Evidence of uniform inefficiency in market portfolios based on dominance tests

18th April 2020

April 2020

Abstract

We find stochastic uniform inefficiency of many widely held (active) portfolios and fund strategies. Uniformity provides robust findings over general classes of utility (loss) functions and unknown distribution of returns. Evidence is based on statistical tests for the null of stochastic uniform inefficiency of a given portfolio. The alternative is that there is at least one portfolio that dominates it. We derive an analytical characterization of stochastic uniform inefficiency. We give the limit distribution for the empirical test statistic, and present a practical implementation with block bootstrap for consistent estimation of p-values. Our test is asymptotically exact and performs well in Monte Carlo experiments.

JEL subject codes: C12,C14,D81,G11.

Keywords: Stochastic dominance; Uniform inefficiency; Contact set; Gaussian process; Block bootstrap; Mutual funds; Optimal portfolio.

1 Introduction

Mutual funds have been one of the fastest growing financial intermediaries in the US. The industry has grown in size to 18 trillion dollars and attracts over 40 percent of US households as investors.[1] There is an ongoing debate about whether actively managed mutual fund managers outperform the passive benchmarks.

This paper's evidence of widespread, uniform, stochastic inefficiency of many important and active funds is rigorous, and remarkable in its robustness relative to any decision maker utility function, and not based on any arbitrary assumptions on the distribution of returns.

Previous literature has proposed various benchmark models for mutual fund performance evaluation.[2] Mutual funds disclose their benchmarks in their prospectus. However, some mutual funds may drift from their disclosed benchmarks[3] and this fact makes it challenging to use the disclosed benchmarks to evaluate a fund's performance. Berk and Van Binsbergen (2015) adopt Vanguard index funds as the passive benchmark and use the dollar value that a mutual fund extracts from capital market as the measure of skill. Crane and Crotty (2018) examined the distribution of index funds alpha and active managed mutual funds alpha. They find that the cross section of index fund alpha (second-order stochastic) uniformly dominates that of active funds.

In this paper, we turn to an idea that is intuitive and, effectively, dates back at least to Malkiel (1995). We consider (passive) mutual funds as the opportunity cost of actively managed funds, and examine whether a particular mutual fund could stochastically dominate a portfolio of (passive) mutual funds. We benchmark fund managers against three types of opportunity sets. First, we include all the equity mutual funds from three well-known mutual fund families (Vanguard, Fidelity, and American) in the opportunity set. Second, we use various market indexes as the opportunity sets. Lastly, following Berk and Van Binsbergen

[1]See the 2019 Investment Company Fact Book at https://www.ici.org/pdf/2019_factbook.pdf

[2]Carhart (1997) uses common factors in stock returns to explain mutual fund performance. Daniel et al. (1997) develop performance measures which use benchmarks based on the characteristics of stocks held by the portfolios that are evaluated.

[3]See for example, Wermers (2012) and Cao et al. (2017)

(2015), we use Vanguard index funds as the benchmark. Investors can invest in these index funds with relatively low transactions cost. Another potential benchmark could be factor portfolios. However, factor portfolios do not account for transaction costs and it is challenging for investors to buy certain factor portfolios. For example, momentum strategy requires high turnover and momentum index funds does not exist. Thus, we adopt Vanguard index funds, market indexes, and mutual funds in three well-known mutual fund families as the benchmark. We test for stochastic uniform inefficiency of a specific fund with respect to all possible portfolios constructed within these three opportunity sets.

