A GMM Test For SSD Efficiency

Thierry Post and Philippe Versijp

ERIM Report Series *Research in Management*	
ERIM Report Series reference number	ERS-2004-024-F&A
Publication	March 2004
Number of pages	15
Email address corresponding author	gtpost@few.eur.nl
Address	Erasmus Research Institute of Management (ERIM) Rotterdam School of Management / Rotterdam School of Economics Erasmus Universiteit Rotterdam P.O. Box 1738 3000 DR Rotterdam, The Netherlands Phone: + 31 10 408 1182 Fax: + 31 10 408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES
RESEARCH IN MANAGEMENT

BIBLIOGRAPHIC DATA AND CLASSIFICATIONS		
Abstract	We develop an empirical test for Second-order Stochastic Dominance (SSD) efficiency of a given investment portfolio relative to all possible portfolios formed from a set of assets. Contrary to the Linear Programming test of Post, Thierry, 2003, Empirical tests for stochastic dominance efficiency, *Journal of Finance* 58, 1905—1932, our test is embedded in the Generalized Method of Moments (GMM) framework. The GMM test has superior statistical properties compared to the original LP test. Using this test, we demonstrate that the anomalous size effect can be explained with risk not captured by variance alone. However, the market portfolio remains SSD inefficient relative to value and momentum portfolios due to the overvaluation of growth stocks and past losers.	
Library of Congress Classification (LCC)	5001-6182	Business
	4001-4280.7	Finance Management, Business Finance, Corporation Finance
	HG 4529.5	Portfolio management
Journal of Economic Literature (JEL)	M	Business Administration and Business Economics
	G 3	Corporate Finance and Governance
	G 12	Asset pricing
European Business Schools Library Group (EBSLG)	85 A	Business General
	220 A	Financial Management
	220 T	Quantitative methods for financial management
Gemeenschappelijke Onderwerpsontsluiting (GOO)		
Classification GOO	85.00	Bedrijfskunde, Organisatiekunde: algemeen
	85.30	Financieel management, financiering
	85.33	Beleggingsleer
Keywords GOO	Bedrijfskunde / Bedrijfseconomie	
	Financieel management, bedrijfsfinanciering, besliskunde	
	Stochastische modellen, Generalisaties, Effectenhandel, Portfolio-analyse	
Free keywords	Stochastic Dominance, Generalized Method of Moments, Asset Pricing, Market Portfolio Efficiency, Size, B/M and Momentum Effects.	

A GMM test for SSD efficiency

Thierry Post* and Philippe Versijp

We develop an empirical test for Second-order Stochastic Dominance (SSD) efficiency of a given investment portfolio relative to all possible portfolios formed from a set of assets. Contrary to the Linear Programming test of Post, Thierry, 2003, Empirical tests for stochastic dominance efficiency, *Journal of Finance* 58, 1905—1932, our test is embedded in the Generalized Method of Moments (GMM) framework. The GMM test has superior statistical properties compared to the original LP test. Using this test, we demonstrate that the anomalous size effect can be explained with risk not captured by variance alone. However, the market portfolio remains SSD inefficient relative to value and momentum portfolios due to the overvaluation of growth stocks and past losers.

A PORTFOLIO IS SECOND-ORDER STOCHASTIC DOMINANCE (SSD) EFFICIENT if no other portfolio is preferred by all nonsatiable and risk averse investors. This efficiency criterion is theoretically appealing, because it avoids structure on investor preferences that does not follow from economic theory, while imposing the economically meaningful conditions of nonsatiation and risk aversion. In this respect, the SSD efficiency criterion has an important comparative advantage relative to the traditional mean-variance efficiency criterion. By focusing on mean and variance exclusively and by allowing for every possible trade-off between these two moments, the mean-variance criterion can fail to detect obvious inefficiencies. Arguably the best illustration of this problem comes from the "grand old man" of SD, Haim Levy (1998, p. 2):

"*[Consider] two alternative investments: x providing \$1 or \$2 with equal probability and y providing \$2 or \$4 with equal probability, with an identical investment of, say, \$1.1. A simple calculation shows that both the mean and the variance of y are greater than the corresponding parameters of x; hence the mean-variance rule remains silent regarding the choice between x and y. Yet, any rational investor would (and should) select y, because the lowest return on y is equal to the largest return on x.*"

Despite its theoretical appeal, the SSD efficiency criterion has not been applied in empirical finance on a broad scale. Paradoxically, due to the availability of large data sets such as the CRSP data set, this research field seems particularly fertile ground for a non-parametric approach. As discussed in Post (2003), this situation can be explained by the traditional SSD tests being relevant only for pairwise comparison of a finite set of choice alternative. These tests do not apply for portfolio choice problems in which infinitely many portfolios can be formed by means of diversification. To deal with this problem, Post developed a Linear Programming test for SSD efficiency that does account for diversification possibilities. He also derived a characterization of the sampling distribution of his test statistic to allow for statistical inference.

While these results provide an important step in the evolution of the SD methodology, they are "*the mere starting point for developing a full framework for SD with full diversification possibilities*" (Post (2003), p. 1926). The most important limitations of the existing test are (1) the lack of statistical power in small samples and (2) the use of a restrictive null hypothesis, which adversely affects the statistical size.

Post's simulations suggest that the SSD-LP test involves little power in typical asset pricing applications such as testing efficiency of the CRSP all-share index relative to the 25 Fama and French portfolios in samples of 120 to 480 monthly observations. The lack of power follows from two features of the test. First, the test focuses on the case with no short sales, which substantially narrows the portfolio possibilities. Second, and related to this, the test focuses on the maximum positive pricing error only. If short selling is excluded, an overvalued asset (or a negative pricing error) does not necessarily imply inefficiency; the evaluated portfolio can be improved only by selling the current holdings of the asset (if any) and not by short selling the asset. However, if short selling is allowed, every overvalued asset implies inefficiency. In this case, considering all pricing errors yields a more powerful test.

