x_3, x_4)}. Let us write Δ corresponding to SSD* for this specific example.

$$\Delta \equiv E_F U(x) - E_G U(x) = \int_{x_0}^{x_1} [G(x)-F(x)]U'(x)dx + \int_{x_3}^{x_4} [G(x)-F(x)]U'(x)dx$$

$$+ \int_{x_1}^{x_3} [G(x)-F(x)]U'(x)dx + \int_{x_4}^{b} [G(x)-F(x)]U'(x)dx$$

Figure 13.4: No FSD: No SSD



As the first two terms are negative (see Figure 13.4) we made these two terms even smaller by substituting $U'(x)$ by Sup $[U'(x)]$. The last two terms are positive. By substituting $U'(x)$ by Inf $[U'(x)]$ we decrease these two terms. While this claim is obvious regarding the last term, regarding the third term the argument is a little more delicate: the plus area denoted by C is larger than the negative area (in absolute terms) denoted by D (note that A + C + D = 0, hence C > $|D|$). As U' is declining, the contribution of the plus area to the Δ is larger than the reduction to Δ induced by area D. Therefore, by substituting $U'(x)$ in this range by Inf $[U'(x)] < U'(x)$ we reduce this positive term. Therefore, if $\Delta^* > 0$ for a given U also Δ corresponding to this utility function must be positive. Note that if U is not concave this argument is wrong. The reason for this assertion is that it is possible is that over area D, U' is larger than over area C, hence the third term is negative. By substituting $U'(x)$ by a constant number, Inf $[U'(x)]$, we change the third term from being negative to a positive term, hence we cannot claim anymore that $\Delta^* > 0 \Rightarrow \Delta > 0$. This example illustrates the need to employ in the proof of proposition 13.2 that $U' > 0$ and $U'' < 0$.

13.4 APPLICATION OF FSD* TO INVESTMENT CHOICES: STOCKS VERSUS BONDS

Suppose that the one-period (annual) rates of return on stocks and bonds are as follows:

Rate of Return: z	5%	7%	9%	12%
Pr(X = z) (Stocks)	0.1	0	0	0.9
Pr(Y = z) (Bonds)	0	0.4	0.6	0

It is easy to show that there is no FSD and no SSD, see the cumulative distributions given by Figure 13.5a. Yet, assuming that returns are identical and independent over time (i.i.d.), as the number of investment periods increases we obtain two cumulative distributions like those given in Figure 13.5b. We have a small area, ε, which induces no FSD and no SSD dominance, see range $x < x_0$. The larger n the smaller ε and we may have FSD*, let alone SSD*, of stocks over bonds. (Note that F_n and G_n in Figure 13.5b crosses left to x_0, but it is impossible to see it in this Figure).

Figure 13.5a: F and G for stocks and bonds
n = 1 period

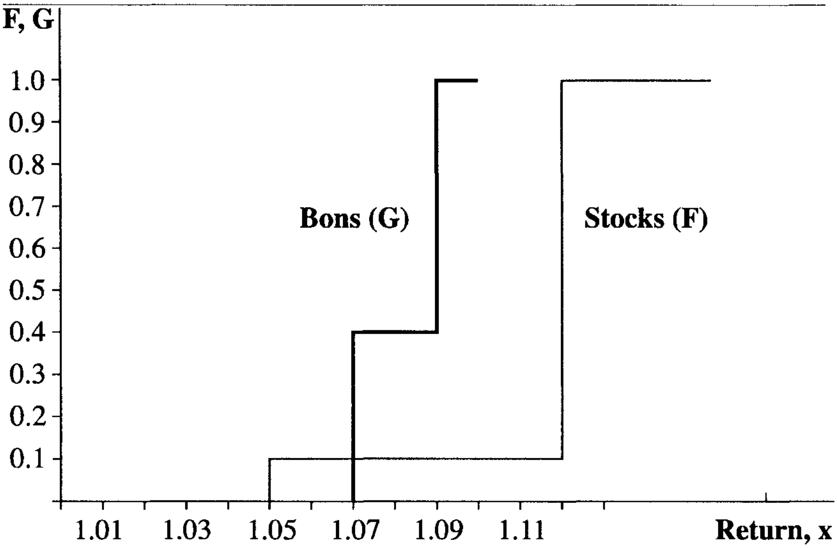
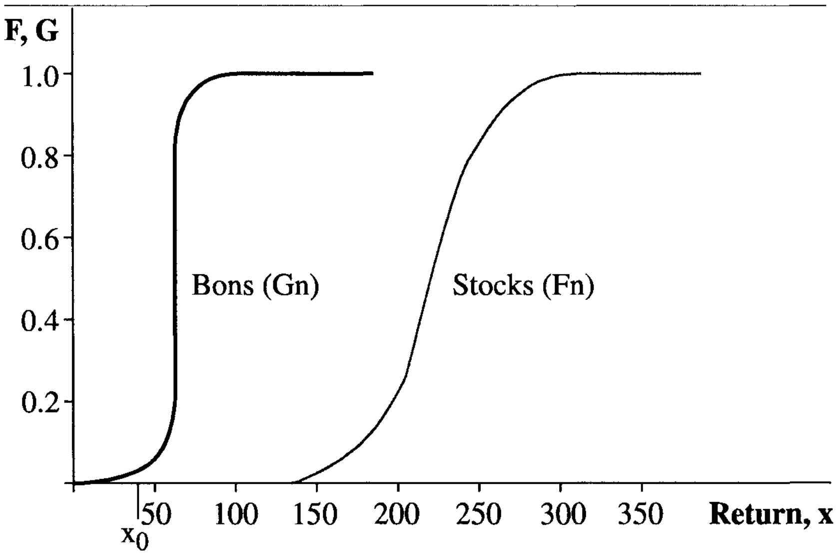


Figure 13.5b F and G for stocks and bonds
n = 50 periods



Indeed, using the above figures for rates of return on stocks and bonds, we calculated $\varepsilon^{(n)}$ for n-period distributions as follows:

$$\varepsilon^{(n)} = \frac{\int_{S_1} [F^{(n)}(t) - G^{(n)}(t)]\,dt}{\int_a^b \left|G^{(n)}(t) - F^{(n)}(t)\right|\,dt}.$$

where $F^{(n)}$ and $G^{(n)}$ are the distribution of $\prod_{i=1}^{n}(1+X_i)$ and $\prod_{i=1}^{n}(1+Y_i)$, respectively, and $\varepsilon^{(n)}$ is the violating area corresponding to the n-period distributions of stocks and bonds divided by the total absolute difference of area enclosed between $F^{(n)}$ and $G^{(n)}$. The following table summarizes the results:

n	$\varepsilon^{(n)}$
1	0.095
2	0.063
5	0.020
10	0.005
50	5.4×10^{-7}

We can see from this table that $\varepsilon^{(n)} \to 0$ as $n \to \infty$, hence the set $U_1^*(\varepsilon)$ expands to include "most" utility functions[3]. This result conforms with what practitioners recommend; as $n \to \infty$ "most" investors should prefer equity to bonds. The fact that $\varepsilon^{(n)}$ decreases as *n* increases has a practical implication for investors. Based on historical data, one can estimate the one-period probability distributions of stocks and bonds. The financial analyst can construct a portfolio F_α when it consists of α percent of stocks and $(1-\alpha)$ percent of bonds (say, for $\alpha = 0, 0.1, 0.2, \ldots, 1$). Then one can construct the *n* period distribution of each of the above portfolios as done above. Then for a given ε, e.g., $\varepsilon = 0.005$, one can find the portfolios corresponding to *n* (investment horizon) which are not AFSD inferior. For example, an investment recipe may be as follows: For $n \leq 5$, invest 40% or less in bonds, for $5 < n \leq 10$ invest 20% or less in bonds, etc. Such a recipe conforms with the common recommendation by practitioners to increase the proportion of equity in the portfolio as the investment horizon increases.

Finally, it is interesting to note that for a myopic utility function, e.g., U(x) = log x, the investment horizon has no effect on choices as long as returns are i.i.d. This does not conform with FSD* or SSD* results given in this Chapter. Let us analyze the properties of the myopic preferences.

[3] The increase in U_1^* is not guaranteed, as for large n, the maximum wealth also increases, $\text{Inf}\ U'(x)$ may decrease, hence inequality (13.7) may be valid for n_1 but not for n_2 where $n_2 > n_1$. However, some experimental evidence showing that this is not the case can be found in Levy, Leshno and Leibovitch, see footnote 4.

With rates of return $x \lesseqgtr 0$, the value $\prod_{t=1}^{n}(1+x_t)$ is the n-period return with a range $a_n = 0$, $b_n > 0$ (b_n may be a very large number). The myopic function is given by $U(x) = x^{1-\alpha}/(1-\alpha)$ with $U'(x) = x^{-\alpha} = 1/x^{\alpha}$ (the log function is a specific case of this myopic function). First note that with n-period investment the expected myopic utility function becomes

$$EU(\cdot) = E\frac{\left(\prod_{i=1}^{n}(1+R_i)\right)^{1-\alpha}}{1-\alpha}$$

and as returns are i.i.d we have

$$E\,U(\cdot) = \left[E(1+R)^{1-\alpha}/(1-\alpha)\right]^{n}$$

Thus, the same diversification policy which is optimum for n=1 is optimum for any n, therefore this function is called a myopic utility function.

How can we reconcile the contradiction between FSD*'s results, showing that the larger n the larger the proportion of equity that should be included in the portfolio and the myopic investment behavior? It is easy to show that the myopic utility function does not belong to $\mathbf{U}_1^*$. To be more specific, Sup [U'(x)]= U'(x = 0)= ∞ (for x = 0) and $\text{Inf}[U'(x)] = \frac{1}{(b_n)^{\alpha}} > 0$

and therefore $\text{Sup}\,U'(x) \le \text{Inf}[U'(x)]\left[\frac{1}{\varepsilon} - 1\right]$ cannot hold for any $0 < \varepsilon < 0.5$.

Thus, the myopic function does not belong neither to $\mathbf{U}_1^*$ nor to $\mathbf{U}_2^*$, which explains why with FSD* the equity becomes more attractive as n increases while with myopic preferences, the diversification between bonds and stocks is not affected by the assumed investment horizon.

13.5 ASD: Experimental Results[4]

We have seen in this chapter that SD rules may create paradoxes in choices and ASD criteria, which allow some ε-area violation resolve the paradoxes. Thus, instead of FSD corresponding to $\mathbf{U}_1$ we have almost FSD (AFSD or FSD*) corresponding to the bounded set $\mathbf{U}_1^*$ and almost SSD (ASSD or SSD*) corresponding to the bounded set $\mathbf{U}_2^*$.

---

[4]The experimental results are taken from Levy, H., M. Leshno and B. Leibovitch, "Bounded Preferences, Paradoxes and Decision Rules," Hebrew University (2005) working paper.

The FSD* and SSD* rules are obtained by imposing some constraints on U', hence $U_1^* \subset U_1$ and $U_2^* \subset U_2$, respectively. These constraints on U' are a function of the relative area violation allowed, denoted by ε. However, in the above analysis and in the derivation of U_i^* (I = 1,2) there is no discussion regarding the magnitude of the allowed ε. In other words, what is the reasonable allowed proportion of area violation, which is denoted by ε? For what value ε, all investors or almost all of them will prefer one option over the other even in the case where SD does not prevail?

To illustrate, consider an example where option A yields \$1 or \$1 million with an equal probability and option B yields \$2 with certainty. Both A and B are in the FSD efficient set. Obviously, not choosing A is considered as a paradox as most, if not all, people, in practice would choose option A . Assume now that rather than \$1 million we would have in option A, \$10,000 with probability of ½. Is it still considered a paradox that A and B are not distinguishable by FSD? What if the \$1 million is replaced by only \$1,000? \$100? or \$10? We define a result of a decision rule as a paradox as long as 100% of the subjects prefer one option over the other despite the fact that by FSD rule (or SSD or MV rules) the two options are indistinguishable, hence are mistakenly included in the efficient set. Alternatively, the two options are mathematically included in the efficient set (in U_i), while in practice only one of the options is in the efficient set (in U_i^*)

The question is how to define U_1^* and U_2^* such that they conform with "all" or at least "most" investors' choices in practice. Namely, what preference although mathematically included in U_1^* and U_2^*, should be ruled out as they do not characterize "all" or "most" investors? We can answer such a question by conducting experiments and learn from the subjects' choices on the reasonable amount of allowed area violation, namely on the reasonable definition of U_1^* and U_2^*.

Suppose that a subject has to choose between two options A and B given below and is asked what should be the minimum value z such that prospect B is preferred over prospect A. Obviously, the larger the selected value z, the smaller the allowed area violation ε. Also note that if an investor allows, say, $\varepsilon = 10\%$ area violation she, *a fortiori*, will allow a lower value, say $\varepsilon = 5\%$. Namely if the investor chooses B, with say z = \$400, *a fortiori*, she will select B with z = \$1,000.

Prospect A		Prospect B	
Probability	Outcome	Probability	Outcome
0.5	\$100	0.5	\$50
0.5	\$200	0.5	\$ z

Note that in this example neither A nor B dominates the other by FSD for z > \$200. Moreover even if $z \to \infty$, such a dominance does not prevail as the two distributions cross. In a large sample of subjects we have found a distribution of values z corresponding to the subjects' choices. The largest value selected (by about 2% of the subjects) was z = \$1,000 which corresponds to the lowest value ε. Thus, for z ≤ \$1,000 all subjects choose Option B despite the no FSD between A and B. A simple calculation reveals that for z = \$1,000, ε = 5.9%. Obviously, for a selection of z < \$1,000 the allowed area violation is even larger, i.e., ε > 5.9%. Thus, if we allow area violation of 5.9% or more there is a dominance by B over A by FSD*, i.e. by *all* subjects. Though for all subjects in our sample, B dominates A, we call this dominance "almost" FSD, as we may have other subjects not participating in our experiment who may require z > \$1,000, namely only a smaller area violation is allowed by them. Yet, we can learn from this experimental study that one should not stick strictly to FSD, and by applying FSD*, paradoxes are resolved and FSD* rule is appropriate to almost all investors. Let us now turn to SSD - ε area violation as determined experimentally.

The following two choices relate to the relationship between SSD and SSD*.

Prospect A		Prospect B	
Probability	Outcome	Probability	Outcome
1/3	\$125	1/3	\$100
1/3	\$150	1/3	\$200
1/3	\$z	1/3	\$300

It is easy to verify that neither A nor B dominates the other by SSD. The subjects were asked what should be the minimum value \$Z such that prospect A would be preferred over B. We obtained a distribution of values z corresponding to the answers of the subjects, with the highest value, once again, is z = \$1,000 which was selected by 2.5% of the subjects.

Drawing the cumulative distributions of A and B, with z = \$1,000 reveals that there is no SSD due to ε = 3.2%, area violation. Yet, 100% of the subjects selected prospect A with z = \$1,000 or less. Therefore, we can safely assert that SSD area violation which is allowed in practice is ε = 3.2% or more. Therefore, with z = \$1,000 we have SSD* of A over B, though we do not have SSD. Of course ε may change from one group of investors to another but by the above two experiments we get a rough idea of the size of ε, which clearly indicates that in practice ASD decision rules and not SD decision rules are relevant.

13.6 SUMMARY

FSD, SSD and M-V investment criteria may reveal no dominance between two options when experimentally or empirically there is an ultimate preference for one of these options – hence paradoxes are created. The paradoxes are induced by the fact that the investment criteria correspond to all preferences (in a given class) including preferences which do not conform with any investor's behavior. We suggest in this chapter decision rules called Almost SD (ASD), which correspond to all preferences, excluding those preferences which mathematically are accepted but do not conform with any investor's behavior. These specific preferences which are excluded can be considered as pathological, unreasonable or simply irrelevant preferences. Thus, Almost FSD replaces FSD, Almost SSD replaces SSD, and Almost M-V replaces M-V rule. The investment criteria AFSD and ASSD allow the establishment of preference despite the fact that there is no FSD or SSD, hence the paradoxes are avoided.

We have the following relationship:

$$\begin{array}{ccc} \text{FSD} & \Rightarrow & \text{AFSD (or FSD*)} \\ \Downarrow & & \Downarrow \\ \text{SSD} & \Rightarrow & \text{ASSD (or SSD*)} \end{array}$$

Key Terms

Almost FSD (AFSD or FSD*)

Almost SSD (ASSD or SSD*)

Almost M-V (AM-V or M-V*)

Area violation - ε

14

NON-EXPECTED UTILITY AND STOCHASTIC DOMINANCE

Most of the economic and finance models that deal with investment decision making under uncertainty are based on the expected utility paradigm. However, experimental studies have shown that subjects often behave in a manner that runs counter to expected utility maximization. Such inconsistencies have been shown to be mainly due to violation of the independent axiom (called also the interchangeability axiom, see Chapter 2). In this chapter, we discuss some of the violations of the expected utility model (for a fuller account, see Machina, [1982 and 1983][1]), and review the modified expected utility theory, well-known as the *generalized expected utility* or *non-expected utility theory*, as well as the competing models that have been developed in order to avoid these violations.

14.1 THE ALLAIS PARADOX

The Allais paradox is a classical example of decision making that runs counter to expected utility maximization. The violation is revealed in a two-part experiment. In part I, a choice is offered between A and B, and in Part II, a choice is offered between C and D, as follows:

PART I:

A: { \$1 million with probability 1}

or:

$$B:\begin{Bmatrix}\$0,\$1\,\text{million, or }\$5\,\text{million with}\\ \text{probabilities of }\ 0.01,\ \ 0.89\ \text{ and }\ 0.1,\text{respectively.}\end{Bmatrix}$$

[1]Machina, Mark A., " 'Expected Utility' Analysis Without Independent Axiom," *Econometrica*, 50, 1982, pp. 270–323, and Machina, M.A., "Generalized Expected Utility Analysis and the Nature of Observed Violations of the Independence Axiom, in Stigum, B., and Wenstøph, F. (eds.) *Foundation of Utility and Risk with Applications*, Reidel, Dordrecht, Holland, 1983.

PART II:

$$C:\begin{cases}\$0, \$1\text{ million with probabilities of}\\ 0.89 \text{ and } 0.11, \text{ respectively.}\end{cases}$$

or:

$$D:\begin{cases}\$0, \$5\text{ million with probabilities of}\\ 0.9 \text{ and } 0.1, \text{respectively.}\end{cases}$$

Results show that in Part I, most investors choose A, and in Part II, most investors choose D. In the following, we show that these decisions are inconsistent and contradict expected utility theory:

The preference of *A over B* implies (all figures in million dollars):

$$1\ U(1) > 0.01\ U(0) + 0.89\ U(1) + 0.1\ U(5), \tag{14.1}$$

and the preference of *D over C* implies:

$$0.9\ U(0) + 0.1\ U(5) > 0.89\ U(0) + 0.11\ U(1). \tag{14.2}$$

Inequality (14.1) can be rewritten as:

$$0.01\ U(0) + 0.1\ U(5) < 0.11\ U(1),$$

and inequality (14.2) can be rewritten as:

$$0.01\ U(0) + 0.1\ U(5) > 0.11\ U(1).$$

This outcome is inconsistent: it suggests either that subjects do not maximize expected utility or that the expected utility model needs to be modified in order to accommodate and explain paradoxical results such as these.

The explanation offered in the literature for this behavior is that subjects overweigh the 0.01 probability of receiving nothing in option B; this explains the preference of A over B in part I of the experiment. This paradox, and many other similar ones revealed in experimental studies, led to the elaboration of the generalized (or non-expected) utility theory which attempts to explain such paradoxes.

14.2 NON-EXPECTED UTILITY THEORY

There are various ways of formulating the axioms of expected utility. In Chapter 2 we provided the simplest set of axioms from which one can derive the rule of expected utility maximization. The FSD criterion did not form part of these axioms but in another formulation the FSD serves as one of the axioms.

Fishburn (1982)[2] suggests the following four preference relationships (denoted by $\succeq$) that are satisfied by expected utility as well as other generalizations and extensions of expected utility theory. The four axioms are:

a) If x_1 and x_2 have the same cumulative distribution, then $x_1 \sim x_2$ where $\sim$ denotes equivalent.

b) The preference $\succeq$ is a *weak order*, which implies that $\succeq$ is *complete, transitive,* and *reflexive.*

c) If x_1 dominates x_2 by FSD, then $x_1 \succeq x_2$.

d) The preference relationship $\succeq$ is continuous.

Thus, by this formulation, the FSD criterion is one of the axioms on which expected utility theory relies. By expected utility theory FSD should not be violated. However, as we shall see below by some extension and generalization of expected utility theory FSD may be violated. Indeed, a number of suggestions have been offered for the modification or generalization of expected utility theory. However, although these modified theories may explain the Allais Paradox, some of these modifications are unacceptable because they violate the FSD criterion (axiom). The expected utility generalization and extension rely heavily on decision weight substitution for probabilities.

Let us now turn to the various modifications of expected utility theory.

a) Probability Weighting

The main explanation of the paradoxes revealed in experimental studies is that investors employ *subjective probabilities* (or decision weights). For example, in the case of the Allais paradox, it has been speculated that the subjects overweigh the 0.01 probability of obtaining $0 in option B and, therefore, mistakenly, select option A. Indeed, the main modification to expected utility theory relies on models that replace the probabilities, p, with *decision weights,* w(p). Experimental studies reveal that the decision weights, w(p), are related to the objective probabilities, p, by an S-shape function (see, for example, Mosteller and Nogee, [1951][3] and Edwards, [1953], [1954][4]). Accordingly, p is replaced by w(p)

---

[2]Fishburn, P.C., "Nontransitive Measurable Utility," *Journal of Math. Psychology*, 26, 1982, pp. 31–67.

[3]Mosteller, F., and Nogee, P., "An Experimental Measurement of Utility,"*Journal of Political Economy*, 59, October 1951, pp. 371–404.

which is a weight function. Moreover, it is suggested that the *subjective expected utility* be calculated. In other words, instead of maximization of $EU(x) = \Sigma p(x)U(x)$, it is assumed that investors maximize $EU^*(x) = \Sigma w(p)\,U(x)$ where the superstar emphasizes that w(p) rather than p are employed.
Maximizing $\Sigma w(p)\,U(x)$ rather than $\Sigma p(x)\,U(x)$ may solve the Allias paradox as well as other paradoxes. However, such a modification in expected utility theory is not acceptable because it may violate the FSD criterion. To see this, consider the following example:

Example:

Assume that before using decision weights, we have the following two options, F and G:

F		G	
Return	Probability (p)	Return	Probability (p)
9	1/10	9	2/10
10	7/10	10	6/10
11	2/10	11	2/10

Because $F(x) \le G(x)$ for all values x and $F(9) < G(9)$ (i.e., we have at least one value with a strict inequality), F dominates G by FSD. We conduct the following probability transformation $w(p) = p^2$ to obtain new functions F* and G* corresponding to F and G.

F^*		G^*	
Return	Decision weights [w(p)]	Return	Decision weights [w(p)]
9	1/100	9	4/100
10	49/100	10	36/100
11	4/100	11	4/100

We have $F^*(10) = 50/100 > G^*(10) = 40/100$; hence, the FSD of F over G is violated by the weight function $w(p) = p^2$. Note that we obtain the violation of FSD because F dominates

[4]Edwards, W., "Probability Preferences in Gambling," *American Journal of Psychology*, 66, 1953, pp. 349–364 and Edwards W., "Probability Preferences Among Bets with Differing Expected Values," *American Journal of Psychology*, 67, 1954, pp. 56–67.

G by FSD but F* does not dominate G* by FSD. One might suspect that the FSD violation is due to the fact that after the transformation, we have Σ w(p) < 1 (i.e., the decision weight function is not a probability measure). This is not the case. Comparison of F** and G** (where F** and G** are derived by normalization from F* and G*, respectively) reveals that with the normalization, $w_N(p) = w(p)/\Sigma w(p)$, where the subscript N indicates that the weights w(p) are normalized, we again find that the FSD is violated. This is illustrated in the next table:

F**		G**	
Return	Cumulative probability with normalized decision weights w(p)	Return	Cumulative probability with normalized decision weights w(p)
9	1/54	9	4/44
10	49/54	10	40/44
11	1	11	1

With this normalization of the decision weights, $F^{**}(10)=50/54\cong 0.926>G^{**}(10)=40/44\cong 0.909$; hence, the FSD of F over G is also violated with $w_N(p)$, the normalized weight function of w(p).

To sum up, replacing p with decision weight w(p) violates the FSD criterion. Fishburn (1978)[5] proves that with two outcome returns, the violation of the FSD criterion can be avoided only with the specific weight function w(p) = p for all p. This implies returning to expected utility maximization; hence the Allias and other paradoxes persist.

b) Prospect Theory's Decision Weights

By Prospect Theory (PT) of Kahaneman & Tversky [1979][6] the decision weight of probability p is w(p). PT's decision weight scheme has several advantages over other decision weights schemes. The advantages are as follows:

1) If $p_i = p_j$ also $w(p_i) = w(p_j)$, i.e., the same decision weight is assigned to identical probabilities. This is in particular crucial for empirical studies in economics and finance where a probability of 1/n is assigned to each observation, n being the number of observations.
2) Though for equally likely outcomes we generally have by PT that w(p) $\neq$ p, still the choices are unaffected by the decision weights. Namely,

[5]See Footnote 2.
[6]Kahaneman, D.K., and Tversky, A., "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, *47*, 1979, pp. 263-291.

for two options F and G the following is intact,

$E_F U(x) \geq E_G U(x)$	$\Leftrightarrow$	$E_F U(x) \geq E_G U(x)$
with objective probabilities		with PT's decision weights

For a proof see Levy & Levy [2002][7]

3) With PT's decision weights, one has the flexibility to have $w(p) \neq p$ for relatively small probabilities and $w(p) \cong p$ for relatively large probabilities.