Stochastic dominance (SD) ranking has been employed to compare the performance of investment portfolios. The best-known tests for SD relationships are between a pair of alternatives. Moreover, recently developed tests apply under more general conditions than pairwise tests. These tests establish if a given portfolio is SD efficient relative to all portfolios. Specifically, in applications of portfolio theory and asset pricing theory, a given portfolio is tested for versions of SD efficiency with respect to all possible portfolios constructed from a set of financial assets (Kuosmanen (2004); Post (2003)). Post and Versijp (2007), Scaillet and Topaloglou (2010) and Linton et al. (2014) have recently proposed tests of this hypothesis as well as methods of inference. Such tests are robust alternatives to the existing Mean-Variance (M-V) portfolio efficiency tests (e.g., Gibbons et al. (1989)), especially since the return distribution is unknown, may be skewed and fat tailed, and the preferences may be more complex than the "quadratic" utility assumption that underlies the M-V assessments. The model-free nature of SD tests is advantageous in this application area, because decision makers disagree about the relevant shape of utility functions of investors and the probability distribution of stock returns.

Post (2016) develops SD optimality tests and uses them to test whether any combination of small cap stocks dominates others. Gibbons et al. (1989); Post and Poti (2016) propose SD tests and analyze the efficiency of a passive stock market index relative to five different sets of benchmark portfolios used in empirical asset pricing. Hodder et al. (2015) construct stock portfolios that are second order stochastically efficient over the CRSP all-share index and evaluate their out-of-sample performance assuming that asset returns are iid. Daskalaki et al. (2017) use first and second SD efficiency methods to test whether commodities provide diversification benefits compared with stocks and bonds.

In the above-mentioned testing procedures, there are differences in the way the null hypothesis is formulated, the type of test statistic employed, the ability of the test to handle dependent observations and correlated samples, and the approach to computing p-values.

The most common approach to test for SD is to posit the null hypothesis of dominance. However, rejecting the null of SD by the first distribution does not imply SD by the other distribution, since it can also happen that the test fails to rank those distributions. However, rejecting the null of stochastic non-dominance delivers an unambiguous result of SD. Davidson and Duclos (2013) suggest that it would be "reasonable" to reverse the null hypothesis of SD to stochastic non-dominance in pairwise tests.[4]

The simplest way for testing pairwise stochastic non-dominance was originally proposed by Kaur et al. (1994) for continuous distributions, and a similar test was proposed in an unpublished article by Howes (1993) for discrete distributions.

In this paper we develop a SD methodology where the null is stochastic uniform inefficiency (as defined precisely below). With a null of inefficiency, rejection is unambiguous since the alternative is uniform efficiency (with respect to SD of a certain order).

Most of the tests proposed in the literature do not explicitly account for dependence in the data, such as autocorrelation in returns or GARCH effects in volatility. Exception are the papers of Davidson and Duclos (2013) and Linton et al. (2005) (for pairwise comparisons) and Scaillet and Topaloglou (2010) and Linton et al. (2014) (for the portfolio case). Our proposed test accommodates dependent samples and time-dependent data. This is important since assets in different portfolios are correlated (or even the same).

[4] Original papers by McFadden (1989) and similar tests by Linton et al. (2005) were constructed to conduct a bi-directional test as a remedy. These tests also considered testing for a "maximal set"' of prospects in which no prospect would be dominant.

Our statistical methodology is in the spirit of Linton et al. (2014) who develop a consistent test for second order SD efficiency, extending the work of Linton et al. (2005) to the portfolio case. This required a substantial modification of the procedure presented in Linton et al. (2005) because of a boundary problem that will be discussed below. Linton et al. (2014) relied on the theory of estimated sets (Chernozhukov et al. (2007)) to evaluate the supremum in the test statistic over the complement of a small enlargement of the "contact set" defined by stochastic non-dominance. Our approach generalizes Davidson (2009) and Davidson and Duclos (2013) to accommodate composite hypotheses such as stochastic uniform inefficiency. Hence, our contribution is also germane to econometric theory.

