Asset prices and omitted moments

A stochastic dominance analysis of market efficiency

Thierry Post

ERIM REPORT SERIES *RESEARCH IN MANAGEMENT*	
ERIM Report Series reference number	ERS-2003-017-F&A
Publication status / version	June 2003
Number of pages	31
Email address corresponding author	gtpost@few.eur.nl
Address	Erasmus Research Institute of Management (ERIM) Rotterdam School of Management / Faculteit Bedrijfskunde Rotterdam School of Economics / Faculteit Economische Wetenschappen Erasmus Universiteit Rotterdam PoBox 1738 3000 DR Rotterdam, The Netherlands Phone: # 31-(0) 10-408 1182 Fax: # 31-(0) 10-408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website:
www.erim.eur.nl

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES
RESEARCH IN MANAGEMENT

BIBLIOGRAPHIC DATA AND CLASSIFICATIONS		
Abstract	We analyze if the value-weighted stock market portfolio is second-order stochastic dominance (SSD) efficient relative to benchmark portfolios formed on market capitalization, book-to-market equity ratio and industry classification. During the period from the mid-1970s to the late 1980s, the market portfolio is significantly mean-variance inefficient. During this period, the market portfolio generally also is significantly SSD inefficient. This suggests that mean-variance inefficiency cannot be explained by omitted return moments like higher-order central moments or lower partial moments.	
Library of Congress Classification (LCC)	5001-6182	Business
	5601-5689	Accountancy, Bookkeeping
	4001-4280.7	Finance Management, Business Finance, Corporation Finance
	HG 4027	Financial management: Statitstics
Journal of Economic Literature (JEL)	M	Business Administration and Business Economics
	M 41	Accounting
	G 3	Corporate Finance and Governance
	C 19	Statistical methods: other
	G 12	Asset pricing
European Business Schools Library Group (EBSLG)	85 A	Business General
	225 A	Accounting General
	220 A	Financial Management
	220 T	Quantitative methods for financial management
Gemeenschappelijke Onderwerpsontsluiting (GOO)		
Classification GOO	85.00	Bedrijfskunde, Organisatiekunde: algemeen
	85.25	Accounting
	85.30	Financieel management, financiering
	85.30	Financieel management, financiering
	31.73	Wiskundige statistiek
Keywords GOO	Bedrijfskunde / Bedrijfseconomie	
	Accountancy, financieel management, bedrijfsfinanciering, besliskunde	
	Effectenhandel, portfolio-analyse, stochastsiche methoden, bootstrap (statistiek), evenwichtsmodellen (economie) , kapitaalmarkt	
Free keywords	stochastic dominance, market efficiency, asset pricing, statistical inference, size and book-to-market effects	

Asset prices and omitted moments

A stochastic dominance analysis of market efficiency

THIERRY POST*

We analyze if the value-weighted stock market portfolio is second-order stochastic dominance (SSD) efficient relative to benchmark portfolios formed on market capitalization, book-to-market equity ratio and industry classification. During the period from the mid-1970s to the late 1980s, the market portfolio is significantly mean-variance inefficient. During this period, the market portfolio generally also is significantly SSD inefficient. This suggests that mean-variance inefficiency cannot be explained by omitted return moments like higher-order central moments or lower partial moments.

Post is corresponding author: Erasmus University Rotterdam, P.O. Box 1738, 3000 DR, Rotterdam, The Netherlands, email: gtpost@few.eur.nl, tel: +31-104081428. This study forms part of a research program on stochastic dominance and financial markets. Details on the program are available at the program homepage: http://www.few.eur.nl/few/people/gtpost/stochastic_dominance.htm*. We appreciate the comments by Roger Bowden, Winfried Hallerbach, Haim Levy, Jan van der Meulen, Nico van der Sar and Jaap Spronk, as well as participants at the 30th EURO Working Group on Financial Modeling, the 29th annual meeting of the European Finance Association, and seminars of Center for Economic Studies (Catholic University of Leuven), Erasmus Center of Financial Research (Erasmus University Rotterdam) and Ghent Finance Center (Ghent University). Martijn van den Assem, Peter van der Spek and Pim van Vliet provided invaluable research assistance. Financial support by Tinbergen Institute, Erasmus Research Institute of Management and Erasmus Center of Financial Research is gratefully acknowledged. Any remaining errors are our own.

Asset prices and omitted return moments

A stochastic dominance analysis of market efficiency

We analyze if the value-weighted stock market portfolio is second-order stochastic dominance (SSD) efficient relative to benchmark portfolios formed on market capitalization, book-to-market equity ratio and industry classification. During the period from the mid-1970s to the late 1980s, the market portfolio is significantly mean-variance inefficient. During this period, the market portfolio generally also is significantly SSD inefficient. This suggests that mean-variance inefficiency cannot be explained by omitted return moments like higher-order central moments or lower partial moments.

THE TRADITIONAL MEAN-VARIANCE CAPITAL ASSET PRICING MODEL (CAPM) by Sharpe (1964) and Lintner (1965) predicts that the value-weighted market portfolio of risky assets is mean-variance efficient, or, equivalently, that there exists an exact positive linear relationship between assets' mean return and market co-variance. Several empirical regularities seem to contradict this prediction (see, e.g., Schwert (2002) for a recent survey of asset pricing anomalies). For example, common market proxies seem significantly mean-variance inefficient relative to portfolios formed on market capitalization (size) and book-to-market equity ratio (BE/ME). Related to this, market co-variance seems to explain only a small portion of the cross-sectional variation in mean return, while size and BE/ME appear to have substantial explanatory power (see, e.g., Fama and French (1992)). Roughly speaking, value stocks and small caps seem to earn above-normal returns, and we can 'beat the market' by assigning these stocks a higher portfolio weight than their relative market capitalization.

The mean-variance CAPM is a relatively simple single period, portfolio-oriented, representative investor model, and it can be extended in many ways. One approach is to change the maintained assumptions on investor preferences. If we do not restrict the shape of the return distribution, then the mean-variance model is consistent with expected utility theory only if utility takes a quadratic form.[1] Extensions can be obtained by using alternative classes of utility. For example, the three-moment CAPM, used by, e.g., Kraus and Litzenberger (1976) and Harvey and Siddique (2000), assumes a cubic utility function (or a third-order Taylor series approximation to the true utility function), which implies that investors care about the first three central moments of the return distribution (mean, variance and skewness). While altering the shape of the utility function, these extended models still assume a representative investor who holds the market portfolio.[2] The most common approach to test these models is by testing the first-order optimality condition (or Euler equation) for the market portfolio. This condition implies an exact linear relationship between assets' co-moments with the market portfolio. For example, the three-moment CAPM predicts an exact linear relationship between mean, co-variance and co-skewness.

A difficulty in changing the preference assumptions is the need to give a parametric specification of the functional form of the utility function. Economic theory gives only minimal guidance for this purpose, and there is a substantial risk of specification error. For example, the three-moment CAPM ignores the central moments of order higher than three (e.g., kurtosis), as well as the lower partial moments (e.g., semi-variance), which generally cannot be expressed in terms of the first three central moments.[3] Another problem associated with low order polynomials is the difficulty to impose the conditions that do have economic meaning. For example, we cannot impose the condition of nonsatiation by restricting a quadratic function to be monotone increasing, and we cannot impose the condition of risk aversion by restricting a cubic function to be concave (see, e.g., Levy (1969)).

To circumvent these problems, we may use rules of stochastic dominance (SD; see, e.g., Levy (1992)). SD rules do not require a parametric specification of the preferences of the decision-makers and the statistical distribution of the choice alternatives. Rather, they impose only general and economically meaningful preference conditions and they effectively consider the entire return distribution rather than a finite set of moments. For example, the popular criterion of second-order SD (SSD) assumes only non-satiation and risk aversion. Hence, the market portfolio must be SSD efficient for all asset-pricing models that use a nonsatiable and risk-averse representative investor, regardless of the specific functional form of the utility function and the specific shape of the return distribution. SSD inefficiency would imply that all such models would fail to rationalize the market portfolio, and that the omission of higher order central moments and partial lower partial moments cannot explain inefficiency.

Post (2003) develops a tractable linear programming test for SSD efficiency of a given portfolio relative to all portfolios formed from a set of assets. To illustrate his test, he tests if the value-weighted US stock market portfolio is SSD efficient relative to benchmark portfolios formed on size and BE/ME. Interestingly, his results suggest that the market portfolio is significantly SSD inefficient. Again, a SSD inefficiency classification is highly interesting, because it effectively rules out explanations based on omitted return moments. Still, Post's analysis is used for the purpose of illustration only and a sound empirical study requires more rigor. The purpose of this study is to provide a more rigorous study of SSD efficiency of the market portfolio:

1. Post uses a 460-month or 39-year sample period (July 1963 to October 2001) and he assumes that the distribution of excess returns remains unchanged for the entire period. However, there exists strong evidence that the return distribution (e.g., the market risk premium) varies through time (see, e.g., Campbell (1987) and Jagannathan and Wang (1996)). Ideally, we would circumvent this problem by developing a conditional test for SSD efficiency that links the ex ante return distribution to the investors' (time-varying) information set. 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 take another approach to account for time-variation. Specifically, we use a rolling window analysis that applies the SSD efficiency test to short subsamples of 60, 90 and 120 months. This approach is far less sensitive to time-variation, since it assumes that the distribution of excess returns is fixed in the subsamples rather than in the full sample.

2. The finding that the market portfolio is mean-variance inefficient relative to size and BE/ME sorted portfolios is sometimes criticized as being the result of data-snooping bias. Specifically, the benchmark portfolios are formed on firm characteristics (size and BE/ME) that are *a priori* known to be correlated with

expected returns (see, e.g., Lo and MacKinlay (1990)). After all, the size and BE/ME effects were well known prior to the influential Fama and French studies (see, e.g., Banz (1981) and Rosenberg *et al.* (1985)). To circumvent this problem, we also test if the market portfolio is SSD efficient relative to benchmark portfolios formed on industry classification. Industry portfolios are not sorted by a variable that is known to be correlated with expected returns, and hence they are less susceptible to data-snooping bias.

