Chapter 2: Multivariate tests for SD efficiency of a given portfolio[1]

Chapter summary: This chapter develops empirical tests for stochastic dominance efficiency of a given investment portfolio, relative to all possible portfolios formed from a given set of assets. These tests use multivariate statistics, which results in superior statistical power properties compared to previous SD efficiency tests. Interestingly, it also increases the comparability with existing mean-variance efficiency tests.

As shown in Chapter 1, A portfolio is second-order stochastic dominance (SSD) efficient if and only if it is optimal for some nonsatiable and risk-averse investor. Similarly, third-order stochastic dominance (TSD) efficiency applies if and only if a portfolio is optimal for some nonsatiable, risk-averse and skewness-loving investor. These efficiency criteria are theoretically appealing, because they impose economically meaningful regularity conditions (nonsatiation, risk aversion and skewness preference), while avoiding further structure that does not follow from economic theory, such as a particular class of statistical distributions.

In this respect, the stochastic dominance (SD) efficiency criteria have an important advantage relative to the traditional mean-variance efficiency criterion. By focusing on the first two central moments exclusively and allowing for every possible trade-off between these two moments, the mean-variance criterion may classify portfolios that are optimal for some investors as inefficient and classify portfolios that are inferior for all investors as efficient.[2] A similar case can be made against parametric extensions such as the mean-variance-skewness framework; moments of order four and higher are excluded and the preferences over the first three moments are not restricted to obey the regularity conditions.

Despite their theoretical appeal, the SD efficiency criteria thus far have not been applied in empirical finance on a broad scale. Paradoxically, due to the availability of large

[1] This chapter is an adaptation of Post-Versijp (2007). In order to obtain a clearer exposition for this thesis, parts of this article are moved to other chapters.

[2] A good illustration of failure of the mean-variance criterion comes from the "grand old man" of SD, Haim Levy (1998, p. 2): "[Consider] 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 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."

data sets of historical returns, this research field seems particularly well suited for a non-parametric approach. This situation can be explained by the traditional SD tests being relevant only for pairwise comparisons of a finite set of choice alternatives. Interesting advances in this field notwithstanding, these pairwise tests do not apply for portfolio choice problems in which infinitely many portfolios can be formed by means of diversification.[3] To deal with this problem, Post (2003), Kuosmanen (2004) and Post and Levy (2005) developed Linear Programming tests for SD efficiency that do account for diversification possibilities.

While these tests provide an important step in the evolution of the SD methodology, they are only a starting point for developing a SD framework for portfolio analysis. The most important limitation of the existing tests is the lack of statistical power, or the inability to detect inefficient portfolios, in small samples. Specifically, Post's simulations show that the linear programming test of SSD efficiency involves little power in typical asset pricing applications. The lack of power follows from the focus of the test on the maximum positive pricing error and ignoring the remaining errors. A multivariate test that considers all pricing errors jointly can be more powerful. Also, the SSD criterion does not impose skewness preference and hence may fail to reject efficiency if the evaluated portfolio is optimal for skewness-averters but inferior for skewness-lovers. A test for TSD efficiency will reject a broader range of portfolios and hence will be more powerful.

The goal of this chapter is to develop more powerful SSD and TSD efficiency tests. In the spirit of the generalized method of moments (GMM; Hansen, 1982), our tests focus on a weighted sum of squared errors, using weights that reflect the joint probability distribution of the errors. However, in contrast to typical GMM tests, our tests avoid a functional form specification while imposing the regularity conditions of nonsatiation and risk aversion. Compared to the linear programming test, the new SSD test has substantially more power, because it considers all pricing errors jointly rather than the maximum positive pricing error only. The TSD test is even more powerful, because it imposes skewness preference in addition to nonsatiation and risk aversion. In fact, the power of this test can be shown to be comparable with that of the well-known Gibbons, Ross and Shanken (1989, henceforth GRS) test for mean-variance efficiency in the case of a normal distribution. For a non-normal distribution, the GRS test generally does not apply, while our tests do.

The focus on the SSD and TSD criteria means that risk aversion is assumed

[3] For a representative sample of state-of-the-art SD tests for pairwise comparison, we refer to Dardanoni and Forcina (1999), Davidson and Duclos (2000), Barrett and Donald (2003) and Linton *et al*. (2005).

throughout this study. For various reasons, we do not cover the first-order stochastic dominance (FSD) criterion, which allows for risk seeking behavior. First, risk aversion is a standard assumption in financial economics, following from diminishing marginal utility of wealth, and consistent with common observations such as risk premiums for risky assets, portfolio diversification and the popularity of insurance contracts. There are indications for local risk seeking behavior at the individual level, witness for example the popularity of lotteries. However, the bulk of the literature on asset pricing and portfolio selection assumes that investors are globally risk averse when forming investment portfolios. Second, a FSD efficiency test adds relatively little value to higher-order tests, for the simple reason that risk seekers generally will hold ill-diversified portfolios. Not surprisingly, Kuosmanen finds that the FSD and SSD criteria yield exactly the same results for testing market portfolio efficiency. Third, the FSD criterion is very general and allows for exotic preference structures, for example utility functions with inflection points and discontinuous jumps. Thus, an empirical test for FSD efficiency will have considerable freedom to fit a utility function to the data. Presumably, this will considerably slow down the rate of convergence of an empirical test. For the sample size in typical applications, an empirical test will generally lack statistical power to allow for a meaningful application. Fourth, our tests are based on the first-order conditions for portfolio optimization. These conditions are necessary and sufficient in case of risk aversion but they are no longer sufficient if we allow for risk seeking.

The remainder of this chapter is structured as follows. Section 2.1 introduces preliminary notation, assumptions and definitions. Section 2.2 introduces our new SSD test, and discusses issues of statistical inference and computation. Next, Section 2.3 extends the analysis of SSD to a TSD test. These first three sections focus on the basic case without portfolio restrictions. Section 2.4 extends the analysis to the case with general linear restrictions on the portfolio weights. Section 2.5 uses computer simulations to gauge the statistical size and power properties of the SD tests relative to the GRS test. Subsequently, Finally, Section 2.6 gives concluding remarks and suggestions for further research. The Appendix gives formal proofs of our theorems.[4]

[4] The original paper included a section which analyzes if the CRSP all-share index (a popular proxy for the stock market portfolio) is efficient relative to benchmark portfolios formed on market beta. The idea is that by comparing the SD results with the results of the GRS test, we can determine if market risk other than variance plays a role in asset pricing. For this thesis, the results of this section has been incorporated into Chapter 4.

2.1 Preliminaries

We consider a single-period, portfolio-based model of investment that satisfies the following assumptions:

ASSUMPTION 1 *Investors are nonsatiable and risk-averse and they choose investment portfolios to maximize the expected utility associated with the return of their portfolios. Throughout the text, we will denote utility functions by* $u:\Re \rightarrow P$, $u \in U_2$*, with* U_2 *for the set of increasing and concave, continuously differentiable, von Neumann-Morgenstern utility functions, and* P *for a nonempty, closed and convex subset of* $\Re$.[5,6]

ASSUMPTION 2 *The investment universe consists of N risky assets and a riskless asset. Investors may diversify between the assets, and we will use* $\boldsymbol{\lambda} \in \Re^N$ *for a vector of portfolio weights. Positive weights reflect long positions and negative weights reflect short positions. If the weights sum to unity, or* $\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{1}_T = 1$*, then all wealth is invested in risky assets, and* $\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{1}_T < 1$ *and* $\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{1}_T > 1$ *refer to lending and borrowing (or long and short positions in the riskless asset) respectively. The evaluated portfolio is denoted by* $\boldsymbol{\tau} \in \Re^N$.

ASSUMPTION 3 *The excess return vector* $\boldsymbol{x} \in R^N$ *is a random vector with a continuous joint cumulative distribution function (CDF)* $G: R^N \rightarrow [0,1]$*, where* R *is a nonempty, bounded and*

---

[5] 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 $\boldsymbol{x}$.

[6] By contrast, Post (2003) does not assume that the utility function is continuously differentiable, so as to allow for, e.g., piecewise linear utility functions. However, in practice, we typically cannot distinguish between a kinked utility function and a smooth utility function with rapidly changing marginal utility. Nevertheless, using subdifferential calculus, we may obtain exactly the same results if utility is not continuously differentiable. Further, Post requires utility to be strictly increasing. To remain consistent with the original definition of SSD, we require a weakly increasing utility function. This is one of our reasons for adopting a new standardization for the gradient vectors.

convex subset of $\Re$.[7] *Our analysis will focus on the product term* $u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}$ *and we assume that the mean vector* $\boldsymbol{\alpha}(u) \equiv \int u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}dG(\boldsymbol{x})$ *is finite and the covariance matrix* $\boldsymbol{\Omega}(u) \equiv \int (u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x} - \boldsymbol{\alpha}(u))(u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x} - \boldsymbol{\alpha}(u))^{\mathrm{T}} dG(\boldsymbol{x})$ *is finite and positive-definite for all* $u \in U_2$.[8]

Under these assumptions, the investors' optimization problem can be summarized as $\max_{\boldsymbol{\lambda} \in \Re^N} \int u(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\lambda})dG(\boldsymbol{x})$. The evaluated portfolio $\boldsymbol{\tau} \in \Re^N$ is optimal for a given utility function $u \in U_2$ if and only if the first-order optimality condition, or Euler equation, is satisfied:

$$\boldsymbol{\alpha}(u) = \mathbf{0}_N \qquad (1)$$

Using the terminology that is common in the asset pricing literature, the marginal utility function $u'(x)$ represents a pricing kernel and $\boldsymbol{\alpha}(u)$ represents a vector of pricing errors. If $\alpha_i(u) > 0$, asset $i \in \mathrm{I}$ is undervalued and its weight in the portfolio should be increased relative to τ_i. Similarly, if $\alpha_i(u) < 0$, asset $i \in \mathrm{I}$ is overvalued and its weight in the portfolio should be decreased.