By extending the results presented in the aforementioned literature, we analytically characterize stochastic uniform inefficiency, define the test statistic and derive its limit distribution under the composite null hypothesis of stochastic uniform inefficiency. Moreover, for the approximation of the asymptotic critical values we contribute to the previous literature by describing a block bootstrap-based test that relies on the estimation of the contact set (Linton and Whang (2010)) and that is asymptotically exact and consistent. We perform Monte Carlo experiments to evaluate the finite sample size and power of the tests under the conditional heteroskedasticity framework of Arvanitis and Topaloglou (2017). Finally, we discuss computational aspects of mathematical programming formulations corresponding to the proposed test statistic and, using a finite approximation, we are able to design a feasible and easy to follow numerical algorithm for the testing procedure.

Our empirical analysis poses a substantial empirical question of wide interest in financial economics and management. We test whether a given (benchmark) fund is stochastically uniformly inefficient to a degree of statistical confidence, or whether we could construct a portfolio of funds that dominate it. The results are rigorous, robust and unambiguous. There is strong evidence of stochastic uniform inefficiency for numerous widely held portfolios and well-established funds in the US, including many "active" portfolios and strategies. To emphasize, uniformity of the rankings, and the very high p-values, suggest that our findings are not fragile with respect to the choice of utility functions, the undelying asset distributions, or specific critical values.

The paper is organized as follows. In Section 2, we define stochastic non-dominance and stochastic uniform inefficiency. We provide a characterization of stochastic uniform inefficiency by properties of suprema of appropriate functions and we use this characterization to describe the relevant statistical hypothesis. We obtain the test statistic as the appropriately scaled empirical analogue of the aforementioned suprema and we then suggest a modification that properly characterizes the null hypothesis, keeping away from the boundary points. Under relevant assumptions, we obtain the limit theory of the employed statistic under the null hypothesis. We design the testing procedure based on approximations of the asymptotic p-values by a block bootstrap method and establish consistency. In Section 3, we provide an approximation of the test that is numerically implementable. In Section 4, we perform Monte Carlo experiments to evaluate the size and power of the test in a framework of conditionally heteroskedasticity. Section 5 describes the main empirical findings of the paper. We conclude in Section 6 and provide the proofs in the Appendix.

2 Stochastic non-dominance and uniform inefficiency

2.1 Concepts and hypothesis structure

Consider $\Delta \equiv \{\lambda \in \mathbb{R}_+^N : \mathbf{1}'\lambda = 1\}$ a general portfolio possibilities set that consists of all convex combinations of N base assets, where the base assets are not restricted to be individual securities. Suppose that F is the joint cumulative density function of the investment returns to the base assets with closed and convex support and let $F(z, \lambda)$ denote $\int_{\mathbb{R}^N} \mathbb{I}\{\lambda' u \leq z\} dF(u)$, the marginal cdf of the linear portfolio $\lambda \in \Delta$. The union of the support of the N marginal distributions is denoted by $\mathcal{Z}$.

The characterization of the notions of second order stochastic non-dominance and uniform inefficiency make use of the following integrated cdf,

$$\mathcal{J}(z,\lambda,F) \equiv \int_{-\infty}^{z} F(u,\lambda)\,du,$$

which corresponds to the expected shortfall (or the first-order lower-partial moment) for arbitrary $\lambda \in \Delta$ and for return threshold $z \in \mathcal{Z}$.

We will consider $\Lambda \subseteq \Delta$ a convex subset of the general portfolio set reflecting any possible further restrictions on Δ. Let $D(z,\kappa,\lambda,F) \equiv \mathcal{J}(z,\kappa,F) - \mathcal{J}(z,\lambda,F)$, the spread of the first-order lower-partial moments between two portfolios $\lambda, \kappa \in \Lambda$ where κ is to be tested for the relevant notion of uniform inefficiency with respect to the members of Λ. Given this, we have the following definitions for second order stochastic non-dominance and uniform inefficiency that are we are occupied with. (Stochastic non-dominance): *Portfolio κ does not strictly second order stochastically dominate portfolio* λ, say $\kappa \not\succ_F \lambda$, iff

$$\exists z \in \mathcal{Z}:\ D(z,\kappa,\lambda,F) > 0 \,\text{or}\, \forall z \in \mathcal{Z}:\ D(z,\kappa,\lambda,F) = 0.$$

The convexity assumption of Λ allows for an equivalent formulation in terms of Expected Utility. Strict second order stochastic non-dominance holds iff λ achieves a higher expected utility for some non-decreasing and concave utility function or achieves the same expected utility for every non-decreasing and concave utility function. Otherwise said strict stochastic non-dominance holds iff λ is strictly preferred to κ by some risk averter, or every risk averter is indifferent between them; see Levy (1992).