3. A practical complication associated with the SSD test is its high sensitivity to sampling error. Post derived the asymptotic sampling distribution of his SSD test statistic under particular statistical assumptions. Unfortunately, a test procedure that uses this distribution involves reasonable power (the probability of correctly rejecting efficiency for an inefficient portfolio) only for large data sets. Presumably, this reflects the use of the least favorable distribution, which minimizes the statistical size (the probability of wrongly rejecting efficiency for an efficient portfolio) at the cost of power. For our purpose, the asymptotic sampling distribution will not be useful, because we focus on small samples of only 60 to 120 monthly observations. Rather, we will use a test procedure that uses the bootstrap methodology (see, e.g., Efron (1979) and Efron and Gong (1983)). A simulation experiment demonstrates that this procedure involves substantially more power than the asymptotic sampling distribution. In fact, the size and power are reasonable even for samples of only 60 observations.

We analyze if erroneous preference and distribution assumptions, or, equivalently, omitted return moments, can explain the empirical failure of mean-variance CAPM. To focus on the role of preference and distribution assumptions, we largely adhere to the remaining assumptions of the mean-variance CAPM: we use a single-period, portfolio-oriented, representative investor model. Of course, there are good reasons to doubt our maintained assumptions, and to believe that our results are affected by these assumptions in a non-trivial way. Still, we believe that our approach is useful, as we have to 'walk before we can run', and the analysis can form the starting point for further research based on more general assumptions (e.g., by considering heterogeneous preferences and beliefs or the multiperiod consumption-investment problem).

The remainder of this study is structured as follows. Section I introduces the notation, definitions and assumptions that will be used throughout the text. Section II analyzes the statistical size and power properties of a test procedure that uses the asymptotic sampling distribution and a test procedure that uses the bootstrap. Section III empirically analyzes SSD efficiency of the value-weighted market portfolio. Finally, Section IV summarizes our conclusions and presents directions for further research.

I. SSD efficiency

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

1. Investors are nonsatiable and risk averse and they select investment portfolios to maximize the expected utility associated with the return of their investment

portfolio. Throughout the text, we will denote utility functions by $u:\Re \to P$, $u \in U_2$, with U_2 for the class of strictly increasing and concave, continuously differentiable, von Neumann-Morgenstern utility functions, and P for a nonempty, closed, and convex subset of $\Re$.[4]

2. The investment universe consists of N assets, associated with returns $\boldsymbol{x} \in \Re^N$. Throughout the text, we will use the index set $\mathrm{I} \equiv \{1,\cdots,N\}$ to denote the different assets. The returns are serially independent and identically distributed (IID) random variables with a continuous joint cumulative distribution function (CDF) $G:\Re^N \to [0,1]$. Further, the returns have means $E[\boldsymbol{x}] = \boldsymbol{\mu}$ and bounded variance-covariance matrix $E[(\boldsymbol{x}-m)(\boldsymbol{x}-m)^{\mathrm{T}}] = \mathrm{W} < \infty$.

3. Investors may diversify between the assets, and we will use $l \in \Re^N$ 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 $\Lambda \equiv \{l \in \Re_+^N : \boldsymbol{e}^{\mathrm{T}} l = 1\}$, with $\boldsymbol{e}$ for a unity vector with dimensions conforming to the rules of matrix algebra.[5]

Under these assumptions, the investors' optimization problem can be summarized as $\max_{l \in \Lambda} \int u(\boldsymbol{x}^{\mathrm{T}} l)dG(\boldsymbol{x})$. Post's (2003) test statistic is based on the first-order condition for this problem. Specifically, a given portfolio $t \in \Lambda$ is optimal for a given utility function $u \in U_2$ if and only if

$$\int \partial u(\boldsymbol{x}^{\mathrm{T}} t)(\boldsymbol{x}^{\mathrm{T}} t - x_i)dG(\boldsymbol{x}) \geq 0 \quad \forall i \in \mathrm{I}, \tag{1}$$

with $\partial u(x)$ for a supergradient at x.[6] This naturally leads to the following measure for SSD efficiency:

$$x(t,G) \equiv \min_{u \in U_2}\left\{\max_{i \in \mathrm{I}}\left\{\int \partial u(\boldsymbol{x}^{\mathrm{T}} t)(x_i - \boldsymbol{x}^{\mathrm{T}} t)dG(\boldsymbol{x})\right\}\right\}. \tag{2}$$

DEFINITION 1 *Portfolio $t \in \Lambda$ is SSD efficient if and only if it is optimal for at least some $u \in U_2$, i.e., $x(t,G) = 0$.*

To test the null of SSD efficiency, i.e., $H_0 : x(t,G) = 0$, we need full information on the CDF $G(\boldsymbol{x})$. In practical applications, $G(\boldsymbol{x})$ generally is not known and information is limited to a discrete set of T time series observations. We assume that observations are independent random draws from the CDF. Throughout the text, we will represent the observations by the matrix $\mathrm{C} \equiv (\boldsymbol{x}_1 \cdots \boldsymbol{x}_T)$, with $\boldsymbol{x}_t \equiv (x_{1t} \cdots x_{Nt})^{\mathrm{T}}$. Since the timing of the draws is inconsequential, we are free to label the observations by their ranking with respect to the evaluated portfolio, i.e., $\boldsymbol{x}_1^{\mathrm{T}} t < \boldsymbol{x}_2^{\mathrm{T}} t < \cdots < \boldsymbol{x}_T^{\mathrm{T}} t$. Using the observations, we can construct the empirical distribution function (EDF):

$$F_{\mathrm{C}}(\boldsymbol{x}) \equiv \mathrm{card}\{t \in \{1,\cdots,T\} : \boldsymbol{x}_t \leq \boldsymbol{x}\}/T, \tag{3}$$

with $\text{card}\{\cdot\}$ for the number of elements of a set. Since the observations are serially IID distributed, $F_{\mathrm{C}}(\boldsymbol{x})$ is a consistent estimator for $G(\boldsymbol{x})$. This suggests that we can use $x(t, F_{\mathrm{C}})$ as a consistent estimator for $x(t, G)$.

For computing $x(t, F_{\mathrm{C}})$, Post (2003) derives the following linear programming formulation:

$$x(t, F_{\mathrm{C}}) = \min_{b \in \mathrm{B}, q} \left\{ q : \sum_{t=1}^{T} b_t (\boldsymbol{x}_t^{\mathrm{T}} t - x_{it}) / T + q \geq 0 \quad \forall i \in \mathrm{I} \right\}, \tag{4}$$

with $\mathrm{B} \equiv \left\{ b \in \Re_+^T : b_1 \geq b_2 \geq \cdots \geq b_T = 1 \right\}$. The optimal solution $b \in \mathrm{B}$, say b^*, represents a supergradient vector $(\partial u^*(\boldsymbol{x}_1^{\mathrm{T}} t) \cdots \partial u^*(\boldsymbol{x}_T^{\mathrm{T}} t))^T$ for the optimal utility function $u^*(t)$. B represents the restrictions on the supergradient vector that follow from the assumptions of nonsatiation and risk aversion, as well as the (harmless) standardization $\partial u(\boldsymbol{x}_T^{\mathrm{T}} t) = 1$.[7]

In our analysis, we will also use a test for mean-variance efficiency. Mean-variance efficiency is equivalent to SSD efficiency with the additional restriction that the utility function takes a quadratic form, i.e., $u \in Q \equiv \left\{ u \in U_2 : u(x) = r_0 x + r_1 x^2 \right\}$.[8] By analogy to (2), this leads to the following measure for mean-variance efficiency

$$z(t, G) \equiv \min_{u \in Q} \left\{ \max_{i \in \mathrm{I}} \left\{ \int \partial u(\boldsymbol{x}^{\mathrm{T}} t)(x_i - \boldsymbol{x}^{\mathrm{T}} t) dG(\boldsymbol{x}) \right\} \right\}. \tag{5}$$

The null of mean-variance efficiency for portfolio $t \in \Lambda$ can be stated as $H_2 : z(t, G) = 0$. By analogy to (4), we may use the sample statistic

$$z(t, F_{\mathrm{X}}) = \min_{q, (r_0, r_1) \in \mathrm{P}} \left\{ q : \sum_{t=1}^{T} (r_0 + 2 r_1 \boldsymbol{x}_t^{\mathrm{T}} t)(\boldsymbol{x}_t^{\mathrm{T}} t - x_{it}) / T + q \geq 0 \quad \forall i \in \mathrm{I} \right\}, \tag{6}$$

with $\mathrm{P} \equiv \left\{ (r_0, r_1) : r_0 + 2 r_1 \boldsymbol{x}_T^{\mathrm{T}} t = 1; r_1 \leq 0 \right\}$, as a consistent estimator for $z(t, G)$.[9,10] Like the SSD sample statistic (4), we can compute this statistic using straightforward linear programming.

II. Statistical inference

Since $F_{\mathrm{C}}(\boldsymbol{x})$ is a consistent estimator for $G(\boldsymbol{x})$, $x(t, F_{\mathrm{C}})$ is a consistent estimator for $x(t, G)$. However, $x(t, F_{\mathrm{C}})$ generally is very sensitive to sampling variation and the test results are likely to be affected by sampling error in a nontrivial way. The applied researcher must therefore have knowledge of the sampling distribution in order to make inferences about the true efficiency classification. Post (2003) derived the asymptotic sampling distribution of $x(t, F_{\mathrm{C}})$ under null that all assets have the same mean, i.e., $H_1 : E[\boldsymbol{x}] = m\boldsymbol{e}$, $m \in \Re$. This null gives a sufficient condition for the true null of efficiency, i.e., $H_0 : x(t, G) = 0$.[11]

THEOREM 1 (Post (2003)) *The p-value* $\Pr[x(t,F_C) > y | H_1]$, $y \geq 0$, *asymptotically equals the integral* $\Gamma(y,S) \equiv (1 - \int_{z \leq ye} d\Phi(z|S))$, *with* $\Phi(z|S)$ *for a N-dimensional multivariate normal distribution function with mean* **0** *and (singular) variance-covariance matrix* $S \equiv (I - et^{T})W(I - te^{T})/T$.