DEFINITION 1 *The evaluated portfolio* $\boldsymbol{\tau} \in \Re^N$ *is second-order stochastic dominance (SSD) efficient if and only if* $\boldsymbol{\alpha}(u) = \mathbf{0}_N$ *for some* $u \in U_2$. *The portfolio is SSD inefficient if and only if it is not optimal, that is,* $\boldsymbol{\alpha}(u) \neq \mathbf{0}_N$, *for all* $u \in U_2$.

This definition asks whether the evaluated portfolio is optimal for some admissible utility function $u \in U_2$. The traditional definition of SSD efficiency asks if there exists a portfolio $\boldsymbol{\lambda} \in \Re^N$ that is preferred to $\boldsymbol{\tau}$ for all utility functions $u \in U_2$. Both definitions are

[7] Since the portfolio possibilities are not restricted, portfolios with negative weights, or short positions, may yield returns outside the domain *R*. Still, for any given portfolio the returns will be bounded from above and from below.

[8] These general assumptions allow for general correlation and heteroskedasticity across the assets, contrary to the existing, pairwise SD tests, which invariantly require some sort of independence and/or homoskedasticity.

equivalent if the portfolio possibilities set is convex (see Post (2003), Theorem 1), as is true in our framework.

The objective of this study is to develop empirical tests for the null of efficiency, or $H_0 : \boldsymbol{\alpha}(u) = \mathbf{0}_N$, $u \in U_2$. We stress it is not our objective to estimate the (marginal) utility function that rationalizes the evaluated portfolio. The (marginal) utility functions used in our study are mere instruments in the analysis of the efficiency classification. One insurmountable problem in estimating the (marginal) utility function using our nonparametric approach is that the return range of the portfolio generally covers only a part of the domain of the (marginal) utility function. For example, we cannot possibly study the sensitivity to large losses of an investor who holds a well-diversified portfolio that never yields large losses without extrapolating from the observed return range.

To test the null, we need full information on the CDF $G(\boldsymbol{x})$.[9] In practical applications, $G(\boldsymbol{x})$ generally is not known and information is limited to a discrete set of T time series observations.

ASSUMPTION 4 *The observations are serially independently and identically distributed (IID) random draws from the CDF.*[10] *Throughout the text, we will represent the observations by the matrix* $\mathbf{X} \equiv (\boldsymbol{x}_1 \cdots \boldsymbol{x}_T)$*, with* $\boldsymbol{x}_t \equiv (x_{1t} \cdots x_{Nt})^{\mathrm{T}}$*. Since the timing of the draws is inconsequential, we are free to label the observations by their ranking with respect to the evaluated portfolio, that is,* $\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{\tau} < \boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{\tau} < \cdots < \boldsymbol{x}_T^{\mathrm{T}}\boldsymbol{\tau}$*.*

---

[9] By contrast, Post (2003) analyzes the sampling distribution under the null that all assets have the same mean return (or equivalently investors are risk-neutral and hence do not demand a risk premium). Clearly, this approach may lead to erroneous rejections of the "true null" of SSD efficiency in cases where the evaluated portfolio is efficient but the assets have different means, or equivalently, investors are risk-averse and therefore demand a risk premium.

[10] The assumption of a serial IID return distribution may be relaxed to a stationary and ergodic distribution, provided we correct for serial correlation and heteroskedasticity in addition to contemporaneous correlation and heteroskedasticity (see for example MacKinlay and Richardson (1991) in a mean-variance framework, and Linton *et al.*, 2005, in a pairwise SD framework). This seems especially useful for applications to high-frequency returns, where strong and predictable patterns of serial correlation and heteroskedasticity have been documented. However, for low-frequency returns, such patterns seem less important. Therefore, we use all available data to recover the patterns of contemporaneous correlation and heteroskedasticity, which generally are very strong and prediactble. We leave the issue of accounting for serial correlation and heteroskedasticity for high-frequency data for further research.

Using the observations, we can construct the following empirical distribution function (EDF):

$$F_{\mathbf{X}}(\boldsymbol{x}) \equiv T^{-1}\sum_{t=1}^{T} 1_{(\boldsymbol{x}_t \leq \boldsymbol{x})} \tag{2}$$

Furthermore, using the gradient vector $\nabla\boldsymbol{u} \equiv (u'(\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{\tau})\cdots u'(\boldsymbol{x}_T^{\mathrm{T}}\boldsymbol{\tau}))^{\mathrm{T}}$, we can construct the following sample counterparts of $\boldsymbol{\alpha}(u)$ and $\boldsymbol{\Omega}(u)$:[11]

$$\hat{\boldsymbol{\alpha}}(u) \equiv T^{-1}\mathbf{X}\nabla\boldsymbol{u} \tag{3}$$

$$\hat{\boldsymbol{\Omega}}(u) \equiv T^{-1}\left(\mathbf{X}\circ(\mathbf{1}_N\nabla\boldsymbol{u}^{\mathrm{T}}) - \hat{\boldsymbol{\alpha}}(u)\mathbf{1}_T^{\mathrm{T}}\right)\left((\nabla\boldsymbol{u}\mathbf{1}_N^{\mathrm{T}})\circ\mathbf{X}^{\mathrm{T}} - \mathbf{1}_T\hat{\boldsymbol{\alpha}}(u)^{\mathrm{T}}\right) \tag{4}$$

Since the observations are assumed to be serially IID draws from $G(\boldsymbol{x})$, $F_{\mathbf{X}}(\boldsymbol{x})$ is a consistent estimator for $G(\boldsymbol{x})$, and $\hat{\boldsymbol{\alpha}}(u)$ and $\hat{\boldsymbol{\Omega}}(u)$ converge in probability to $\boldsymbol{\alpha}(u)$ and $\boldsymbol{\Omega}(u)$ respectively. Furthermore, it follows from the Levy-Lindenberg central limit theorem that $\hat{\boldsymbol{\alpha}}(u)$ obeys an asymptotic multivariate normal distribution with mean $\boldsymbol{\alpha}(u)$ and covariance matrix $T^{-1}\boldsymbol{\Omega}(u)$, i.e., $\hat{\boldsymbol{\alpha}}(u) \xrightarrow{d} N(\boldsymbol{\alpha}(u), T^{-1}\boldsymbol{\Omega}(u))$. Finally, the sample covariance matrix $\hat{\boldsymbol{\Omega}}(u)$ obeys an asymptotic Wishart distribution with scale matrix $\boldsymbol{\Omega}(u)$ and degrees of freedom parameter (T-1). These properties are useful for statistical inference about H_0, as will be shown in the next section.

2.2 The SSD Test

A. THE TEST STATISTIC

We propose the following test statistic to test if $\boldsymbol{\tau} \in \Re^N$ is SSD efficient:

[11] Slightly departing from the conventional notation, the Hadamard operator "$\circ$" is used here for element-by-element multiplication of the elements of every column (row) of a matrix with a given column (row) vector.

$$J_2 \equiv \min_{u \in U_2^*} \hat{\boldsymbol{\alpha}}(u)^{\mathrm{T}} \hat{\boldsymbol{\Omega}}(u)^{-1} \hat{\boldsymbol{\alpha}}(u), \tag{5}$$

where

$$U_2^* \equiv \left\{ u \in U_2 : T^{-1} \nabla \boldsymbol{u}^{\mathrm{T}} \mathbf{1}_T = 1 \right\} \tag{6}$$

is the subset of SSD utility functions for which the sample mean of marginal utility (evaluated at the market return) equals unity.[12] The restriction on mean marginal utility standardizes utility such that the optimal solution is empirically distinguishable from the trivial solution $u(x) = 0$, which reflects the indifferent investor. Since utility functions are unique up to the level of a positive linear transformation, the standardization does not affect the efficiency classification, the p-values or critical values.