Strict (non-) dominance is the irreflexive part of the preorder and weak orders of stochastic (non-)dominance are obtained by eliminating the requirement of strict inequality at some point.

The related concept of stochastic uniform inefficiency applies the pairwise stochastic non-dominance concept to every feasible alternatives in Λ. (Stochastic uniform inefficiency):

Portfolio κ is *second order stochastically uniformly inefficient over* Λ*, iff*

$$\exists \lambda \in \Lambda, \exists z \in \mathcal{Z} : \; D\left(z, \kappa, \lambda, F\right) > 0 \, \mathrm{or} \, \forall z \in \mathcal{Z} : \; D\left(z, \kappa, \lambda, F\right) = 0.$$

Or equivalently κ is *second order stochastically uniformly inefficient* iff there is some $\lambda \in \Lambda$ such that λ is not strictly second order stochastically dominated by κ. Although the distinction betwen strong and weak dominance is generally not robust in empirical applications when comparing two distinct portfolios, it is relevant for SD efficiency tests. However, the hypothesis structure that will be presented below is with respect to strict SD to avoid the problems that arise when the null hypothesis does not contain elements of the boundary between the two composite hypotheses. Under the null hypothesis of weak uniform inefficiency we may end up with test statistics that have low power properties because of an asymptotic non-tightness near the boundary (see Section 2.4 below).

Definition 2 essentially constitutes a hypothesis for inefficiency. Testing the null hypothesis of inefficiency can be very interesting in the framework of statistical inference since efficiency is a logically cumbersome hypothesis, and tests for our proposed alternative hypothesis can have better properties.

The general strategy is to test the null hypothesis H_0 that portfolio κ is *second order stochastically uniformly inefficient* according to Definition 2 in the sense that there exists a portfolio in Λ that is not strictly second order stochastically dominated by κ against the general alternative hypothesis H_A (the negation of H_0) that is the evaluated portfolio is efficient in the sense that there is no other element in Λ that strictly second order stochastically dominates it.

As a measure of deviations from second order stochastic uniform inefficiency we consider a modification of the population functional used in Linton et al. (2014) to test for stochastic efficiency. This functional satisfies ≥ 0 under the null hypothesis, it can be readily empirically approximated and is going to be employed for our testing procedure,

$$\xi\left(\kappa, F\right) \equiv \sup_{\lambda \in \mathit{\Lambda}} \sup_{z \in \mathcal{Z}} D\left(z, \kappa, \lambda, F\right). \tag{1}$$

However, similarly to the case of efficiency testing of Linton et al. (2014) the above functional does not allow to obtain a consistent test given that it assumes the value of zero under both the null hypothesis and its negation. To see this for each $\lambda \in \mathit{\Lambda}$, define the three following subsets of $\mathcal{Z}$ (Linton et al. (2014)),

$$\mathcal{Z}_{\lambda}^{-} = \left\{z \in \mathcal{Z} \,:\, D\left(z, \kappa, \lambda, F\right) < 0\right\},$$

$$\mathcal{Z}_{\lambda}^{=} = \left\{z \in \mathcal{Z} \,:\, D\left(z, \kappa, \lambda, F\right) = 0\right\},$$

$$\mathcal{Z}_{\lambda}^{+} = \left\{z \in \mathcal{Z} \,:\, D\left(z, \kappa, \lambda, F\right) > 0\right\}.$$