We may use this theorem by comparing the *p*-value for the observed value of $x(t,F_C)$ with a predefined level of significance $a \in [0,1]$; we may reject efficiency if $\Gamma(x(t,F_C),S) \leq a$. Alternatively, we may reject efficiency if the observed value of $x(t,F_C)$ is greater than or equal to the critical value $\Gamma^{-1}(a,S) \equiv \inf_{y \geq 0}\{y : \Gamma(y,S) \leq a\}$.

Post demonstrates that a test procedure that uses this asymptotic sampling distribution involves reasonable power in real-life applications only for large data sets. Presumably, this reflects the fact that the theorem uses the least favorable distribution, which minimizes the statistical size at the cost of power. Further, H_1 does not give a necessary condition for H_0, and hence the *p*-values and critical values under H_1 are likely to underestimate the true values under H_0. Consequently, a test procedure that uses the asymptotic sampling distribution under H_0 will involve even less power. Bootstrapping (see, e.g., Efron (1979) and Efron and Gong (1983)) is an alternative approach to sampling error. Since the bootstrap does not focus on the least favorable distribution, it potentially offers more power in small samples. In addition, the bootstrap can evaluate the true null of SSD efficiency rather than the null of equal means.

In order to analyze the size and power properties of a test procedure that uses the bootstrap, we extend the simulation experiment of Post (2003, Section IIIC). The simulations involve 26 benchmark assets with a multivariate normal return distribution. The joint population moments are equal to the sample moments of the monthly excess returns of the one-month U.S. Treasury bill and the 25 Fama and French U.S. common stock portfolios formed on market capitalization (size) and book-to-market equity ratio (BE/ME), for the sample period from July 1963 to October 2001. To provide some feeling for the data, Figure 1 shows a mean-variance diagram including the individual assets (the clear dots), as well as the mean-variance frontier for the case without the riskless asset (AB) and the case with the riskless asset (OP1B). The figure also includes the tangency portfolio (P1) and the equal weighted average of all 25 risky assets (P2). The tangency portfolio is efficient and we may analyze the size of a test procedure by the relative frequency of cases in which this portfolio is wrongly classified as inefficient. By contrast, the equal weighted portfolio is inefficient; it is possible to achieve a substantially higher mean given the standard deviation, and to achieve a substantially lower standard deviation given the mean. Hence, we may analyze the power of a test procedure by its ability to correctly classify the equal weighted portfolio as inefficient.

[Insert Figure 1 about here]

We assess the size and power of the following two alternative test procedures:

A. A test procedure that uses the asymptotic sampling distribution under the null of equal means (H_1). Computing p-values $\Gamma(x(t, F_C), S)$ requires the unknown variance-covariance matrix W. We estimate its elements w_{ij}, $i, j \in I$, in a distribution-free and consistent manner using the sample equivalents $\hat{w}_{ij} \equiv \sum_{t=1}^{T}(x_{it} - \sum_{t=1}^{T} x_{it}/T)(x_{jt} - \sum_{t=1}^{T} x_{jt}/T)/T$. We reject efficiency if and only if $\Gamma(x(t, F_C), \hat{S}) \le a$, with $\hat{S} \equiv (I - \boldsymbol{e}\boldsymbol{t}^{T})\hat{W}(I - \boldsymbol{t}\boldsymbol{e}^{T})/T$.[12] This procedure was used also in Post (2003), and the simulation results are taken directly from that study.

B. A test procedure that uses the bootstrap. Key to the success of bootstrapping 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. Under the assumption that the return distribution is serially IID (see Section I), the EDF $F_C(\boldsymbol{x})$ is a consistent estimator for the CDF $G(\boldsymbol{x})$. This suggests that bootstrap pseudo-samples would be simply obtained by randomly sampling with replacement from the EDF. We generate 1,000 pseudo-samples $\hat{C}$ in this way and compute the test statistic $x(t, F_{\hat{C}})$ for each pseudo-sample. To correct for possible bias, we then compute the bootstrap bias-corrected estimators $x^*(t, F_{\hat{C}}) \equiv x(t, F_{\hat{C}}) - 2\overline{x}(t, F_{\hat{C}}) + 2x(t, F_C)$, with $\overline{x}(t, F_{\hat{C}})$ for the average value of $x(t, F_{\hat{C}})$ over the pseudo-samples.[13] Finally, we compute the bootstrap p-value as the relative frequency of pseudo-samples in which the evaluated portfolio is classified as efficient, i.e., $x^*(t, F_{\hat{C}}) = 0$. We reject efficiency if and only if the p-value is smaller than the desired level of significance (a).

To assess the size and power of these procedures, we draw 1,000 random samples from the multivariate normal population distribution through Monte-Carlo simulation. For each random sample, we apply each of above two test procedures to the efficient tangency portfolio (P1) and the inefficient equal weighted portfolio (P2). 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 a sample size (T) of 25 to 4,000 observations and for a significance level (a) of 2.5, 5, and 10 percent.

Figure 2 and Figure 3 show the size and power of the three test procedures. Again, $x(t, F_C)$ converges to $x(t, G)$, and we expect minimal Type I and Type II error in large samples. Indeed, for both procedures the size goes to zero and the power goes to unity as we increase the sample size. However, in small samples, the two procedures yield very different results. For the asymptotic test procedure (Procedure A), the size is much lower than the nominal significance level. Again, this presumably reflects the conservative nature of tests that are based on the least favorable distribution. By contrast, the size of the bootstrap procedure (Procedure B) in small samples is more comparable with the nominal level of significance (a). For the asymptotic sampling distribution, minimizing Type I errors comes at the cost of Type II errors, and we need large samples to obtain reasonable power. For example, using a ten percent significance level, Procedure A achieves a power of about 60 percent for samples of about 500 observations. By contrast, the bootstrap involves substantially more power. For example, using the ten percent significance level, Procedure B yields a 60 percent rejection rate already for samples of 25 observations. Of course, this benefit has to be balanced against the additional computational burden associated with bootstrapping. However, this is not a major issue given the powerful computer hardware and software currently available.

[Insert Figure 2 and 3 about here]

III. Efficiency of the market portfolio

We analyze if the Fama and French market portfolio is SSD efficient. This market portfolio is the value-weighted average of all non-financial common stocks listed on NYSE, AMEX, and Nasdaq, and covered by CRSP and COMPUSTAT. We use two sets of benchmark portfolios. The first set includes the 25 Fama and French benchmark portfolios formed on size and BE/ME and the one-month U.S. Treasury bill (used also in the simulations in Section II). The second set consists of the 30 Fama and French benchmark portfolios formed on industry classification and the one-month U.S. Treasury bill. As discussed in the introduction, this benchmark set is less susceptible to data-snooping issues than the size and BE/ME portfolios. We use monthly excess returns for the period from July 1963 to December 2002.[14] Tables I and II give descriptive statistics for the full sample period for the market portfolio and both sets of benchmark portfolios.

[Insert Table I and II about here]

As discussed in the introduction, there exists strong evidence that the return distribution varies through time. For example, there is substantial cyclical variation in popular proxies for the market risk premium, like the credit spread (used in, e.g., Jagannathan and Wang (1996)). We are skeptical about the possibility of developing a conditional test for SSD efficiency that links the ex ante return distribution to the investors' (time-varying) information set. There are already large problems for conditional tests for mean-variance efficiency. First, the appropriate set of conditioning variables and the appropriate functional form of the relationship between these variables and the ex ante return distribution generally do not follow from economic theory. This introduces a serious risk of misleading data mining results. Second, there is no guarantee that the relationship between the ex ante distribution and the conditioning variables is sufficiently stable through time to lead to an improvement over unconditional models. These problems are even greater for a test for SSD efficiency, which involves all return moments rather than mean and variance only (the 'curse of dimensionality'). We therefore take another approach to time-variation. Specifically, we use a rolling window analysis that applies the SSD test to a series of short subsamples. Obviously, selecting the appropriate window size involves a difficult trade-off between the number of observations and the comparability of the observations in the subsamples. In this study, we use a window size of 60 months, 90 months and 120 months. More precisely, we consider all 415 subsamples of 60 months, beginning with July 1963 to June 1968 and ending with January 1998 to December 2002, all 385 subsamples of 90 months, beginning with July 1963 to December 1970 and ending with July 1995 to December 2002, and all 355 periods of 120 months, beginning with July 1963 to June 1973 and ending with January 1993 to December 2002. We think that the return distribution is much more stable within these subsamples than in the full sample. To illustrate this point, Figure 4 shows the full sample variance and the within-subsample variance of the US credit spread. Interestingly, the within-subsample variance is substantially smaller than the full sample variance. For example, for 60-month subsamples, the median within-sample variance is only 22 percent of the full sample variance. This suggests that the most important variations in the return distribution occur between the subsamples rather than within the subsamples, and that observations from the same subsample generally obey approximately the same distribution. Still, there is substantial variation in the subsample variance across different periods. The subsample variance is especially low during the late 1960s, the late 1980s and the 1990s. By contrast, the subsample variance is substantially higher during the 1970s and the early 1980s.

[Insert Figure 4 about here]

Of course, the benefit of comparable observations comes at the cost of additional sampling error for small subsamples. Given the low power of the asymptotic test procedure in small samples, we therefore use the bootstrap procedure discussed in Section II. The simulation experiment in that section suggests that this procedure involves acceptable statistical size and power properties even for samples of 60 to 120 observations. For each subsample, we test if the market portfolio is significantly SSD efficient. For the sake of comparison, we also test for mean-variance efficiency, using test statistic (6). To compute *p*-values for this test statistic, we use the same bootstrap procedure as for the SSD sample statistic.

