Empirical Tests for Stochastic Dominance Efficiency
Author(s): Thierry Post
Source: *The Journal of Finance*, Vol. 58, No. 5 (Oct., 2003), pp. 1905-1931
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/3648178
Accessed: 24/11/2009 09:23

THE JOURNAL OF FINANCE • VOL. LVIII, NO. 5 • OCTOBER 2003

Empirical Tests for Stochastic Dominance Efficiency

THIERRY POST*

ABSTRACT

We derive empirical tests for the stochastic dominance efficiency of a given portfolio with respect to all possible portfolios constructed from a set of assets. The tests can be computed using straightforward linear programming. Bootstrapping techniques and asymptotic distribution theory can approximate the sampling properties of the test results and allow for statistical inference. Our results could provide a stimulus to the further proliferation of stochastic dominance for the problem of portfolio selection and evaluation. Using our tests, the Fama and French market portfolio is significantly inefficient relative to benchmark portfolios formed on market capitalization and book-to-market equity ratio.

THE THEORY OF STOCHASTIC DOMINANCE (SD), developed by Hadar and Russell (1969), Hanoch and Levy (1969), Rothschild and Stiglitz (1970), and Whitmore (1970), among others, gives a systematic framework for analyzing economic behavior under uncertainty. SD has seen considerable theoretical development and empirical application in the last decades, in various areas of economics, finance, and statistics (see, e.g., Levy (1992, 1998)). It is useful for positive analysis (where the objective is to analyze the decision rules actually used by decision makers) as well as normative analysis (where the objective is to support practical decision making). The theoretical attractiveness of SD lies in its nonparametric orientation. SD criteria do not require a full parametric specification of the preferences of the decision maker and the statistical distribution of the choice alternatives, but rather they rely on a set of general assumptions.

The SD literature involves a multitude of different criteria, associated with different sets of preference assumptions. Higher order criteria involve more discriminating power than lower order ones, because they induce a larger reduction of the set of efficient choice alternatives. However, that power has to be balanced against the stringency of the additional preference assumptions. In general,

*Post is an Associate Professor of Finance at the Erasmus University Rotterdam in the Netherlands. This study forms part of a research program on stochastic dominance. Details on the program are available at the program homepage http://www.few.eur.nl/few/people/gtpost/stochastic.dominance.htm. I would like to express my appreciation to Martijn van den Assem, Alan Genz, Timo Kuosmanen, Haim Levy, and Jaap Spronk, as well as the editor, Rick Green, and two anonymous referees, for providing helpful comments and stimulating discussion. I gratefully acknowledge financial support by Tinbergen Institute and Erasmus Research Institute of Management. Any remaining errors are my own.

striking that balance requires a careful consideration of the structure and the context of the decision problem under consideration. For the sake of compactness, we focus on the popular criterion of second-order SD (SSD). The assumptions associated with this criterion (monotonicity and concavity of the utility function) have a compelling economic interpretation (nonsatiation and risk aversion). Still, nothing excludes the generalization of our analysis to other SD criteria, including first-order SD (FSD), third-order SD (TSD), and higher-order SD (see Section V).

The problem of selecting and evaluating investment portfolios is an interesting application area for SD because (1) economic theory does not forward strong predictions about investor preferences and asset return distributions, and (2) large, high-quality data sets are available, so that nonparametric analysis can let the data "speak for themselves." Still, the focus of the research in this area has predominantly been on mean-variance analysis (MVA), for example, Kandel and Stambaugh (1987, 1989), and Gibbons, Ross, and Shanken (1989), and extensions of that framework (e.g., Yamakazi and Konno (1991), and Young (1998)). Unfortunately, MVA is consistent with expected utility theory if and only if investor preferences and/or return distributions obey highly restrictive conditions (see, e.g., Hanoch and Levy (1969)). An important reason why SD has not seen the proliferation that one might expect (based on its theoretical attractiveness) is that until recently, SD efficiency could only be tested pairwise. This restriction limits the scope of SD tests, because investors generally can diversify between a set of assets and they effectively face infinitely many choice alternatives. There are currently no computationally tractable empirical tests for SD efficiency with full diversification possibilities. Levy (1992, p. 583) concludes that this problem is "the main drawback of the SD framework." The purpose of this study is to fill this gap; we develop tractable linear programming (LP) tests for the SD efficiency of a given portfolio relative to all possible portfolios created from a set of assets.

For applying SD criteria to empirical data, simple crossing algorithms have been developed that check the difference of the empirical distribution functions (EDFs) of the choice alternatives (e.g., Levy (1992), Appendix A). This approach is computationally very efficient; one first constructs the EDFs from the ordered outcomes of the alternatives and next checks whether the EDFs cross for any of the observed values. Unfortunately, this approach is restricted to pairwise comparison of a finite number of choice alternatives, and it cannot be applied to problems with full diversification possibilities. The problem is that the ordering of the outcomes (and hence the EDF) of a diversified portfolio of alternatives cannot be determined in a straightforward way from the orderings of the individual alternatives. Therefore, the ordering of each portfolio has to be determined individually. This is computationally impossible if infinitely many portfolios have to be considered, as is true if full diversification is possible. Also, in case of more than two choice alternatives, pairwise dominance is not a necessary condition for SD inefficiency. Bawa et al. (1985) provide LP tests for applying the concept of convex stochastic dominance (Fishburn (1974)), which does provide a necessary and sufficient condition if more than two choice alternatives are compared. However, this approach also considers only a finite number of choice alternatives, and it cannot account for full diversification. Recently, some authors have suggested ways to test whether a given portfolio is SD efficient if full diversification is possible. The Marginal Conditional Stochastic Dominance (MCSD) framework by Shalit and Yitzhaki (1994) provides tests for identifying assets whose portfolio weights should be altered in order to construct alternative portfolios that SD dominate the evaluated portfolio. These tests give necessary but not sufficient conditions for SD efficiency, and they may fail to identify inefficient portfolios. Parallel to our study, Kuosmanen (2001) develops necessary *and* sufficient tests. He includes the ordering of the outcomes as model variables and provides a LP relaxation of the resulting mixed integer linear programming model. Unfortunately, the number of model variables increases progressively with the number of observations (for SSD, the number of variables is a second-order polynomial; for TSD the order is four). The computational complexity of LP problems (as measured by the number of arithmetic operations, run time, and working memory requirements) increases progressively with the number of model variables. Hence, the Kuosmanen approach is computationally prohibitive for samples that are sufficiently large to allow for powerful analysis (i.e., hundreds or thousands of observations; see Section III). We therefore conclude that there currently are no computationally tractable necessary and sufficient tests for SD with full diversification.

Our tests are derived from the optimality conditions for portfolio optimization in the expected utility framework. The expected utility framework has a number of well-known limitations. However, SD criteria are also economically meaningful for a range of nonexpected utility theories of choice behavior under uncertainty (see, e.g., Machina (1982), and Starmer (2000)). Basically, the tests check whether we can construct piecewise-linear utility functions that rationalize the evaluated portfolio. In this respect, our approach has a strong analogy with the nonparametric approach to consumer analysis by Afriat (1967) and Varian (1982). In financial economics, Dybvig and Ross (1982), Varian (1983), and Green and Srivastava (1985, 1986), among others, have proposed related approaches to deriving optimality conditions and to recovering investor preferences.

The remainder of this paper is organized as follows. Section I derives a useful characterization of SSD efficiency. Section II gives our LP tests for SSD efficiency. We present a primal test in terms of piecewise-linear utility functions and a dual test in terms of Bowden's (2000) Ordered Mean Difference performance measure. Apart from the computational problems associated with full diversification possibilities, another problem in practical applications of SD is the sensitivity of the results to sampling error. Section III discusses how bootstrapping techniques and asymptotic distribution theory can approximate the sampling distribution of the test results and allow for statistical inference. Section IV uses our approach to examine whether the Fama and French market portfolio is efficient relative to benchmark portfolios formed on market capitalization and book-to-market equity ratio. Finally, Section V gives conclusions and suggests directions for future research. The Appendix gives the formal proofs of our theorems.

I. Second-Order Stochastic Dominance

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

1. Investors select investment portfolios to maximize the expected utility associated with the return of their investment portfolio.[1] Throughout the text, we will denote utility functions by $u : \Re \rightarrow P$, $u \in U$, with U for the class of von Neumann–Morgenstern utility functions, and P for a nonempty, closed, and convex subset of $\Re$.[2]
2. The investment universe consists of N assets, associated with returns $\boldsymbol{x} \in \Re^N$.[3] Throughout the text, we will use the index set $\mathrm{I} \equiv \{1, \ldots, N\}$ to denote the different assets. The returns are serially independent and identically distributed (IID) random variables with a continuous joint cumulative distribution function (CDF) $G : \Re^N \rightarrow [0,1]$.[4] Further, the returns have means $E[\boldsymbol{x}] = \boldsymbol{\mu}$ and variance-covariance matrix $E[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}] = \boldsymbol{\Omega}$.
3. Investors may diversify between the assets, and we will use $\boldsymbol{\lambda} \in \Re^N$ for a vector of portfolio weights. We consider the case where short sales are not allowed, and the portfolio weights belong to the portfolio possibilities set

[1] This portfolio-based perspective treats investment as an isolated problem. A consumption-based perspective would treat investment as part of a more general problem where the objective is to maximize the expected intertemporal utility from consumption. Since a large part of personal income comes from nonfinancial assets, for example, salary and wages, we would have to extend our analysis to include nonfinancial assets, for instance, human capital. Unfortunately, empirical data on consumption and nonfinancial assets generally are of less quality than data on financial assets (see, e.g., Breeden, Gibbons, and Litzenberger (1989)).

[2] 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 $\boldsymbol{x}$.

[3] Our analysis allows for including a riskless asset in the analysis, and thus covers not only SSD but also the more powerful SSDR (i.e., SSD with a riskless asset; see, e.g., Levy (1998), Section 4.3).

[4] There is substantial evidence that the distribution of assets returns (e.g., interest rates, risk premia, volatilities, and correlation coefficients) varies through time. This problem is especially relevant for applications that use data for long time periods. In such cases, the observations generally are not serially IID random variables and the empirical distribution function (EDF) is not a statistically consistent estimate for the CDF. Unfortunately, the search for a satisfactory specification of the return dynamics is still far from accomplished. In fact, Ghysels (1998) finds that ill-specified conditional asset pricing models in many cases yield greater pricing errors than unconditional models. For this reason, we use an unconditional model here. Still, further research could relax the IID assumption. In the empirical application, we account for time variation by adjusting the raw return observations for the variation in the riskless rate. A more involved approach is to use econometric time-series estimation techniques to estimate a conditional CDF. Our empirical tests can then be applied to random samples from the estimated CDF rather than the EDF.