The test statistic J_2 is a variation to the classic Hotelling T^2 statistic used in multivariate statistical analysis. The objective function is a weighted sum of squares of the empirical pricing errors $\hat{\boldsymbol{\alpha}}(u)$, with the weights taken from the inverted covariance matrix $\hat{\boldsymbol{\Omega}}(u)^{-1}$. The higher the volatility or the correlation with the other errors, the lower the weight assigned to a given error (all other things remaining constant). This orientation is fundamentally different from the linear programming tests, which consider the largest positive error only.

Typical GMM tests of the Euler equation use a similar J-statistic. However, in contrast to our test, such tests use a parametrically specified utility function with a few—if any—unknown parameters. The Gibbons, Ross and Shanken (GRS, 1989) test for mean-variance efficiency can also be interpreted in this way. The GRS test is motivated by a normal return distribution. However, for a non-normal distribution, the GRS test statistic can also be interpreted as (a transform of) the J-statistic for a quadratic utility function with the parameters fixed by the sample mean and standard deviation of the evaluated portfolio.[13] Our

---

[12] By contrast, Post (2003) sets the sample minimum of marginal utility to equal unity, reflecting his assumption of strictly (rather then weakly) increasing utility. The present standardization is chosen to match the one (implicitly) used in standard studies in the literature on portfolio efficiency and asset pricing, including GRS.

[13] Rather than using an explicit utility function, the GRS test compares the Sharpe ratio of the market portfolio with that of the ex post tangency portfolio. This approach can be reformulated equivalently in terms of a

test deviates from this approach by allowing for non-quadratic utility and non-normal distributions in a nonparametric manner. Rather than using a parametrically specified utility function, the functional form of the utility function is left unspecified. Naturally, this approach will come at the cost of a loss of statistical power in small samples. Nevertheless, the utility function is restricted to be "economically meaningful", that is, it must obey nonsatiation and risk aversion. These restrictions will help to increase power relative to an unrestricted parametric approach.

B. STATISTICAL PROPERTIES

Based on the known distribution of the empirical pricing errors $\hat{\boldsymbol{\alpha}}(u)$ and sample covariance matrix $\hat{\boldsymbol{\Omega}}(u)$ for every admissible utility function $u \in U_2$, we can derive some useful statistical properties of our SSD efficiency test statistic.

THEOREM 1 *The SSD test statistic converges in probability to zero, that is,* $\lim_{T \to \infty} \Pr[J_2 > \varepsilon] = 0$, $\forall \varepsilon > 0$, *if and only if the evaluated portfolio is SSD efficient.*

Thus, our test is statistically consistent. This result may seem surprising given that the (marginal) utility functions are of infinite dimension. However, every admissible utility function can be represented by an elementary, representative one-parameter utility function that takes the form of the negative of the first-order lower partial moment (expected loss below a threshold); see Russell and Seo (1989) and Bowden (2005), among others. The proof to Theorem 1 in the Appendix uses this insight.

For the purpose of statistical inference in finite samples, we need to characterize the sampling distribution of the test statistic. Unfortunately, the imposed utility restrictions make it difficult to characterize the exact sampling distribution; this is a problem encountered in parametric models as well, since the restrictions may be binding or non-binding depending on the return distribution in an unknown manner. Nevertheless, it is possible to derive

quadratic utility function with parameter values determined by the sample mean and standard deviation of the evaluated portfolio.

conservative asymptotic p-values and critical values.[14] Specifically, we may derive the following result:

THEOREM 2 *The asymptotic null distribution of the SSD test statistic is bounded from above by a central chi-squared distribution with N degrees of freedom, that is,* $\Pr[J_2T > y|H_0] \leq 1-\chi^2_N(y)$.

This theorem can be used for conservative statistical inference. Specifically, the asymptotic *p*-value associated with the observed value of the test statistic is always smaller than or equal to $1-\chi^2_N(J_2T)$ and we can be at least $100\cdot\chi^2_N(J_2T)$ percent certain that efficiency is violated. Equivalently, the critical value is always smaller than or equal to $(\chi^2_N)^{-1}(1-\alpha)$. Thus, we can reject efficiency at a confidence level of at least $(1-\alpha)$ if $1-\chi^2_N(J_2T)\leq\alpha$, or if $J_2T \geq (\chi^2_N)^{-1}(1-\alpha)$.

This conservative approach is consistent with the convention of rejecting the null only if the p-value is smaller than a prespecified significance level. In fact, the empirical size (relative frequency of Type I error of wrongly rejecting efficiency) will be smaller than or equal to the nominal significance level α. Naturally, the obvious question is: How much statistical power does this conservative approach have? To answer this question, Section V will perform a simulation study. Interestingly, the results are very encouraging.

In contrast to our test, GRS derive the exact sampling distribution for their mean-variance statistic, but are only able to do so by assuming a normal return distribution and by treating the returns for the evaluated portfolio as exogenously fixed, leading to a conditional distribution. Also, they fix the model parameters using the sample mean and standard deviation of the evaluated portfolio and their model does not involve any free parameters. By contrast, our aim is to develop a test that can deal with non-normal distributions. Since the exact sampling distribution is not known, we resort to the asymptotic sampling distribution. Furthermore, we account for the sampling variation of the evaluated portfolio and consider the unconditional distribution rather than conditioning on the returns for the evaluated

[14] A conservative approach to statistical inference is not uncommon in non-parametric tests for optimizing behaviour; see e.g. Varian (1985). This is not surprising given the focus of non-parametric tests on minimizing type I error (wrongly rejecting the null).

portfolio. Finally, our model has free parameters and requires parameter restrictions to ensure a well-behaved kernel. To avoid the task of quantifying the effect of the free parameters and the restrictions, we resort to the conservative bounding distribution.

C. Computational Issues

At first sight, the test statistic ξ_2 seems computationally intractable, because the utility functions $u \in U_2^*$ are of infinite dimension. However, the gradient vector $\nabla \boldsymbol{u}$ is the only aspect of the utility function that is actually used for computing the test statistic. This gradient vector is of finite dimensions (*T*) and all admissible gradient vectors can be represented by the following *T*-dimensional polytope:

$$B_2 \equiv \left\{\nabla \boldsymbol{u} : u \in U_2^*\right\} = \left\{\boldsymbol{\beta} \in \Re_+^T : T^{-1}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{1}_T = 1; \Delta\beta_t \geq 0 \quad t = 2,\cdots,T\right\} \tag{7}$$

where $\Delta\beta_t \equiv \beta_t - \beta_{t-1}$. Given the ordering of the data (see Assumption 4), the restrictions $\Delta\beta_t \geq 0 \quad t = 2,\cdots,T$ suffice to impose the risk aversion condition. Using B_2, the test statistic can be reformulated as the following problem of finite dimensions:[15,16]

$$J_2 = \min_{\boldsymbol{\beta} \in B_2} T^{-2}(\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1}(\mathbf{X}\boldsymbol{\beta}) \tag{8}$$

An advantage of this formulation is that the pricing errors are linear functions of the model variables. Unfortunately, the weighting matrix $\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1}$ is a more complex function of the model variables. However, we can deal with the weighting matrix by using an iterative approach in the spirit of Hansen (1982), Ferson and Foerster (1994) and Hansen, Heaton, and Yaron (1996). Let

---

[15] In contrast to the earlier notation, the covariance matrix is now defined over the gradient vector rather than the utility function. Thus, $\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})$ should be understood as $\hat{\boldsymbol{\Omega}}(u) : \nabla \boldsymbol{u} = \boldsymbol{\beta}$.

[16] As discussed in Subsection A above, the problem can be simplified by analyzing only simple two-parameter utility functions that are defined over the mean return and a first-order LPM. However, this approach does not necessarily reduce the computational burden. To capture the threshold parameter we need to include a series of integer variables, one for every observed return level, to indicate if the threshold is located at a given return level.

$$J_2(\mathbf{W}) \equiv \min_{\boldsymbol{\beta} \in B_2} T^{-2} (\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}} \mathbf{W} (\mathbf{X}\boldsymbol{\beta}) \tag{9}$$

and

$$\boldsymbol{\beta}^*(\mathbf{W}) \equiv \arg\min_{\boldsymbol{\beta} \in B_2} T^{-2} (\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}} \mathbf{W} (\mathbf{X}\boldsymbol{\beta}) \tag{10}$$

with **W** for a prespecified positive-definite weighting matrix.

Computing $J_2(\mathbf{W})$ requires solving a quadratic objective function under linear constraints. This is a convex quadratic programming problem that can be solved using standard mathematical programming techniques. We can compute J_2 by solving $J_2(\mathbf{W})$ iteratively for various specifications of **W**. Solving the problem for an initial positive definite $\mathbf{W}_1$ will yield an initial consistent estimator $\boldsymbol{\beta}^*(\mathbf{W}_1)$ for the gradient vector, which can then be used to set $\mathbf{W}_2=\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta}^*(\mathbf{W}_1))^{-1}$ for the second iteration and/or to compute p-values. We may then stop and approximate J_2 by $J_2(\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta}^*(\mathbf{W}_2))^{-1})$ or conduct further iterations, each using $\mathbf{W}_s=\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta}^*(\mathbf{W}_{s-1}))^{-1}$, s = 3,4,…, possibly until convergence, to obtain a more efficient estimator. To reduce the computational burden of our simulations, we will use a two-stage estimator in this study.