For the size and BE/ME portfolios, Figure 5 shows the *p*-values for the mean-variance efficiency test. Overall, the evidence against mean-variance efficiency is strong. In 73 percent of the 60-month subsamples, 69 percent of the 90-month subsamples, and 85 percent of the 120-month subsamples, the market is classified as inefficient with at least 90 percent confidence. However, the evidence against efficiency differs strongly across different time periods. The evidence is overwhelming during the 1970s and the 1980s. By contrast, we cannot reject efficiency for the late 1960s. Still, the evidence against efficiency during this period increases for 90-month subsamples and it is strong for 120-month subsamples. The mean-variance test also does not reject efficiency during 1990s. During this period, the evidence against efficiency remains weak also for subsamples of 90 months and 120 months. This result confirms Schwert's (2002) finding that the size effect and BE/ME effect seem to have disappeared after the papers that highlighted them were published. Schwert raises the possibility that the anomalies never existed but rather reflect sample selection bias, as well as the possibility that the market became efficient due to the activities of practitioners who implemented the strategies implied by the academic papers. Our analysis raises a third possibility. Specifically, the market may be conditionally efficient for the full sample, but it may appear unconditionally efficient only during the subperiods with low variability for the return distribution. In this respect, the *p*-values in Figure 5 closely match the variability of the credit spread in Figure 4; we generally cannot reject efficiency if the subsample variance of the credit spread is low, and we can reject efficiency if the subsample variance is high.

Figure 6 shows the *p*-values for the SSD efficiency test. As expected, these *p*-values generally exceed the *p*-values for the mean-variance test and the evidence against efficiency is weaker; after all, mean-variance efficiency implies SSD efficiency.[15] Still, the SSD results are remarkably similar to the mean-variance results.

Again, the evidence against efficiency is strongest during the period from the early 1970s to the late 1980s. Also, for the late 1960s, the evidence against efficiency is weak for subsamples of 60 months, but the evidence increases for 90-month subsamples and it becomes very strong for 120-month subsamples. Further, we again cannot reject efficiency for the late 1990s, even for subsamples of 120 months. Brief, during the 1970s and 1980s, the market portfolio generally seems SSD inefficient, and no rational, nonsatiable and risk averse investor would hold this portfolio. Since the SSD criterion effectively considers the entire return distribution (rather than mean and variance only), this suggests that the size and BE/ME effects cannot be explained by omitted higher order central moments and partial lower partial moments.

[Insert Figure 5 and 6 about here]

Figure 7 and Figure 8 display the results for the industry portfolios. The results differ somewhat from the results for the size and BE/ME portfolios. Most notably, for 60-month subsamples, the evidence against efficiency during the late 1960s is substantially stronger, while the evidence during the late 1970s and early 1980s is substantially weaker. Also, for 120-month subsamples, the evidence against efficiency during the late 1960s and early 1970s is substantially weaker. Still, the results are similar to the results for the size and BE/ME portfolios in two important respects. First, for the 90-month and 120-month subsamples, the market portfolio seems significantly inefficient for the period from the mid-1970s to the late 1980s. Also, the market portfolio seems efficient during the 1990s. Second, while the efficiency classifications sometimes differ across the two sets of benchmark portfolios, the outcomes are remarkably robust across the two efficiency criteria; the SSD results again very closely match the mean-variance results. Again, this suggests that mean-variance inefficiency of the market portfolio cannot be explained by omitted return moments.

[Insert Figure 7 and 8 about here]

Our results do not solve the puzzle of mean-variance inefficiency; the results merely suggest that one possible explanation, the omission of return moments, is unlikely to solve this puzzle. Several alternative explanations remain to be explored (see, e.g., the concluding remarks 3, 4 and 5 below).

VI. Conclusions

1. This study provides the first rigorous analysis of the efficiency of the value-weighted market portfolio using the SSD efficiency criterion rather than the traditional mean-variance criterion. During the period from the mid-1970s to the late 1980s, the market portfolio is significantly mean-variance inefficient relative to benchmark portfolios formed on size, BE/ME and industry. During this period, the market portfolio generally also is significantly SSD inefficient. Since SSD effectively considers the entire return distribution, this suggests that the mean-variance inefficiency during these periods cannot be explained by the omission of higher order central moments and partial lower partial moments.

2. Our findings contrast with existing empirical results for models that account for omitted return moments. For example, several studies report superior

performance for the three-moment CAPM relative to the mean-variance CAPM. These contrasting results may reflect the failure of these studies to appropriately impose the conditions of nonsatiation and risk aversion. There are compelling economic arguments in favor of these conditions, and in addition risk aversion is needed to test efficiency by means of the first-order condition (an exact linear relationship between assets' co-moments). Still, the studies typically fail to impose these conditions, and hence they may overstate the explanatory power of omitted moments. For example, Post, Levy and Van Vliet (2003) show that the results of the three-moment CAPM studies typically imply an inverse S-shaped utility function with risk seeking for gains, and that the market portfolio is far from efficient, even though there is a strong linear relationship between mean, co-variance and co-skewness. Another complication arises from the criterion used to determine if omitted moments are relevant. Sometimes, a significant reduction in the violations of the first-order condition (or 'pricing errors') is presented as evidence that return moments are 'priced'. However, the first-order condition is an 'all-or-nothing' concept and the relevant criterion is if the market portfolio is efficient, or, equivalently, if the violations are significantly different from zero. As demonstrated by Roll and Ross (1994) in the context of the mean-variance CAPM, small deviations from efficiency can involve large pricing errors, and, similarly, large deviations from efficiency can involve small pricing errors. Hence, we cannot conclude that moments are priced from a significant reduction of the pricing errors, unless the pricing errors are reduced to values that are not significantly different from zero.

3. The inefficiency classifications closely match the variability of the return distribution, as measured by the variance of the credit spread. Specifically, we generally cannot reject efficiency if the variability is low, and we generally can reject efficiency if the variability high. This may be interpreted as evidence in favor of conditional efficiency of the market portfolio and the use of the credit spread as an ex ante proxy for the return distribution, as in Jagannathan and Wang (1996). Still, further research is needed, e.g., using additional conditioning variables, before we can draw firm conclusions. We remain skeptical about developing a full-fledged conditional test for SSD efficiency (see Section III). Rather, we have better hopes for applying the current unconditional test to subsamples of observations that can be assumed to obey approximately the same return distribution, e.g., because they belong to the same stage of the business cycle.

4. Contrary to the predictions of representative investor models, many investors (both individual and institutional) actually hold highly undiversified portfolios (see, e.g., Levy (1978)). Perhaps we have to move from models with a representative investor to models with heterogeneous investors who hold different, possibly highly undiversified portfolios, in order to solve the puzzle of why the market portfolio seems inefficient. Dybvig and Ross (1982) demonstrate that the SSD efficient set generally is not convex, and hence, there is no guarantee that the market portfolio is SSD efficient if different investors hold different portfolios of risky assets. Consequently, a test for SSD efficiency of the market portfolio generally is not relevant in the context of a model with heterogeneous investors. Rather, we need a test for SSD *spanning* that tests if all assets are included in some SSD efficient portfolio (not necessarily with a weight

that equals the relative market capitalization). Compared with the existing tests for mean-variance spanning (see, e.g., Huberman and Kandel (1987)), such a test would account for the full return distribution rather than its first two central moments only. Still, there is a *revealed preference* argument to expect that the market portfolio is efficient even if we allow for heterogeneous investors. Specifically, passive mutual funds and exchange-traded funds that track broad value-weighted equity indexes became an important investment vehicle during the 1990s. In other words, many actual investors reveal a preference for market indexes, and it is difficult to rationalize their portfolio choice if we assume that these indexes are inefficient. Of course, we could directly analyze the efficiency of actual funds rather than the Fama and French market portfolio. Still, many actual funds, including total market index funds based on the very broad Wilshire 5000 index (e.g., the Vanguard Total Stock Market Index Fund) are likely to be very highly correlated with the market portfolio. Hence, the popularity of index funds and exchange-traded funds suggests that the market portfolio is efficient during the 1990s. Interestingly, during this period, the evidence against mean-variance efficiency and SSD efficiency is indeed very weak.

5. Our SSD test builds on a single-period model where investors wish to hedge against reductions in the end-of-period value of their investment portfolio. Perhaps we have to move to intertemporal models where long-term investors also wish to hedge against deteriorations of the investment opportunity set for future periods (see, e.g., Merton (1973) and Campbell (1993)) in order to understand why the market portfolio seems inefficient in single-period models. The existing intertemporal models predominantly rely on mean-variance analysis, and hence they inherit the possible omission of return moments and the difficulties in imposing the conditions of nonsatiation and risk aversion (see point 6 below). In this respect, it seems interesting to develop an SD-type test that allows for testing intertemporal efficiency in a nonparametric fashion. Of course, to avoid the 'curse of dimensionality', developing such a test will require at least some prior structure on investor preferences and the return distribution.

6. Our analysis does not show marked differences between the mean-variance efficiency test and the SSD efficiency test. In our opinion, it would be a mistake to interpret this as an argument for using mean-variance tests rather than SSD tests. First, omitted risk moments may be important for studies that evaluate other portfolios and/or use other benchmark assets and/or sample periods. In such cases, the SSD criterion circumvents the possible specification error associated with the mean-variance criterion. Second, the SSD test imposes the economically meaningful conditions of non-satiation and risk aversion, while the existing mean-variance tests typically fail to impose these assumptions. For example, the Gibbons, Ross and Shanken (GRS; 1989) test checks if there exists an exact linear relationship between mean and market co-variance, or, alternatively, if the market portfolio maximizes the absolute value for the Sharpe ratio. This test gives only a necessary condition for mean-variance efficiency and it may fail to detect inefficiency, because the market risk premium is allowed to be so high that the implied utility function becomes decreasing for a range (a violation of nonsatiation) and the market risk premium is allowed to be negative (a violation of risk aversion). This problem is even greater for conditional tests of mean-

variance efficiency, as we then have to guarantee that the market risk premium is 'well-behaved' for all possible realizations of the conditioning variables.

References

Banz, Rolf W., 1981, The relation between return and market value of common stocks, *Journal of Financial Economics* 9, 3-18.

Bawa, Vijay S., 1975, Optimal rules for ordering uncertain prospects, *Journal of Financial Economics* 2, 95-121.