$\Lambda \equiv \{\boldsymbol{\lambda} \in \Re_+^N : \boldsymbol{e}^{\mathrm{T}}\boldsymbol{\lambda} = 1\}$, with $\boldsymbol{e}$ for a unity vector with dimensions conforming to the rules of matrix algebra.[5]

A given portfolio $\boldsymbol{\tau} \in \Lambda$ is optimal for an investor with utility $u \in U$ if and only if

$$\int u(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})dG(\boldsymbol{x}) = \max_{\boldsymbol{\lambda} \in \Lambda} \int u(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\lambda})dG(\boldsymbol{x}).^{6} \tag{1}$$

In practical applications, full information about utility functions typically is not available, and this condition cannot be verified directly. This provides the rationale for using SD criteria that rely on a set of general assumptions rather than a full specification of the utility function. The SSD criterion restricts attention to the class of nonsatiable and risk-averse investors or the class of strictly increasing and concave utility functions $U_2 \subseteq U$.[7] Note that we do not assume that the utility function is continuously differentiable. However, utility is concave and hence everywhere continuous and superdifferentiable.[8] Throughout the text, we will denote supergradients at x by $\partial u(x)$.

Apart from the utility function, also the CDF generally is not known in practical applications. Rather, information typically is limited to a discrete set of T, $T > N$, time-series observations, which are here treated as independent random draws from the CDF. 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, by assumption, 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, $\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{\tau} < \boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{\tau} < \cdots < \boldsymbol{x}_T^{\mathrm{T}}\boldsymbol{\tau})$[9,10] Using

[5] We focus on the simplex Λ because short selling typically is difficult to implement in practice due to margin requirements and explicit or implicit restrictions on short selling for institutional investors (see, e.g., Sharpe (1991)). Still, it is possible to generalize our analysis to cases where (bounded) short sales are allowed (as well as cases where additional restrictions are imposed on the portfolio weights). Our analysis is based on the first-order optimality conditions for optimizing a concave function over a polyhedron (see the proof to Theorem 2). We may generalize our analysis by replacing Λ with a more general polyhedral portfolio possibilities set. We basically have to check whether a tangent hyperplane exists that envelops all extreme points of the portfolio set. One approach is to enumerate all extreme points and to replace the set I with the set of all extreme points.

[6] We use $dG(\boldsymbol{x})$ for the probability measure associated with $G(\boldsymbol{x})$.

[7] SSD sometimes assumes concavity without strictly increasing utility (see, e.g., Huang and Litzenberger (1988)). Our analysis can be generalized in a straightforward way to include this alternative SSD definition, as well as other SD criteria; see Section V.

[8] Roughly speaking, $u(x)$ is superdifferentiable if, for all $x \in \Re$, there exist tangency lines that support the graph of $u(x)$ from above. We do not assume continuous differentiability so as to allow for nonsmooth functions, including piecewise-linear functions (see Section II).

[9] Since we assume a continuous return distribution, ties do not occur. Still, the analysis can be extended in a straightforward way to cases where ties do occur, for example, due to a discrete return distribution or due to measurement problems or rounding (see the discussion at the end of Section II).

[10] Note that we rank only based on the returns of the evaluated portfolio. By contrast, the existing crossing algorithms (see, e.g., Levy (1992), Appendix A) also rank based on the benchmark portfolio, which introduces computational problems if the benchmark portfolio is not known beforehand, but rather has to be selected from infinitely many candidate portfolios.

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

$$F(\boldsymbol{x}) \equiv \operatorname{card}\{t \in \Theta : \boldsymbol{x}_t \leq \boldsymbol{x}\}/T, \tag{2}$$

with $\Theta \equiv \{1, \cdots, T\}$ for the index set of ranked observations, and $\operatorname{card}\{\cdot\}$ for the number of elements of a set. Our empirical tests use the EDF in place of the CDF. In Section III, we discuss how to account for sampling error.

Using the above notation and assumptions, SSD for pairwise comparison can be defined as follows.

Definition 1: *Portfolio $\boldsymbol{\lambda} \in \Lambda$ dominates portfolio $\boldsymbol{\tau} \in \Lambda$ by SSD if and only if, for all utility functions $u \in U_2$, $\boldsymbol{\lambda}$ has a higher expected utility than $\boldsymbol{\tau}$, that is,*

$$\int u(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\lambda})dF(\boldsymbol{x}) - \int u(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})dF(\boldsymbol{x}) > 0 \quad \forall u \in U_2 \Leftrightarrow \tag{3}$$

$$\sum_{t \in \Theta} \left(u\left(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda}\right) - u\left(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda}\right)\right)/T > 0 \quad \forall u \in U_2 \tag{4}$$

Note that this definition of SSD uses strict inequalities for all $u \in U_2$. By contrast, the traditional definition uses weak inequalities with a strict inequality for at least one $u \in U_2$. This difference is important from a theoretical perspective, and one can think of examples where the two definitions give different efficiency classifications. For example, using Definition 1, $\boldsymbol{\lambda} \in \Lambda$ does not dominate mean-preserving spreads of $\boldsymbol{\lambda}$, because risk-neutral investors are indifferent between alternatives that have identical means. By contrast, dominance does exist using the traditional definition, because all strictly risk-averse investors do prefer $\boldsymbol{\lambda}$ to mean-preserving spreads. However, from an empirical perspective, the definitions are indistinguishable, because arbitrary small data perturbations to the evaluated portfolio can make the classifications consistent. Related to this, data sets where this theoretical issue has a decisive impact are extremely unlikely for return distributions that are continuous by approximation.[11]

The following is a straightforward generalization of Definition 1 to the case where full diversification is allowed.

Definition 2: *Portfolio $\boldsymbol{\tau} \in \Lambda$* is SSD inefficient if and only if some portfolio $\boldsymbol{\lambda} \in \Lambda$ *SSD dominates it. Alternatively, portfolio $\boldsymbol{\tau} \in \Lambda$ is SSD efficient if and only if no portfolio $\boldsymbol{\lambda} \in \Lambda$ SSD dominates it.*

Interestingly, this definition can be rephrased in terms of a minimax formulation.[12]

Theorem 1: *Portfolio $\boldsymbol{\tau} \in \Lambda$ is SSD inefficient if and only if, for all utility functions $u \in U_2$, the maximum expected utility is greater than the expected utility of $\boldsymbol{\tau}$, that is,*

[11] Similar arguments are used in production analysis to establish the empirical equivalence of strong and weak measures of productive efficiency in case input–output data are generated by a continuous distribution (see, e.g., Kuntz and Scholtes (2000)).

[12] Similar minimax formulations exist for many weak measures of productive efficiency, including the well-known Debreu (1951) and Farrell (1957) measures.

$$\min_{u\in U_2}\left\{\max_{\lambda\in\Lambda}\left\{\int u(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\lambda})dF(\boldsymbol{x})-\int u(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})dF(\boldsymbol{x})\right\}\right\}=0\Leftrightarrow \tag{5}$$

$$\min_{u\in U_2}\left\{\max_{\lambda\in\Lambda}\left\{\sum_{t\in\Theta}(u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda})-u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau}))/T\right\}\right\}=0. \tag{6}$$

Alternatively, portfolio $\boldsymbol{\tau}\in\Lambda$ is SSD efficient if and only if it is optimal relative to some utility functions $u\in U_2$, that is,

$$\min_{u\in U_2}\left\{\max_{\lambda\in\Lambda}\left\{\int u(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\lambda})dF(\boldsymbol{x})-\int u(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})dF(\boldsymbol{x})\right\}\right\}=0\Leftrightarrow \tag{7}$$

$$\min_{u\in U_2}\left\{\max_{\lambda\in\Lambda}\left\{\sum_{t\in\Theta}(u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda})-u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau}))/T\right\}\right\}=0. \tag{8}$$

II. Linear Programming Tests

This section presents two linear programming (LP) tests for SSD efficiency: (1) a primal test and (2) an equivalent dual test. The primal test asks if we can construct piecewise-linear utility functions $p\in U_2$ that rationalize the evaluated portfolio $\boldsymbol{\tau}\in\Lambda$. All relevant piecewise-linear utility functions $p\in U_2$ can be constructed from a series of T linear support lines characterized by intercept coefficients $\boldsymbol{\alpha}\equiv(\alpha_1\cdots\alpha_T)^{\mathrm{T}}\in\Re^T$ and (normalized) slope coefficients $\boldsymbol{\beta}\in\mathrm{B}\equiv\{\boldsymbol{\beta}\in\Re_+^T:\beta_1\geq\beta_2\geq\cdots\geq\beta_T=1\}$ as

$$p(x|\boldsymbol{\alpha},\boldsymbol{\beta})\equiv\min_{t\in\Theta}(\alpha_t+\beta_t x). \tag{9}$$

The dual test involves the Ordered Mean Difference (OMD):

$$\rho_t(\boldsymbol{\lambda},\boldsymbol{\tau})\equiv\sum_{s=1}^{t}(\boldsymbol{x}_s^{\mathrm{T}}\boldsymbol{\lambda}-\boldsymbol{x}_s^{\mathrm{T}}\boldsymbol{\tau})/t\qquad t\in\Theta. \tag{10}$$

The OMD represents a running mean for the difference between the return of the evaluated portfolio $\boldsymbol{\tau}$ and a benchmark portfolio $\boldsymbol{\lambda}\in\Lambda$.[13] Bowden (2000) introduced this statistic to measure (in pairwise fashion) the performance of the evaluated portfolio relative to a given benchmark portfolio.

[13] Since the return observations are labeled by their ranking with respect to the evaluated portfolio $\boldsymbol{\tau}$, the OMD differs from the cumulated difference between the quantiles of the EDFs of $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$. The quantiles of the EDF of $\boldsymbol{\lambda}$ require ranking by $\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda}$ rather than $\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau}$; the OMD uses a "less favorable" ranking and hence gives an upper bound for the cumulated difference between the quantiles.

A. Central Theorem

Consider the primal test statistic

$$\xi(\boldsymbol{\tau}) \equiv \min_{\boldsymbol{\beta}\in \mathrm{B},\theta} \left\{ \theta : \sum_{t\in\Theta} \beta_t(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau} - x_{it})/T + \theta \geq 0 \quad \forall i \in \mathrm{I} \right\} \tag{11}$$

and the associated dual test statistic

$$\psi(\boldsymbol{\tau}) \equiv \max_{\boldsymbol{\lambda}\in\Lambda}\{\rho_T(\boldsymbol{\lambda},\boldsymbol{\tau}) : \rho_t(\boldsymbol{\lambda},\boldsymbol{\tau}) \geq 0 \ \forall t \in \{1,\cdots,T-1\}\}. \tag{12}$$

Using these test statistics, we can state the central result of this paper.