Iterated GMM procedures frequently start with the identity matrix $\mathbf{W}_1=\mathbf{I}_N$ as the initial weighting matrix. In this study, we will however use $\mathbf{W}_1=\hat{\boldsymbol{\Omega}}(\mathbf{1}_T)^{-1}$, with $\hat{\boldsymbol{\Omega}}(\mathbf{1}_T) = T^{-1}\left(\mathbf{X} - T^{-1}\mathbf{X}\ \mathbf{1}_T\mathbf{1}_T^{\mathrm{T}}\right)\left(\mathbf{X}^{\mathrm{T}} - T^{-1}\mathbf{1}_T\mathbf{1}_T^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\right)$ for the covariance matrix of returns. In principle, every positive definite $\mathbf{W}_1$ will yield consistent first-stage estimates, because the alphas converge to zero if and only if the evaluated portfolio is efficient and every positive definite weighting matrix yields a zero value for the test statistic if and only if the alphas are zero. However, $\mathbf{I}_N$ is problematic in our context because it generally does not represent the inverted covariance matrix $\hat{\boldsymbol{\Omega}}(\nabla \boldsymbol{u})$ for an admissible utility function $u \in U_2$. By contrast, $\hat{\boldsymbol{\Omega}}(\mathbf{1}_T)$ corresponds to $\hat{\boldsymbol{\Omega}}(\nabla \boldsymbol{u})$ for the risk-neutral investor $u(x) = x$, which is admissible. Interestingly, the covariance matrix of returns $\hat{\boldsymbol{\Omega}}(\mathbf{1}_T)$ comes close to the second-moment matrix of returns $T^{-1}\left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right)$ proposed by Hansen and Jagannathan (1997, p.570) in the context of testing mean-variance efficiency.

The number of model variables in our test equals the number of time-series observations (T). At first sight, this seems to compromise the statistical consistency and power of our test. However, the model variables are restricted such that they represent the gradient vector of a well-behaved utility function. This substantially reduces the flexibility of the model variables. In fact, we know that the model can be reformulated in terms of a single parameter: the threshold rate of return of a representative utility function; see Section 2.2B and the proof to Theorem 1 in the Appendix. Unfortunately, the pricing errors are non-linear functions of this threshold parameter. By contrast, the pricing errors are linear in our model variables—the marginal utility levels. Thus, we effectively include T model variables to linearize pricing errors that are non-linear functions of a single parameter.

2.3 The TSD Test

We have thus far considered the SSD efficiency criterion associated with nonsatiable and risk-averse investors. It is relatively straightforward to generalize our results to the third-order stochastic dominance (TSD; Whitmore, 1970) criterion, which also assumes that investors prefer positively skewed return distributions.[17] Interestingly, empirical evidence suggests that investors indeed display this kind of skewness preference (e.g. Arditti, 1967, Kraus and Litzenberger, 1976, Cooley, 1977, Friend and Westerfield, 1980, and Harvey and Siddique, 2000). Let $U_3 \subset U_2$ represent all nonsatiable and risk-averse investors with a convex marginal utility function (skewness preference).

DEFINITION 2 *The evaluated portfolio* $\boldsymbol{\tau} \in \Re^N$ *is third-order stochastic dominance (TSD) efficient if and only if* $\boldsymbol{\alpha}(u) = \mathbf{0}_N$ *for some* $u \in U_3$. *The portfolio is TSD inefficient if and only if it is not optimal, that is,* $\boldsymbol{\alpha}(u) \neq \mathbf{0}_N$, *for all* $u \in U_3$.

By analogy to (5), we may use the following test statistic to test if the given portfolio $\boldsymbol{\tau}$ is TSD efficient:

[17] As discussed in the concluding remarks, generalizations to SD rules of order four and higher are also possible. However, these higher-order rules are not discussed here because they involve little additional power (fourth-order), or have, as far as we are aware, no economic justification (fifth-order and higher).

$$J_3 \equiv \min_{u \in U_3^*} \hat{\boldsymbol{\alpha}}(u)^{\mathrm{T}} \hat{\boldsymbol{\Omega}}(u)^{-1} \hat{\boldsymbol{\alpha}}(u), \tag{11}$$

with

$$U_3^* \equiv \left\{u \in U_3 : T^{-1} \nabla \boldsymbol{u}^{\mathrm{T}} \mathbf{1}_T = 1\right\} \tag{12}$$

It is straightforward to demonstrate that the statistical properties derived for J_2 in Section 2.2 also apply for J_3:[18]

THEOREM 3 *If and only if the evaluated portfolio is TSD efficient, the TSD test statistic converges in probability to zero, that is,* $\lim_{T \to \infty} \Pr[J_3 > \varepsilon] = 0$, $\forall \varepsilon > 0$.

THEOREM 4 *The asymptotic null distribution of the TSD test statistic is bounded from above by a chi-squared distribution with N degrees of freedom, that is,* $\Pr[J_3 T > y | H_0] \leq 1 - \chi_N^2(y)$.

Thus, the same conservative asymptotic p-values and critical values that apply for J_2 also apply for J_3. Of course, $U_3 \subset U_2$ and hence $J_2 \leq J_3$ and the TSD test will yield higher rejection rates.

By analogy to (8), we may reformulate J_3 as follows:

$$J_3 = \min_{\boldsymbol{\beta} \in B_3} T^{-2} (\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}} \hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1} (\mathbf{X}\boldsymbol{\beta}), \tag{13}$$

with

$$\mathrm{B}_3 \equiv \left\{\boldsymbol{\beta} \in B_2 : \Delta\beta_t (\Delta\mathbf{x}_t^{\mathrm{T}} \boldsymbol{\tau})^{-1} \geq \Delta\beta_{t-1} (\Delta\mathbf{x}_{t-1}^{\mathrm{T}} \boldsymbol{\tau})^{-1} \quad t = 3, \cdots, T\right\} \tag{14}$$

B_3 is the subset of the positive and decreasing gradients (B_2) for which marginal utility is convex. The term $\Delta\beta_t (\Delta\mathbf{x}_t^{\mathrm{T}} \boldsymbol{\tau})^{-1}$ can be regarded as an approximation to the second-order

[18] Proofs are omitted for the sake of brevity.

derivative $u''(z_t)$ at $z_t \equiv 0.5(\boldsymbol{x}_{t-1}^{\mathrm{T}}\boldsymbol{\tau} + \boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau})$. This term needs to be increasing in order to guarantee skewness preference.

By analogy to the procedure described in Section 2.2C, the minimization problem (13) can again be solved by iteratively solving the embedded convex quadratic programming problem $\min_{\boldsymbol{\beta} \in B_3} T^{-2}(\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{W}(\mathbf{X}\boldsymbol{\beta})$ for various specifications of the weighting matrix $\mathbf{W}$.

2.4 Investment Restrictions

We have thus far assumed that the investors face no restrictions on the portfolio weights, or $\boldsymbol{\lambda} \in \Re^N$; see Assumption 2. Many capital market theories assume a perfect capital market without restrictions. Furthermore, when analyzing the value-weighted market portfolio, many restrictions such as short selling restrictions are not binding, because the market portfolio assigns a strictly positive weight to every risky asset. Still, for analyzing actual investment portfolios, it is useful to construct efficiency tests that do account for practical investment restrictions. Such tests are relevant also for restrictions that involve the market portfolio of risky assets, such as restrictions on riskless lending and borrowing.

In this section, we generalize our results to the case with general linear weight restrictions. Many practical investment constraints have a linear form, including short-selling constraints, position limits and restrictions on risk factor loadings (or "betas"). Furthermore, non-linear restrictions often can be approximated with high precision by a set of linear restrictions. With restrictions, the portfolio possibilities can be represented by a polyhedron of general form

$$\Lambda \equiv \left\{\boldsymbol{\lambda} \in \Re^N : \mathbf{A}_1\boldsymbol{\lambda} \le \mathbf{b}_1, \mathbf{A}_2\boldsymbol{\lambda} = \mathbf{b}_2\right\} \tag{15}$$

with $\mathbf{A}_1 \in \Re^{M_1 \times N}$ and $\mathbf{b}_1 \in \Re^{M_1}$ for the coefficients of M_1 linear inequality restrictions and $\mathbf{A}_2 \in \Re^{M_2 \times N}$ and $\mathbf{b}_2 \in \Re^{M_2}$ for the coefficients of M_2 linear equality restrictions. Two interesting special cases are (i) the case with no riskless asset, or $\Lambda = \left\{\boldsymbol{\lambda} \in \Re^N : \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{1}_N = 1\right\}$, and (ii) the case with no short selling allowed, or $\Lambda = \left\{\boldsymbol{\lambda} \in \Re_+^N : \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{1}_N \le 1\right\}$.

