The Econometrics Journal

Econometrics Journal (2014), volume **17**, pp. S59–S74.
doi: 10.1111/ectj.12016

Testing for the stochastic dominance efficiency of a given portfolio

OLIVER LINTON[†], THIERRY POST[‡] AND YOON-JAE WHANG[§]

[†]*Faculty of Economics, University of Cambridge, The Austin Robinson Building, Sidgwick Ave, Cambridge, CB3 9DD, UK*.
E-mail: obl20@cam.ac.uk

[‡]*Graduate School of Business, Koç University, Rumelifeneri Yolu, 34450, Sariyer, Istanbul, Turkey*.
E-mail: thierrypost@hotmail.com

[§]*Department of Economics, Seoul National University, Shillim Dong, Seoul, 151 746, Korea*.
E-mail: whang@snu.ac.kr

First version received: August 2012; final version accepted: August 2013

Summary 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.

Keywords: *Linear programming, Portfolio choice, Stochastic dominance, Subsampling.*

1. INTRODUCTION

The portfolio choice problem is the cornerstone of portfolio theory and asset pricing. There are two main approaches to this: the mean variance (MV) approach and the more general stochastic dominance (SD) approach. In the MV 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, e.g., Markowitz (1959) and Gibbons et al. (1989). The SD approach makes much weaker assumptions about the distribution of returns and preferences. However, the practical implementation of SD analysis has proven to be more difficult because of a number of mathematical and statistical issues. 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 progress with regards to computational and statistical issues, which has advanced the position of the SD method; see Levy (2006) for an overview and bibliography.

We propose a test of whether a given portfolio is efficient with respect to the SD criterion in comparison with a polyhedral 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 have provided a method of inference based on the first-order optimality conditions of the investor's expected utility maximization problem. Their statistical approach uses a conservative bounding distribution, which might compromise statistical power or the ability to detect inefficient portfolios in small samples. They have also used a sampling scheme that assumes serially i.i.d.

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society. Published by John Wiley & Sons Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA, 02148, USA.

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

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 the use of a modification of the Kolmogorov–Smirnov test statistic of McFadden (1989) and Klecan et al. (1991). Recently, Linton et al. (2005a) (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 several non-trivial conceptual issues. The null hypothesis in LMW was of stochastic maximality in a finite set (i.e., 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 involves pairwise comparison, it is not appropriate for the situation where an investor might combine a set of basis assets into a portfolio. In addition, financial theory generally focuses on the concept of efficiency or non-dominance rather than maximality. For example, under the standard capital asset pricing model, the market portfolio is efficient but not maximal; it does not dominate any of the other efficient portfolios on the so-called capital market line.

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 because of boundary problems, an issue raised by 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 behaviour of these estimated sets and derived quantities. Our theory is related to the recent work of Chernozhukov et al. (2007). There is also an issue of computation because it is necessary to search over a large set of portfolios. When the number of base assets considered is small, grid search can be an effective way to solve the required optimization problem. For larger problems, some iterative optimization algorithms such as that of Nelder and Mead (1965) might be better suited. We leave the further development of an efficient algorithm for large-scale applications for a follow-up project. Here, we address the statistical performance of our test statistics. We provide the limiting distribution of our test statistic under the null hypothesis of SD efficiency, and we also give 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.

For various reasons, we focus on the SD criteria of order two and higher, meaning that risk aversion is assumed throughout this study. Kopa and Post (2009) have provided, in a portfolio context, an extensive analysis of first-order SD (FSD), which allows for risk seeking. First, risk aversion is a standard assumption in financial economics, 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 (e.g., the popularity of lotteries). However, the bulk of the literature on asset pricing and portfolio selection assumes that investors are globally risk averse. Secondly, FSD generally involves mixed integer programming techniques. In large-scale applications, the computational burden quickly becomes prohibitive, especially if the statistical inference is based on resampling methods. Thirdly, the FSD criterion is very general and allows for 'exotic' preference structures (e.g., utility functions with inflection points and discontinuous jumps). Thus, an empirical test for FSD efficiency will have considerable freedom to fit an unrealistic utility function to the data. Presumably, this will considerably slow down the rate of convergence of an empirical test.

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society.

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

2. THE NULL HYPOTHESIS

We consider a single-period portfolio decision under risk model. Individuals choose 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 portfolio of X. The benchmark portfolio is taken as given and follows from theory or observation (e.g., the aggregate market portfolio held by the representative investor or the actual portfolio held by a given individual investor). 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. In this general case, we can enumerate the V vertices of the polytope and replace the K base assets with these vertices. However, restrictions on value-at-risk and tracking error introduce non-linearities and estimation issues that require modifications of our model.

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 $u' \geq 0$ (non-satiation). Also, let $\mathcal{U}_2$ denote the class of all utility functions in $\mathcal{U}_1$ for which $u'' \leq 0$ (risk aversion), and let $\mathcal{U}_3$ be the set of functions in $\mathcal{U}_2$ for which $u''' \leq 0$ (prudence).