Notice that under the null hypothesis that $\kappa \not\succ_F \lambda$ we have that both $\mathcal{Z}_{\lambda}^{+} \neq \emptyset$ and $\mathcal{Z}_{\lambda}^{-} = \emptyset$. But if also $\mathcal{Z}_{\lambda}^{=} \neq \emptyset$ holds then $\sup_z D\left(z, \kappa, \lambda, F\right) = 0$. Thus, the supremum over the entire support in (1) will fail to distinguish between the null and the alternative hypothesis on the boundary. To allow for the separation between the two hypotheses keeping away from the boundary points and following Linton et al. (2014), for arbitrary $\delta > 0$ we define the δ-enlargement of the set $\mathcal{Z}_{\lambda}^{=}$ as

$$\left(\mathcal{Z}_{\lambda}^{=}\right)^{\delta} = \left\{z + \eta \in \mathcal{Z} :\, z \in \mathcal{Z}_{\lambda}^{=} \text{ and } |\eta| < \delta\right\}. \tag{2}$$

Then the modified population functional to be employed for our testing procedure is

$$\xi\left(\delta, \kappa, F\right) \equiv \sup_{\lambda \in \Lambda} \sup_{z \in \mathrm{A}_{\lambda}^{\delta}} D\left(z, \kappa, \lambda, F\right), \tag{3}$$

where the supremum is taken overA_{λ}^{δ} , the complement of $\left(\mathcal{Z}_{\lambda}^{=}\right)^{\delta}$ in $\mathcal{Z}$, i.e.

$$\begin{cases} \mathcal{Z} \backslash \left(\mathcal{Z}_{\lambda}^{=}\right)^{\delta}, & \text{if } \mathcal{Z}_{\lambda}^{=} \neq \mathcal{Z} \\ \mathcal{Z}, & \text{if } \mathcal{Z}_{\lambda}^{=} = \mathcal{Z} \end{cases}.$$

Given the continuity and the integrability of F under the null hypothesis of uniform inefficiency, $\xi(\delta, \kappa, F) \geq 0$ for each $\delta \geq 0$, while under the alternative $\xi(\delta, \kappa, F) < 0$ for some $\delta > 0$.

2.2 Statistical tests

In practice, generally F and A_{λ}^{δ} are unknown and have to be estimated from the data. Hence, the results above cannot usually be directly employed but given the availability of statistical information we may construct analogous testing procedures.

To this end, consider a process $(Y_t)_{t \in \mathbb{Z}}$ taking values in $\mathbb{R}^N$ which in a financial framework represents the returns of N financial base assets upon which portfolios can be constructed via convex combinations. Consider the following auxiliary sequences (c_T) and (δ_T) where $c_T = c_0 \sqrt{\frac{\log T}{T}}$, with $c_0 > 0$, and where δ_T is a sequence of positive constants satisfying $\lim_{T \to \infty} \delta_T = 0$ while $\delta_T > c_T$, $\forall T \geq 1$ and let $\hat{F}$ denote the *empirical* cdf associated with the random element $(Y_t)_{t=1,\ldots,T}$.

If we define the following empirical associated sets,

$$\hat{\mathcal{Z}}_{\lambda}^{=} = \left\{ z \in \mathcal{Z} : \left| D\left(z, \kappa, \lambda, \hat{F}\right) \right| \leq c_T \right\},$$

$$\left(\hat{\mathcal{Z}}_{\lambda}^{=}\right)^{\delta_T} = \left\{ z + \eta \in \mathcal{Z} \,:\, z \in \hat{\mathcal{Z}}_{\lambda}^{=} \text{ and } |\eta| < \delta_{\mathrm{T}} \right\},$$

and $\hat{A}_{\lambda}^{\delta_T}$ as

$$\hat{A}_{\lambda}^{\delta_T} = \begin{cases} \mathcal{Z} \backslash \left(\hat{\mathcal{Z}}_{\lambda}^{=}\right)^{\delta_T}, & \text{if } \hat{\mathcal{Z}}_{\lambda}^{=} \neq \mathcal{Z} \\ \mathcal{Z}, & \text{if } \hat{\mathcal{Z}}_{\lambda}^{=} = \mathcal{Z} \end{cases},$$