Bawa, Vijay S. and Eric B. Lindenberg, 1977, Capital Market Equilibrium in a Mean-Lower Partial Moment Framework, *Journal of Financial Economics* 4, 189-200.

Berk, Jonathan, 1997, Necessary conditions for the CAPM, *Journal of Economic Theory* 73, 245,257.

Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373-99.

Campbell, John Y., 1993, Intertemporal asset pricing without consumption data, *American Economic Review* 83, 487—512.

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

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427-465.

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.

Hanoch, Giora and Haim Levy, 1970, Efficient Portfolio Selection with Quadratic and Cubic Utility, *Journal of Business* 43 (2), 181-189.

Harlow, W.V. and Rao, Ramesh K. S., 1989, Asset Pricing in a Generalized Mean-Lower Partial Moment Framework: Theory and Evidence, *The Journal of Financial and Quantitative Analysis* 24, 285-311.

Harvey, Campbell R., and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *Journal of Finance* 55, 1263-1295.

Huberman, Gur and Shmuel Kandel, 1987, Mean-Variance Spanning, *Journal of Finance* 42 (4), 873-888.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The Conditional CAPM and the Cross-Section of Expected Returns, *Journal of Finance* 51, 3-53.

Kraus, Alan and Robert Litzenberger, 1976, Skewness preference and the valuation of risk assets, *Journal of Finance* 31, 1085-1100.

Levy, Haim, 1969, Comment: A utility function depending on the first three moments, *Journal of Finance* 24, 715-719.

Levy, Haim, 1978, Equilibrium in an imperfect market: a constraint on the number of securities in a portfolio, *American Economic Review* 68, 643-658.

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

Lintner, John, 1965, Security prices, risk and maximal gains from diversification, *Journal of Finance* 20, 587-615.

Lo, Andrew W. and A. Craig MacKinlay, 1990, Data-Snooping Biases in Test of Financial Asset Pricing Models, *Review of Financial Studies* 3, 431-468.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867—87.

Post, Thierry, 2003, Empirical Tests for Stochastic Dominance Efficiency, forthcoming in *Journal of Finance*.

Post, Thierry, Haim Levy and Pim van Vliet, 2003, Risk Aversion and Skewness Preference, Erasmus Research Institute of Management (ERIM) Working Paper, Report ERS-2003-01-F&A.

Roll, Richard, Shephan A. Ross, On the Cross-Sectional Relation Between Expected Returns and Betas, *Journal of Finance* 49, 101-121.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive evidence of market inefficiency, *Journal of Portfolio Management* 11, 9-17.

Rubinstein, Mark, 1974, An aggregation theorem for securities markets, *Journal of Financial Economics* 1, 225-244.

Schwert, G. William, 2002, Anomalies and Market Efficiency, Chapter 17, *Handbook of the Economics of Finance*, edited by George Constantinides, Milton Harris, and René M. Stulz, North-Holland Publishing Company.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425-442.

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

Wang, Zhenyu, 1998, Efficiency loss and constraints on portfolio holdings, *Journal of Financial Economics* 48, 359-375.

Table I

Descriptive Statistics Size and BE/ME Portfolios

Monthly excess returns (month-end to month-end) from July 1963 to December 2002 (474 months) for the value-weighted Fama and French market portfolio and 25 value-weighted benchmark portfolios based on market capitalization (size) and/or book-to-market equity ratio (BE/ME). Excess returns are computed from the raw return observations by subtracting the return on the one-month US Treasury bill. All data are obtained from the data library on the homepage of Kenneth French.

		Mean	St. Dev.	Skewness	Kurtosis	Minimum	Maximum
Market Portfolio		0.410	4.509	-0.485	1.964	-23.09	16.05
Benchmark Portfolios							
BE/ME	*Size*						
Growth	Small	0.135	8.292	0.007	2.284	-34.35	38.83
2	Small	0.703	7.103	0.023	3.168	-31.31	36.73
3	Small	0.784	6.127	-0.093	3.141	-29.01	27.72
4	Small	0.995	5.695	-0.141	3.548	-29.19	27.58
Value	Small	1.070	5.967	-0.113	3.950	-29.51	31.92
Growth	2	0.305	7.575	-0.287	1.479	-33.32	29.94
2	2	0.567	6.124	-0.487	2.631	-32.37	26.07
3	2	0.830	5.421	-0.521	3.630	-28.27	26.61
4	2	0.892	5.183	-0.411	3.751	-27.05	26.57
Value	2	0.926	5.771	-0.294	4.042	-29.78	29.1
Growth	3	0.334	6.916	-0.310	1.250	-30.03	23.24
2	3	0.646	5.516	-0.618	2.954	-29.49	23.7
3	3	0.659	4.989	-0.619	2.877	-25.54	21.64
4	3	0.792	4.739	-0.357	2.912	-23.04	22.75
Value	3	0.937	5.385	-0.449	3.999	-27.76	27.06
Growth	4	0.459	6.170	-0.170	1.516	-26.02	25.01
2	4	0.435	5.211	-0.558	3.105	-29.45	20.17
3	4	0.635	4.904	-0.428	3.129	-26.13	23.02
4	4	0.758	4.693	-0.061	2.211	-18.19	24.29
Value	4	0.850	5.422	-0.242	2.663	-24.59	26.2
Growth	Big	0.386	4.890	-0.171	1.516	-22.23	21.63
2	Big	0.415	4.607	-0.346	1.796	-23.14	16.05
3	Big	0.455	4.386	-0.270	2.415	-22.46	18.11
4	Big	0.559	4.315	0.031	1.480	-15.51	18.94
Value	Big	0.523	4.803	-0.179	0.851	-19.23	15.39

Table II

Descriptive Statistics Industry Portfolios

Monthly excess returns (month-end to month-end) from July 1963 to December 2002 (474 months) for the value-weighted Fama and French market portfolio and 30 value-weighted benchmark portfolios based on industry classification. Excess returns are computed from the raw return observations by subtracting the return on the one-month US Treasury bill. All data are obtained from the data library on the homepage of Kenneth French.

	Mean	St. Dev.	Skewness	Kurtosis	Minimum	Maximum
Market Portfolio	0.410	4.509	-0.485	1.964	-23.09	16.05
Benchmark Portfolios						
Industry						
Food	0.656	4.662	-0.078	2.178	-18.06	19.77
Beer	0.628	5.273	0.096	2.053	-18.74	25.18
Smoke	0.941	6.162	-0.106	2.092	-21.66	28.19
Games	0.600	7.216	-0.288	1.795	-33.02	30.36
Books	0.634	5.580	-0.261	1.508	-26.4	22.68
Hshld	0.522	4.936	-0.246	1.461	-22.25	16.8
Clths	0.401	6.617	-0.069	2.797	-32.14	31.84
Hlth	0.663	5.168	0.085	2.416	-20.9	29.07
Chems	0.391	5.311	-0.078	2.477	-28.6	20.76
Txtls	0.409	6.208	-0.534	2.731	-33.22	22.38
Cnstr	0.451	5.710	-0.222	2.392	-28.83	24.48
Steel	0.203	6.380	-0.105	2.177	-31.76	25.91
FabPr	0.380	5.967	-0.376	2.197	-31.96	21.77
ElcEq	0.543	6.697	0.371	3.559	-32.69	37.57
Autos	0.336	6.099	-0.173	1.689	-28.7	22.15
Carry	0.642	6.539	-0.259	1.449	-31.05	23.44
Mines	0.377	6.944	-0.204	1.306	-32.92	20.53
Coal	0.486	7.867	0.489	3.203	-30.74	45.92
Oil	0.517	5.237	0.090	1.660	-19.31	23.41
Util	0.277	4.120	0.146	1.114	-12.95	18.36
Telcm	0.357	4.975	-0.118	1.994	-19.19	21.63
Servs	0.607	7.226	-0.112	1.028	-28.61	25.64
BusEq	0.456	6.879	-0.229	1.404	-28.55	20.48
Paper	0.438	5.181	-0.078	2.628	-27.67	21.53
Trans	0.409	6.123	-0.262	1.230	-28.4	18.66
Whlsl	0.617	5.953	-0.373	2.362	-31.59	17.32
Rtail	0.595	5.715	-0.159	2.133	-29.59	26.53
Meals	0.655	6.684	-0.487	2.248	-32.1	28.1
Fin	0.549	5.116	-0.200	1.384	-20.77	20.18
Other	0.347	5.901	-0.352	1.303	-27.93	19.66

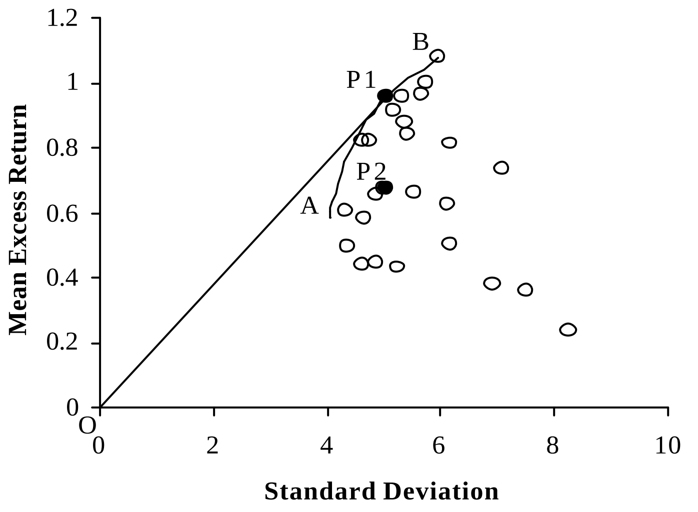

Figure 1. Mean-variance diagram 25 benchmark assets. Diagram for the mean excess returns and standard deviations of the 25 risky assets (the clear dots), as well as the efficient tangency portfolio (P1) and the inefficient equally weighted test portfolio (P2). The 25 assets obey a multivariate normal 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.