The first-order Karush-Kuhn-Tucker (KKT) optimality conditions for the generalized investment problem $\max_{\boldsymbol{\lambda} \in \Lambda} \int u(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\lambda})dG(\boldsymbol{x})$ are:

$$\boldsymbol{\alpha}(u) - \mathbf{A}_1'^{*}\boldsymbol{\rho}_1 + \mathbf{A}_2'\boldsymbol{\rho}_2 = \mathbf{0}_N \tag{16}$$

with $\mathbf{A}_1^*$ for a $(M_1^* \times N)$ matrix constructed from the rows of $\mathbf{A}_1$ for which the inequality $\mathbf{A}_1\boldsymbol{\lambda} \leq \mathbf{b}_1$ is binding, and $\boldsymbol{\rho}_1 \in \Re_+^{M_1^*}$ and $\boldsymbol{\rho}_2 \in \Re^{M_2}$ for Lagrange multipliers. Nonbinding inequality restrictions are not included, because the Lagrange multipliers must equal zero in this case. Thus, such restrictions do not affect the KKT conditions for the alphas.

Thus, in contrast to the unrestricted case (1), non-zero alphas generally are allowed in the case with restrictions. The admissible alphas can be described by the following polyhedral cone:

$$C \equiv \{\boldsymbol{z} \in \Re^N : \boldsymbol{z} = \mathbf{A}_1'^{*}\boldsymbol{\rho}_1 - \mathbf{A}_2'\boldsymbol{\rho}_2, (\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) \in \Re_+^{M_1^*} \times \Re^{M_2}\} \tag{17}$$

Thus, portfolio $\boldsymbol{\tau} \in \Lambda$ is efficient if and only if the alphas lie within the cone, that is, $H_0 : \boldsymbol{\alpha}(u) \in C$ for some admissible utility function. In the case with no binding restrictions, that is, $\mathbf{A}_1^* = \varnothing$ and $\mathbf{A}_2 = \varnothing$, we find $C = \mathbf{0}_N$ and the null reduces to the Euler equation (1).

To test efficiency under restrictions, we will use the following generalization of (8) and (13):

$$J_i^R \equiv \min_{\substack{\boldsymbol{\beta} \in B_i \\ \boldsymbol{z} \in C}} (T^{-1}(\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}} - \boldsymbol{z})^{\mathrm{T}} \hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1}(T^{-1}\mathbf{X}\boldsymbol{\beta} - \boldsymbol{z}) \tag{18}$$

with $i \in \{2,3\}$ for the order of the relevant SD rule. The test statistic measures the smallest distance between the errors and the cone of admissible errors.

The computational strategy that was used for the unrestricted case also applies here. For a given the weighting matrix $\mathbf{W}$, the embedded minimization problem $\min_{\substack{\boldsymbol{\beta} \in B_i \\ \boldsymbol{z} \in C}} (T^{-1}(\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}} - \boldsymbol{z})^{\mathrm{T}} \mathbf{W}(T^{-1}\mathbf{X}\boldsymbol{\beta} - \boldsymbol{z})$ is a convex quadratic programming problem and the entire problem can be solved by iteratively solving this problem using the optimal solution from the previous iteration to specify the relevant weighting matrix.

Compared with the unrestricted case, determining the sampling distribution of the test statistic is further complicated by the M_1^* binding inequalities, for which the Lagrange multipliers are restricted to be nonnegative. However, we can obtain an upper bound for the sampling distribution using the known effect of the M_2 equalities (which involve unrestricted Lagrange multipliers). If the binding inequality restrictions were dropped and $\boldsymbol{\beta}$ were fixed at the gradient vector of a given utility function that rationalizes the evaluated portfolio, the resulting test statistic $\min_{z \in C}(T^{-1}(\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}} - z)^{\mathrm{T}} \hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1}(T^{-1}\mathbf{X}\boldsymbol{\beta} - z)$ would obey a central chi-square distribution with $(N - M_2)$ degrees of freedom; every equality restriction lowers the number of degrees of freedom by one. Since the minimization across Lagrange multipliers for the binding inequality restrictions and across gradient vectors lowers the test statistic, $J_i^R T$ is bounded from above by this distribution.

2.5 Simulation

The existing SSD efficiency test suffers from low power in typical empirical applications, as demonstrated in the simulation experiment of Post (2003, Section IIIC) based on the returns of the well-known 25 Fama and French stock portfolios formed on market capitalization and book-to-market-equity ratio. In part, the lack of power reflects the difficulty of estimating a 25-dimensional return distribution. It is likely that the power increases (at an increasing rate) as the length of the cross-section is reduced to for example ten benchmark portfolios, which is a common choice in asset pricing tests. Furthermore, as discussed above, the lack of power may reflect the focus on the maximum positive pricing error. The new SSD efficiency test accounts for all errors and is expected to involve substantially more power. Furthermore, the TSD efficiency test imposes skewness preference and hence is expected to be even more powerful. To shed some light on the statistical properties of our tests in finite samples, this section reports the results of a simulation experiment.

A. Simulation Setup

Instead of the 25 Fama and French portfolios, this study uses ten single-sorted portfolios formed on market beta. We focus on these portfolios for two reasons. First, sorting stocks on beta maximizes the spread in betas and hence reduces the probability of erroneous rejection of the null of mean-variance efficiency (Type I error). Second, time-variation of the return distribution can severely bias the results of unconditional asset pricing tests (see for instance Jagannathan and Wang, 1996).[19] Hence, the large sample properties of our tests

---

[19] One approach to correct for time-variation is to use a conditional efficiency test; see for example, Jagannathan and Wang (1996). Unfortunately, conditional tests are no panacea, because economic theory gives minimal

apply only to benchmark portfolios for which long, stationary samples are available in practice. Unfortunately, the 25 Fama and French portfolios seem severely affected by cyclical time-variation (see for example Lettau and Ludvigson, 2001). By contrast, beta portfolios by construction have a more stable distribution, as a stock migrates to another benchmark portfolio if its beta changes significantly through time.

Panel A of Table I gives descriptive statistics for the monthly returns of the beta decile portfolios in the sample from January 1933 to December 2002 (840 months). The skewness and kurtosis statistics suggest that the returns do not obey a normal distribution. Nevertheless, in the simulations, we use a normal distribution with joint population moments equal to the first two sample moments of the portfolios. This means that we effectively take away the rationale for using SD criteria rather than the mean-variance criterion. For a normal distribution, the SD criteria reduce to the mean-variance criterion. Thus, we first analyze the statistical properties of our SD efficiency tests under relatively unfavorable conditions where the tests are necessarily inferior to mean-variance tests.

---

guidance for modelling the evolution of investor preferences and the return distribution, which creates a serious risk of specification error. Second, it is difficult to ensure that conditional tests obey the regularity conditions of nonsatiation and risk aversion for all possible states-of-the-world. Failure to ensure that the utility function is well behaved can seriously reduce the power of a conditional test.

Table I

Descriptive Statistics Benchmark Portfolios

The table shows descriptive statistics for the benchmark portfolios formed on market beta (Panel A) and the monthly excess returns of the CRSP index, as well as the EP, TP and LP test portfolios constructed for our simulations (Panel B). The reported kurtosis is the excess kurtosis. The beta portfolios are constructed from the CRSP tapes. In December of each year, all stocks that fulfill our data requirements are placed in ten portfolios based on the previous 60-month betas. A minimum of 12 months of return observations is needed for a stock to be included on formation date. Each portfolio includes an equal number of stocks. The sample period runs from January 1933 to December 2002 (T=840). Excess returns are computed from the raw return observations by subtracting the return on the one-month US Treasury bill from Ibbotson. We thank Pim van Vliet for making the data available. All data described in Panel A can be found at his online datacenter: http://www.few.eur.nl/few/people/wvanvliet/datacenter.

Panel A : The 10 benchmark portfolios						
	Mean	Stdev.	Skewness	Kurtosis	Min	Max
Low β	0.670	3.822	-0.754	5.230	-24.577	15.718
2	0.698	4.015	-0.018	3.926	-20.573	24.222
3	0.756	4.631	0.648	10.175	-25.003	41.292
4	0.659	4.832	0.255	6.269	-25.943	34.332
5	0.918	5.669	1.041	13.370	-29.333	55.762
6	0.833	6.094	0.592	8.279	-28.615	48.932
7	0.809	6.538	0.574	8.773	-32.573	53.842
8	0.768	7.470	0.774	9.264	-30.395	61.832
9	0.833	8.306	0.689	7.941	-36.583	64.262
High β	0.794	9.653	0.814	8.516	-37.133	83.692
Panel B : The CRSP market index and the test portfolios						
	Mean	Stdev.	Skewness	Kurtosis	Min	Max
Market	0.714	4.937	0.156	6.181	-23.673	38.172
EP	0.774	5.699	0.560	9.025	-28.020	47.953
TP	0.960	4.264	-0.022	4.571	-21.870	27.730
LP	0.960	5.721	1.361	9.156	-15.139	53.311

After analyzing the normal case, we will turn to the empirical, non-normal return distribution that is also used in Section VI below. This allows us to gauge the added value of the SD efficiency tests in cases where the mean-variance criterion is not consistent with expected utility theory.

