Chapter 1: Introduction

Chapter summary: This chapter gives a general introduction, as well as some necessary background to the issues addressed in the following chapters. We also discuss the theoretical foundations of market portfolio efficiency and the relation with various Stochastic Dominance rules; in the Mean-Variance framework these issues are largely known, yet there has been relatively little work on these in a Stochastic Dominance framework.

1.1 A brief outline

'How should I invest my money?' This question has been on the minds of investors for an unknown, but obviously long time, and at least since Markowitz (1952) on those of academic researchers as well.[1] The main problem is how to balance risk and return, for almost all assets offer potential rewards as well as losses. A large part of the literature in this broad field assumes that through the concept of utility an optimal trade-off can be identified; the extent and circumstances of the gains and losses can be measured and balanced, at least to some degree.

Yet even to arrive at that result certain assumptions have to be made; assumptions regarding the nature of utility, the similarity of investors (even if a single investor has no influence on the market, the investors as a whole will have an impact on the market), the characteristics of the assets, or the markets on which they are traded.

Assumptions like these are needed for a structured discussion, as required by academic rigor. However, it may be beneficial to take a critical look at those maintained assumptions. This thesis will touch only a niche of this field, that of Stochastic Dominance (SD) analysis, which proposes a certain way to regard risk and hence utility functions. SD does not offer a final answer to the problems of asset pricing or investments, and taken in isolation will often offer less answers than some alternative approaches. Yet it can prove to be a useful tool, one well worth developing and expanding. This was the aim of the research underlying this thesis, for there were – and are still – many possibilities for improvement.

[1] It would probably go too far to name Markowitz as the first researcher who took an interest in this topic, but his seminal paper can be counted as the birth of portfolio theory.

The structure of this introductory chapter is as follows:

Firstly, in section 1.2 we offer a brief discussion of the place of SD in the field of Finance, in particular Asset Pricing. The section also offers comments on the limitations of SD analysis, as well as problems as of yet unsolved.

In section 1.3 the basic concepts of SD, necessary for later chapters, are introduced. This will be an incomplete overview by necessity – the interested reader will on several occasions be referred to other authors who have much more space to cover these topics – yet I hope this thesis will be readable as a stand-alone publication, so some concepts needs to be mentioned.

Section 1.4 will give the motivation of the tests developed in Chapters 2 and 3, and discuss the efficiency of the market in an SD framework. While much of this material is in fact known for decades, it has seen very little application to an SD setting; an omission that is (partly) rectified in this section.

Finally, Section 1.5 will give the structure of the remainder of this thesis.

1.2 Why SD?

The investment problem is, at least since and Markowitz (1952), described in terms of risk and return.[2] The nature of the latter raises relatively few questions; provided one knows the value of an asset (or portfolio of assets) at the beginning and end of a certain time span, the return is easily calculated. In a forward-looking setting an expectation of the return of the assets will be used, but this expectation is often nothing else than the (average) historical return, so we only have to calculate a mean.

However, the nature of risk is subject to more debate. The general definition is not in question (it can be found in many forms, but Levy (1998, p.6) offers this brief yet precise version: "A risky position is a situation in which there is more than one financial outcome, say x_1, x_2, …, x_n and, for at least one value x_i, $0 < p(x_i) < 1$, where p denotes a probability of x_i occurring"). However, how to measure risk is subject of a protracted debate, as well as refinements of the definition.[3] The classical approach is to let the outcomes (x_i) and probabilities p enter a Von-Neumann-Morgenstern utility

[2] Of course, risk and return as concepts predate Markowitz, see for example Hicks (1939, p.126). The point is that earlier treatments did not use a portfolio context.

[3] A good starting point is section 3 of Levy (1992).

function (Von Neumann-Morgenstern, 1944) which is state-independent (i.e. the outcome may differ between different states, but the utility function may not[4]):
A von Neumann-Morgenstern expected utility function is obtained if:[5]

a) There is an assignment of numbers ($u_1, u_2, \ldots, u_n$) to the n outcomes (returns, denoted by x) such that for every set of outcomes[6] L we have:

$$U(L) = U(x_1)p_1 + U(x_2)p_2 + ... + U(x_n)p_n$$

and,

b) The values for U are contained within $\Re$, the set of real numbers.

