Testing for Stochastic Dominance Efficiency

Thierry Post, Oliver Linton and Yoon-Jae Whang

ERIM REPORT SERIES *RESEARCH IN MANAGEMENT*	
ERIM Report Series reference number	ERS-2005-033-F&A
Publication	May 2005
Number of pages	25
Persistent paper URL	
Email address corresponding author	gtpost@few.eur.nl
Address	Erasmus Research Institute of Management (ERIM) RSM Erasmus University / Erasmus 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

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

ERASMUS **R**ESEARCH **I**NSTITUTE OF **M**ANAGEMENT

REPORT SERIES
RESEARCH IN MANAGEMENT

ABSTRACT AND KEYWORDS	
Abstract	We propose a new test of the stochastic dominance efficiency of a given portfolio over a class of portfolios. We establish its null and alternative asymptotic properties, and define a method for consistently estimating critical values. We present some numerical evidence that our tests work well in moderate sized samples.
Free Keywords	Stochastic Dominance, Portfolio Diversification, Asset Pricing, Portfolio Analysis
Availability	The ERIM Report Series is distributed through the following platforms: Academic Repository at Erasmus University (DEAR), DEAR ERIM Series Portal Social Science Research Network (SSRN), SSRN ERIM Series Webpage Research Papers in Economics (REPEC), REPEC ERIM Series Webpage
Classifications	The electronic versions of the papers in the ERIM report Series contain bibliographic metadata by the following classification systems: Library of Congress Classification, (LCC) LCC Webpage Journal of Economic Literature, (JEL), JEL Webpage ACM Computing Classification System CCS Webpage Inspec Classification scheme (ICS), ICS Webpage

Testing for Stochastic Dominance Efficiency

Oliver Linton*
London School of Economics

Thierry Post†
Erasmus University

Yoon-Jae Whang‡
Korea University

PRELIMINARY AND INCOMPLETE

May 13, 2005

Abstract

We propose a new test of the stochastic dominance efficiency of a given portfolio over a class of portfolios. We establish its null and alternative asymptotic properties, and define a method for consistently estimating critical values. We present some numerical evidence that our tests work well in moderate sized samples.

1 Introduction

We propose a test of whether a given portfolio is efficient with respect to the stochastic dominance criterion in comparison with a set of portfolios formed from a given finite set of assets. The stochastic dominance criteria represent economically meaningful restrictions, but avoid further restrictions like those imposed in mean variance analysis. Post (2003) and Post and Versijp (2004) have recently proposed tests of the same hypothesis and provide a method of inference based on a duality representation of the investor's expected utility maximization problem. Their approach uses a conservative bounding distribution, which may compromise statistical power or the ability to detect inefficient portfolios in small samples. We propose a more standard approach the inference problem.

*Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: `lintono@lse.ac.uk`. Research supported by the ESRC.

†Erasmus University, Rotterdam, P.O. Box 1738, 3000 Rotterdam, The Netherlands. E-mail address: `gtpost@few.eur.nl`.

‡Department of Economics, Korea University, Seoul 136-701, Korea. Email Address : `whang@korea.ac.kr`.

We propose a more standard statistical approach to the problem. Specifically we suggest to use a modification of the Kolmogorov-Smirnov test statistic of McFadden (1989) and Klecan, McFadden, and McFadden (1991). Recently, Linton, Maasoumi, and Whang (2004) (hereafter LMW) have provided a comprehensive theory of inference for this class of test statistics for the standard pairwise comparison of prospects. We extend their work to the portfolio case. This entails a nontrivial computational issue, which we propose to solve using a nested linear programming algorithm. We provide the limiting distribution under the null hypothesis of SD efficiency, and some results on asymptotic power. We propose to use either the subsampling method or a recentered bootstrap method for obtaining the critical values. We evaluate the performance of our method on simulated data.

We focus on the stochastic dominance criteria of order two and higher. For various reasons, we do not cover the first-order criterion, which allows for risk seeking behaviour: (1) risk aversion is a standard assumption in financial economics and FSD seems less relevant than SSD for testing portfolio efficiency, (2) the definition of FSD efficiency in a portfolio context is ambiguous, (3) our computational strategy breaks down for local risk seekers, (4) the FSD criterion is very general and presumably lacks statistical power for the typical sample size.

2 Null Hypothesis

Let $X_t = (X_{1t}, \ldots, X_{Kt})^\top$ for $t = 1, \ldots, T$ be observations on a set of K assets, and let Y_t be some benchmark asset; Y_t could be a portfolio of X_t. We consider other portfolios with return $X_t^\top \lambda$, where $\lambda = (\lambda_1, \ldots, \lambda_K)^\top$, $\Lambda = \{\lambda \in \mathbb{R}^K : e^\top \lambda = 1\}$, and $e = (1, \ldots, 1)^\top$. The approach applies also for a portfolio possibilities set with the shape of a general polytope, allowing for general linear constraints, such as short selling constraints, position limits and restrictions on risk factor loadings. Let Λ_0 be some subset of Λ reflecting whatever additional restrictions if any are imposed on Λ. Let $\mathcal{U}_1$ denote the class of all von Neumann-Morgenstern type utility functions, u, such that $u' \geq 0$, (increasing). Also, let $\mathcal{U}_2$ denote the class of all utility functions in $\mathcal{U}_1$ for which $u'' \leq 0$ (strict concavity).

Definition 1. (SSD Efficiency) *The asset Y_t is SSD efficient if and only if some $u \in \mathcal{U}_2$, $E[u(Y_t)] \geq E[u(X_t^\top \lambda)]$ for all $\lambda \in \Lambda_0$.*

Likewise one can define third order efficiency. Let $F_\lambda(\cdot)$ and $F_Y(\cdot)$ be the c.d.f.'s of $X_t^\top \lambda$ and Y_t, respectively. For a given integer $s \geq 1$, define the s -th order integrated c.d.f. of $X_t^\top \lambda$ to be

$$G_\lambda^{(s)}(x) = \int_{-\infty}^{x} G_\lambda^{(s-1)}(y)dy,$$

where $G_\lambda^{(0)}(\cdot) = F_\lambda(\cdot)$, and likewise for $G_Y^{(s)}(x)$. A portfolio $X_t^\top \lambda$ s-order dominates Y_t if and only if $G_\lambda^{(s)}(x) - G_Y^{(s)}(x) \leq 0$ for all x with strict inequality for at least one x in the support $\mathcal{X}$. For $s \geq 2$

this definition is equivalent to definition 1, but not so for $s = 1$, see Post (2005) for discussion. Thus our results are only meaningful for $s \geq 2$, although we retain the general definition. For notational simplicity, we sometimes let the dependence on s of the quantities introduced below be implicit , i.e., we write $G_\lambda^{(s)}$ as G_λ and so on. We wish to test the null hypothesis that Y_t is *s-th order SD efficient* in the sense that there does not exist any portfolio in $\{X_t^\top \lambda : \lambda \in \Lambda_0\}$ that dominates it, where Λ_0 is a compact subset of Λ. This hypothesis has previously been tested by Post (2003) and Post and Versijp (2004) among others. In order to test this hypothesis we must provide a scalar valued population functional that divides the null from alternative.

2.1 First Functional

Suppose we consider the functional

$$\sup_{\lambda \in \Lambda_0} \inf_{x \in \mathcal{X}} \left[G_Y(x) - G_\lambda(x) \right] \tag{1}$$

that is essentially a modification of the functional used in LMW to test for stochastic dominance between fixed alternatives. This functional satisfies (1)≤ 0 under the null hypothesis. Unfortunately, there are some elements of the alternative for which (1)$= 0$. In fact, the null hypothesis is quite complex, and to characterize it we introduce some further notation. For each λ define the three subsets

$$\begin{aligned} A_\lambda^- &= \{x \in \mathcal{X} : G_Y(x) - G_\lambda(x) < 0\} \\ A_\lambda^= &= \{x \in \mathcal{X} : G_Y(x) - G_\lambda(x) = 0\} \\ A_\lambda^+ &= \{x \in \mathcal{X} : G_Y(x) - G_\lambda(x) > 0\} . \end{aligned}$$

If $X_t^\top \lambda$ dominates Y_t, then $A_\lambda^- = \varnothing$, and A_λ^+ is nonempty. However, it can be that both $A_\lambda^=$ and A_λ^+ are nonempty in which case $\inf_{x \in \mathcal{X}} (G_Y(x) - G_\lambda(x)) = 0$. The supremum over the entire support fails to distinguish between weak and strict inequality. This is perhaps less of an issue in testing dominance of one outcome over another, since the reverse comparison will identify that $\inf_{x \in \mathcal{X}} (G_\lambda(x) - G_Y(x)) < 0$. However, it does matter here. Specifically, suppose that $A_\lambda^=$ and A_λ^+ are non-empty and $A_\lambda^- = \varnothing$ for some λ's. For these λ's, we have $\inf_{x \in \mathcal{X}} (G_Y(x) - G_\lambda(x)) = 0$ even though $X_t^\top \lambda$ dominates Y_t. If the other λ's are such that we have only $A_\lambda^=$ and A_λ^- non-empty so that $\inf_{x \in \mathcal{X}} (G_Y(x) - G_\lambda(x)) < 0$ for those values, then we obtain that (1)$= 0$. The following example in Figure 1 shows that X strictly dominates Y but $\inf_{x \in \mathcal{X}} [G_Y(x) - G_X(x)] = 0$.