DEFINITION 2.1. (SSD Efficiency). *The asset Y is SSD efficient if and only if for some $u \in \mathcal{U}_2$, $E[u(Y)] \geq E[u(X^\top\lambda)]$ for all $\lambda \in \Lambda_0$.*

Likewise, we can define third-order efficiency replacing $\mathcal{U}_2$ by $\mathcal{U}_3$. This is the definition of portfolio efficiency used by Post (2003), which generalized the definition of convex SD of Fishburn (1974) and Bawa et al. (1985) to account for diversification.

Let $F_\lambda(\cdot)$ and $F_Y(\cdot)$ be the cumulative distribution functions (CDFs) of $X^\top\lambda$ and Y, respectively. For a given integer $s \geq 2$, define the sth-order integrated CDF of $X^\top\lambda$ to be

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

where $G_\lambda^{(1)}(\cdot) = F_\lambda(\cdot)$, and likewise for $G_Y^{(s)}(x)$. A portfolio $X^\top\lambda$ s-order dominates Y 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 $\mathcal{X}$. One way to interpret this condition is that portfolio $X^\top\lambda$ has a more favourable $(s-1)$th-order lower partial moment than portfolio Y for all relevant return levels. For $s \geq 2$, this definition is equivalent to Definition 2.1, but not so for $s = 1$; see Kopa and Post (2009) for a 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 following null hypothesis.

$\mathbf{H}_0$: *Y is sth-order SD efficient according to Definition 2.1 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 the efficiency hypothesis against the general alternative hypothesis $\mathbf{H}_A$ that is the negation of $\mathbf{H}_0$ (i.e., the evaluated portfolio is inefficient). In our case with a convex choice set and SD criteria of order

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society.

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

two and higher, inefficiency of Y implies that there exists a portfolio with weights $\lambda_0 \in \Lambda_0$ that dominates Y (a result that does not apply in the case of a discrete choice set and/or the first-order criterion). However, our focus here is on statistical inference regarding the efficiency classification of a given portfolio rather than the (often non-robust) dominance relation between individual portfolios. Also, we do not attempt to characterize the entire set of efficient portfolios, which is generally uncountable, non-convex and non-robust, although such a characterization (presumably by means of mixed-integer linear constraints) can be of interest for a stochastic programming approach to portfolio construction. Similarly, a characterization of all utility functions (presumably piecewise polynomials) that support the efficiency classification of a given portfolio is beyond the scope of this paper.

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. We then replace F by an estimate $\widehat{F}$, compute the empirical functional $d(\widehat{F})$ and reject for large positive values of $d(\widehat{F})$. To carry out a statistical test, we have to choose the cut-off point c_α to have certain properties, but we shall address this later.

Consider the functional

$$d = \sup_{\lambda \in \Lambda_0} \inf_{x \in \mathcal{X}} (G_Y(x) - G_\lambda(x)). \tag{2.1}$$

This is essentially a modification of the functional used in LMW to test for stochastic maximality.[1] This functional satisfies (2.1) ≤ 0 under the null hypothesis. Unfortunately, there are some elements of the alternative for which (2.1) $= 0$ and so we cannot obtain a consistent test from this functional. Kroll and Levy (1980) have considered a similar example, where Y is $U[0, 1]$ and X is $U[0, 2]$ so that X dominates Y. They have proved 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 there is 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, we define the following 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 $X^\top \lambda$ dominates Y, then $A_\lambda^- = \varnothing$, and A_λ^+ is non-empty. However, it can be that both $A_\lambda^=$ and A_λ^+ are non-empty, 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, because 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^+

[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

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

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

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society.

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

are non-empty and $A_\lambda^- = \varnothing$ for some λ. For these λ, we have $\inf_{x\in\mathcal{X}}(G_Y(x) - G_\lambda(x)) = 0$, even though $X^\top\lambda$ dominates Y. If the other λ 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 find that (2.1) $= 0$.

We next suggest some modifications of (2.1) 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 $\mathcal{X}$:

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

$$B_\lambda^\epsilon = \begin{cases} \mathcal{X}\backslash(A_\lambda^=)^\epsilon & \text{if} \quad A_\lambda^= \neq \mathcal{X} \\ \mathcal{X} & \text{if} \quad A_\lambda^= = \mathcal{X}. \end{cases} \tag{2.2}$$

Then let

$$d_*(\epsilon, F) = \sup_{\lambda\in\Lambda_0} \inf_{x\in B_\lambda^\epsilon} (G_Y(x) - G_\lambda(x)). \tag{2.3}$$