The next question is then 'What is the form of this utility function *U*?' – we need to model risk to be able to compare one set of outcomes with another. The classical method uses the *variance* of the returns (again, this goes back to Markowitz (1952)) to model risk, and assumes variance decreases an investor's utility. This is called the mean-variance (MV) framework, and is the basis for e.g. the Capital Asset Pricing Model (CAPM; see Sharpe (1964), Linter (1965), Mossin (1966), Black (1972a); the textbook by Cochrane (2005) is a good starting point, Van Vliet (2004) discusses various interesting aspects. But despite its widespread use, variance as a risk-measure has various problems:

- All deviations from the mean are taken to increase risk, disregarding the difference between upward and downward deviations (gains and losses; while it is in principle assumed that losses decrease utility, while gains increase it, the MV criterion ignores this).
- If an investor uses variance as a risk-measure throughout, there will be a point where the investor will violate one of the so-called 'first-principles'[7], namely that of non-satiation; he will prefer less over more.

[4] This is in itself a somewhat restrictive assumption, which some authors drop as well.

[5] The following is adapted from Mas-Collel et al. (1995), p.173.

[6] A set of outcomes could be e.g. an asset or a portfolio.

[7] These regularity conditions are derived from micro-economic utility theory and reflect certain characteristics of the utility function which are assumed by a large part of the research community, namely nonsatiation (positive first derivative of the utility function) and risk-aversion (negative second derivative of the utility function). However, these are not universally accepted, especially global risk-aversion is challenged, see e.g. Hartley and Farrell (2002) and the sources cited therein.

An apt illustration of this problem has been given by Levy (1998, p.2), who compares two alternative investments:

> "x providing \$1 or \$2 with equal probability and y providing \$2 or \$4 with equal probability, with an identical investment of, say, \$1.1. A simple calculation shows that both the mean and 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 highest return on x."

In this example an investor who sufficiently dislikes variance will choose x, despite that he will be always as well, or better, off when investing in y. That is also the reason why, if the MV-framework would hold, it is possible for both assets to require the same initial investment. Yet if every investor does exhibit nonsatiation (part of the 'rationality' mentioned in the quote, and one of the 'first principles') every investor will choose y.[8]

In itself this would be enough to question the use of the MV-framework, yet there are more objections, namely the conditions that have to be met to achieve compatibility between Expected Utility Theory (EUT) and the MV-framework. In fact, there is only a limited set of circumstances under which the MV-model applies.

1. If the utility function is quadratic, i.e. of the form $U = \alpha x^2 + \beta x$. By design this means only the first two moments of the return distribution are relevant, so mean and variance are the only measures of return and risk needed. It also implies that non-satiation will be violated at some point, as a quadratic function cannot be increasing over its entire domain.
2. If the return distribution belongs to a certain class, namely those that are characterised by the location and scale parameters, and the scale parameter is proportional to the variance. See Bawa (1975, p 111 – 118), who shows that

[8] Of course, in a situation where every investor buys y, the price of y will increase relative to that of x, on a free market. Yet the market responds to the preferences, so under the MV-rule this need not happen.

for these sets of distributions the decision rule can be phrased in terms of comparing the mean and the scale parameters. If the later is proportional to the variance, the MV-criterion is applicable. The classes of admissible return distributions include the normal, t-, lognormal, exponential and uniform distributions, though there may be additional constraints on the range of the returns (again, see Bawa (1975).

The above offers the motivation to search for alternatives to variance; SD is just one of these, there are many different ways to model the utility function which are not affected by the disadvantages mentioned above. An example is the use of Lower Partial Moment (LPM) utility functions. This approach is developed by among others Bawa (1975), Bawa-Lindenberg (1977) and Harlow & Rao (1989). The MLPM (mean –LPM) framework only takes returns below a threshold into account when modelling risk.[9] The formula for the LPM of portfolio $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\lambda}$ is:

$$\mathrm{LPM}_{\mathrm{n}}(\tau, \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\lambda}) = \int_{-\infty}^{\tau} (\tau - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\lambda})^{\mathrm{n}} \,\mathrm{d}F(\boldsymbol{x})$$

The investor then minimizes the LPM, given a prespecified required return of τ. Note that the LPM formulation forces us to choose n (the order of the LPM) and the target rate τ. Especially the latter may be a difficult choice; one could make a case for such values as zero, a risk-free return, inflation (however measured), or a required return on investment to meet some kind of goal. This kind of flexibility (for the MPLM model incorporates a variety of other specifications, with n=2 and τ equal to the risk-free rate we would obtain the same results as an MV specification[10] (Harlow & Rao (1989)) also raises questions. For example, how a market consisting of investors with different target rates would behave is not completely clear.

An other potentially awkward fact is that alternatives like LPM still require an explicit model of the utility function, a topic on which we have only limited, and sometimes contradictory evidence. Examples include Harvey and Siddique (2000) who

[9] Though strictly speaking, we need a return distribution that belongs to the two-parameter location-scale family to justify using only LPMs based on a single target rate τ. See Harlow & Rao (1989) and Bawa and Goroff (1982)

[10] But only in case the return distributions are such that a MV framework is valid.

effectively propose a cubic utility function and Dittmar (2002) who suggests that a fourth order polynomal is needed; or compare for instance the work of Friedman and Savage (1948), or Kahneman and Tversky (1979) who have different reasons to propose (local) risk- seeking (see Hartley and Farrell (2002) for a recent discussion) with much of the Mean-Variance and SD literature which assumes global risk-aversion.