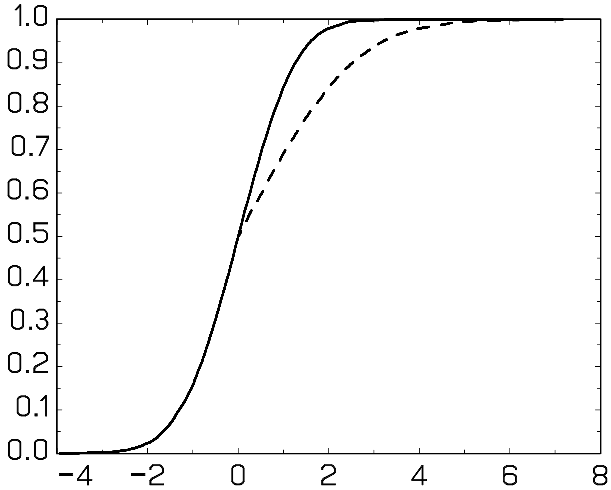

Figure 1. Shows the c.d.f of Y (solid line) and X (dashed line)

We next suggest some modifications of (1) that properly characterize the null hypothesis.

2.2 Our Functional

For each $\epsilon > 0$, define the ϵ-enlargement of the set $A_\lambda^=$,

$$(A_\lambda^=)^\epsilon = \{x + \eta \in \mathcal{X} : x \in A_\lambda^= \text{ and } |\eta| < \epsilon\},$$

and let

$$B_\lambda^\epsilon = \begin{cases} \mathcal{X} \backslash (A_\lambda^=)^\epsilon & \text{if } A_\lambda^= \neq \mathcal{X} \\ \mathcal{X} & A_\lambda^= = \mathcal{X}. \end{cases} \tag{2}$$

Then let

$$d_*(\epsilon) = \sup_{\lambda \in \Lambda_0} \inf_{x \in B_\lambda^\epsilon} \left[G_Y(x) - G_\lambda(x) \right]. \tag{3}$$

Under the null hypothesis, $d_*(\epsilon) \leq 0$ for each $\epsilon \geq 0$, while under the alternative hypothesis we have $d_*(\epsilon) > 0$ for some $\epsilon > 0$. The idea is that you prevent the inner infimum ever being zero through equality on some part of $\mathcal{X}$. This functional divides the null from alternative.

For later discussion, we shall also need the following partition of Λ_0 :

$$\Lambda_0 = \Lambda_1 \cup \Lambda_2, \text{ where } \Lambda_1 \cap \Lambda_2 = \varnothing, \ \Lambda_1 = \Lambda_0^- \cup \Lambda_0^=, \ \Lambda_2 = \Lambda_0^+ \cup \Lambda_0^{\simeq} \tag{4}$$

$$\Lambda_0^{=} = \{\lambda \in \Lambda_0 : G_Y(x) = G_\lambda(x)\ \forall x \in \mathcal{X}\} \tag{5}$$

$$\Lambda_0^{-} = \left\{\lambda \in \Lambda_0 : \inf_{x\in\mathcal{X}} [G_Y(x) - G_\lambda(x)] < 0\right\} \tag{6}$$

$$\Lambda_0^{+} = \left\{\lambda \in \Lambda_0 : \inf_{x\in\mathcal{X}} [G_Y(x) - G_\lambda(x)] > 0\right\} \tag{7}$$

$$\Lambda_0^{\simeq} = \left\{\lambda \in \Lambda_0 : \inf_{x\in\mathcal{X}} [G_Y(x) - G_\lambda(x)] = 0,\ \inf_{x\in B_\lambda^\epsilon} [G_Y(x) - G_\lambda(x)] > 0 \text{ for some } \epsilon > 0\right\}. \tag{8}$$

Under the null hypothesis, $\Lambda_2 = \varnothing$ and hence $\Lambda_0 = \Lambda_1$. Under the alternative hypothesis, $\Lambda_1 = \varnothing$ and $\Lambda_0 = \Lambda_2$.

3 Test Statistics

The general approach is to define empirical analogues of (3) as our test statistics. Let $k_T = c_0 \cdot (\log T/T)^{1/2}$ and let ϵ_T denote a sequence of positive constants satisfying Assumption 2 below, where c_0 is a positive constant. Define

$$\widehat{A}_\lambda^{=} = \left\{x \in \mathcal{X} : \left|\widehat{G}_Y(x) - \widehat{G}_\lambda(x)\right| \le k_T\right\}, \tag{9}$$

$$\left(\widehat{A}_\lambda^{=}\right)^{\epsilon_T} = \left\{x + \eta \in \mathcal{X} :\ x \in \widehat{A}_\lambda^{=}, |\eta| < \epsilon_T\right\}, \tag{10}$$

$$\widehat{B}_\lambda^{\epsilon_T} = \begin{cases} \mathcal{X}\backslash(\widehat{A}_\lambda^{=})^{\epsilon_T} & \text{if } \widehat{A}_\lambda^{=} \neq \mathcal{X} \\ \mathcal{X} & \text{if } \widehat{A}_\lambda^{=} = \mathcal{X} \end{cases}. \tag{11}$$

Then, define

$$W_T = \sup_{\lambda\in\Lambda_0} \inf_{x\in\widehat{B}_\lambda^{\varepsilon_T}} Q_T(\lambda, x), \tag{12}$$

where

$$Q_T(\lambda, x) = \sqrt{T}\left[\widehat{G}_Y(x) - \widehat{G}_\lambda(x)\right], \tag{13}$$

$$\widehat{G}_{T\lambda}(x) = \int_{-\infty}^{x} \widehat{G}_{T\lambda}^{(s-1)}(y)dy,\ \widehat{F}_{T\lambda}(x) = \frac{1}{T}\sum_{t=1}^{T} 1(X_t^\top \lambda \le x),$$

and likewise for $\widehat{G}_Y(x)$. This is our proposed test statistic; rejection is for large positive values. Notice that to compute (12) requires potentially high dimensional optimization of a discontinuous non-convex/concave objective function. In the next section we discuss how to compute the various infimums and supremums in (12).

4 Computational Strategy

The supremum over the scalar x is computed by a grid search. The objective function $Q_T(\lambda, x)$ can be written as

$$Q_T(\lambda, x) = \frac{1}{(s-1)!\sqrt{T}} \sum_{t=1}^{T} \left\{ (x - Y_t)^{s-1} 1\left(Y_t \leq x\right) - (x - X_t^\top \lambda)^{s-1} 1\left(X_t^\top \lambda \leq x\right) \right\},$$

see Davidson and Duclos (2000). When $s = 1$, $Q_T(\lambda, x)$ is neither continuous in x nor in λ. When $s = 2$, this function is not differentiable or convex in $\lambda \in \mathbb{R}^K$, but it is continuous in x. When $s = 3$, the objective function is differentiable in x but not in λ. Therefore, one cannot use standard derivative-based algorithms like Newton-Raphson to find the optima. One could replace the empirical c.d.f.'s by smoothed empirical c.d.f. estimates in order to impose additional regularity on the optimization problem so that derivative based iterative algorithms could be used. There is a well-established literature in econometrics concerning this class of non-smooth optimization estimators, see Pakes and Pollard (1989). Nevertheless, it is a difficult problem computationally to achieve the maximum over λ with high accuracy when K is large in the non-smooth case. In the next subsection, we show how to write the optimization problem (in the second order dominance case $s = 2$) as a one-dimensional grid search with embedded linear programming.

4.1 Profiling on the SD Efficient Set

Every SSD efficient portfolio is optimal for some increasing and concave utility function. Russell and Seo (1989) show that each increasing and concave utility function can be represented by an elementary, two-piece linear utility functions characterized by a single scalar threshold parameter, say μ:

$$u_\mu(x) = \min\{x - \mu, 0\}.$$

Thus every efficient portfolio is the solution to the following problem

$$\max_{\lambda \in \Lambda} \frac{1}{T} \sum_{t=1}^{T} \min\{X_t^\top \lambda - \mu, 0\}$$

for some value of μ. It is straightforward to show that this problem is equivalent to the following linear programming problem:

$$\max_{\theta \in \mathbb{R}^T, \lambda \in \mathbb{R}^K} \frac{1}{T} \sum_{t=1}^{T} \theta_t \tag{14}$$

$$\theta_t \leq \sum_{j=1}^{K} \lambda_j X_{jt} - \mu, \ t = 1, \ldots, T \tag{15}$$

$$\theta_t \leq 0,\ t = 1, \ldots, T \tag{16}$$

$$\sum_{j=1}^{K} \lambda_j = 1 \tag{17}$$

$$\lambda_j \geq 0,\ j = 1, \ldots, K, \tag{18}$$

where $\theta = (\theta_1, \ldots, \theta_T)$ and $\lambda = (\lambda_1, \ldots, \lambda_K)$.

Let $\widehat{\lambda}(\mu), \widehat{\theta}(\mu)$ be the solution to (14)-(18) for each μ. In this problem, θ_t captures the discontinuous term $\min\{X_t^\top \lambda - \mu, 0\}$. Specifically, due to the maximization orientation in (14), constraint (15) and/or (16) will be binding and hence $\widehat{\theta}_t = \min\{X_t^\top \widehat{\lambda} - \mu, 0\}$ at the optimum. In brief, the SSD efficient set reduces to a one-dimensional manifold and the elements can be identified by solving the LP problem (14)-(18) for different values of the single threshold parameter μ. We then compute every λ from $\widehat{\lambda}(\mu)$ for $\mu \in M$, where M is some set of values for μ (under no short-selling we can take $M = [\mu_{\min}, \mu_{\max}]$, where $\mu_{\min}, \mu_{\max}$ are the minimum and maximum expected returns of the individual assets respectively). The infimum and supremum in (12) can be computed by a grid search.