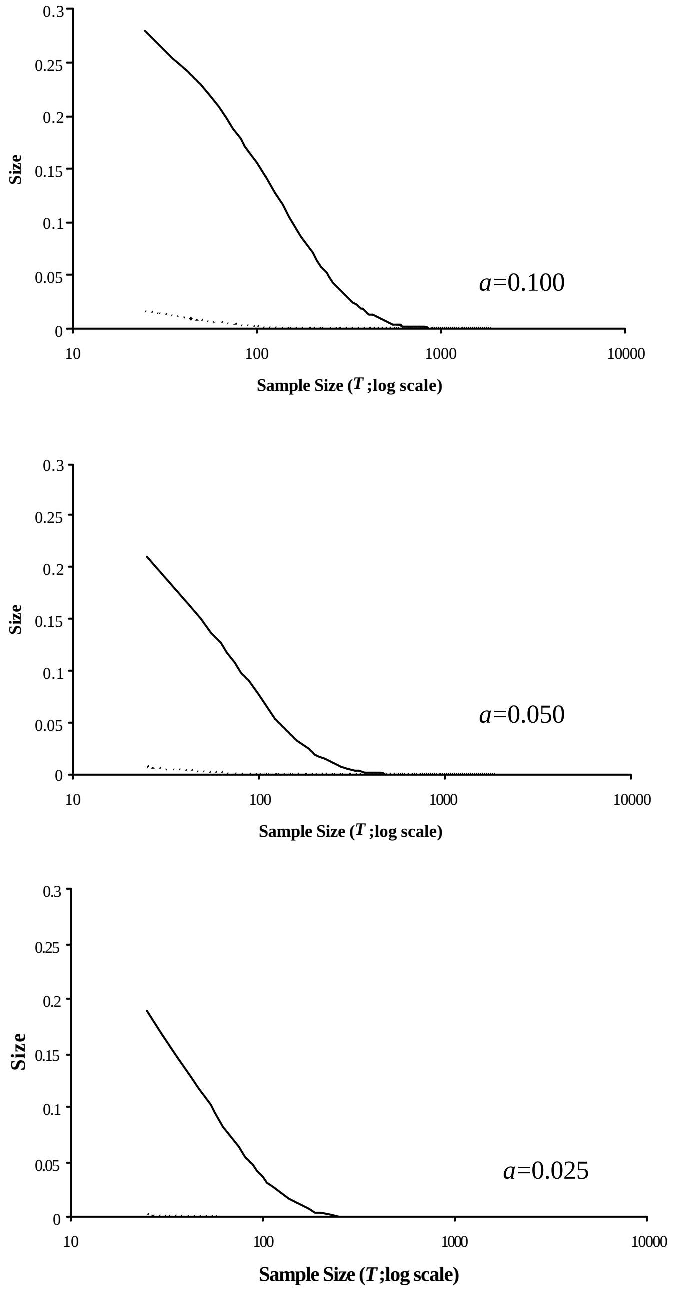

Figure 2. Size of competing test procedures. The figure shows the statistical size of the two competing test procedures. The dotted line shows the results for the procedure that uses the asymptotic sampling distribution under H_1 (Procedure A). The solid line shows the results for the bootstrap procedure (Procedure B). The figure displays the size for a sample size (T) of 25 to 4,000 and for a significance level (a) of 2.5, 5, and 10 percent. The results are based on 1,000 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 tangency portfolio (P1) is wrongly classified as inefficient.

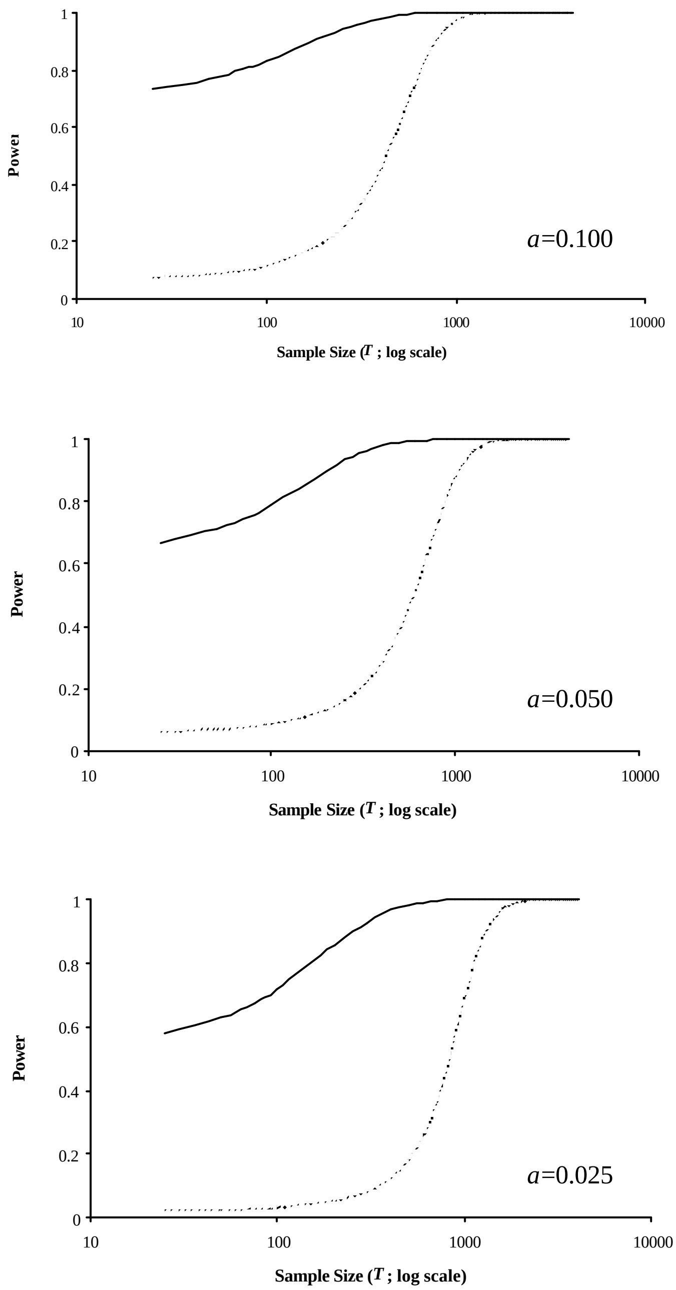

Figure 3. Power of competing test procedures. The figure shows the statistical power of the two competing test procedures. The dotted line shows the results for the procedure that uses the asymptotic sampling distribution under H_1 (Procedure A). The solid line shows the results for the bootstrap procedure (Procedure B). The figure displays the power for a sample size (T) of 25 to 4,000 and for a significance level (a) of 2.5, 5, and 10 percent. The results are based on 1,000 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 portfolio (P2) is correctly classified as inefficient.

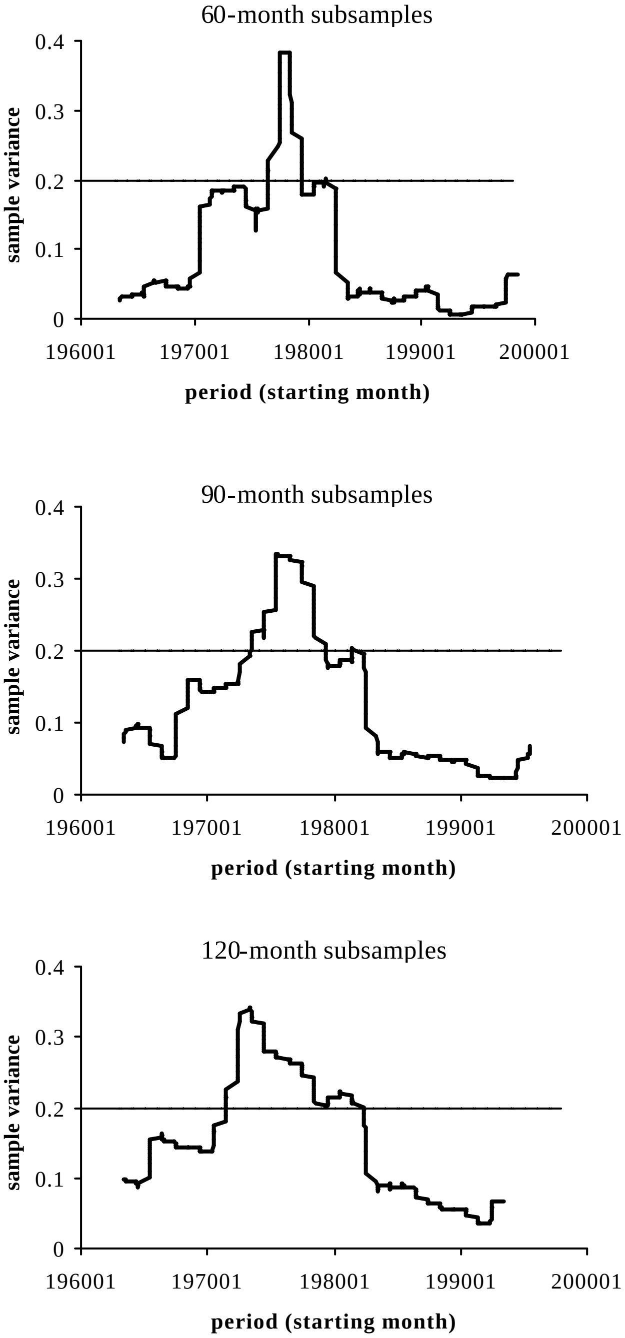

Figure 4. Sample variation in the credit spread. The figure shows the sample variance of the US default premium for subsamples (the heavy line) as well as the full sample from July 1963 to December 2002 (the thin line). Results are shown for all 415 subsamples of 60 months (beginning with Jul 1963-Jun 1968 and ending with Jan 1998-Dec 2002), 385 subsamples of 90 months (beginning with Jul 1963-Dec 1970 and ending with Jul 1995-Dec 2002), and 355 subsamples of 120 months (beginning with Jul 1963-Jun 1973 and ending with Jan 1993-Dec 2002). The credit spread is computed as the difference between the monthly Moody's Seasoned Aaa Corporate Bond Yield and the Moody's Seasoned Baa Corporate Bond Yield. Data are taken from the FRED data library of the Federal Reserve Bank of St. Louis: http://research.stlouisfed.org/fred/.