Non-parametric methods offer the benefit of releasing the researcher of the obligation to make a judgement call on the proper specification of the utility function; this is instead determined by the data, subject to the restrictions the researcher may want to impose.

This brings us to the SD methodology. SD is a non-parametric method, which does not model an explicit utility function, but takes into account all possible forms of this function which conform to a set of restrictions. For the basic forms of SD analyis, these restrictions come from micro-economic principles; they are discussed in detail in the next section. Important is the point that *SD allows a data-driven, non-parametric modelling of risk, while directly imposing the restrictions on (marginal) utility.* SD requires the researcher to make very few choices that are based on judgement calls or may even end up being arbitrary, and prevents the model from contradicting first principles.

The above may seem that SD analysis offers all benefits and no drawbacks. While the benefits are obvious, the drawbacks are present as well. Some of these have been (partially) overcome recently, others remain. The main problems hindering widespread application of SD are (or were):

Diversification

The main problem for many years was that SD rules could only be applied when comparing 2 distributions, i.e. 2 assets. This means that, if there is any diversification and we want to test for the efficiency of an asset or portfolio relative to a range of assets, the test should be repeated for every possible combination in that range. This requires an infinite number of comparisons and is therefore not feasible. One needed a methodology that enabled researchers to take diversification into account in a single test. The first ones of this kind were developed by Post (2003) and Kuosmanen (2004). Chapter 2 of this thesis is based on Post-Versijp (2007) which replaces the

Post (2003) test with a more powerful version. However, the approach has not yet been extended to all possible SD criteria (DARA-SD and variations which allow for risk-seeking; see also section 1.3).

Computation

The early SD literature was partially based on algorithms designed to identify dominated alternatives. Even without a complete coverage of the possibilities of diversification, the large datasets typically available in Finance (as an extreme example, the CRSP database for monthly returns on the US markets currently has over 2.7 million entries) demanded large amounts of computer-time. Even today this issue still has not disappeared; while the examples used in this thesis can all be replicated on a 'normal' PC, the computational burden rises exponentially with the amount of assets and the number of periods. Certain problems are still well out of our reach due to lack of computing power.

Identification of 'optimal' choices

Another factor which makes SD harder to 'sell' is the fact that it deals with classes of utility functions that are quite wide, and hence cannot identify 'optimal' choices. In MV analysis it is often possible to determine the portfolio weights that would maximise utility, as the utility is known. Even in a more general specification, the MV utility function can be constructed as to have only a single unknown parameter that controls the trade-off between risk and variance. SD, or any non-parametric method, will find it nigh-on impossible to match this property.

The best SD could potentially do is identify dominated alternatives, namely those that are *not* optimal for *all* investors, and hence can be discarded. This approach is unlikely to reduce the set of options to a similar extent as parametric approaches, and may not even be feasible for all SD criteria (e.g. for DARA SD, see below, such a test is to the best of our knowledge not available)

1.3 A short overview of SD concepts

To facilitate the discussion in later chapters, this section offers a brief recap of some central concepts in our SD tests.

The orders of Stochastic Dominance

As stated in section 1.2, SD is based directly on the certain micro-economic principles.[11] The origins are outside Finance; see Levy (1992) for some early references. In decision theory SD was introduced by Allais (1953), Quirk and Saposnik (1962) and Fishburn (1964), the applications to Finance started a few year later (see below). The most commonly invoked principles are:

- non satiation: investors prefer more over less; marginal utility is positive.
- risk aversion: investors prefer a certain outcome over an uncertain one with the same (expected) payoff; marginal utility is decreasing
- skewness preference: investors prefer positively skewed distributions; marginal utility is convex.

For each of these, there is a corresponding SD order, which states if, when comparing assets (specifically their outcomes and probability distributions), there is one which would dominate the other, meaning that *every investor conforming to this rule will prefer the dominating choice, regardless of any other feature or difference of the utility functions*. The customary SD rules are:

FSD

First-order Stochastic Dominance (FSD; Hadar and Russell (1969), Hanoch and Levy (1969), Rothschild and Stiglitz (1970)) only requires marginal utility to be positive, meaning that investors prefer more over less. Mathematically, FSD states that X dominates Y if, and only if:[12]

$$E[U(x)] - E[U(y)] \geq 0 \qquad \forall\, U \in U_1$$

Where U_1 denotes all utility functions with positive marginal utility. This can be rewritten as:

[11] As opposed to variance. Variance and SD are not compatible in general; the MV-criterion is neither necessary nor sufficient for dominance, unless special restrictions are placed on the return distributions.

