cemmap

centre for microdata methods and practice

Testing for the stochastic dominance efficiency of a given portfolio

Oliver Linton
Yoon-Jae Whang

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP27/12

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

Testing for the Stochastic Dominance Efficiency of a given Portfolio*

Oliver Linton† Yoon-Jae Whang‡

University of Cambridge Seoul National University

First Version, December 2005, This Version, August 2012

Abstract

We propose a new statistical 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

The portfolio choice problem is a cornerstone of finance. There are two main approaches to this, the mean variance approach and the stochastic dominance approach. In the mean variance approach strong assumptions are made about the distribution of returns and/or preferences of the investor. The rules for practical computation and statistical inference are well established, see for example Markowitz (1952) and Gibbons, Ross, and Shanken (1989). The stochastic dominance approach makes much weaker assumptions about the distribution of returns and/or preferences. On the other hand, the practical implications of SD analysis has proven to be more difficult. The portfolio problem is especially difficult, because we have to consider infinitely many portfolios, while the standard SD rules rely on pairwise comparison of the individual alternatives. Recently, there has been significant

*We thank participants at the conference in honor of Haim Levy held at CEMMAP, November 4-5, 2005 for comments. Whang thanks the National Research Foundation of Korea grant funded by the Korean Goverment (NRF-2011-342-B00004).

†Faculty of Economics, University of Cambridge, UK. E-mail: `obl20@cam.ac.uk`. Thanks to the ERC for financial support.

‡Department of Economics, Seoul National University, Seoul 151-746, Korea. Email Address : `whang@snu.ac.kr`.

progress on computational and statistical issues that have advanced the position of the stochastic dominance method. See Levy (2006) for an overview and bibliography.

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. Post (2003) and Post and Versijp (2007) 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. They also used a sampling scheme that assumed iid observations and hence does not allow for the GARCH effects often seen in high frequency returns.

We propose an alternative 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 (2005) (hereafter LMW) have provided a comprehensive theory of inference for a class of test statistics for the standard pairwise comparison of prospects. We extend their work to the portfolio case. This entails a nontrivial conceptual and computational issue. The null hypothesis in LMW was of stochastic maximality in a finite set, i.e., that there was at least one prospect that weakly stochastically dominated some of the others. The alternative was two-sided and the number of prospects considered was finite. Because this only involved pairwise comparison it is not appropriate for the situation where an investor may combine a set of basis assets into a portfolio. We consider the null hypothesis that a given portfolio is not dominated by any other feasible portfolio. This requires a substantial modification to the test statistics of LMW due to boundary problems, an issue raised in Kroll and Levy (1980). Specifically, we estimate a 'contact set' and compute the supremum in the test statistic only over the complement of a small enlargement of this set. For this we need to develop new theory for the behavior of these estimated sets and derived quantities. Our theory is related to recent work of Chernozhukov, Hong, and Tamer (2007). There is also an issue of computation because one has to search over a very large set of portfolios. We propose to solve this computational issue using a nested linear programming algorithm. We provide the limiting distribution of our test statistic under the null hypothesis of SD efficiency, and give also some results on asymptotic power. We propose to use the subsampling method for obtaining the critical values, and we establish that this is consistent under general conditions. We evaluate the performance of our method on simulated data.

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. We discuss this issue below. In general a portfolio may be second order SD efficient but not mean variance efficient and vice versa so the two concepts might yield different predictions.

2 The Null Hypothesis

We consider a single-period portfolio decision under uncertainty model. Individuals chose portfolios of assets to maximize the expected utility of the returns to their portfolio. Let $X = (X_1, \ldots, X_K)^\top$ be the vector of returns on a set of K assets, and let Y be the return on some benchmark asset that is a portfolio of X. We consider portfolios with return $X^\top \lambda$, 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), and let $\mathcal{U}_3$ be the set of functions in $\mathcal{U}_2$ for which $u''' \leq 0$.

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

Likewise one can define third order efficiency replacing $\mathcal{U}_2$ by $\mathcal{U}_3$. This is the definition of portfolio efficiency used in Post (2003). Bawa, Bodurtha, Rao, and Suri (1985) distinguish between the admissible set of portfolios, which is a subset of the choice set that contains only portfolios that are not pairwise dominated by any other portfolio, and the optimal set, which is a subset of the admissible set that will be chosen by some utility function in the class.

Let $F_\lambda(\cdot)$ and $F_Y(\cdot)$ be the c.d.f.'s of $X^\top \lambda$ and Y, respectively. For a given integer $s \geq 1$, define the s -th order integrated c.d.f. of $X^\top \lambda$ 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 $X^\top \lambda$ 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 is *s-th order SD efficient* according to definition 1 in the sense that there does not exist any portfolio in $\{X^\top \lambda : \lambda \in \Lambda_0\}$ 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 (2007) among others. In the next section we discuss the general approach for testing this hypothesis.