We can do with the LP approach, also for higher-order criteria, because the efficient set then is a subset of the SSD efficient set. Explain that we need to search only over the set of SSD efficient portfolios. This is obviously true for the SSD criterion ($s = 2$), but also for the higher-order criteria ($s > 2$), as the efficient set in these cases is a subset of the SSD efficient set.

4.2 Starting Values on the Mean Variance Frontier

An alternative approach is to use a standard Nelder Mead algorithm. This may work in greater generality for higher order and other kinds of dominance criteria. For this algorithm to work well in high dimensional cases one needs good starting values. We propose to obtain these by grid searching over the mean variance (MV) efficient frontier. The MV efficient set is a natural starting point, because for the normal distribution the SD efficient sets reduce to the MV efficient set. The set of mean variance efficient portfolios can be computed in terms of the unconditional mean μ and the covariance matrix Σ of the vector X_t. For given μ_p there exists a unique portfolio $\lambda(\mu_p)$ that minimizes the variance σ_p^2 of the portfolios that achieve return μ_p. The set of mean variance efficient portfolio weights are indexed by the target portfolio return μ_p, specifically

$$\lambda_p = g + h\mu_p, \tag{19}$$

where the vectors $g(\mu, \Sigma), h(\mu, \Sigma)$ satisfy

$$g = \frac{1}{D}\left[B\Sigma^{-1}i - A\Sigma^{-1}\mu\right] \text{ and } h = \frac{1}{D}\left[C\Sigma^{-1}\mu - A\Sigma^{-1}i\right],$$

with the scalars $A = i^\top \Sigma^{-1} \mu$, $B = \mu^\top \Sigma^{-1} \mu$, $C = i^\top \Sigma^{-1} i$, and $D = BC - A^2$, see Campbell, Lo, and McKinlay (1997, p185). Therefore, one takes a grid of values of μ_p and obtains λ_p for this grid and then compute the test statistic. To impose that there is no short selling it suffices to search in the range $M = [\mu_{\min}, \mu_{\max}]$. The optimal value of λ_p can be used as a starting value in some more general optimization algorithm.

5 Discussion

We focus on stochastic dominance criteria of order two and higher, meaning that risk aversion is assumed throughout this study. For various reasons, we do not cover the first-order criterion, which allows for risk seeking behaviour.

First, risk aversion is a standard assumption in financial economics, being consistent with common observations such as risk premiums for risky assets, portfolio diversification and the popularity of insurance contracts. There are indications for local risk seeking behaviour at the individual level, witness for example the popularity of lotteries. However, the bulk of the literature on asset pricing and portfolio selection assumes that investors are globally risk averse when forming investment portfolios. Apart from this, a FSD test for portfolio efficiency adds relatively little value to a SSD test, for the simple reason that risk seekers generally will hold ill-diversified portfolios. Not surprisingly, Kuosmanen (2004) finds that the FSD and SSD criteria yield exactly the same results for testing market portfolio efficiency.

Second, as is shown in Post (2005), the definition of FSD efficiency in a portfolio context is ambiguous. The stochastic dominance rules of order two and higher assume a concave utility function and hence expected utility is a quasiconcave function of the portfolios weights. In this case, we can invoke Sion's (1958) Minimax Theorem to show the equivalence between two definitions of efficiency: (1) a portfolio is efficient if and only if no other portfolio dominates it and (2) a portfolio is efficient if and only if it is the optimal solution for some investor in the class admitted by the relevant SD rule; see Post (2003; Theorem 1). The first-order criterion allows for risk seeking and expected utility generally is not quasiconcave in this case. Therefore, the two definitions generally diverge, with definition (1) being less restrictive than definition (2); a portfolio may be nondominated but still be nonoptimal for all investors. Until the ambiguity surrounding the definition of FSD efficiency in a portfolio context is resolved, it seems premature to develop procedures for statistical inference for this efficiency criterion.

Third, our computational strategy breaks down for local risk seekers. As discussed in Section 4, for the second-order criterion, the efficient set reduces to a one-dimensional manifold and the efficient portfolios can be identified by solving a simple LP problem. In case of higher-order criteria, the efficient set is a subset of SSD efficient set and the same approach can be used. However, the same approach does not apply for FSD. In this case, the elementary Russell-Seo utility functions take the form of discontinuous step functions. Portfolio optimization for these utility functions requires mixed integer programming techniques and generally involves multiple optimal solutions. For this reason, our computational strategy seems inappropriate for the FSD criterion.

Fourth, the FSD criterion is very general and allows for "exotic" preference structures, for example utility functions with inflection points and discontinuous jumps. Thus, an empirical test for FSD efficiency will have considerable freedom to fit a utility function to the data. Presumably, this will considerably slow down the rate of convergence of an empirical test. For the sample size in typical applications, an empirical test will lack statistical power to allow for a meaningful application.

6 Asymptotic Properties

6.1 Null Distribution

To discuss the asymptotic null distribution of our test statistic, we need the following assumptions:

Assumption 1. (i) $\{(X_t^\top, Y_t)^\top : t = 1, \ldots, T\}$ *is a strictly stationary and* α*- mixing sequence with* $\alpha(m) = O(m^{-A})$ *for some* $A > (q-1)(1+q/2)$*, where* $X_t = (X_{1t}, \ldots, X_{Kt})^\top$ *and* q *is an even integer that satisfies* $q > 2(K+1)$*. (ii) The supports of* X_{kt} *and* Y_t *are compact* $\forall k = 1, \ldots, K$*. (iii) The distributions of* X_t *and* Y_t *are absolutely continuous with respect to Lebesgue measure and have bounded densities.*

Assumption 2. (i) $\{\epsilon_T : T \geq 1\}$ *is a sequence of positive constants such that* $\lim_{T\to\infty} \epsilon_T = 0$ *and* $\epsilon_T > k_T \; \forall T \geq 1$. (ii) *For each* $x \in \mathcal{X}$*, constant* $C_1 > 0$ *and* $\lambda \in \Lambda_0$ *such that* $A_\lambda^= \neq \varnothing$*, we have:*

$$|G_Y(x) - G_\lambda(x)| \geq C_1 \min\left\{ \inf_{x' \in A_\lambda^=} |x - x'|, \epsilon_T \right\}$$

for T *sufficiently large.*

Assumption 2 requires that the function $G_Y(\cdot) - G_\lambda(\cdot)$ is monotonic on a ϵ_T - neighborhood of the boundary $\partial A_\lambda^=$ of $A_\lambda^=$. It is satisfied when $G_Y(x)$ and $G_\lambda(x)$ have derivatives that are not equal on the local neighborhood of $\partial A_\lambda^=$ because by Taylor expansion $G_Y(x) - G_\lambda(x) \simeq [g_Y(x') - g_\lambda(x')][x' - x]$ for x close to x', hence we can bound $|G_Y(x) - G_\lambda(x)|$ from below for x close to $A_\lambda^=$, while for x far from $A_\lambda^=$ the minimum is eventually dominated by ϵ_T which can be made arbitrarily small.

Define the empirical process in λ and x to be

$$\nu_T(\lambda, x) = \sqrt{T}\left[\widehat{G}_Y(x) - \widehat{G}_\lambda(x) - G_Y(x) + G_\lambda(x)\right]. \tag{20}$$

Let $\widetilde{\nu}(\cdot,\cdot)$ be a mean zero Gaussian process on $\Lambda_0 \times \mathcal{X}$ with covariance function given by

$$C((\lambda_1, x_1), (\lambda_2, x_2)) = \lim_{T\to\infty} E\nu_T(\lambda_1, x_1)\nu_T(\lambda_2, x_2). \tag{21}$$

Then, the limiting null distribution of our test statistic is given in the following theorem.

Theorem 1. *Suppose Assumptions 1 and 2 hold. Then, under the null hypothesis, we have*

$$W_T \Rightarrow \Upsilon = \begin{cases} \sup_{\lambda\in\Lambda_0^=} \inf_{x\in\mathcal{X}} [\widetilde{\nu}(\lambda, x)] & \text{if } \Lambda_0^= \neq \varnothing \\ -\infty & \text{if } \Lambda_0^= = \varnothing \end{cases},$$

where $\Lambda_0^=$ *is defined in (5).*

Theorem 1 shows that the asymptotic null distribution of W_T is non-degenerate when $\Lambda_0^= \neq \varnothing$ and depends on the joint distribution function of $(X_t^\top, Y_t)^\top$. The latter implies that the asymptotic critical values for W_T can not be tabulated once and for all. However, we define below various simulation procedures to estimate them from the data.

6.2 Critical Values

6.2.1 Subsampling

We can use a subsampling method to obtain consistent critical values. The subsampling method has been proposed by Politis and Romano (1994) and works in many cases under very general settings, see, e.g., Politis, Romano, and Wolf (1999). The subsampling is useful in our context because our null hypothesis consists of complicated system of inequalities which is hard to mimic using the standard bootstrap. Furthermore, the subsampling-based test described below has an advantage of being asymptotically similar on the boundary of the null hypothesis, see below and LMW(2004) for details.

The subsampling procedure is based on the following steps:

(i) Calculate the test statistic W_T using the original full sample $\mathcal{W}_T = \{Z_t = (X_t^\top, Y_t)^\top : t = 1, \ldots, T\}$.

(ii) Generate subsamples $\mathcal{W}_{T,b,t} = \{Z_t, \ldots, Z_{t+b-1}\}$ of size b for $t = 1, \ldots, T-b+1$.

(iii) Compute test statistics $W_{T,b,t}$ using the subsamples $\mathcal{W}_{T,b,t}$ for $t = 1, \ldots, T-b+1$.