Apart from the lack of power, the test also suffers from the use of a "wrong null". Specifically, Post analyzes the sampling distribution under the null that all assets have the same mean return. Clearly, this approach may lead to erroneous rejections of the "true null" of SSD efficiency in cases where the evaluated portfolio is efficient but the assets have different means. Of course, using the true null would further reduce the power of the Post approach of focusing on the maximum positive pricing error. Indeed, Post and Van Vliet (2004) show that using the true null yields low power even for narrow cross-section of the ten Fama and French portfolios formed on book-to-market equity-ratio (B/M).

The goal of this paper is to circumvent the above limitations by developing a SSD test within the framework of Generalized Method of Moments (GMM; Hansen (1982)). Contrary to typical parametric GMM tests, our test avoids functional specification while imposing the standard regularity conditions of nonsatiation and risk aversion. Contrary to Post (2003), our SSD-GMM test has more statistical power, because it considers all pricing errors rather than the maximum positive pricing error only. Further, as in Post and van Vliet (2004), our test is based on the true null of efficiency rather than the restrictive null of equal means.

Using the novel SSD-GMM test, we will test if the CRSP all-share index, a popular proxy for the market portfolio, is SSD efficient relative to all portfolios formed from benchmark portfolios based on size, B/M and momentum. This analysis enables us to investigate if existing asset pricing puzzles that occur in the mean-variance framework can be explained by market risk not captured by variance.

The remainder of this text is structured as follows. Section I introduces preliminary notation, assumptions and definitions. Section II discusses our novel SSD-GMM test. Next, Section III turns to the computational aspects of using the test in practice. Section IV extends Post's simulations in order to gauge the statistical size and power properties of the SSD-GMM test. Subsequently, Section V uses the test to analyze if the CRSP all-share index is SSD efficient and if existing asset pricing puzzles can be explained by market risk other than variance. Finally, Section VI gives concluding remarks and suggestions for further research. The Appendix gives the formal proofs for our theorems.

I. Preliminaries

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

Assumption 1 (Investor preferences) Investors are nonsatiable and risk averse and they choose investment portfolios to maximize the expected utility associated with the return of their portfolios. Throughout the text, we will denote utility functions by $u:\Re \rightarrow P$, $u \in U_2$, with

U_2 for the set of increasing and concave, once continuously differentiable, von Neumann-Morgenstern utility functions, and P for a nonempty, closed, and convex subset of $\Re$.[1,2]

Assumption 2 (Portfolio possibilities) The investment universe consists of N risky assets and a riskless asset. Throughout the text, we will use the index set $\mathrm{I} \equiv \{1, \cdots, N\}$ to denote the different risky assets. Investors may diversify between the assets, and we will use $l \in \Re^N$ for a vector of portfolio weights. If the weights sum to unity, or $l^{\mathrm{T}}\boldsymbol{e}=1$, then all wealth in invested in risky assets, and $l^{\mathrm{T}}\boldsymbol{e}<1$ and $l^{\mathrm{T}}\boldsymbol{e}>1$ refer to lending and borrowing (or long and short positions in the riskless asset) respectively. The evaluated portfolio is denoted by $t \in \Re^N$.

Assumption 3 (Return distribution) The excess returns $\boldsymbol{x} \in \Re^N$ are serially independent and identically distributed (IID) random variables with a continuous joint cumulative distribution function (CDF) $G: \Re^N \to [0,1]$. The utility-weighted moments $m(u,G) \equiv \int u'(\boldsymbol{x}^{\mathrm{T}} t)\boldsymbol{x} dG(\boldsymbol{x})$ and $\mathrm{W}(u,G) \equiv \int u'(\boldsymbol{x}^{\mathrm{T}} t)^2 \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} dG(\boldsymbol{x})$ are finite, or $m(u,G) < \infty$ and $\mathrm{W}(u,G) < \infty$, for all $u \in U_2$.

Under these assumptions, the investors' optimization problem can be summarized as $\max_{l \in \Re^N} \int u(\boldsymbol{x}^{\mathrm{T}} l) dG(\boldsymbol{x})$. The evaluated portfolio $t \in \Re^N$ is optimal for a given utility function $u \in U_{SSD}$ if and only if the Euler equation is satisfied:

$$m(u,G) = \mathbf{0} \tag{1}$$

In this respect, $m(u,G)$ can be seen as a vector of pricing errors. If $m_i(u,G) > 0$, asset $i \in \mathrm{I}$ is undervalued and its weight in the portfolio should be increased relative to t_i. Similarly, if $m_i(u,G) < 0$, asset $i \in \mathrm{I}$ is overvalued and its weight in the portfolio should be decreased relative to t_i.

Definition 1 (SSD Efficiency) The evaluated portfolio $t \in \Re^N$ is SSD efficient if and only if $m(u,G) = \mathbf{0}$ for some $u \in U_2$. The portfolio is inefficient if and only if it is not optimal, that is, $m(u,G) \neq \mathbf{0}$, for all $u \in U_2$.

To test the null of efficiency, or $H_0 : m(u,G) = \mathbf{0}$, 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.

Assumption 4 (Data set) The observations are serially independently and identically distributed (IID) 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, that is, $\boldsymbol{x}_1^{\mathrm{T}} t < \boldsymbol{x}_2^{\mathrm{T}} t < \cdots < \boldsymbol{x}_T^{\mathrm{T}} t$.