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 to 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 the 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, there might be slow enough convergence in ϵ so that it is possible to distinguish the null from the 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 the null from the 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 Chernozhukov et al. (2007) for a 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 (2.3) 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 4.2, where c_0 is a positive constant. Define

$$\widehat{A}_\lambda^= = \{x \in \mathcal{X} : |\widehat{G}_Y(x) - \widehat{G}_\lambda(x)| \leq k_T\} \tag{3.1}$$

$$\left(\widehat{A}_\lambda^=\right)^{\epsilon_T} = \left\{x + \eta \in \mathcal{X} : \; x \in \widehat{A}_\lambda^=, |\eta| < \epsilon_T\right\} \tag{3.2}$$

$$\widehat{B}_\lambda^{\epsilon_T} = \begin{cases} \mathcal{X}\backslash\left(\widehat{A}_\lambda^=\right)^{\epsilon_T} & \text{if} \quad \widehat{A}_\lambda^= \neq \mathcal{X} \\ \mathcal{X} & \text{if} \quad \widehat{A}_\lambda^= = \mathcal{X} \end{cases} \tag{3.3}$$

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society.

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

$$Q_T(\lambda, x) = \sqrt{T}(\widehat{G}_Y(x) - \widehat{G}_\lambda(x)) \tag{3.4}$$

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

$$d_T = \sup_{\lambda\in\Lambda_0} \inf_{x\in\widehat{B}_\lambda^{\varepsilon_T}} Q_T(\lambda, x), \tag{3.5}$$

and likewise for $\widehat{G}_Y(x)$. This is our proposed test statistic; rejection is for large positive values. Note that to compute (3.5) requires potentially high dimensional optimization of a discontinuous non-convex/concave objective function.

We next discuss briefly the computational issues. The supremum over the scalar x in (3.5) is computed by a grid search; the main issue is with regards to the optimization over λ, which might be high dimensional. The objective function $Q_T(\lambda, x)$ can be written as

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

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, we cannot use standard derivative-based algorithms, such as Newton–Raphson, to find the optima (in any case, these methods do not work when the solution might be on the boundary of the parameter space). We could replace the empirical CDFs by smoothed empirical CDF 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 principle, it is possible to use one of the many algorithms appropriate for non-smooth optimization, such as the algorithm of Nelder and Mead (1965) or more recent developments. This method does not require any particular structure. For this algorithm to work well in high-dimensional cases, good starting values are necessary. One proposal is to obtain these by grid searching over the 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 MV 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 MV 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 $g = (1/D)(B\Sigma^{-1}i - A\Sigma^{-1}\mu)$ and $h = (1/D)(C\Sigma^{-1}\mu - A\Sigma^{-1}i)$, 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 et al. (1997, p. 185). Therefore, we take a grid of values of μ_p, obtain λ_p for this grid and then compute the test statistic. To impose the condition 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.

In an earlier version of this paper (Linton et al., 2005b), we proposed an alternative computationally desirable method based on linear programming, and in a future paper we will implement a large-scale application based on this idea.

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society.

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

4. 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.

4.1. Null distribution

We 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

$$\Lambda_0^= = \{\lambda \in \Lambda_0 : G_Y(x) = G_\lambda(x)\ \forall x \in \mathcal{X}\} \tag{4.1}$$

$$\Lambda_0^- = \Big\{\lambda \in \Lambda_0 : \inf_{x\in\mathcal{X}}(G_Y(x) - G_\lambda(x)) < 0\Big\}$$

$$\Lambda_0^+ = \Big\{\lambda \in \Lambda_0 : \inf_{x\in\mathcal{X}}(G_Y(x) - G_\lambda(x)) > 0\Big\}$$

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

Under the null hypothesis, Λ_2 is empty and hence $\Lambda_0 = \Lambda_1$. However, under the alternative hypothesis, Λ_2 is non-empty.

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

ASSUMPTION 4.1. *(a) $\{(X_t^\top, Y_t)^\top : t = 1, \ldots, T\}$ 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 $q > 2(K+1)$. (b) The supports of X_{kt} and Y_t are compact $\forall k = 1, \ldots, K$. (c) The distributions of X_t and Y_t are absolutely continuous with respect to the Lebesgue measure and have bounded densities.*

ASSUMPTION 4.2. *(a) $\{\epsilon_T : T \geq 1\}$ is a sequence of positive constants such that*

$$\lim_{T\to\infty} \epsilon_T = 0 \text{ and } \epsilon_T > k_T\ \forall T \geq 1.$$

(b) For each $x \in \mathcal{X}$, constant $C_1 > 0$ and $\lambda \in \Lambda_0$ such that $A_\lambda^= \neq \varnothing$, we have

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

for T sufficiently large.

Assumption 4.2 requires that the function $G_Y(\cdot) - G_\lambda(\cdot)$ is monotonic on a ϵ_T-neighbourhood 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 neighbourhood 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

$$v_T(\lambda, x) = \sqrt{T}(\widehat{G}_Y(x) - \widehat{G}_\lambda(x) - G_Y(x) + G_\lambda(x)). \tag{4.2}$$

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society.