(iv) Approximate the sampling distribution of W_T by

$$\widehat{S}_{T,b}(w) = \frac{1}{T-b+1} \sum_{t=1}^{T-b+1} 1\left(W_{T,b,t} \leq w\right).$$

(v) Get the α-th sample quantile of $\widehat{S}_{T,b}(\cdot)$, i.e.,

$$s_{T,b}(\alpha) = \inf\{w : \widehat{S}_{T,b}(w) \geq \alpha\}.$$

(vi) Reject the null hypothesis at the significance level α if $W_T > s_{T,b}(\alpha)$.

The above subsampling procedure can be justified in the following sense: Let $b = \widehat{b}_T$ be a data-dependent sequence satisfying

Assumption 3. *$P[l_T \leq \widehat{b}_T \leq u_T] \to 1$ where l_T and u_T are integers satisfying $1 \leq l_T \leq u_T \leq T$, $l_T \to \infty$ and $u_T/T \to 0$ as $T \to \infty$.*

Then, the following theorem shows that our test based on the subsample critical value has asymptotically correct size.

Theorem 2. *Suppose Assumptions 1-3 hold. Then, under the null hypothesis, we have*

$$\text{(a) } s_{T,\widehat{b}_T}(\alpha) \xrightarrow{p} \begin{cases} s(\alpha) & \text{if } \Lambda_0^= \neq \varnothing \\ -\infty & \text{if } \Lambda_0^= = \varnothing \end{cases}$$

$$\text{(b) } P[W_T > s_{T,\widehat{b}_T}(\alpha)] \to \begin{cases} \alpha & \text{if } \Lambda_0^= \neq \varnothing \\ 0 & \text{if } \Lambda_0^= = \varnothing \end{cases}$$

as $T \to \infty$, where $s(\alpha)$ denotes the α-th quantile of the asymptotic null distribution $\sup_{\lambda \in \Lambda_0^=} \inf_{x \in \mathcal{X}} [\widetilde{\nu}(\lambda, x)]$ *of W_T given in Theorem 1.*

6.2.2 Bootstrap

We next define an alternative to our subsampling procedure based on full-sample bootstrap applied to a recentered test statistic.

(i) Calculate the test statistic W_T using the original full sample $\mathcal{W}_T = \{Z_t = (X_t^\top, Y_t)^\top : t = 1, \ldots, T\}$.

(ii) Generate the bootstrap sample $\mathcal{W}_T^* = \{Z_t^* : t = 1, \ldots, T\}$ M -times, where M is the number of the bootstrap samples, see below for various possible ways to draw the bootstrap samples.

(iii) Compute the recentred test statistic W_T^* using the bootstrap sample $\mathcal{W}_T^*$: i.e.,

$$W_T^* = \sup_{\lambda \in \Lambda_0} \inf_{x \in \mathcal{X}} Q_T^*(\lambda, x),$$

where

$$\begin{aligned} Q_T^*(\lambda, x) &= \sqrt{T}\left[\widehat{G}_Y^*(x) - \widehat{G}_\lambda^*(x) - E^*\left(\widehat{G}_Y^*(x) - \widehat{G}_\lambda^*(x)\right)\right], \quad (22) \\ \widehat{G}_\lambda^*(x) &= \int_{-\infty}^{x} \widehat{G}_\lambda^{*(s-1)}(y)dy, \;\; \widehat{F}_\lambda^*(x) = \frac{1}{T}\sum_{t=1}^{T} 1(X_t^{*\top}\lambda \leq x), \end{aligned}$$

and likewise for $\widehat{G}_Y^*(x)$ and $E^*(\cdot)$ denotes the expectation relative to the distribution of the bootstrap sample $\mathcal{W}_T^*$ conditional on the original sample $\mathcal{W}_T$, see below for details.

(iv) Approximate the sampling distribution of W_T by

$$\widehat{H}_T(w) = \frac{1}{M}\sum_{m=1}^{M} 1\left(W_{T,m}^* \leq w\right),$$

where $W_{T,m}^*$ denotes the value of W_T^* computed from the m -th bootstrap sample for $m = 1, \ldots, M$.

(v) Get the α-th sample quantile of $\widehat{H}_T(\cdot)$, i.e.,

$$h_T(\alpha) = \inf\{w : \widehat{H}_T(w) \geq \alpha\}.$$

(vi) Reject the null hypothesis at the significance level α if $W_T > h_T(\alpha)$.

When the data are independent over time, the bootstrap sample can be generated by drawing the vector $Z_t^* = (X_{1t}^*, \ldots, X_{Kt}^*, Y_t^*)^\top$ randomly with replacement from the empirical joint distribution of the vectors $\{Z_t : t = 1, \ldots, T\}$. Drawing Z_t^* as a vector will enable the bootstrap sample to preserve the general mutual dependence among $K+1$ assets that may exist in the original sample. In step (iii) above, the recentering in $Q_T^*(\lambda, x)$ can be done with

$$E^*\left(\widehat{G}_Y^*(x) - \widehat{G}_\lambda^*(x)\right) = \widehat{G}_Y(x) - \widehat{G}_\lambda(x)$$

This recentering crucial and is used to impose the least favorable case of the null restriction, i.e.,

$$G_Y(x) = G_\lambda(x) \; \forall x \in \mathcal{X}, \; \forall \lambda \in \Lambda_0. \tag{23}$$

The idea of recentering has also been suggested in other contexts by Hall and Horowitz (1996), Chernozhukov (2002) and LMW (2004), among others.

In the time series case, the bootstrap procedure should be modified to account for the temporal dependence. We briefly describe the non-overlapping (viz., Carlstein (1986)) and overlapping (viz., Künsch (1989)) block bootstrap procedures. The observations to be bootstrapped are the vectors $\{Z_t : t = 1, \ldots, T\}$ as before. Let L denote the length of the blocks satisfying $L \propto T^\gamma$ for some $0 < \gamma < 1$. With non-overlapping blocks, block 1 is observations $\{Z_j : j = 1, \ldots, L\}$, block 2 is observations $\{Z_{L+j} : j = 1, \ldots, L\}$, and so forth. There are B different blocks, where $BL = T$. With overlapping blocks, block 1 is observations $\{Z_j : j = 1, \ldots, L\}$, block 2 is observations $\{Z_{1+j} : j = 1, \ldots, L\}$, and so forth. There are $T - L + 1$ different blocks. The bootstrap sample $\{Z_t^* : t = 1, \ldots, T\}$ are obtained by sampling B blocks randomly with replacement from either the

B non-overlapping blocks or the $T-L+1$ overlapping blocks and laying them end-to-end in the order sampled. In the case of non-overlapping bootstrap, the recentering (22) may be done with $E^*\left(\widehat{G}_Y^*(x)-\widehat{G}_\lambda^*(x)\right) = \widehat{G}_Y(x)-\widehat{G}_\lambda(x)$ as in the independent sampling case. However, when the overlapping block bootstrap is used, we need to recenter the statistic with

$$E^*\left(\widehat{G}_Y^*(x)-\widehat{G}_\lambda^*(x)\right) = \widehat{G}_{Y,OB}(x)-\widehat{G}_{\lambda,OB}(x), \text{ where}$$

$$\begin{aligned} \widehat{G}_{\lambda,OB}(x) &= \int_{-\infty}^{x} \widehat{G}_{\lambda,OB}^{(s-1)}(y)dy,\ \widehat{F}_{\lambda,OB}(x)=\frac{1}{T}\sum_{t=1}^{T}\omega(t,L,T)1(X_t^\top\lambda\leq x)\ , \\ \omega(t,L,T) &= \begin{cases} t/L & \text{if } t\in[1,L-1] \\ 1 & \text{if } t\in[L,T-L+1] \\ (T-t+1)/L & \text{if } t\in[T-L+2,T]\,. \end{cases}, \end{aligned}$$

and likewise for $\widehat{G}_{Y,OB}(x)$.

We now compare the subsampling and bootstrap procedures. Under suitable regularity conditions, it is not difficult to show that the asymptotic size of the test based on bootstrap critical value $h_T(\alpha)$ is α if the least favorable case (23) is true. Therefore, in this case, we might prefer bootstrap to subsampling since the former uses the full sample information and hence may be more efficient in finite samples. However, as we have argued in other context (see LMW (2004, Section 6.1)), the least favorable case (23) is only a special case of the boundary, i.e., $\Lambda_0^{=}\neq\varnothing$, of the null hypothesis $\mathbf{H}_0$, whereas the test statistic W_T has a non-degenerate limit distribution everywhere on the boundary. This implies that the bootstrap-based test is not asymptotically similar on the boundary, which in turn implies that the test is biased. On the other hand, the subsample-based test is unbiased and asymptotically similar on the boundary and may be preferred in this sense. In practice, one might wish to employ both approaches to see if the results obtained are robust to the choice of resampling schemes, as we did in our empirical applications below.

6.3 Asymptotic Power

In this section, we discuss consistency and local power properties of our test.