THEOREM 2: *Portfolio $\boldsymbol{\tau} \in \Lambda$ is SSD inefficient if and only if $\xi(\boldsymbol{\tau}) > 0$ or, equivalently, $\psi(\boldsymbol{\tau}) > 0$. Alternatively, portfolio $\boldsymbol{\tau} \in \Lambda$ is SSD efficient if and only if $\xi(\boldsymbol{\tau}) = 0$ or, equivalently, $\psi(\boldsymbol{\tau}) = 0$.*

The test statistics $\xi(\boldsymbol{\tau})$ and $\psi(\boldsymbol{\tau})$ involve linear objective functions and linear constraints, and they can be computed using straightforward linear programming. Full LP formulations are included as (P) and (D) in the proof to Theorem 2. The primal problem involves only T variables and $N+T-1$ constraints, and the dual problem involves $N+T-1$ variables and T constraints. The problems are feasible and bounded. A possible primal solution is $\boldsymbol{\beta}=\boldsymbol{e}$, associated with the upper bound $\xi(\boldsymbol{\tau}) = \max_{i\in\mathrm{I}}\{\Sigma_{t\in\Theta}(x_{it} - \boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau})/T\}$. (This solution effectively represents the utility function $u(x)=x$ or the risk-neutral investor who cares about expected return only.) A possible dual solution is $\boldsymbol{\lambda}=\boldsymbol{\tau}$, associated with the lower bound $\psi(\boldsymbol{\tau})=0$. For small data sets up to hundreds of observations and/or assets, the problems can be solved with minimal computational burden, even with desktop PCs and standard solver software (e.g., LP solvers included in spreadsheets). Still, the computational complexity, as measured by the required number of arithmetic operations, and hence the run time and memory space requirement, increases progressively with the number of variables and constraints. Therefore, specialized hardware and/or solver software is recommended for large-scale problems involving thousands of observations and/or assets.[14]

B. Interpretation of the Test Results

Roughly speaking, the primal test statistic $\xi(\boldsymbol{\tau})$ can be interpreted as the "least amount of disutility" that any nonsatiated and risk-averse investor would suffer by holding on to the evaluated portfolio $\boldsymbol{\tau}$. The dual test statistic $\psi(\boldsymbol{\tau})$ gives the maximum possible increase in average return, that is, $\rho_T(\boldsymbol{\lambda},\boldsymbol{\tau})$, achieved by a portfolio $\boldsymbol{\lambda} \in \Lambda$ that "outperforms" the evaluated portfolio in terms of the OMDs. These test statistics give necessary and sufficient conditions that can separate efficient portfolios from inefficient ones. *However, the test statistics do not represent*

[14] For an elaborate introduction in LP, we refer to Chvatal (1983). In practice, very large LPs can be solved efficiently by both the simplex method and interior-point methods. An elaborate guide to LP solver software can be found at the homepage of the Institute for Operations Research and Management Science (INFORMS), http://www.informs.org/.

meaningful performance measures that can be used for ranking portfolios based on the "degree of efficiency." The point is that SSD considers an entire class of investors and one cannot compare the utility or disutility of one investor to that of another.

In addition to the value of the test statistic, the primal model gives information on the shape of the optimal piecewise-linear utility function. Specifically, the optimal value for each β_t variable represents the slope of the tth support line of the utility function, and we may recover a complete utility function from these variables (see the proof to Theorem 2). Further, the dual model identifies a benchmark portfolio in addition to the value of the test statistic. Specifically, the optimal values of the dual variables λ_i, $i \in \mathrm{I}$, represent the shadow prices for the primal constraints $\sum_{t\in\Theta} \beta_t(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau} - x_{it})/T + \theta \geq 0,\ i \in \mathrm{I}$. These constraints are binding for the assets that constitute the portfolio that maximizes the value of the test statistic, and the optimal $\boldsymbol{\lambda}$ gives the composition of this portfolio. *We stress that the primal piecewise-linear utility function and dual solution portfolio are mere byproducts of our analysis and we cannot claim any confidence in their economic interpretation or statistical estimation.* For example, there typically exist multiple optimal solutions for the slope coefficients if the evaluated portfolio is SSD efficient, and, in addition, the interpretation of the piecewise-linear utility function is not clear if the evaluated portfolio is SSD inefficient. Similarly, the solution portfolio generally does not SSD dominate the evaluated portfolio, and moreover it may, in fact, be inefficient. Another complication stems from sampling error. While it is possible to perform powerful and accurate inference about the *test statistic* (at least in large samples; see Section III), the estimated utility function and benchmark portfolio are substantially more sensitive to estimation error.

The primal model involves T different slope coefficients. However, in many applications, most slope coefficients will take the same value, that is, $\beta_t = \beta_{t+1}$ for many $t \in \{1, \ldots, T-1\}$. Hence, piecewise-linear utility functions constructed from the optimal slope coefficients will consist of only a few line segments and they will exhibit only a few "kinks." The dual formulation offers a direct explanation for this phenomenon in terms of the "stickiness" of running means. It follows from the proof to Theorem 2 that $t\rho_t(\boldsymbol{\lambda}, \boldsymbol{\tau})/T$ represents the shadow price for the primal constraint $\beta_t - \beta_{t+1} \geq 0$, $t \in \{1, \ldots, T-1\}$. Complementary slackness implies that $\beta_t - \beta_{t+1} > 0$ only if $\rho_t(\boldsymbol{\lambda}, \boldsymbol{\tau}) = 0$, and $\beta_t - \beta_{t+1} = 0$ if $\rho_t(\boldsymbol{\lambda}, \boldsymbol{\tau}) > 0$. The OMD is a running mean, and it is very "sticky," as $\rho_{t+1}(\boldsymbol{\lambda}, \boldsymbol{\tau}) = t\rho_t(\boldsymbol{\lambda}, \boldsymbol{\tau})/(t+1) + \boldsymbol{x}_{t+1}^{\mathrm{T}}(\boldsymbol{\lambda} - \boldsymbol{\tau})/(t+1)$. Therefore, if the evaluated portfolio is inefficient, then there typically exist few $t \in \{1, \ldots, T-1\}$ for which $\rho_t(\boldsymbol{\lambda}, \boldsymbol{\tau}) = 0$ and hence $\beta_t - \beta_{t+1}$ usually equals zero. This phenomenon helps to explain why increasing the sample size increases power (see Section III), even though the number of primal variables and dual constraints increases at the same rate. Since the primal variables (dual constraints) are strongly interrelated, introducing additional observations does not "proportionally" loosen (tighten) the primal (dual) model.

C. Treatment of Ties

As discussed in Section I, we label the return observations by their ranking with respect to the evaluated portfolio. Multiple orderings exist if the observations for the evaluated portfolio include ties, for example, due to a discrete return distribution or due to measurement problems or rounding (or if we analyze a riskless asset). A simple solution to the problem of ties is available. Specifically, if $\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau} = \boldsymbol{x}_{t+1}^{\mathrm{T}}\boldsymbol{\tau}$ for some $t \in \{1, \cdots, T-2\}$, then we can drop the primal constraint $\beta_t \geq \beta_{t+1}$ and replace it with the constraint $\beta_t \geq \beta_{t+2}$ plus the constraint $\beta_{t-1} \geq \beta_{t+1}$ if $t \in \{2, \cdots, T-2\}$. In addition, if $\boldsymbol{x}_{T-1}^{\mathrm{T}}\boldsymbol{\tau} = \boldsymbol{x}_T^{\mathrm{T}}\boldsymbol{\tau}$, then we can replace the constraint $\beta_{T-1} \geq \beta_T = 1$ with the constraint $\beta_{T-1}, \beta_T \geq 1$. (The adjustments to the dual test statistic can be obtained from the LP dual of adjusted primal test statistic.)

The following example with two time periods, one risky asset and one riskless asset, can provide some feeling for Theorem 2, as well as the treatment of ties:

	x_1	x_2
$t=1$	$r-a$	r
$t=2$	$r+b$	r

where r is the riskless return and a, $b>0$. In this case, we know that the risky asset, x_1, is SSD efficient if and only if $b \geq a$, and the risk-free asset, x_2, is SSD efficient for all a, $b>0$ (e.g., Arrow (1971)). Theorem 2 states that x_1 is efficient if and only if we can find β_1 and β_2 such that $\beta_1 \geq \beta_2 = 1$ and $\beta_1((r-a)-r) + \beta_2(r+b)-r) \geq 0 \Leftrightarrow b\beta_2 - a\beta_1 \geq 0$. Such β_1 and β_2 can be found if and only if $b \geq a$. If this condition holds, then x_1 is optimal for all risk-neutral investors, for example, with utility function $u(x) = x$. We now turn to testing efficiency for the riskless asset, x_2. The returns of x_2 are tied, and we therefore have to drop the constraint $\beta_1 \geq \beta_2 = 1$, and replace it with β_1, $\beta_2 \geq 1$. Theorem 2 then states that x_2 is efficient if and only if we can find coefficients β_1, $\beta_2 \geq 1$ such that $\beta_1(r-(r-a)) + \beta_2(r-(r+a)) \geq 0$. For all a, $b>0$, we can find such weights, and hence x_2 is always efficient. For example, x_2 is the optimal portfolio for all investors with utility function

$$u(x) = \begin{cases} (1-c)r + cx & x \leq r \\ x & x \geq r \end{cases},$$

with $c \geq b/a$. These results exactly conform to the known results for this particular case.

III. Statistical Inference

We have thus far discussed SD efficiency relative to the EDF $F(\boldsymbol{x})$ rather than the CDF $G(\boldsymbol{x})$ (see Section I). Generally, the EDF is very sensitive to sampling variation and the test results are likely to be affected by sampling error in a nontrivial way. For example, if we compare two alternatives with the same population distribution, then we know that there exists no pairwise dominance relationship in the population. Still, the probability of finding a dominance relationship based on two independent random samples of 1,000 observations can be as high as 50 percent (see Dardanoni and Forcina (1999)). In addition, the outcomes of various simulation studies (Kroll and Levy (1980), Stein and Pfaffenberger (1986), among others) cause serious doubt about the reliability of SD applications that rely in a naïve way on the EDF without accounting for sampling error. The applied researcher must therefore have knowledge of the sampling distribution in order to make inferences about the true efficiency classification. In the SD literature, two approaches have been developed to approximate the sampling distribution: (1) bootstrapping and (2) analytical asymptotic analysis.

A. Bootstrapping

The bootstrap method, first introduced by Efron (1979) and Efron and Gong (1983), is a well-established tool to analyze the sensitivity of empirical estimators to sampling variation in situations where the sampling distribution is difficult or impossible to obtain analytically. Bootstrapping is based on the idea of repeatedly simulating the CDF, usually through resampling, and applying the original estimator to each simulated sample or pseudo-sample so that the resulting estimators mimic the sampling distribution of the original estimator. Key to the success of the bootstrap is the selection of an appropriate approximation for the CDF. If the approximation is statistically consistent, then the bootstrap distribution gives a statistically consistent estimator for the original sampling distribution. In the context of our SD tests, the EDF is an appropriate approximation for the CDF; under the assumption that the return distribution is serially IID (see Section I), the EDF is a consistent estimator of the true CDF. This suggests that bootstrapping samples would be simply obtained by randomly sampling with replacement from the EDF along the lines of the "correlation model" proposed by Freedman (1981) in a regression framework. Nelson and Pope (1991) demonstrated in a convincing way that this approach can quantify the sensitivity of the EDF to sampling variation, and that SD analysis based on the bootstrapped EDF is more powerful than analysis based on the original EDF.[15] The computational ease of the crossing algorithms allows for substituting brute computational force to overcome the analytical intractability of SD. Interestingly, the tractable LP structure of our tests suggests that this is possible also in the case with full diversification possibilities.

B. Analytical Asymptotic Analysis

The alternative approach is to derive an analytical characterization of the asymptotic sampling distribution (see, e.g., Beach and Davidson (1983), Dardanoni and Forcina (1999), Davidson and Duclos (2000)). The literature thus far invariantly deals with comparing a finite number of choice alternatives, and it is not immediately clear how to generalize the existing results to our case with full diversification possibilities.[16] The purpose of this section is to provide intuition

[15] Interestingly, bootstrapping is used successfully also for the nonparametric approach to analyzing productive efficiency, which has a structure similar to SD analysis (see Simar and Wilson (1998), among others).