The main disadvantage of PT's decision weights is that it may violate FSD. To illustrate this consider two prospects F and G as follows:

F		G	
Return	Probability	Return	Probability
5	½	10	1
10	½		

Obviously G dominates F by FSD. Select PT's decision weight as follows:

$$w(\tfrac{1}{2}) = \tfrac{3}{4}$$

$$w(1) = 1$$

It is easy to find a monotonic utility function U_0 such that,

$$\tfrac{3}{4}\, U_0(5) + \tfrac{3}{4}\, U_0(10) > U(10)$$

hence, for this $U_0 \in \mathbf{U}_1$ option F is preferred, which violates FSD. Tversky & Kahaneman (T&K) [8] who realized this drawback of PT's decision weight suggested in [1992] Cumulative Prospect Theory (CPT) as a substitute for PT, where the main difference relates to the decision weights. By CPT's decision weights, FSD is not violated. To this we turn next.

[7]See Levy, H. and M. Levy, "Experimental Test of the Prospect Theory Value Function: A Stochastic Dominance Approach," *Organizational Behavior and Human Decision Processes*, *89*, 2002, pp. 1058–1081.

[8]Tversky, A. and Kahaneman, D.K., "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty, 5,* 1992, pp. 297-323.

c) CPT's Decision Weights

T&K [1997] estimate experimentally the decision weights function separately for negative outcomes and separating for positive outcomes. These suggested formulas for decision weights determination is as follows:

$$\left.\begin{aligned} w^{*-}(P) &= \frac{P^{\delta}}{[P^{\delta}+(1-P)^{\delta}]^{1/\delta}} \\ w^{*+}(P) &= \frac{P^{\gamma}}{[P^{\gamma}+(1-P)^{\gamma}]^{1/\gamma}} \end{aligned}\right\} \qquad (14.3)$$

where $\gamma = 0.61$, and $\delta = 0.69$, P is the cumulative (objective) probability, and $w^{*}(P)$ is the cumulative decision weight, where $w^{*-}(P)$ relates to the negative outcomes and $w^{*+}(P)$ relates to the positive outcomes. These weighting functions have a reverse S-shape. From these cumulative weights function one can derive the individual outcomes decision weights (in the discrete case). It can be shown that with CPT's decision weights FSD is not violated. Other researchers who followed T&K's research, estimate the parameters of the reverse S-shape weighting function and obtain a little different values for δ and γ (see, for example, Wu and Gonzales [1996][9] and Abdellaoui [2000][10]). Prelec [1998][11] states a set of axioms from which he derives several w(P) forms. With his main formula for decision weights, $w(P)=\exp\{-(\ln P)^{\alpha}$, $(0<\alpha<1)$, he obtains a decision weight function which is characterized by similar properties of CPT's decision weight function. In particular, it also has no flexibility, as for a given cumulative probability P, w(P) is determined regardless of the left tail of the distribution of outcomes (see the Figure on p. 498 in Prelec [1998], see footnote 11).

In all the above mentioned studies the weighting function has an inverse S-shape, hence probability decision weights in the center of the distribution tend to be smaller than the objective probabilities and the opposite holds with regard to the left and right ends of the distribution. There is a strong experimental support that in some situations, especially in the case of "long shots", decision weights rather than objective probabilities are employed and that indeed the decision weight function has an inverse S-shape. Indeed, it is worth noting that formula (14.3) was estimated by T&K mainly with bets with small probabilities, e.g., 0.1. One of the basic issues regarding CPT's decision weights is whether one can generalize this formula and apply this probability weighting function to other bets, e.g., bets with relatively large probabilities, say, $p \geq 0.25$, and particularly to equally likely

[9]Wu, G., and Gonzales, R., "Curvature of the Probability Weighting Function," *Management Science, 42, 12,* 1996, pp. 1676-1690.

[10]Abdellaoui, M., "Parameter Free Elicitation of Utility and Probability Weighting Functions," *Management Science, 2000, 46*, pp. 1497-1512.

[11]Prelec, D. "The Probability Weighting Function," *Econometrica, 66,* 1998, pp. 497-527.

outcome bets i.e., $p_i=1/n$ when n = 2,3,4,... We claim that the CPT's decision weight function cannot be generalized and employed in many important cases, and it is particularly inappropriate in the equally likely events. For example, employing CPT's decision weights to outcomes –\$500, –\$300, +\$500, +\$1000 with an equal probability of p = ¼, implies that the decision weights are .29, .16, .13 and .29, respectively, i.e., CPT advocates decision weights which are very hard to accept in this equally likely event case.

Thus, CPT's decision weight formula has the advantage of not violating FSD, yet it has its drawbacks: no flexibility in determining w(p) and unreasonable decision weights in the equally likely case. Note that not all researchers agree that equally likely probabilities should be replaced with decision weights. For example, Viscusi [1989][12] reveals evidence that in such a case w(1/n)=1/n. Similarly, by PT the decision weight in the above example is identical for all values as long as p=¼ for all outcomes. Thus, to some extent PT's decision weights are in line with Viscusi's approach, as using decision weights rather than probability does not change the choices (see footnote 7). A completely different result emerges with CPT's and Prelec's decision weights. In these models the decision weight function reveals an inverse S-shape which may strongly affect the choices. From the above discussion, we can see that the theoretical advantages of CPT over PT, namely not violating FSD, has its cost: In some important cases, CPT, in our view, determines decision weights which are very hard to accept: the probability of \$500 decreases from .25 to .13 (see above example) and the probability of \$1,000 increases from 0.25 to .29. Assigning to \$1,000 more than double(!) decision weight than to \$500 has no empirical proof or an intuitive explanation. A similar argument is intact with Prelec's decision weights. This extreme decision weights do not occur with PT, as in our example we would have in the case $p_i=¼$, the same decision weights to all outcomes.

d) Rank Dependent Expected Utility (RDEU) and FSD

We have seen above that transformation of probabilities may violate the FSD criterion. To avoid this violation and paradoxes such as the Allias paradox, it is suggested by RDEU model that a probability transformation be carried out on the cumulative distributions F(x) and G(x) rather than on the individual probabilities: In other words, it is assumed that investors compare $F^*(x) = T(F(x))$ and $G^*(x) = T(G(x))$ where T is a monotonic non-decreasing transformation, $T'(\cdot) \geq 0$ (see Yaari, [1987][13]; Tversky and Kahaneman [1992][14]

[12]Viscusi, W.K., "Prospective Reference Theory: Toward an Explanation of Paradoxes," *Journal of Risk and Uncertainty, 2,* 1989, 235-264.

[13]Yaari, M., "The Dual Theory of Choice Under Risk," *Econometrica*, 55, 1987, pp. 95–115.

[14]Teversky, A. and D. Kahaneman, "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 1992, pp. 297–323.

and Quiggin [1991][15]). The decision model in which F is replaced with weights T(F) is called the *rank dependent expected utility* (RDEU) model. RDEU model does not violate FSD because the following Relationship exists: $F(x) \leq G(x) \Rightarrow T(F(x)) \leq T(G(x))$. This property holds because T is monotonic with $T'(\cdot) \geq 0$. The transformation should also fulfill the constraint $T(1) = 1$. In the discrete case, RDEU suggests maximization of a function V(w,p) where:

$$V(w,p)=\sum_{i=1}^{n} U(x_i)w_i(p)\,,$$

where:

$$w_i(p)=T(\sum_{j=1}^{i} p_j)-T(\sum_{j=1}^{i-1} p_j)$$
$$= T(F(x_i)) - T(F(x_{i-1}))$$

where T is a transformation such that $T[0,1] \rightarrow [0,1]$.[16]

Thus, non-expected utility models are needed to explain the observed behavior of subjects in experimental studies. It seems that RDEU is the most promising modification in expected utility because it explains subjects' behavior and unlike the weighting of individual probabilities (as done in PT), it does not violate the FSD criterion.

e. Configural Decision Weights (CDW)

Birnbaum and Navarrete [1998][17] (B&N) and Birnbaum [2004][18] suggest decision-making models called configural weight (CW) models. By these models if option F has an outcome of $1,000 with a probability of, say, 0.1, and option G has two outcomes of $1000, each of which with a probability of 0.05, according to CW models the decision weights assigned to the two branches of G are together larger than the decision weight assigned to 0.10, corresponding to $1000 of prospect F. The CW models are conceptually different than the other decision weighting schemes as CW depends on the structure of the branches of the

[15]Quiggin, J., *Generalized Expected Utility Theory, The Rank Dependent Model*, Kluwer Academic Publishers, Boston, 1993.

[16] Levy & Wiener show that for SSD or TSD not to be violated by the transformation, the requirement $T''(\cdot) \leq 0$ and $T'''(\cdot) \geq 0$, respectively should be added. See, Levy, H., and Wiener, Z., "Stochastic Dominance and Prospect Dominance with Subjective Weighting Functions," *Journal of Risk and Uncertainty*, *16*, 1998, pp. 147–163.

[17]Birnbaum, M.H., and Navarrete, J.B., "Testing Descriptive Utility Theories: Variations of Stochastic Dominance and Cumulative Independence," *Journal of Risk and Uncertainty, 17*, 1998, pp 49-78.

[18]Birnbaum, M.H., "New Paradoxes of Risky Decision-Making, 2004, California State University, Fullerton, Working Paper.

uncertain outcome. Thus, by splitting a given outcome to two branches we may, irrationally, affect choices.

14.3 DECISION WEIGHTS AND FSD VIOLATION

According to CPT's, Prelec's and RDEU decision weights, FSD should not be violated. Namely, if for all $U \in \mathbf{U}_1$

$$E_F U(x) \geq E_G U(x) \text{ also } E_{F^*} U(x) \geq E_{G^*} U(x), \text{ where}$$

F and G are the prospects with objective probabilities and F* and G* are the prospects with decision weights. Yet, while with RDEU the functions F* and G* are cumulative probabilities this is not the case with CPT's and Prelec's decision weights as with these two models we may have $\sum w_i(p) \lesseqgtr 1$.

Obviously, with EUT when probabilities are not distorted FSD should not be violated and any choice of G where $F(x) \leq G(x)$ for all x is considered as an irrational choice. With PT as illustrated above, FSD may be violated and with CW models, when options have some specific branches, FSD is *predicted* to be violated.

The issue whether, in practice, FSD is violated or not, is an empirical or experimental question. In this section we present some experimental findings regarding FSD violation. We present some FSD experimental tests with and without monetary payoff.

The following Table represents an experimental study where several groups of subjects have to choose from two prospects.

F		G	
Outcome in $	Probability	Outcome in $	Probability
−100	½	−100	¼
+400	½	−75	¼
		+400	½

Obviously, G dominates F by FSD. However, to see it transparently one needs to split the first outcome of F and write it as {(−100, ¼), (−100, ¼ } instead of (−100, ½). We find experimentally that to reveal the FSD of G over F with no splitting, as presented to the subjects, is not a simple and obvious task, as we observe in this experiment FSD violations. It is important to emphasize that the FSD violations exist also in some cases even where monetary payoff was involved.

Table 14.1 presents the various groups of subjects and the choices.[19] Where monetary payoff was involved, a lottery was conducted in front of the subjects and they observe the realized outcome corresponding to their choice. Namely, if the subject selects, say, F, and the lottery shows 400 she gets the prize of about $9. A negative outcome implies a proportional loss. The subject cannot lose in the experiment.[20]

Table 14.1 The choices in the Experiment (in %)

Group	Subjects	Number of Subjects N	Choice of Prospect		Total
			F	G	
I	Undergraduate Business Students, **no** monetary Payoff	58	15.5	84.5	100.0
II	Mutual funds managers and financial analysts, **no** monetary payoff	42	7.1	92.9	100.0
III	Second year MBA Students No exposure to FSD Criterion **with** monetary payoff	23	21.7	78.3	100.0
IV	Second year MBA students, all studies FSD **with** monetary Payoff	27	22.2	77.8	100.0
V	Advanced MBA students and Ph.D candidates: all studied expected utility and FSD, **with** monetary payoff	15	13.3	86.7	100.0
Total	Aggregate across all groups	165	15.2	84.8	100.0

[19]This table reports only one experiment's results of a more extensive study. For more details, see Levy, H., "First degree Stochastic Dominance Violations: Decision Weights and Bounded Rationality," 2005, Working Paper, the Hebrew University of Jerusalem.

[20]The payoff is determined by the outcome offer deleting one zero and stating it in Israeli Shekels. Thus, –$100 is involved with a loss of 10/4.5 ≅ $2.2 where 4.5 is the exchange rate between Israeli Shekels and US dollars. The subjects received an initial endowment such that even if –100 occurs, they end up with zero net balance.

We see from this experiment that even with a very simple choice there is a substantial amount of FSD violations. The lowest proportion of FSD violations is done by group II, which is composed from mutual fund managers and financial analysts (7.1% of FSD violations). It is interesting to note that the existence of a monetary payoff, or the degree of knowledge of the subjects in expected utility theory and stochastic dominance theory did not affect much the results. Yet, unlike in N&B's study, in this experiment all participants are business students and most of them are specializing in finance. Over all, we have 7.1%–22% of FSD violations with an average across all groups of 15.2% of FSD violations. Obviously we find significantly less than 50% violations, hence EUT, CPT and RDEU which advocate no FSD violations cannot be rejected solely based on these results. A possible explanation for the selection of the FSD inferior option by about 15% of the subjects may be related to the fact that G is characterized by two possible negative outcomes and F has only one negative outcome. Of course, splitting the −100 of F to two outcomes of −100 and with a probability of ¼, would eliminate this framing effect. Another possible explanation for the FSD violations is by the configural decision weight model of Birnbaum. By this model the branch (−100, ¼) (−75, ¼) gets relatively high weight which makes G inferior. Thus, we find that splitting the outcome (−100, ½) to (−100, ¼), (100, ¼) presumably could not be done by about 15% of the subjects which induces them to choose the inferior FSD choice.

Thus, with a relatively simple FSD dominance when only splitting of the probability is needed to have a transparent FSD, we find that across all subjects 15.2% of FSD violations. Recall that in this experiment we have choice in line with Birnbaum's recipe for FSD violation, albeit not as complicated as the one suggested by him, which may explain why we got only 15.2% of FSD violations. The violations of FSD seem to be due to the computational difficulties and in particular due to the difficulty to simplify the choices.

To sum up, the FSD is violated by about 15% of the subjects which may have two possible interpretations.

a) CW's or PT's decision weights are employed which may explain the FSD violations.
b) RDEU or CPT decision weights are employed, or EUT is valid, and the violations observed are due to bounded rationally.

We tend to accept explanation b) as other experiments, not reported here, reveal that the more complicated the choice the larger the proportion of FSD violations. Moreover, the FSD violations are obtained with and without "branches" which, in turn, are the main reason why CW models predict FSD violation.

14.4 TEMPORARY AND PERMANENT ATTITUDE TOWARD RISK

Experiments have shown that if one has, say, \$10 and looses \$1, he/she feels worse than having \$9 without losing or gaining anything. This contradicts expected utility theory because, in both cases, subjects will have \$9 and, therefore, by the expected utility theory, they ought to be indifferent between the two options. Such observed results led Kahaneman & Tversky (K&T) (1979)[21] to formulate *Prospect Theory* (PT) in terms of *changes* in wealth rather than total wealth. We suggest in this section a dynamic decision making process, which bridges the gaps between PT and EUT approach, i.e., between the employment of change of wealth and total wealth.

The main experimental findings in support of PT and CPT, are as follows:

1) The majority of the subjects violate expected utility exactly as shown by the Allais paradox.

2) Subjects commonly assign values to changes in wealth rather than to total wealth which, again, contradicts the expected utility paradigm: Subjects maximize the expected *value function* V, which is a function of *change* in wealth, rather than expected utility, U, which is a function of the *total* wealth.

3) The value function V is S-shaped; it is concave for gains (risk aversion) and convex for losses (risk seeking). The value function is steeper for losses than for gains. Yet, the steepness of the value function V depends on wealth, w.

4) Decision weights, w(p), are different from corresponding probabilities p.

Levy and Wiener's (L&W) model (1997)[22] bridges the gaps between expected utility theory (EUT) and PT. According to this model, decision making based on changes in wealth (PT) reflects *temporary attitude toward risk* (TATR), and decision making based on final wealth (EUT) reflects *permanent attitude towards risk* (PATR). However, even TATR is a function of wealth, w: the value function can be written as $V_w(x)$, where the changes in wealth x are emphasized. The TATR can be described by a *path-dependent utility* function, and the PATR, by the von- Neumann and Morgenstern utility function. The combination of TATR and PATR can be used to explain the observed (seemingly) paradoxical experimental results, in particular, those obtained by K&T (1979, 1992) and Thaler and Johnson (1990)[23], as well as

---

[21] Kahneman, D. and Tversky, A., "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 1979, pp. 263–291.

[22] Levy, H., and Wiener, Z., "Prospect Theory and Utility Theory: Temporary and Permanent Attitude Toward Risk," Working paper, The Hebrew University of Jerusalem, 1997.

[23] Thaler, R.H., and E.J. Johnson, "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choices," *Management Science*, 36, 1990.

the well-documented phenomenon of short-term overreaction of stock prices. This approach allows for explanation of the aforementioned paradoxes in the framework of the expected utility paradigm. The integration of these two competing theories suggests that there is a value function $V_w^*(x)$ corresponding to initial wealth, w and change in wealth, x. This function has two components:

$$V_w^*(x) = U(w) + V_w(x).$$

The above function is also equal to the two dimensional path-dependent utility function $U^*(w, x)$. By writing $U^*(w, x)$, the two-parameters w and x are emphasized, and comparison with the von-Neumann and Morgenstern utility function U(w+x) is feasible. By writing $V_w^*(x)$, comparison with K&T value function, V(x), is feasible. Because $V_w^*(x) \equiv U^*(w,x)$, the terms, path value function and path-dependent utility function can be used interchangeably.

Note that $U^*(w, x) \neq U(w+x)$. The asterisk is added to emphasize that this is not the von-Neumann and Morgenstern utility function, but a path-dependent two-parameter utility function. The function $U^*(w,x)$ has the following path-dependent properties:

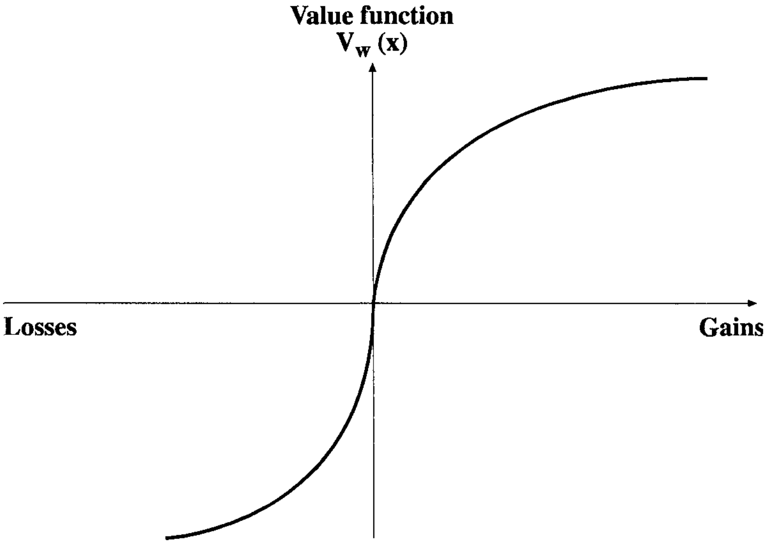
$$U^*(w, 0) < U^*(w-x, x)$$

$$U^*(w, 0) > U^*(w+x, -x), \text{ where } x > 0.$$

Note that the investor's final wealth is w in all cases; the path utility level is determined by the way in which w is achieved.

Prospect Theory (PT) emphasizes changes of wealth and not final wealth, but it does not ignore initial wealth. K&T claim that initial wealth serves as a reference point, and investors evaluate the changes in wealth relative to this reference point.

Figure 14.1a: The Value Function



14.1b: The Path-dependent utility function

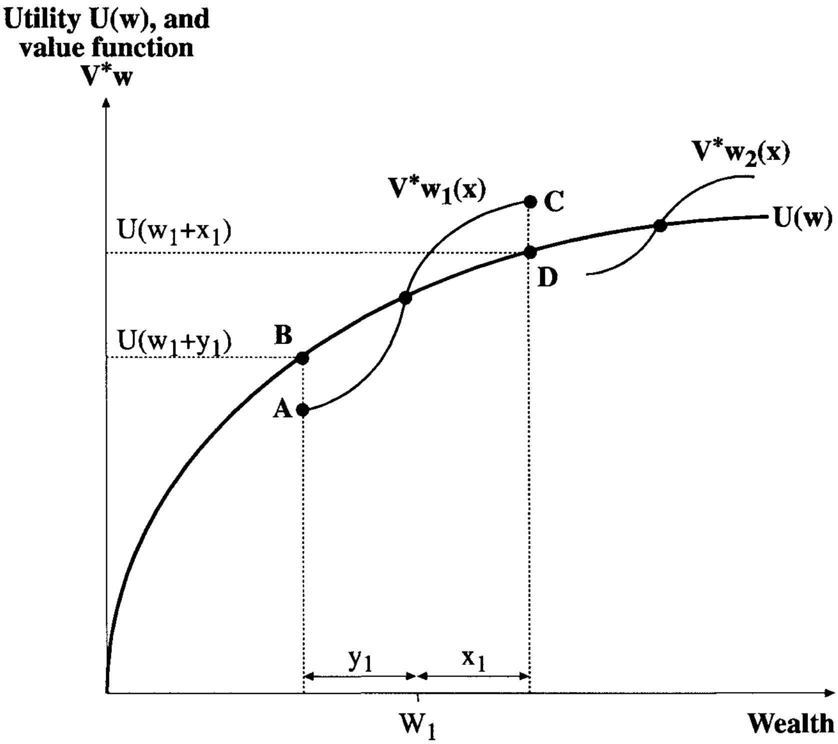


Figure 14.1a illustrates the K&T value function V(x) with zero as the reference point (V(0)=0). This value function is S-shaped, but the slopes right and left of $x = 0$ are likely to depend on the investor's initial wealth. Thus, $V_w(x)$ in Figure 14.1a measures the value function as a function of changes in wealth. Figure 14.1b illustrates the path-dependent utility function $V_w^*(x)$ (which is equal to $U^*(w, x)$), where we use w as the reference point. In other words, we add the values given in Figure 14.1a to U(w) as follows:

$$V_w^*(x) = U(w) + V_w(x).$$

Now, if $x \neq 0$, we have two different functions – the value function $V_w^*(x)$ and the utility function U(w+x) as described in Figure 14.1b. Functions $V_{w_1}^*(x)$ and $V_{w_2}^*(x)$ are two hypothetical value functions for two reference points $w_1 < w_2$. The function $V_{w_1}^*$ as a function of x is S-shaped (for each value of w). For x=0, we have $V_w^*(0) = U(w)$. According to K&T, the value function $V^*(x)$ (as well as V(x)) has the following properties for all values of w: $V^{*\prime}(x) > 0$ for all x; $V^{*\prime\prime}(x) > 0$ for $x < 0$ and, $V^{*\prime\prime}(x) < 0$ for $x > 0$ (prime denotes derivative with respect to x). Friedman and Savage (1948)[24] and Markowitz (1952)[25] offer several explanations for the combination of risk seeking and risk aversion corresponding to various domains of x. Our explanation is in the spirit of K&T's PT (1979).

The value function represents the Temporary Attitude Towards Risk (TATR). In particular, if for wealth w_1 an unexpected income x_1 is obtained, the value function will be $V_{w_1}^*(x_1)$ (see point C, Figure 14.1). If an unexpected amount y_1 has to be paid (or is lost), the value function will be $V_{w_1}^*(-y_1)$ (see point A). Put differently, in the short run, decrease in income from w_1 to $(w_1 - y_1)$ will be somewhat painful; hence, the value function $V_{w_1}^*(-y_1)$ will be lower than

$U(w_1 - y_1)$ (see Figure 14.1b). To illustrate, suppose that \$10,000 is invested in the stock market and the stock price goes down by 20%. Even though the investor is left with wealth amounting to \$8,000, the blow of losing \$2,000 will make him/her feel as if less than \$8,000 is left. However, this feeling will be temporary. After a short while (hours, days, weeks) the investor will adapt to the new wealth position; in other words, feels that he/she has \$8,000, and not less. This will signify that the investor has come to terms with the loss (see K&T, 1979, p.287). Thus, after the overreaction subsidies, there will be a shift from point A to point B, and from point C to point D (in the case of gain), reflecting the Permanent Attitude Towards Risk

[24]Friedman, M., and Savage, L.J., "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy,* August 1948.