If the alternative hypothesis is true, $\Lambda_0=\Lambda_0^{+}\cup\Lambda_0^{\simeq}$. When Λ_0^{+} is empty, we need the following assumption for consistency of our test:

Assumption 4. *When* $\Lambda_0=\Lambda_0^{\simeq}$, $\underline{\lim}_{T\to\infty}(T/u_T)^{1/2}\Delta_\lambda(\epsilon_T)>0$ *for some* $\lambda\in\Lambda_0$, *where* $\Delta_\lambda(\epsilon)=\inf_{x\in B_\lambda^\epsilon}(G_Y(x)-G_\lambda(x))$ *and* u_T *is defined in Assumption 3.*

For each $\lambda\in\Lambda_0^{\simeq},\Delta_\lambda(\epsilon)$ is a non-decreasing in ϵ, $\Delta_\lambda(\epsilon)>0\ \forall\epsilon>0$ and $\Delta_\lambda(0)=0$. Therefore, from a Taylor expansion $(T/u_T)^{1/2}\Delta_\lambda(\epsilon_T)\simeq(T/u_T)^{1/2}\epsilon_T(\partial\Delta_\lambda(0)/\partial\epsilon)$, Assumption 4 holds if ϵ_T

goes to zero at a rate not too fast and and the derivative of $\Delta_\lambda(\epsilon)$ is strictly positive at $\epsilon = 0$ for some $\lambda \in \Lambda_0^{\approx}$.

Then, we have:

Theorem 3. *Suppose that Assumptions 1-4 hold. Then, under the alternative hypothesis, we have*

$$P[W_T > s_{T,\widehat{b}_T}(\alpha)] \to 1 \text{ as } T \to \infty.$$

Next, we determine the power of the test W_T against a sequence of contiguous alternatives converging to the boundary $\Lambda_0^= \neq \varnothing$ of the null hypothesis at the rate $1/\sqrt{T}$. That is, consider the set of portfolio weights

$$\Lambda_{0T} = \left\{ \lambda + c/\sqrt{T} : \lambda \in \Lambda_0^=, \; c \in \mathbb{R}^K \right\}.$$

Let $F_{\lambda_T}(\cdot) = G_{\lambda_T}^{(0)}(x)$ be the c.d.f.'s of $X_t^\top \lambda_T$ for $\lambda_T \in \Lambda_{0T}$. Also, for $s \geq 1$, define

$$G_{\lambda_T}^{(s)}(x) = \int_{-\infty}^{x} G_{T\lambda_T}^{(s-1)}(y)dy.$$

As before, we abbreviate the superscript s for notational simplicity. Then, we assume that the functionals $G_{\lambda_T}(x)$ and $G_Y(x)$ satisfy the following local alternative hypothesis:

$$\mathbf{H}_a : G_Y(x) - G_{\lambda_T}(x) = \frac{\delta_{Y\lambda}(x)}{\sqrt{T}} \text{ for } \lambda_T \in \Lambda_{0T} \text{ and } \lambda \in \Lambda_0^=, \tag{24}$$

where $\delta_{Y\lambda}(\cdot)$ is a real function such that $\inf_{x \in \mathcal{X}}[\delta_{Y\lambda}(x)] > 0$.

The asymptotic distribution of W_T under the local alternatives is given in the following theorem:

Theorem 4. *Suppose Assumptions 1 and 2 (with Λ_0 replaced by Λ_{0T}) hold. Then, under the sequence of local alternatives $\mathbf{H}_a$, we have*

$$W_T \Rightarrow \sup_{\lambda \in \Lambda_0^=} \inf_{x \in \mathcal{X}} \left[\widetilde{\nu}(\lambda, x) + \delta_{Y\lambda}(x) \right],$$

where $\widetilde{\nu}(\lambda, x)$ is defined as in Theorem 1.

The result of Theorem 4 implies that asymptotic local power of our test based on the subsample critical value is given by

$$\lim_{T \to \infty} P[W_T > s_{T,\widehat{b}_T}(\alpha)] = P\left[L_0 > s(\alpha)\right], \tag{25}$$

where L_0 denotes the limit distribution given in Theorem 4 and $s(\alpha)$ denotes the α-th quantile of the asymptotic null distribution of W_T given in Theorem 1. Also, our test is aymptotically local unbiased because, by Anderson's lemma (see Bickel et. al. (1993, p.466)), the right hand side of (25) is less than α.

7 Numerical Results

The existing SSD efficiency test suffers from low power in typical empirical applications, as demonstrated in the simulation experiment of Post (2003, Section IIIC) based on the returns of the well-known 25 double-sorted Fama and French stock portfolios formed on market capitalization and book-to-market-equity ratio. In part, the lack of power reflects the difficulty of estimating a 25-dimensional return distribution. It is likely that the power increases (at an increasing rate) as the length of the cross-section is reduced to for example ten benchmark portfolios, which is a common choice in asset pricing tests.

Instead of the 25 Fama and French portfolios, this study uses ten single-sorted portfolios formed on market beta. We focus on these portfolios for two reasons. First, sorting stocks on beta maximizes the spread in betas and hence minimizes the probability of erroneous rejection of the null of mean-variance efficiency (Type I error). Second, time-variation of the return distribution can severely bias the results of unconditional asset pricing tests (see for instance Jagannathan and Wang (1996)). Hence, the large sample properties of our tests apply only to benchmark portfolios for which long, homogenous samples are available in practice. Unfortunately, the 25 Fama and French portfolios seem severely affected by time-variation. By contrast, beta portfolios by construction have a more stable distribution, as a stock migrates to another benchmark portfolio if its beta changes significantly through time.

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

[Insert Table I about here]

We will first apply our procedures to two test portfolios in random samples drawn from the multivariate normal population distribution. The equal weighted portfolio (EP) is known to be SSD, TSD and mean-variance inefficient relative to the normal population distribution (in this case with normal distributions). 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 lies in the null hypothesis.

7.1 Preliminary Results

In the first part of our simulations, we draw random samples from the multivariate normal population distribution with moments taken from the beta-sorted portfolios. For every random sample, we apply our test procedures for second order and third order stochastic dominance to both test portfolios. This experiment is performed for a sample size $T \in \{50, 100, 200, 500, 1000, 2000\}$. Below we show some preliminary results for the special case of two portfolios (numbers 2 and 9 in terms of β) in which case we just perform a grid search over 100 linear combinations of these assets. We take $k_T = 0.3\sqrt{\log(T)/T}$ and $\epsilon_T = 2 \times k_T$. These results are based on $ns = 400$ replications. We show the median p-value across simulations against sample size. The p-values are computed by comparing the test statistic with 200 recentered bootstrap resamples. Recall that the equally weighted portfolio (EP) is inefficient according to second order and third order dominance, while the tangency portfolio (TP) is efficient. The results are shown in Figures 2-5 below.