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

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

Also, using the gradient vector $\nabla u(\mathrm{C}^{\mathrm{T}} t) \equiv (u'(\boldsymbol{x}_1^{\mathrm{T}} t) \cdots u'(\boldsymbol{x}_T^{\mathrm{T}} t))^{\mathrm{T}}$, we can construct the following sample counterparts of $m(u,G)$ and $\mathrm{W}(u,G)$:

$$m(u, F_{\mathrm{C}}) = T^{-1} \nabla u(\mathrm{C}^{\mathrm{T}} t)^{\mathrm{T}} \mathrm{C} \tag{3}$$

$$\mathrm{W}(u, F_{\mathrm{C}}) = T^{-1} \nabla u(\mathrm{C}^{\mathrm{T}} t)^{\mathrm{T}} \mathrm{C}\mathrm{C}^{\mathrm{T}} \nabla u(\mathrm{C}^{\mathrm{T}} t) \tag{4}$$

Since the observations are assumed to be serially IID, $F_{\mathrm{C}}(\boldsymbol{x})$ is a consistent estimator for $G(\boldsymbol{x})$, and $m(u, F_{\mathrm{C}})$ and $\mathrm{W}(u, F_{\mathrm{C}})$ converge to $m(u,G)$ and $\mathrm{W}(u,G)$ respectively.

II. The SSD-GMM Test

To test if the market portfolio t is SSD efficient, we will use the following test statistic:

$$x(F_{\mathrm{C}}) \equiv \min_{u \in U_2^*} T m(u, F_?) \mathrm{W}(u, F_?)^{-1} m(u, F_?) \tag{5}$$

In this expression,

$$U_2^* \equiv \left\{ u \in U_2 : \nabla u(\mathrm{C}^{\mathrm{T}} t)^{\mathrm{T}} \boldsymbol{e} = 1 \right\} \tag{6}$$

is the subset of SSD utility functions for which the sample average of marginal utility (evaluated at the market return) equals unity. Imposing the latter restriction may seem very unusual. The objective is to standardize utility such that the optimal solution is empirically distinguishable from $u(x) = 0$. Since utility functions are unique up to the level of a positive linear transformation, the standardization does not affect the efficiency classification.

The test statistic $x(F_{\mathrm{C}})$ gives a weighted average of the empirical pricing errors $m(u, F_?)$, with the weights depending on the empirical variance-covariance matrix $\mathrm{W}(u, F_?)$. The higher the variance or the covariance with the pricing errors of other assets, the lower the weight assigned to a given asset (all other things remaining constant). The test statistic basically is a nonparametric variant of the traditional J-statistic used in GMM. Rather than using a parametrically specified utility function with a few unknown parameters, the functional form of the utility function is left unspecified. Naturally, this approach will come at the cost of a loss of statistical power in small samples. Note however, that the utility function is restricted to be “economically meaningful”, that is, it must obey nonsatiation and risk aversion. This restriction will help to increase power relative to an unrestricted parametric approach. The latter approach may fail to detect inefficiency if the market portfolio is optimal relative a utility function that is decreasing or convex over a range.[3]

For the purpose of statistical inference, we need to characterize the sampling distribution of the test statistic. The nonparametric approach and the imposed utility restrictions make it difficult to characterize the exact sampling distribution[TTP1]. Nevertheless, it is possible to derive conservative asymptotic p-values and critical values. Specifically, we may derive the following result:

THEOREM 1 (Sampling distribution) Asymptotically, the null distribution of $x(F_C)$ is bounded from above by a chi-squared distribution with N degrees of freedom, that is, $P[x(F_C) > y|H_0] < 1 - c_N^2(y)$.

This theorem can be used for conservative statistical inference. Specifically, the asymptotic *p*-value associated with the observed value of the test statistic $x(F_C)$ is always smaller than or equal to $1 - c_N^2(x(F_C))$ percent and we can be at least $c_N^2(x(F_C))$ percent certain that efficiency is violated. Equivalently, the critical value is always smaller than or equal to $(c_N^2)^{-1}(1-a)$. Thus, we can reject efficiency at a confidence level of at least $(1-a)$ if $1 - c_N^2(x(F_C)) \leq a$, or if $x(F_C) \geq (c_N^2)^{-1}(1-a)$. This conservative approach is consistent with the convention of rejecting the null only if the *p*-value is smaller than a prespecified significance level. In fact, the statistical size (relative frequency of Type I error of wrongly rejecting efficiency) will be smaller than the nominal significance level a. Obviously, the natural question is: How much statistical power does this approach have? To answer this question, Section 4 will extend Post’s (2003) simulation study.

III. Computational Issues

At first sight, the test statistic $x(F_C)$ seems computationally intractable, because of two undesirable features. First, the utility functions $u \in U_2^*$ are of infinite dimension. Second, the inverted variance-covariance matrix $W(u, F_?)^{-1}$ is a complex function of the utility function *u*. However, there are ways to circumvent these problems.