[25]Markowitz, H. M., "Portfolio Selection," *Journal of Finance*, 7, 1952, pp. 77–91.

(PATR). Thus, the value function, V^* , reflects the investor's value function at the time of making the decision. Some time has to pass (to adjust to the new wealth position) before the shift from the value function V^* to the utility function $V^*_{w+x}(0) = U(w+x)$ occurs. Experiments have shown that investors base their choices on value function V^* rather than utility function U. This suggests that investors are myopic: they focus on short-term values, oblivious to the effect of the shift back to utility function U in the long-term, after the overreaction (due to the loss) or joy (due to the gain) passes. According to the path-dependent utility function, the value function is employed in arriving at the decision and, after the initial reaction to gain or loss, the investor shifts to the base utility function U, in preparation for a new decision which will be based, once again, on a new value function, V.

L&W (1997) suggest that investor preference is given by a path-dependent utility function $U^*(w, x)$ given by: $U^*(w, x) \equiv V^*_w(x) = U(w) + V_w(x)$ where U^* reflects wealth as well as changes in wealth, U(w) is the utility function where the initial wealth, w, is the reference point, and $V_w(x)$ is K&T's value function measuring the *additional* value due to *changes* of wealth, x. TATR implies that the investor makes decisions by the $U^*(w, x)$ function which is analogous to K&T's claim that investors make decisions by V(x). However, after x is realized, and after the passage of time during which the investor adjusts to changes in wealth, we have:

$$V^*_{w+x}(0)=U(w+x)$$

which reflects the investor's PATR (the intersection point of $V^*_w(\cdot)$ and $U(\cdot)$, see Figure 14.1b). If the investor is then faced with a new investment opportunity involving a random change in wealth y, this process will be repeated and the investor's preference will be given by:

$$U^*(w+x, y) \equiv V^*_{w+x}(y)=U(w+x)+V_{w+x}(y) \qquad (14.3)$$

where the decision is made by V^* at reference point $w_1 = w + x$, reflecting the TATR. Thus, Prospect Theory, by adopting a path-dependent utility function U(w, x) with TATR and PATR concepts, can be integrated into classical expected utility theory, and most observations in experimental studies can be explained in this framework.

14.5 SUMMARY

Although the theory of expected utility theory (EUT) is still the dominant model in the economics of uncertainty, it has some limitations, which have led researchers to modify the theory, or to develop competing theories.

The main source of criticism of the expected utility theory is derived from results obtained in experimental studies which suggest that investors do not behave in accordance with maximization of expected utility: It seems that there is overweighing of very small probabilities (the Allais paradox), and that, contrary to expected utility theory, investors base their decisions on changes in wealth rather than total wealth.

The non-expected utility theory was developed in order to accommodate results such as these. The suggestion by PT and CW models that decision weights w(p) should replace the probability p was found to be unacceptable because such a transformation violates the FSD criterion. A rank dependent expected utility (RDEU) theory was suggested to overcome this drawback whereby the transformation is conducted on the cumulative probability, that is, instead of comparing two cumulative distributions F and G, decision makers compare $F^* = T(F)$ and $G^* = T(G)$, where T is a monotonic transformation with $T' > 0$. Also, CPT's decision weight model does not violate FSD. The advantages of such a transformation is that it resolves some of the paradoxes produced in experimental studies without violating the FSD criterion.

Experimental studies reveal that about 15% of the choices violate FSD. These violations seem to be induced by bounded rationally rather than a theoretical model which predicts FSD violation. Moreover, the simpler the choice, the less FSD violations are observed.

Prospect theory (PT) is based on the claim that investors maximize expected value function, $V_w(x)$, where w denotes wealth and x denotes change in wealth, rather than maximizing the expected utility U(w+x). Levy & Wiener's (1996) model bridges the gap between prospect theory and expected utility theory. According to this theory, decisions by investors are made in two stages: in the short run, they behave according to the value function $V_w(x)$ (temporary attitude toward risk (TATR)) and in the long run, they behave according to the utility function U(w+x) (Permanent attitude toward risk (PATR)).

KEY TERMS

The Allais Paradox

Generalized Utility Theory

Non-Expected Utility Theory

Weak Order

Decision Weights

Subjective Expected Utility

Rank Dependent Expected Utility (RDEU)

Prospect Theory (PT)

Value Function

Temporary Attitude Toward Risk (TATR)

Permanent Attitude Toward Risk (PATR)

Path-Dependent Utility Function

Cumulative Prospect Theory (CPT)

Configured Weight (CW) model

Bounded Rationality

FSD Violations

15

STOCHASTIC DOMINANCE AND PROSPECT THEORY

In the year 2003 Daniel Kahaneman won the Nobel Prize for Economics mainly for his joint contribution with the late Amos Tversky, called Prospect Theory (PT) and its latest modified version called Cumulative Prospect Theory (CPT).

The main features of PT and CPT are:

a) Preference is S-shape with reference point at x = 0

b) Investor maximizes the expected value of V(x) when V is the preference and x is the *change* of wealth

c) Decision makers employ decision weight w(p) rather than objective probability p, where by CPT the weight function w(P) has a reverse S-shape where P is the cumulative probability.

Items b) and c) are in sharp contradiction to von-Neumann and Morgenstern expected utility theory. In addition, item a) above implies that risk aversion does not prevail, which is in contradiction to most equilibrium models in economics which assume risk aversion. In this chapter we focus on factors a) and b) above, i.e., we develop decision rules which assume no change in probability. However, factor c) (i.e., decision weights) is incorporated in some specific theoretical cases and mainly in the experimental studies given at the end of the chapter. The decision rule presented here is called Prospect Stochastic Dominance, yet it refers only to factors a) and b) of PT given above, and can be extended to incorporate factor c) only in very specific cases.

The purpose of this chapter is as follows:

1. To examine whether FSD rule is valid within CPT framework.

2. To develop Prospect Stochastic Dominance (PSD) rule which corresponds to all S-shape preferences.

3. As Markowitz [1952] (see footnote 3) suggests that preferences are reverse S-shape, we also develop Markowitz Stochastic Dominance rule (MSD) corresponding to all reverse S-shape preferences.

4. Assuming that CPT and alternatively Markowitz's preference are valid, to examine their impact on the equilibrium CAPM of Sharpe [1964] and Lintner [1965], (see Chapter 12).

5. To test whether CPT is a valid theory.

6. To test whether risk-aversion prevails.

As CPT relies on experimental results we also analyze CPT experimentally. However, the difference between traditional experimental studies, which support CPT, and our studies is that we use here SD approach, while the studies which support CPT employ the certainty equivalent (CE) approach with the "certainty effect" drawback which affects the results. As CPT is the modified version of PT in this chapter we refer only to CPT.

15.1 CPT AND FSD RULE

By CPT the value function is an S-shape, therefore all SD rules which assume risk aversion in the whole range are irrelevant. However, FSD, which assumes only that $U' > 0$ is relevant. As for FSD we have,

$$F(w+x) \leq G(w+x) \Leftrightarrow F(x) \leq G(x)$$

relying on change of wealth, x, as advocated by CPT and not on total wealth, w+x, as advocated by EU theory, does not affect the FSD dominance relationship. The CPT's decision weights formula, as well as other monotonic decision weight functions e.g., the RDEU's decision weight formula also do not affect the FSD (see previous chapter). Therefore, FSD is theoretically intact also with CPT (and RDEU).

15.2 PROSPECT STOCHASTIC DOMINANCE (PSD)

In Chapter 3 we defined FSD, SSD and TSD. In this section, we define the notion of *prospect stochastic dominance* (PSD) which corresponds to S-shaped value and utility functions. In the derivation of PSD, it is assumed, as in Chapter 14, that the utility function is path-dependent $U^*(w, x)$, and that $V_w(x)$ is an S-shaped value function. With no constraints on the relationship between $V_w(x)$ and w (as long as it remains S-shaped), it is possible to derive conditions for dominance of F over G for all (S-shaped) value functions. First we assume a value function, $V_w(x)$, as advocated by K&T (see Figure 14.1a in the previous chapter) and then we show that the results remain intact for path-dependent utility function $U^*(w, x)$. The results also hold for the von-Neumann and Morgenstern utility function U(w+x) with the inflection point at the initial wealth, w.

Suppose that we have two uncertain options with density functions f(x) and g(x), and that the corresponding cumulative distribution functions are F(x) and G(x), where x denotes change in wealth (gains or losses). Then, by PT, F will be preferred over G if and only if the following holds:

$$E_F V_w(x) = \int V_w(x)dF(x) \geq \int V_w(x)dG(x) = E_G V_w(x) \quad (15.1)$$

where $V_w(x)$ is the K&T (S-shaped) value function, and the precise shape of this function depends on current wealth, w (see Figure 14.1a). Define the set of all functions $V_w(x)$ with $V'_w(x) \geq 0$ and $V''_w(x) \geq 0$ for $x < 0, V''_w(x) \leq 0$ for $x > 0$ by $\mathbf{V}_s$, where the subscript s stands for S-shaped function. In the following theorem, we derive PSD which provides the necessary and sufficient conditions for preference of F over G for all value functions, $V_w(\cdot) \in \mathbf{V}_s$. If F and G are the cumulative distributions stated in terms of the objective distributions, we implicitly assume no probability distortion. However, F and G may be stated also in terms of subjective probabilities, a case when decision weights are incorporated. However, in this case PSD has a meaning only if F and G are still probability distributions and all investors employ the same decision weight function. Alternatively, all investors can be divided into groups, each of which employ a given decision weight function. In such a case we derive the PSD efficient set separately for each group of investors.

Theorem 15.1: Let $V_w(\cdot)$ be an S-shape value function, $V_w(x) \in \mathbf{V}_s$. Then:

$$\int_y^x [G(t)-F(t)]dt \geq 0 \text{ for all pairs } y<0<x \Leftrightarrow E_F V_w(x) \geq E_G V_w(x) \quad (15.2)$$

for all $V_w(x) \in \mathbf{V}_s$, where F and G are the cumulative distribution functions of f and g, respectively. If all subjects employ the same decision weight formula, then F and G can be considered as the distorted cumulative probability functions, provided the decision weight function is a probability measure. If eq. (15.2) holds, we say that F dominates G by PSD.

Condition (15.2) can be rewritten as

$$\int_x^0 [G(t)-F(t)]dt \geq 0 \text{ for } x < 0$$

and

$$\int_0^x [G(t)-F(t)]\, dt \geq 0 \text{ for } x > 0$$

Proof: For simplicity, we assume that gains and losses are bounded by [a, b][1], $a < 0 < b$.

a) *Sufficiency*:

Define by $\Delta \equiv E_F V(x) - E_G V(x)$, $x \in [a,b]$. Then:

[1] The proof, as the proofs of the SD criteria, holds also for the unbounded case (see Hanoch and Levy, *Review of Economic Studies*, *36*, 1969, pp. 335–346).

$$\Delta = \int_a^b V_w(x)dF(x) - \int_a^b V_w(x)dG(x) = \int_a^b V_w(x)d[F(x)-G(x)]$$

Integrating by parts yields:

$\Delta = V_w(x)\left[F(x) - G(x)\right]\Big|_a^b - \int_a^b [F(x) - G(x)]\, V'_w(x)\, dx$. As a and b define the range of x,

we have F(b) = G(b) = 1, and F(a) = G(a) = 0.

Thus, the first term is equal to zero and we are left with:

$$\Delta = \int_a^b [G(x) - F(x)]\, V'_w(x)\, dx = \int_a^0 [G(x) - F(x)]\, V'_w(x)\, dx + \int_0^b [G(x) - F(x)]\, V'_w(x)\, dx \tag{15.3}$$

Integrating once again by parts each of the above two terms on the right-hand side of eq. (15.3) yields:

$$\Delta = V'_w(x)\int_a^x [G(t)-F(t)]dt\Big|_a^0 - \int_a^0 V''_w(x)\int_a^x [G(t)-F(t)dtdx$$
$$+V'_w(x)\int_0^x [G(t)-F(t)]dt\Big|_0^b - \int_0^b V''_w(x)\int_0^x [G(t)-F(t)]dtdx$$

Or:

$$\Delta = V'_w(0)\int_a^0 [G(t)-F(t)]dt - \int_a^0 V''_w(x)\int_a^x [G(t)-F(t)]dtdx \tag{15.4}$$
$$+ V'_w(b)\int_0^b [G(t)-F(t)]dt - \int_0^b V''_w(x)\int_0^x [G(t)-F(t)]dtdx \ .$$

For the range $0 < x \leq b$, $V''_w(x) \leq 0$; hence, if in this range $\int_0^x [G(t) - F(t)]dt \geq 0$ for all x, then the sum of the last two terms on the right-hand side of eq.(15.4) are non-negative (recall that $V'_w(b) \geq 0$). With regard to the range $a \leq x < 0$, $V''_w(x) \geq 0$, it is tempting to require that $\int_0^x [G(t) - F(t)]dt \leq 0$ in this range to insure that $\Delta \geq 0$. However, this would be an error because although this condition would guarantee a non-negative second term on the right-hand side of eq. (15.4), the first term on the right-hand side of eq. (15.4) may be negative with this condition, which will not guarantee that $\Delta \geq 0$. Therefore, to find the condition that guarantees non-negative Δ, we have to rewrite the second term on the right-hand side of eq. (15.4) as follows:

$$-\int_a^0 V''_w(x)\int_a^x [G(t)-F(t)]dtdx = \tag{15.5}$$
$$= -\int_a^0 V''_w(x)\int_a^0 [G(t)-F(t)]dtdx + \int_a^0 V''_w(x)(\int_x^0 [G(t)-F(t)]dtdx).$$

The first term on the right-hand side of eq. (15.5) can be rewritten as:

$$-\int_a^0 [G(t)-F(t)]dt \int_a^0 V''_w(x)dx = -V''_w(x)\Big|_a^0 (\int_a^0 [G(t)-F(t)]dt) \tag{15.6}$$
$$= -V'_w(0)\int_a^0 [G(t)-F(t)]dt + V'_w(a)\int_a^0 [G(t)-F(t)]dt \ .$$

Substituting the two terms of eq. (15.6) for the first term in eq. (15.5) and then collecting all these results eq. (15.4) becomes

$$\Delta = V'_w(a)\int_a^0 [G(t)-F(t)]dt + \int_a^0 V''_w(x)\int_x^0 [G(t)-F(t)]dtdx$$
$$+ V'_w(b)\int_0^b [G(t)-F(t)]dt - \int_0^b V''_w(x)\int_0^x [G(t)-F(t)]dtdx \ . \tag{15.7}$$

From eq. (15.7), we see that if:

$$\int_x^0 [G(t)-F(t)]dt \geq 0 \ \text{ for all } \ x \leq 0 \tag{15.8}$$

and

$$\int_0^x [G(t)-F(t)]dt \geq 0 \ \text{ for all } \ x \geq 0, \tag{15.9}$$

we obtain $\Delta \geq 0$, namely, F dominates G by PSD, or F dominates G for all S-shaped value functions (recall that $V''_w(x) \leq 0$ for $x > 0$ and $V''_w(x) \geq 0$ for $x < 0$ and $V'_w(b) \geq 0$, $V'_w(a) \geq 0$).

Finally, conditions (15.8) and (15.9) are equivalent to the condition:

$$\int_y^x [G(t)-F(t)]dt \geq 0 \ \text{ for } \textit{any} \text{ pair of x and y such that } y < 0 < x,$$

which completes the proof.

The necessity side of the proof is similar to the necessity proof of other SD rules, hence is omitted here.

Corollary 1: F dominates G by PSD if and only if F dominates G for all path-dependent utility functions $U^*(w, x)$ given by eq. (14.2).

The proof is straightforward. To see this, note that because $U^*(w, x) = U(w) + V_w(x)$, and $U^{*\prime}(w,x)=V_w'(x)$, $U^{*\prime\prime}(w,x) = V_w''(x)$ (all derivations are with respect to x). Thus, the PSD proof with $U^*(w, x)$ is the same as in Theorem 15.1: simply substitute U^* for V everywhere.

Discussion: Expected utility theory is defined on wealth (w + x) and not on change of wealth as in the value function $V_w(x)$ of prospect theory. Thus, it would seem that the results of Theorem 15.1 do not apply to utility theory. This is not the case, and the results of Theorem 15.1 hold for all S-shaped value functions $V_w(x)$ as well as for all utility functions of the form U(w+x) such that $U'' \geq 0$ for all $x < 0$ and $U'' < 0$ for $x > 0$. The magnitude of U'' may be a function of w exactly as the magnitude of V_w'' is a function of w. Similarly, in utility theory, F and G should be defined on terminal wealth (w + x) and not on x as in the proof of Theorem 15.1. However, because:

$$\int_y^x [G(t) - F(t)]dt > 0 \Leftrightarrow \int_{y-w}^{x-w} [G(w+t) - F(w+t)]dt > 0 \ ,$$

w can be ignored. From these two properties we can conclude the following:
Theorem 15.2: F dominates G by prospect theory (PSD) for all S-shaped value functions, if and only if F dominates G in the expected utility framework for all S-shaped utility functions of the form U(w+x), such that $U''(w+x) \geq 0$ for $x < 0$ and $U''(w+x) \leq 0$ for $x > 0$, and $U' > 0$ for all x.

The proof is straightforward: In Theorem 15.1, substitute U(w+x) for $V_w(x)$ and F(w+x), and G(w+x) for F(x) and G(x), respectively.

Thus, PSD with value function, V can be viewed as a special case of expected utility with an S-shaped utility function where the inflection point is at current wealth w, with risk seeking for $x < 0$ and risk aversion for $x > 0$.

Example:

Suppose that we have the following two alternative prospects, F and G:

G		F	
Return	Probability	Return	Probability
-5	½	-10	¼
0	¼	0	¼
2.5	¼	+10	¼
		+20	¼

As can be seen from this example, neither F nor G dominates the other by FSD (because the two cumulative distributions intersect). However, there is a PSD of F over G (for all S-shaped value functions and for all S-shaped utility functions). To see this, note that:

$$\int_{x}^{0}[G(t)-F(t)]dt \geq 0 \quad \text{for all } x \leq 0 \qquad (15.10)$$

and:

$$\int_{0}^{x}[G(t)-F(t)]dt \geq 0 \quad \text{for all } x \geq 0 \qquad (15.11)$$

(and there is at least one strict inequality). Alternatively, for any pair $y \leq 0$ $x \leq 0$, the integral $\int_{y}^{x}[G(t)-F(t)]dt \geq 0$, which implies that F dominates G by PSD.

It is interesting to note that in the above example, neither F nor G dominates the other by Second degree Stochastic Dominance (SSD). To see this, recall that F dominates G by SSD if and only if $\int_{a}^{x}(G(t)-F(t))dt \geq 0$ for all values x (with at least one strict inequality). This condition does not hold for $x < -5$. Hence, F does not dominate G by SSD. It is easy to show that G also does not dominate F by SSD (for $x = 10$, the integral $\int_{a}^{10}[F(t)-G(t)]dt < 0)$.

As expected, if F dominates G by FSD, then F dominates G by PSD. To see this, recall that FSD dominance implies that $F(w+x) \leq G(w+x)$ for all x, which is the same as $F(x) \leq G(x)$ for all x. However, because $V'(x) \geq 0$, by equation (15.3), if F dominates G by FSD, $\Delta \geq 0$ for all value functions V_w with $V'_w(\cdot) \geq 0$; hence, F dominates G by PSD. It is easy to show that SSD, unlike FSD, does not imply PSD.

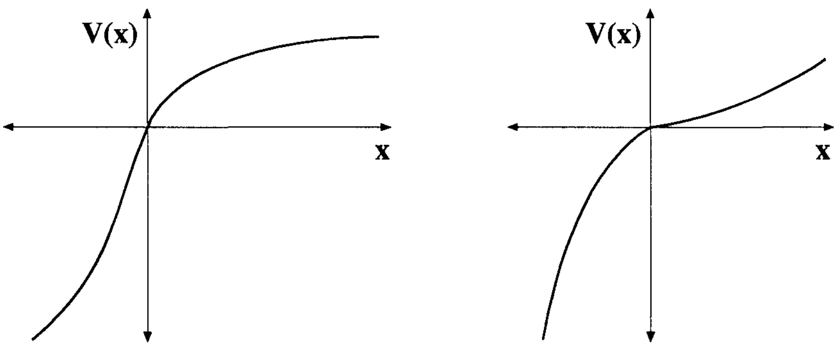
Finally, note that in the proof of PSD, we assume that F and G are the original distributions and we do not conduct transformations on the probabilities. However, if a transformation is carried out on the cumulative distribution (as advocated by CPT, in Tversky and Kahneman [1992[2]]) with $F^* = T(F)$ and $G^* = T(G)$, T being a monotonic non-decreasing transformation, then the PSD still holds but, this time, the condition is that $\int_{x}^{y}[G^*(t)-F^*(t)]dt \geq 0$ for all pairs $y < 0$ and $x > 0$. This is valid only if all investors carry out the same transformation. Thus, while FSD is intact for any monotone decision weight function, this is not the case with PSD.

[2]See footnote 8 in Chapter 14.

15.3 MARKOWITZ'S STOCHASTIC DOMINANCE

Figure 15.1 contrasts the S-shape preference of CPT and the reverse S-shape preference advocated by Markowitz[3]. Of course, both functions relate to a relatively small range [A,B] around zero (which corresponds to an experiment runs for several thousands of dollars) and the functions may have more inflection points outside the range (A,B).

Figure 15.1: The S-shape and reverse S-shape utility functions



We develop below a stochastic dominance rule corresponding to all preferences with a reverse S-shape. As this type of preference was suggested by Markowitz, we call it Markowitz Stochastic Dominance (MSD).

Theorem 15.3[4]: Let $V(x) \in V_M$ which is the class of all continuously and twice differentiable Markowitz utility functions such that $V' \geq 0$ for all x, with $V'' \leq 0$ for $x<0$ and $V'' \geq 0$ for $x>0$. Then F dominates G for all $V(x) \in V_M$ if and only if:

$$\int_a^x [G(t)-F(t)]dt \geq 0 \qquad \text{for all } x < 0 \qquad (15.12)$$

and

$$\int_x^b [G(t)-F(t)]dt \geq 0 \qquad \text{for all } x > 0 \qquad (15.13)$$

Proof: Let us first formulate our proof in terms of change of wealth, x, rather

---

[3]See Markowitz, H.M., "The Utility of Wealth," *Journal of Political Economy*, pp.151-156.

[4]The proof of MSD relies on the proof given in Levy, M. and Levy, H., "Prospect Theory: Much Ado About Nothing," *Management Science, 48*, 2002, pp. 1334–1349.

than total wealth, w+x, and then show that the dominance is invariant to the value w. As before, assume that the outcomes of Prospects F and G have lower and upper bounds a and b.

$$\Delta \equiv E_F V(x) - E_G V(x)$$

It has been shown before (see eq. (3.1) that Δ can be rewritten as,

$$\Delta = \int_a^b [G(x)-F(x)]\, V'(x)\, dx$$
$$= \int_a^0 [G(x)-F(x)]\, V'(x)\, dx + \int_0^b [G(x)-F(x)] V'(x) dx,$$

Integrating by parts, the two terms on the right-hand side yield:

$$\Delta = V'(x) \int_a^x [G(t)-F(t)]dt \Big|_a^0 - \int_a^0 V''(x) \int_a^x [G(t)-F(t)]dt\, dx$$
$$+ V'(x) \int_0^x [G(t)-F(t)]dt \Big|_0^b - \int_0^b V''(x) \int_0^x [G(t)-F(t)]dt\, dx$$

As some of the terms (i.e., the cases x = a, and x = 0) are equal to zero Δ can be rewritten as:

$$\Delta = V'(0) \int_a^0 [G(t)-F(t)]\, dt - \int_a^0 V''(x) \int_a^x [G(t)-F(t)]\, dt\, dx \qquad (15.14)$$
$$+ V'(b) \int_0^b [G(t)-F(t)]\, dt - \int_0^b V''(x) \int_0^x [G(t)-F(t)]\, dt\, dx.$$

Because $V' \geq 0$ and $V'' \leq 0$ for $x < 0$, the condition

$\int_a^x [G(t)-F(t)]dt \geq 0$ ensures that the first two terms on the right-hand side of

Δ are nonnegative. (Note that we assume that the utility function is twice differentiable, and that $V' \geq 0$ for all x. If the utility functions is not differentiable at a given point x_0, approximations can be used without altering the results).

The third term of eq. (15.14) is also positive. However, the fourth term may be negative. Therefore, conditions (15.12) and (15.13) do not guarantee that $\Delta > 0$, and some more algebraic manipulation in (15.14) are needed to show that indeed $\Delta > 0$.