2.1 General Strategy

Let F be the joint distribution of X. The general approach is to find a functional $d(F) \in \mathbb{R}$ such that $d(F) \leq 0$ when F satisfies the null hypothesis, while $d(F) > 0$ when F does not satisfy the null hypothesis. One then replaces F by an estimate $\widehat{F}$ and computes the empirical functional $d(\widehat{F})$ and rejects for large positive values of $d(\widehat{F})$. To carry out a statistical test one has to choose the cut-off point c_α to have certain properties, but we shall address this later.

Consider the functional

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

This is essentially a modification of the functional used in LMW to test for stochastic maximality.[1] This functional satisfies (1)≤ 0 under the null hypothesis. Unfortunately, there are some elements of the alternative for which (1)$= 0$ and so one cannot get a consistent test from this functional. Kroll and Levy (1980) consider a similar example where Y is $U[0,1]$ and X is $U[0,2]$ so that X dominates Y. They prove that $\Pr[\min_{1 \leq i \leq n} X_i < \min_{1 \leq i \leq n} Y_i] \to 1/3$ as $n \to \infty$ so that one has approximately at least one third chance of finding no dominance based on samples on X, Y.

The null and alternative hypotheses we are testing are quite complex, and to characterize them we introduce some further notation. For each λ define the three subsets of $\mathcal{X}$:

$$A_\lambda^- = \{x : G_Y(x) - G_\lambda(x) < 0\} \;\; ; \;\; A_\lambda^= = \{x : G_Y(x) - G_\lambda(x) = 0\} \;\; ; \;\; A_\lambda^+ = \{x : G_Y(x) - G_\lambda(x) > 0\}.$$

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 not an issue in testing the hypothesis of stochastic maximality, 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)$= 0$.

[1] Their null hypothesis was that there exists at least one prospect from a finite set that dominates some of the others. They considered the functional

$$d^* = \min_{\lambda \neq \mu} \sup_{x \in \mathcal{X}} \left[G_\mu(x) - G_\lambda(x) \right],$$

where λ, μ are chosen from a finite set. Under their null hypothesis $d^* \leq 0$, while under their alternative $d^* > 0$.

We next suggest some modifications of (1) that properly characterize the null hypothesis. This modification involves keeping away from the boundary points.

For each $\epsilon > 0$, define the ϵ-enlargement of the set $A_\lambda^=$ and its complement in $\mathcal{X}$:

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

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

Then let

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

Under the null hypothesis, $d_*(\epsilon, F) \leq 0$ for each $\epsilon \geq 0$, while under the alternative hypothesis we have $d_*(\epsilon, F) > 0$ for some $\epsilon > 0$. The idea is that you prevent the inner infimum ever being zero through equality on some part of $\mathcal{X}$. Now consider $\overline{d}(F) = \sup_{\epsilon \in [0, \overline{\epsilon}]} d_*(\epsilon, F)$ for some $\overline{\epsilon} > 0$. This functional divides the null from alternative. An alternative approach is based on the idea that even in cases where $\lim_{\epsilon \to 0} d_*(\epsilon, F) = 0$ under the alternative, one may have slow enough convergence in ϵ so that one can distinguish null from alternative for these cases based on the 'contact rate'. That is, we can expect $d_*(\epsilon, F) \simeq \Phi(F)\epsilon^\alpha$ as $\epsilon \to 0$ for some $\alpha > 0$ and $\Phi(F)$, where $\Phi(F) = 0$ under the null hypothesis and $\Phi(F) > 0$ under the alternative hypothesis. This higher order difference is enough to identify null from alternative as we show below.

In practice we have to estimate the set B_λ^ϵ from the data, which we do below in a simple way. See Chernozhukhov, Han, and Tamer (2007) for discussion of set estimation problems.

3 Test Statistics

We suppose now that we have a time series of observations on the assets, $X_t = (X_{1t}, \ldots, X_{Kt})^\top$ and Y_t for $t = 1, \ldots, T$. 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:

$$\widehat{A}_\lambda^= = \left\{x \in \mathcal{X} : \left|\widehat{G}_Y(x) - \widehat{G}_\lambda(x)\right| \leq k_T\right\} \tag{4}$$

$$\left(\widehat{A}_\lambda^=\right)^{\epsilon_T} = \left\{x + \eta \in \mathcal{X} : \ x \in \widehat{A}_\lambda^=, |\eta| < \epsilon_T\right\} \tag{5}$$

$$\widehat{B}_\lambda^{\epsilon_T} = \begin{cases} \mathcal{X} \backslash (\widehat{A}_\lambda^=)^{\epsilon_T} & \text{if } \widehat{A}_\lambda^= \neq \mathcal{X} \\ \mathcal{X} & \text{if } \widehat{A}_\lambda^= = \mathcal{X} \end{cases} \tag{6}$$

$$Q_T(\lambda, x) \quad = \quad \sqrt{T}\left[\widehat{G}_Y(x) - \widehat{G}_\lambda(x)\right] \tag{7}$$

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

$$d_T = \sup_{\lambda \in \Lambda_0} \inf_{x \in \widehat{B}_\lambda^{\varepsilon_T}} Q_T(\lambda, x), \tag{8}$$

and likewise for $\widehat{G}_Y(x)$. This is our proposed test statistic; rejection is for large positive values. Notice that to compute (8) requires potentially high dimensional optimization of a discontinuous non-convex/concave objective function. In the next section we discuss how to compute the test statistic (8).