[16] Presumably, this is because much of the research on SD is aimed at ordering distributions of poverty, welfare, and inequality, an application area that typically involves the comparison of a small set of distributions (e.g., for a set of countries) and diversification between the distributions is not allowed.

about the size and power of our SSD test as the number of observations and the number of assets vary. For this purpose, we analyze the asymptotic sampling distribution of our test statistic $\xi(\boldsymbol{\tau})$ under two simplifying assumptions. (Duality implies that the results apply with equal strength to the dual statistic $(\boldsymbol{\tau})$.)

The first assumption relates to the null hypothesis. Ideally, the null hypothesis would simply state that the evaluated portfolio is efficient, that is,

$$H_0 : \min_{u \in U_2} \left\{ \max_{\boldsymbol{\lambda} \in \Lambda} \left\{ \int u(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\lambda}) dG(\boldsymbol{x}) - \int u(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau}) dG(\boldsymbol{x}) \right\} \right\} = 0. \tag{13}$$

Unfortunately, it is impossible to reject this null in practice, because there exist $G(\boldsymbol{x})$ with a very high probability of classifying an efficient portfolio as highly inefficient.[17] To circumvent this problem, we will use the more restrictive null that all assets have the same mean, that is, $H_1{:}E[\boldsymbol{x}] = \mu\boldsymbol{e}$, $\mu \in \Re$, and the variance-covariance matrix is bounded, that is, $E[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}] = \boldsymbol{\Omega} < \infty$.[18] In general, H_1 gives a sufficient but not necessary condition for H_0. In fact, under the null, all portfolios $\boldsymbol{\lambda} \in \Lambda$ are efficient, because the utility function $u(x) = x$ is feasible and the least disutility associated with any portfolio is zero (as measured relative to this utility function). *However, rejection of H_1 generally does not imply rejection of H_0. Further research is needed to analyze the conditions needed to guarantee that H_1 is sufficiently close to H_0, or alternatively to extend our results to a less restrictive null.*

The second assumption relates to the shape of the return distribution $G(\boldsymbol{x})$ under the null. Our approach will be to focus on the least favorable distribution, that is, the distribution that maximizes the size or relative frequency of Type I error (rejecting the null when it is true). The least favorable distribution yields highly conservative critical values and p-values for the test statistic $\xi(\boldsymbol{\tau})$. This approach gives protection against Type I error; for each $G(\boldsymbol{x})$, the size is always smaller than the size for the least favorable distribution.[19] *However, this approach comes at the cost of a high frequency of Type II error (accepting the null when it is not true) or a low power (one minus the relative frequency of Type II error).*

Under these assumptions, and using $\mathbf{C} \equiv (\mathbf{I} - \boldsymbol{e}\boldsymbol{\tau}^{\mathrm{T}})$, we can summarize the asymptotic sampling distribution of $\xi(\boldsymbol{\tau})$ as follows.[20]

[17] For example, consider a risky asset with returns x_{1t}, $t \in \Theta$, with probability p of finding $x_{1t} = r + a$, $a > 0$ and probability $(1 - p)$ of finding $x_{1t} = r + b$, $b < 0$, and a riskless asset with returns $x_{2t} = r$, $t \in \Theta$. In this case, we know that $\boldsymbol{\tau} = [0 \quad 1]^{\mathrm{T}}$ is SSD efficient relative to the CDF. Still, $\boldsymbol{\tau}$ will be classified as efficient relative to the EDF only if $x_{1t} = r + b$ for at least one observation $t \in \Theta$. Hence, for every $T < \infty$, there exist p such that $\Pr[\xi(\boldsymbol{\tau}) = a] = \Pr[x_{1t} = r + a \quad \forall t \in \Theta] = p^T$ approximates unity.

[18] The SD literature typically uses the even more restrictive null that all choice alternatives are contemporaneously IID. We cannot use this null for our test, because a portfolio of multiple assets generally is not independent of the individual assets and it does not have the same distribution as the individual assets, even if the assets are contemporaneously IID.

[19] SD is typically used to reduce the number of choice alternatives that are considered in a follow-up analysis that is intended to select the optimal portfolio. In this context, a Type I error is problematic because it may exclude the optimal choice alternative. By contrast, a Type II error merely increases the number of alternatives for further analysis.

[20] We use $\mathbf{I}$ for an identity matrix with dimensions conforming to the rules of matrix algebra.

THEOREM 3. *For the asymptotic least favorable distribution, the p-value* $\Pr[\xi(\tau) \geq y|H_1]$, $y \geq 0$, *equals the integral* $\Gamma(y) \equiv (1 - \int_{\boldsymbol{z} \leq y\boldsymbol{e}} d\Phi_{\boldsymbol{\Sigma}}(\boldsymbol{z}))$, with $\Phi_{\boldsymbol{\Sigma}}(\boldsymbol{z})$ *for a N-dimensional multivariate normal distribution function with zero means and singular variance-covariance matrix* $\boldsymbol{\Sigma} \equiv (\mathbf{C}\ \boldsymbol{\Omega}\ \mathbf{C}^{\mathrm{T}})/T$.

The theorem shows the crucial role of the length of the time series (T) and the length of the cross section (N); the p-values $\Gamma(y)$, $y \geq 0$, decrease as the time series grows, and they increase as the cross section grows. For small time series and large cross sections, the p-values are very large and a naïve approach to the test statistic (reject efficiency if and only if $\xi(\tau) > 0$) is unlikely to yield anything but noise. A more sound approach is to compare the p-value for the observed value of $\xi(\tau)$ with a predefined level of significance $a \in [0, 1]$; we may reject efficiency if $\Gamma(\xi(\tau)) \leq a$.[21] Alternatively, we may reject efficiency if the observed value of $\xi(\tau)$ is greater than or equal to the critical value

$$\Gamma^{-1}(a) \equiv \inf_{y \geq 0}\{y : \Gamma(y) \leq a\}. \tag{14}$$

Two results are useful for implementing this approach in practice. First, computing the p-value requires the unknown variance-covariance matrix $\boldsymbol{\Omega}$. We may estimate its elements ω_{ij}, $i, j \in \mathrm{I}$, in a distribution-free and consistent manner using the sample equivalents

$$\hat{\omega}_{ij} \equiv \sum_{t \in \Theta}((x_{it} - \sum_{t \in \Theta} x_{it}/T)(x_{jt} - \sum_{t \in \Theta} x_{jt}/T))/T. \tag{15}$$

Second, the integration variables $z \in \Re^N$ obey a singular multivariate normal distribution. To evaluate the integral $\Gamma(y)$, it is useful to formulate these variables as linear combinations of independent standard normal variables. Assume that the risky assets are linearly independent and that no riskless asset is included, so the variance-covariance matrix $\boldsymbol{\Omega}$ is nonsingular. We may then use the decomposition

$$\boldsymbol{z} = \mathbf{C}\mathbf{D}\boldsymbol{w}, \tag{16}$$

with $\boldsymbol{w} \in \Re^N$ for an independent standard normal random vector, and $\mathbf{D} \in \Re^{N \times N}$ for a lower triangular Cholesky factor of $\boldsymbol{\Omega}$, so $\boldsymbol{\Omega} = \mathbf{D}\mathbf{D}^{\mathrm{T}}$. If a riskless asset is included as the Nth asset, then $\boldsymbol{\Omega}$ is singular. We may then use $\boldsymbol{w} \in \Re^{N-1}$ and $\mathbf{D} \in \Re^{N \times (N-1)}$ with the first $(N-1)$ rows equal to the Cholesky factor of the nonsingular $(N-1) \times (N-1)$ variance-covariance matrix of risky assets, and with zeros for the Nth row. Using the above decomposition, we may express the integral equivalently as $\Gamma(y) = (1 - \int_{\mathbf{CD}\boldsymbol{w} \leq y\boldsymbol{e}} d\Phi_{\mathbf{I}}(\boldsymbol{w}))$. This integral can be approximated using standard simulation techniques (see, e.g., footnote 25) and/or numerical integration techniques (see, e.g., Genz and Kwong (2000)).

[21] The significance level refers to the size under the asymptotic least favorable distribution. Under the true return distribution $G(\boldsymbol{x})$, the size is generally smaller than the significance level; see, for example, the simulation experiment below.

C. Simulation Experiment

To further analyze the sampling distribution, we perform a simulation experiment. We analyze 25 risky assets with a multivariate normal return distribution and a single riskless asset. The joint population moments are equal to the sample moments of the monthly excess returns of the 25 Fama and French benchmark portfolios and the one-month U.S. Treasury bill used in Section IV (see Table I for descriptive statistics). Figure 1 shows a mean-variance diagram including the individual assets (the clear dots), as well as the mean-variance frontiers for

Table I

Descriptive Statistics

Monthly percentage excess returns (month-end to month-end) from July 1963 to October 2001 (460 months) for the Fama and French market portfolio and the 25 Fama and French benchmark portfolios formed on market capitalization (size) and book-to-market equity ratio (B/M) are shown. Excess returns are computed from the raw return observations by subtracting the return on the one-month U.S. Treasury bill. All data are obtained from the data library on the homepage of Kenneth French at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french.

No.	B/M	Size	Mean	Variance	Skewness	Kurtosis	Minimum	Maximum
Market portfolio			0.462	4.461	−0.498	2.176	−23.09	16.05
Benchmark portfolios								
1	Low	Small	0.235	8.246	0.003	2.466	−34.32	38.82
2	2	Small	0.733	7.064	0.037	3.338	−31.30	36.67
3	3	Small	0.815	6.140	−0.087	3.233	−29.17	27.86
4	4	Small	0.998	5.724	−0.149	3.636	−29.47	27.43
5	High	Small	1.075	5.943	−0.098	4.128	−29.45	31.83
6	Low	2	0.359	7.494	−0.316	1.673	−33.21	29.62
7	2	2	0.622	6.120	−0.472	2.778	−32.50	26.25
8	3	2	0.842	5.410	−0.488	3.739	−28.10	26.78
9	4	2	0.913	5.154	−0.385	3.979	−27.09	26.70
10	High	2	0.966	5.660	−0.243	4.245	−29.83	29.43
11	Low	3	0.380	6.908	−0.342	1.365	−29.96	23.39
12	2	3	0.665	5.511	−0.624	3.051	−29.65	23.35
13	3	3	0.673	4.962	−0.621	2.984	−25.56	21.14
14	4	3	0.818	4.710	−0.327	3.029	−23.20	22.61
15	High	3	0.953	5.297	−0.423	4.384	−27.79	27.30
16	Low	4	0.506	6.146	−0.182	1.683	−26.02	25.01
17	2	4	0.433	5.218	−0.571	3.317	−29.62	20.46
18	3	4	0.651	4.855	−0.423	3.348	−26.13	22.87
19	4	4	0.820	4.622	0.035	2.394	−18.30	24.57
20	High	4	0.880	5.347	−0.198	2.697	−25.36	26.20
21	Low	Big	0.442	4.841	−0.177	1.676	−22.15	21.70
22	2	Big	0.441	4.588	−0.334	1.897	−23.21	16.05
23	3	Big	0.493	4.344	−0.243	2.639	−22.47	18.22
24	4	Big	0.603	4.289	0.056	1.574	−15.35	18.85
25	High	Big	0.583	4.645	−0.146	0.944	−19.28	15.39