Let us rewrite the fourth term as follows:

$$-\int_0^b V''(x)\int_0^x [G(t)-F(t)]\,dt\,dx$$

$$=-\int_0^b V''(x)\int_0^b [G(t)-F(t)]dt\,dx+\int_0^b V''(x)\int_x^b [G(t)-F(t)\,dt\,dx$$

$$=-\int_0^b [G(t)-F(t)]dt\int_0^b V''(x)dt+\int_0^b V''(x)\int_x^b [G(t)-F(t)]\,dt\,dx\ ,$$

$$=-\int_0^b [G(t)-F(t)]dt[V'(x)]_0^b+\int_0^b V''(x)\int_x^b [G(t)-F(t)]\,dt\,dx\ ,$$

$$=-V'\int_0^b [G(t)-F(t)]\,dt+V'(0)\int_0^b [G(t)-F(t)]dt$$

$$+\int_0^b V''(x)\int_x^b [G(t)-F(t)]\,dt\,dx\ .$$

Substituting these three terms instead of the fourth term on the right-hand side of Δ yields

$$\Delta=V'(0)\int_a^b [G(t)-F(t)]dt-\int_a^0 V''(x)\int_a^x [G(t)-F(t)]dt\,dx$$

$$+\int_0^b V''(x)\int_x^b [G(t)-F(t)]\,dt\,dx\ . \tag{15.15}$$

Because by the theorem conditions

$$\int_a^x [G(t)-F(t)]dt\geq 0 \text{ for } x<0 \text{ and} \tag{15.16}$$

and

$$\int_x^b [G(t)-F(t)]dt>0 \text{ for } x>0 \tag{15.17}$$

we can conclude that $\int_a^b [G(t)-F(t)]dt\geq 0$. Thus, the first term on the right-hand side of Δ is nonnegative. Because $V''\leq 0$ for $x<0$ and $V''\geq 0$ for $x>0$, the condition of the theorem guarantee that $\Delta \geq 0$.

Finally, note that if the utility function is $V(w + x)$ and the inflection point is at $x = 0$, the proof is kept unchanged because $F(w+x)$ and $G(w+x)$ are simply shifted to the right by w with no change in the area enclosed between F and G.

Necessity. It can be easily shown that if $\int_a^{x_0}[G(t)-F(t)-F]dt<0$ for some $x_0<0$, then there is some $V\in V_M$ for which $\Delta_0<0$. To show this, employ the same necessity proof of Hanoch and Levy (1969) for second-degree stochastic dominance. By a similar argument, one can show that $\int_x^b [G(t)-F(t)]\ dt > 0$ for $x > 0$ is also a necessary condition for MSD dominance.

Discussion

Under PSD we need that the accumulated area from zero to left and right should always be non-negative (eg. 15.8 and 15.9). With MSD we require that the accumulated area from the right-end point (point b) to any point $x \geq 0$, and the accumulated area from the left-end point (point a) to any point $x \leq 0$ will be non-negative (see 15.13) and (15.14). Thus, it seems that PSD and MSD have opposite requirements. Also because PSD assumes S-shape preference and MSD reverse S-shape preference, one is tempted to believe that if F dominates G by PSD, then G dominates F by MSD. The numerical example given in section (15.2) reveals that this is not the case: F dominates G by PSD, yet as simple calculations reveal, G does not dominate F by MSD.

As a necessary condition for PSD as well as for MSD is that the dominating prospect must have equal or larger mean than the inferior prospect, a necessary condition for F to dominate G by PSD and G to dominate F by MSD is that $E_F(x) = E_G(x)$. In the above example $E_F(x) > E_G(x)$, hence F may dominate G by one of these two rules but G cannot dominate F neither by PSD nor by MSD.

15.4 CPT, M-V AND THE CAPM

In the next section we provide experimental results regarding SSD, PSD and MSD, i.e., we test whether risk aversion, PT's S-shape or Markowitz's reverse S-shape preference is supported. In this section we assume that PT or CPT are valid and contrast them with the M-V analysis and the CAPM which implicitly or explicitly assume risk aversion.

Levy & Levy [2004] have shown that the M-V efficient set "almost" coincides with the PT's efficient set. This is summarized in the following Theorems taken from Levy & Levy [2004].[5]

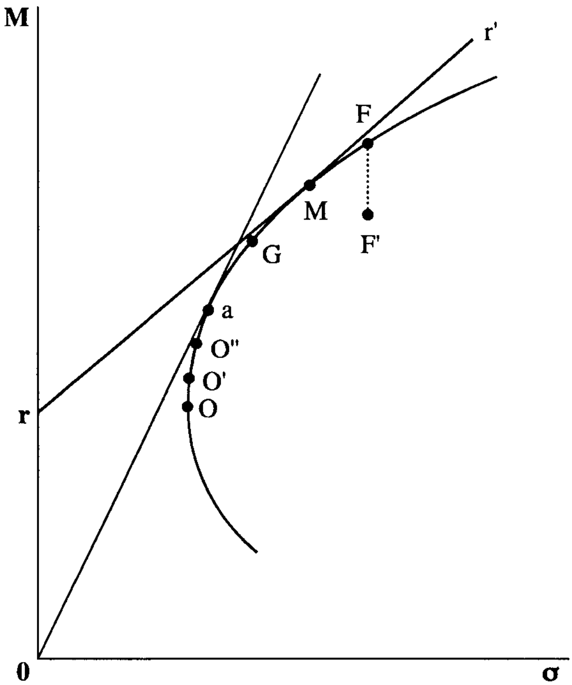
Theorem 15.4 Assume that distributions of return are normal, hence the M-V rule is optimal. Suppose that the objective probability distributions are employed. Then (i) the PSD-efficient set is a subset of the MV-efficient set, and (ii) the segment of the M-V efficient set which is excluded from the PSD-efficient set is at most the segment between the minimum variance portfolio and the point of tangency from the origin to the frontier (segment Oa in Figure 15.2).

While the formal proof is given in Levy & Levy [2004] the intuitive explanation is as follows: For any M-V efficient portfolio there is an efficient portfolio located

[5]See, Levy, H., and Levy, M., "Prospect Theory and Mean-Variance Analysis," *Review of Economic Studies, 17,* 2004, pp. 1015–1041.

vertically above it. For example, for portfolio F′ there is a portfolio F which dominates it by the M-V. However, as with normal distributions it implies that the cumulative distribution of F is located to the right of distribution F′ and F and F′ do not cross (as $(\sigma_{F'} = \sigma_F)$, we conclude that F dominates F′ by FSD. However as FSD ⇒ PSD, F dominates F′ also by PSD. Therefore, any M-V inefficient portfolio is also PSD inefficient. Levy & Levy [2004] show that the PSD efficient set may be a subset of the M-V efficient set as the segment oa (see Figure 15.2) may be PSD inefficient.

Figure 15.2: The M-V and PSD efficient sets



So far we assume PT's S-shape value function but employ objective distribution. Namely, F and F′ (see Figure 15.2) are assumed to be normal. Now let us introduce also decision weights which are a basic ingredient of CPT. The M-V and PSD relationship with decision weights is given in Theorem 15.5.

Theorem 15.5: Suppose that the objective distributions are normal. Furthermore, suppose that the objective probabilities are subjectively distorted (or decision weights are employed) by any transformation that does not violate FSD, for example, the Cumulative Prospect Theory transformation or RDEU's decision weights. Then, the PSD-efficient set is a subset of the MV-efficient set.

Proof. The proof that the PSD-efficient set is a subset of the MV-efficient frontier is as follows: No portfolio interior to the MV-efficient frontier, such as portfolio F' in Figure 15.2 can be PSD efficient, because it is FSD dominated by portfolio F on the frontier. As we are considering FSD-maintaining probability transformations, the FSD dominance of F over F' with the objective probabilities implies that F dominates F' for every individual with an increasing utility function, even if she subjectively distorts the probability distributions. Namely, if F dominates F' by FSD, then also T(F) dominates $T(F')$ by FSD. As FSD $\Rightarrow$ PSD, portfolio F' is dominated by portfolio F, also by PSD. Thus, the PSD-efficient set is a subset of the MV-efficient set.

A few remarks are called for:

1) Though the distributions of returns with objective distribution are assumed to be normal, when decision weights are incorporated, the new distributions are generally not normal and may be skewed. Yet, the PSD efficient set with skewness is a subset of the MV-efficient set.

2) The Efficient set with Markowitz's reverse S-shape is also a subset of Markowitz's M-V set. As in the proof of FSD we employ the fact that FSD $\Rightarrow$ PSD and as also FSD $\Rightarrow$ MSD, all the results are intact for any non-decreasing utility function including Markowitz's reverse S-shape function.

3) CPT and CAPM: One may add to Figure 15.2 a straight line rising from the risk-free interest rate r. With borrowing and lending for any portfolio below line rr' there is a portfolio on line rr' which dominates it by FSD. As FSD $\Rightarrow$ PSD and FSD $\Rightarrow$ MSD, all investors including $U \in U_S$ and $U \in U_M$, will invest in a mix of portfolio M (market portfolio) and the riskless asset.

Thus, we have a separation Theorem and the CAPM holds also with CPT and with Markowitz's preference. Levy, DeGiorgi and Hens [2003][6] show in another paper that the CAPM holds with CPT as long as borrowing is restricted. They also show[7] in another paper that with CPT *equilibrium* does not exist. They suggest another utility function similar to the S-shape function in the neighborhood of x = 0, but it is less linear than the CPT's utility function for very high stakes. With this utility function equilibrium is possible.

[6]Levy, H., E. DeGiorgi, and T. Hens., "Two Paradigms and Two Nobel Prizes in Economics: A Contradiction or Coexistencd." Working paper, 2003, Hebrew University.

[7]DeGiorgi, E., T. Hens and Levy, H., "Prospect Theory and the CAPM: a contradiction or Coexistence," 2003, working paper.

15.5 TESTING THE COMPETING THEORIES: SD APPROACH

Experimental studies have been employed to test various theories and in particular whether risk aversion prevails or S-shape function prevails. Until 2001 the common method was the Certainty Equivalent (CE) approach and in 2001, the Stochastic Dominance approach was introduced by Levy and Levy.[8] We will first discuss the CE and SD approaches and then summarize the experimental findings regarding the various competing theories.

a) Testing the Competing Theories: SD Approach

Suppose that a subject faces an uncertain prospect yielding \$500 and \$1500 with equal probability. She is asked what is the certain amount such that she is indifferent between getting the certain amount or getting the uncertain prospect. If this certain amount, X_{CE}, is more than \$1,000, we conclude that she has risk-seeking preference. If $X_{CE} = \$1,000$, she is risk neutral and if $X_{CE} < \$1,000$ she is a risk-averter. This methodology is the standard approach and has been used to study preferences. The CE approach has been used also by PT advocates. Using the CE approach has three main drawbacks:

i) Only uncertain prospect with only two outcomes is possible. Otherwise, no conclusion can be reached regarding the shape of the preference.

ii) The "Certainty Effect". The certainty equivalent approach, employed by Kahneman and Tversky and by many other researchers following their path, involves one prospect with an outcome which is certain. This is quite troubling, because a certain outcome has been documented to have a dramatic effect on subjects' choices. This is well known as the "certainty effect" (see, for example, Battalio et al., 1990[9], Tversky and Kahneman, 1981[10]). Schneider and Lopes (1986)[11] find support for the S-shape preference only when a prospect with a riskless component is involved.

iii) Another problematic aspect of this methodology is that it typically employs either positive prospects or negative prospects, but cannot be employed to the more general and realistic case of mixed prospects. To see this, suppose that one experimentally finds that subjects are indifferent between \$40 with certainty and a mixed prospect, which yields either –\$100 or \$200 with equal probability. From this observation, one cannot reach conclusions regarding the shape of the

---

[8]See, Levy, M., H. Levy, "Testing for risk aversion: A stochastic dominance approach," *Economic Letters*, 2001, pp.233-240.

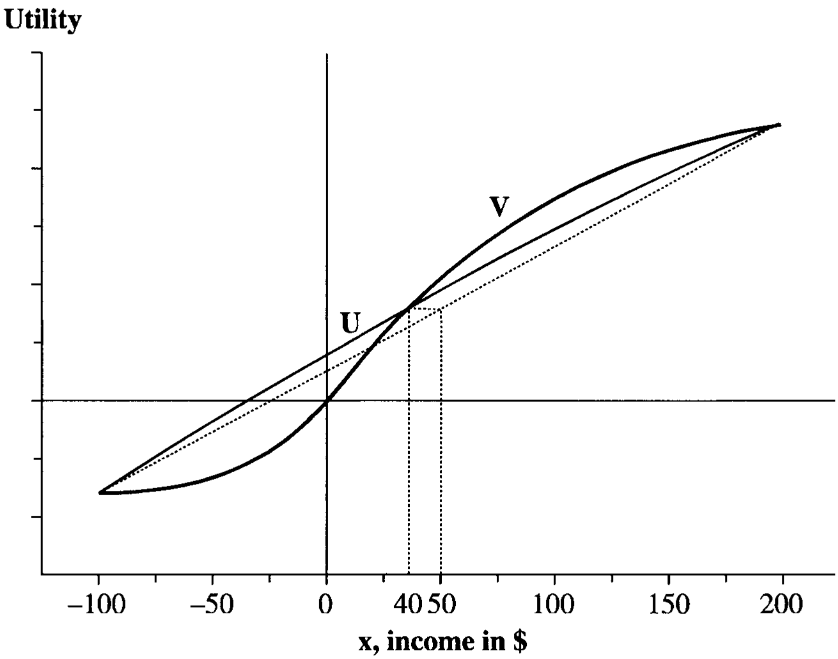
[9]Battalio, R.C., Kagel, H., & Jiranyakul, K., "Testing between alternative models of choice under uncertainty: some initial results." *Journal of Risk and Uncertainty, 3*, 1990, pp.25-50.

[10]Tversky, A., & Kahneman, D. "The framing of decisions and the psychology of choice." *Science, 211*, 1981, pp.453-480.

[11]Schneider, S.L., & Lopes, L.L., "Reflection in preferences under risk: Who and when may suggest why." *Journal of Experimental Psychology: Human Perception and Performance, 12*, 1986, pp.535-548.

preferences. Fig. 15.3 presents two possible preferences, which are consistent with this result. As can be seen, both a function which is concave in the whole range or a value function with a risk-seeking segment can fit this result. Thus, the CE approach is confined to either two negative outcomes or two positive outcomes but not to mixed outcomes. As in practice, the return on most investments are mixed (stocks, bonds, etc.) the CE approach does not conform with realistic choices in the capital market. The SD approach introduced by Levy & Levy 2001 (see footnote 8), overcomes all the three drawbacks of the CE approach.

Figure 15.3: The certainty equivalent with various utility functions



b) The Stochastic Dominance Approach

The SD approach to test preference has the following advantages over the CE approach: it can be applied to two uncertain prospects, hence the "certainty effect" is neutralized, and it can be applied to prospects of mixed outcomes. Hence, the superiority of SD approach over CE approach is overwhelming.

The basic idea is as follows: Suppose that the subject faces two uncertain options F and G. Furthermore, suppose that F dominates G, by say, SSD. If P% of the subjects choose G, we can conclude that *at least* P% are not risk averse. The reason is that FDG by SSD $\Leftrightarrow E_F U(x) \geq E_G U(x)$ for all $U \in U_2$ and as P% select G, at least P% of the subjects are not risk averters. We say *at least* because from those 1−P% who selected F, we may have some $U_0 \notin \mathbf{U}_2$ which justifies this selection; hence we may have more than P% who are not risk averters. Using this technique, we can test whether $U \in \mathbf{U}_2$, $U \in \mathbf{U}_s$ or $U \in \mathbf{U}_M$. To this we turn next.

c) Are People Risk Averse? (SSD)

The most common assumption in economic models is that people are risk averse. One can design an experiment to test whether risk aversion prevails. This has been done in several studies, and we report here the main results. The first study which employs SSD to test risk aversion has been conducted by Levy & Levy. [12] Let us elaborate. 194 subjects had to choose between F and G as follows:

F		G	
Gain or Loss	Probability	Gain or Loss	Probability
-500	¼	0	½
+500	¼		
+1000	¼	+1500	½
+2,000	¼		

Source: Levy & Levy (see footnote 12).

Simple calculation reveals that G dominates F by SSD. Yet 54% of the subjects selected F, hence *at least* 54% are not risk-averse. Similar results are obtained with students, practitioners (mutual fund managers) and with experiments with and without financial payoff. Thus, we conclude that there is evidence that a large proportion of people are not risk averse. As decision weights are not incorporated in our calculation, it is assumed here that large probabilities $p \geq 0.25$, are not distorted, at least not in the symmetrical case presented above.

d) Is CPT Valid Theory? (PSD)

As dominance by SD rules, including PSD rule, can be stated either in terms of change of wealth or total wealth, there are two remaining main factors to be tested: the S-shape value function and the reverse S-shape weighting function of CPT. There are several experimental studies, which rejected these two components of CPT. We provide here only one of them taken from Levy & Levy [2002].[13]

[12] See Levy, M. and Levy, H., "Testing for risk aversion: a stochastic dominance approach," *Economic Letters*," *71*, 2001, pp.233–240.

[13] Levy, H., and Levy, M., "Experimental test of the prospect theory value function: A stochastic dominance approach," *Organizational Behavior and Human Decision Processes*, (2002) *89*, pp. 1058–1081.

Table 15.1: The Two Tasks

Task I: Suppose that you decided to invest $10,000 either in Stock F or in Stock G. Which stock would you choose, *F* or *G*, when it is given that the *dollar gain or loss* one month from now will be as follows:

F		G	
Gain or Loss	Probability	Gain or Loss	Probability
-1600	¼	-1000	¼
-200	¼	-800	¼
1200	¼	800	¼
1600	¼	2000	¼

Please write F or G:

Task II: Suppose that you decided to invest $10,000 either in Stock F or in Stock G. Which stock would you choose, *F* or *G*, when it is given that the *dollar gain or loss* one month from now will be as follows:

F		G	
-875	.5	-1000	.4
2025	.5	1800	.6

Please write F or G:

Source: Levy & Levy (see footnote 13)

Table 15.1 provides the two tasks in the experiment with 84 subjects. In Task I we have equally likely outcomes with probability of 0.25 for each outcome, hence we ignore the decision weights on the assumption that probabilities are not distorted in such a case. As some may have other views and claim that even in such a case probabilities are distorted, in Task II we have a joint hypothesis of the CPT where both the S-shape value function as well as the reverse S-shape weighting function are tested simultaneously.

The decision weights of CPT in Task II are $w^-(0.5) = .454$, $w^+(0.5) = .421$, $w^-(0.4) = .392$ and $w^+(0.6) = .474$, respectively.[14] Drawing the cumulative distribution,

[14] Notice that for mixed prospects the decision weights do not generally add up to 1 (Tversky & Kahneman, 1992, p. 301). In Task II, the decision weights add up to .875 for *F* and to .866 for *G*. We assign the probability which is "missing" to the outcome 0, which of course does not affect the results by CPT V(0) = 0. If all outcomes are either positive or negative we obtain by CPT that $\sum w(p) = 1$. In such cases, experimental findings strongly reject CPT. For more details, see Levy, H., "Cumulative Prospect Theory: New Evidence," 2005, working paper, Hebrew University of Jerusalem.

one can easily find that in Task I, F dominates G by PSD, while in Task II, G dominates F by PSD, with and without decision weights. The results of this experiment are given in Table 15.2.

Table 15.2: The results of the Experiment[a]
(in %)

Task	F	G	Indifferent	Total
(N = 84)				
I	38	62	0	100
II	66	34	0	100

[a] Total number of subjects: 84. The numbers in the table are rounded to the nearest integer. In Task *F* dominates *G* by PSD. In Task II *G* dominates *F* by PSD.

Source: Levy & Levy [2002], see footnote 13.

As we see in Task I, 62% selected G and in Task II, 66% selected F, hence *at least* 62%–66% reject CPT. In Task II, it is a joint hypothesis, hence either the S-shape value function or the weighting function or both are rejected.

Other studies with only positive outcomes or only negative outcomes strongly reject CPT and support Markowitz's utility function as the option which dominates by MSD is selected by a high proportion of subjects.

15.6 SSD, PSD, MSD AND THE EFFICIENCY OF THE MARKET PORTFOLIO

When the investors make decisions by the M-V rule and the CAPM underlying assumptions are intact, it is predicted that the market portfolio, i.e., the mean rate of return and standard deviation of a portfolio diversified across all assets (where the investment proportions are equal to the market value proportions) will lie on the Capital Market Line (CML). Hence, passive mutual funds (index funds) that track the value weighted portfolio are expected to be M-V efficient. However, empirical evidence reveals that this is not the case and the market portfolio is highly M-V inefficient. Several factors can explain these inefficiency results.

a) The preference is not quadratic, hence the M-V rule is not intact.

b) Distributions of return are not normal, or risk aversion does not globally hold. Indeed, empirical studies detect skewness of returns, and there are some behavioral studies which claim that risk aversion does not globally hold (see above discussion).[15]

If the distributions of return are not solely a function of mean and variance, and in particular if skewness is an important factor in asset pricing, one would indeed expect that the market portfolio will be M-V inefficient. However, if risk aversion globally prevails and distributions are not symmetric, one would expect the market portfolio to be SSD efficient, though it is M-V inefficient.

In a recent paper, Post[16] employs Linear Programming (LP) procedure to test whether the market portfolio is SSD efficient relative to a benchmark portfolio. Post rejects the market portfolio SSD efficiency.[17] Pim extends Post's analysis and analyzes whether the value weighted CRSP total return index (a proxy to the market portfolio) is efficient.[18] He finds that indeed the market portfolio is M-V inefficient, but it is TSD efficient. By using TSD efficiency rather than SSD efficiency, the emphasis is shifted to skewness preference as TSD assumes that $U''' > 0$ and this, in turn, implies skewness preference.

The difference between the results of Post and the results of Pim may be attributed to the fact that Post's null hypothesis is restrictive as equal means are assumed, while in Pim's case, the null hypothesis, is that the market portfolio is TSD efficient, with no restrictions on the means. Also the different period covered in these two studies may induce a difference in the conclusions regarding market efficiency.

Post and Levy[19] test whether the market portfolio is SSD, PSD or MSD efficient. They test whether the value weighted market portfolio is efficient relative to benchmark portfolios formed by size, BE/ME (book value relative to to market

[15] For the importance of skewness see Levy, H., "A utility function depending on the first three moments," *Journal of Finance*, 1969, 24, pp. 715–719, Arditti, F.D., "Skewness and the required return on equity," *Journal of Finance*, 1967, 22, pp. 19-36, Kraus, A. and R.H. Litzenberger, "Skewness Preferences and the valuation of risk assets," *Journal of Finance*, 1976, 31, pp. 1085-1100, and Harvey, C., and A. Siddique," "Conditonal Skewness in asset pricing tests, *Journal of Finance*, 2000, 55, pp. 1263-1295.

[16] Post, Thierry, "Empirical test for stochastic dominance efficiency," *Journal of Finance*, 2003, 58, pp. 1905-1931.

[17] There are several working papers on this issue. Yitzhaki and Mayshar follow a distributional approach to SSD while Post employs necessary and sufficient utility function restriction. Both methods reduce to a linear programming problem. Bodurtha develops algorithms that efficiently identify improvements to dominated choices or preference functions choosing undominated choices. For more details on works dealing with this issue, see: Yitzhaki, S., and J. Mayshar, "Characterizing Efficient Portfolios," Hebrew University of Jerusalem, 1997, working paper, Kuosmanen, T., "Stochastic Dominance Efficiency Tests Under Diversification," Helsinki School of Economics, 2001, working paper, and Bodurtha, J.N., "Second-Order Dominance Dominated, Undominated and Optimal Portfolios, Georgetown University., 2003, working paper.

[18] Pim van Vliet, "Downside Risk and Empirical Asset Pricing ," Erim Ph.D. Series Research in Management, 49, 2004.

[19] Post, T. and Levy, H., "Does Risk Seeking Drive Stock Prices," *Review of Financial and Studies*, forthcoming.

value) and momentum. The first benchmark portfolio relies on Fama and French's[20] 25 portfolios constructed as the intersection of five quantile portfolios formed on size and five quantile portfolios formed on BE/ME. The second benchmark portfolio relies on 27 portfolios given in Carhart *et al.*[21] and used in Carhart.[22] The portfolios are formed based on size, BE/ME and mometum.

Post and Levy extended the LP procedure of Post to test whether the market portfolio is efficient relative to these benchmark portfolios.

The main conclusion of Post and Levy are:

a) They cannot reject the MSD efficiency of the market portfolio, indicating that the data conform with preference, which is reverse S-shape as suggested by Markowitz (see footnote 3). Namely, reverse S-shape preference may rationalize the market portfolio.

b) SSD and PSD are rejected, which implies that global risk aversion or S-shape preference suggested by Prospect Theory cannot rationalize the market portfolio.

c) Investors are risk averse for losses and risk seeking for gains; therefore, they are willing to pay premium for stock which provides downside protection in bear markets, and upside potential in bull markets.

Yet, Post & Levy implicitly assume that each historical observation has a probability of 1/n, n being the number of observations. If one uses the probability distortion formulas as suggested by Cumulative Prospect Theory (CPT) then the results of Post and Levy are also consistent with CPT as the probability distortion formula is a reverse S-shape. However, if the probabilities are not distorted in the uniform probability case, the CPT's S-shape preference cannot rationalize the market portfolios, and the empirical evidence supports Markowitz's reverse S-shape preference.