4 Computational Strategy

We next discuss our suggested computational strategy in detail. The supremum over the scalar x in (8) is computed by a grid search, the main issue is with regard to the optimization over λ, which may be high dimensional. 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 (in any case, these methods do not work when the solution may be on the boundary of the parameter space). 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. We next 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.

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:

$$\max_{\theta\in\mathbb{R}^T,\lambda\in\mathbb{R}^K} \frac{1}{T}\sum_{t=1}^{T} \theta_t \tag{9}$$

$$\theta_t \leq \sum_{j=1}^{K} \lambda_j X_{jt} - \mu,\ t = 1,\ldots,T \tag{10}$$

$$\theta_t \leq 0,\ t = 1,\ldots,T \tag{11}$$

$$\sum_{j=1}^{K} \lambda_j = 1 \tag{12}$$

$$\lambda_j \geq 0,\ j = 1,\ldots,K, \tag{13}$$

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 (9)-(13) 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 (9), constraint (10) and/or (11) will be binding and hence $\widehat{\theta}_t(\mu) = \min\{X_t^\top\widehat{\lambda}(\mu) - \mu, 0\}$ 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 (9)-(13) for different values of the single threshold parameter μ. We then compute $Q_T(\lambda, x)$ for every λ from $\{\widehat{\lambda}(\mu) : \mu \in M\}$, 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 (8) can be computed by a grid search. We can also take this approach for higher-order criteria, because the efficient set then is a subset of the SSD efficient set. This approach works well for moderate sized samples and for single replications. For Monte Carlo studies with large sample sizes it becomes too time consuming. The standard simplex algorithm (employed in GAUSS/MATLAB type software) is exponential in the dimensions $(T + K - 1)$, and should be replaced by a polynomial time algorithm.

An alternative approach is to use one of the many algorithms appropriate for non-smooth optimization like the Nelder Mead or more recent developments. This method does not require any particular structure. 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 set and the MV efficient set coincide. 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 $\lambda_p = g + h\mu_p$, where the vectors $g(\mu,\Sigma), h(\mu,\Sigma)$ satisfy $g = \frac{1}{D}\left[B\Sigma^{-1}i - A\Sigma^{-1}\mu\right]$ 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 Asymptotic Properties

In this section we give the asymptotic properties of the test statistic under the null and alternative hypothesis. We also present the subsampling method for obtaining critical values and establish that our test is consistent against all alternatives under our conditions.

5.1 Null Distribution

We shall need the partition $\Lambda_0 = \Lambda_1 \cup \Lambda_2$, where $\Lambda_1 \cap \Lambda_2 = \varnothing$, $\Lambda_1 = \Lambda_0^- \cup \Lambda_0^=$, and $\Lambda_2 = \Lambda_0^+ \cup \Lambda_0^{\simeq}$ with:

$$\Lambda_0^= = \{\lambda \in \Lambda_0 : G_Y(x) = G_\lambda(x)\ \forall x \in \mathcal{X}\} \quad (14)$$

$$\Lambda_0^- = \left\{\lambda \in \Lambda_0 : \inf_{x\in\mathcal{X}} [G_Y(x) - G_\lambda(x)] < 0\right\} \quad (15)$$

$$\Lambda_0^+ = \left\{\lambda \in \Lambda_0 : \inf_{x\in\mathcal{X}} [G_Y(x) - G_\lambda(x)] > 0\right\} \quad (16)$$

$$\Lambda_0^{\simeq} = \left\{\lambda \in \Lambda_0 : \inf_{x\in\mathcal{X}} [G_Y(x) - G_\lambda(x)] = 0,\ \inf_{x\in B_\lambda^\epsilon} [G_Y(x) - G_\lambda(x)] > 0 \text{ for some } \epsilon > 0\right\}. \quad (17)$$

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

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

$$\nu_T(\lambda, x) = \sqrt{T}\left[\widehat{G}_Y(x) - \widehat{G}_\lambda(x) - G_Y(x) + G_\lambda(x)\right]. \tag{18}$$

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

$$C((\lambda_1, x_1), (\lambda_2, x_2)) = \lim_{T\to\infty} E\nu_T(\lambda_1, x_1)\nu_T(\lambda_2, x_2). \tag{19}$$

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*

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

where $\Lambda_0^=$ *is defined in (14).*

Theorem 1 shows that the asymptotic null distribution of d_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 d_T can not be tabulated once and for all. However, we define below various simulation procedures to estimate them from the data.

5.2 Critical Values

5.2.1 Subsampling

We propose 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) and Horowitz (2003). 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 for details. It is also much more computationally convenient than full resampling.