We will compare the statistical properties of three alternative test procedures: (1) our SSD efficiency test, (2) our TSD efficiency test and (3) the GRS test for mean-variance efficiency. As discussed in Section 2.2A, for a non-normal distribution, the GRS test can be interpreted as a test for the Euler Equation (1) for a quadratic utility function with parameters fixed by the sample mean and standard deviation of the evaluated portfolio. Thus, compared to the SD efficiency tests, the GRS test imposes a specific functional form (normal distribution or quadratic utility) and fixes the model parameters without explicit use of the regularity conditions.

By comparing the SSD and TSD efficiency tests, we analyze the additional power obtained by imposing skewness preference, which increases the rejection rate. Comparing the TSD and GRS tests will show the loss of power due to using a nonparametric approach, which reduces the rejection rate. In principle, the SD tests can be even more powerful than the GRS test, because the latter does not explicitly impose the regularity conditions. However, in practice the effect of relaxing the functional form is of greater importance, as we will see below.

We will first apply the three procedures to two test portfolios in random samples drawn from the multivariate normal population distribution. The equal weighted portfolio (EP) is known to be SSD, TSD and mean-variance inefficient relative to the normal population distribution.[20] Hence, we may analyze the statistical power of the competing test procedures by their ability to correctly classify EP as inefficient. By contrast, the ex ante tangency portfolio (TP) is SSD efficient and we may analyze the empirical size by the relative frequency of random samples in which this portfolio is wrongly classified as inefficient.

Next, we will move to the empirical distribution, and TP (used to gauge statistical size) is replaced with an alternative portfolio that is SSD and TSD efficient but mean-variance inefficient, resulting in a portfolio that can illustrate the added value of the SD tests. This portfolio, is selected by minimizing the second-order lower partial moment (LPM):

$$LPM_2 \equiv \frac{1}{T}\sum_{t=1}^{T}\max(m - x_t, 0)^2 \tag{19}$$

with a target rate of return (m) of -10%, subject to the mean return being equal to that of TP.[21] In case of a normal distribution, minimizing the LPM_2 is equivalent to minimizing variance

[20] It is possible to achieve a substantially higher mean given the standard deviation of EP and hence EP is mean-variance inefficient. Since we assume a normal distribution in the simulations, the SSD and TSD criteria coincide with the mean-variance criterion and EP is also SSD and TSD inefficient.

[21] We will refer to this portfolio as the LPM portfolio, or LP. Lower partial moments are generally considered more meaningful measures of risk than the variance, because variance treats upside and downside deviations in a symmetric way, which is inconsistent with decreasing absolute risk aversion. By contrast, LPMs focus on the downside exclusively, and may use another reference point than the mean to measure deviations. For a thorough treatment of LPMs and their relation with the SD rules, see for instance Bawa (1975).

and LP equals TP. The differences between TP and LP are therefore a direct consequence of the nonnormality of the empirical distribution. LP clearly lies in the interior of the mean-variance frontier. However, this portfolio involves a thin left tail and a fat right tail (see Table I, Panel B), and hence is much less risky than suggested by its variance. Panel B of Table I gives descriptive statistics for the EP, TP and LP test portfolios, as well as the market index used in Section V. Figure 1 further illustrates our simulation conditions by means of a mean-standard deviation diagram.

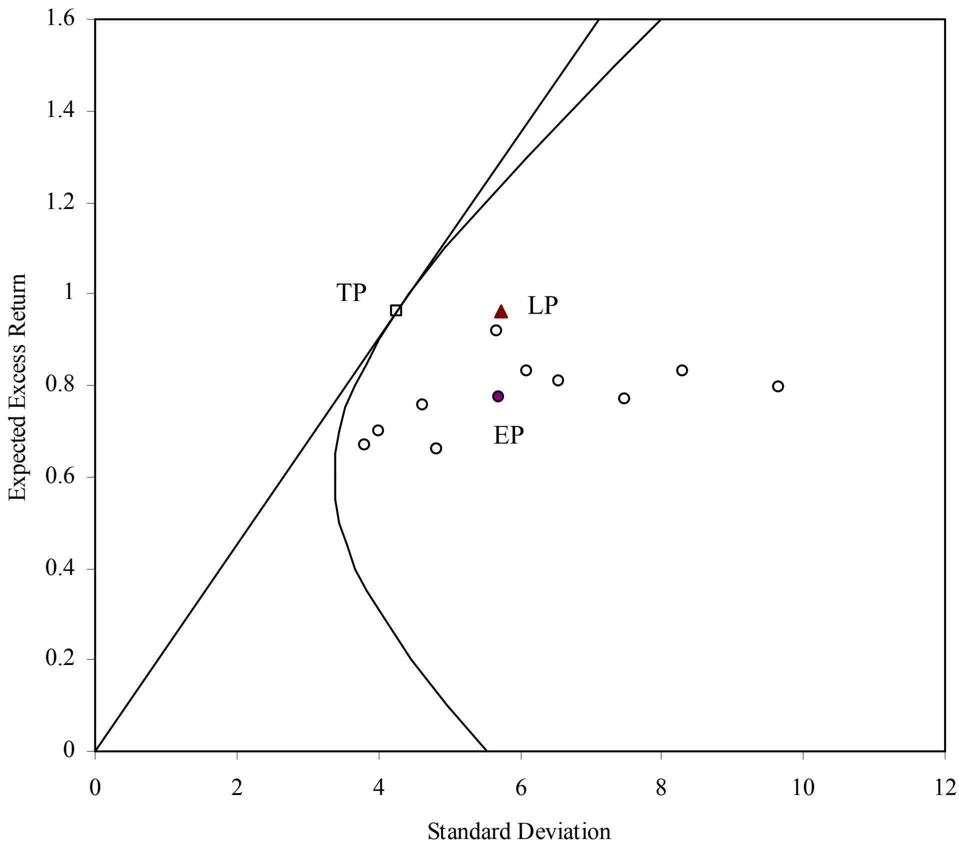

Figure 1

Mean-Standard Deviation Diagram

The figure shows the mean-standard deviation diagram based on the monthly excess returns of the ten beta portfolios in the sample from January 1933 to December 2002 (840 months). The diagram includes the individual benchmark portfolios (the clear dots), the equal weighted test portfolio (EP, the filled dot), the tangency portfolio (TP, the clear square), the lower partial moment portfolio (LP) and the efficient frontier, all allowing for short selling.

B. THE NORMAL DISTRIBUTION

In the first part of our simulations, we draw 10,000 random samples from the multivariate normal population distribution through Monte-Carlo simulation. For every random sample, we apply all three test procedures to both test portfolios. This experiment is performed for a sample size (T) of 50 to 2,000 observations and for a significance level (α) of 2.5, 5, and 10 percent.

Table II summarizes the results. The size of the two SD tests exceeds the nominal significance level for small samples, which reflects that the asymptotic p-values apply only for large samples. In small samples, estimation error for the return distribution increases the pricing errors above the expected level based on the asymptotic distribution. As the sample size increases, the estimate of the return distribution becomes more accurate. Hence, the pricing errors fall at a higher rate than given by the asymptotic rule of T^{-1}, causing the size to fall. Ultimately, the size converges to levels below the nominal significance level as the number of time series observations (T) increases, for example 0.8% for the SSD test for α=10% and T=1000. This reflects the use of conservative p-values and critical values, which do not account for the number of model parameters and the parameter restrictions (see Section 2.2). In contrast, the GRS test does not place restrictions on the parameters and the exact sampling distribution of the test statistic is known. Indeed, the table shows that the size of the GRS test equals the nominal significance level, save some simulation error. Of course, for a non-normal distribution, the size of the GRS test generally will deviate from the nominal significance level (see Part C below).

Table II

Statistical properties of the competing test procedures under the normal distribution

The table displays the size and power for various numbers of time-series observations (T) and for various levels of significance (α). Panel A represents the new SSD test, while panel B gives the results for the new TSD test, and Panel C shows the results of the GRS test for mean-variance efficiency. The results are based on 10,000 random samples from a multivariate normal distribution with joint moments equal to the sample moments of the monthly excess returns of the ten beta portfolios for the period from January 1933 to December 2002. Size is measured as the relative frequency of random samples in which the efficient tangency portfolio (TP) is wrongly classified as inefficient. Power is measured as the relative frequency of random samples in which the inefficient equally weighted portfolio (EP) is correctly classified as inefficient.