The experimental results discussed above, showing that CPT is rejected, i.e., the joint null hypothesis asserting that the preference is an S-shape and that probability distortion is a reverse S-shape, as suggested by CPT's formula, is rejected, casts doubt on one of the interpretations of Post & Levy asserting that their results are also consistent with CPT, and support their main conclusion asserting that the results conform with Markowitz's preference.

---

[20] Fama, E.F., and K.R. French, "The cross-section of expected stock returns," *Journal of Finance*, 1992, 47, pp. 427–465.

[21] Cahart, M.M., R.J. Krail, R.J. Stevens and K.E. Welch, "Testing the conditional CAPM, 1996, unpublished manuscript, University of Chicago.

[22] Cahart, M.M., "On the persistence in mutual fund performance," *Journal of Finance*, 1997, 52, pp. 57–82.

15.7 SUMMARY

Prospect Theory (PT) and Cumulative Prospect Theory (CPT) challenge EU theory. The experimental studies which support PT and CPT are obtained by employing the certainty equivalent (CE) approach which suffers from the "certainty effect." Moreover, these studies are confined to bets with only two outcomes, which must be either positive or negative, but not mixed.

In this chapter we suggest SD rules to test CPT, a paradigm which does not suffer from the above drawbacks of the CE approach. Prospect Stochastic Dominance (PSD) and Markowitz's Stochastic Dominance (MSD) corresponding to S-shape function and reverse S-shape function are presented. These decision rules are generally employed with cumulative distributions, F and G, derived from objective probabilities. However, they can be employed also with subjective cumulative distributions in some specific cases. Using these rules, CPT is rejected as 62%–66% of the subjects select, say, prospect G despite the fact that another prospect, say prospect F dominates it by PSD.

Key Terms

Prospect Theory (PT)

Cumulative Prospect Theory (CPT)

Decision Weights

Stochastic Dominance (SD)

Prospect Stochastic Dominance (PSD)

Markowitz Stochastic Dominance (MSD)

S-shape Value Function

Reverse S-shape Utility Function

Capital Asset Pricing Model (CAPM)

Certainty Equivalent (CE)

16

FUTURE RESEARCH

In this concluding chapter we propose and discuss directions of future research in the area of investment decision making under uncertainty. Most of these research suggestions are strictly related to the stochastic dominance (SD) paradigm, and some relate to SD, Mean-Variance (M-V) paradigms and Prospect Theory. Let us briefly discuss these research ideas.

16.1 PORTFOLIO CONSTRUCTION AND STOCHASTIC DOMINANCE EQUILIBRIUM

The three main advantages of the M-V rule over the stochastic dominance (SD) rules are:

i) Unlike the SD model, the M-V analysis provides a method to construct *a portfolio* from various *individual* securities.

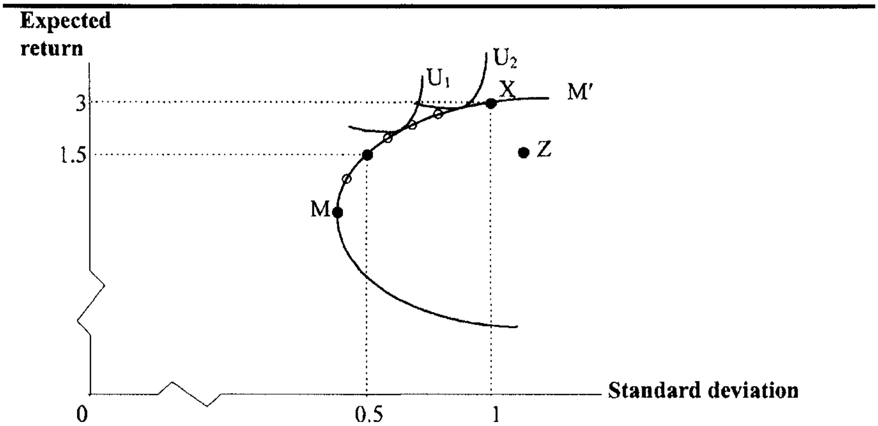
ii) Unlike the SD model, the M-V paradigm provides an equilibrium price, and the well-known CAPM risk-return linear relationship.

iii) It is relatively simple to apply M-V analysis to risky assets whereas in the case of SD, such applications are quite complex.

Let us elaborate on each of the above SD deficiencies and suggest further research that may help overcome them.

i) By SD criteria, we can tell whether one portfolio dominates the other but not how to construct a dominating portfolio from individual risky assets. In particular, we do not know how the various pair-wise correlations affect the SD efficient combination of assets. In the M-V space, we know that the efficient frontier includes all M-V efficient combinations of assets given by the segment MM′ (see Figure 16.1). Moreover, we know the asset combinations of all these efficient portfolios. It may seem that all these M-V efficient portfolios are also FSD and SSD efficient because they tangent to some indifference curve corresponding to a quadratic utility function, which, in turn, pertains to utility such as U_1 and U_2. However, this would be a false conclusion. To see this, consider a portfolio Y with return 1 and return 2 with equal probability and another portfolio X with return 2 and return 4 with equal probability. Strictly speaking, both X and Y will be in the M-V efficient set and, therefore, they will be located on the M-V efficient frontier. However, portfolio X dominates Y by FSD, as well as by SSD. Thus, Y is located in the M-V efficient set, in spite of the fact that it is FSD and SSD inefficient (see Figure 16.1).

Figure 16.1: The M-V and SD efficient set



From this example we see that portfolios can be M-V efficient but FSD and SSD inefficient.

Can one find FSD and SSD efficient diversification strategies? Is it possible for some *subset* (e.g., curve MM′ excluding portfolio such as Y) of the M-V efficient frontier to be included in the FSD or SSD efficient set? Indeed, the optimum rule for quadratic utility function can be used to eliminate portfolios such as Y, above. This rule advocates that portfolio X will dominate portfolio Y if both of the following two conditions hold:

1) $$E(x) \geq E(y)$$

2) $$[E(x) - E(y)]\,[2\cdot \text{Max}\,(x, y) - (E(x) + E(y))] - (\sigma_x^2 - \sigma_y^2) \geq 0$$

where x and y stand for the return on the two portfolios, X and Y, respectively.

(for more details see Hanoch & Levy 1970[1]).

In our specific example, condition 1) and condition 2) both hold because,

1) $$3 > 1\tfrac{1}{2} \quad \text{and,}$$

2) $$[3 - 1.5]\,[2\cdot 4 - (1.5 + 3)] - [1 - \tfrac{1}{4}] = 4.5 > 0.$$

[1] Hanoch, G. and H. Levy, "Efficient Portfolio Selection with Quadratic and Cubic Utility," *J. Business*, 43, 1976, pp. 181–189.

Scanning the M-V frontier with this rule, we can safely conclude that the portfolios left in the M-V efficient set are also FSD and SSD efficient because each of them maximizes the expected utility of some quadratic function, U_q, which pertains to $\mathbf{U}_1$ and to $\mathbf{U}_2$ (but not to $\mathbf{U}_3$). Thus, each of the efficient portfolios left (which together establish a subset of the M-V efficient set), maximizes the expected utility of some quadratic utility function. Hence, this M-V subset of efficient portfolios must be also included in the FSD and SSD efficient sets. Thus, a subset of the M-V efficient portfolio (e.g., the segment MM′ less the "wholes" depicted by circles, see Figure 16.1) are included in the FSD as well as SSD efficient set. This technique allows us to find some of the SD efficient diversification strategies but not all of them. The quadratic utility criterion represents one technique for identifying SSD efficient diversification strategies. The Mean-Gini criterion (see Yizhaki 1982)[2] can also be applied to identify SSD efficient portfolios. However, more research is needed to develop additional techniques for identifying additional SD efficient portfolios. Three research avenues are suggested:

a) The development of other rules, in addition to the quadratic utility criterion and the Mean-Gini criterion, to find SSD efficient diversification strategies (e.g., a technique to find all efficient investment strategies for well-known and commonly used utility functions, such as, x^{α}/α (for various α), $-e^{-\alpha x}$, etc.

b) Analysis of the size of the SSD efficient strategies obtained by applying rules suggested in a) above, relative to the "true" size of the SSD efficient diversification strategies. This can be attempted by a simulation or by assuming that some information is known: Note, however, finding the efficient strategies should not rely on this information.

c) Unlike the M-V efficient portfolio, by SD, some efficient portfolios may be located inside the M-V frontier. For example, portfolio Z may be SSD efficient but is not M-V efficient (see Figure 16.1). Thus, future research is called for to develop techniques to figure out how to locate more SD efficient combinations. Research in this direction may tell us how to add more asset combinations to the SD efficient sets and how they are related to the M-V efficient set.

ii) Another issue warranting more research is the elaboration of a risk-return return relationship that would be in line with the CAPM. Note that Rothchild and Stiglitz (1970)[3] and Kroll, Leshno, Levy and Spector (1995)[4] defined situations in which one asset is more risky than the other without quantifying risk. By analogy to the CAPM beta, we need to find a way to measure the "SSD risk"

---

[2] Yitzhaki, S., "Stochastic Dominance, Mean-Variance and Gini's Mean Difference," *Amer. Economic Rev.*, 72, 1982, pp. 178–185.

[3] Rothschild, M. and J.E. Stiglitz, "Increasing Risk. I. A Definition," *J. Economic Theory*, 3, 1970, pp. 66–84.

[4] Kroll, Y., Leshno, M., Levy, H. and Spector, Y., "Increasing Risk, Decreasing Absolute Risk Aversion and Diversification," *Journal of Math. Economics*, 24, 1995, pp. 537–556.

or "SSD contribution to portfolio risk". Then we need to price the risk and to establish the risk-return relationship in the SSD framework. This is not an easy task and assumptions regarding preferences or the random variables may be needed in order to obtain such results. It is doubtful whether a two-dimensional figure can be established in which SSD risk would be given on one axis and the expected return on the other axis, as in the CAPM framework. However, research in this direction should be attempted.

iii) Finally, constructing an efficient M-V portfolio is much simpler than constructing SSD efficient portfolios. A relatively simple algorithm exists for M-V portfolio selection but not for SSD selection. We need to establish a simple method that would tell us which asset to add and which to exclude in constructing the SSD efficient portfolio. The *marginal stochastic dominance* established by Shalit and Yitzhaki (1984)[5] is a step in this direction. However, much more research in this area is needed, in particular in simplifying the security selection procedure.

16.2 RISK ATTITUDE AND EQUILIBRIUM

Suppose that one assumes risk aversion and normal distribution of returns. Then, the CAPM follows, implying that the market portfolio with the actual market weights should be MV efficient. Virtually all empirical studies reveal that this is not the case and the market portfolio is MV inefficient. The inefficiency may be due to the fact that distributions are not normal or due to the lack of risk aversion.

One can employ SD criteria to test empirically whether the efficiency of the market portfolio can be rationalized by SSD, TSD, PSD or MSD. The advantage of these rules is that they do not rely on the normality assumption, hence higher moments, and, in particular, skewness are implicitly incorporated into the analysis. Moreover, if the market portfolio can be justified by PSD, it provides support for Prospect Theory's preference. On the other hand, if the market portfolio is rationalized by MSD, we have support for Markowitz's reverse S-shape preference.

Post[6][2003] was the first one to employ this approach. Several other studies[7] followed this line of research. We believe that this is a new research area with a lot of potential.

---

[5] Shalit, H. and Yitzhaki, S., "Marginal Conditional Stochastic Dominance," *Management Science*, 40, No.5, 1984.

[6] Post, Thierry, "Empirical Test for stochastic dominance efficiency," *Journal of Finance*, 2003, 58, pp. 1905-1931.

[7] See Kuosmanen, T., "Stochastic Dominance Efficiency Tests Under Diversification," Helsinki School of Economics, 2001, Bodurtha, J.N., "Second-Order Dominance Dominated, Undominated and Optimal Portfolios, Georgetown University., 2003, working paper, Pim van Vliet, "Downside Risk and Empirical Asset Pricing," Erim Ph.D. Series Research in Management, 49, 2004, and Post, T. and Levy, H., "Does Risk Seeking Drive Stock Prices," *Review of Financial Studies*, 2005.

16.3 THE STOCHASTIC DOMINANCE RULES AND LENGTH OF THE INVESTMENT HORIZON

M-V and SD focus mostly on one-period analysis. In other words, the holding period is assumed to be fixed and the analysis focuses on whether one portfolio dominates the other for this specific holding period. In practice, however, individual investors plan differential holding periods. Because the distributions of returns change with changes in the investment horizon, the differential planned holding period may affect the investment decision. Therefore, the size and the content of the efficient set may change with changes in the assumed investment horizon. An efficient portfolio for, say, a one-year holding period may be inefficient for, say, a two-year holding period. With regard to this issue, M-V and SD analysis can be extended in the following directions:

a) analysis of changes in the size and content of the efficient set with changes in the investment horizon: Tobin (1965)[8] analyzes the relationship between the M-V efficient set and the investment horizon and Levy (1973)[9] analyzes this issue in the SD framework (for FSD and SSD). However, much more has to be done in this SD framework (for FSD and SSD). However, much more has to be done in this area, in particular in light of the fact that the results obtained by Tobin contradict those obtained by Levy. Tobin shows that under the assumption of identical independent distributions over time (i.i.d.), the M-V efficient set increases with increase in the investment horizon whereas Levy shows that, with FSD and SSD efficient sets, the opposite holds. The main reason for this contradiction is that the M-V rule cannot be an optimal rule simultaneously for various horizons: if the one-period return x_1 is normally distributed and another one-period return x_2 is normally distributed, the product x_1x_2, a two-period return, will not be normally distributed. Thus, the normality assumption is violated rendering the M-V analysis wrong, and the SSD analysis remains intact because it is distribution-free. An important related research goal would be to seek out the distribution of x_1x_2 (where the distribution of x_1 and x_2 taken separately is normal) to establish an optimal investment rule for the distribution of x_1, x_2, and to compare it to the SD rules.

b) The investment horizon analysis could also be extended to include the riskless asset and, in particular, to incorporate information from the prevailing yield curve. To be more specific, suppose that one investor invests for holding period n=1 and another for holding period n=2. In addition, suppose that for n=1, the riskless interest rate is 5%, and for the longer period, n=2, the riskless interest rate is 12%. In such a case, it is not obvious how the one-period and two-period efficient sets are related. The efficient set,

---

[8] Tobin, J., "The Theory of Portfolio Selection," in F. Hahn and F. Brechling, *The Theory of Interest Rates* (MacMillan, New York).

[9] Levy, H., "Stochastic Dominance, Efficiency Criteria, and Efficient Portfolios: The Multi-Period Case," *Amer. Economic Rev.*, 63, 1973, pp. 986-994.

the investment horizon, and the shape of the yield curve are clearly related but the precise relationship has yet to be studied.

c) Practitioners and some academics commonly believe that optimal portfolio composition changes with the investor's age. To be more specific, young investors who have, say, 30 years left to retirement should invest a higher proportion of their assets in risky assets (stocks) relative to older investors who have, say, five years left to retirement. In other words, the claim is that optimal portfolio composition is a function of the assumed holding period, here number of years left to retirement. Does this approach imply that stocks become less risk in the long run (due to diversification across time) and, therefore, young investors can afford to buy a higher proportion of risky stocks than older investors? Not all would agree with such a claim. Merton and Samuelson (1974)[10], believe that at least for myopic utility functions, investment horizon (or investor age) is irrelevant in determining optimal portfolio composition.

As the horizon increases, both the mean and the variance of rates of return on stocks increase sharply relative to the increase in the mean and variance of rates of return on bonds (which are considered to be less risky than stocks). Hence, the rationale for the belief that stocks become less risky as the horizon increases is not clear. Any such claim is obviously not intact in the M-V paradigm. On the contrary, in the M-V paradigm, stocks become rather more risky. Much more research is needed on this issue focusing on non-myopic utility functions.

Another issue is related to the assets' performance. The performance measures, in particular, Sharpe's performance measures (1966)[11] and Treynor's performance measures (1965)[12] change with the horizon in some systematic way. (For more details, see Levy, 1972[13]; Levhari & Levy, 1977[14] and Hodges, Taylor and Yoder [1997])[15]. However, in the SD paradigm (i.e., the expected utility paradigm), stocks may be less risky as the horizon increases and even dominate risky bonds. Thus, research analyzing the relationship between M-V, SD, performance measures and risk as a function of the length of the investment horizon may be helpful in answering such questions. This issue is theoretically interesting and important in practice because investors may wish to change their investment policy as they grow older.

---

[10] Merton, R.C., and Samuelson, P.A., "Fallacy of the Log Normal Approximation to Optimal Portfolio Decision-Making Over Many Periods," *Journal of Financial Economics*, 1, 1974, pp. 67-94.

[11] Sharpe, W.F., "Mutual Fund Performance," *Journal of Business*, 39, 1966, pp. 119–38.

[12] Treynor, J.L., "How to Rate Management Investment Funds," *Harvard Business Review* (July-August 1965).

[13] Levy, H., "Portfolio Performance and the Investment Horizon," *Management Science*, 18, 1972, pp. 645–53.

[14] Levhari, D. and H. Levy, "The Capital Asset Pricing Model and the Investment Horizon," *Review of Economics and Statistics*, 59, 1977, pp. 92–104.

[15] Hodges, C.W., Taylor, W.R.L. and Yoder, J.A., "Stocks, Bonds, The Sharpe Ratio, and the Investment Horizon," *Financial Analyst Journal*, December 1997.

d) Bodie studied the question whether stocks become less or more risky as the investment horizon increases using the Black-Scholes option pricing model. According to Bodie (1995)[16], risk can be measured by the value of a put option and he shows that the longer the horizon, the larger the put value; hence, stocks become more risky as the horizon increases. However, other studies dispute this result. Levy and Cohen (1998)[17] show the following two results:

   i) Bodie's put value measure of risk does not generally increase with increase in the investment horizon; the behavior of the put value is a function of the selected striking price. Thus, even by Bodie's model, stocks are not necessarily more risky with increase in the investment horizon.

   ii) Risk is determined by the whole distribution (of terminal wealth) rather than its left tail. When the whole distribution is considered and lognormal distribution of terminal wealth is assumed, stocks become rather less risky as the investment horizon increases.

The dispute as to whether the stock-bond mix should be changed with increased age remains unsolved. Levy and Cohen show a clear-cut result supporting the practitioners' view but their results are limited to lognormal distribution. Research extending their analysis to other distributions would be of great importance.

The above discussion indicates that the M-V model is inappropriate for the analysis of the effect of the holding period on the efficient set because it provides an optimal rule for a single-holding period or for multi-period holding, but not for both (due to violation of the assumption of normality). Therefore, stochastic dominance rules, which do not rely on the assumption of normality, are more appropriate for analyzing the holding period issue. More research on the horizon effect, employing SD criteria, may help resolve the dispute regarding the role of the investor's age in determining optimal portfolio composition.

16.4 UNCERTAIN INVESTMENT HORIZON

Virtually all research in portfolio composition assumes that the holding period is known with certainty. The above M-V and SD studies discuss the possible effects of changes in this horizon on the size of the efficient set and the composition of efficient portfolios. In reality, the investment horizon is never certain. An investor may plan to invest for, say, two years. Yet, after one year, due to an emergency or other sudden need for money, the portfolio may have to be liquidated. Hence,

[16]Bodie, Zvi, "On the Risks of Stocks in the Long Run," *Financial Analyst Journal*, May/June 1995, pp. 18–22.

[17]Levy, H. and Cohen, A., "On the Risk of Stocks in the Long Run: Revisited," *The Journal of Portfolio Management*, Spring 1998, pp. 60–69.

investment decision rules and, in particular, stochastic dominance rules for uncertain investment horizons should be developed. To illustrate, suppose now that an investor intends to hold his/her portfolio for n years but knows in advance that there is a probability P_i that the portfolio will be liquidated due to sudden needs for cash after i years of investment, where $\sum_{i=1}^{n} P_i = 1$. M-V and SD efficient sets need to be developed for such scenarios. Next, the sensitivity of the resulting efficient set to changes in the probabilities P_i should be analyzed with a theoretical model for dominance corresponding to such a case. Then the relationship between the size of the efficient set corresponding to this case and Tobin's (1965) and Levy (1973)[18] efficient sets can be analyzed theoretically and empirically (or by simulation). Preliminary research in this area has been done by Cohen (1995)[19].

16.5 RISK INDEX

With a certain investment horizon, the M-V risk index is well defined, but in the SD framework such a risk index has yet to be developed. It would be of interest to find a risk index, both in the M-V and SD frameworks, corresponding to the uncertain investment horizon case. It should first be assumed that all investors face homogeneous uncertainty regarding the investment horizon and then heterogeneity should be allowed.

16.6 STOCHASTIC DOMINANCE AND INCREASING INTEREST RATE

Levy and Kroll developed stochastic dominance rules with a riskless asset (SDR). They also analyzed the case where the borrowing rate is higher than the lending rate. However, in practice, the borrowing rate may be an increasing function, $r_b = r(B)$, of the amount borrowed, B, namely $r'_b(B) > 0$. Stochastic dominance rules for such a case (as well as M-V rules) need to be developed. Moreover, SD rules are needed for a constant lending rate r_l and a borrowing rate r_b higher than r_l, where r_b is a function of B.

16.7 TRUNCATED DISTRIBUTIONS AND STOCHASTIC DOMINANCE

Stochastic dominance rules for specific distributions and, in particular, for normal and lognormal distributions are available in the literature. However, the latter distributions raise some problems: the normal distribution is inappropriate because *actual* rates of return are bounded by −100% whereas the normal distributions are unbounded. Lognormal distributions are inappropriate because a linear combination (a portfolio) of lognormal random variables will not be lognormally distributed. Thus, it can be assumed either that a portfolio return is

---

[18] Levy, H., "Stochastic Dominance, Efficiency Criteria, and Efficient Portfolios: The Multi-Period Case," *Amer. Economic Rev.*, 1993b, pp. 986–994.

[19] Cohen, A., "Portfolio Selection with Stochastic Investment Horizons," The Hebrew University of Jerusalem, October 1995.

lognormally distributed or that the individual asset return is lognormally distributed, but not that both are lognormally distributed. The lognormal and normal distribution limitations in portfolio analysis characterize both M-V analysis and SD analysis. However, Levy and Markowitz (1979)[20] and Kroll, Levy and Markowitz (1984)[21] have shown that if the range of returns is not "too large," the M-V rule can be used even if the normality or lognormality assumption is violated (see also H. Markowitz, Nobel Laureate speech, 1991)[22].

Direct research complementing the work of Levy and Markowitz and Kroll, Levy and Markowitz in the SD framework is called for. To be more specific, we suggest the following two areas of research.

a) The development of SD rules for a truncated normal distribution where the truncation is at rate of return −100% and more importantly, rules for a combination of such truncated distributions. Levy (1982)[23] has taken a first step in this direction, but much has still to be done, in particular, in analyzing combinations of truncated normal distributions.

b) Levy (1973) established the FSD and SSD rule for lognormal distributions. However, as noted above, a linear combination of lognormal distributions do not distribute logormally. Research in this area faces two interesting challenges: first, to analyze the portfolio distribution composed from lognormal random variables and then to find stochastic dominance rules for this unknown distribution. Secondly, if these two distributions are truncated such that the range of returns is not "too large," to ascertain whether the developed SD results are consistent with the work of Levy and Markowitz (1979) and Kroll, Levy, and Markowitz (1984).

16.8 EMPLOYING STOCHASTIC DOMINANCE CRITERIA IN OTHER RESEARCH AREAS

SD is superior to M-V when two distributions (portfolios) need to be compared; however, it is inferior to the M-V rule when efficient portfolios need to be constructed from individual assets. In many areas, outside portfolio construction, SD is superior to M-V. For example, suppose that a choice between two statistical estimators is needed (e.g., the mean and the median of a distribution). Then, the best estimator can be selected by applying SD rules to the monetary values induced by the selection of each of these two estimators. SD can be developed and used in many areas such as economics, medicine, statistics, and agriculture

[20]Levy, H., and Markowitz, H.M., "Approximating Expected Utility by a Function of Mean and Variance," *American Economic Review*, 69, 1979, pp. 308–17.

[21]Kroll, Y., Levy, H., and Markowitz, H., "Mean-Variance Versus Direct Utility Maximization," *The Journal of Finance*, March 1984, pp. 47–61.

[22] Markowitz, H.M., "Foundations of Portfolio Theory," *Journal of Finance*, Vol. 46, 2, 1991, pp. 469–478.

[23]Levy, H., "Stochastic Dominance Rules for Truncated Normal Distributions: A Note," *Journal of Finance*, 37, 1982, pp. 1299–1303.

where selection does not involve a combination of actions (see also Chapter 8). For example, in a recent paper, Stinnett and Mullahy (1997)[24] suggest that SD be employed rather than the conventional cost benefit ratio in evaluating various medical interventions.[25] We envisage more research in the application of SD rules to such problems (e.g., using medicine A or medicine B), and other problems that do not involve the construction of a portfolio from the individual assets.

16.9 REFINING THE STOCHASTIC DOMINANCE CRITERIA

One of the main criticisms of SD rules (as well as of the M-V rule) is that they may be unable to rank two investments even though it is obvious which one "most" investors would choose. To illustrate, suppose that we have two investments x and y as follows:

x		y	
return($)	**Probability**	**return($)**	**probability**
1	ε	1.1	½
10^6	$1-\varepsilon$	2	½

where $\varepsilon \rightarrow 0$.