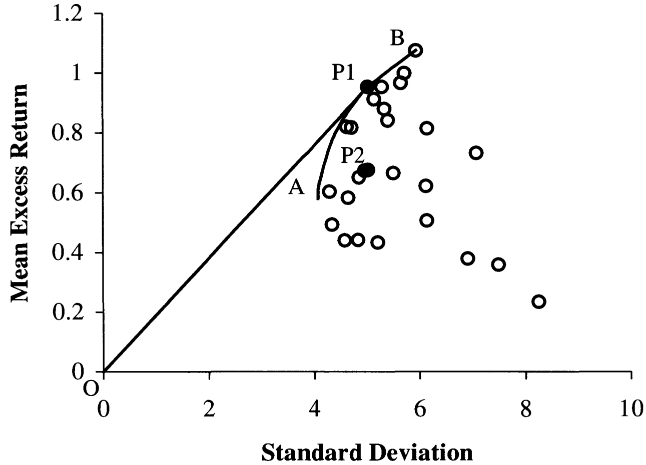

Figure 1. Mean-variance diagram. Diagram for the mean excess returns and standard deviations of the 25 risky assets (the clear dots), as well as the efficient test portfolio (P1) and the inefficient equally weighted test portfolio (P2). The 25 assets obey a multivariate normal return distribution with joint population moments equal to the sample moments of the monthly excess returns of the 25 Fama and French benchmark portfolios. The curve AB represents the efficient frontier of risky assets without short selling. If we include the riskless asset, then OP1B represents the efficient frontier.

the case without the riskless asset (AB) and the case with the riskless asset (OP1B).[22]

We analyze two different test portfolios constructed from the 26 assets. The first test portfolio (P1) is efficient; it is found by maximizing the mean subject to the constraint that the standard deviation is smaller than or equal to five percent. This portfolio consists of only three components: (1) 46.3 percent of risky asset 5, (2) 11.9 percent of risky asset 15, and (3) 41.8 percent of risky asset 19. The second test portfolio (P2) is found as the equally weighted average of all 25 risky assets. This portfolio is inefficient; it is possible to achieve a higher mean given the standard deviation, and to achieve a lower standard deviation given the mean. Both test portfolios are included in Figure 1 (the dark dots).

In this case, we know that P1 is efficient in the population and P2 is inefficient. Sampling error complicates the empirical determination of these known efficiency classifications. A Type I error occurs if P1 is wrongly classified as inefficient and a Type II error occurs if P2 is wrongly classified as efficient. We assess the size (the relative frequency of Type I error) and power (one minus the relative frequency of Type II error) of three test procedures:

[22] Note that this is the frontier for the case where short sales are not allowed. Again, we focus on the case where portfolio possibilities are described by all convex combinations of the individual assets (see Section I). Further, the frontier of risky assets (AB) tightly envelops the individual assets. This reflects the high correlation between the Fama and French benchmark portfolios.

1. A naïve SSD test procedure that rejects efficiency if and only if the test statistic is strictly positive, that is, $\xi(\tau) > 0$.
2. An asymptotic SSD test procedure that rejects efficiency if and only if the asymptotic least favorable p-value is smaller than or equal to the desired level of significance (a), that is, $\Gamma(\xi(\tau)) \leq a$.[23]
3. An asymptotic mean-variance (MV) test procedure that computes the maximal additional average return achieved by a portfolio with a variance that is lower than or equal to that of the evaluated portfolio, that is,

$$\zeta(\tau) \equiv \max_{\lambda \in \Lambda} \left\{ \sum_{t \in \Theta} (\boldsymbol{x}_t^{\mathrm{T}} \boldsymbol{\lambda} - \boldsymbol{x}_t^{\mathrm{T}} \boldsymbol{\tau})/T : \sum_{t \in \Theta} \left(\boldsymbol{x}_t^{\mathrm{T}} \boldsymbol{\lambda} - \sum_{t \in \Theta} \boldsymbol{x}_t^{\mathrm{T}} \boldsymbol{\lambda}/T\right)^2 / \boldsymbol{T} \right.$$
$$\left. \leq \sum_{t \in \Theta} \left(\boldsymbol{x}_t^{\mathrm{T}} \boldsymbol{\tau} - \sum_{t \in \Theta} \boldsymbol{x}_t^{\mathrm{T}} \boldsymbol{\tau}/T\right)^2 / \boldsymbol{T} \right\}, \tag{17}$$

and that rejects efficiency if and only if $\Gamma(\zeta(\tau)) \leq a$.[24]

To assess the statistical goodness of these procedures, we draw 10,000 random samples from the multivariate normal population distribution through Monte-Carlo simulation. For each random sample, we compute the test statistics $\xi(\tau)$ and $\zeta(\tau)$ and the p-values $\Gamma(\xi(\tau))$ and $\Gamma(\zeta(\tau))$ for P1 and P2.[25] Finally, for each procedure, we compute the size as the rejection rate for P1 and the power as the rejection rate for P2. This experiment is performed for samples of 25 to 2,000 observations and for significance levels of 2.5, 5, and 10 percent.

Figure 2 shows the rejection rates for the naïve test procedure. For every sample size, the procedure correctly rejects efficiency for P2 in *all* samples (i.e., the power equals unity). However, the rejection rate for the efficient P1 (i.e., the size of the test) is also very high, even for samples with as many as 2,000 observations. These results suggest that the naïve test procedure yields only noise in practical applications.

Figure 3 shows the size of the asymptotic SSD and MV test procedures. For both tests, the size is much lower than the level of significance. Presumably, this finding reflects the conservative nature of tests that are based on the least favor-

[23] For the simulated population distribution, H_1 (equal means) differs from H_0 (efficiency). Therefore, the simulation results may indicate whether the results obtained for H_1 are useful also for H_0 for nontrivial return distributions.

[24] We may use the asymptotic least favorable distribution of $\xi(\tau)$ also for $\zeta(\tau)$, because $\zeta(\tau)$ is bounded from above in a similar way as $\xi(\tau)$, that is, $\zeta(\tau) \leq \max_{i \in \mathrm{I}}\{\sum_{t \in \Theta} (x_{it} - \boldsymbol{x}_t{}^{\mathrm{T}}\boldsymbol{\tau})/T\}$. Ideally, we would exploit the information that the return distribution is normal to further restrict the sampling distribution of both test statistics; this information could substantially lower the p-values and increase the power.

[25] We approximate the integral $\Gamma(y) \equiv (1 - \int_{\boldsymbol{z} \leq y\boldsymbol{e}} d\Phi_{\Sigma}(\boldsymbol{z}))$using Monte-Carlo simulation. Specifically, we generate 10,000 independent standard normal random vectors $\boldsymbol{w}_s \in \Re^{N-1}$, $s \in \{1, \ldots, 10{,}000\}$, using the RNDN function in Aptech Systems' GAUSS software. Next, each random vector $\boldsymbol{w}_s$ is transformed into a multivariate normal vector $\boldsymbol{z}_s \in \Re^N$ with variance-covariance matrix $\boldsymbol{\Sigma}$, using $\boldsymbol{z}_s = \mathbf{CD}\boldsymbol{w}_s$. Finally, $\Gamma(y)$ is approximated by the relative frequency of transformed vectors $\boldsymbol{z}_s$, $s \in \{1, \cdots, 10{,}000\}$, that fall outside the integration region $\{\boldsymbol{z} \in \Re^N : \boldsymbol{z} \leq y\boldsymbol{e}\}$.

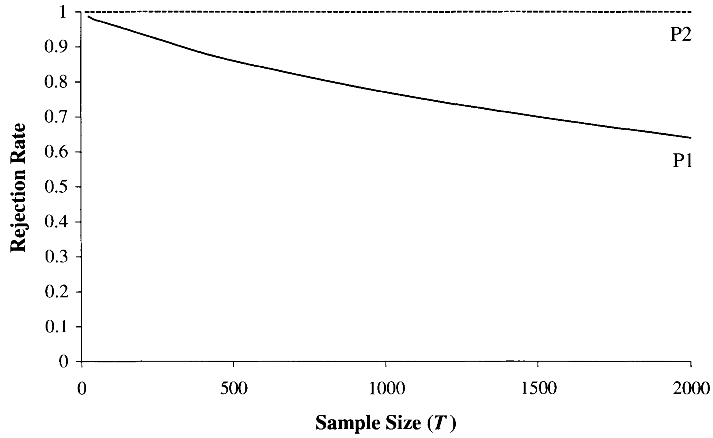

Figure 2. Rejection rates of the naïve test procedure. Rejection rates of the naïve test procedure that rejects efficiency if and only if the test statistic is strictly positive. The figure displays the rejection rates for a sample size (T) of 25 to 2,000. The dark descending line shows the results for the efficient test portfolio (P1). The gray straight line shows the results for the inefficient test portfolio (P2). 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 25 Fama and French benchmark portfolios and the U.S. Treasury bill.

able distribution. The MV test uses the prior information that the returns have a joint normal distribution, and hence the size for this test comes closest to the significance level. Figure 4 shows the power of the asymptotic test procedures. For this kind of application, we need at least 500 observations to obtain reasonable power for the SSD test. Also, for small samples, the power substantially falls short of that of the MV test. The MV test successfully exploits the prior information that the return distribution is normal, while the SSD test ignores this information. Briefly, in small samples, the SSD test lacks power, both in absolute terms and relative to the potential of the MV test. However, the good news is that the SSD test is powerful in large samples and with less Type I error (and with less risk of specification error) than the MV test.[26]

IV. Empirical Application

To further illustrate our approach, we analyze whether the Fama and French market portfolio is SSD efficient relative to all possible portfolios constructed from 25 Fama and French benchmark portfolios and the one-month U.S. Treasury

[26] A thorough comparison between MVA and SSD should be more rigorous than is possible here. First, the approaches of using prior information about the return distribution and using prior information about investor preferences are complements rather than substitutes; the analysis generally benefits from including both types of information. Second, a fair comparison would also consider other simulation conditions, including other (nonnormal) CDFs, other numbers of assets, and other correlation structures. Finally, a comparison would also benefit from comparison with other efficiency criteria, both distribution-based criteria (e.g., mean-variance-skewness analysis) and preference-based criteria (e.g., third-order SD).

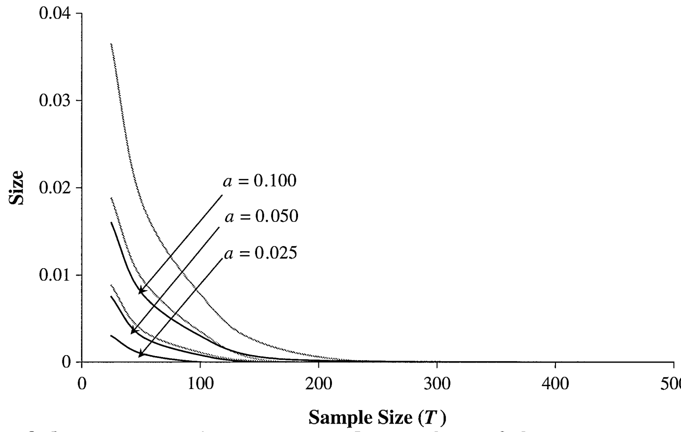

Figure 3. Size of the asymptotic test procedures. Size of the asymptotic test procedures that reject efficiency if and only if the asymptotic least favorable p-value (from Theorem 3) is smaller than or equal to the level of significance (a). The figure displays the size for levels of significance of 2.5, 5, and 10 percent, and with a sample size (T) of 25 to 500. The dark lines show the results for the SSD test, and the gray lines show the results for the MV test. 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 25 Fama and French benchmark portfolios and the U.S. Treasury bill. Size is measured as the relative frequency of random samples in which the efficient test portfolio (P1) is wrongly classified as inefficient.

bill (a riskless asset).[27] The market portfolio is the value-weighted average of all NYSE, AMEX, and Nasdaq stocks. The benchmark portfolios are the intersections of five portfolios formed on market capitalization (size) and five portfolios formed on book-to-market equity ratio (B/M). We use data on monthly returns (month-end to month-end) from July 1963 to October 2001 (460 months) obtained from the data library on the homepage of Kenneth French.[28] Our data set includes observations for a 38-year period and it is not realistic to assume that the observations are serially IID over the sample period. To account for time variation of the return distribution, we correct the raw return observations for changes in the riskless rate. Table I shows some descriptive statistics for the excess returns.