Panel A: SSD						
	Size			Power		
	α=10%	α=5%	α=2.5%	α=10%	α=5%	α=2.5%
T=50	0.118	0.075	0.050	0.208	0.141	0.101
T=100	0.050	0.027	0.015	0.183	0.117	0.075
T=200	0.023	0.010	0.006	0.274	0.185	0.126
T=500	0.014	0.006	0.003	0.641	0.532	0.430
T=1,000	0.008	0.003	0.001	0.940	0.893	0.844
T=2,000	0.007	0.004	0.000	0.999	0.991	0.989
Panel B: TSD						
	Size			Power		
	α=10%	α=5%	α=2.5%	α=10%	α=5%	α=2.5%
T=50	0.235	0.161	0.114	0.322	0.240	0.175
T=100	0.122	0.073	0.042	0.289	0.199	0.136
T=200	0.072	0.037	0.020	0.400	0.288	0.215
T=500	0.054	0.027	0.012	0.765	0.668	0.563
T=1,000	0.035	0.019	0.010	0.974	0.953	0.930
T=2,000	0.020	0.010	0.007	1.000	1.000	1.000
Panel C: GRS						
	Size			Power		
	α=10%	α=5%	α=2.5%	α=10%	α=5%	α=2.5%
T=50	0.101	0.052	0.027	0.179	0.102	0.059
T=100	0.107	0.057	0.029	0.284	0.183	0.117
T=200	0.104	0.053	0.029	0.494	0.370	0.274
T=500	0.102	0.054	0.031	0.885	0.816	0.735
T=1,000	0.109	0.059	0.030	0.997	0.993	0.985
T=2,000	0.109	0.056	0.029	1.000	1.000	1.000

For all three procedures, the statistical power goes to unity as we increase the number of time series observations. However, in small samples, the tests may lack power. Clearly, we cannot expect reliable non-parametric estimates of a 10-dimensional return distribution based on only a few observations. The good news is that the new SSD procedure is substantially more powerful than the existing linear programming procedure of Post (2003); see Figure 4 in his article.[22] Evidently, accounting for all pricing errors rather than the maximum positive error only substantially improves the power of the SSD test. Ignoring the results for very small samples (where the size is poor), the power of the new SSD test remains below that of the GRS procedure. As expected, the effect of using a nonparametric approach (which lowers the rejection rates) outweighs the effect of imposing the regularity conditions (which raises the rejection rates). Nevertheless, the power remains at a level that is acceptable for empirical research.

The TSD procedure, which imposes skewness preference in addition to nonsatiation and risk aversion, substantially improves power relative to the SSD test. In fact, the power of the TSD test is comparable to that of the GRS test. For example, for T=200 and α=10%, the power of the TSD test is 40.0%, which is reasonably close to the corresponding figure of 49.4% for the GRS test. Since the TSD test in contrast to the GRS test also applies for non-normal distributions, the relatively small loss in power is particularly encouraging.

C. THE EMPIRICAL DISTRIBUTION

The results for the normal distribution are silent on the added value of the SD tests in the important case where the return distribution is not normal. For this reason, we repeat the above simulations for the empirical, non-normal distribution; see Table III. Rather than sampling from a normal population distribution, we now take random samples with replacement from the original dataset $\{\boldsymbol{x}_t\}_{t=1}^{T}$. Recall that we also replace the mean-variance tangency portfolio TP with LP, the portfolio that minimizes the LPM_2 (rather than variance) given the mean of the TP. Since LP is known to be SSD and TSD efficient but mean-variance inefficient, it is ideal for gauging the added value of the SD tests in cases where the GRS test is inappropriate.

Interestingly, the power of the three tests for the empirical distribution is comparable

[22] In fact, the power of the SSD-LP test (not reported here in order to keep the presentation compact) in our present setup is worse than reported in Post (2003), which uses a different dataset. Presumably, this is caused by the large negative pricing errors for the high-beta portfolios, which are ignored by the SSD-LP procedure.

to that for the normal distribution. We again see that only limited harm is done in terms of power by using SD tests. By contrast, the size results are very different from the normal case.

We first focus on the size of the SSD and TSD tests. For both tests, the size level is substantially higher than under the normal distribution. This reflects that estimation error is more important for the tail of the distribution (which determines the efficiency classification of LP) than for the means and covariance terms (which are relevant under the normal distribution). In other words, the sample tail frequently differs strongly from the population tail and differences generally lead to rejections of efficiency for LP (Type I error). The SSD test is less sensitive to this problem than the TSD test, because it has more flexibility to fit a non-convex function to the tail observations.

Apart from a higher size level, the size function also falls at a much lower rate than under the normal distribution. In fact, the size increases locally and achieves a local maximum between T=500 and T=1000. At first sight, this finding seems counterintuitive. However, the increasing size does not mean that the test statistic actually increases as the sample size increases. Rather, it means that the test statistic falls at a lower rate than the critical value. This is possible because the tail of the distribution (which determines the efficiency classification of LP) converges at a lower rate than the means and covariance terms (which determine the asymptotic distribution). This effect at some point outweighs the effect of higher accuracy for the means and covariance terms in a larger sample.

Despite the high level and the local increase of the size function, the size function falls again as T becomes large and drops below the nominal significance level in large enough samples (reflecting the use of conservative p-values).

The GRS test yields markedly different results: its size is much higher than for the SD tests and approaches unity already for medium-sized samples. For example, for T=500, the GRS test has a size of 74.3%, while the SSD and TSD tests have a size of 19.8% and 33.4% respectively. These differences are substantially larger than the differences in power between the SD tests and the GRS test. Thus, the theoretical advantage of the SD tests—considering the entire return distribution rather than mean and variance alone—is also present in real-life applications: for samples from a realistic return distribution and with a realistic sample size, the gain in size seems to outweigh the loss in power.

Table III

Statistical properties of the competing test procedures under the empirical distribution

The table displays the size and power for various numbers of time-series observations (T) and for various levels of significance (α) using the empirical, non-normal distribution as the population distribution. The results are based on the original dataset of monthly returns of the ten beta portfolios from January 1933 to December 2002. We drew 10,000 random samples with replacement. Panel A represents the new SSD test, while panel B gives the results for the new TSD test, and Panel C shows the results of the GRS test for mean-variance efficiency. Size is measured as the relative frequency of random samples in which the SD-efficient lower partial moment portfolio (LP) is wrongly classified as inefficient. Power is measured as the relative frequency of random samples in which the inefficient equally weighted portfolio (EP) is correctly classified as inefficient.

Panel A: SSD						
	Size			Power		
	α=10%	α=5%	α=2.5%	α=10%	α=5%	α=2.5%
T=50	0.210	0.145	0.105	0.233	0.164	0.117
T=100	0.168	0.108	0.070	0.208	0.137	0.090
T=200	0.174	0.111	0.071	0.280	0.190	0.127
T=500	0.198	0.135	0.092	0.583	0.473	0.373
T=1,000	0.178	0.129	0.090	0.876	0.811	0.746
T=2,000	0.100	0.075	0.066	0.993	0.986	0.975
Panel B: TSD						
	Size			Power		
	α=10%	α=5%	α=2.5%	α=10%	α=5%	α=2.5%
T=50	0.350	0.263	0.196	0.365	0.271	0.201
T=100	0.275	0.192	0.135	0.345	0.242	0.173
T=200	0.293	0.208	0.148	0.445	0.336	0.251
T=500	0.334	0.249	0.185	0.739	0.642	0.540
T=1,000	0.320	0.251	0.211	0.931	0.896	0.826
T=2,000	0.241	0.198	0.182	0.994	0.993	0.992
Panel C: GRS						
	Size			Power		
	α=10%	α=5%	α=2.5%	α=10%	α=5%	α=2.5%
T=50	0.170	0.090	0.048	0.176	0.098	0.055
T=100	0.258	0.163	0.102	0.302	0.195	0.124
T=200	0.406	0.295	0.208	0.511	0.386	0.290
T=500	0.743	0.644	0.541	0.885	0.817	0.737
T=1,000	0.963	0.931	0.890	0.996	0.990	0.981
T=2,000	1.000	0.999	0.998	1.000	1.000	1.000

2.6 Concluding Remarks

In this chapter, we developped multivariate statistical tests for second-order and third-order stochastic dominance efficiency of a given portfolio. In contrast to the existing linear programming tests, our tests consider all pricing errors jointly, rather than the maximum positive error only. The test statistics can be computed by iterating a convex quadratic programming problem that can be solved using standard mathematical programming techniques. We have also shown how to conduct statistical inference using conservative asymptotic p-values and critical values derived from a chi-squared distribution. Our simulations show that this approach has superior statistical power properties compared to the existing linear programming approach. In fact, the power of our TSD test is comparable with that of the GRS test in both the case of a normal distribution and for the empirical distribution. The small loss in power seems an acceptable price for the superior statistical size for non-normal distributions, where the GRS test does not apply due to the inadequacy of the mean-variance framework. We have also shown how the tests can be extended to account for general linear restrictions on the portfolio weights.