[12] As is customary, in the following 'if, and only if' will often be abbreviated to 'iff'.

$$G(x) - F(x) \geq 0 \qquad \forall\, x$$

Where F(.) and G(.) are cumulative probability distribution functions (CDFs). See Hanoch & Levy (1969), Fishburn and Vickson (1978), Levy (1998). This notation underscores the fact that SD uses the complete distribution of the return, rather than just the first and second moments (mean and variance).

SSD

Second-order Stochastic Dominance (SSD; again see Hadar and Russell (1969), Hanoch and Levy (1969), Rothschild and Stiglitz (1970)) requires marginal utility to be positive and decreasing, adding risk aversion to the requirement of non-satiation. Under SSD, X dominates Y iff:

$$E[U(x)] - E[U(y)] \geq 0 \qquad \forall\, U \in U_2$$

Where U_2 denotes all utility functions with positive and decreasing marginal utility, This can be rewritten as:

$$\int_{-\infty}^{x} G(x) - F(x) dU(x) \geq 0 \qquad \forall\, x$$

where F(.) and G(.) are cumulative probability distribution functions (CDFs).

TSD

Third-order Stochastic Dominance (TSD; Whitmore, (1970)) adds skewness preference to SSD. TSD states that X dominates Y iff:

$$E[U(x)] - E[U(y)] \geq 0 \qquad \forall\, U \in U_3, \text{ and}$$
$$E[F(x)] - E[G(x)] \geq 0$$

Where U_3 denotes all utility functions with positive, decreasing and convex marginal utility, and F(x) and G(y) are cumulative probability distribution functions (CDFs). Again, we rewrite this as:

$$\int_{-\infty}^{x}\int_{-\infty}^{v} G(\mathrm{x}) - F(\mathrm{x}) dU(\mathrm{x}) dv \geq 0 \quad \forall\, \mathrm{x}, \text{ and}$$

$$E[F(\mathrm{x})] - E[G(\mathrm{x})] \geq 0$$

TSD is often motivated as a necessary condition for DARA, see below.

Higher order SD

In the same vein as TSD builds on SSD, one can construct ever higher orders of SD. A natural extension is fourth-order stochastic dominance (4SD), which assumes kurtosis aversion in addition to non-satiation, risk aversion and skewness preference.

Also note that each successive order further limits the range of admissible utility functions. This has the advantage of eliminating more possibilities when it comes to the efficient set (see below). On the other hand, higher orders are harder to justify using economic arguments: only for 4SD there is such an economic explanation (Kurtosis aversion is a necessary condition for decreasing absolute prudence; which is the phenomenon one can defend with economic arguments (Kimball, 1990); there is to the best of our knowledge none for 5SD and upwards. Also, higher order SD criteria are more difficult in empirical work; one needs a very broad return interval to draw conclusions about the utility specifications that these criteria propose (or rule out).

Convex SD

Convex Stochastic Dominance (Fishburn (1974) is a form of SD that restricts the undominated (i.e. admissible) set by introducing the possibility that an asset or portfolio can be dominated by a convex combination of other assets, rather than by the assets themselves.[13] It can be applied to any order of SD discussed so far, for the sake of brevity we'll only provide the formula for CSSD (Convex Second order SD):

Under CSSD, X dominates Y iff:

$$\int_{-\infty}^{\boldsymbol{x}} G(\boldsymbol{x}) \mathrm{dU}(\boldsymbol{x}) - \sum_{\mathrm{i}=1}^{\mathrm{i}=\mathrm{n}} \lambda_{\mathrm{i}} \int_{-\infty}^{\boldsymbol{x}} F(\boldsymbol{x}) \mathrm{dU}(\boldsymbol{x}) \geq 0 \qquad \forall\, \boldsymbol{x}$$

[13] This goes a long way towards diversification, but diversification need not be restricted to convex combinations of assets.

CSSD eliminates more portfolios from the admissible set and is hence a more stringent criterion than the corresponding 'normal' SD order. Bawa et al (1985) show that for their data the number of undominated portfolios decreases by roughly a third.

Other criteria

There have also been researchers who developed SD criteria based on different principles, such a Decreasing Absolute Risk Aversion (DARA; Pratt (1964)), which results in DARA-SD (Vickson 1975, Levy (1998)), which is a more stringent criterion than TSD[14], and SD for risk-seeking investors (Post and Levy (2005)), which is compatible with FSD, but incompatible with higher orders. While they present interesting avenues for research, the tests we propose in Chapters 2 and 3 are not (readily) adaptable to these situations due to mathematical problems, hence we will not focus on them.