Our test results suggest that the market portfolio is SSD inefficient. The value of the test statistic is $\xi(\tau) = 0.434$. Of course, this finding is based on empirical data rather than the true (unknown) return distribution, and it is likely to be affected by sampling error in a nontrivial way (see, e.g., Figure 2). As discussed in Section III, bootstrapping is one approach to assess the sampling properties of the test results. We generate 10,000 pseudo-samples through random sampling with replacement from the EDF, and compute $\xi(\tau)$ for each pseudo-sample. Figure

[27] The Fama and French market portfolio cannot be constructed exactly as a convex combination of the 25 benchmark portfolios. Therefore, we included the market portfolio as the 27th asset. This approach does not affect the test results if the market portfolio is inefficient. However, it forces the test statistic to take a zero value if the market portfolio is efficient.

[28] The data library is found at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french.

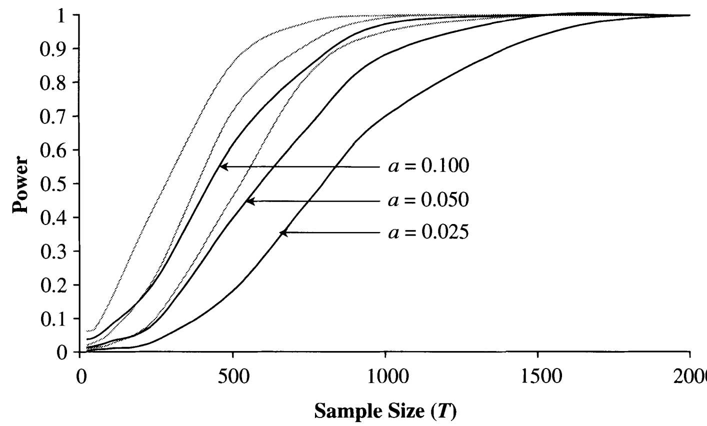

Figure 4. Power of the asymptotic test procedures. Power of the asymptotic test procedures that reject efficiency if and only if the asymptotic least favorable p-value (from Theorem 3) is smaller than or equal to the level of significance (a). The figure displays the power for levels of significance of 2.5, 5, and 10 percent, and with a sample size (T) from 25 to 2,000. The dark lines show the results for the SSD test, and the gray lines show the results for the MV test. 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 25 Fama and French benchmark portfolios and the U.S. Treasury bill. Power is measured as the relative frequency of random samples in which the inefficient equally weighted test portfolio (P2) is correctly classified as inefficient.

5 shows the resulting bootstrap distribution. The market portfolio is classified as efficient, that is, $\xi(\tau)=0$, in only 4 out of 10,000 pseudo-samples. In addition, the BCa method (see, e.g., Efron (1987)), which corrects for possible bias and skewness in the test statistic, gives a bootstrap p-value far below one percent.

The asymptotic test procedure developed in Section III gives another approach to sampling error. Again, the procedure is based on the least favorable distribution, and it may lack power for a sample of 460 observations (see, e.g., Figure 4). Still, the asymptotic least favorable p-value is as low as three percent in this case, that is, $\Gamma(0.434)=0.031$.[29]

We may ask if the inefficiency classification can be explained by additional transaction costs for benchmark portfolios that include relatively more small caps than the market portfolio. If the transaction costs are known, then we can directly adjust the return observations. If the transaction costs are not known, then extensions of our LP tests can compute the critical level of additional transaction costs required to reverse the classification. For example, the test statistic

$$\psi'(\tau,\alpha) \equiv \max_{\lambda\in\Lambda,\delta}\{\delta : \rho_T(\lambda,\tau) - \delta \geq \Gamma^{-1}(a);$$

$$\rho_t(\lambda,\tau) - \delta \geq 0\,\forall t \in \{1,\ldots,T-1\}\} \tag{18}$$

[29] Strictly speaking, we can only reject H_1 (equal means). Again, further research is needed to analyze the conditions needed to guarantee that H_1 is sufficiently close to H_0 (efficiency), or alternatively to extend our results to a less restrictive null.

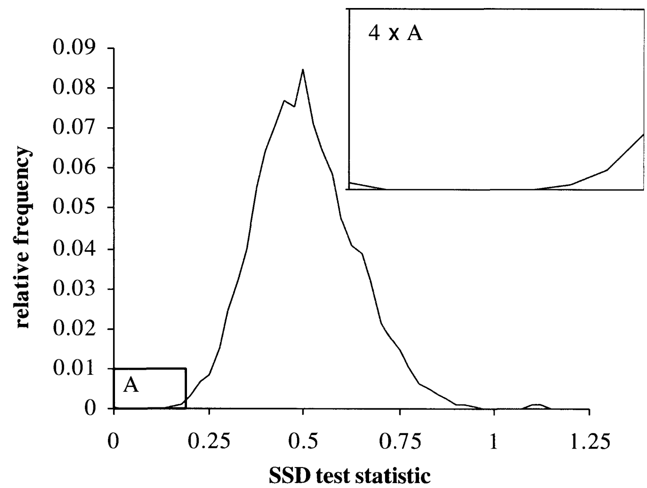

Figure 5. The bootstrap distribution of the SSD test statistic. We generate 10,000 pseudo-samples through random sampling with replacement from the EDF, and compute the SSD statistic relative to each of the samples. The Fama and French market portfolio is classified as efficient in only 4 out of 10,000 random pseudo-samples. The box in the top right corner enlarges the left tail of the distribution.

gives the maximum possible downward shift of all return observations (while keeping $\mathbf{X}^T\tau$ constant) that preserves the inefficiency classification at significance level α, that is, $\psi(\tau) \geq \Gamma^{-1}(a)$. For our application, we find $\psi'(\tau,0.10) = 0.086$, that is, transaction costs of at least 8.6 basis points per month or 103.2 basis points per year are required to reverse the inefficiency classification at a 90 percent confidence level. In a similar fashion, we may compute the critical level of additional transaction costs using bootstrapping. For our application, we find that additional transaction costs of at least 250 basis points per year are needed to achieve a bootstrap p-value of 10 percent or more. These estimates are substantially above the level of additional costs typically reported by index funds that follow small cap indexes relative to index funds that follow broad market indexes. For example, the Vanguard 500 Index Fund (based on the S&P 500 Index) currently has an expense ratio of 18 basis points *per annum*, while the Vanguard Small Cap Index (based on the Russell 2000 Index) has an expense ratio of 27 basis points.

Briefly, our findings suggest that the market portfolio is SSD inefficient to a statistically and economically significant degree. This result could be interpreted as evidence against market efficiency or alternatively as evidence against single-period, portfolio-oriented models of investment. Still, we stress that this application is used for the purpose of illustration only and we cannot draw firm conclusions at this stage. A sound empirical study should be more rigorous than is possible here. For example, we simply adopt the Fama and French market portfolio of U.S. stocks and we do not assess the effects of including foreign stocks or nonfinancial assets. Further, more sophisticated corrections for time variation of the return distribution may be required. Finally, we do not analyze the sensitivity of our results to the return horizon and the sample period. We leave these issues for further research.

V. Concluding Remarks and Suggestions

We have derived necessary and sufficient empirical tests for SSD efficiency of a given portfolio with respect to all possible portfolios constructed from a set of assets. The primal test checks whether we can construct a piecewise-linear utility function that rationalizes the evaluated portfolio. The dual test checks whether we can construct a benchmark portfolio that outperforms the evaluated portfolio in terms of the ordered mean difference.

The test statistics can be computed by solving LP problems. These problems generally are relatively small, because the number of primal variables and dual constraints equals the number of time series observations, and the number of primal constraints and dual variables roughly equals the number of time series observations plus the number of assets. Hence, the problems generally can be solved with minimal computational burden, even with desktop PCs and standard solver software. For example, the computations for our application (460 observations, 26 individual choice alternatives) used the simplex module of Aptech Systems' GAUSS software, operated on a desktop PC with a 1700 MHz Pentium IV microprocessor and with 512 MB of working memory available. The run time was less than one second per LP problem. With the current exponential growth of computer power, the computational burden can be expected to drop much further in the foreseeable future. Still, the computational complexity of the model increases as the number of time series observations and/or the number of assets increases, and specialized hardware and solver software is recommended for large-scale problems involving thousands of time series observations and/or assets (or if higher-order criteria are used; see below).

The computational ease is an important advantage relative to the Kuosmanen (2001) approach (see the introductory section). Kuosmanen focuses on identifying an efficient solution portfolio that dominates the evaluated portfolio. By contrast, our approach focuses on identifying (a set of supergradients of) the utility function that puts the evaluated portfolio in the best possible light. Sion's (1958) minimax theorem implies that both approaches are equivalent in terms of the efficiency classification (efficient or inefficient) of the evaluated portfolio (see the proof to Theorem 1). However, our dual solution portfolio generally does not dominate the evaluated portfolio and it may even be inefficient. In our opinion, this is not a substantial loss of information. First, it is relatively simple to identify efficient or even dominating portfolios. For example, one can take a specific utility function and find the portfolio that maximizes the expected utility for this function (see, e.g., Yitzhaki (1982), Kroll, Levy and Markowitz (1984), and the discussion in Levy (1998), Section 14.1). Further, Shalit and Yitzhaki (1994) suggest an iterative approach to finding efficient portfolios that dominate the evaluated portfolio. Second, the solution portfolio comes as a byproduct to the efficiency classification, and it is generally very sensitive to sampling variation. Third, the solution portfolio (even if it can be computed and if it is robust) is only one element of the efficient set, and there is no prior reason to prefer this portfolio to other efficient portfolios. For selecting the optimal portfolio, one typically needs more information than is assumed in SD.