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

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

$$C((\lambda_1, x_1), (\lambda_2, x_2)) = \lim_{T\to\infty} E\nu_T(\lambda_1, x_1)\nu_T(\lambda_2, x_2). \tag{4.3}$$

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

THEOREM 4.1. *Suppose Assumptions 4.1 and 4.2 hold. Then, under the null hypothesis, we have*

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

where $\Lambda_0^=$ is defined in (4.1).

Theorem 4.1 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 cannot be tabulated once and for all. However, we define in Section 4.2 various simulation procedures to estimate them from the data.

4.2. Critical values

4.2.1. 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 et al. (1999). The subsampling is useful in our context because our null hypothesis consists of a 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.

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

STEP 2. Generate subsamples $\mathcal{W}_{T,b,t} = \{Z_t, \ldots, Z_{t+b-1}\}$ of size b for $t = 1, \ldots, T-b+1$.

STEP 3. Compute test statistics $d_{b;T,t}$ using the subsamples $\mathcal{W}_{T,b,t}$ for $t = 1, \ldots, T-b+1$.

STEP 4. Approximate the sampling distribution of d_T by

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

STEP 5. Get the α-sample quantile of $\widehat{S}_{T,b}(\cdot)$, i.e.,

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

STEP 6. 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 Step 2 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 the following.

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society.

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

ASSUMPTION 4.3. $\Pr(l_T \leq \widehat{b}_T \leq u_T) \to 1$ *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 the asymptotically correct size.

THEOREM 4.2. *Suppose Assumptions 4.1–4.3 hold. Then, under the null hypothesis, we have*

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

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

as $T \to \infty$, *where* $s(\alpha)$ *denotes the* α*-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 4.1.*

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 the bootstrap critical value $h_T(\alpha)$ is α if the least favourable case (when the marginal distributions all coincide) is true. Therefore, in this case, we might prefer bootstrap to subsampling because the former uses the full sample information and hence might be more efficient in finite samples. However, as we have argued in another context (see LMW, Section 6.1), the least favourable 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). However, the subsample-based test is unbiased and asymptotically similar on the boundary and might be preferred in this sense. In practice, it might be desirable to employ both approaches to see if the results obtained are robust to the choice of resampling schemes.

4.3. Asymptotic power

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

If the alternative hypothesis is true, $\Lambda_2 = \Lambda_0^{+} \cup \Lambda_0^{\simeq}$ is non-empty. When Λ_0^{+} is empty, we need the following assumption for the consistency of our test.

ASSUMPTION 4.4. *When* $\Lambda_2 = \Lambda_0^{\simeq}$, $\underline{\lim}_{T\to\infty}(T/u_T)^{1/2}\Delta_\lambda(\epsilon_T) > 0$ *for some* $\lambda \in \Lambda_2$, *where* $\Delta_\lambda(\epsilon) = \inf_{x \in B_\lambda^\epsilon}(G_Y(x) - G_\lambda(x))$ *and* u_T *is defined in Assumption 4.3.*

For each $\lambda \in \Lambda_0^{\simeq}$, $\Delta_\lambda(\epsilon)$ is 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 4.4 holds if ϵ_T goes to zero at a rate not too fast 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 4.3. *Suppose that Assumptions 4.1–4.4 hold. Then, under the alternative hypothesis, we have*

$$\Pr\left(d_T > s_{T,\widehat{b}_T}(\alpha)\right) \to 1 \text{ as } T \to \infty.$$

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society.

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

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

$$\Lambda_{0T} = \left\{\lambda(1 + c/\sqrt{T}) : \lambda \in \Lambda_0^=, \quad c \in \mathbb{R}^K\right\}.$$

Let $F_{\lambda_T}(\cdot) = G^{(1)}_{\lambda_T}(\cdot)$ be the CDFs of $X_t^\top \lambda_T$ for $\lambda_T \in \Lambda_{0T}$. Also, for $s \geq 2$, 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:

$$\mathbf{H}_a : G_Y(x) - G_{\lambda_T}(x) = \frac{\delta_{\lambda,T}(x)}{\sqrt{T}} \text{ for } \lambda_T \in \Lambda_{0T} \text{ and } \lambda \in \Lambda_0^=, \tag{4.4}$$

where $\delta_{\lambda,T}(\cdot)$ is a real function such that $\delta_{\lambda,T}(x) = \delta_\lambda(x) + o(1)$ uniformly in $x \in \mathcal{X}$. For example, when $s = 1$, $\delta_\lambda(x) = -c\int_{-\infty}^x z'(\partial F_\lambda(z)/\partial z)dz$ under suitable conditions, where F_λ denotes the joint CDF of $X_t'\lambda$.

The asymptotic distribution of d_T under the local alternatives is given in the following theorem.

THEOREM 4.4. *Suppose Assumptions 4.1 and 4.2 (with Λ_0 replaced by Λ_{0T}) hold. Then, under the sequence of local alternatives $\mathbf{H}_a$, we have*

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

where $\widetilde{\nu}(\lambda, x)$ is defined as in Theorem 4.1.