SD Efficiency

The concept of efficiency can be defined for each of the SD orders mentioned above. Under relatively mild assumptions to be discussed in Chapter 2, the investors' optimization problem can be summarized as

$$\max_{\boldsymbol{\lambda}\in\Re^N} \int \mathrm{u}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\lambda})\mathrm{dG}(\boldsymbol{x})$$

It is possible that a certain portfolio will be stochastically dominated for all utility functions belonging to the class under consideration (e.g., functions with positive marginal utility for SSD). Such a portfolio is deemed inefficient; no investor will invest in it. Portfolios that are efficient cannot be ruled out for an investor conforming to that SD-criterion; if there is just one well behaved utility function for which the portfolio is optimal, it is defined as efficient. Note that this slightly deviates from the more traditional definition of efficiency, which asks if there exists a portfolio $\boldsymbol{\lambda} \in \Re^N$ that is preferred to $\boldsymbol{\tau}$ for all utility functions $u \in U_k$ (where k = 1,2,3,… depending on which SD criterion is used), but in the setup of the following chapters the two definitions will coincide; also see Post (2003), Theorem 1.

[14] Except for the case of equal means, where they coincide. See Fishburn and Vickson (1978).

To arrive at a formal definition of SD efficiency, one must first establish the optimality condition. An evaluated portfolio[15] $\boldsymbol{\tau} \in \Re^N$ is optimal for a given utility function $u \in U_k$, k = 1,2,3,…, if and only if the first-order optimality condition, or Euler equation, is satisfied:

$$\boldsymbol{\alpha}(u) \equiv \int u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}dG(\boldsymbol{x}) = \mathbf{0}_N$$

This enables us to give the formal definition of efficiency:

DEFINITION 1 *The evaluated portfolio* $\boldsymbol{\tau} \in \Re^N$ *is nth-order stochastic dominance efficient if and only if* $\boldsymbol{\alpha}(u) = \mathbf{0}_N$ *for some* $u \in U_k$. *The portfolio is k-th order SD inefficient if and only if it is not optimal, that is,* $\boldsymbol{\alpha}(u) \neq \mathbf{0}_N$, *for all* $u \in U_k$.

Two fund Separation under SD

Two-fund separation (TFS; see, e.g., Tobin (1958), Cass and Stiglitz (1970) and Ross (1978)) is a related hypothesis to that of efficiency. TFS applies if *all* well-behaved investors mix the same portfolio of risky assets (which has to be the value-weighted market portfolio if all assets are to be held) with the riskless asset. This means that TFS is a much stricter condition than simply efficiency of the market portfolio; instead of a single well-behaved investor, we now need all well-behaved investors to hold the market portfolio. Equivalently, TFS can be defined as follows:

DEFINITION 2 *Two-fund separation applies if and only if the market portfolio* $\boldsymbol{\tau} \in \Re_+^N : \boldsymbol{\tau}^{\mathrm{T}}\mathbf{1}_N = 1$, *is the optimal portfolio for all well-behaved investors who do not lend or borrow, that is,*

$$\boldsymbol{\alpha}(u) = \mathbf{0}_N \quad \forall u \in U_2 : \boldsymbol{\alpha}(u)^{\mathrm{T}}\boldsymbol{\tau} = 0 \tag{1}$$

[15] We use τ to denote a vector of portfolio weights; the product of these weights and the set of assets yields the evaluated portfolio. Also, throughout the text, we will use $\Re^N$ for an *N*-dimensional Euclidean space, and $\Re_+^N$ denotes the positive orthant. Further, to distinguish between vectors and scalars, we use a bold font for vectors and a regular font for scalars. Finally, all vectors are column vectors and we use $\boldsymbol{x}^{\mathrm{T}}$ for the transpose of $\boldsymbol{x}$.

Note that this definition requires positive weights for each asset, and that these weights must sum to one; this is to ensure all assets are indeed held. Lending or borrowing does not materially affect this issue, as any investor who lends or borrows still has to hold the market.

One way to test the TFS is by examining the actual investment portfolios of investors (for instance, Blume and Friend, 1975) or participants in controlled laboratory experiments (for example, Kroll *et al.*, 1988). However, the observed behavior of individuals may not represent the behavior of the aggregated capital market. Surprisingly, as far as we know, the literature currently does not propose empirical tests for TFS at the market level. Interestingly, a simple variation to our SSD efficiency test can test TFS.