From a theoretical perspective, SD is an attractive approach to analyzing economic behavior under uncertainty. In nonparametric fashion, it allows the data "to speak for themselves" (at least in large samples), rather than being forced to speak the idiom of a prior parametric specification of investor preferences and asset return distributions. By extending the SD approach to the case where full diversification between the choice alternatives is possible, we hope to provide a stimulus to the further proliferation of SD for the problem of portfolio selection and evaluation (as well as other choice problems under uncertainty that involve diversification possibilities). Still, this paper forms the mere starting point for developing a full framework for SD with full diversification possibilities. Further research could deal with the following subjects:

1. Our analysis of SSD efficiency is based on the first-order optimality conditions for maximizing a concave function over a polyhedron (see the proof to Theorem 2). Similar optimality conditions apply for all *pseudo-concave* functions (see, e.g., Mangasarian (1969)). All known SD criteria, including FSD, TSD, and higher-order SD criteria, assume (explicitly or implicitly) that utility is pseudo-concave, and hence they can be treated in a similar fashion as SSD. Specifically, we can generalize the primal test statistic (10) by replacing B with a more general polyhedron that represents the marginal properties of the utility functions under consideration. (The generalized dual test statistic can be obtained as the LP dual of the generalized primal test statistic.) The criteria of FSD and TSD can illustrate this approach. FSD drops the assumption of concavity. However, it maintains the assumption that utility is monotone increasing and hence pseudo-concave. We can test for FSD efficiency by dropping the constraint that the slope coefficients are decreasing, that is, by replacing B with
$$\mathrm{B}_{FSD} \equiv \{\boldsymbol{\beta} \in \Re_{+}^{T} : \beta_t \geq 1\, \forall t \in \Theta\}. \tag{19}$$
TSD (Whitmore (1970)) adds the assumption that investors prefer positively skewed distributions or that marginal utility is convex. We can test for TSD efficiency by adding the constraint that the slope coefficients decrease at a diminishing rate, that is, by using
$$\mathrm{B}_{TSD} \equiv \left\{\boldsymbol{\beta} \in \mathrm{B} : \frac{(\beta_{t+1} - \beta_t)}{(\boldsymbol{x}_{t+1}^{\mathrm{T}}\boldsymbol{\tau} - \boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau})} \leq \frac{(\beta_{t+2} - \beta_{t+1})}{(\boldsymbol{x}_{t+2}^{\mathrm{T}}\boldsymbol{\tau} - \boldsymbol{x}_{t+1}^{\mathrm{T}}\boldsymbol{\tau})}\, \forall t \in \{1, \cdots, T-2\}\right\}. \tag{20}$$
Further research could extend our results along these lines.
2. Our tests are useful for testing if a given portfolio is SD efficient relative to a given portfolio possibilities set. In many cases, it is also interesting to analyze the effect of a change in the portfolio possibilities set, for example, by the introduction of new assets (e.g., via IPOs) or the relaxation of invest-

ment restrictions for existing assets (e.g., liberalization in emerging markets). For this purpose, it is useful to develop SD tests for "spanning" (no investor benefits from an expansion of the investment possibilities) and "intersection" (some but not all investors benefit from additional investment possibilities). Thus far, the literature on spanning and intersection predominately focused on mean-variance analysis (see, e.g., Huberman and Kandel (1987)). Further research could extend our SD efficiency tests to spanning and intersection.

3. The asymptotic sampling distribution derived in Section III is based on a restrictive null hypothesis (H_1) in place of the true null of efficiency (H_0). The results of our simulations (which use a population distribution that violates H_1) suggest that the results derived for H_1 may apply also for H_0 for nontrivial return distributions. Still, further research is needed to analyze the conditions needed to guarantee that H_1 is sufficiently close to H_0, or, alternatively, to extend our results to a less restrictive null. Further, the asymptotic test procedure is based on the least favorable distribution, and it involves low power in small samples (see Figure 4). Fortunately, large data sets are available for many applications in financial economics. In addition, we could apply econometric time-series techniques to obtain an estimate for the CDF that is statistically more efficient than the EDF. We could then apply our test to a large random sample from the estimated CDF rather than the raw data. This approach effectively uses prior distribution information to generate artificial return observations. Still, it is desirable to develop more powerful tests, for example, tests that explicitly minimize the probability of Type II error rather than Type I error, and tests that are based on a particular class of return distributions.

Appendix

Proof of Theorem 1: The class of utility functions, U_2, and the feasible set, Λ, are both convex. Therefore, Sion's (1958) minimax theorem implies that we can without harm change the order of the two optimization operators:

$$\begin{aligned}&\min_{u\in U_2}\left\{\max_{\lambda\in\Lambda}\left\{\sum_{t\in\Theta}(u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda})-u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau}))/T\right\}\right\}=\\ &\max_{\lambda\in\Lambda}\left\{\min_{u\in U_2}\left\{\sum_{t\in\Theta}(u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda})-u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau}))/T\right\}\right\}.\end{aligned}\tag{A1}$$

Hence, if $\boldsymbol{\tau}\in\Lambda$ is SSD inefficient, that is, $\max_{\lambda\in\Lambda}\{\Sigma_{t\in\Theta}(u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda})-u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau}))/T\}>0\quad\forall u\in U_2$, then there exists a portfolio $\boldsymbol{\lambda}\in\Lambda$ that SSD dominates $\boldsymbol{\tau}$, that is, for which $\min_{u\in U_2}\{\Sigma_{t\in\Theta}(u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda})-u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau}))/T\}>0$. Similarly, if $\boldsymbol{\tau}\in\Lambda$ is SSD efficient, that is, $\exists u\in U_2:\max_{\lambda\in\Lambda}\{\Sigma_{t\in\Theta}(u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda})-u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau}))/T\}=0$, then $\min_{u\in U_2}\{\Sigma_{t\in\Theta}(u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda})-u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau}))/T\}\leq 0$ for all $\boldsymbol{\lambda}\in\Lambda$, and there does not exist a portfolio $\boldsymbol{\lambda}\in\Lambda$ that SSD dominates $\boldsymbol{\tau}$. Q.E.D.

Proof of Theorem 2: We will use the first-order optimality conditions for maximizing a concave function over a polyhedral set (see, e.g., Hiriart-Urruty and Lemaréchal (1993), Thm. VII:1.1.1 and Cond. VII: 1.1.3). Specifically, $\boldsymbol{\tau}$ is an optimal portfolio, that is, $\boldsymbol{\tau} = \arg\max_{\lambda \in \Lambda} \Sigma_{t\in\Theta} u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda})/T$ for $u \in U_2$ if and only if all assets $i \in \mathrm{I}$ are enveloped by the tangent hyperplane defined by a supergradient vector $\partial u(\boldsymbol{\tau}) \equiv (\partial u(\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{\tau}) \cdots \partial u(\boldsymbol{x}_T^{\mathrm{T}}\boldsymbol{\tau}))^{\mathrm{T}}$, that is,

$$\sum_{t\in\Theta} \partial u(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau})(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau} - x_{it})/T \geq 0 \qquad \forall i \in \mathrm{I}. \tag{A2}$$

This condition is similar to the well-known Kuhn–Tucker conditions for selecting an optimal portfolio if short sales are not allowed. The Kuhn–Tucker conditions apply for continuously differentiable utility functions. Condition (A2) is a generalization to superdifferentiable utility functions (recall that we use piecewise-linear utility functions, which generally are not continuously differentiable).

If $\boldsymbol{\tau}$ is optimal relative to some $u \in U_2$, then it is also optimal relative to the standardized utility function $v \equiv (u/\partial u(\boldsymbol{x}_T^{\mathrm{T}}\boldsymbol{\tau})) \in U_2$. By construction, $\partial v(\boldsymbol{\tau})$ is a feasible solution to the primal problem, that is, $\partial v(\boldsymbol{\tau}) \in \mathrm{B}$. Optimality condition (A2) implies that this solution is associated with a solution value of zero. Hence, we find the necessary condition; $\boldsymbol{\tau}$ is SSD efficient only if $\xi(\boldsymbol{\tau}) = 0$.

To establish the sufficient condition, use $\boldsymbol{\beta}^* \equiv (\beta_1^* \cdots \beta_T^*)^{\mathrm{T}} \in \mathrm{B}$ for the optimal solution to the primal problem. The piecewise-linear function $p(x|\boldsymbol{\alpha},\boldsymbol{\beta}^*)$ connects the points $(p(x|\boldsymbol{\alpha},\boldsymbol{\beta}^*),z_t)$, $z_t \in [\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau}, \boldsymbol{x}_{t+1}^{\mathrm{T}}\boldsymbol{\tau}]$, $t \in \{1, \ldots, T-1\}$, if we choose α_t such that $\alpha_t + \beta_t{}^* z_t = \alpha_{t+1} + \beta_{t+1}^* z_t$ for all $t \in \Theta$. This can be achieved, for example, by setting $\alpha_t = \alpha_t^* \equiv \Sigma_{s=t}^{T-1}(\beta_{s+1}^* - \beta_s^*)z_s$, $t \in \{1, \cdots, T-1\}$, and $\alpha_T = 0$. We then find

$$p(x|\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = \begin{cases} \sum\limits_{s=1}^{T-1} (\beta_{s+1}^* - \beta_s^*)z_s + \beta_1^* x & x \leq z_1 \\ \sum\limits_{s=2}^{T-1} (\beta_{s+1}^* - \beta_s^*)z_s + \beta_2^* x & z_1 \leq x \leq z_2 \\ \vdots & \\ (1 - \beta_{T-1}^*)z_{T-1} + \beta_{T-1}^* x & z_{T-2} \leq x \leq z_{T-1} \\ x & x \geq z_{T-1} \end{cases} . \tag{A3}$$

By construction, $p(x|\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*)$ is strictly increasing and concave and hence it belongs to U_2. If $\xi(\boldsymbol{\tau}) = 0$ and hence $\Sigma_{t\in\Theta}\beta_t^*(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau} - x_{it})/T \geq 0 \quad \forall i \in \mathrm{I}$, then optimality condition (A2) implies that $\boldsymbol{\tau}$ is optimal relative to $p(x|\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*)$, that is, $\boldsymbol{\tau} = \arg\max_{\lambda\in\Lambda} \Sigma_{t\in\Theta} p(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda}|\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$. Hence, we find the sufficient condition; portfolio $\boldsymbol{\tau} \in \Lambda$ is SSD efficient if $\xi(\boldsymbol{\tau}) = 0$.