The subsampling procedure is based on the following steps:

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

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

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

(iv) Approximate the sampling distribution of d_T by

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

(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 $d_T > s_{T,b}(\alpha)$.

The circular block version (Kläver (2005) involves an edge modification in (ii) that wraps the sample around. 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*

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

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

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 d_T given in Theorem 1.*

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 (when the marginal distributions all coincide) 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 (Section 6.1)), the least favorable case 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 d_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, see Lehmann (1959, Chapter 4. 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.

5.3 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 $\lambda \in \Lambda_0^{\simeq}$.

Then, we have the following result.

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

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

Next, we determine the power of the test d_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 $F_{\lambda_T}(\cdot) = G^{(0)}_{\lambda_T}(x)$ be the c.d.f.'s of $X_t^\top \lambda_T$ for $\lambda_T \in \Lambda_{0T}$. Also, for $s \geq 1$, define $G^{(s)}_{\lambda_T}(x) = \int_{-\infty}^{x} G^{(s-1)}_{T\lambda_T}(y)dy$. 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:

$$\mathbf{H}_a : G_Y(x) - G_{\lambda_T}(x) = \frac{\delta_{Y\lambda}(x)}{\sqrt{T}} \text{ for } \lambda_T \in \Lambda_{0T} \text{ and } \lambda \in \Lambda_0^=, \tag{20}$$

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 d_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*

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

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

$$\lim_{T\to\infty} P[d_T > s_{T,\widehat{b}_T}(\alpha)] = P\left[L_0 > s(\alpha)\right], \tag{21}$$

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 d_T given in Theorem 1. Also, our test is asymptotically local unbiased because, by Anderson's lemma (see Bickel et. al. (1993, p.466)), the right hand side of (21) is less than α.

6 Simulation Study

We report the results of a small simulation study based on multivariate normal distributions with moments taken from the beta-sorted portfolios reported in Post and Versijp (2007). For every random sample, we apply our test procedures for second order and third order stochastic dominance to both test portfolios, the equally weighted portfolio (EP) and the tangency portfolio (TP). Recall that the equally weighted portfolio is inefficient according to second order and third order dominance, while the tangency portfolio is efficient. The experiments are performed for sample sizes $T \in \{50, 100, 200, 500, 1000, 2000\}$. Below we show some 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 400 simulations against sample size. The p-values are computed by comparing the test statistic with 200 recentered bootstrap resamples. The results are shown in Figures 1-4 below.

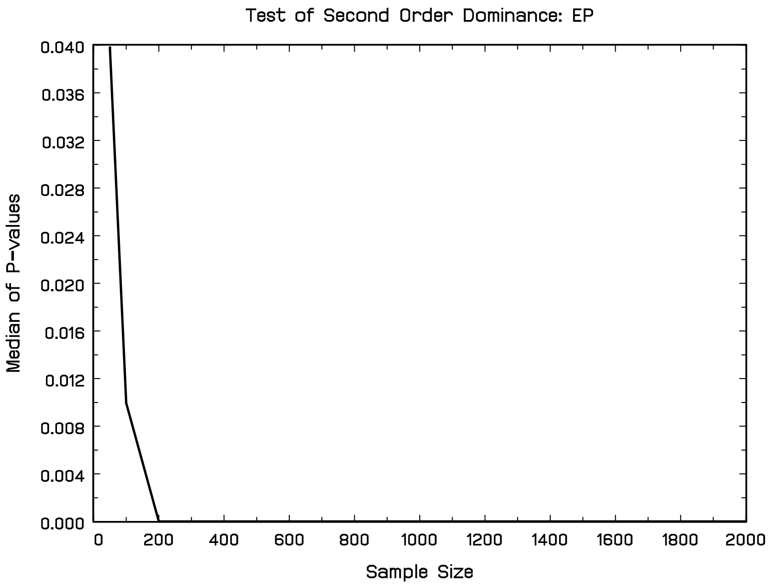

Figure 1. (Alternative hypothesis)