The gradient vector $\nabla u(C^T t)$ is the only aspect of the utility function that is actually used for computing the test statistic. This gradient vector is of finite dimensions (*T*) and all relevant gradient vectors can be represented by a *T*-dimensional polyhedron:

$$B \equiv \left\{\nabla u(C^T t) : u \in U_2^*\right\} = \left\{b \in \Re_+^T : b_1 \geq b_2 \geq \cdots \geq b_T ; T^{-1} b^T \boldsymbol{e} = 1\right\} \quad \textbf{(7)}$$

Note that the ordering of the gradient vector ($b_1 \geq b_2 \geq \cdots \geq b_T$) reflects the risk aversion condition and the ordering of the data ($\boldsymbol{x}_1^{\mathrm{T}} t < \boldsymbol{x}_2^{\mathrm{T}} t < \cdots < \boldsymbol{x}_T^{\mathrm{T}} t$; see Assumption 4). Using B, the test statistic can be reformulated as follows:

$$x(F_{\mathrm{C}}) = \min_{b \in \mathrm{B}} (b^{\mathrm{T}} \mathrm{C})(b^{\mathrm{T}} \mathrm{C} \mathrm{C}^{\mathrm{T}} b)^{-1} (\mathrm{C}^{\mathrm{T}} b) \qquad \textbf{(8)}$$

This leaves us with the problem of working with the inverted matrix $(b^{\mathrm{T}} \mathrm{C} \mathrm{C}^{\mathrm{T}} b)^{-1}$, which is a complex function of b. We can deal with this problem by using an iterative approach in the spirit of Hansen (1982), Ferson and Foester (1994) and Hansen, Heaton, and Yaron (1996). Let

$$x_{?}(F_{\mathrm{C}}) \equiv \min_{b \in \mathrm{B}} (b^{\mathrm{T}} \mathrm{C})(\boldsymbol{?}^{\mathrm{T}} \mathrm{C} \mathrm{C}^{\mathrm{T}} \boldsymbol{?})^{-1} (\mathrm{C}^{\mathrm{T}} b) \qquad \textbf{(9)}$$

and

$$b_g^* \equiv \arg\min_{b \in \mathrm{B}} (b^{\mathrm{T}} \mathrm{C})(\boldsymbol{?}^{\mathrm{T}} \mathrm{C} \mathrm{C}^{\mathrm{T}} \boldsymbol{?})^{-1} (\mathrm{C}^{\mathrm{T}} b) \qquad \textbf{(10)}$$

with g for a prespecified vector. Note that computing $x_{?}(F_{\mathrm{C}})$ requires solving a quadratic objective function under linear constraints. This is a Convex Quadratic Programming (CQP) problem that can be solved using straightforward mathematical programming techniques.[4]

We can estimate $x(F_{\mathrm{C}})$ by computing $x_{?}(F_{\mathrm{C}})$ iteratively for different values of g . We may start with the initial vector $g_1 = \boldsymbol{e}$ [TTP2]. This will yield an initial consistent estimator $b_{g_1}^*$ for the gradient vector, which is then used as g_2 in the second iteration. We may then stop and approximate $x(F_{\mathrm{C}})$ by $x_{?_2}(F_{\mathrm{C}})$ or conduct further iterations, each using $g_s = b_{g_{s-1}}^*$, s=3,…, possibly until convergence, to obtain a more efficient estimator. To reduce the computational burden of our simulations, we will use a two-stage estimator in this study.

IV. Simulation[F3]

Using a simulation experiment, Post (2003, Section IIIC) demonstrates that his SSD test procedure involves low power in small and medium-sized samples generated from a normal distribution fitted to the returns of the well-known 25 Fama and French stock portfolios formed on size and B/M. Post and Van Vliet (2004) show that that using the correct null of SSD efficiency rather than the null of equal means further reduces the power of the SSD test. In part, the lack of power in the Post experiment reflects the difficulty of estimating a 25-dimensional multivariate return distribution. It is likely that the power increases (at an increasing rate) as the length of the cross-section is reduced to for example ten benchmark portfolios, which is common in asset-pricing tests. Further, as discussed in the introductory section, the lack of power may reflect the focus on the maximum positive pricing error. The novel SSD-GMM test procedure focuses on all errors and hence may be expected to involve substantially more power.

To shed some light on the statistical properties of the SSD-GMM test procedure, we extend the Post and Van Vliet (2004) simulation experiment. The simulations involve ten assets with a multivariate normal return distribution. The joint population moments are equal to the sample moments of the monthly excess returns of the ten Fama and French stock portfolios formed on B/M, in the sample from July 1963 to October 2001.

We will analyze the statistical properties of the SSD-GMM test procedure and the Post and Van Vliet (2004) test. Comparing these two test procedures will reveal the value added of accounting for all pricing errors. We will apply both procedures to two test portfolios in random samples drawn from this multivariate normal distribution. The equal weighted portfolio (EP) is known to be SSD inefficient relative to the normal population distribution.[5] Hence, we may analyze the statistical power of the competing test procedures by their ability to correctly classify EP as inefficient. By contrast, the ex ante tangency portfolio (TP) is SSD efficient and we may analyze the statistical size by the relative frequency of random samples in which this portfolio is wrongly classified as inefficient.[6]

We draw 10,000 random samples from the multivariate normal population distribution through Monte-Carlo simulation. For every random sample, we apply the both test procedures to the efficient TP and the inefficient EP. For both procedures, we compute the size as the rejection rate for TP and the power as the rejection rate for EP. This experiment is performed for a sample size (T) of 25 to 2,000 observations and for a significance level (a) of 2.5, 5, and 10 percent.

Table I summarizes the results. For both procedures, the size is generally substantially smaller than the nominal level of significance a, and it converges to zero. In fact, the size is smaller than one percent for samples as small as 250 observations and with a level of significance as high as ten percent. Presumably, this reflects our use of conservative p-values and critical values (see Section II).

As discussed in Section II, $x(F_C)$ converges to $x(G)$, and we expect minimal Type II error in large samples. Indeed, for both procedures, the power goes to unity as we increase the sample size. However, in small samples, the tests lack power. Clearly, we cannot expect reliable non-parametric estimates of a 10-dimensional return distribution based on a few observations. The good news is that the SSD-GMM procedure is substantially more powerful than the SSD-LP procedure. For example, using a ten percent significance level, the SSD-GMM procedure achieves a rejection rate of 25 percent for samples of about 200 observations. By contrast, for this sample size, the SSD-LP procedure yields a power of only about 12.5% and the SSD-LP test requires at least 350 observations to achieve a rejection rate of 25%. Clearly, accounting for all pricing errors rather than the maximum positive error only substantially improves the power of the SSD test.