A natural extension of our SSD and TSD tests is a test for fourth-order stochastic dominance (4SD), which assumes kurtosis aversion in addition to non-satiation, risk aversion and skewness preference. Kurtosis aversion is a necessary condition for decreasing absolute prudence (Kimball, 1990). Arguably, any realistic utility function should exhibit this latter property, which says that bearing one risk should make an investor less willing to bear independent other risks. Kurtosis aversion boils down to requiring the second-order derivative of utility to be concave (or the third-order derivative to be decreasing). Like risk aversion and skewness preference, this restriction can be imposed by means of linear restrictions on the gradient vector.[23] Adding these restrictions to those already imposed in the TSD tests yields a 4SD test. Unfortunately, the resulting test gives somewhat disappointing results when included in our simulation experiment; the increase in power when moving from TSD to 4SD is relatively small compared to the difference between SSD and TSD. For this reason, we have not included 4SD in this study. SD criteria of order five and higher are also not included, because we are not aware of convincing economic arguments for these criteria.

---

[23] Specifically, we need the restrictions $\Delta(\Delta\beta_t(\Delta\mathbf{x}_t^{\mathrm{T}}\boldsymbol{\tau})^{-1})(\Delta z_t)^{-1} \leq \Delta(\beta_{t-1}(\Delta\mathbf{x}_{t-1}^{\mathrm{T}}\boldsymbol{\tau})^{-1})(\Delta z_{t-1})^{-1} \quad t = 4,\cdots,T$ to ensure that the second-order derivative is concave. In these restrictions, $\Delta(\Delta\beta_t(\Delta\mathbf{x}_t^{\mathrm{T}}\boldsymbol{\tau})^{-1})(\Delta z_t)^{-1}$ serves as a proxy for the third-order derivative $u'''(0.5(z_{t-1}+z_t))$.

We hope that our results contribute to the further proliferation of the stochastic dominance methodology. As discussed in the introductory section, the dominant mean-variance methodology generally is not economically meaningful; it may fail to detect inefficiency for portfolios that no nonsatiable risk-averter would select (especially if the portfolios have relatively fat left tails) and may reject efficiency for portfolios that are perfectly good solutions for some nonsatiable risk-averters (especially if the portfolios have relatively thin left tails). Furthermore, empirical finance seems a particularly fertile ground for the stochastic dominance methodology, because the large, high quality data sets that are available allow the "data to speak for themselves" in a nonparametric fashion.

Appendix

Proof of THEOREM 1 To prove consistency, we first simplify the problem by (i) replacing the weighting matrix $\hat{\mathbf{\Omega}}(u)^{-1}$ with an identity matrix $\mathbf{I}_N$ and (ii) replacing the utility functions $u \in U_2$ with elementary one-parameter utility functions $u_z(x) \equiv p(z)^{-1}(x-z)1_{x\leq z}$, for some threshold $z \in Z$, $Z \equiv (\underline{z}, \bar{z}]$. Here, $\underline{z} > -\infty$ and $\bar{z} < \infty$ bound the return interval for the evaluated portfolio, i.e., $\Pr[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau} < \underline{z}] = 0$ and $\Pr[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau} > \bar{z}] = 0$, and $p(z) = E[u'_z(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})] = E[1_{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau}\leq z}]$ is selected to ensure that the expected value of marginal utility $u'_z(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau}) = p(z)^{-1}1_{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau}\leq z}$ is one.

The replacement of the weighting matrix does not reduce generality. Since the covariance matrix $\hat{\mathbf{\Omega}}(u)$ is positive definite, the weighting matrix $\hat{\mathbf{\Omega}}(u)^{-1}$ is positive definite, too. Therefore, $\hat{\boldsymbol{\alpha}}(u)^{\mathrm{T}}\hat{\mathbf{\Omega}}(u)^{-1}\hat{\boldsymbol{\alpha}}(u) = 0 \Leftrightarrow \hat{\boldsymbol{\alpha}}(u) = \mathbf{0}_N \Leftrightarrow \hat{\boldsymbol{\alpha}}(u)^{\mathrm{T}}\mathbf{W}\hat{\boldsymbol{\alpha}}(u) = 0$ for any positive definite weighting matrix **W**, including $\mathbf{W}=\mathbf{I}_N$. The replacement of the utility functions also does not reduce generality. As shown by Russell and Seo (1989) and Bowden (2005), a portfolio is SSD efficient if and only if it is optimal for some utility function $u_z(x)$, $z \in Z$. Using these two simplifications, it suffices to prove that $\underline{J}_2 \equiv \min_{z\in Z}\hat{\boldsymbol{\alpha}}(u_z)^{\mathrm{T}}\hat{\boldsymbol{\alpha}}(u_z)$ converges to zero under the null and to a strictly positive value under the alternative.

For a given $z \in Z$, let $J(z) \equiv \hat{\boldsymbol{\alpha}}(u_z)^{\mathrm{T}}\hat{\boldsymbol{\alpha}}(u_z)$. Since the observations are serially IID, the $\hat{\boldsymbol{\alpha}}(u_z)$ converge to $\boldsymbol{\alpha}(u_z)$, and hence $J(z)$ converges to $J^P(z) \equiv \boldsymbol{\alpha}(u_z)^{\mathrm{T}}\boldsymbol{\alpha}(u_z)$ for all $z \in Z$, where the superscript P denotes the population. In case of efficiency, we have $J^P(z) = 0$ for some $z \in Z$. Since $0 \leq \underline{J}_2 \leq J(z)$ for all $z \in Z$, this implies that $\underline{J}_2$ converges to zero in case of efficiency; $\lim_{T\to\infty}\Pr[\underline{J}_2 > \varepsilon] = 0$ for any $\varepsilon > 0$.

It remains to be shown that $\underline{J}_2$ converges to a strictly positive value in case of inefficiency, i.e., $J^P(z) > c$ for all $z \in Z$ and for some $c > 0$. Partition the domain Z into $K < \infty$ intervals of equal length, i.e., $\Delta z \equiv (\bar{z} - \underline{z})/K$, and with midpoints $\underline{z} + (k-0.5)\Delta z$, $k = 1,\cdots,K$. Let $\underline{\underline{J}}_2 \equiv \min_{k=1,\cdots,K} J(\underline{z} + (k-0.5)\Delta z)$. We can obtain an arbitrarily good approximation, e.g., $\underline{\underline{J}}_2 - \underline{J}_2 < \varepsilon$, with $\varepsilon > 0$ an arbitrarily small number, provided we set K to be sufficiently large. Hence, we may focus on the behavior of $\underline{\underline{J}}_2$ instead of that of $\underline{J}_2$

without loss of generalization. Using the inclusion-exclusion principle for the union of sets, we find $\Pr[\underline{\underline{J}}_2 \le y] = \Pr[\bigcup_{k=1}^{K}\{J(\underline{z}+(k-0.5)\Delta z) \le y\}] \le \sum_{k=1}^{K}\Pr[J(\underline{z}+(k-0.5)\Delta z) \le y]$, for any $y \ge 0$. Since the individual terms $J(\underline{z}+(k-0.5)\Delta z)$, $k = 1,\cdots,K$, converge to $J^P(\underline{z}+(k-0.5)\Delta z) > c$, the individual probabilities $\Pr[J(\underline{z}+(k-0.5)\Delta z) \le y]$ converge to zero and $\lim_{T\to\infty}\Pr[\underline{\underline{J}}_2 \le y] = 0$ for all $y < c$. *Q.E.D.*

Proof of T*HEOREM* *2* Let $J(u) \equiv \hat{\boldsymbol{\alpha}}(u)^{\mathrm{T}}\hat{\boldsymbol{\Omega}}(u)^{-1}\hat{\boldsymbol{\alpha}}(u)$ for given $u \in U_2^*$ and let $U_2^{**} \equiv \{u \in U_2^* : \boldsymbol{\alpha}(u) = \mathbf{0}_N\}$, which is always nonempty under the null. We can derive the asymptotic distribution of $J(u)$ for any given $u \in U_2^{**}$ from known results. The empirical pricing errors $\hat{\boldsymbol{\alpha}}(u)$ obey an asymptotic multivariate normal distribution with mean $\boldsymbol{\alpha}(u)$ and covariance matrix $T^{-1}\boldsymbol{\Omega}(u)$, i.e., $\hat{\boldsymbol{\alpha}}(u) \xrightarrow{d} N(\boldsymbol{\alpha}(u), T^{-1}\boldsymbol{\Omega}(u))$. Therefore, $J(u)T$ obeys an asymptotic non-central chi-squared distribution with *N* degrees of freedom and non-centrality parameter $\mu(u) \equiv T\boldsymbol{\alpha}(u)^{\mathrm{T}}\boldsymbol{\Omega}(u)^{-1}\boldsymbol{\alpha}(u)$. For all $u \in U_2^{**}$, we have $\boldsymbol{\alpha}(u) = \mathbf{0}_N$ and $\mu(u) = 0$ and thus $J(u)T$ obeys an asymptotic central chi-squared distribution with *N* degrees of freedom, that is, $J(u)T \xrightarrow{d} \chi_N^2 \quad \forall u \in U_2^{**}$. Since $J_2 \le J(u)$ for all $u \in U_2^{**}$, we therefore find that the asymptotic null distribution of J_2 is bounded from above by the chi-squared distribution. *Q.E.D.*