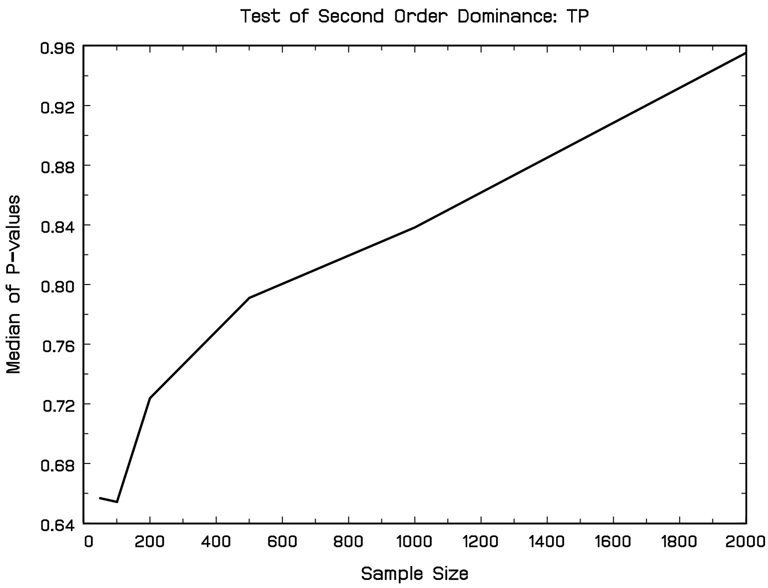

Figure 2. Null hypothesis

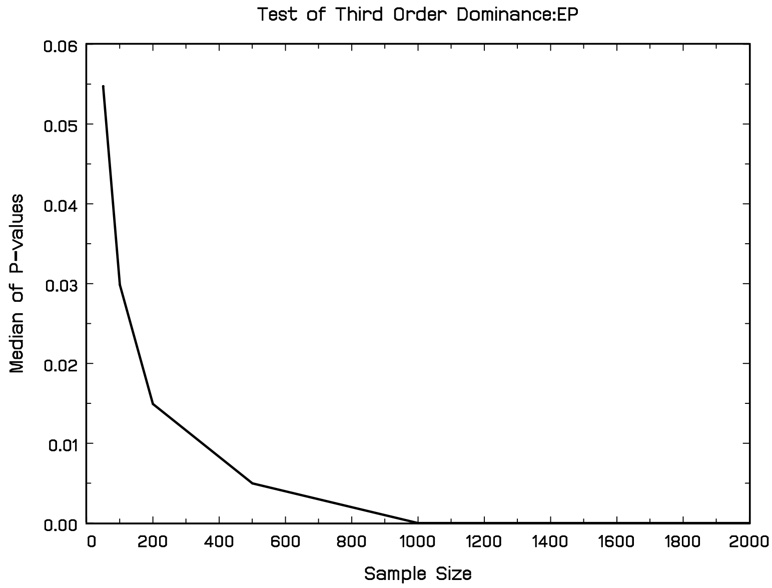

Figure 3. Alternative hypothesis

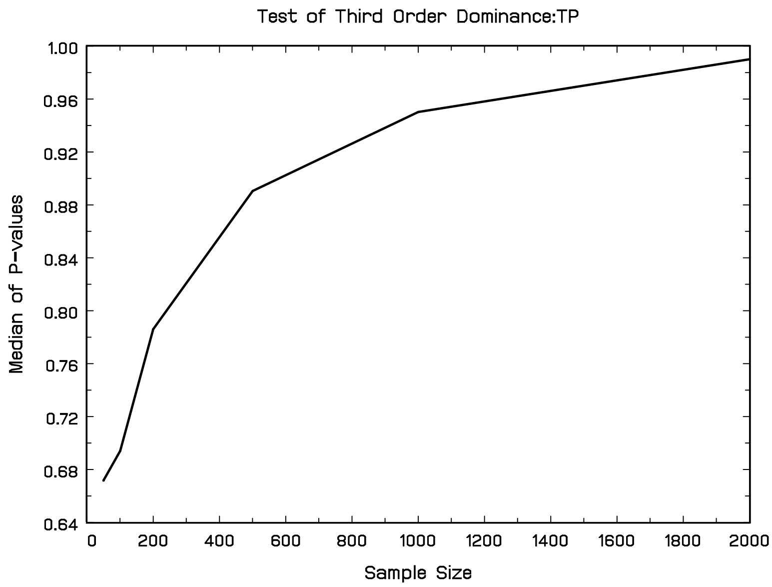

Figure 4. Null hypothesis

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. We show also the distribution of the test statistics in the two case for the largest sample sizes; they have quite different shapes and locations.