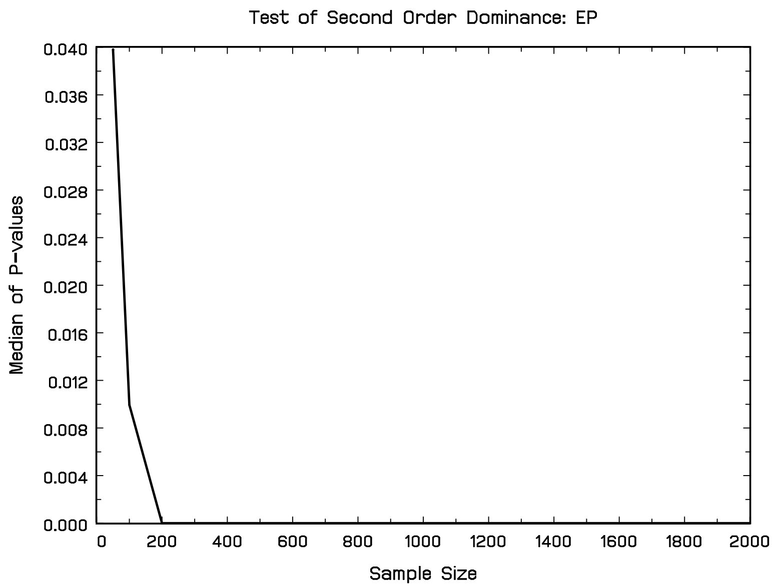

Figure 2. Alternative hypothesis

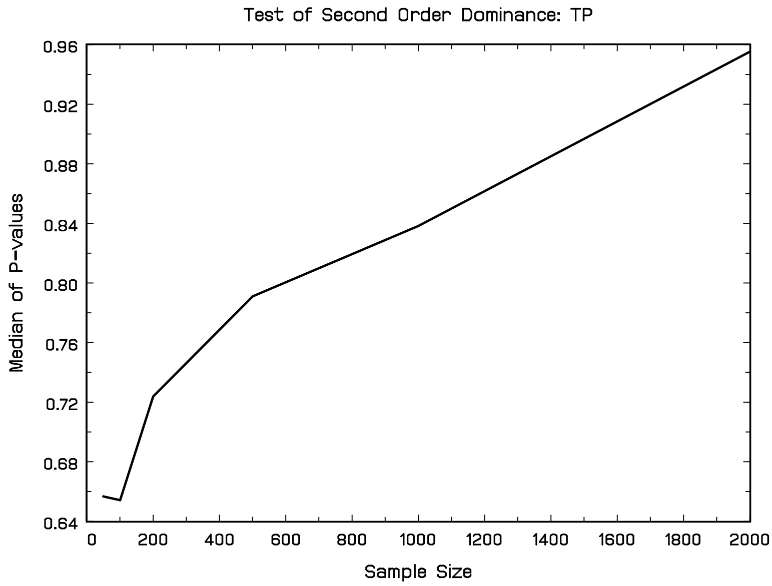

Figure 3. Null hypothesis

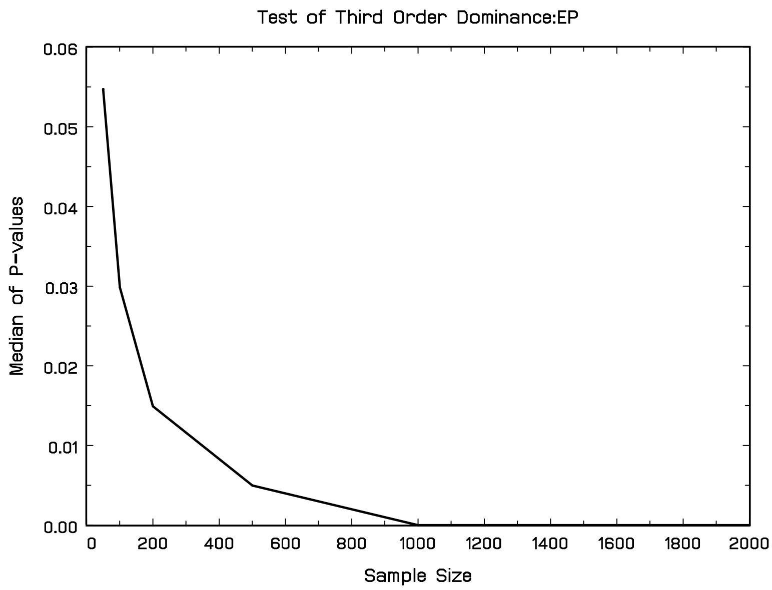

Figure 4. Alternative hypothesis

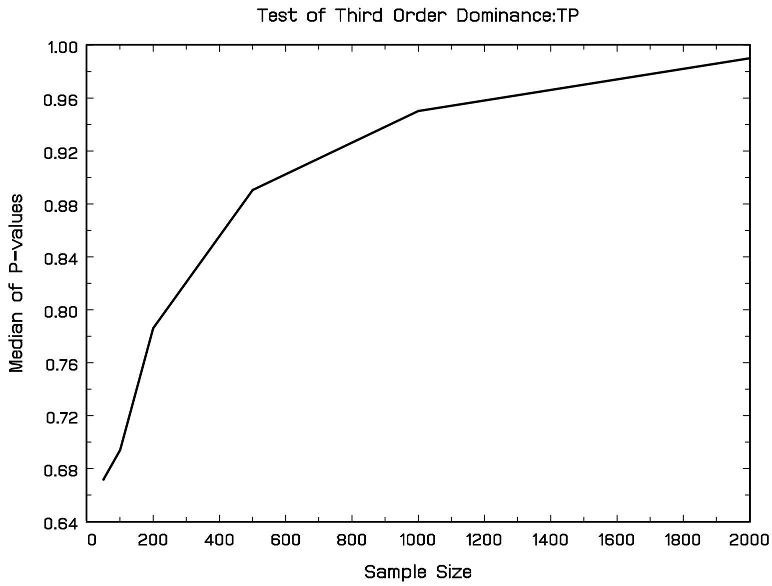

Figure 5. Null hypothesis

These results seem to be encouraging: under the null hypothesis median p-values tend to one and under the alternative hypothesis median p-values tend to zero with sample size.

8 Appendix

Lemma 1. *Suppose Assumption 1 holds, Then, we have*

$$\nu_T(\cdot,\cdot) \Rightarrow \widetilde{\nu}(\cdot,\cdot). \tag{26}$$

Proof of Lemma 1. For lemma 1, we need to verify (i) finite dimensional (fidi) convergence and (ii) the stochastic equicontinuity result: that is, for each $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\overline{\lim_{T\to\infty}} \left\| \sup_{\rho^*((\lambda_1,x_1),(\lambda_2,x_2))<\delta} |\nu_T(\lambda_1,x_1) - \nu_T(\lambda_2,x_2)| \right\|_q < \varepsilon, \tag{27}$$

where the pseudo-metric on $\Lambda_0 \times \mathcal{X}$ is given by

$$\begin{aligned} & \rho^*\left((\lambda_1,x_1),(\lambda_2,x_2)\right) \\ = \; & \Big\{E\Big[(x_1-Y_t)^{s-1}1\left(Y_t\le x_1\right)-(x_1-X_t^\top\lambda_1)^{s-1}1\left(X_t^\top\lambda_1\le x_1\right) \\ & -(x_2-Y_t)^{s-1}1\left(Y_t\le x_2\right)+(x_2-X_t^\top\lambda_2)^{s-1}1\left(X_t^\top\lambda_2\le x_2\right)\Big]^2\Big\}^{1/2}. \end{aligned}$$

The fidi convergence result holds by the Cramer-Wold device and a CLT for bounded random variables (see Hall and Heyde (1980, Corollary 5.1)) since the underlying random sequence $\{(X_t^\top, Y_t)^\top : t \ge 1\}$

is strictly stationary and α - mixing with $\sum_{m=1}^{\infty} \alpha(m) < \infty$ by Assumption 1. On the other hand, the stochastic equicontinuity condition (27) holds by Theorem 2.2 of Andrews and Pollard (1994) with $Q = q$ and $\gamma = 2$. To see this, note that their mixing condition is implied by Assumption 1(i). Also, let

$$\begin{aligned} \mathcal{F} &= \{f_t(\lambda, x) : (\lambda, x) \in \Lambda_0 \times \mathcal{X}\}, \text{where} \\ f_t(\lambda, x) &= (x - Y_t)^{s-1} 1 (Y_t \leq x) - (x - X_t^\top \lambda)^{s-1} 1 \left(X_t^\top \tau \leq x\right) .. \end{aligned}$$

Then, $\mathcal{F}$ is a class of uniformly bounded functions that satisfy the L^2-continuity condition: that is, for some constants $C_1, C_2 < \infty$,

$$\begin{aligned} & E \overset{*}{\sup} \left[f_t(\lambda_1, x_1) - f_t(\lambda, x)\right]^2 \\ & \leq C_1 \left\{ E \overset{*}{\sup} \left[(x_1 - Y_t)^{s-1} - (x - Y_t)^{s-1}\right]^2 + E \overset{*}{\sup} \left[1 (Y_t \leq x_1) - 1 (Y_t \leq x)\right]^2 \right. \\ & \left. + E \overset{*}{\sup} \left[(x_1 - X_t^\top \lambda_1)^{s-1} - (x - X_t^\top \lambda)^{s-1}\right]^2 + E \overset{*}{\sup} \left[1 \left(X_t^\top \lambda_1 \leq x_1\right) - 1 \left(X_t^\top \lambda \leq x\right)\right]^2 \right\} \\ & \leq C_2 \cdot r, \end{aligned}$$

where $\sup^*$ denotes the supremum taken over $(\lambda_1, x_1) \in \Lambda_0 \times \mathcal{X}$ for which $\|\lambda_1 - \lambda\| \leq r_1$, $|x_1 - x| \leq r_2$ and $\sqrt{r_1^2 + r_2^2} \leq r$, the first inequality holds by several applications of Cauchy-Schwarz inequality and Assumption 1(ii) and the second inequality holds by Assumptions 1(iii). This implies that the bracketing condition of Andrews and Pollard (1994, p.121) holds because the L^2-continuity condition implies that the bracketing number satisfies $N(\varepsilon, \mathcal{F}) \leq C_3 \cdot (1/\varepsilon)^{K+1}$ .This establishes Lemma 1. ■

Lemma 2. *Suppose Assumptions 1 and 2 hold. Then, we have*

$$P\left(B_\lambda^{2\epsilon_T} \subset \widehat{B}_\lambda^{\epsilon_T} \subset B_\lambda^{\epsilon_T}\right) \to 1 \ \forall \lambda \in \Lambda_0$$

as $T \to \infty$.

Proof of Lemma 2. It suffices to show that for each $\lambda \in \Lambda_0$,

$$P\left((A_\lambda^=)^{\epsilon_T} \subset \left(\widehat{A}_\lambda^=\right)^{\epsilon_T}\right) \;\to\; 1 \tag{28}$$

$$P\left(\left(\widehat{A}_\lambda^=\right)^{\epsilon_T} \subset (A_\lambda^=)^{2\epsilon_T}\right) \;\to\; 1. \tag{29}$$

Suppose $A_\lambda^= \neq \mathcal{X}$. (If $A_\lambda^= = \mathcal{X}$, (29) trivially holds and (28) holds by the same argument as (30) below.) We first establish (28). Consider λ such that $A_\lambda^= \neq \varnothing$. (Otherwise, (28) holds trivially.) Let $x_0^* \in (A_\lambda^=)^{\epsilon_T}$. Then, $x_0^* = x_0 + \eta_{0T}$ for some $x_0 \in A_\lambda^=$ and a fixed sequence $|\eta_{0T}| < \epsilon_T$. Now (28) holds since

$$\begin{aligned} P\left(x_0^* \in \left(\widehat{A}_\lambda^=\right)^{\epsilon_T}\right) &\geq P(x_0 \in \widehat{A}_\lambda^=) \\ &= P\left(\left|\widehat{G}_\lambda(x_0) - \widehat{G}_Y(x_0) - G_\lambda(x_0) + G_Y(x_0)\right| \leq k_T\right) \\ &= P\left(|O_p(1)| \leq (\log T)^{1/2}\right) \to 1, \end{aligned} \tag{30}$$

where the second equality holds by the fidi convergence result of Lemma 1.