then the test statistic for stochastic uniform inefficiency is obtained as the appropriately scaled empirical analogue of the functional appearing in (3)

$$\xi_T \equiv \sup_{\lambda\in\Lambda} \sup_{z\in\hat{A}_{\lambda}^{\delta_T}} D\left(z,\kappa,\lambda,\sqrt{T}\hat{F}\right). \tag{4}$$

The underlying optimizations are generally analytically intractable, and we resort to numerical techniques. First, we derive the asymptotic distribution of the test statistic under the null hypothesis, which will facilitate the design of the testing procedure.

2.3 Null distribution

As remarked in Linton et al. (2014) the following assumption is needed for the subsequent derivation of the null distribution of the test statistic.

Assumption 1. *For any $\lambda \in \Lambda$ for which $\mathcal{Z}_{\lambda}^{=} \neq \emptyset$, for all $C > 0$, $\forall z \in \mathcal{Z}$,we have that for T sufficiently large,*

$$\left|D\left(z,\kappa,\lambda,F\right)\right| \geq C \min\left\{\inf_{z'\in\mathcal{Z}_{\lambda}^{=}} \left|z'-z\right|, \delta_T\right\}.$$

This is the modification of the relevant part of Assumption 4.2 of Linton et al. (2014) in our framework and it enforces a monotonicity property to $D\left(\cdot,\kappa,\lambda,F\right)$ on the boundary of $\mathcal{Z}_{\lambda}^{=}$. It holds for example when $D\left(\cdot,\kappa,\lambda,F\right)$ is (potentially one-sided) differentiable on the boundary (which holds given the assumption below) with non-vanishing derivative there.

The following assumption enables the validity of a mixing CLT while it is more general than Assumption 4.1 of Linton et al. (2014) since it does not require compactness for the support (see Arvanitis (2015)).

Assumption 2. *For some $0 < \varepsilon$, $\mathbb{E}\left[\left\|Y_0\right\|^{2+\varepsilon}\right] < +\infty$. $\left(Y_t\right)_{t\in\mathbb{Z}}$ is stationary and strongly mixing with mixing coefficients $a_T = O(T^{-a})$ for some $a > 1 + \frac{2}{\epsilon}$, $0 < \epsilon < 2$, as $T \to \infty$.*

Furthermore,

$$\mathbb{V} = \mathbb{E}\left[(Y_0 - \mathbb{E}Y_0)(Y_0 - \mathbb{E}Y_0)^T\right] + 2\sum_{t=1}^{\infty}\mathbb{E}\left[(Y_0 - \mathbb{E}Y_0)(Y_t - \mathbb{E}Y_t)^T\right]$$

is positive definite.

The mixing part holds for many stationary models used in the context of financial econometrics. Prominent examples are the strictly stationary versions of (possibly multivariate) ARMA or several GARCH and stochastic volatility type of models. The moment existence condition also holds in models such as the aforementioned. The positive definiteness of the "long-run" covariance matrix is satisfied, for isntance, if $(Y_t)_{t\in\mathbb{Z}}$ is a vector martingale difference process and the elements of Y_0 are linearly independent random variables.

Consider the following subsets of the parameter space Λ,

$$\Lambda^{=} \equiv \{\lambda \in \Lambda : D(z, \kappa, \lambda, F) = 0, \forall z \in \mathcal{Z}\}.$$

Notice that $\Lambda^{=} \neq \emptyset$ since $\Lambda^{\star} \equiv \{(z, \kappa, \kappa), \kappa \in \Lambda, z \in \mathbb{R}\} \subseteq \Lambda^{=}$.