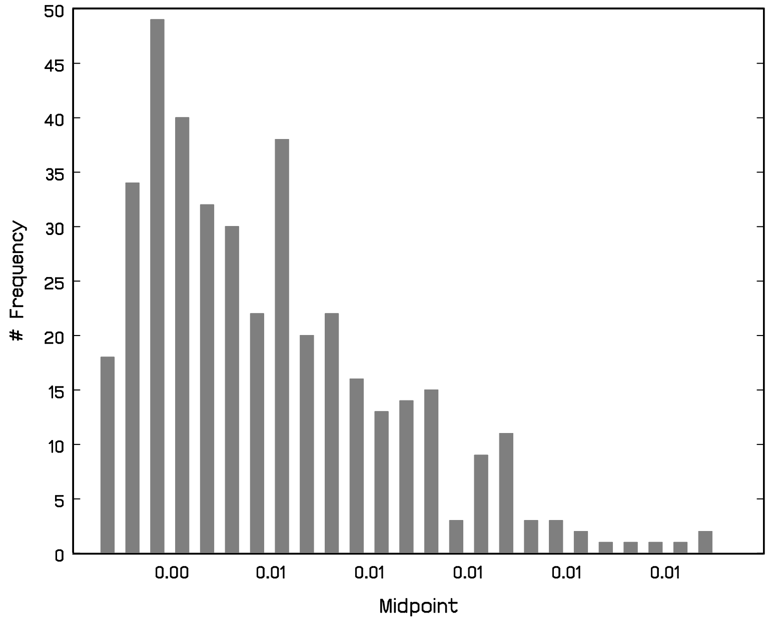

Figure 5. Histogram of test statistic for EP case

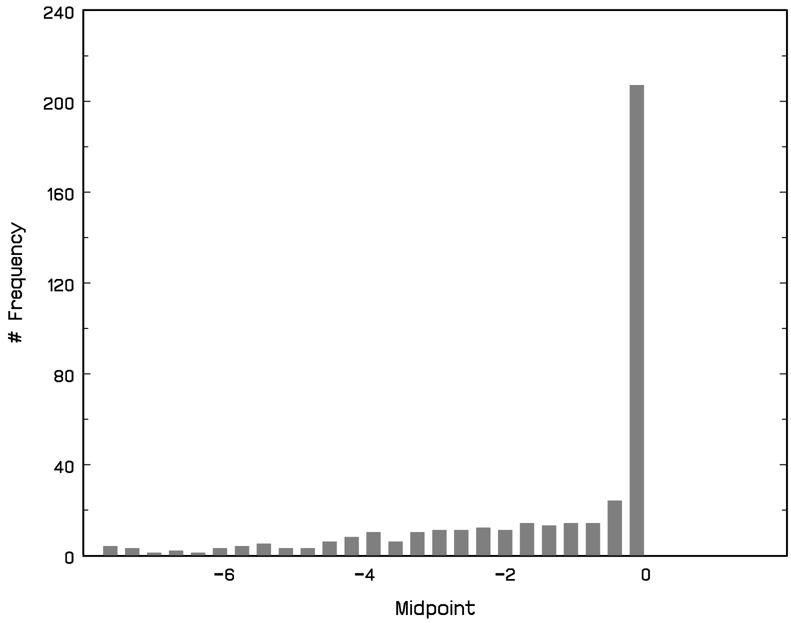

Figure 6. Histogram for test statistic, TP case.

7 Conclusions

We have proposed a statistical test of the efficiency in the stochastic dominance sense of a given portfolio. We have shown that the test is consistent against alternatives and show that it works reasonably well in small samples in a simple bivariate case based on plausible parameter values. Implementing the test for higher dimensions remains a formidable challenge, although we have suggested some techniques that may help.

8 Appendix

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

$$\nu_T(\cdot,\cdot) \Rightarrow \widetilde{\nu}(\cdot,\cdot). \tag{22}$$

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

$$\overline{\lim_{T\to\infty}} \left\| \sup_{\rho^*((\lambda_1,x_1),(\lambda_2,x_2))<\delta} |\nu_T(\lambda_1,x_1) - \nu_T(\lambda_2,x_2)| \right\|_q < \varepsilon, \tag{23}$$

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

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

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 (23) 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

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

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

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

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 $\lambda \in \Lambda_0$,

$$P\left((A_\lambda^{=})^{\epsilon_T} \subset \left(\widehat{A}_\lambda^{=}\right)^{\epsilon_T}\right) \ \to \ 1 \tag{24}$$

$$P\left(\left(\widehat{A}_\lambda^{=}\right)^{\epsilon_T} \subset (A_\lambda^{=})^{2\epsilon_T}\right) \ \to \ 1. \tag{25}$$

Suppose $A_\lambda^{=} \neq \mathcal{X}$. (If $A_\lambda^{=} = \mathcal{X}$, (25) trivially holds and (24) holds by the same argument as (26) below.) We first establish (24). Consider λ such that $A_\lambda^{=} \neq \varnothing$. (Otherwise, (24) 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 (24) holds since

$$\begin{aligned} P\left(x_0^* \in \left(\widehat{A}_\lambda^{=}\right)^{\epsilon_T}\right) \ &\geq \ P(x_0 \in \widehat{A}_\lambda^{=}) \\ &= \ P\left(\left|\widehat{G}_\lambda(x_0) - \widehat{G}_Y(x_0) - G_\lambda(x_0) + G_Y(x_0)\right| \leq k_T\right) \\ &= \ P\left(|O_p(1)| \leq (\log T)^{1/2}\right) \to 1, \end{aligned} \tag{26}$$

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

We next establish (25). 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 (26) 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

$$\sup_{\lambda \in \Lambda_0^{=}} \inf_{x \in \widehat{B}_\lambda^{\varepsilon_T}} Q_T(\lambda, x) \ \Rightarrow \ \Upsilon \tag{27}$$

$$\sup_{\lambda \in \Lambda_0^{-}} \inf_{x \in B_\lambda^{2\varepsilon_T}} \nu_T(\lambda, x) - \sup_{\lambda \in \Lambda_0^{-}} \inf_{x \in B_\lambda^{0}} \nu_T(\lambda, x) \ = \ o_p(1). \tag{28}$$