We next establish (29). Let $x_1^* \in \left(\widehat{A}_\lambda^=\right)^{\epsilon_T}$, i.e., $x_1^* = x_1 + \eta_{1T}$ for some $x_1 \in \widehat{A}_\lambda^=$ and fixed sequence $|\eta_{1T}| < \epsilon_T$. It suffices to show that $P\left(x_1 \in (A_\lambda^=)^{\epsilon_T}\right) \to 1$. Let $C_1 > 1$ be a constant. Then, we have: $wp \to 1$,

$$\begin{aligned} & |G_Y(x_1) - G_\lambda(x_1)| \\ \leq & \left|\widehat{G}_Y(x_1) - G_Y(x_1)\right| + \left|\widehat{G}_\lambda(x_1) - G_\lambda(x_1)\right| + \left|\widehat{G}_Y(x_1) - \widehat{G}_\lambda(x_1)\right| \\ \leq & C_1 k_T, \end{aligned}$$

where the first inequality holds by triangular inequality and the second inequality holds using the fidi convergence result as in (30) and the fact that $x_1 \in \widehat{A}_\lambda^=$. Now, by Assumption 2, since $\epsilon_T > k_T$, we have:

$$\inf_{x' \in A_\lambda^=} |x_1 - x'| < \epsilon_T \quad wp \to 1,$$

which implies that $P\left(x_1 \in (A_\lambda^=)^{\epsilon_T}\right) \to 1$, as required. ■

Proof of Theorem 1. Below, we shall establish

$$\sup_{\lambda \in \Lambda_0^=} \inf_{x \in \widehat{B}_\lambda^{\varepsilon_T}} Q_T(\lambda, x) \Rightarrow \Upsilon \tag{31}$$

$$\sup_{\lambda \in \Lambda_0^-} \inf_{x \in B_\lambda^{2\varepsilon_T}} \nu_T(\lambda, x) - \sup_{\lambda \in \Lambda_0^-} \inf_{x \in B_\lambda^0} \nu_T(\lambda, x) = o_p(1). \tag{32}$$

Then, Theorem 1 holds because of the following arguments: For any $w \in \mathbb{R}$, we have

$$\begin{aligned} & \overline{\lim_{T\to\infty}} \left| P\left(W_T \leq w\right) - P\left(\sup_{\lambda \in \Lambda_0^=} \inf_{x \in \widehat{B}_\lambda^{\varepsilon_T}} Q_T(\lambda, x) \leq w\right)\right| \\ \leq & \overline{\lim_{T\to\infty}} P\left(\sup_{\lambda \in \Lambda_0^-} \inf_{x \in \widehat{B}_\lambda^{\varepsilon_T}} Q_T(\lambda, x) > w\right) & (33) \\ \leq & \overline{\lim_{T\to\infty}} P\left(\sup_{\lambda \in \Lambda_0^-} \inf_{x \in B_\lambda^{2\varepsilon_T}} Q_T(\lambda, x) > w\right) & (34) \\ \leq & \overline{\lim_{T\to\infty}} P\left(\sup_{\lambda \in \Lambda_0^-} \inf_{x \in B_\lambda^{2\varepsilon_T}} \nu_T(\lambda, x) > w + T^{1/4}\right) & (35) \\ = & \overline{\lim_{T\to\infty}} P\left(\sup_{\lambda \in \Lambda_0^-} \inf_{x \in B_\lambda^0} \nu_T(\lambda, x) > w + T^{1/4}\right) & (36) \\ = & 0, & (37) \end{aligned}$$

where (33) holds by the fact that $\Lambda_0 = \Lambda_0^- \cup \Lambda_0^=$ under the null hypothesis and the general inequality $|P(\max(X,Y) \leq x) - P(Y \leq x)| \leq P(X > x)$ for any rv's X and Y, (34) holds by Lemma 2, (35) follows from the result $\overline{\lim}_{T\to\infty} \sup_{\lambda \in \Lambda_0^-} \inf_{x \in B_\lambda^{2\varepsilon_T}} T^{1/4}\left(G_Y(x) - G_\lambda(x)\right) < -1$, (36) holds by (32), and

(37) holds since $\sup_{\lambda\in\Lambda_0^-}\inf_{x\in B_\lambda^0}\nu_T(\lambda,x)=O_p(1)$ using Lemma 1 and continuous mapping theorem. This result and (31) combine to yield Theorem 1.

We now establish (31) and (32). Let $w\in\mathbb{R}$. Then, by Lemma 2, we have

$$\begin{aligned}&\left|P\left(\sup_{\lambda\in\Lambda_0^=}\inf_{x\in\widehat{B}_\lambda^{\varepsilon_T}}Q_T(\lambda,x)\le w\right)-P\left(\sup_{\lambda\in\Lambda_0^=}\inf_{x\in\mathcal{X}}Q_T(\lambda,x)\le w\right)\right|\\ \le\ & P\left(\widehat{B}_\lambda^{\epsilon_T}\neq\mathcal{X}\text{ for }\lambda\in\Lambda_0^=\right)\to 0.\end{aligned}$$

Therefore, (31) holds by Lemma 1, continuous mapping theorem and the fact

$$\sup_{\lambda\in\Lambda_0^=}\inf_{x\in\mathcal{X}}Q_T(\lambda,x)=\sup_{\lambda\in\Lambda_0^=}\inf_{x\in\mathcal{X}}\left[\nu_T(\lambda,x)\right].$$

Next, consider (32). Let $\mathcal{Z}\subset\mathbb{R}$ be a compact set containing zero. Define the stochastic process $l_T(\cdot,\cdot,\cdot)$ on $\Lambda_0^-\times\mathcal{X}\times\mathcal{Z}$ to be $l_T(\lambda,x,z)=\nu_T(\lambda,x+z)$. Then, by an argument similar to Lemma 1, $l_T(\cdot,\cdot,\cdot)$ is stochastic equicontinuous on $\Lambda_0^-\times\mathcal{X}\times\mathcal{Z}$, which in turn implies that

$$\begin{aligned}&\sup_{\lambda\in\Lambda_0^-}\inf_{x\in B_\lambda^{2\varepsilon_T}}\nu_T(\lambda,x)-\sup_{\lambda\in\Lambda_0^-}\inf_{x\in B_\lambda^0}\nu_T(\lambda,x)\\ =\ &\sup_{\lambda\in\Lambda_0^-}\inf_{x\in B_\lambda^0,\ |z|\le 2\epsilon_T}l_T(\lambda,x,z)-\sup_{\lambda\in\Lambda_0^-}\inf_{x\in B_\lambda^0}l_T(\lambda,x,0)\\ =\ &o_p(1),\ \text{ as required.}\end{aligned}$$

This now completes the proof of Theorem 1. ■

Proof of Theorem 2. The proof is similar to the proof of Theorem 2 of LMW(2004), see also Politis et. al (1999, Theorem 3.5.1). ■

Proof of Theorem 3. Under the alternative hypothesis, $\Lambda_0=\Lambda_0^+\cup\Lambda_0^{\simeq}$. Let

$$\begin{aligned}\widehat{S}_{T,b}^0(w)&=\frac{1}{T-b+1}\sum_{t=1}^{T-b+1}1\left(b^{-1/2}W_{T,b,t}\le w\right)\\ S_b^0(w)&=P\left(b^{-1/2}W_{T,b,1}\le w\right).\end{aligned}$$

Using the inequality of Bosq (1998, Theorem 1.3) and Assumption 3 (see also LMW (2004, proof of Theorem 2)), we can establish the uniform convergence result:

$$\sup\left|\widehat{S}_{T,b}^0(w)-S_b^0(w)\right|\xrightarrow{p}0. \tag{38}$$

Therefore, (38) and the pointwise convergence result $b^{-1/2}W_{T,b,1}\xrightarrow{p}d_*(0)$ yield:

$$s_{T,\widehat{b}_T}^0(\alpha)=\inf\{w:\widehat{S}_{T,b}^0(w)\ge\alpha\}\to d_*(0)\ge 0, \tag{39}$$

where $d_*(\cdot)$ is defined in (3). Note that $d_*(0)$ is strictly positive if $\Lambda_0^+ \neq \varnothing$, while $d_*(0) = 0$ if $\Lambda_0^+ = \varnothing$. Therefore,

$$\begin{aligned}
& P\left(W_T > s_{T,\widehat{b}_T}(\alpha)\right) \\
\geq\ & P\left(\sup_{\lambda \in \Lambda_0^+ \cup \Lambda_0^{\approx}} \inf_{x \in B_\lambda^{\varepsilon_T}} \left[\nu_T(\lambda, x) + T^{1/2}\left(G_Y(x) - G_\lambda(x)\right)\right] > \widehat{b}_T^{1/2} s^0_{T,\widehat{b}_T}(\alpha)\right) + o(1) \\
\geq\ & P\left(\sup_{\lambda \in \Lambda_0^+ \cup \Lambda_0^{\approx}} \inf_{x \in B_\lambda^{\varepsilon_T}} \left[\nu_T(\lambda, x) + T^{1/2}\left(G_Y(x) - G_\lambda(x)\right)\right] > u_T^{1/2} s^0_{T,\widehat{b}_T}(\alpha)\right) + o(1) \\
\geq\ & P\left(\sup_{\lambda \in \Lambda_0^+ \cup \Lambda_0^{\approx}} \inf_{x \in B_\lambda^{\varepsilon_T}} \left(\frac{T}{u_T}\right)^{1/2} \left[T^{-1/2}\nu_T(\lambda, x) + \left(G_Y(x) - G_\lambda(x)\right)\right] > d_*(0)\right) + o(1) \qquad (40)
\end{aligned}$$

where the first inequality holds by Lemma 2 and the second inequality holds by Assumption 3, and the last inequality holds by (39). Now consider the right hand side of (40). Note that

$$T^{-1/2}\nu_T(\lambda, x) \xrightarrow{p} 0 \qquad (41)$$

by Lemma 1. Also,

$$\underline{\lim}_{T\to\infty} \sup_{\lambda \in \Lambda_0^+ \cup \Lambda_0^{\approx}} \left(T/u_T\right)^{1/2} \Delta_\lambda(\epsilon_T) > d_*(0) \qquad (42)$$