1.4 Testing for market portfolio efficiency: what makes the market special?

The Stochastic Dominance tests described in this book are all related to the market portfolio. The applications of the efficiency tests all have the null hypothesis of the market portfolio being efficient; the two-fund separation test asks if the data could have been generated by a representative investor – who of course would hold the market as his portfolio, since all assets need to be held by someone. Yet neither test requires the use of the market portfolio from a purely econometric perspective. We could test for efficiency of any random portfolio, or 2-fund separating involving any other portfolio, using the same instruments. The special focus on the market portfolio requires an explanation.

There are several reasons why one would have a special interest in the market portfolio's efficiency:

- Comparability with the Mean-Variance framework. The standard MV setup with quadratic utility or normal returns leads to an efficient market portfolio. Since the SD framework employs different assumptions, it is of interest to see if the results change.

- The theory of complete markets. If one assumes complete markets, efficiency of the market portfolio follows. Rejecting the null of efficiency forces the researcher to either question this assumption or the quality of his data.
- Convexity of the efficient set. If the efficient set is convex, market efficiency would obtain without further assumptions on the set of tradable assets.
- Homogeneity of preferences. If preferences are similar enough between individual investors, market portfolio efficiency may result.

Yet one may ask if those arguments, mostly developed using assumptions differing from those embedded in the concepts of SD, are valid in our setup. Methodologically this question is extremely relevant, as it directly concerns the null-hypothesis of many applications. And aside from the arguments based on theory and comparability, there is also the 'revealed preference' argument stating that, given the popularity of assets that mimic a proxy for the market portfolio (for example Exchange Traded Funds and broadly diversified mutual funds), efficiency of said portfolio is likely (or, at least, cannot be dismissed out of hand). Hence a situation where the market portfolio cannot be efficient would be a huge negative in assessing our advances.

Furthermore, a situation in where the market portfolio may or may not be efficient would pose problems in the development of the SD field as well, as it would be hard to justify taking the efficiency of any other (random) portfolio as a null-hypothesis. Ideally, at least one of the generally accepted SD criteria would be a sufficient condition for market efficiency, or, failing that, some new SD criterion would have the same property. Unfortunately, this ideal may not be reached, as is explained below.

1.4.1 Complete markets

Following Dybvig & Ingersoll (1982, p. 235) we define complete markets as follows: "Complete markets: Each competitive investor can obtain any pattern of returns through the purchase of marketed assets (subject only to his own budget constraint). If the number of outcome states is finite, markets are complete if the number of marketed assets with linearly independent returns is equal to the number of states"

It can be shown that under these conditions the resulting decisions are indistinguishable from those that would have been made by a single 'representative' investor endowed with the total wealth. This argument has developed over time, important building blocks[16] are: Arrow (1953), Debreu (1959), Wilson (1968), Rubinstein (1974), Dybvig & Ingersoll (1982).

It should be noted that under complete markets, there is no restriction on the utility function except being of the Von Neumann-Morgenstern variety, and exhibiting risk-aversion. Hence, complete markets are not only compatible with SSD and higher order criteria, but offer a sufficient condition for market portfolio efficiency. Yet this assumption can be attacked based on its realism; it's self-evident that not all risks can be traded on the financial markets, and it is hard to say to what extent the results would be affected should there be even a small violation of completeness. This adds a certain all-or-nothing dimension to the problem, making the assumption of completeness somewhat hard to defend. However, it should be noted that this has not stopped academics from using this assumption, in a similar manner the problem of having no perfect proxy for the market portfolio (which should theoretically include all assets, including paintings, rare coin collections and everything else with a less-than-perfect correlation with other assets) is customarily ignored.

If we drop the assumption of complete markets, the picture becomes a lot vaguer. As Detemple and Gottardi (1998) show, under incomplete markets aggregation obtains only under (locally[17]) homothetic preferences or under quadratic utility. The original arguments are detailed below. Note that aggregation in itself is a sufficient condition for market portfolio efficiency, but not a necessary one. For later reference it is also worthwhile to mention that efficiency under incomplete markets does not necessarily translate to two-fund separation.

1.4.2 Convexity of the efficient set

Among the various factors that can lead to market portfolio efficiency, the convex efficient set takes a special place. If the efficient set[18] would be convex, this means

[16] Unfortunately, there's to the best of the author's knowledge no review paper that presents the result in a concise way.

[17] This is an extension to the results in the next section, though not one that is of any assistance to our problems.