By FSD, SSD and TSD (as well as M-V), neither x nor y dominates the other as long as $\varepsilon > 0$. However, no investor is likely to select y. The reason why SD and M-V cannot distinguish between x and y is that there may be a utility function U_0 where $U_0 \in \mathbf{U}_1$ (and which also belongs to $\mathbf{U}_2$ and $\mathbf{U}_3$) with a relatively very large utility weight on low income with zero marginal utility on a high income (for example, suppose that for $x \geq 2$, $U_0' = 0$). This pathological utility function is an obstacle in distinguishing between x and y by SD rules. We, therefore, need to define $\mathbf{U}_i^*$ for i=1, 2, 3 such that $\mathbf{U}_1^* \subset \mathbf{U}_1$, $\mathbf{U}_2^* \subset \mathbf{U}_2$ and $\mathbf{U}_3^* \subset \mathbf{U}_3$. The subset $\mathbf{U}_i^*$ includes all the utility functions except for the pathological utility functions. Thus, we suggest two tasks for future research:

a) Identification of the set of pathological utility function (i.e., to define $\mathbf{U}_i^*$).

b) Establishment of SD rules corresponding to $\mathbf{U}_i^*$.

[24] See Chapter 8, footnote 26.
[25] See also Leshno, M. and Levy, H., "Stochastic Dominance and Medical Decision Making," *Health Care Management Science*, 2004.

Leshno and Levy (2002)[26] have initiated research in this direction, but much more has to be done. In particular, we need to analyze and compare the U_i^* and the U_i efficient sets, to add riskless assets to these SD rules and to develop SDR rules corresponding to U_i^*.

16.10 STOCHASTIC DOMINANCE AND OPTION VALUATION

The value of a put and a call option is generally determined by the Black and Scholes formula, or its variations and extensions. However, this formula makes many assumptions, which clearly do not hold in reality. Alternative models to evaluate options which require fewer assumptions can be used, but they also yield price bounds rather than one equilibrium price. Levy (1985, 1988)[27] and Perrakis (1984, 1986)[28] use the SD paradigm to derive upper and lower bounds on option value. A. Levy (1988)[29] extends this analysis to include a portfolio of put, call, and the underlying stocks (for more details on these studies, see Chapter 8). More research in this area, allowing options, the market portfolio, and the underlying stocks to be held with and without the riskless asset could provide narrower bounds on the option value. This can be accomplished also by employing TSD with and without the riskless asset. Thus, the end result would be six bounds corresponding to FSD, SSD and TSD, each with and without the riskless asset.

16.11 EXPERIMENTAL STOCHASTIC DOMINANCE CRITERIA

Prospect Theory and Non-Expected utility models (see Chapter 14) were developed on the basis of the laboratory experimental finding that subjects tend to violate the expected utility paradigm. SD criteria were developed on the basis of the expected utility paradigm. An experimental study that directly tests the acceptance of SD criteria by investors would be of great interest. For example, if subjects are presented with investments F and G where F dominates G by, say, FSD, what proportion of the subjects would choose project F (i.e., not violate the FSD criterion)? And if violation is found, by how far should we increase the distance between F and G to prevent such violation? If we find, for example, that $F + a \leq G$ $(a > 0)$ must hold at least for some range to avoid such violations, we may define a new *experimental* FSD criterion that is inconsistent with expected utility predictions but consistent with investor behavior. By the same token, experimental studies can be conducted to examine whether subjects are able to distinguish dominance by SSD, TSD and PSD as well as the corresponding stochastic dominance criteria with the riskless asset.

---

[26]Leshno, M., and Levy, H., "Preferred by 'All' and Preferred by 'Most' Decision Makers: Almost Stochastic Dominance," *Management Science*, 2002. For more details, see Chapter 14.

[27]See Chapter 8, footnotes 11 and 12.

[28] See Chapter 8, footnotes 13 and 14.

[29]See Chapter 8, footnote 18.

BIBLIOGRAPHY

Abdellaoui, M., "Parameter Free Elicitation of Utility and Probability Weighting Functions," *Management Science*, 2000, 46, pp. 1497-1512.

Aboudi, R., and D. Thon, "Efficient Algorithms for Stochastic Dominance Tests Based on Financial Market Data," Working Paper, 1989.

Aboudi, R., and D. Thon, "Efficient Algorithms for Stochastic Dominance Tests Based on Financial Market Data," *Management Science*, Vol. 40, April 1994, 508–15.

Aboudi, R., and D. Thon, "Second-degree Stochastic Dominance Decisions and Random Initial Wealth with Applications to the Economics of Insurance," *Journal of Risk and Insurance*, Vol. 62, March 1995, 30–49.

Adler, Andre and Rosalsky, Andres, "Some General Strong Laws for Weighted Sums of Stochastically Dominated Random Variables," *Stochastic Analysis and Applications*, 5, 1987, 1–16.

Aharony, J. and M. Loeb, "Mean-Variance vs. Stochastic Dominance: Some Empirical FIndings on Efficient Sets," *J. Banking and Finance*, 1 1977, 92–102.

Alarie, Y., G. Dionne, and L. Eeckhoudt, "Increases in Risk and the Demand for Insurance," in G. Dionne, ed., *Contributions to Insurance Economics*, Boston: Kluwer Academic Publishers, 1992, 275–290.

Ali, M.M., "Stochastic Dominance and Portfolio Analysis," *J. Financial Economics*, 2, 1975, 205–229.

Allais, M., "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine," *Econometrica*, 21, 1953, 503–546.

Allais, M., "The Foundations of a Positive Theory of Choice Involving Risk and a Criticism of the Postulates and Axioms of the American School," *Utility Hypotheses and the Allais Paradox*, Allais, M. and M. Machina (Eds.), D. Reidel, Dordrecht, 1979.

Allais, M., "Allais Paradox," in *The New Palgrave: Utility and Probability*, The Macmillan Press, 1990, 3–9.

Angus, John E. and Herring, Susan, K., "A Smooth Test for First-Order Stochastic Dominance," *Journal of Applied Statistical Science*, 1, 1994, 299–311.

Any, J.S. and J.H. Chua, "Mutual Funds: Different Strokes for Different Folks?" *J. Portfolio Management*, 8, 1982, 43–47.

Arditti, F.D., "Rate and the Required Return on Equity," *Journal of Finance*, March 1967, pp. 19-36.

Arditti, F.D., "Skewness and the required return on equity," *Journal of Finance*, 1967, 22, pp. 19-36.

Arditti, F.D., "Another Look at Mutual Fund Performance," *J. Financial and Quantitative Anal.*, 6, 1971, 909–912.

Arditti, F.D., and Y. Peles, "Leverage Bankruptcy and the Cost of Capital," In *Natural Resources, Uncertainty and General Equilibrium Systems: Essays in Memory of Rafael Lusky*, A.S. Blinder and P. Friedman (Eds.). Academic Press, New York, 1977.

Arditti, F.D., and Levy, H., "A Model of Parallel Team Strategy in Product Development," *American Economic Review*, December 1980, 1089–1097.

Arditti, F.D., and Levy, H., and Sarnat, M., "Taxes, Capital Structure and Cost of Capital: Some Extensions," *The Quarterly Review of Economics and Business*, Summer 1977, 89–95.

Arrow, K.J., *Aspects of the Theory of Risk Bearing*, Yrjo Jahnssonin Saatio, Helsinki, 1965.

Arrow, K.J., *Essays in the Theory of Risk-Bearing*, Markham Publishing Company, Chicago, 1971.

Arrow, K.J., and M. Kurz, *Public Investment, the Rate of Return and Optimal Fiscal Policy*, John Hopkins Press, Baltimore, MD., 1970.

Ashton, D.J., "Stochastic Dominance and Mean-Variance Rules in the Selection of Risky Investments," *J. Business Finance and Accounting*, 9, 1982, 471–481.

Atkinson, A.B., "On the Measurement of Inequality," *J. Economic Theory*, 2, 1970, 244–263.

Barlev, B., and Levy, H., "Missue and Optimum Inspecting Strategy in Agency Problems," *Metroeconomica*, 1996, 82–104.

Barlev, B., H., W. Denny and H. Levy, "Using Accounting Data for Portfolio Management," *J. Portfolio Management*, 14, 1988, 70–77.

Baron, D.P., "Default Risk, Homemade Leverage and the Modigliani-Miller Theorem," *Amer. Economic Rev.*, 64, 1974, 176–82.

Baron, D.P., "Firm Valuation, Corporate Taxes, and Default Risk," *J. Finance*, 30, 1975, 1251–1264.

Battalio, R.C., Kagel, H., & Jiranyakul, K., "Testing between alternative models of choice under uncertainty: some initial results." *Jouurnal of Risk and Uncertainty*, 3, 1990, pp. 250-50.

Baumol, W.J., "An Expected Gain Confidence Limit Criterion for Portfolio Selection," *Management Science*, 10, 1963, 174–182.

Bawa, V.S., "Optimal Rules for Ordering Uncertain Prospects," *J. Financial Economics*, 2, 1975, 95–121.

Bawa, V.S., "Safety-First, Stochastic Dominance and Optimal Portfolio Choice," *J. Financial and Quantitative Anal.*, 13, 1978, 255–71.

Bawa, V.S., "Stochastic Dominance: A Research Bibliography," *Management Science*, 28, 1982, 698–712.

Bawa, V.S., E.B. Lindenberg and L.C. Rafsky, "An Efficient Algorithm to Determine Stochastic Dominance Admissible Set," *Management Science*, 25, 1979, 609–622.

Bawa, V.S., J. Bondurtha, M.R. Rao and H.L. Suri, "On Determination of Stochastic Dominance Optimal Set," *J. Finance*, 40, 1985, 417–431.

Ben-Horin, M., "Stochastic Dominance and Truncated Sample Data," *J. Financial Res.*, 13, 1990, 105–113.

Ben-Horin, M., and H. Levy, "Evaluating Estimators Using Stochastic Dominance Rules: The Variance of a Normal Distribution," *Comm. Statist.*, 1982.

Ben-Horin, M., and H. Levy, "Stochastic Dominance and Parameters Estimation: The Case of Symmetric Stable Distribution," *Insurance Math. and Economics*, 3, 1984, 133–138.

Benartzi, Shlomo and Richrd H. Thaler, "Myopic Loss Aversion and the Equity Premium Puzzle," *The Quarterly Journal of Economics*, 110(1), 1995, 73–92.

Bernstein, Peter L., "What Rate of Return Can You Reasonably Expect...or What Can the Long Run Tell Us about the Short Run," *The Financial Analysts Journal*, March/April 1997, 20–28.

Bey, R.P., "Estimating the Optimal Stochastic Dominance Efficient Set with Mean Semi-Variance Algorithms," *J. Financial and Quantitative Anal.*, 14, 1979, 1059–1070.

Bey, R.P., R.C. Burgess and R.B. Kearn, "Moving Stochastic Dominance: An Alternative Method for Testing Market Efficiency," *J. Financial Res.*, Vol. 7, 1984, 185–196.

Bey, R.P., and K.M. Howe, "A Gini's Mean Difference and Portfolio Selection. An Empirical Evaluation," *J. Financial and Quantitative Anal.*, 19, 1984, 329–338.

Bey, R.P., and R.B. Porter, "Portfolio Applications: Comparison of Selection Models, *Stochastic Dominance. An Approach to Decision-Making Under Risk*, Whitmore, G.A., and Findlay, M.C.(Eds.), D.C. Heath, Lexington, MA, 1978, 263–292.

Birnbaum, M.H., and Navarrete, J.B., "Testing Descriptive Utility Theories: Variations of Stochastic Dominance and Cumulative Independence," *Journal of Risk and Uncertainty*, 17, 1998, pp.49-78.

Birnbaum, M.H., "New Paradoxes of Risky Decision-Making," 2004, California State University, Fullerton, Working Paper.

Bishop, J. A., S. Chakraborti and P.D. Thistle, "Asymptotically Distribution-Free Statistical Inference for Generalized Lorenz Curves," *Rev. Economics and Statist.*, 71, 1989, 725–727.

Black, F., and Scholes, M., "The Princing of Options and Corporate Liabilities," *Journal of Political Economy*, May/June 1973, 637–654.

Black, J.M. and Bulkley, G., "A Ratio Criterion for Signing the Effect of an Increase in Uncertainty," *International Economic Review*, 30, 1989, 119–130.

Blackwell, D., "Equivalent Comparison of Experiments," *Ann. Math. Statist.*, 24, 1953, 264–272.

Blackwell, D., and M.A. Girshick, *Theory of Games and Statistical Decisions*, Wiley, New York, 1954.

Blayback, B.K. and G. Karaphillis, "A Selection Technique for Capital Investment Decisions: An Industrial Application of Stochastic Dominance," *Rev. Business and Economic Res.*, 18, 1983, 1–12.

Bodie, Zvi, "On the Risks of Stocks in the Long Run," *Financial Analysts Journal*, May/June 1995, 18–22.

Bodurtha, J.N., "Second-Order Dominance Dominated, Undominated and Optimal Portfolios, Georgetown University, 2003, working paper.

Booth, J.R., H. Tehranian and G.L. Trennepohl, "Efficiency Analysis and Option Portfolio Selection," *J. Financial and Quantitative Anal.*, 20 1985, 435–450.

Borch, V.S., "A Note on Uncertainty and Indifference Curves," *Rev. Economic Studies*, 36, 1969, 1–4.

Boyer, Marcel, Georges Dionne and Richard Khilstrom, "Insurance and the Value of Publicly Available Insurance," In Thomas B. Fomby and Tae Kun Seo (Eds.), *Studies in the Economics of Uncertainty. In Honor of Josef Hadar*, Springer Verlag, New York, 1989, 137–155.

Bradley, M.G. and D.E. Lehman, "Instrument Effects and Stochastic Dominance," *Insurance, Math. and Economics*, 7, 1988, 185–191.

Breeden, David T., "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Oppotunities," *Journal of Financial Econ.*, 7, 1979, 265–296.

Brennan, Michael and Alan Kraus, "Efficient Financing Under Asymmetric Information," *J. Finance*, 1987, 1225–1243.

Brooks, R., "Investment Decision-Making with Derivative Securities," *Financial Rev.*, 24, 1989, 511–527.

Brooks, R., and H. Levy, "Efficiency Analysis with Index Options," Working Paper, University of Florida, 1989.

Brooks, R., and H. Levy, "Portfolio Insurance: Does it Pay?" *Advances in Futures and Options Research*, 1993, JAI Press, 329–353.

Brooks, R., and J. Yoder, "Using Stochastic Dominance to Evaluate the Performance of Portfolios with Options," *Financial Analysts J.*, 43, 1987, 79–82.

Broske, Mary S. and Haim Levy, "The Stochastic Dominance Estimation of Default Probability," In Thomas B. Fomby and Tae Kun Seo (Eds.), *Studies in the Economics of Uncertainty. In Honor of Josef Hadar*, Springer Verlag, New York, 1989, 91–112.

Brumelle, S.L., "When Does diversification Between Two Investments Pay?" *J. Financial and Quantitative Anal.* 9, 1974, 473–482.

Buckley, J., "Stochastic Dominance: An Approach to Decision Making Under Risk," 6, 1986, 35–41.

Burgess, R.C., "Portfolio Applications: Comparison of Selection Models," *Stochastic Dominance. An Approach to Decision-Making Under Risk*, Whitmore, G.A., and M.C. Findley (Eds.), D.C. Heath, Lexington, MA, 1978, 163–195.

Butler, Kirk C., and Dale L. Domain, "Risk Diversification and the Investment Horizon," *The Journal of Portfolio Management*, Spring 1991, 41–48.

Cahart, M.M., "On the persistence in mutual fund performance," *Journal of Finance*, 1997, 52, pp. 57-82.

Cahart, M.M., R.J. Krail, R.J. Stevens and K.E. Welch, "Testing the conditional CAPM, 1996, unpublished manuscript, University of Chicago.

Cass, D. and J.E. Stiglitz, "The Structure of Investor Preferences and Assets Returns, and Separability in Portfolio Allocation: A Contribution to the Pure Theory of Mutual Funds," *J. Economic Theory*, 2, 1970, 122–160.

Chaing, R., "Some Results on Bond Yield and Default Probability," *Southern Economic J.*, 53, 1987, 1037–1051.

Cheng, H.C., M.J. Magill and W. Safer, "Some Results on Comparative Statics Under Uncertainty," *Int. Econ. Rev.*, 28, 1987, 493–507.

Cheung, C.S., C.C.Y. Kwan and P.C.Y. Yip, "The Hedging Effectiveness of Options and Futures: A Mean-Gini Approach," *J. Futures Markets*, 10, 1990, 61–73.

Chew, Soo Hong, "A Generalization of the quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Paradox Resolving the Allais Paradox," *Econometrica*, 51, 1983, 1065–1092.

Chew, Soo Hong, Edi Karni and Zvi Safra, "Risk Aversion in the Theory of Expected Utility with Rank Dependent Probabilities," *Journal of Economic Theory*, 42, 1987, 370–381.

Chew Soo Hong and Itzhak Zilcha, "Invariance of the Efficient Sets when the Expected Utility Hypothesis is Relaxed," *Journal of Economic Behavior and Organization*, 13, 1990, 125–131.

Choi, E.K., and S.R. Johnson, "Stochastic Dominance and Uncertain Price Prospects," *Southern Economic J.*, 55, 1989, 699–709.

Clarke, R.G., "Stochastic Dominance of Portfolio Insurance Strategies," Working Paper, Brigham Young University, 1988.

Cohen, Allon, "Portfolio Selection with Stochastic Investment Horizons," October 1995, MA Thesis, The Hebrew University of Jerusalem.

Constantinides, G.M., "Transaction Costs and the Volatility Implied by Option Prices," Working paper, January 1998, University of Chicago.

Constantinides, G.M. and T. Zariphopoulou, "Bounds on Prices of Contingent Claims in an Intertemporal Economy with Proportional Transaction Costs and General Preferences," Working paper, 1997a, University of Chicago and University of Wisonsin, Madison.

Constantinides, G.M. and T. Zariphopoulou, "Bounds on Option Prices in an Intertemporal Setting with Proportional Transaction Costs and Multiple Securities," Working paper, 1997b, University of Chicago and University of Wisconsin, Madison.

Cootner, P.H., (Ed.) *The Random Character of Stock Market Prices*, MIT Press, Cambridge, MA, 1967.

Cramer, H., *Mathematical methods of statistics*, (Princeton University Press, Princeton, NJ).

Cummins, J. David and Weis, Mary, A., "The Stochastic Dominance of No-fault Automobile Insurance," *The Journal of Risk and Insurance*, Vol. 60, June 93, 230–64.

Dasgupta, P., A. Sen and D. Starrett, "Notes on the Measurement of Inequality," *J. Economic Theory*, 6, 1973, 180–187.

Davis, M.H.A. and J.M.C. Clark, "Analysis of Financial Models Including Transaction Costs," *Royal Society Discussion Meeting in Mathematical Models in Finance*, London, November 10–11, 1993.

DeGiorgi, E., T. Hens and Levy, H., "Prospect Theory and the CAPM: a contradiction or Coexistence," 2003, working paper.

Demers, Fanny and Demers, Michel, "Increases in Risk and the Optimal Deductible," *The Journal of Risk and Insurance*, Vol. 58, Dec. 91, 670–99.

Deshpande, J.V. and H. Singh, "Testing for Second Order Stochastic Dominance," *Comm. Statist. Part A: Theory and Methods*, 14, 1985, 887–897.

Dexter, A.S., J.S. W. Yu and W.T. Ziemba, "Portfolio Selection in a Lognormal Market When the Investor has a Power Utility Function: Computational Results," *Stochastic Programming*, Dempster, M.A. (Ed.), Academic Press, London, 1980, 507–523.

Diamond, P.A. and J.E. Stiglitz, "Increases in Risk and Risk Aversion," *J. Economy Theory*, 8, 1974, 337–360.

Dickinson, J. P., "The Reliability of Estimation Procedures in Portfolio Analysis," *J. Financial and Quantitative Anal.*, 9, 1974, 447–462.

Dionne, G., L. Eeckhoudt and E. Briys, "Proportional Risk Aversion and Saving Behavior under Uncertainty," in H. Houbergé, ed., *Risk, Information and Insurance*, Boston: Kluwer Academic Publishers, 1989, 67–90.

Dionne, G., L. Eeckhoudt and C. Gollier, "Increases in Risk and Linear Payoffs." *International Economic Review*, 34, 1993, 309–319.

Doherty, Neil A. and Harris Schlesinger, "A Note on Risk Premiums with Random Initial Wealth," *Insurance Math. and Economics*, 5, 1986, 183–185.

Dybvig, P., "Distributional Analysis of Portfolio Analysis," *J. Financial and Quantitative Anal.*, 9, 1974, 447–462.

Eaton, P., "Stochastic Dominance in Two-Stage Problem with Linear Criteria," *Engineering Economist*, 33, 1988, 211–234.

Eden, B., "Stochastic Dominance in Human Capital," *J. Political Economics*, 87, 1980, 135–145.

Edwards, W., "Probability Preferences in Gambling," *American Journal of Psychology*, 66, 1953, pp. 349-364.

Edwards, W., "Probability Preferences Among Bets with Differing Expected Values," *American Journal of Psychology*, 67, 1954, pp. 56-67.

Eeckhoudt, L. and P. Hansen, "Minimum and Maximum Prices, in Certainty and the Theory of the Competitive Firm, *American Economic Review*, Vol. 70, 1980, 1064–1068.

Efron, Bradley and Gail Gong, "A Leisurely Look at the Bootstrap the Jackknife and Cross Validation," *The Amer. Statist.*, 1983, 36–48.

Ekern, S., "Time Dominance Efficiency Analysis," *J. Finance*, 36, 1981, 1023–1034.

Ellsberg, D., "Risk, Ambiguity, and the Savage Axioms," *Quart. J. Economics*, 75, 1961, 643–669.

Elton, E.J., and M.J. Gruber, "Portfolio Theory When Investment Relatives are Lognormally Distributed," *Journal of Finance*, 29, December 1974, 1265–73.

Epps, T.W., "Stochastic Dominance with Quadratic Utilities of Weak Condition," *J. Economics and Business*, 31, 1978, 70–73.

Epstein, L.G. and S.M. Tanny, "Increasing Generalized Correlation: A Definition and Some Consequences," Institute for Policy Analysis, Working Paper No. 7803, University of Toronto, 1978.

Evans, R. and R. Weinstein, "Ranking Occupations as Risky Income Prospects," *Industrial and Labor Rev.*, 35, 1982, 252–259.

Falk, H. and H. Levy, "Market Reaction to Quarterly Earnings' Announcements: A Stochastic Dominance Based Test of Market Efficiency," *Management Sci.*, 35, 1989, 425–446.

Fama, E.F. and J.D. MacBeth, "Risk, Return and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, June 1973, 605–36.

Fama, E.F., and K. R. French, " The cross-section of expected stock returns," *Journal of Finance*, 1992, 47, pp. 427-465.

Figlewski, S., "Options Arbitrage in Imperfect Markets," *Journal of Finance*, 44, 1989, 1289–1311.

Fishburn, P.C., *Decision and Value Theory*, Wiley, New York, 1964.

Fishburn, P.C., "Convex Stochastic Dominance with Continuous Distributions," *Journal of Economic Theory*, 1974, 7, pp. 143-158.

Fishburn, P.C., "Stochastic Dominance Without Transitive Preferences," *Management Science*, 24, 1978, 1268–1277.

Fishburn, P.C., "Stochastic Dominance and Moments Distributions," *Math. Oper. Res.*, 5, 1980, 94–100.

Fishburn, P.C., "Subjective Expected Utility: A Review of Normative Theories," *Theory and Decision*, 13, 1981, 139–199.

Fishburn, P.C., "Moment-Preserving Shifts and Stochastic Dominance," *Math. Oper. Res.*, 7, 1982a, 629–634.

Fishburn, P.C., "Nontransitive Measurable Utility," *J. Math. Psychology*, 26, 1982b, 31–67.

Fishburn, P.C., "SSB Utility Theory and Decision-Making Under Uncertainty," *Math. Social Sciences*, 8, 1984, 253–285.

Fishburn, P.C., "Foundations of Decision Analysis: Along the Way," *Management Science*, 1989a, 387–424.

Fishburn, P.C., "Stochastic Dominance in Nonlinear Utility Theory," in Thomas B. Fomby and Tae Kun Seo (Eds.), *Studies in the Economics of Uncertainty. In Honor of Josef Hadar*, Springer Verlag, New York, 1989b, 3–20.

Fishburn, P.C., "Interdependent Preferences," *Utility and Probability*, The Macmillan Press, United Kingdom, 1990a, 121–127.

Fishburn, P.C., "Representation of Preferences," *Utility and Probability*, The Macmillan Press, United Kingdom, 1990b, 217–226.

Fishburn, P.C., and R.B. Porter, "Optimal Portfolios with One Safe and One Risky Asset: Effects of Changes in Rates of Return and Risk," *Management Science*, 22, 1976, 1064–1073.