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

$$
\begin{aligned}
& \overline{\lim_{T\to\infty}} \left| P\left(d_T \le w\right) - P\left(\sup_{\lambda\in\Lambda_0^=} \inf_{x\in\widehat{B}_\lambda^{\varepsilon_T}} Q_T(\lambda,x) \le w\right) \right| \\
\le\ & \overline{\lim_{T\to\infty}} P\left(\sup_{\lambda\in\Lambda_0^-} \inf_{x\in\widehat{B}_\lambda^{\varepsilon_T}} Q_T(\lambda,x) > w\right) && (29)\\
\le\ & \overline{\lim_{T\to\infty}} P\left(\sup_{\lambda\in\Lambda_0^-} \inf_{x\in B_\lambda^{2\varepsilon_T}} Q_T(\lambda,x) > w\right) && (30)\\
\le\ & \overline{\lim_{T\to\infty}} P\left(\sup_{\lambda\in\Lambda_0^-} \inf_{x\in B_\lambda^{2\varepsilon_T}} \nu_T(\lambda,x) > w + T^{1/4}\right) && (31)\\
=\ & \overline{\lim_{T\to\infty}} P\left(\sup_{\lambda\in\Lambda_0^-} \inf_{x\in B_\lambda^{0}} \nu_T(\lambda,x) > w + T^{1/4}\right) && (32)\\
=\ & 0, && (33)
\end{aligned}
$$

where (29) holds by the fact that $\Lambda_0 = \Lambda_0^- \cup \Lambda_0^=$ under the null hypothesis and the general inequality $|P(\max(X,Y) \le x) - P(Y \le x)| \le P(X > x)$ for any rv's X and Y, (30) holds by Lemma 2, (31) follows from the result $\overline{\lim}_{T\to\infty} \sup_{\lambda\in\Lambda_0^-} \inf_{x\in B_\lambda^{2\varepsilon_T}} T^{1/4}\left(G_Y(x) - G_\lambda(x)\right) < -1$, (32) holds by (28), and (33) 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 (27) combine to yield Theorem 1.

We now establish (27) and (28). Let $w \in \mathbb{R}$. Then, by Lemma 2, we have

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

Therefore, (27) 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 (28). 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, 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

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

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

$$\sup\left|\widehat{S}_{T,b}^0(w) - S_b^0(w)\right| \xrightarrow{p} 0. \tag{34}$$

Therefore, (34) and the pointwise convergence result $b^{-1/2}d_{T,b,1} \xrightarrow{p} d_*(0)$ yield:

$$s_{T,\widehat{b}_T}^0(\alpha) = \inf\{w : \widehat{S}_{T,b}^0(w) \geq \alpha\} \to d_*(0) \geq 0, \tag{35}$$

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,

$$\begin{aligned} & P\left(d_T > s_{T,\widehat{b}_T}(\alpha)\right) \\ \geq\; & P\left(\sup_{\lambda\in\Lambda_0^+\cup\Lambda_0^{\simeq}} \inf_{x\in B_\lambda^{\varepsilon_T}} \left[\nu_T(\lambda,x) + T^{1/2}\left(G_Y(x) - G_\lambda(x)\right)\right] > \widehat{b}_T^{1/2} s_{T,\widehat{b}_T}^0(\alpha)\right) + o(1) \\ \geq\; & P\left(\sup_{\lambda\in\Lambda_0^+\cup\Lambda_0^{\simeq}} \inf_{x\in B_\lambda^{\varepsilon_T}} \left[\nu_T(\lambda,x) + T^{1/2}\left(G_Y(x) - G_\lambda(x)\right)\right] > u_T^{1/2} s_{T,\widehat{b}_T}^0(\alpha)\right) + o(1) \\ \geq\; & P\left(\sup_{\lambda\in\Lambda_0^+\cup\Lambda_0^{\simeq}} \inf_{x\in B_\lambda^{\varepsilon_T}} \left(\frac{T}{u_T}\right)^{1/2} \left[T^{-1/2}\nu_T(\lambda,x) + \left(G_Y(x) - G_\lambda(x)\right)\right] > d_*(0)\right) + o(1) \end{aligned} \tag{36}$$

where the first inequality holds by Lemma 2 and the second inequality holds by Assumption 3, and the last inequality holds by (35). Now consider the right hand side of (36). Note that

$$T^{-1/2}\nu_T(\lambda,x) \xrightarrow{p} 0 \tag{37}$$

by Lemma 1. Also,

$$\underline{\lim}_{T\to\infty} \sup_{\lambda\in\Lambda_0^+\cup\Lambda_0^{\simeq}} \left(T/u_T\right)^{1/2}\Delta_\lambda(\epsilon_T) > d_*(0) \tag{38}$$