[18] Portfolios that would be optimal for an investor with a well-behaved utility function; it is possible that every possible utility function would result in a different portfolio.

that any combination of efficient portfolios would be efficient as well. This is of special importance, as the individual optimisation alone would yield a state indistinguishable from one where there would be only one (representative) investor, without the need to make further assumptions about for instance the completeness of the market. Efficiency of the market would in that case depend on the characteristics of the investors and (possibly) the return distributions, but not on other properties of the market, such as the space of tradable assets. Examples of this can be found in the works cited above, where it was shown that under certain return distributions (e.g. the normal distribution) the MV framework would hold and result in convexity of the efficient set. Bodurtha and Shen (2001) discuss some links between the return distribution and SD rules. However, if a restriction on the distribution(s) of returns is needed, it would not be clear ex ante if that restriction would be met in empirical settings, and how to deal with the almost unavoidable deviations. Empirical returns seldom conform to statistical distributions; see for example the descriptive statistics in Chapter 4.

Convexity based on the return distributions is not the only route. An interaction with the utility functions, or convexity based on conditions on the utility specification alone, are also possible.

In general, a convex efficient set would imply the following (in terms of pricing errors):

if: $$\begin{aligned} E[u'_1(x'\lambda_1)x] = 0 \quad u'_1 \in U^* \\ E[u'_2(x'\lambda_2)x] = 0 \quad u'_2 \in U^* \end{aligned}$$

then

$$E[u'_3(w_1 x'\lambda_1 + w_2 x'\lambda_2)x] = 0 \quad u'_3 \in U^*;\ w_1 + w_2 = 1;\ w_1 \geq 0 \tag{2}$$

where U* refers to the set of utility functions conforming to the SD-criterion of choice, x denotes the return distribution of the portfolio, and λ denotes portfolio weights. If we can prove that theoretically (2) should hold, we can then test the theory by looking at nothing more than market efficiency, which is relatively straightforward – assuming one is satisfied with the proxy for the market portfolio that's being used, which is in most cases a factor that needs to be taken into consideration. If (2) holds for a SD criterion, market efficiency necessarily follows, and hence would provide a powerful test of these models.

Disappointingly, neither the FSD nor SSD restrictions alone lead to a convex efficient set (Dybvig & Ross, 1982), without further restrictions on the return distribution. To illustrate the problem, it's worthwhile to briefly recapitulate the argument made by these authors. Dybvig & Ross make their case in a 3 asset, 4 state setup (the minimum amount of assets and states-of-the-world for the argument to hold), using the following returns:

$$\boldsymbol{X} = \begin{bmatrix} 66 & 82 & 78 \\ 44 & 52 & 48 \\ 52 & 38 & 48 \\ 50 & 50 & 48 \end{bmatrix}$$

Each state has equal probability and both asset 1 and asset 2 are supported by a generalized marginal utility vector – in other words, there are well-behaved investors for which these assets would be optimal. Dybvig and Ross only assume FSD at this stage, though the example is compatible with SSD as well. So both $\mathbf{x_1}$ and $\mathbf{x_2}$ are in the efficient set. However, the combination ½$\mathbf{x_1}$ + ½$\mathbf{x_2}$ is not. Not only does it run foul of Dybvig & Ross' theorem 2*, if we look at the portfolio $\boldsymbol{w}$* = (-1, -0.4, 2.4) we get the following:

$$\begin{aligned} &\mathrm{E}[\mathrm{U}(\boldsymbol{w}^* \boldsymbol{X})] = \mathrm{E}[\mathrm{U}(74, 50.4, 48, 45.2)] = \mathrm{E}[\mathrm{U}(74, 48, 45.2, 50.4)] \\ &> \mathrm{E}[\mathrm{U}(74, 48, 45, 50)] = \mathrm{E}[\mathrm{U}(½\mathbf{x_1} + ½\mathbf{x_2})] \qquad \forall\, \mathrm{U} \in \mathrm{U}_1 \end{aligned}$$

So for all FSD and SSD admissible utility functions the efficient set is shown to be non-convex. This non-convexity also holds in the presence of short selling restrictions.

More precisely, the last part of the argument is airtight, the convex combination of the two original portfolios is FSD dominated, which cannot lead to efficiency under any SD rule. Hence Dybvig & Rubinstein's assertion the example is valid for all monotone utility functions.

However, there might be a way to circumvent this problem, namely by restricting the efficiency of the original portfolios. Perhaps an SD-rule can be found which reduces the efficient set in such a way only a convex set is left (The rule has to shrink the efficient set compared to SSD, so it must impose more conditions on top of that criterion). Furthermore, restricting the return distributions may lead to some progress as well, though the question of how to deal with empirical distributions that do (exactly) not conform to such restrictions remains.

1.4.3 Homogeneity of preferences.

Another way to arrive at a representative investor model is through homogeneity of preferences. This approach is related to convexity of the efficient set, but the demands are more stringent: here the starting point is utility functions that are such that each investor invests in the *same portfolio* (though possibly with different percentages invested in that portfolio and the riskless assets). The convexity argument does *not* require all investors to invest in the same portfolio, only that the combinations are also optimal for a well-behaved investor due to either the nature of the utility function, the return distribution, or their interaction – but no-one needs to hold the same portfolio, or even the market.