because, if $\Lambda_0^+ \neq \varnothing$, $\lim_{T\to\infty} \Delta_\lambda(\epsilon_T) = d_*(0) > 0\ \forall \lambda \in \Lambda_0^+$ and $\underline{\lim}_{T\to\infty} \left(T/u_T\right)^{1/2} > 1$ by Assumption 4 and, if $\Lambda_0^+ = \varnothing$, $\underline{\lim}_{T\to\infty} \sup_{\lambda\in\Lambda_0^{\approx}} \left(T/u_T\right)^{1/2} \Delta_\lambda(\epsilon_T) > 0 = d_*(0)$ by Assumption 4. Therefore, (40), (41), and (42) imply that

$$P\left(W_T > s_{T,\widehat{b}_T}(\alpha)\right) \to 1,$$

as required. ■

Proof of Theorem 4. Define the empirical process in $(\lambda, z, x) \in \Lambda_0^= \times \mathcal{Z} \times \mathcal{X}$ to be:

$$\nu_T^*(\lambda, z, x) = \sqrt{T}\left[\widehat{G}_Y(x) - \widehat{G}_{T,\lambda+z}(x) - G_Y(x) + G_{\lambda+z}(x)\right],$$

where $\mathcal{Z}$ is a compact set containing zero and $G_Y(x) - G_{\lambda_T}(x) = G_Y(x) - G_{\lambda+c/\sqrt{T}}(x)$ satisfies the local alternative hypothesis (24). Similarly to Lemma 1, we can show that the stochastic process $\{\nu_T^*(\cdot,\cdot,\cdot) : T \geq 1\}$ is stochastically equicontinuous on $\Lambda_0^= \times \mathcal{Z} \times \mathcal{X}$. Therefore, since

$$Q_T(\lambda_T, x) = \nu_T^*(\lambda, c/\sqrt{T}, x) + \delta_{Y\lambda}(x), \qquad (43)$$

we have,

$$\begin{aligned}
& \sup_{\lambda_T \in \Lambda_{0T}} \inf_{x \in \widehat{B}_\lambda^{\varepsilon_T}} Q_T(\lambda_T, x) - \sup_{\lambda \in \Lambda_0^=} \inf_{x \in \mathcal{X}} \left[\nu_T^*(\lambda, 0, x) + \delta_{Y\lambda}(x)\right] \\
=\ & \sup_{\lambda \in \Lambda_0^=, c/\sqrt{T} \in \mathcal{Z}} \inf_{x \in \mathcal{X}} \left[\nu_T^*(\lambda, c/\sqrt{T}, x) + \delta_{Y\lambda}(x)\right] - \sup_{\lambda \in \Lambda_0^=} \inf_{x \in \mathcal{X}} \left[\nu_T^*(\lambda, 0, x) + \delta_{Y\lambda}(x)\right] \qquad (44) \\
=\ & o_p(1), \qquad (45)
\end{aligned}$$

where (44) holds $wp \to 1$ since $P\left(\widehat{B}_{\lambda}^{\epsilon_T} = \mathcal{X}\right) \to 1$ for $\lambda \in A_0^{=}$ by Lemma 2 and (45) holds by the stochastic equicontinuity of $\{\nu_T^*(\cdot,\cdot,\cdot) : T \geq 1\}$. Now, the result of Theorem 4 holds by the weak convergence of $\nu_T^*(\cdot, 0, \cdot) + \delta_{Y\lambda}(\cdot)$ to $\widetilde{\nu}(\cdot,\cdot) + \delta_{Y\lambda}(\cdot)$ and continuous mapping theorem. ■

References

[1] Bosq, D. (1998), *Nonparametric statistics for stochastic processes: Estimation and prediction*, Lecture Note in Statistics 110, 2nd edition, Springer, New York.

[2] Davidson R. and J-Y. Duclos (2000), "Statistical inference for stochastic dominance and for the measurement of poverty and inequality", *Econometrica* 68, 1435-1464.

[3] Hall, P. and Heyde, C. C. (1980), *Martingale limit theory and its applications*, Academic Press, New York.

[4] Jagannathan, R., and Z. Wang, (1996), "The conditional CAPM and the cross-section of expected returns," *Journal of Finance* 51, 3-53.

[5] Kaur, A., B. L. S. Prakasa Rao, and H. Singh (2004), "Testing for second order dominance of two distribuions," *Econometric Theory* 10, 849-866.

[6] Klecan, L., R. McFadden, and D. McFadden (1991), "A robust test for stochastic dominance," Working paper, Dept. of Economics, MIT.

[7] Kuosmanen, T., 2004, 'Efficient Diversification According to Stochastic Dominance Criteria', Management Science 50, 1390-1406.

[8] Linton, O., E. Maasoumi, and Y. -J. Whang (2004): "Consistent Testing for Stochastic Dominance under General Sampling Schemes," Forthcoming in *Review of Economic Studies.*

[9] McFadden, D. (1989), "Testing for stochastic dominance," in Part II of T. Fomby and T.K. Seo (eds.) *Studies in the Economics of Uncertainty* (in honor of J. Hadar), Springer-Verlag.

[10] Pakes, A., and D. Pollard (1989): "Simulation and the asymptotics of optimization estimators," *Econometrica* 57, 1027-1057.

[11] Post, T., (2003): "Empirical tests for stochastic dominance efficiency," *Journal of Finance* 58, 1905-1932.

[12] Post, T., (2005): "Wanted: A test for FSD optimality of a given portfolio," http://ssrn.com/abstract=720882

[13] Post, T., and P. Versijp (2004): "GMM tests for SD efficiency of a given portfolio," manuscript,EUR.

[14] Russell, W.R. and T.K. Seo, 1989, Representative sets for stochastic dominance rules, in: T.B. Fomby and T.K. Seo (eds.), *Studies in the economics of uncertainty: In honor of Josef Hadar*, New York: Springer Verlag, 59-76.

[15] Sion, M., 1958, 'On general minimax theorems', Pacific Journal of Mathematics 8, 171-176.

9 Tables and Figures

Table I. Descriptive Statistics Benchmark Portfolios

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

Panel A : The 10 benchmark portfolios

	Mean	Stdev.	Skewness	Kurtosis	Min	Max
1	0.670	3.822	-0.754	5.230	-24.577	15.718
2	0.698	4.015	-0.018	3.926	-20.573	24.222
3	0.756	4.631	0.648	10.175	-25.003	41.292
4	0.659	4.832	0.255	6.269	-25.943	34.332
5	0.918	5.669	1.041	13.370	-29.333	55.762
6	0.833	6.094	0.592	8.279	-28.615	48.932
7	0.809	6.538	0.574	8.773	-32.573	53.842
8	0.768	7.470	0.774	9.264	-30.395	61.832
9	0.833	8.306	0.689	7.941	-36.583	64.262
10	0.794	9.653	0.814	8.516	-37.133	83.692

Panel B : The CRSP market index and the test portfolios

	Mean	Stdev.	Skew	Kurtosis	Min	Max
Mkt	0.714	4.937	0.156	6.181	-23.673	38.172
EP	0.774	5.699	0.560	9.025	-28.020	47.953
TP	0.960	4.264	-0.022	4.571	-21.870	27.730
LP	0.960	5.721	1.361	9.156	-15.139	53.311

Publications in the Report Series Research* in Management

ERIM Research Program: "Finance and Accounting"

2005

Royal Ahold: A Failure Of Corporate Governance
Abe De Jong, Douglas V. Dejong, Gerard Mertens en Peter Roosenboom
ERS-2005-002-F&A
http://hdl.handle.net/1765/1863

Capital Structure Policies in Europe: Survey Evidence
Dirk Brounen, Abe de Jong and Kees Koedijk
ERS-2005-005-F&A
http://hdl.handle.net/1765/1923

A Comparison of Single Factor Markov-Functional and Multi Factor Market Models
Raoul Pietersz, Antoon A. J. Pelsser
ERS-2005-008-F&A
http://hdl.handle.net/1765/1930

Efficient Rank Reduction of Correlation Matrices
Igor Grubišić and Raoul Pietersz
ERS-2005-009-F&A
http://hdl.handle.net/1765/1933

Generic Market Models
Raoul Pietersz and Marcel van Regenmortel
ERS-2005-010-F&A
http://hdl.handle.net/1765/1907

The price of power: valuing the controlling position of owner-managers in french ipo firms
Peter Roosenboom and Willem Schramade
ERS-2005-011-F&A
http://hdl.handle.net/1765/1921

The Success of Stock Selection Strategies in Emerging Markets: Is it Risk or Behavioral Bias?
Jaap van der Hart, Gerben de Zwart and Dick van Dijk
ERS-2005-012-F&A
http://hdl.handle.net/1765/1922

Sustainable Rangeland Management Using a Multi-Fuzzy Model: How to Deal with Heterogeneous Experts' Knowledge
Hossein Azadi, Mansour Shahvali, Jan van den Berg and Nezamodin Faghih
ERS-2005-016-F&A
http://hdl.handle.net/1765/1934

A Test for Mean-Variance Efficiency of a given Portfolio under Restrictions
Thierry Post
ERS-2005-016-F&A

Testing for Stochastic Dominance Efficiency
Oliver Linton, Thierry Post and Yoon-Jae Whang
ERS-2005-033-F&A

Wanted: A Test for FSD Optimality of a Given Portfolio
Thierry Post
ERS-2005-034-F&A

How Domestic is the Fama and French Three-Factor Model? An Application to the Euro Area
Gerard A. Moerman
ERS-2005-035-F&A
http://hdl.handle.net/1765/6626

* 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