In fact, this simulation may underestimate the potential improvement of power. Specifically, the maximum positive pricing error relative to the population distribution in our simulations is relatively high. The SSD-GMM test can be expected to perform even better in cases with for instance large negative pricing errors, which will not be detected by the SSD-LP test. [F4]

[Insert Table I about here]

V. Empirical Application

We analyze if the CRSP all-share index is mean-variance and SSD efficient. This value-weighted index consists of all common stocks listed on NYSE, AMEX, and NASDAQ. To proxy the investment universe of individual assets, we use three well-known sets of ten benchmark portfolios that capture three important asset pricing puzzles: the size, B/M and momentum effects. The first two sets are the decile portfolios formed on market capitalization and the decile portfolios formed on book-to-market-equity ratio (B/M). Fama and French

(1992) (1993)) give a detailed [TTP5]description of the selection procedures used to construct these benchmark portfolios. In our analysis of the size and B/M portfolios, we focus on the period from January 1933 to December 2002 (840 months). Third, we use ten portfolios sorted on price momentum as described in Jegadeesh and Titman (2001). This benchmark set ranges from January 1965 to December 1998 (408 months).[7] Table I gives descriptive statistics for the three data sets.

[Insert Table II about here]

Clearly, it makes little sense to assume that the return distribution remains stable during a 70-year period; there exists a wealth of evidence to suggest that the risk profile of stocks and the risk preferences of investors change through time. To control for structural variation in risk and risk premiums, we employ a rolling window analysis. With one-month steps, we consider all 120-month samples from January 1933 to December 2002 (721 samples in total). In every subsample, we apply our SSD-GMM test as well as a standard Gibbons, Ross, and Shanken (1989) test for mean-variance efficiency. Figure 2 shows the rolling p-values for both tests.

[Insert Figure 1 about here]

The three asset pricing puzzles are clearly present in the figure. Specifically, severe violations of mean-variance efficiency occur for the size, B/M and momentum portfolios, with GRS p-values reaching levels far below 10% and sometimes below 1% or even 0.1%. For the size portfolios, the violations are concentrated especially in the 1980s. Mean-variance inefficiency relative to the value portfolios is most pronounced in the 1950s, the late 1960s, early 1970s and the 1980s. We find little evidence for size and B/M effects during the 1970s. By contrast, Post and Van Vliet find strong violations of market portfolio efficiency during this decade. This difference reflects the maximum positive pricing error (the focal point of Post and Van Vliet) being much larger than the remaining pricing errors; based on a weighted average of all errors (the focus of the SSD-GMM test), we cannot reject market portfolio efficiency. The momentum effect is present in the entire sample period (1970s, 1980s and 1990s).

Interestingly, the size effect disappears if we use the SSD test; the SSD p-values are generally far above ten percent, with an incidental drop to about 10% around 1990. This result confirms part of the findings of Post and Van Vliet; the size effect can be explained by market risk not captured by variance, most notably the left tail of the return distribution (market crash risk). However, contrary to Post and Van Vliet, we cannot (fully) rationalize the value and momentum effects. For the early 1970s and the 1980s, the value effect disappears if we use the SSD-GMM test, but for the 1950s and the late 1960s, the effect remains unexplained, with SSD p-values far below ten percent. Similarly, the SSD test cannot explain market portfolio inefficiency relative to the momentum portfolios during the entire sample period. In fact, the evidence against SSD efficiency in the momentum data set is overwhelming, with p-values reaching the 0,01% level and never exceeding the 10% level.

Again, the remarkable difference with the results of Post and van Vliet (who cannot reject SSD efficiency relative to B/M and momentum portfolios) can be explained by their focus on the maximum positive pricing error. For the value portfolios, the maximum positive pricing error is relatively small during the 1950s and the late 1960s and hence the market portfolio passes the SSD-LP test. However, the B/M 1 portfolio of growth stocks involves a large negative pricing error and hence the SSD-GMM test classifies the market portfolio as inefficient. [F6]Similarly, the maximum positive pricing error for the Momentum 10 portfolio of past winners is relatively small, while the Momentum 1 portfolio of past losers involves a large negative pricing error.

In summary, we conclude that (1) the size effect can be explained by market risk not captured by variance, (2) the value and momentum effects cannot be fully explained by market risk and (3) Post and van Vliet fail to detect the latter effects due to the use of a less powerful test that focuses on the maximum positive pricing error. However, the Post and Van Vliet results are still relevant if short selling is difficult to implement due to margin requirements and explicit or implicit restrictions faced by institutional investors. We cannot fully exploit the large negative pricing errors of growth stocks and past losers without short selling, and adding value stocks and past winners to the portfolio comes at the cost of increasing the downside risk of the portfolio (as shown in Post and Van Vliet).

VI. Concluding Remarks

We develop an empirical test for Second-order Stochastic Dominance (SSD) efficiency that is embedded in the Generalized Method of Moments (GMM) framework. In contrast to Post's (2003) Linear Programming test, the GMM test considers all pricing errors rather than the maximum positive error only. The test statistic can be computed by iterating a Convex Quadratic Programming problem that can be solved using straightforward mathematical programming techniques. Theorem 1 allows for statistical inference using conservative asymptotic p-values and critical values. Our simulations show that this approach has superior statistical properties compared to the original LP test. Using this test, we demonstrate that the anomalous size effect can be explained with risk not captured by variance alone. By contrast, the market portfolio remains SSD inefficient relative to value and momentum portfolios due to the overvaluation of growth stocks and past losers. Nevertheless, the Post and Van Vliet (2004) results are still relevant if short selling is excluded, because short selling is required to fully exploit overvalued stocks.