Yet if a homogeneity condition holds, and every investor does hold the same portfolio, market clearance automatically leads to the result that this portfolio must indeed be the market. Consequence is that everyone holds the market, so convexity of the efficient set is a moot issue: there is only one efficient portfolio of risky assets. Also, note that convexity of the efficient set is based on an interaction of the utility functions (i.e. SD-order) and the return distribution, while the homogeneity argument only involves utility functions and would hold irrespective of the distribution of asset-returns.

The seminal paper on this issue is Rubinstein (1974), who gave sufficient conditions for this type of aggregation. Brennan & Kraus (1978) showed that the same conditions are also necessary[19], thereby excluding the possibility that a broader category of utility specifications could lead to the same result.

[19] More specifically, Brennan & Kraus (1978) show that the necessary condition is that the Engel curves of all investors must be parallel straight lines, for which necessary conditions are either the same 'cautiousness' and beliefs for all investors, or exponential utility for all investors. These happen to be the same conditions derived as sufficient by Rubinstein (1974), minus the fact that the assumption of equal beliefs – which Rubinstein uses throughout – can be dropped in case of exponential utility.

Adapting Rubinstein (1974)'s text to our own notation, utility functions that lead to aggregation must be solutions of the differential equation:

$$-U'(x)/U''(x) = A + Bx \qquad (3)$$

where A and B are fixed parameters. This equation has three solutions depending on the value of B,

(I) $U(x) \sim -A(e^{xA^{-1}})$ (B = 0)

(II) U(x) ~ ln(A + x) (B = 1)

(III) $U(x) \sim \dfrac{b}{1-b}(A + Bx)^{1-b}$ (B ≠ 0, 1)

where $b \equiv B^{-1}$ and ~ means 'is equivalent up to an increasing linear transformation to'. Note that these functions are subject to several regularity conditions, which can be found in Rubinstein's footnote 2.

The utility specifications above are not universally compatible with FSD, SSD and TSD. First of all, existence is sometimes an issue, as the logarithm of a negative number or a negative number taken to a negative power cannot be meaningfully defined in this context. Even then, the parameters need to be chosen with care to prevent violation of nonsatiation; similar to the quadratic utility function of the CAPM, marginal utility can become negative unless the parameters are specifically matched to the dataset – in general, (A+Bx) should be positive – to prevent this. As shown above this may not be possible if other restrictions are imposed as well.[20,21]

So the search for a SD criterion that necessitates aggregation does not have a 'happy ending'. We have necessary and sufficient conditions for aggregation, but they are parametric, and even when these utility specifications are not in violation of the SD

[20] Quadratic utility is a subcase of solution III with B=-1.

[21] At first sight, (3) seems very silimar to the DARA specification. However, the DARA SD criterion will not be of any help here, as DARA requires:

$$\frac{\partial \pi(x)}{\partial x} < 0, with\, \pi(x) = -U'(x)/u''(x)$$

Using the utility specifications in (3), this means A+Bx should be negative throughout, which leads to negative marginal utility and a violation of the (F)SD regularity conditions.

rules, they are a good deal more stringent. One of the prime advantages of the SD approach, namely the lack to parameterize the utility function, is lost. Given the fact that it is possible to arrive at an efficient market without every individual investor having the same optimal portfolio (e.g. by complete markets or perhaps convexity) this route proves to be rather unsatisfactory. Also the assumption of investors having the same beliefs – which is made throughout in the above – is only justifiable as an ad-hoc way to exclude problems. Recent work in this subfield tends to focus on aggregation under heterogeneous beliefs instead (see e.g. Calvet et al. (2001) and their references).

1.5 Structure of the thesis

The second Chapter will introduce a test for the SD efficiency of a given portfolio relative to a set of assets and all the possible combinations created by diversification. The test uses the SSD and TSD criteria, but could easily be expanded to 4SD. Contrary to earlier tests, it accounts for all pricing errors. We investigate its properties through simulation. Chapter two is adapted from Post-Versijp (2007).

Chapter three will offer a new test for Two-fund separation (TFS). Using a methodology that looks at the largest pricing error (for technical reasons an approach similar to chapter 2 is not possible) we investigate if a given portfolio would be optimal for *all* possible investors belonging to an SD class. This Chapter is adapted from Post-Versijp (2005).

Chapter four will investigate how these tests fare when used on empirical data; we investigate the market efficiency relative to a wide variety of benchmark portfolios formed on different 'anomalies', as well as if TFS applies. We use these tests to draw conclusions about the representative investor model.

Chapter 5 concludes, followed by the customary summary in Dutch.