The result of Theorem 4.4 implies that asymptotic local power of our test based on the subsample critical value is given by

$$\lim_{T\to\infty} \Pr(d_T > s_{T,\widehat{b}_T}(\alpha)) = \Pr(L_0 > s(\alpha)), \tag{4.5}$$

where L_0 denotes the limit distribution given in Theorem 4.4 and $s(\alpha)$ denotes the α-quantile of the asymptotic null distribution of d_T given in Theorem 4.1. The asymptotic local power of the test is determined by $\delta_\lambda(\cdot)$ and the test is asymptotically locally unbiased if $\delta_\lambda(x)$ is a non-negative function and is bound bounded away from zero on (a strict subset of) $\mathcal{X}$ for all $\lambda \in \Lambda_0^=$.

5. CONCLUSIONS

We have proposed a statistical test of the efficiency in the SD sense of a given portfolio relative to a polyhedral set of feasible portfolios formed from a discrete set of base assets. We have shown that the test is consistent against alternatives. In an earlier version of this paper (Linton et al., 2005b), we have shown that it works reasonably well in small samples in a simple bivariate Monte Carlo simulation based on plausible parameter values. Implementing the test for higher dimensions (a large number of base assets) remains a formidable challenge, and we are pursuing this in a separate paper (in progress).

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society.

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

ACKNOWLEDGEMENTS

We thank participants at the conference in honour of Haim Levy held at CEMMAP, 4–5 November 2005, for comments. Oliver Linton thanks the ERC for financial support. Yoon-Jae Whang thanks the National Research Foundation (NRF-2012S1A3A2033467) and the Institute of Economic Research, Seoul National University, for financial support.

REFERENCES

Andrews, D. W. K. and D. Pollard (1994). An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review 62*, 119–32.

Bawa, V., J. Bodurtha, M. R. Rao and H. L. Suri (1985). On determination of stochastic dominance optimal set. *Journal of Finance 40*, 417–31.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction* (2nd ed.). *Lecture Note in Statistics, Volume 110*. New York, NY: Springer.

Campbell, J. Y., A. W. Lo and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

Chernozhukov, V., H. Hong and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica 75*, 1243–84.

Davidson, R. and J.-Y. Duclos (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica 68*, 1435–64.

Fishburn, P. C. (1974). Convex stochastic dominance with continuous distribution functions. *Journal of Economic Theory 7*, 143–58.

Gibbons, M. R., S. A. Ross and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica 57*, 1121–52.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and its Applications*. New York, NY: Academic Press.

Kläver, H. (2005). Testing for stochastic dominance using circular block methods. Working paper, Graduate School of Risk Management, University of Cologne.

Klecan, L., R. McFadden and D. McFadden (1991). A robust test for stochastic dominance. Working paper, Department of Economics, MIT.

Kopa, M. and T. Post (2009). A portfolio efficiency test based on FSD. *Journal of Financial and Quantitative Analysis 44*, 1103–24.

Kroll, Y. and H. Levy (1980). Sampling errors and portfolio efficient analysis. *Journal of Financial and Quantitative Analysis 15*, 655–88.

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

Levy, H. (2006). *Stochastic Dominance: Investment Decisionmaking under Uncertainty* (2nd ed.). Berlin: Springer.

Linton, O., E. Maasoumi and Y-J. Whang (2005a). Consistent testing for stochastic dominance under general sampling schemes. *Review of Economic Studies 72*, 735–65 (LMW).

Linton, O., T. Post and Y-J. Whang (2005b). Testing for the stochastic dominance efficiency of a given portfolio. Working paper, London School of Economics.

McFadden, D. (1989). Testing for stochastic dominance. In T. Fomby and T. K. Seo (Eds.), *Studies in the Economics of Uncertainty: In Honor of J. Hadar*, 113–34. Berlin: Springer.

Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investments*. New York, NY: Wiley.

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society.

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *Computer Journal 7*, 308–13.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*, 1027–57.

Politis, D. N. and J. P. Romano (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics 22*, 2031–50.

Politis, D. N., J. P. Romano and M. Wolf (1999). *Subsampling*. New York, NY: Springer.

Post, T. (2003). Empirical tests for stochastic dominance efficiency. *Journal of Finance 58*, 1905–32.

Post, T. and P. Versijp (2007). Multivariate tests for stochastic dominance efficiency of a given portfolio. *Journal of Financial and Quantitative Analysis 42*, 489–516.