Fishburn, P.C., and R.G. Vickson, "Theoretical Foundations of Stochastic Dominance," *Stochastic Dominance. An Approach to Decision-Making Under Risk*, Whitmore, G.A. and Findlay, M.C. (Eds.), D.C. Heath, Lexington, MA, 1978, 39–114.

Fishburn, P.C., and R.D. Willig, "Transfer Principles in Income Redistribution," *J. Public Economics*, 25, 1984, 323–328.

Flesaker, B. and L.P. Hughston, "Contingent Claim Replication in Continuous Time with Transaction Costs." Working paper, Merrill Lynch International, 1994.

Franck, W.E., "A Likelihood Ratio Test for Stochastic Ordering," *J. Amer. Statist. Assoc.*, 70, 1984, 686–691.

Frankfurter, G.M. and H.E. Phillips, "Efficient Algorithms for Conducting Stochastic Dominance Tests on Large Numbers of Portfolios, A Comment," *J. Financial and Quantitative Anal.*, 10, 1975, 177–179.

Friedman, Milton and Leonard J. Savage, "The Utility Analysis of Choices Involving Risk, *The Journal of Political Economy*, 56(4), 1948, 279–304.

Gan, C.R.B. Wharton and T.P. Zacharias, "Risk-Efficiency Analysis of Export Earnings. An Application of Stochastic Dominance," *Amer. J. Agricultural Economics*, 70, 1988, 1204.

Garrone, François and Solnik, Bruno, "A Global Approach to Money Management," *The Journal of Portfolio Management*, Summer 1976, 5–14.

Gavish, B., "A Relaxation Algorithm for Building Undominated Portfolios," *J. Banking and Finance*, 1, 1977, 143–50.

Ghandhi, D.K. and A. Saunders, "The Superiority of Stochastic Dominance over Mean-Variance Efficiency Criteria: Some Clarifications," *J. Business Finance and Accounting*, 8, 51–59.

Gilboa, Itzhak, "Expected Utility with Purely Subjective Non-Additive Probabilities, *Journal of Mathematical Economics*, 16, 1987, 65–88.

Gollier, C., "The Comparative Statics of An Increase in Risk Revisited," (Mimeo, Groupe HEC, Department of Finance, F–78351, Jouy-en-Josas CEDEX, France.

Gould, J.P., "Risk Stochastic Preferences and the Value of Information," *J. Economic Theory*, 8, 1974, 64–84.

Grannan, E.R. and G.H. Swindle, "Minimizing Transaction Costs of Option Hedging Strategies," *Mathematical Finance*, 6, 1996, 341–364.

Gunthorpe, Deborah and Haim Levy, "Portfolio Composition and the Investment Horizon," *Financial Analysts Journal*, January-February 1994, 51–56.

Hadar, J. and W. Russell, "Rules for Ordering Uncertain Prospects," *Amer. Economic Rev.*, 59, 1969, 25–34.

Hadar, J. and W. Russell, "Stochastic Dominance and Diversification," *J. Economic Theory*, 3, 1971, 288–305.

Hadar, J. and W. Russell, "Diversification of Interdependent Prospects," *J. Economic Theory*, 7, 1974, 231–40.

Hadar, J. and W. Russell, "Applications in Economic Theory Analysis," *Stochastic Dominance. An Approach to Decision-Making Under Risk*, Whitmore, G.A., and Findlay, M.C. (Eds.), D.C. Heath, Lexington, MA, 1978, 39–114.

Hadar, J., W. Russell and K. Seo, "Gain from Diversification," *Rev. Economic Studies*, 44, 1977, 363–368.

Hadar, J., and T. K. Seo, "Stochastic Dominance and the Case for Diversification," *Research in Finance*, H. Levy (Ed.), 2, 1980, 99–100.

Hadar, J., and T. K. Seo, "Asset Proportions in Optimal Portfolios," *Rev. Economic Studies*, 55, 1988, 459–468.

Hadar, J., and T. K. Seo, "The Effects of Shifts in a Return Distribution on Optimal Portfolio, *Internat. Economic Rev.*, 31, 1990, 721–736.

Hammond, J.S., "Simplifying the Choice Between Uncertain Prospects: Where Preference is Non-Linear," *Management Science*, 20, 1974, 1046–1072.

Hanoch, G. and H. Levy, "The Efficiency Analysis of Choices Involving Risk," *Rev. Economic Studies*, 36, 1969, 335–346.

Hanoch, G. and H. Levy, "Efficient Portfolio Selection with Quadratic and Cubic Utility," *J. Business*, 43, 1970, 181–189.

Hardy, G.H., J.E. Littlewood and G. Polya, *Inequalities*, University Press, Cambridge, MA, 1959.

Harlow, W.V., "Asset Allocation in a Downside Risk Framework," *Financial Analysts Journal*, September/October 1991, 28–40.

Harvey, C., and A. Siddique, "Conditional Skewness in asset pricing tests," *Journal of Finance*, 2000, 55, pp. 1263-1295.

Harris, Thomas R. and H.P. Mapp, "A Stochastic Dominance Comparison of Water-Conserving Irrigation Strategies," *Amer. J. Agricultural Economics*, 68, 1986, 298–305.

Henin, C. and W.F. Rentz, "Call Purchases, Stock Purchases, and Subjective Stochastic Dominance," *J. Business Finance and Accounting*, 11, 1984, 127–138.

Henin, C. and W.F. Rentz, "Subjective Stochastic Dominance, Put Writing, and Stock Purchases with Extension to Option Pricing and Portfolio Composition," *Management Science*, 31, 1985, 919–927.

Hodges, C. W., Walton, R.L.T., and Yoder, J.A., "Stocks, Bonds, the Sharpe Ratio, and the Investment Horizon," *Financial Analysts Journal*, November/December 1997, 74–80.

Hogan, W.W. and J.M. Warren, "Computation of the Efficient Boundary in the E-S Portfolio Selection Model, *J. Financial and Quantitative Anal.*, 7, 1972, 1881–1896.

Horen, J. and S.Y. Wu, "Vertical Transactions under Uncertainty," In Thomas B. Fomby and Tae Kun Seo (Eds.), *Studies in the Economics of Uncertainty. In Honor of Josef Hadar*, Springer Verlag, New York, 1989, 157–186.

Huang, C.C., I. Vertinsky and W.T. Ziemba, "On Multiperiod Stochastic Dominance," *J. Financial and Quantitative Anal.*, 13, 1978, 1–13.

Hughes, J.S., "Agency Theory and Stochastic Dominance," *J. Financial and Quantitative Anal.*, Vol. 17, 1982, 341–361.

Ingersoll, J.E., *Theory of Financial Decisions Making*, Rowman & Littlefield Publishers, Totowa, NJ, 1987.

Itagaki, Takao, "Optimal Tariffs and Quotas under Uncertain International Transfer," In Thomas B. Fomby and Tae Kun Seo (Eds.), *Studies in the Economics of Uncertainty. In Honor of Josef Hadar*, Springer Verlag, New York, 1989, 187–199.

Jarrow, R., "The Relationship Between Arbitrage and First Order Stochastic Dominance," *J. Finance*, 41, 1986, 915–921.

Jean, W.H., "Comparison of Moment and Stochastic Dominance Ranking Methods," *J. Financial and Quantitative Anal.*, 10, 1975, 151–161.

Jean, W.H., "The Geometric Mean and Stochastic Dominance," *J. Finance*, 35, 1980, 151–158.

Jean, W.H., "The Harmonic Mean and Other Necessary Conditions for Stochastic Dominance," *J. Finance*, 39, 1984, 527–534.

Jean, W.H., and B.P. Helms, "Geometric Mean Approximations," *J. Financial and Quantitative Anal.* 18, 1983, 287–293.

Jean, W.H., and B.P. Helms, "Stochastic Dominance as a Decision Model," *Quart. J. Business and Economics*, 25, 1986, 65–101.

Jean, W.H., and B.P. Helms, "The Identification of Stochastic Dominance Efficient Set by Moment Combination Ordering," *J. Banking and Finance*, 12, 1988a, 243–253.

Jean, W.H., and B.P. Helms, "Moment Orderings and Stochastic Dominance Tests," *J. Business Finance and Accounting*, 15, 1988b, 573–584.

Jewitt, I., "A Note on Comparative Statistics and Stochastic Dominance," *J. Math. Economics*, 15, 1986, 249–254.

Johnson, K.H., and R.C. Burgess, "The Effect of Sample Size on the Accuracy of EV and SSD Efficiency Criteria," *J. Financial and Quantitative Anal.*, 10, 1975, 813–820.

Jorion, Philippe, Value at Risk, Mc-Graw Hill, New York 1997.

Joy, O.M. and R.B. Porter, "Stochastic Dominance and Mutual Fund Performance," *J. Financial and Quantitative Anal.*, 9, 1974, 24–31.

Kahneman, D. and A. Tversky, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 1979, 263–291.

Kallberg, J.G. and W.T. Ziemba, "Remarks on Optimal Portfolio Selection," *Methods Oper. Res.*, 44, 1981, 507–520.

Kallberg, J.G. and W.T. Ziemba, "Comparison of Alternative Utility Function in Portfolio Selection Problems," *Management Science*, 29, 1983, 1257–1276.

Karamata, J., "Sur une inegalité relative aux fonctions convexes," *Publ. Math. Univ. de Belgrade*, 1, 1932, 145–148.

Katzner, Donald W., "The 'Comparative Statics' of the Shackle-Vickers Approach to Decision-Making in Ignorance," In Thomas B. Fomby and Tae Kun Seo (Eds.), *Studies in the Economics of Uncertainty. In Honor or Josef Hadar*, Springer Verlag, New York, 1989, 21–43.

Kearns, R.B. and R.C. Burgess, "On Effectiveness Algorithms for Estimating Stochastic Dominance Efficient Set," *J. Financial and Quantitative Anal.*, 14, 1979, 547–552.

Keppe, H.J. and M. Weber, "Stochastic Dominance with Incomplete Information on Probabilities," *European J. Oper. Res.*, 43, 1989, 350–355.

Kira, D. and W.T. Ziemba, "Equivalence Among Alternative Portfolio Selection Criteria," *Financial Decision Making Under Uncertainty*, H. Levy and M. Sarnat (Eds.), Academic Press, New York, 1977, 151–161.

Kira, D. and W.T. Ziemba, "The Demand for Risky Assets," *Management Science.*, 26, 1980, 1158–1165.

Kraus, A., and R.H. Litzenberger, "Skewness Preferences and the valuation of risk assets," *Journal of Finance*, 1976, 31, pp. 1085-1100.

Kroll, Y., "Efficiency Analysis of Deductible Insurance Policies," *Insurance, Math. and Economics*, 2, 1983, 119–137.

Kroll, Y., "The Analysis of Risky Investment: A State-Contingency Approach," *J. Banking and Finance*, 8, 1984, 509–524.

Kroll, Y., and H. Levy, "Stochastic Dominance with a Riskless Asset: An Imperfect Market," *J. Financial and Quantitative Anal.*, 14, 1979, 179–204.

Kroll, Y., and H. Levy, "Sampling Errors and Portfolio Efficient Analysis," *J. Financial and Quantitative Anal.*, 15, 1980a, 655–688.

Kroll, Y., and H. Levy, "Stochastic Dominance Criteria: A Review and Some New Evidence," *Review in Finance*, Vol. 2, JAP Press, Greenwich, CT, 1980b, 163–227.

Kroll, Y., and H. Levy, "A Parametric Approach to Stochastic Dominance: The Lognormal Case," *Management Science*, 32, 1986, 283–288.

Kroll, Y., and H. Levy, "Increasing Risk and Decreasing Risk Aversion," Working Paper, Hebrew University, 1988.

Kroll, Y., and H. Levy, "Investment, Capital Structure and Cost of Capital Revisited," In Thomas B. Fomby and Tae Kun Seo (Eds.), *Studies in the Economics of Uncertainty. In Honor of Josef Hadar*, Springer Verlag, New York, 1989, 201–224.

Kroll, Y., H. Levy, and H. Markowitz, "Mean-Variance Versus Direct Utility Maximization," *J. Finance*, March 1984, 47–61.

Kroll, Y., Leshno, M., Levy, H. and Spector, Y., "Increasing Risk, Decreasing Absolute Risk Aversion and Diversification," *Journal of Mathematical Economics*, 24, 1975, 537–556.

Kuosmanen, T., "Stochastic Dominance Efficiency Tests Under Diversification," Helsinki School of Economics, 2001, working paper.

Laffont, J.J., "Risk, Stochastic Preferences and Value of Information: A Comment," *J. Economic Theory*, 12, 1976, 483–487.

Lahneman, Daniel and Amos Tversky, "Prospect Theory of Decisions Under Risk," *Econometrica*, 47(2), 1979, 263–291.

Landsberger, M. and Meilijson, I., "Production and Portfolio Decisions Under Uncertainty," (Mimeo, Department of Economics, Haifa University, Israel), 1988.

Landsberger, M. and Meilijson, I., "Demand for Risky Financial Assets: A Portfolio Analysis," *J. Economic Theory*, 50, 1990, 204–213.

Landsberger, M. and Meilijson, I., "A Tale of two Tails: An Alternative Characterization of Comparative Risk," *Journal of Risk and Uncertainty*, 3, 1990a, 65–82.

Landsberger, M. and Meilijson, I., "Lotteries, Insurance and Star-Shaped Utility Functions," *Journal of Economic Theory*, 52, 1990b, 1–17.

Landsberger, M. and Meilijson, I., "Mean-Preserving Portfolio Dominance," *Review of Economic Studies*, 60, 1993, 479–485.

Latane, H.A., "Criteria for Choice Among Risky Ventures," *J. Political Economy*, 67, 1959, 144–155.

Lebreton, M., "Stochastic Dominance: A Bibliographical Rectification and Restatement of Whitmore's Theorem," *Math. Social Sci.*, 13, 1987, 73–79.

Lee, Y.W., "Diversification and time: Do investment horizons matter?, *Journal of Portfolio Management*, 16 (3), 1990, 21–26.

Lee, J.B, R.D. Lacewell, and J.R. Ellis, "Evaluation of Production and Financial Risk. A Stochastic Dominance Approach," *Canada Journal of Agricultural Economics*, 35, 1987, 109–126.

Lee, W.Y. and R.K.S. Rao, "Mean Lower Partial Moment Valuation and Lognormally Distributed Returns," *Management Science*, 34, 1988, 446–453.

Lehmann, E., "Ordered Families of Distributions," *Ann. Math. Statist.* 26, 1955, 399–419.

Leibowitz, Martin L., and Stanley Kogelman, "Asset Allocation under Shortfall Constraints," *The Journal of Portfolio Management*, Winter 1991, 18–23.

Leshno, M. and H. Levy," "Preferred by All" and Preferred by "Most" Decision Makers: Almost Stochastic Dominance," *Management Science*, 2002, 48, pp. 1074-1085.

Leshno, M., Levy, H., and Spector, Y., "A Comment on Rothschild and Stiglitz's 'Increasing Risk: I. A Definition,'" *Journal of Economic Theory*, Vol. 77, No. 1, 1997, 223–228.

Levhari, D. and H. Levy, "The Capital Asset Pricing Model and the Investment Horizon," *Rev. Economics and Statist.* 59, 1977, 92–104.

Levhari, D., J. Paroush and B. Peleg, "Efficiency Analysis of Multivariate Distributions," *Rev. Economic Studies*, 42, 1975, 87–91.

Levhari, D. and T. N. Srinivasan, "Optimal Saving Under Uncertainty," *Rev. Economic Studies*, 36, 1969, 153–164.

Levy, A., "Option Equilibrium in an Incomplete Market with Risk Aversion," Working Paper, Bank of Israel and Hebrew University, 1988.

Levy, H., "A utility function depending on the first three moments," *Journal of Finance*, 1969, 24, pp. 715-719.

Levy, H., "Portfolio Performance and the Investment Horizon, *Management Science*, 18, 1972, 645–653.

Levy, H., "The Demand for Assets Under Condition of Risk," *Journal of Finance*, March 1973, 79–96.

Levy, H., "Stochastic Dominance Among Log-Normal Prospects," *Internat. Economic Review*, 14, 1973a, 601–614.

Levy, H., "Stochastic Dominance, Efficiency Criteria, and Efficient Portfolios: The Multi-Period Case," *Amer. Economic Review*, 63, 1973b, 986–994.

Levy, H., "Multi-Period Consumption Decision Under Conditions of Uncertainty," *Management Science*, 22, 1976, 1258–1267.

Levy, H., "The Definition of Risk: An Extension," *J. Economic Theory*, 14, 1977, 232–234.

Levy, H., "Multi-Period, Stochastic Dominance with One-Period Parameters, Liquidity Preference and Equilibrium in the Log-Normal Case," In *Natural Resources, Uncertainty and General Equilibrium Systems. Essays in Memory of Rafael Lusky*, A.S. Blinder and P. Friedman (Eds.), Academic Press, New York, 1979a.

Levy, H., "Does Diversification Always Pay?" In Elton, E.J., and Gruber, M.J. (Eds.), *Portfolio Theory, 25 Years After. Essays in Honor of Harry Markowitz, TIMS Studies in the Management Sciences*, 11, 1979b.

Levy, H., "Stochastic Dominance Rules for Truncated Normal Distributions: A Note," *Journal of Finance*, 37, 1982, 1299–1303.

Levy, H., "Upper and Lower Bounds of Put and Call Option Value: Stochastic Dominance Approach," *Journal of Finance*, 40, 1985, 1197–1217.

Levy, H., "Option Valuation Bounds: A Comparative Analysis," *Studies in Banking and Finance*, 5, 1988, 199–220.

Levy, H., "Stochastic Dominance," *Utility and Probability*, The Macmillan Press, United Kingdom, 1990, 251–254.

Levy, H., "The Mean-Coefficient of Variation Rule: The Lognormal Case," *Management Science*, June 1991, 745–747.

Levy, H., "Absolute and Relative Risk Aversion: An Experimental Study, *Journal of Risk and Uncertainty*, 8, 1994, 289–302.

Levy, H., "Investment Diversification and Investment Specialization and the Assumed Holding Period, *Applied Mathematical Finance*, 3, 1996, 117–134.

Levy, H., "Cumulative Prospect Theory: New Evidence," 2005, working paper, Hebrew University of Jerusalem.

Levy, H., "First degree Stochastic Dominance Violations: Decision Weights and Bounded Rationality," 2005, Working Paper, The Hebrew University of Jerusalem.

Levy, H., and Cohen, Allon, "On the Risk of Stocks in the Long Run: Revisited," *the Journal of Portfolio Management*, forthcoming, Spring 1998, 60–69.

Levy, H., and D. Gunthorpe, "Optimal Investment Proportions in Senior Securities and Equities Under Alternative Holding Periods, *Journal of Portfolio Management* 19(4), 1993, 30–36.

Levy, H., and R. Brooks, "An Empirical Analysis of Term Premiums Using Stochastic Dominance," *J. Banking and Finance*, 12, 1989, 245–260.

Levy, H., and G. Hanoch, "Relative Effectiveness of Efficiency Criteria for Portfolio Selection," *J. Financial and Quantitative Anal.* 5, 1970, 63–76.

Levy, H., and Y. Kroll, "Stochastic Dominance with Riskless Assets," *J. Financial and Quantitative Anal.*, 11, 1976, 743–773.

Levy, H., and Y. Kroll, "Ordering Uncertain Options with Borrowing and Lending," *Journal of Finance*, 33, 1978, 552–573.

Levy, H., and Y. Kroll, "Efficiency Analysis with Borrowing and Lending: Criteria and Their Effectiveness," *Rev. Economics and Statistics*, 61, 1979a, 25–30.

Levy, H., and Y. Kroll, "Investment Decision Rules, Diversification and Investor's Initial Wealth," *Econometrica*, 41, 1979b, 125–130.

Levy, H., and Y. Kroll, "Sampling Errors and Portfolio Efficiency Analysis," *Journal of Financial and Quantitative Anal.*, 15, 1980a, 655–688.

Levy, H., and Y. Kroll, "Sample vs. Population Mean-Variance, Efficient Portfolios," *Management Science*, 26, 1980b, 1108–1116.

Levy, H., and Z. Lerman, "Testing P/E Ratios Filters with Stochastic Dominance," *Journal of Portfolio Management*, 11, 1988, 241–254.

Levy, H., and Z. Lerman, "Testing the Predictive Power of Ex-Post Efficient Portfolios," *Journal of Financial Research*, 11, 1988, 241–254.

Levy, H., and A. Levy, "Stochastic Dominance and the Investment Horizon with Riskless Asset," *Rev. Economic Studies*, 49, 1982, 427–438.

Levy, H., and A. Levy, "Multivariate Decision Making," *Journal of Economic Theory*, 32, 1984a, 36–51.

Levy, H., and A. Levy, "Ordering Uncertain Options Under Inflation: A Note," *Journal of Finance*, 39, 1984b, 1223–1229.

Levy, H., and A. Levy, "Option Valuation: An Extension of the Binomial Model," *Advances in Futures and Options*, Vol. 5, Fabozzi, F.J. (Ed.), JAI Press, London, England, 1991, 49–69.

Levy, H., and C. Ofir, "New Product Screening via the Intention to Buy Scale," *Decision Science*, 17, 1986, 65–78.

Levy, H., and J. Paroush, "Multiperiod Stochastic Dominance," *Management Science*, 21, 1974, 428–435.

Levy, H., and J. Paroush, "Toward Multivariate Efficiency Criteria," *Journal of Economic Theory*, 7, 1975, 129–142.

Levy, H., and M. Sarnat, "Portfolio Selection and Investors' Utility: A Graphical Analysis," *Applied Economics*, 2, 1970a, 113–119.

Levy, H., and M. Sarnat, "Alternative Efficiency Criteria: an Empirical Analysis," *Journal of Finance*, 25, 1970b, 1153–1158.

Levy, H., and M. Sarnat, "A Note on Portfolio Selection and Investor's Wealth," *Journal of Financial and Quantitative Analysis*, 6, 1971a, 639–642.

Levy, H., and M. Sarnat, "Two-Period Portfolio Selection and Investor's Discount Rates," *Journal of Finance*, 26, 1971b, 757–761.

Levy, H., and M. Sarnat, "Investment Performance in an Imperfect Securities Market," *Financial Analysts Journal*, 28, 1972a, 77–81.

Levy, H., and M. Sarnat, "Safety First—An Expected Utility Principle," *Journal of Financial and Quantitative Analysis*, 7, 1972b, 1829–1834.

Levy, H., and M. Sarnat, *Portfolio and Investment Selection: Theory and Practice*, Prentice Hall International, Englewood Cliffs, NJ, 1984.

Levy, H., and P.A. Samuelson, "The Capital Asset Pricing Model with Diverse Holding Period," *Management Science*, 38, 1992, 1529–1542.

Levy, H., and Schwarz, G., "Correlation and the Time Interval Over which the Variables are Measured," *Journal of Econometrics*, 76, 1997, 341–350.

Levy, H., and Wiener, Z., "Prospect Theory and Utility Theory: Temporary and Permanent Attitude Toward Risk," Working paper, The Hebrew University of Jerusalem, 1997.

Levy, H., and Wiener, Z., "Stochastic Dominance and Prospect Dominance with Subjective Weighting Functions," *Journal of Risk and Uncertainty*, 16, 1998, pp. 147-163.

Levy, H. and M. Levy, "Experimental Test of the Prospect Theory Value Function: A Stochastic Dominance Approach," *Organizational Behavior and Human Decision Processes*, 89, 2002, pp. 1058-1081.

Levy, H. and Levy, M., "Prospect Theory and Mean-Variance Analysis," *Review of Economic Studies*, 17, 2004, pp. 1015-1041.

Levy, H., E. DeGiorgi, and T. Hens, "Two Paradigms and Two Nobel Prizes in Economics: A Contradiction or Coexistence," Working paper, 2003, Hebrew University.

Levy, H., Leshno, M., and Hechet, Y., "Third Degree Stochastic Dominance: An Algorithm and Empirical Study," 2004, working paper, Hebrew University, Jerusalem, Israel.

Levy, M. and Levy, H., "Testing for risk aversion: a stochastic dominance approach," *Economic Letters*," 71, 2001, pp. 233-240.

Levy, M. and Levy, H., "Prospect Theory: Much Ado About Nothing," *Management Science*, 48, 2002, pp. 1334-1349.

Levy, H., G.A. Whitmore and M. Ben-Horim, "Selection of Estimators for Portfolio Construction," Working Paper, Hebrew University, 1989.

Levy, H., M. Leshno and B. Leibovitch, "Bounded Preferences, Paradoxes and Decision Rules," Hebrew University (2005) working paper.

Levy, M., "Is Stochastic Dominance Efficient Set Really Efficient? A Joint Stochastic Dominance Analysis," Hebrew University, Jan. 2004, working paper.

Levy, Moshe, Haim Levy and Sorin Solomon, "A Microscopic Model of the Stock Market: Cycles, Booms, and Crashes," *Economics Letters*, 45, 1994, 103–111.

Li, Yuming and William T. Ziemba, "Rules for Diversification for All Risk Averters," *J. Economics and Business*, 1990, 165–170.

Lichtenstein, S. and P. Slovic, "Reversals of Preferences between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology*, 89, 1971, 46–55.