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

References

Fama, Eugene F., and Kenneth R. French, 1993, Common risk-factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3—56.

Fama, Eugene F., and Kenneth R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55—84.

Ferson, W. E., and S. R. Foerster, 1994, Finite Sample Properties of the Generalized Method of Moments in Tests of Conditional Asset Pricing Models, *Journal of Financial Economics* 36, 29—55.

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

Hansen, L. P., 1982, Large Sample Properties of Generalized Method of Moments Estimator, *Econometrica* 50, 1029—1054.

Hansen, L. P., J. Heaton, and A. Yaron, 1996, Finite-Sample Properties of Some Alternative GMM Estimators, *Journal of Business and Economic Statistics* 14, 262—280.

Jegadeesh, Narasimhan, and Sheridan Titman, 2001, Profitability of momentum strategies: An evaluation of alternative explanations, *Journal of Finance* 56, 699—720.

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

Post, Thierry, 2003, Empirical tests for stochastic dominance efficiency, *Journal of Finance* 58, 1905—1932.
Post, Thierry and Levy, Haim, 2002, Does Risk Seeking Drive Asset Prices? A Stochastic Dominance Analysis of Aggregate Investor Preferences, Tinbergen Institute Working Paper No. 02-070/2; and ERIM Report Series Reference No. ERS-2002-50-F&A, http://ssrn.com/abstract=320767.
Post, Thierry and van Vliet, Pim, 2004, Downside Risk and Asset Pricing, ERIM working paper, http://ssrn.com/abstract=503142.

Appendix

Proof of THEOREM 1 (sampling distribution) It follows directly from the definition of our test statistic that

$$x(F_C) \leq \min_{u \in U_2^*: m(u,G)=\mathbf{0}} z(u, F_C) \equiv T m(u, F_C)^{\mathrm{T}} \mathrm{W}(u, F_C)^{-1} m(u, F_C) \qquad \text{(i)}$$

Note that under the null, the set $\{u \in U_2^* : m(u,G) = 0\}$ is nonempty. Also, since F_C converges to G, the inequality (i) asymptotically becomes an equality.

We can derive the asymptotic distribution of $z(u,F_C)$ for any given $u \in U_2^* : m(u,G) = \mathbf{0}$ from known results. Since the observations are serially IID, it follows from the Levy-Lindenberg central limit theorem that $m(u,F_C) = T^{-1}(\nabla u(\mathrm{C}^{\mathrm{T}} t)^{\mathrm{T}} \mathrm{C})$ obeys an asymptotic multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $T^{-1}\mathrm{W}(u,G)$, i.e., $m(u,F_C) \xrightarrow{d} N(\mathbf{0}, T^{-1}\mathrm{W}(u,G))$. Using this result and $\mathrm{W}(u,F_C) \xrightarrow{p} \mathrm{W}(u,G)$, we find that $T^{-1/2} m(u,F_C)^{\mathrm{T}} \mathrm{W}(u,F_C)^{-1/2}$ obeys an asymptotic multivariate standard normal distribution, i.e., $T^{-1/2} m(u,F_C)^{\mathrm{T}} \mathrm{W}(u,F_C)^{-1/2} \xrightarrow{d} N(\mathbf{0}, \mathrm{I}_N)$. Consequently, $z(u,F_C)$ obeys[F7] an asymptotic chi-squared distribution with N degrees of freedom, that is,

$$z(u,F_C) \xrightarrow{d} c_N^2 \quad \forall u \in U_2^*(t) : m(u,G) = \mathbf{0}. \qquad \text{(ii)}$$

Combining (i) and (ii), we find that the asymptotic null distribution of $x(F_C)$ is bounded from above by the chi-squared distribution. **Q.E.D.**

Table I

Statistical properties of the competing test procedures [TTP8]

The table displays the statistical size and power for various numbers of time-series observations (T) and for a significance level (a) of 2.5, 5 and 10 percent. Panel A represents the SSD-LP test of Post and Van Vliet (2004) test, which excludes short sales and focuses on the maximum positive pricing error only, and Panel B gives the results for the novel SSD-GMM test which includes short sales and account for all pricing errors. The results are based on 10,000 random samples from a multivariate normal distribution with joint moments equal to the sample moments of the monthly excess returns of the ten B/M portfolios for the period from July 1963 to October 2001. Size is measured as the relative frequency of random samples in which the efficient tangency portfolio (TP) is wrongly classified as inefficient. Power is measured as the relative frequency of random samples in which the inefficient equally weighted portfolio (EP) is correctly classified as inefficient.

Panel A: SSD-LP						
	Statistical Size			Statistical Power		
	a=10%	a=5%	a=2.5%	a=10%	a=5%	a=2.5%
T=25	0.059	0.030	0.016	0.050	0.018	0.007
T=50	0.047	0.023	0.012	0.057	0.021	0.008
T=100	0.030	0.013	0.006	0.074	0.029	0.011
T=200	0.012	0.004	0.002	0.123	0.052	0.021
T=500	0.001	0.000	0.000	0.436	0.256	0.127
T=1,000	0.000	0.000	0.000	0.929	0.881	0.779
T=2,000	0.000	0.000	0.000	1.000	1.000	1.000
Panel B: SSD-GMM						
	Statistical Size			Statistical Power		
	a=10%	a=5%	a=2.5%	a=10%	a=5%	a=2.5%
T=25	0.159	0.099	0.062	0.092	0.057	0.039
T=50	0.119	0.074	0.041	0.106	0.066	0.045
T=100	0.067	0.041	0.018	0.140	0.090	0.061
T=200	0.021	0.013	0.003	0.235	0.159	0.106
T=500	0.001	0.000	0.000	0.671	0.571	0.427
T=1,000	0.000	0.000	0.000	0.980	0.972	0.941
T=2,000	0.000	0.000	0.000	1.000	1.000	1.000