The statistic $\psi(\boldsymbol{\tau})$ is derived using linear duality theory. The following is a full LP formulation for $\xi(\boldsymbol{\tau})$:

$$\min_{\theta,\boldsymbol{\beta}} \theta \tag{P}$$

$$\text{s.t.} \sum_{t=1}^{T} \beta_t(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau} - x_{it})/T + \theta \geq 0 \; i = 1, \ldots, N \, [\lambda_i]$$

$$\begin{aligned}
&\beta_{t-1} - \beta_t \geq 0 \, t = 1, \ldots, T-1 \, [\rho_t] \\
&\beta_t \geq 0 \quad t = 1, \ldots, T-1 \\
&\beta_T = 1 \\
&\theta \quad \text{free}
\end{aligned}$$

The shadow prices to the constraints are given within brackets. This information is useful for interpreting the dual formulation of (P):

$$\max_{\boldsymbol{\lambda}, \rho_1, \cdots, \rho_{T-1}} \sum_{i=1}^{N} (\lambda_i x_{iT} - \boldsymbol{x}_T^{\mathrm{T}}\boldsymbol{\tau})/T + \rho_{T-1} \tag{D}$$

$$\text{s.t. } \rho_1 + (\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{\tau} - \sum_{i=1}^{N} \lambda_i x_{i1})/T \leq 0 \, [\beta_1]$$

$$\rho_t - \rho_{t-1} + (\boldsymbol{x}_t^{\mathrm{T}}\tau - \sum_{i=1}^{N} \lambda_i x_{it})/T \leq 0 \; t = 2, \cdots, T-1 \; [\beta_t]$$

$$\sum_{i=1}^{N} \lambda_i = 1 \; [\theta]$$

$$\begin{aligned}
&\lambda_i \geq 0 \quad i = 1, \ldots, N \\
&\rho_t \geq 0 \quad t = 1, \ldots, T-1
\end{aligned}$$

Again, the primal variables that correspond to each of the dual constraints are given within brackets, so as facilitate the interpretation and the relationship between the primal and the dual. The first $T-1$ inequality constraints are always binding (the primal positivity constraints $\beta_t \geq 0$, $t = 1, \cdots, T-1$, are redundant because the primal also requires $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_T = 1$). Hence, the optimal solution is $\rho_t^* = \Sigma_{s=1}^{t}(\boldsymbol{x}_s^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{x}_s^{\mathrm{T}}\boldsymbol{\tau})/T = t\rho_t(\boldsymbol{\lambda}, \boldsymbol{\tau})/T$ for $t = 1, \cdots, T-1$. Substituting these expressions in (P), and dividing each $\rho_t{}^*$ term by T/t, we directly obtain (11). Since (P) and (D) are feasible and bounded, it follows from the Duality Theorem of Linear Programming that they have the same optimal solution value, that is, $\psi(\boldsymbol{\tau}) = \xi(\boldsymbol{\tau})$. Q.E.D.

Proof of Theorem 3: Known results can derive the exact asymptotic sampling distribution of $\omega(\boldsymbol{\tau}) \equiv \max_{i \in \mathrm{I}}\left\{\Sigma_{t \in \Theta}(x_{it} - \boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau})/T\right\}$. Under the null $H_1 : E[\boldsymbol{x}] = \mu\boldsymbol{e}$, the elements of $\mathbf{C}\boldsymbol{x}_t = ((x_{1t} - \boldsymbol{x}_t{}^{\mathrm{T}}\boldsymbol{\tau}) \cdots (x_{Nt} - \boldsymbol{x}_t{}^{\mathrm{T}}\boldsymbol{\tau})^{\mathrm{T}}$, $t \in \Theta$, are serially IID random vectors with zero mean and variance-covariance matrix $\mathbf{C}\boldsymbol{\Omega}\mathbf{C}^{\mathrm{T}}$. Hence, the Lindeberg–Levy central limit theorem implies that the vector

$\mathbf{CX}\boldsymbol{e}/T = (\Sigma_{t\in\Theta}(x_{1t} - \boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau})/T \cdots \Sigma_{t\in\Theta}(x_{Nt} - \boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau})/T)^{\mathrm{T}}$ obeys an asymptotically joint normal distribution with zero mean and variance-covariance matrix $\boldsymbol{\Sigma} \equiv (\mathbf{C\Omega C}^{\mathrm{T}})/T$. Hence, $\omega(\boldsymbol{\tau})$ asymptotically behaves as the largest order statistic of N random variables with a multivariate normal distribution, and $\Pr[\omega(\boldsymbol{\tau}) > y|H_1] = 1 - \Pr[\mathbf{CX}\boldsymbol{e}/T \leq y\boldsymbol{e}]$ asymptotically equals the multivariate normal integral $\Gamma(y) \equiv (1 - \int_{\boldsymbol{z}\leq y\boldsymbol{e}} d\Phi_{\boldsymbol{\Sigma}}(\boldsymbol{z}))$. Since the unity vector is a feasible solution to the primal problem, that is, $\boldsymbol{e} \in \mathrm{B}$, we know that $\xi(\boldsymbol{\tau}) \leq \omega(\boldsymbol{\tau})$, and hence $\Pr[\xi(\boldsymbol{\tau}) > y|H_1]$ is bounded from above by $\Pr[\omega(\boldsymbol{\tau}) > y|H_1]$ for all return distributions $G(\boldsymbol{x})$ with bounded variance-covariance matrix. Moreover, there exist $G(\boldsymbol{x})$ for which $\xi(\boldsymbol{\tau})$ approximates $\omega(\boldsymbol{\tau})$, and therefore the asymptotic distribution of $\omega(\boldsymbol{\tau})$ also represents the asymptotic least favorable distribution for $\xi(\boldsymbol{\tau})$. Q.E.D.

REFERENCES

Afriat, Sydney N., 1967, The construction of a utility function from expenditure data, *International Economic Review* 8, 67–77.

Arrow, Kenneth, 1971, *Essays in the Theory of Risk-Bearing* (North-Holland Publishing Company, Amsterdam).

Bawa, Vijay, James Bodurtha Jr., M. R. Rao, and Hira Suri, 1985, On determination of stochastic dominance optimal sets, *Journal of Finance* 40, 417–431.

Beach, Charles, and Russell Davidson, 1983, Distribution-free statistical inference with Lorenz curves and income shares, *Review of Economic Studies* 50, 723–735.

Bowden, Roger, 2000, The ordered mean difference as a portfolio performance measure, *Journal of Empirical Finance* 7, 195–223.

Breeden, Douglas T., Michael R. Gibbons, and Robert H. Litzenberger, 1989, Empirical tests of the consumption-oriented CAPM, *Journal of Finance* 44, 231–262.

Chvatal, Vasek, 1983, *Linear Programming* (Freeman, New York).

Dardanoni, Valentino, and Antonio Forcina, 1999, Inference for the Lorenz curve orderings, *Econometrics Journal* 2, 49–75.

Davidson, Russell, and Jean-Yves Duclos, 2000, Statistical inference for stochastic dominance and for the measurement of poverty and inequality, *Econometrica* 68, 1435–1464.

Debreu, Gerard, 1951, The coefficient of resource utilization, *Econometrica* 19, 273–292.

Dybvig, Phillip, and Stephen Ross, 1982, Portfolio efficient sets, *Econometrica* 50, 1525–1546.

Efron, Bradley, 1979, Bootstrap methods: Another look at the jackknife, *Annals of Statistics* 7, 1–26.

Efron, Bradley, 1987, Better bootstrap confidence intervals, *Journal of the American Statistical Association* 82, 171–200.

Efron, Bradley, and Gail Gong, 1983, A leisurely look at the bootstrap, the jackknife, and cross validation, *The American Statistician* 37, 36–48.

Farrell, Michael J., 1957, The measurement of productive efficiency, *Journal of the Royal Statistical Society Series A* 120, 253–281.

Fishburn, Peter C., 1974, Convex stochastic dominance with continuous distribution functions, *Journal of Economic Theory* 7, 143–158.

Freedman, David A., 1981, Bootstrapping regression methods, *Annals of Statistics* 9, 1218–1228.

Genz, Alan, and Koon-Shing Kwong, 2000, Numerical evaluation of singular multivariate normal distributions, *Journal of Statistical Computational and Simulation* 68, 1–21.

Ghysels, Eric, 1998, On stable factor structures in the pricing of risk: Do time-varying betas help or hurt?, *Journal of Finance* 53, 549–573.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

Green, Richard C., and Sanjay Srivastava, 1985, Risk aversion and arbitrage, *Journal of Finance* 40, 257–268.

Green, Richard C., and Sanjay Srivastava, 1986, Expected utility maximization and demand behavior, *Journal of Economic Theory* 38, 313–323.

Hadar, Josef, and William R. Russell, 1969, Rules for ordering uncertain prospects, *American Economic Review* 59, 25–34.

Hanoch, Giora, and Haim Levy, 1969, The efficiency analysis of choices involving risk, *Review of Economic Studies* 36, 335–346.

Hiriart-Urruty, Jean-Baptiste, and Claude Lemaréchal, 1993, *Convex Analysis and Minimization Algorithms I & II* (Springer Verlag, Berlin).

Huang, Chi-fu, and Robert H. Litzenberger, 1988, *Foundations for Financial Economics* (North Holland, Amsterdam).

Huberman, Gur, and Shmuel Kandel, 1987, Mean-variance spanning, *Journal of Finance* 42, 873–888.

Kandel, Shmuel, and Robert Stambaugh, 1987, On correlations and inferences about mean variance efficiency, *Journal of Financial Economics* 18, 61–90.

Kandel, Shmuel, and Robert Stambaugh, 1989, A mean-variance framework for tests of asset pricing models, *Review of Financial Studies* 2, 125–156.

Kroll, Yoram, and Haim Levy, 1980, Sampling errors and portfolio efficient analysis, *Journal of Financial and Quantitative Analysis* 15, 655–688.

Kroll, Yoram, Haim Levy, and Harry Markowitz, 1984, Mean variance versus direct utility maximization, *Journal of Finance* 39, 47–61.

Kuntz, Ludwig, and Stefan Scholtes, 2000, Measuring the robustness of empirical efficiency valuations, *Management Science* 46, 807–823.

Kuosmanen, Timo, 2001, Stochastic dominance efficiency tests under diversification, Working paper W-283, Helsinki School of Economics.

Levy, Haim, 1992, Stochastic dominance and expected utility: Survey and analysis, *Management Science* 38, 555–593.

Levy, Haim, 1998, *Stochastic Dominance* (Kluwer Academic Publishers, Norwell, MA).

Machina, Mark J., 1982, Expected utility analysis without the independence axiom, *Econometrica* 50, 277–323.

Mangasarian, Olvi L., 1969, *Nonlinear Programming* (McGraw-Hill, New York).

Nelson, Ray D., and Rulon Pope, 1991, Bootstrapped insights into empirical applications of stochastic dominance, *Management Science* 37, 1182–1194.

Rothschild, Michael, and Joseph E. Stiglitz, 1970, Increasing risk I: A definition, *Journal of Economic Theory* 2, 225–243.

Shalit, Haim, and Shlomo Yitzhaki, 1994, Marginal conditional stochastic dominance, *Management Science* 39, 670–684.

Sharpe, William, 1991, Capital asset prices with and without negative holdings, *Journal of Finance* 64, 489–509.

Simar, Leopold, and Paul Wilson, 1998, Sensitivity analysis of efficiency scores: How to bootstrap in non-parametric frontier models, *Management Science* 44, 49–61.

Sion, Maurice, 1958, On general minimax theorems, *Pacific Journal of Mathematics* 8, 171–176.

Starmer, Chris, 2000, Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk, *Journal of Economic Literature* 38, 332–382.

Stein, William, and Roger Pfaffenberger, 1986, Estimation of error probabilities in stochastic dominance, *Journal of Statistical Computation and Simulation* 25, 137–165.

Varian, Hal, 1982, The non-parametric approach to demand analysis, *Econometrica* 50, 945–973.

Varian, Hal, 1983, Nonparametric tests of models of investor behavior, *Journal of Financial and Quantitative Analysis* 18, 269–278.

Whitmore, G. Alexander, 1970, Third-degree stochastic dominance, *American Economic Review* 60, 457–459.

Yamakazi, Hiroaki, and Hiroshi Konno, 1991, Mean absolute deviation portfolio optimization model and its application to Tokyo stock market, *Management Science* 37, 519–531.

Yitzhaki, Shlomo, 1982, Stochastic dominance mean-variance, and Gini's mean difference, *American Economic Review* 72, 178–185.

Young, Martin, 1998, A minimax portfolio selection rule with linear programming solution, *Management Science* 44, 673–683.