APPENDIX

LEMMA A.1. *Suppose Assumption 4.1 holds, Then, we have*

$$\nu_T(\cdot,\cdot) \Rightarrow \widetilde{\nu}(\cdot,\cdot). \tag{A.1}$$

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

$$\overline{\lim_{T\to\infty}} \| \sup_{\rho^*((\lambda_1,x_1),(\lambda_2,x_2))<\delta} |\nu_T(\lambda_1,x_1) - \nu_T(\lambda_2,x_2)| \|_q < \varepsilon, \tag{A.2}$$

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

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

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), because 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 4.1. However, the stochastic equicontinuity condition (A.2) 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 4.1(a). Also, let

$$\mathcal{F} = \{f_t(\lambda,x) : (\lambda,x) \in \Lambda_0 \times \mathcal{X}\},$$

where

$$f_t(\lambda,x) = (x - Y_t)^{s-1}1(Y_t \le x) - (x - X_t^\top\lambda)^{s-1}1(X_t^\top\tau \le x).$$

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

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

Here, $\sup^*$ denotes the supremum taken over $(\lambda_1,x_1) \in \Lambda_0 \times \mathcal{X}$ for which $\|\lambda_1 - \lambda\| \le r_1$, $|x_1 - x| \le r_2$ and $\sqrt{r_1^2 + r_2^2} \le r$, the first inequality holds by several applications of Cauchy–Schwarz inequality and Assumption 4.1(b) and the second inequality holds by Assumption 4.1(c). This implies that the bracketing

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society.

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

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 A.1. □

LEMMA A.2. *Suppose Assumptions 4.1 and 4.2 hold. Then, we have*

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

as $T \to \infty$.

Proof: It suffices to show that for each $\lambda \in \Lambda_0$,

$$\Pr((A_\lambda^{=})^{\epsilon_T} \subset (\widehat{A}_\lambda^{=})^{\epsilon_T}) \to 1 \tag{A.3}$$

$$\Pr((\widehat{A}_\lambda^{=})^{\epsilon_T} \subset (A_\lambda^{=})^{2\epsilon_T}) \to 1. \tag{A.4}$$

Suppose $A_\lambda^{=} \neq \mathcal{X}$. (If $A_\lambda^{=} = \mathcal{X}$, (A.4) trivially holds and (A.3) holds by the same argument as (A.4).) We first establish (A.3). Consider λ such that $A_\lambda^{=} \neq \varnothing$. (Otherwise, (A.3) 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 (A.3) holds because

$$\begin{aligned}
\Pr(x_0^* \in (\widehat{A}_\lambda^{=})^{\epsilon_T}) &\geq \Pr(x_0 \in \widehat{A}_\lambda^{=}) \\
&= \Pr(|\widehat{G}_\lambda(x_0) - \widehat{G}_Y(x_0) - G_\lambda(x_0) + G_Y(x_0)| \leq k_T) \\
&= \Pr(|O_p(1)| \leq (\log T)^{1/2}) \to 1,
\end{aligned} \tag{A.5}$$

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

Next, we establish (A.4). Let $x_1^* \in (\widehat{A}_\lambda^{=})^{\epsilon_T}$, 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 $\Pr(x_1 \in (A_\lambda^{=})^{\epsilon_T}) \to 1$. Let $C_1 > 1$ be a constant. Then, we have, $wp \to 1$,

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

where the first inequality holds by triangular inequality and the second inequality holds using the fidi convergence result as in (A.5) and the fact that $x_1 \in \widehat{A}_\lambda^{=}$. Now, by Assumption 4.2, because $\epsilon_T > k_T$, we have $\inf_{x' \in A_\lambda^{=}} |x_1 - x'| < \epsilon_T$, $wp \to 1$, which implies that $\Pr(x_1 \in (A_\lambda^{=})^{\epsilon_T}) \to 1$, as required. □

Proof of Theorem 4.1: Here, we establish

$$\sup_{\lambda \in \Lambda_0^{=}} \inf_{x \in \widehat{B}_\lambda^{\varepsilon_T}} Q_T(\lambda, x) \Rightarrow \Upsilon \tag{A.6}$$

$$\sup_{\lambda \in \Lambda_0^{-}} \inf_{x \in B_\lambda^{2\varepsilon_T}} \nu_T(\lambda, x) - \sup_{\lambda \in \Lambda_0^{-}} \inf_{x \in B_\lambda^{0}} \nu_T(\lambda, x) = o_p(1). \tag{A.7}$$

Then, Theorem 4.1 holds because of the following arguments. For any $w \in \mathbb{R}$, we have

$$\begin{aligned}
&\overline{\lim_{T \to \infty}} |\Pr(d_T \leq w) - \Pr(\sup_{\lambda \in \Lambda_0^{=}} \inf_{x \in \widehat{B}_\lambda^{\varepsilon_T}} Q_T(\lambda, x) \leq w)| \\
&\leq \overline{\lim_{T \to \infty}} \Pr(\sup_{\lambda \in \Lambda_0^{-}} \inf_{x \in \widehat{B}_\lambda^{\varepsilon_T}} Q_T(\lambda, x) > w) \qquad \text{(A.8)} \\
&\leq \overline{\lim_{T \to \infty}} \Pr(\sup_{\lambda \in \Lambda_0^{-}} \inf_{x \in B_\lambda^{2\varepsilon_T}} Q_T(\lambda, x) > w) \qquad \text{(A.9)}
\end{aligned}$$

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society.