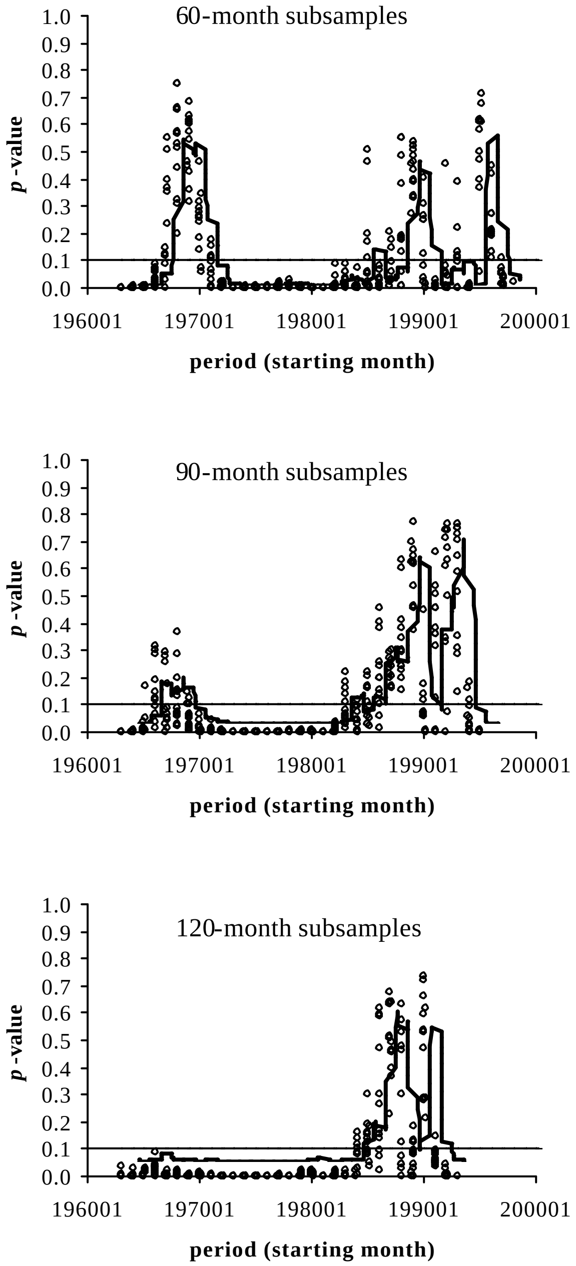

Figure 5. Bootstrap *p*-values for mean-variance efficiency relative to size and BE/ME portfolios. The figure shows the bootstrap *p*-values for the null hypothesis of mean-variance efficiency for the Fama and French market portfolio relative to the 25 Fama and French benchmark portfolios formed on size and BE/ME and the one-month T-bill. Results are shown for all 415 subsamples of 60 months (beginning with Jul 1963-Jun 1968 and ending with Jan 1998-Dec 2002), 385 subsamples of 90 months (beginning with Jul 1963-Dec 1970 and ending with Jul 1995-Dec 2002), and 355 subsamples of 120 months (beginning with Jul 1963-Jun 1973 and ending with Jan 1993-Dec 2002). The *p*-values are computed using 1,000 pseudo-samples for each subsample, and after correcting for possible bias in the test statistic. The *p*-values for the individual month are represented by the dots. The figure also displays the 12-month moving average (the heavy line) and the significance level of ten percent (the thin line).

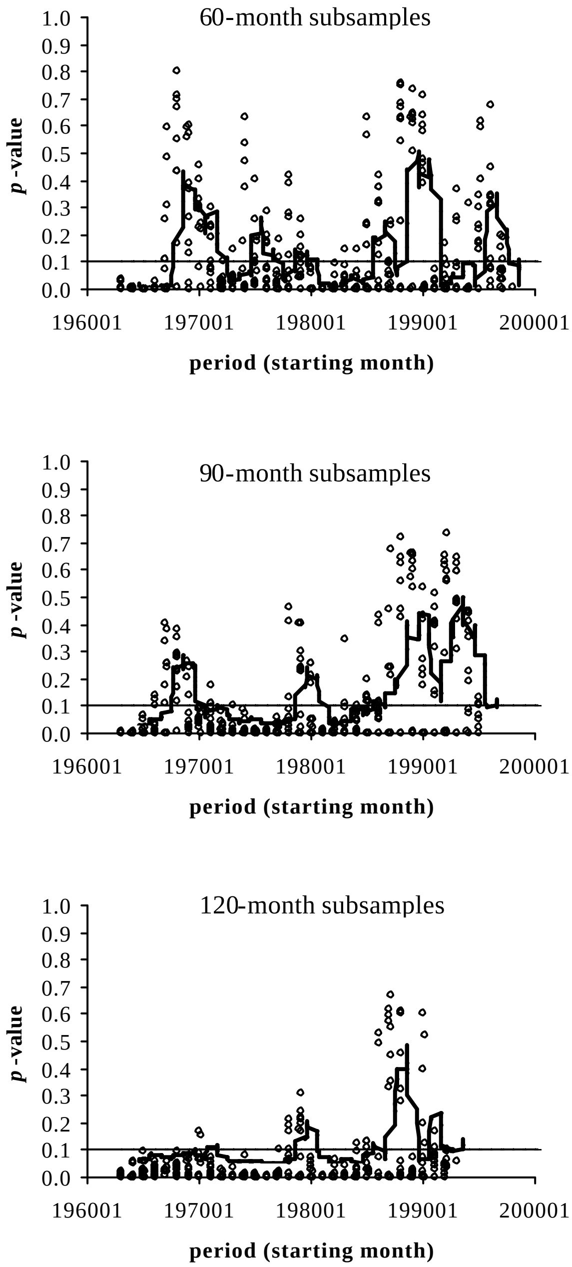

Figure 6. Bootstrap *p*-values for SSD efficiency relative to size and BE/ME portfolios. The figure shows the bootstrap *p*-values for the null hypothesis of SSD efficiency for the Fama and French market portfolio relative to the 25 Fama and French benchmark portfolios formed on size and BE/ME and the one-month T-bill. Results are shown for all 415 subsamples of 60 months (beginning with Jul 1963-Jun 1968 and ending with Jan 1998-Dec 2002), 385 subsamples of 90 months (beginning with Jul 1963-Dec 1970 and ending with Jul 1995-Dec 2002), and 355 subsamples of 120 months (beginning with Jul 1963-Jun 1973 and ending with Jan 1993-Dec 2002). The *p*-values are computed using 1,000 pseudo-samples for each subsample, and after correcting for possible bias in the test statistic. The *p*-values for the individual month are represented by the dots. The figure also displays the 12-month moving average (the heavy line) and the significance level of ten percent (the thin line).

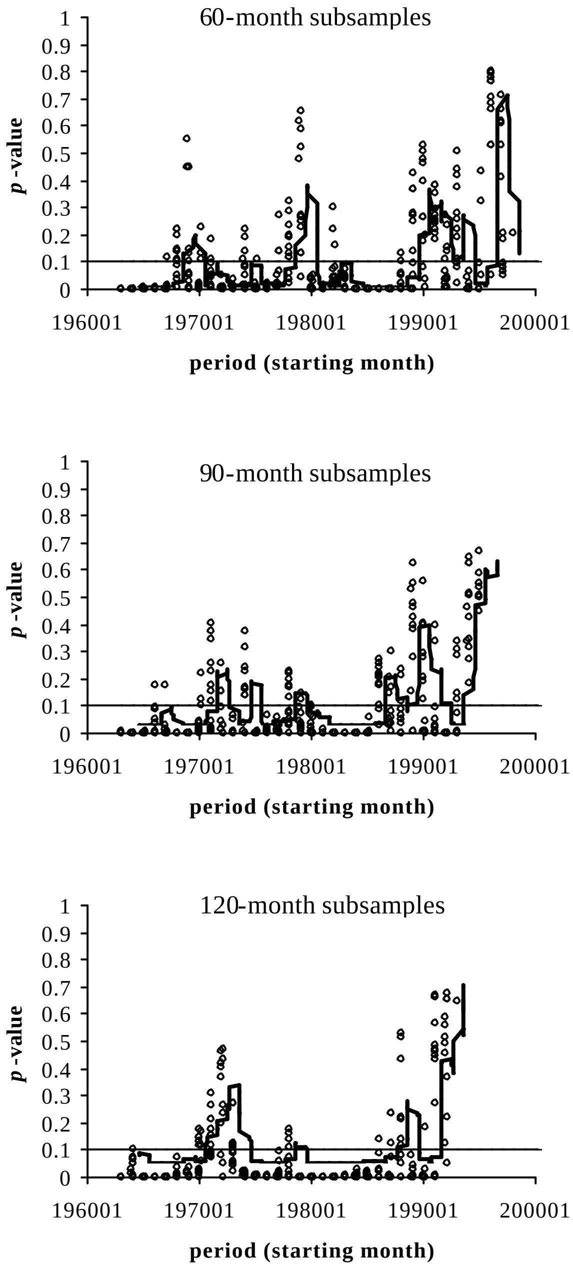

Figure 7. Bootstrap *p*-values for mean-variance efficiency relative to industry portfolios. The figure shows the bootstrap *p*-values for the null hypothesis of mean-variance efficiency for the Fama and French market portfolio relative to the 30 Fama and French benchmark portfolios formed on industry classification and the one-month T-bill. Results are shown for all 415 subsamples of 60 months (beginning with Jul 1963-Jun 1968 and ending with Jan 1998-Dec 2002), 385 subsamples of 90 months (beginning with Jul 1963-Dec 1970 and ending with Jul 1995-Dec 2002), and 355 subsamples of 120 months (beginning with Jul 1963-Jun 1973 and ending with Jan 1993-Dec 2002). The *p*-values are computed using 1,000 pseudo-samples for each subsample, and after correcting for possible bias in the test statistic. The *p*-values for the individual month are represented by the dots. The figure also displays the 12-month moving average (the heavy line) and the significance level of ten percent (the thin line).