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 $\Lambda_0^+ = \varnothing$, $\underline{\lim}_{T\to\infty}\sup_{\lambda\in\Lambda_0^{\simeq}}\left(T/u_T\right)^{1/2}\Delta_\lambda(\epsilon_T) > 0 = d_*(0)$ by Assumption 4. Therefore, (36), (37), and (38) imply that

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

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 (20). 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

$$Q_T(\lambda_T, x) = \nu_T^*(\lambda, c/\sqrt{T}, x) + \delta_{Y\lambda}(x), \tag{39}$$

we have,

$$\begin{aligned}
& \sup_{\lambda_T \in \Lambda_{0T}} \inf_{x \in \widehat{B}_\lambda^{\epsilon_T}} Q_T(\lambda_T, x) - \sup_{\lambda \in \Lambda_0^=} \inf_{x \in \mathcal{X}} \left[\nu_T^*(\lambda, 0, x) + \delta_{Y\lambda}(x)\right] \\
= & \sup_{\lambda \in \Lambda_0^=, c/\sqrt{T} \in \mathcal{Z}} \inf_{x \in \mathcal{X}} \left[\nu_T^*(\lambda, c/\sqrt{T}, x) + \delta_{Y\lambda}(x)\right] - \sup_{\lambda \in \Lambda_0^=} \inf_{x \in \mathcal{X}} \left[\nu_T^*(\lambda, 0, x) + \delta_{Y\lambda}(x)\right] \qquad (40) \\
= & \; o_p(1), \qquad (41)
\end{aligned}$$

where (40) 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 (41) 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] Andrews, D. W. K. and D. Pollard (1994). An introduction to functional central limit theorems for dependent stochastic processes. International Statistical Review 62, 119-132.

[2] Bawa, V., J. Bodurtha, M.R. Rao, and H.L. Suri (1985). On determination of stochasitc dominance optimal set. Journal of Finance 40, 417-431.

[3] Bawa, V., and E. B. Lindenberg, 1977, Capital market equilibrium in a mean-lower partial moment framework, Journal of Financial Economics 5, 189-200.

[4] Bickel, P.J., C.A.J. Klaassen, Y. Ritov and J.A. Wellner (1993). Efficient and Adaptive Estimation for Semiparametric Models. Springer Verlag, Berlin.

[5] Black, Fischer, Michael C. Jensen, and Myron Scholes, 1972, The capital asset pricing model: some empirical tests, Studies in the theory of capital markets (Praeger).

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

[7] Campbell, J. Y., A. W. Lo, & A. C. MacKinlay (1997). The Econometrics of Financial Markets. Princeton University Press, Princeton.

[8] Chernozhukov, V., Hong, H., and E. Tamer (2007). Estimation and Confidence Regions for Parameter Sets in Econometric Models *Econometrica* 75, No. 5, 1243–1284.

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

[10] Fama, Eugene F., and James D. MacBeth, 1973, Risk, return and equilibrium: Empirical tests, Journal of Political Economy 71, 43-66.

[11] Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, Econometrica 57, 1121-1152.

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

[13] Hogan, William W., and James M. Warren, 1974, Toward the development of an equilibrium capital-market model based on semivariance, Journal of Financial and Quantitative Analysis 9, 1-11.

[14] Horowitz, J.L. (2003). The Bootstrap in Econometrics. Statistical Science, 18, 211-218.

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

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

[17] Kläver, H., 2005. Testing for stochastic dominance using circular block methods. Graduate School of Risk Management. University of Cologne.

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

[19] Kroll, Y. and H. Levy (1980). Sampling errors and portfolio efficient analysis. Journal of Financial and Quantitative Analysis 15, 655-688.

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

[21] Lehmann, E.L., (1959): *Testing Statistical Hypotheses.* Wiley, New York.

[22] Levy, H. (2006) Stochastic Dominance: Investment Decisionmaking under Uncertainty. Springer Verlag, Berlin 2nd edition.

[23] Linton, O., E. Maasoumi, and Y. -J. Whang (2005): “Consistent Testing for Stochastic Dominance under General Sampling Schemes,” *Review of Economic Studies* 72, 735-765.

[24] 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.

[25] Markowitz, H. (1952). Portfolio Selection. Journal of Finance 7, 77-91.

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

[27] Politis,D.N. and J.P.Romano (1994), ‘Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions’, Ann. Statist., vol. 22, No. 4, pp. 2031-2050.

[28] Politis, D.N., J.P.Romano, M.Wolf, (1999). SUBSAMPLING, Springer-Verlag, New York.

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

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

[31] Post, T., and P. Versijp (2007) Multivariate tests for stochastic dominance efficiency of a given portfolio, Journal of Financial and Quantitative Analysis, 42 (2), 489-516

[32] Price, Kelly, Barbara Price, and Timothy J. Nantell, 1982, Variance and lower partial moment measures of systematic risk: some analytical and empirical results, Journal of Finance 37, 843-855.

[33] Post, Thierry and van Vliet, Pim, "Empirical Tests of the Mean-semivariance CAPM" (June 2005). http://ssrn.com/abstract=557220

[34] 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.

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