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

$$\leq \overline{\lim_{T\to\infty}} \Pr(\sup_{\lambda\in\Lambda_0^-} \inf_{x\in B_\lambda^{2\varepsilon_T}} \nu_T(\lambda, x) > w + T^{1/4}) \tag{A.10}$$

$$= \overline{\lim_{T\to\infty}} \Pr(\sup_{\lambda\in\Lambda_0^-} \inf_{x\in B_\lambda^{0}} \nu_T(\lambda, x) > w + T^{1/4}) \tag{A.11}$$

$$= 0. \tag{A.12}$$

Here, (A.8) holds by the fact that $\Lambda_0 = \Lambda_0^- \cup \Lambda_0^=$ under the null hypothesis and the general inequality $|\Pr(\max(X, Y) \leq x) - \Pr(Y \leq x)| \leq \Pr(X > x)$ for any random variables X and Y, (A.9) holds by Lemma A.2, (A.10) follows from the result

$$\overline{\lim_{T\to\infty}} \sup_{\lambda\in\Lambda_0^-} \inf_{x\in B_\lambda^{2\varepsilon_T}} T^{1/4}(G_Y(x) - G_\lambda(x)) < -1,$$

(A.11) holds by (A.7), and (A.12) holds because $\sup_{\lambda\in\Lambda_0^-} \inf_{x\in B_\lambda^0} \nu_T(\lambda, x) = O_p(1)$ using Lemma A.1 and continuous mapping theorem. This result and (A.6) combine to yield Theorem 4.1.

We now establish (A.6) and (A.7). Let $w \in \mathbb{R}$. Then, by Lemma A.2, we have

$$|\Pr(\sup_{\lambda\in\Lambda_0^=} \inf_{x\in\widehat{B}_\lambda^{\varepsilon_T}} Q_T(\lambda, x) \leq w) - \Pr(\sup_{\lambda\in\Lambda_0^=} \inf_{x\in\mathcal{X}} Q_T(\lambda, x) \leq w)|$$

$$\leq \Pr(\widehat{B}_\lambda^{\epsilon_T} \neq \mathcal{X} \text{ for } \lambda \in \Lambda_0^=) \to 0.$$

Therefore, (A.6) holds by Lemma A.1, the continuous mapping theorem and the fact that

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

Next, consider (A.7). 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, using an argument similar to Lemma A.1, $l_T(\cdot, \cdot, \cdot)$ is stochastic equicontinuous on $\Lambda_0^- \times \mathcal{X} \times \mathcal{Z}$, which in turn implies that

$$\sup_{\lambda\in\Lambda_0^-} \inf_{x\in B_\lambda^{2\varepsilon_T}} \nu_T(\lambda, x) - \sup_{\lambda\in\Lambda_0^-} \inf_{x\in B_\lambda^0} \nu_T(\lambda, x)$$

$$= \sup_{\lambda\in\Lambda_0^-} \inf_{x\in B_\lambda^0, |z|\leq 2\epsilon_T} l_T(\lambda, x, z) - \sup_{\lambda\in\Lambda_0^-} \inf_{x\in B_\lambda^0} l_T(\lambda, x, 0)$$

$$= o_p(1), \text{ as required.}$$

This now completes the proof of Theorem 4.1. □

Proof of Theorem 4.2: 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 4.3: Under the alternative hypothesis, $\Lambda_0^+ \cup \Lambda_0^\approx \neq \varnothing$. Let

$$\widehat{S}_{T,b}^0(w) = \frac{1}{T - b + 1} \sum_{t=1}^{T-b+1} 1(b^{-1/2} d_{b;T,t} \leq w)$$

$$S_b^0(w) = \Pr(b^{-1/2} d_{T,b,1} \leq w).$$

Using the inequality of Bosq (1998, Theorem 1.3) and Assumption 4.3 (see also LMW, Proof of Theorem 2), we can establish the uniform convergence result:

$$\sup |\widehat{S}_{T,b}^0(w) - S_b^0(w)| \xrightarrow{p} 0. \tag{A.13}$$

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society.

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

Therefore, (A.13) and the pointwise convergence result $b^{-1/2}d_{T,b,1} \xrightarrow{p} d_*(0)$ yield

$$s^0_{T,\widehat{b}_T}(\alpha) = \inf\{w : \widehat{S}^0_{T,b}(w) \geq \alpha\} \to d_*(0) \geq 0, \tag{A.14}$$

where $d_*(\cdot)$ is defined in (2.3). Note that $d_*(0)$ is strictly positive if $\Lambda_0^+ \neq \varnothing$, while $d_*(0) = 0$ if $\Lambda_0^+ = \varnothing$ and $\Lambda_0^{\approx} \neq \varnothing$. Therefore,