Lichtenstein, S. and P. Slovic, "Response-Induced Reversals of Preference in Gambling: An Extended Replication in Las Vegas," *J. Experimental Psychology*, 101, 1973, 16–20.

Litner, J., "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Rev. Economics and Statist*, 47, 1965, 13–37.

Lo, Andres, W. and A. Craig MacKinlay, "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test," *Review of Financial Studies*, 1, 1988, 41–66.

Luce, R. Duncan and Peter C. Fishburn, "Rank-and-Sign-Dependent Linear Utility Models for Finite First-Order Gambles," *Journal of Risk and Uncertainty*, 4, 1991, 29–59.

Machina, Mark , " 'Expected Utility' Analysis without Independent Axiom," *Econometrica*, 50, 1982, 277–323.

Machina, M. , "Generalized Expected Utility Analysis and the Nature of Observed Violations of the Independence Axiom, in Stigum, B., and Weinstøph, F. (eds.) *Foundation of Utility and Risk with Applications*, Reidel, Dordrecht, Holland, 1983.

Machina, Mark, "Choice Under Uncertainty: Problems Solved and Unsolved," *Journal of Economic Perspectives*, 1987.

Machina, Mark., and Michael Rothschild, "Risk," *Utility and Probability*, The Macmillan Press, United Kingdom, 1990, 227–239.

Markowitz, H.M., "Portfolio Selection," *Journal of Finance*, 7, 1952a, 77–91.

Markowitz, H.M., "The Utility of Wealth," *Journal of Political Economy*, 60, 1952b, 151–156.

Markowitz, H.M., *Portfolio Selection*, John Wiley & Sons, New York, 1959.

Markowitz, H.M., "Investment for the Long Run: New Evidence for an Old Rule," *Journal of Finance*, 31, 1976, 1273–1286.

Markowitz, H.M., "An Algorithm for Finding Undominated Portfolios," In *Financial Decisions Making Under Uncertainty*, H. Levy and M. Sarnat (Eds.), Academic Press, New York, 1977, 3–10.

Markowitz, H.M., *Mean Variance Analysis in Portfolio Choice and Capital Markets*, Basil Blackwell, New York, 1987.

Marshall, A.W. and I. Olkin, *Inequalities: Theory of Majorization and its Applications*, Academic Press, London, 1979.

Masse, P., *Optimal Investment Decision Rules for Action and Criteria for Choice*, Prentice Hall, Englewood Cliffs, NJ, 1962.

May, K.O., "Instransitivity, Utility, and the Aggregation of Preference Patterns," *Econometrica*, 22, 1954, 1–13.

McCaul, B.A., J.B. Hastie, T.D. Knight and J.R. Wilson, "Stochastic Dominance over Potential Portfolios. Caution Regarding Covariance," *Amer. J. Agricultural Economics*, 69, 1987, 1095.

McFadden, Daniel, "Testing for Stochastic Dominance," In Thomas B. Fomby and Tae Kun Seo (Eds.), *Studies in the Economics of Uncertainty. In Honor of Josef Hadar*, Springer Verlag, New York, 1989, 113–134.

Menezes, C.C. Geiss and J. Tressler, "Increasing Downside Risk," *Amer. Economic Rev.*, 70, 1980, 921–932.

Merril, Craig and Steven Thorley, "Time Diversification: Perspectives from Option Pricing Theory," *Financial Analysts Journal*, May/June 1996, 13–19.

Merril, Craig and Steven Thorley, "Time Diversification and Option Pricing Theory: Another Perspective: Response," *Journal of Portfolio Management*, Summer 1997, 61–63.

Merton, Robert C., "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science*, 1973, 141–183.

Merton, Robert C., and Paul A. Samuelson, "Fallacy of the Log-Normal Approximation to Portfolio Decision-Making Over Many Periods," *Journal of Financial Economics*, May 1974, 67–94.

Meyer, J., "Increasing Risk," *Journal of Economic Theory*, 11, 1975, 119–132.

Meyer, J., "Further Application of Stochastic Dominance to Mutual Fund Performance," *Journal of Financial and Quantitative Analysis*, 12, 1977a, 235–242.

Meyer, J., "Choice Among Distributions," *Journal of Economic Theory*, 14, 1977b, 326–336.

Meyer, J., "Second Degree Stochastic Dominance with Respect to a Function," *International Economic Review*, 18, 1979, 477–487.

Meyer, J., "Stochastic Dominance and Transformations of Random Variables," In Thomas B. Fomby and Tae Kun Seo (Eds.), *Studies in the Economics of Uncertainty. In Honor of Josef Hadar*, Springer Verlag, New York, 1989, 45–57.

Meyer, J., and M.B. Ormiston, "The Comparative Statics of Cumulative Distribution Function Changes for the Class of Risk Averse Agents," *Journal of Economic Theory*, 31, October 1983, 153–169.

Meyer, J., and M.B. Ormiston, "Strong Increases in Risk and Their Comparative Statics," Vol. 26, No. 2, June 1985, 425–437.

Mikesell, C.L., J.R. Williams and J.H. Long, "Conservation Tillage of Grain-Sorghum and Soybeans. A Stochastic Dominance Analysis," *American Journal of Agricultural Economics*, 69, 1987, 1095.

Miller, M., "Debt and Taxes," *Journal of Finance*, 1977, 261–75.

Modigliani, F. and M.H. Miller, "Reply to Heins and Sprenkle," *American Economic Review*, 59, 1969, 592–595.

Moseler, K., "Statistical Decisions Between Portfolios: Mean Variance Analysis Revisited," *Engineering and Process Economics*, 4, 1979, 257–268.

Moseler, K., "Stochastic Dominance Decision Rules When the Attributes are Utility Independent," *Management Science*, 30, 1984, 1311–1322.

Moseler, K., and Scarsini, Marco, "Some Theory of Stochastic Dominance," *Stochastic Orders and Decision Under Risk*, 1991, 261–284.

Mossin, J., *Theory of Financial Markets* (Englewood Cliffs: Prentice Hall, 1973).

Mosteller, F. , and Nogee, P., "An Experimental Measurement of Utility," *Journal of Political Economy,* 59, October 1951, pp. 371-404.

Nelson, Ray D. and Rulon D. Pope, "Bootstrapped Insights into Empirical Applications of Stochastic Dominance," *Management Science*, 37, 1991, 1182–1194.

Ng, M.C., "A remark on third stochastic dominance," *Management Science*, 46, 2000, pp. 870-873.

Ohlson, J.A. and W.T. Ziemba, "Portfolio Selection in a Lognormal Market When the Investor has a Power Utility Function," *Journal of Financial and Quantitative Analysis*, 11, 1976, 57–71.

Okuney, J., "A Comparative Study of the Gini's Mean Difference and Mean-Variance in Portfolio Analysis," *Accounting and Finance*, 28, 1988, 1–15.

Oldenkamp, Bart and Ton C.F. Vorst, "Time Diversification and Option Pricing Theory: Another Perspective," *Journal of Portfolio Management*, Summer 1997, 56–60.

Ormiston, M.B. and E. Schee, "Comparative Statics Under Uncertainty for a Class of Economic Agents," *Journal of Economic Theory*, vol. 61, 1993, 412–422.

Paroush, J., "A Note on Relative Efficiency and Risk Aversion," *Journal of Banking and Finance*, 5, 1981, 277–280.

Pearman, A.D. and Z.W. Kimietowicz, "Stochastic Dominance with Linear Partial Information," *Journal of Operational Research*, 23, 1986, 57–63.

Perrakis, S., "Option Pricing Bounds in Discrete Time: Extensions and the Price of the American Put," *Journal of Business*, 59, 1986, 119–142.

Perrakis, S., "Pricing and Replication of Short-Lived Index Options Under Transaction Costs," Working paper, University of Ottawa, 1997.

Perrakis, S., and P. Ryan, "Option Pricing Bounds in Discrete Time," *Journal of Finance*, 39, 1984, 519–525.

Perrakis, S., and J. Zerbins, "Identifying the SSD Portion of the E-V Frontier. A Note," *Journal of Financial and Quantitative Analysis*, 13, 1978, 167–171.

Philippatos, G.C. and N. Gressis, "Conditions of Equivalence Among E-V, SSD and E-H Portfolio Selection Criteria: The Case for Uniform, Normal and Lognormal Distributions," *Management Science*, 21, 1975, 617–625.

Philippatos, G.C. and C.J. Wilson, "Entrophy, Market Risk, and the Selection of Efficient Portfolios," *Applied Economics*, 4, 1972, 209–220.

Pim van Vliet, "Downside Risk and Empirical Asset Pricing," Erim Ph.D Series Research in Management, 49, 2004.

Pope, Rulon D. and Rod F. Ziemer, "Stochastic Efficiency Normality and Sampling Errors in Agricultural Risk Analysis," *American Journal of Agricultural Economics*, 1984, 31–40.

Porter, R.B., "An Empirical Comparison of Stochastic Dominance and Mean-Variance Choice Criteria," *Journal of Financial and Quantitative Analysis*, 8, 1973, 587–608.

Porter, R.B., "Semi-Variance and Stochastic Dominance: A Comparison," *American Economic Review*, 64, 1974, 200–204.

Porter, R.B., "Portfolio Applications: Empirical Studies," *Stochastic Dominance. An Approach to Decision-Making Under Risk*, Whitmore, G.A., and Findlay, M.C. (Eds.), D.C. Heath, Lexington, MA, 1978, 117–162.

Porter, R.B., and J.E. Gaumnitz, "Stochastic Dominance vs. Mean-Variance Portfolio Analysis: An Empirical Evaluation," *American Economic Review*, 62, 1972, 438–46.

Porter, R.B., and R.C. Pfaffenberger, "Efficient Algorithms for Conducting Dominance Tests on Large Numbers of Portfolios: Reply," *Journal of Financial and Quantitative Analysis*, 10, 1975, 181–185.

Porter, R.B., J.R. Wart and D.L. Ferguson, "Efficient Algorithms for Conducting Stochastic Dominance Tests on Large Numbers of Portfolios," *Journal of Financial and Quantitative Analysis*, 8, 1973, 71–82.

Post, Thierry, ""Empirical test for stochastic dominance efficiency," *Journal of Finance*, 2003, 58, pp. 1905-1931.

Post, G.V. and J.D. Diltz, "A Stochastic Dominance Approach to Risk Analysis of Computer Systems," *Management Science Quarterly*, 10, 1986, 363–375.

Post, T. and Levy, H., "Does Risk Seeking Drive Stock Prices," *Review of Financial Studies*, forthcoming.

Pratt, J.W., "Risk Aversion in the Small and in the Large," *Econometrica*, 32, 1964, 122–136.

Pratt, J.W., "Utility Functions, Interest Rates, and the Demand for Bonds," In Thomas B. Fomby and Tae Kun Seo (Eds.), *Studies in the Economics of Uncertainty. In Honor of Josef Hadar*, Springer Verlag, New York, 1989, 225–231.

Prelac, D., "The Probability Weighting Function," *Econometrica*, 66, 1998, pp. 497-527.

Quiggin, J., *Generalized Expected Utility Theory, The Rank Dependent Model*, Kluwer Academic Publishers, Boston, 1993.

Quirk, J.P., and R. Saposnik, "Admissibility and Measurable Utility Function," *Review of Economic Studies*, 29, 1962, 140–146.

Riechers, R.K., J.G. Lee and R.K. Heitschmidt, "Evaluating Alternative Stocking Rate Tactics. A Stochastic Dominance Approach," *American Journal of Agricultural Economics*, 70, 1988, 1186.

Ritchken, P.H., "On Option Pricing Bounds," *Journal of Finance*, 40, 1985, 1219–1233.

Ritchken, P.H., Y. Agarwal and A. Gupta, "Stochastic Dominance Algorithms Using Piecewise Linear Approximations," *Decision Science*, 16, 1985, 370–380.

Ritchken, P.H., and S. Kuo, "Option Bounds with Finite Revision Opportunities," *Journal of Finance*, 43, 1988, 301–308.

Ritchken, P.H., and S. Kuo, "On Stochastic Dominance and Decreasing Absolute Risk Averse Option Pricing Bounds," *Management Science*, 35, 1989, 51–59.

Robinson, L.J. and P.J. Barry, "Risk Efficiency Using Stochastic Dominance," *Journal of Finance*, 43, 1978, 301–308.

Rothschild, M. and J.E. Stiglitz, "Increasing Risk. I. A Definition," *Journal of Economic Theory*, 2, 1970, 225–243.

Rothschild, M. and J.E. Stiglitz, "Increasing Risk. II. Its Economic Consequences," *Journal of Economic Theory*, 3, 1971, 66–84.

Rothschild, M. and J.E. Stiglitz, "Addendum to Increasing Risk. I. A Definition," *Journal of Economic Theory*, 5, 1972, 306.

Rothschild, M. and J.E. Stiglitz, "Some Further Results on the Measurement of Inequality," *Journal of Economic Theory*, 6, 1973, 188–204.

Roy, A.D., "Safety-First and the Holding of Assets," *Econometrica*, 20, 1952, 431–449.

Russell, Gregory-Allen and Haim Shalit, "The Estimation of Systematic Risk under Differentiated Risk Aversion: A Mean-Extended Gini Approach," *Review of Quantitative Finance and Accounting*, 1998, forthcoming.

Russell, W.R. and T.K. Seo, "Ordering Uncertain Prospects: The Multivariate Utility Functions Case," *Review of Economic Studies*, 45, 1978, 605–611.

Russell, W.R. and T.K. Seo, "Efficient Portfolios by Stochastic Dominance," *Southern Economic Journal*, 46, 1980, 877–882.

Russell, W.R. and T.K. Seo, "Representative Sets for Stochastic Dominance Rules," *Studies in the Economics of Uncertainty. In Honor of Josef Hadar*, Springer Verlag, New York, 1989, 59–76.

Russell, W.R. and P.E. Smith, "Taxation, Risk-Taking, and Stochastic Dominance," *Southern Economic Journal*, 45, 1970, 522–533.

Samuelson, Paul A., "Risk and Uncertainty: A Fallacy of Large Numbers," *Scientia*, April/May 1963, pp. 1–6.

Samuelson, Paul A., "General Proof that Diversification Pays," *Journal of Financial Quantitative Analysis*, 2, 1967, 1–13.

Samuelson, Paul A., "The Fallacy of Maximizing the Geometric Mean in Long Sequences of Investing or Gambling," *Proceedings of the National Academy of Science*, October 1971, 2493–96.

Samuelson, Paul A., "The Judgment of Economic Science on Rational Portfolio Management: Timing and Long-Horizon Effects," *The Journal of Portfolio Management*, Fall 1989, 4–12.

Samuelson, Paul A., "The Long-Term Case for Equities and How It Can Be Oversold," *The Journal of Portfolio Management*, Fall 1994, 15–24.

Sandmo, A., "On the Theory of Competitive Firm under Price Uncertainty," *American Economic Review*, 61, 1971, 65–73.

Sarnat, M., "On the Use of Risk Analysis for the Evaluation of Industrial R&D Expenditures," *Managerial and Decision Economics*, 8, 1987, 255–258.

Saunders, A., C. Ward and R. Woodward, "Stochastic Dominance and the Performance of U.K. Unit Trusts," *Journal of Financial and Quantitative Analysis*, 15, 1980, 323–330.

Scarsini, M., "Dominance Condition for Multivariate Utility Functions," *Management Science*, 34, 1988, 454–460.

Segal, U., "Nonlinear Decision Weights with the Independence Axiom," University of California, Los Angeles, manuscript, 1984.

Schneider, S.L., & Lopes, L.L., "Reflection in preferences under risk: Who and when may suggest why." *Journal of Experimental Psychology: Human Perception and Performance, 12,* 1986, pp. 535-548.

Schneller, I.M., "Regression Analysis for Multiplicative Phenomena and its Implication for the Measurement of Investment Risk," *Management Science*, 22, 1975, 422–426.

Shalit, H., and S. Yitzhaki, "Mean Gini, Portfolio Theory and the Pricing of Risky Assets," *Journal of Finance*, 39, 1984, 1449–1468.

Shalit, H., and S. Yitzhaki, "Marginal Conditional Stochastic Dominance," *Management Science*, 40, No. 5, May 1994, 670–684.

Sharpe, W.F., "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 1964, 425–442.

Sharpe, W.F., "Mutual Fund Performance," *Journal of Business*, Vol. 39, No. 1, January 1966, 19–38.

Sharpe, W.F., "The Sharpe Ratio," *Journal of Portfolio Management*, Vol. 21, No. 1, Fall 1994, 49–58.

Sherman, S., "On a Theorem of Hardy, Littlewood, Polya, and Blackwell," *Proc. National Academy of Science*, 37, 1951, 826–831.

Siegel, Jeremy J., "The Equity Premium: Stock and Bond Returns Since 1802," *Financial Analysts Journal*, January 1992, 28–38.

Stein, W.E. and R.C. Pfaffenberger, "Estimation of Error Probabilities in Stochastic Dominance," *Journal of Statistics, Computation and Simulation*, 25, 1986, 137–165.

Stein, W.E., R.C. Pfaffenberger, and D.W. Frence, "Sampling Errors in First Degree Stochastic Dominance," *J. Financial Res.*, 10, 1987, 259–268.

Stein, W.E., R.C. Pfaffenberger, and P.C. Kumer, "On the Estimation Risk in First Degree Stochastic Dominance: A Note," *J. Financial and Quantitative Anal.*, 18, 1984, 471–476.

Stevens, G.V.G., "On Tobin's Multiperiod Portfolio Theorem," *Review of Economic Studies*, 39, 1972, 461–468.

Stinnett, A.A., and Mullahy, J., "A New Framework for the Analysis of Uncertainty in Cost-Effectiveness Analysis," Working Paper, 1998.

Taylor, Walton R.L., and Yoder, James A., "Mutual Fund Trading Activity and Investor Utility," *Financial Analysts Journal*, Vol. 50, May/June 94, 66–69.

Tehranian, H., "Empirical Studies in Portfolio Performance Using Higher Degree of Stochastic Dominance," *Journal of Finance*, 35, 1980, 159–171.

Tehranian, H., and B.P. Helms, "An Empirical Comparison of Stochastic Dominance Among Lognormal Prospects," *Journal of Financial and Quantitative Analysis*, 17, 1982, 217–226.

Tesfatsion, L., "Stochastic Dominance and the Maximum of Expected Utility," *Review of Economic Studies*, 43, 1976, 301–315.

Thaler, R.H., and E.J. Johnson, "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choices," *Management Science*, 36, 1990, pp. 643-660.

Thistle, Paul D., "Negative Moments, Risk Aversion and Stochastic Dominance," *Journal of Financial and Quantitative Analysis*, Vol. 28, June 93, 301–311.

Thistle, Paul D., "Ranking Distributions with Generalized Lorenz Curves," *Southern Economic Journal*, 56, 1989, 1–12.

Thon, Dominique and Lars Thorlund-Petersen, "Stochastic Dominance and Friedman-Savage Utility Functions," *Math. Social Sciences*, 16, 1988, 305–317.

Tobin, J., "Liquidity Preferences as Behavior Toward Risk," *Rev. Economic Studies*, 25, 1958, 65–86.

Tobin, J., "The Theory of Portfolio Selection," in F.Y. Hahn and F.P. Brechling, eds., *The Theory of Interest Rates*, London, 1965.

Toft, K.B., "On the Mean-Variance Tradeoff in Option Replication with Transaction Costs," *Journal of Financial and Quantitative Analysis*, 31, 1996, 233–263.

Tolley, H.D. and M.R. Kosorok, "An Empirical Method of Comparing Risks using Stochastic Dominance," Working paper, Brigham Young University, 1989.

Tolley, H.D. and R.D. Pope, "Testing for Stochastic Dominance," *American Journal of Agricultural Economics*, 70, 1988, 259–268.

Trennepohi, G.J. Booth and Hassan Tehnrnian, "An Empirical Analysis of Insured Portfolio Strategies Using Listed Options," *Journal of Financial Research*, 11, 1988, 1–12.

Tversky, Amos, "Intransitivity of Preferences," *Psychological Review*, 76, 1969, 31–48.

Tversky, A., and Kahneman, D., "The framing of decisions and the psychology of choice." *Science*, 211, 1981, pp. 453-480.

Tversky, Amos, and Daniel Kahneman, "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *The Quarterly Journal of Economcs*, 107, 4, 1991, 1039–1061.

Tversky, Amos, and Daniel Kahneman, "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 1992, 297–323.

Tversky, Amos, and Peter Wakker, "Risk Attitudes and Decision Weights," *Econometrica*, 63(6), 1995, 1255–1280.

Varey and D. Kahneman, "Experiences Extended Over Time: Evaluation of Moments and Episodes," *Journal of Behavioral Decision Making*, 5, 1992, 169–185.

Vickson, R.G., "Stochastic Dominance Tests for Decreasing Absolute Risk-Aversion. II. General-Random Variables," *Management Science*, 23, 1977, 478–489.

Vickson, R.G., and M. Altman, "On the Relative Effectiveness of Stochastic Dominance Rules: Extension to Decreasingly Risk-Averse Utility Functions," *Journal of Financial and Quantitative Analysis*, 12, 1977, 73–84.

Viscusi, W.K., "Prospective Reference Theory: Toward an Explanation of Paradoxes," *Journal of Risk and Uncertainty*, 2, 1989, 235-264.

von Neuman, J. and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ (2nd Ed.), 1947 (3rd Ed.), 1953.

Weeks, J.K. and T.R. Wingler, "A Stochastic Dominance Ordering of Scheduling Rules," *Decision Sciences*, 10, 1979, 245–257.

Whitmore, G.A., "Third Degree Stochastic Dominance," *American Economic Review*, 60, 1970, 457–459.

Whitmore, G.A., "Stochastic Dominance for the Class of Completely Monotonic Utility Functions," In Thomas B. Fomby and Tae Kun Seo (Eds.), *Studies in the Economics of Uncertainty. In Honor of Josef Hadar*, Springer Verlag, New York, 1989, 77–88.

Whitmore, G.A., M. Ben-Horim and H. Levy, "Using Stochastic Dominance to Identify Admissible Estimates and Predictors," Working Paper, McGill University, 1988.

Whitmore, G.A., and M.C. Findlay (Eds.) *Stochastic Dominance: An Approach to Decision-Making Under Risk*, Lexington Books, Toronto, 1978.

Whitt, W., "Comparing Probability Measures on a Set with an Intransitive Preference Relation," *Management Science*, 25, 1979, 505–511.

Whitt, W., "On Stochastic Bounds for the Delay Distribution in GI/G/S Queue," *Oper. Research*, 29, 1981, 604–608.

Williams, J.R., "A Stochastic Dominance Analysis of Tillage and Crop Insurance Practices in a Semiarid Region," *American Journal of Agricultural Economics*, 70, 1988, 112–120.

Wingender, J. and J.E. Groff, "On Stochastic Dominance Analysis of Day-of-the-Week Return Patterns," *Journal of Financial Research*, 12, 1989, 51–55.

Wu, G., and Gonzales, R., "Curvature of the Probability Weighting Function," *Management Science*, 42, 12, 1996, pp. 1676-1690.

Yaari, M., "Convexity in the Theory of Choice Under Risk," *Quarterly Journal of Economics*, 79, 1965, 278–290.

Yaari, M., "Some Remarks on Measures of Risk Aversion and on Their Uses," *Journal of Economic Theory*, 1, 1969, 315–329.

Yaari, M., "The Dual Theory of Choice Under Risk," *Econometrica*, 55(1), 1987, 95–115.

Yitmaz, M.R., "The Case of Risk and Return Models for Multiattribute Decisions with Decomposable Utilities," *Journal of Financial and Quantitative Analysis*, 18, 1983, 279–285.

Yitzhaki, S., "Stochastic Dominance, Mean-Variance, and Gini's Mean Difference," *American Economic Review*, 72, 1982, 178–185.

Yitzhaki, S., and F. Olkin, "Concentration Curves," Working Paper, Hebrew University, Jerusalem, Israel, 1988.

Yitzhaki, S., and J. Mayshar, "Characterizing Efficient Portfolios," Hebrew University of Jerusalem, 1997, working paper.

Ziemba, W.T., "Solving Nonlinear Programming Problems with Stochastic Objective Functions," *Journal of Financial and Quantitative Analysis*, 8, 1972, 1809–1827.

Ziemba, W.T., "Choosing Investment Portfolios When the Returns have Stable Distributions," *Mathematical Programming in Theory and Practice*, P.L. Hammer and G. Zoutendijk (Eds.), Amsterdam, North-Holland Publishing Company, 1974, 443–482.

Ziemba, W.T., "Portfolio Applications: Computational Aspects," *Stochastic Dominance: An Approach to Decision-Making Under Risk*, Whitmore, G.A., and Findlay, M.C. (Eds.), D.C. Heath, Lexington, MA, 1978, 199–260.

Ziemba, W.T., C. Parkan and R. Brooks-Hill, "Calculation of Investment Portfolios with Risk Free Borrowing and Lending," *Management Science*, 21, 1974, 209–222.

Zilcha, I. and S.H. Chew, "Invariance of the Efficient Sets When the Expected Utility Hypothesis is Relaxed," *Journal of Economic Behavior and Organization*, 13, 1990, 125–131.

INDEX