Table II
Descriptive Statistics Data Sets

The table shows descriptive statistics for the monthly excess returns of the value-weighted CRSP all-share market portfolio and the size, B/M and momentum data sets. The sample period is from January 1933 to December 2002 (T=840) for the size and B/M data sets and from January 1965 to December 1998 (T=408) for the momentum data set. Excess returns are computed from the raw return observations by subtracting the return on the one-month US Treasury bill from Ibbotson. The size and B/M data are taken from the homepage of Kenneth French, the momentum data are the courtesy of Narasimhan Jegadeesh.

Panel A: Size and B/M Portfolios (T=840 months)

		Mean	Stdev.	Skewness	Kurtosis	Min	Max
	Market	0.714	4.937	0.156	9.18	-23.67	38.17
10 Size Portfolios	Small	1.328	9.425	2.988	29.49	-34.59	95.97
	2	1.173	8.368	2.399	28.24	-32.93	33.30
	3	1.107	7.432	1.599	21.00	-29.65	95.19
	4	1.047	7.041	1.311	16.88	-30.07	78.59
	5	1.019	6.706	1.086	15.77	-28.89	64.25
	6	0.962	6.298	0.799	13.05	-26.89	57.52
	7	0.928	6.105	0.774	13.77	-29.07	53.07
	8	0.854	5.638	0.379	10.11	-24.90	54.42
	9	0.811	5.327	0.656	12.95	-23.80	45.86
	Big	0.657	4.665	0.093	7.95	-22.99	49.10
10 B/M Portfolios	Growth	0.630	5.366	0.193	7.32	-23.30	38.45
	2	0.695	5.104	-0.192	6.68	-25.19	71.70
	3	0.676	4.967	-0.196	6.87	-26.47	28.74
	4	0.723	5.365	1.210	19.30	-24.26	27.24
	5	0.843	4.982	0.752	13.89	-24.58	56.29
	6	0.898	5.276	0.621	13.40	-26.20	46.15
	7	0.912	5.778	1.424	18.55	-25.62	48.79
	8	1.077	5.910	1.074	14.79	-29.08	59.17
	9	1.136	6.943	1.470	18.83	-30.87	52.43
	Value	1.164	8.215	1.496	18.97	-45.76	62.24

Panel B: Momentum Portfolios (T=408 months)

		Mean	Stdev.	Skewness	Kurtosis	Min	Max
	Market	0.513	4.472	-0.505	5.42	-23.09	16.05
10 Mom. Portfolios	Loser	-0.105	6.878	0.001	5.20	-28.59	28.98
	2	0.374	5.674	-0.085	5.50	-22.68	20.05
	3	0.520	5.156	-0.150	5.80	-23.01	26.82
	4	0.586	4.873	-0.334	6.47	-25.18	25.13
	5	0.606	4.709	-0.491	6.87	-25.90	23.85
	6	0.640	4.701	-0.694	6.93	-26.61	22.51
	7	0.666	4.770	-0.923	7.26	-27.80	20.72
	8	0.750	4.995	-1.041	7.30	-29.18	18.98
	9	0.860	5.391	-1.038	6.73	-30.27	17.31
	Winner	1.123	6.561	-0.882	5.52	-32.74	15.84