$$\begin{aligned}
&\Pr(d_T > s_{T,\widehat{b}_T}(\alpha)) \\
&\geq \Pr(\sup_{\lambda\in\Lambda_0} \inf_{x\in B_\lambda^{\epsilon_T}} (\nu_T(\lambda, x) + T^{1/2}(G_Y(x) - G_\lambda(x))) > \widehat{b}_T^{1/2} s^0_{T,\widehat{b}_T}(\alpha)) + o(1) \\
&\geq \Pr(\sup_{\lambda\in\Lambda_0} \inf_{x\in B_\lambda^{\epsilon_T}} (\nu_T(\lambda, x) + T^{1/2}(G_Y(x) - G_\lambda(x))) > u_T^{1/2} s^0_{T,\widehat{b}_T}(\alpha)) + o(1) \\
&\geq \Pr(\sup_{\lambda\in\Lambda_0} \inf_{x\in B_\lambda^{\epsilon_T}} (u_T^{-1/2}\nu_T(\lambda, x) + (T/u_T)^{1/2}(G_Y(x) - G_\lambda(x))) > d_*(0)) + o(1)
\end{aligned} \tag{A.15}$$

where the first inequality holds by Lemma A.2, the second inequality holds by Assumption 4.3, and the last inequality holds by (A.14). Now, consider the right-hand side of (A.14). Note that

$$u_T^{-1/2}\nu_T(\lambda, x) \xrightarrow{p} 0 \tag{A.16}$$

uniformly in λ and x using Lemma A.1. Also,

$$\underline{\lim}_{T\to\infty} \sup_{\lambda\in\Lambda_0} (T/u_T)^{1/2}\Delta_\lambda(\epsilon_T) > d_*(0), \tag{A.17}$$

because, if $\Lambda_0^+ \neq \varnothing$, $\lim_{T\to\infty} \Delta_\lambda(\epsilon_T) = d_*(0) > 0$ $\forall\lambda \in \Lambda_0^+$ and $\underline{\lim}_{T\to\infty}(T/u_T)^{1/2} > 1$ by Assumption 4.4 and, if $\Lambda_0^+ = \varnothing$ and $\Lambda_0^{\approx} \neq \varnothing$, $\underline{\lim}_{T\to\infty} \sup_{\lambda\in\Lambda_0^{\approx}}(T/u_T)^{1/2}\Delta_\lambda(\epsilon_T) > 0 = d_*(0)$ by Assumption 4.4. Therefore, (A.15), (A.16), and (A.17) imply that

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

as required. □

Proof of Theorem 4.4: Define the empirical process in $(\lambda, z, x) \in \Lambda_0^{=} \times \mathcal{Z} \times \mathcal{X}$ to be

$$\nu^*_T(\lambda, z, x) = \sqrt{T}(\widehat{G}_Y(x) - \widehat{G}_{T,\lambda(1+z)}(x) - G_Y(x) + G_{\lambda(1+z)}(x)),$$

where $\mathcal{Z}$ is a compact set containing zero and $G_Y(x) - G_{\lambda_T}(x) = G_Y(x) - G_{\lambda(1+c/\sqrt{T})}(x)$ satisfies the local alternative hypothesis (4.4). Similarly to Lemma A.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, because

$$Q_T(\lambda_T, x) = \nu^*_T(\lambda, c/\sqrt{T}, x) + \delta_\lambda(x), \tag{A.18}$$

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society.

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

we have

$$\sup_{\lambda_T \in \Lambda_{0T}} \inf_{x \in \widehat{B}_\lambda^{\epsilon_T}} Q_T(\lambda_T, x) - \sup_{\lambda \in \Lambda_0^=} \inf_{x \in \mathcal{X}} (\nu_T^*(\lambda, 0, x) + \delta_\lambda(x))$$

$$= \sup_{\lambda \in \Lambda_0^=, c/\sqrt{T} \in \mathcal{Z}} \inf_{x \in \mathcal{X}} (\nu_T^*(\lambda, c/\sqrt{T}, x) + \delta_\lambda(x)) - \sup_{\lambda \in \Lambda_0^=} \inf_{x \in \mathcal{X}} (\nu_T^*(\lambda, 0, x) + \delta_\lambda(x)) \tag{A.19}$$

$$= o_p(1), \tag{A.20}$$

where (A.19) holds $wp \to 1$ because $\Pr(\widehat{B}_\lambda^{\epsilon_T} = \mathcal{X}) \to 1$ for $\lambda \in A_0^=$ by Lemma A.2 and (A.20) holds by the stochastic equicontinuity of $\{\nu_T^*(\cdot, \cdot, \cdot) : T \geq 1\}$. Now, the result of Theorem 4.4 holds by the weak convergence of $\nu_T^*(\cdot, 0, \cdot) + \delta_\lambda(\cdot)$ to $\widetilde{\nu}(\cdot, \cdot) + \delta_\lambda(\cdot)$ and continuous mapping theorem. □

Downloaded from academic.oup.com/ectj/article/17/2/S59/5056439 by Koc University Law School - IHRL Working Group user on 06 November 2026

© 2013 The Author(s). The Econometrics Journal © 2013 Royal Economic Society.