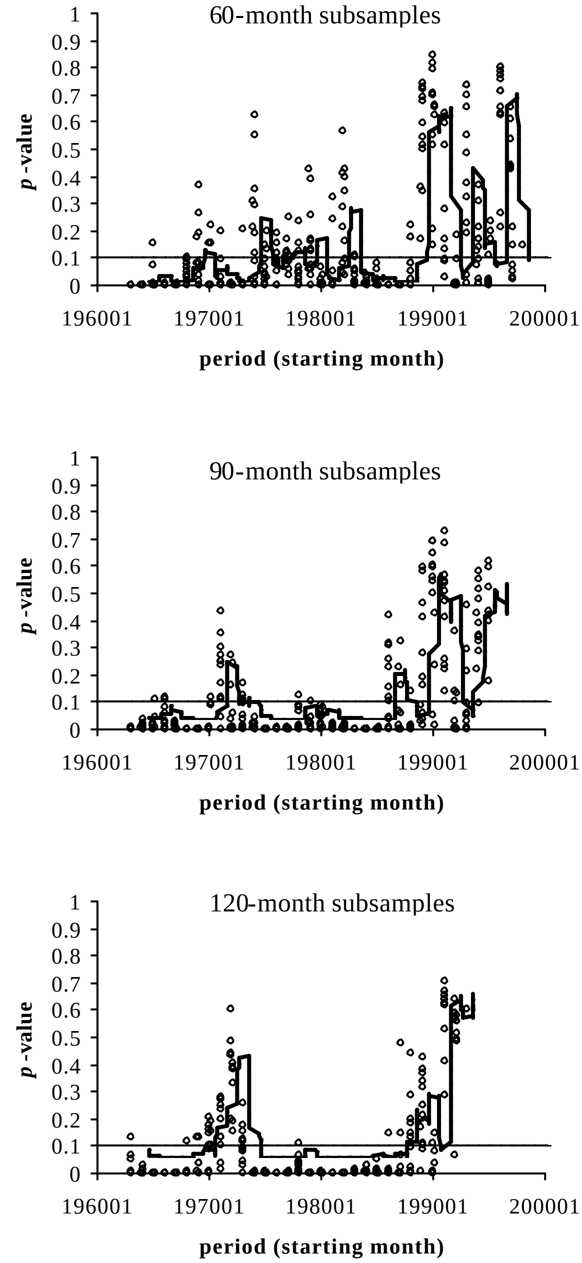

Figure 8. Bootstrap *p*-values for SSD efficiency relative to industry portfolios. The figure shows the bootstrap *p*-values for the null hypothesis of SSD efficiency for the Fama and French market portfolio relative to the 30 Fama and French benchmark portfolios formed on industry classification and the one-month T-bill. Results are shown for all 415 subsamples of 60 months (beginning with Jul 1963-Jun 1968 and ending with Jan 1998-Dec 2002), 385 subsamples of 90 months (beginning with Jul 1963-Dec 1970 and ending with Jul 1995-Dec 2002), and 355 subsamples of 120 months (beginning with Jul 1963-Jun 1973 and ending with Jan 1993-Dec 2002). The *p*-values are computed using 1,000 pseudo-samples for each subsample, and after correcting for possible bias in the test statistic. The *p*-values for the individual month are represented by the dots. The figure also displays the 12-month moving average (the heavy line) and the significance level of ten percent (the thin line).

Footnotes

[1] Less restrictive assumptions are obtained if we do restrict the shape of the return distribution; see, e.g., Berk (1997).

[2] The use of a representative investor is typically motivated by assuming that the market is complete, or, alternatively, that the preferences and beliefs of the actual investors are 'sufficiently similar' (see, e.g., Rubinstein (1974)).

[3] The higher-order central moments become relevant if we add higher-order polynomial terms to the investor's utility function. Similarly, lower partial moments become relevant if the utility function is not continuously differentiable but rather exhibits kinks, e.g., because the investor uses target rates of return. See, e.g., Bawa (1975), Bawa and Lindenberg (1977) and Harlow and Rao (1989) for a detailed discussion of lower partial moments in asset pricing.

[4] 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 $\boldsymbol{x}$.

[5] The simplex Λ excludes short sales. 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) and Wang (1998)). Still, we may generalize our analysis to include (bounded) short selling. The SSD test is based on the first-order optimality conditions for optimizing a concave objective function over a convex set. The analysis can be extended to a general polyhedral portfolio possibilities set. We basically have to check whether there exists an increasing hyperplane that supports the extreme points of the portfolio possibilities set. One approach is to enumerate all extreme points and to include all extreme points as virtual assets.

[6] We use supergradients rather than gradients, because the SSD criterion does not assume that utility is continuously differentiable and it allows for kinked utility functions. In fact, this is why the SSD creterion is consistent with lower partial moments, which become relevant if the utility function is kinked.

[7] Since utility functions are unique up to a positive linear transformation, this standardization does not affect the results.

[8] This is the Hanoch and Levy (1970) definition of mean-variance efficiency, which guarantees that the mean-variance efficient set is a proper subset of the SSD efficient set.

[9] We basically substitute $b_t = (r_0 + 2r_1 \boldsymbol{x}_t^{\mathrm{T}} t)$ in (4); mean-variance analysis assumes that marginal utility is a linear function. The restriction $(r_0, r_1) \in \mathrm{P}$ suffices to guarantee $b \in \mathrm{B}$.

[10] We use this test rather than, e.g., the conventional Gibbons, Ross and Shanken (GRS; 1989) test for mean-variance efficiency. The GRS test gives only a necessary condition for mean-variance efficiency (see remark 5 in the Conclusions). By contrast, our test gives a necessary and sufficient condition.

[11] Under the null, all portfolios $l \in \Lambda$ are efficient, because they maximize the expected value of the utility function $u(x) = x$, i.e., the risk neutral investor is indifferent between portfolios with equal means

[12] We approximate $\Gamma(x(t, F_{\mathrm{C}}), \hat{\mathrm{S}})$, using Monte-Carlo simulation. Specifically, we generate 1,000 independent standard normal random vectors $\boldsymbol{w}_s \in \Re^{N-1}$, $s \in \{1, \cdots, 10{,}000\}$, 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 $\hat{\mathrm{S}}$, using $\boldsymbol{z}_s = (\mathrm{I} - \boldsymbol{e}t^{\mathrm{T}})\hat{\mathbf{D}}\boldsymbol{w}_s$, with $\hat{\mathbf{D}} \in \Re^{N \times N}$ for a lower triangular Cholesky factor of $\hat{\mathrm{W}}$, so $\hat{\mathrm{W}} = \hat{\mathbf{D}}\hat{\mathbf{D}}^{\mathrm{T}}$. Finally, $\Gamma(x(t, F_{\mathrm{C}}), \hat{\mathrm{S}})$ is approximated by the relative frequency of the transformed vectors $\boldsymbol{z}_s$, $s \in \{1, \cdots, 10{,}000\}$, that fall outside the integration region $\{\boldsymbol{z} \in \Re^N : \boldsymbol{z} \le y\boldsymbol{e}\}$.

[13] In this way, the estimators $x^*(t, F_{\hat{\mathrm{C}}})$ are centered at $2x(t, F_{\mathrm{C}}) - \bar{x}(t, F_{\hat{\mathrm{C}}})$, which is the bias-corrected value for the original estimator $x(t, F_{\mathrm{C}})$. The raw bootstrap distribution of the SSD test statistic generally involves positive bias, and not correcting for bias lowers the bootstrap *p*-values and hence increases size and power.

[14] The data set starts in 1963 because the COMPUSTAT data used to construct the benchmark portfolios are biased towards big historically successful firms for the earlier years (see Fama and French (1992)).

[15] By construction, $z(t, F_{\hat{C}}) \geq x(t, F_{\hat{C}})$. Hence, $z^{*}(t, F_{\hat{C}}) < x^{*}(t, F_{\hat{C}})$ occurs only if the bootstrap procedure yields a higher estimate for the bias in the mean-variance test statistic than for the bias in the SSD statistic, i.e., $(\overline{z}(t, F_{\hat{C}}) - z(t, F_{C})) > (\overline{x}(t, F_{\hat{C}}) - x(t, F_{C}))$.

Publications in the Report Series Research* in Management

ERIM Research Program: "Finance and Accounting"

2003

COMMENT, Risk Aversion and Skewness Preference
Thierry Post and Pim van Vliet
ERS-2003-009-F&A
http://hdl.handle.net/1765/319

International Portfolio Choice: A Spanning Approach
Ben Tims, Ronald Mahieu
ERS-2003-011-F&A
http://hdl.handle.net/1765/276

Portfolio Return Characteristics Of Different Industries
Igor Pouchkarev, Jaap Spronk, Pim van Vliet
ERS-2003-014-F&A
http://hdl.handle.net/1765/272

Asset prices and omitted moments
A stochastic dominance analysis of market efficiency
Thierry Post
ERS-2003-017-F&A

A Multidimensional Framework for Financial-Economic Decisions
Winfried Hallerbach & Jaap Spronk
ERS-2003-021-F&A
http://hdl.handle.net/1765/321

A Range-Based Multivariate Model for Exchange Rate Volatility
Ben Tims, Ronald Mahieu
ERS-2003-022-F&A
http://hdl.handle.net/1765/282

Macro factors and the Term Structure of Interest Rates
Hans Dewachter and Marco Lyrio
ERS-2003-037-F&A
http://hdl.handle.net/1765/324

The effects of decision flexibility in the hierarchical investment decision process
Winfried Hallerbach, Haikun Ning, Jaap Spronk
ERS-2003-047-F&A

Takeover defenses and IPO firm value in the Netherlands
Peter Roosenboom, Tjalling van der Goot
ERS-2003-049-F&A

---

* A complete overview of the ERIM Report Series Research in Management:
http://www.erim.eur.nl

ERIM Research Programs:
LIS Business Processes, Logistics and Information Systems
ORG Organizing for Performance
MKT Marketing
F&A Finance and Accounting
STR Strategy and Entrepreneurship