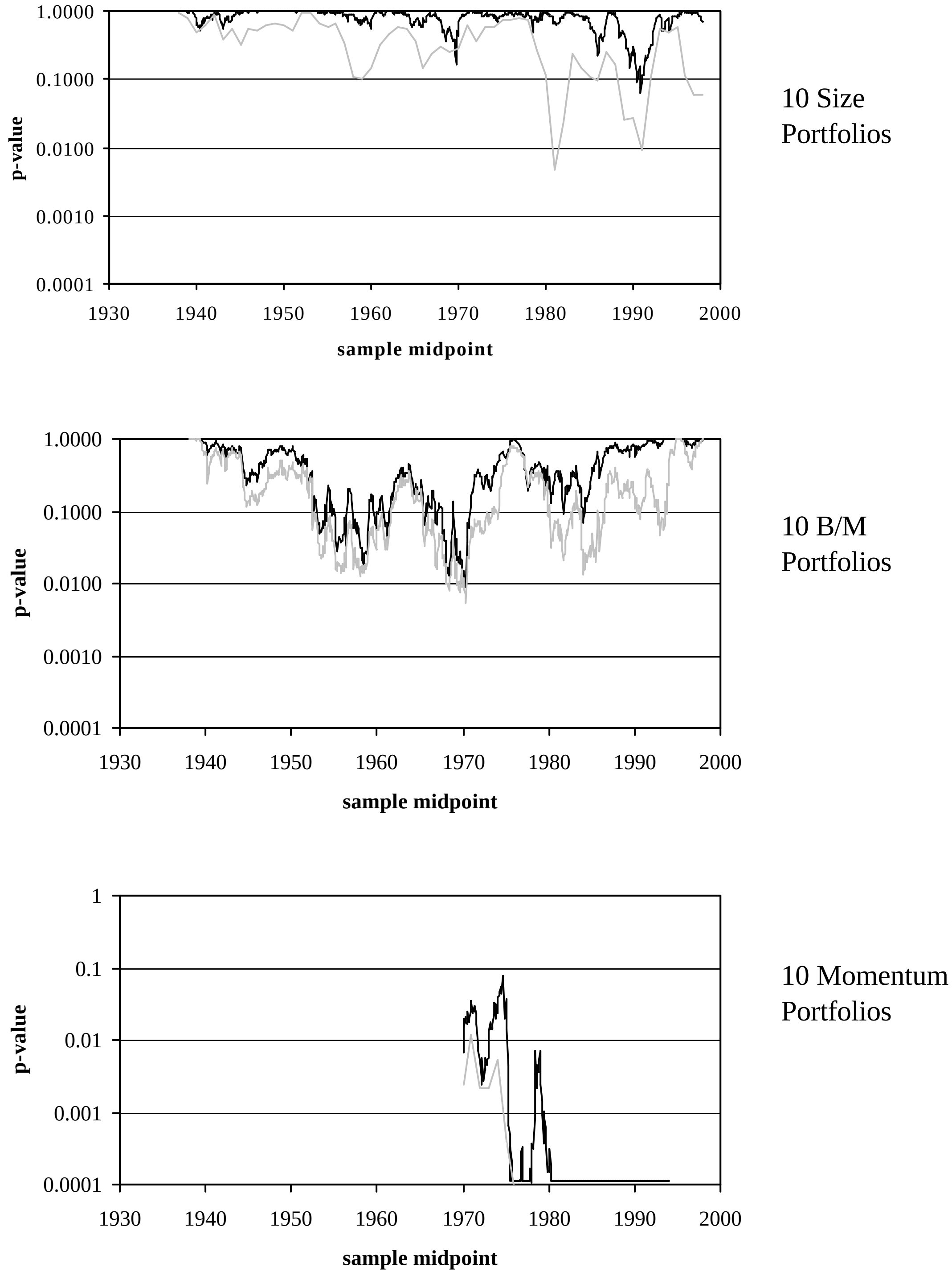

Figure 1
Rolling window analysis of efficiency tests

This figure shows the p-values for the GRS and SSD-GMM efficiency tests in each of the three data sets using a rolling 120-months period (1-month steps). The grey line represents the p-value of the GRS test and the dark line shows the SSD-GMM p-value. To read the figure, consider the January 1980 [TTP9]observation in the size data set, which represents the 120-month period from January 1975 to December 1984. For this period, the stock market is mean-variance inefficient (p=0.005) but SSD efficient (p=0.676).

Footnotes

[1] 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 $\boldsymbol{x}$.

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

[3] We do not claim that utility is always concave. However, we need this condition in order to justify using the first-order condition, or Euler equation (1), to test if the evaluated portfolio is the global maximum. If utility is non-concave, then the first-order condition also applies for possible local optima and for a possible global minimum. Post and Levy (2002) further discuss the se of non-concave utility functions in asset pricing tests.

[4] In this study, we will use the QPROG module of Aptech System's GAUSS software to solve this problem.

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

[6] Note that the SSD-GMM and SSD-LP tests rely on different efficient sets and involve different tangency portfolios. The SSD-LP test excludes short sales and the relevant tangency portfolio consists of 18.22%, 2.04% and 79.74% invested in the fifth, sixth and eight B/M portfolios respectively. The SSD-GMM test allows for short sales and the weights of the tangency portfolio are as follows: (1) -22.26%, (2) 42.56%, (3) -70.41%, (4) -109.98%, (5) -33.39%, (6) 78.38%, (7) 35.40%, (8) 68.56%, (9) 89.22% and (10) 21.93%.

[7] We thank Narasimhan Jegadeesh for sharing his data with us.

Page: 5
[TTP1]Also true for parametric GMM with inequality constraints. Use of bootstrapped p-values?
Page: 6
[TTP2] “OLS”
Pagina: 6
[F3] We may consider adding the GRS or MV-GMM procedures. Of course, they will have an “unfair” advantage due to the normal distribution
Pagina: 7
[F4] Klopt dit wel?
Page: 8
[TTP5]Worden de decile portfolio daadwerkelijk besproken in deze refs?
Pagina: 8
[F6] Klopt dit wel?
Pagina: 10
[F7] Hier nog een verwijzing aan naar een stelling?
Page: 11
[TTP8] Add MV-GMM results
Page: 13
[TTP9] Checken!

Publications in the Report Series Research* in Management

ERIM Research Program: "Finance and Accounting"

2004

Corporate Finance In Europe Confronting Theory With Practice
Dirk Brounen, Abe de Jong and Kees Koedijk
ERS-2004-002-F&A
http://hdl.handle.net/1765/1111

Downside Risk And Asset Pricing
Thierry Post and Pim van Vliet
ERS-2004-018-F&A

An Alternative Decomposition Of The Fisher Index
Winfried G. Hallerbach
ERS-2004-022-F&A
http://hdl.handle.net/1765/1220

A Gmm Test For Ssd Efficiency
Thierry Post And Philippe Versijp
ERS-2004-024-0F&A

The effects of systemic crises when investors can be crisis ignorant
Erik Kole, Kees Koedijk & Marno Verbeek
ERS-2004-027-F&A
http://hdl.handle.net/1765/1270

The Eco-Efficiency Premium Puzzle
Jeroen Derwall, Nadja Günster, Rob Bauer and Kees Koedijk
ERS-2004-043-F&A
http://hdl.handle.net/1765/1296

Shareholders' Voting at General Meetings: Evidence from the Netherlands
Abe De Jong, Gerard Mertens And Peter Roosenboom
ERS-2004-039-F&A
http://hdl.handle.net/1765/1332

Do banks influence the capital structure choices of firms?
Petra Daniševská, Abe de Jong & Marno Verbeek
ERS-2004-040-F&A
http://hdl.handle.net/1765/1333

---

* A complete overview of the ERIM Report Series Research in Management:
https://ep.eur.nl/handle/1765/1

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