The following proposition establishes the asymptotic distribution of the test statistic under the null. We denote $\rightsquigarrow$ for convergence in distribution as $T \to \infty$. Suppose that Assumptions 1 and 2 hold and that H_0 is true. Then,

$$\xi_T \rightsquigarrow \xi_\infty,$$

$$\xi_\infty \equiv \sup_{\lambda\in\Lambda^{=}} \sup_{z\in\mathcal{Z}} D(z, \kappa, \lambda, \mathcal{B}_F),$$

where $\mathcal{B}_F$ is a well defined centered Gaussian process with uniformly continuous sample *paths defined on* $\mathbb{R}^N$ *and covariance kernel given by*

$$Cov(\mathcal{B}_F(x), \mathcal{B}_F(y)) = \sum_{t \in \mathbb{Z}} Cov\left(\mathbb{I}_{Y_0 \leq x}, \mathbb{I}_{Y_t \leq y}\right).$$

When $\Lambda^{=} = \{\kappa\}$, then obviously the distribution of ξ_∞ is degenerate at zero.

When the distribution of ξ_∞ is not degenerate it is possible that it has a discontinuous quantile function $q\left(\xi_\infty, 1-\alpha\right)$. Lemma 6 implies that such a discontinuity can occur only at zero and that the $1-\alpha$ quantile is a continuity point for the cdf under the condition $\underline{\lim}_{T \to \infty} \mathbb{P}\left(q\left(\xi_T, 1-\alpha\right) < -\delta\right) = 1$, for some $\delta > 0$.

Given the asymptotic null distribution we can develop a test procedure where the basic decision rule is to reject H_0 against H_A iff $\xi_T < q\left(\xi_\infty, 1-\alpha\right)$ for any significance level $(1-\alpha)$. Since the law of ξ_∞ depends on the generally unknown F via the covariance kernel of $\mathcal{B}_F$ and the aforementioned partition of Λ, the result cannot be directly used for the construction of the asymptotic rejection region. Nevertheless, a feasible decision rule can be established by the use of some resampling procedure.

2.4 Simulations with block bootstrap

The difficulty for a bootstrap test flows from an inability to impose the (composite) null hypothesis. The common practice of considering only the "least favorable" boundary configuration to bound the size of the test leads to the problem of nonsimilarity on the boundary (see Linton et al. (2005) and Linton et al. (2014)). The usual recentering method imposes restrictions that may not hold outside the least favourable case, and hence the bootstrap-based test is biased. Instead, we employ the method proposed by Linton and Whang (2010) to obtain asymptotically valid critical values by using the estimated boundary points. Weak dependence is accounted for with block bootstrap, as in the relevant framework of Scaillet and Topaloglou (2010) and Arvanitis and Topaloglou (2017).[5]

[5]See, for example, Linton et al. (2014) and Arvanitis (2015) for discussions on the behavior of procedures involving subsampling in similar frameworks.

Let b_T,l_T denote integers such that $T = b_T l_T$ where b_T is the number of blocks and l_T is the block size. As in Arvanitis and Topaloglou (2017) we consider only the non-overlapping case and we make the following assumption concerning the choice of l_T.

Assumption 3. *For some* $0 < q < \frac{1}{3}$ *and for some* $0 < h < \frac{1}{3} - q$ *, as* $T \to \infty$*,* l_T *satisfies*

$$T^h \ll l_T \ll T^{\left(\frac{1}{3}-q\right)}.$$

In what follows, let $(Y_t^\star)_{t=1,\ldots,T}$ denote the bootstrap sample in the context of the non-overlapping blocks methodology, and let $\hat{F}^\star$ denote its empirical distribution. Denote by $\mathbb{P}_T^\star$ the relevant probability distribution that represents the law of $(Y_t^\star)_{t=1,\ldots,T}$ conditional on $(Y_t)_{t=1,\ldots,T}$.

Given a sequence $k_T \to 0$ and $k_T\sqrt{T} \to \infty$ where $k_T > c_T$, we construct an estimated contact set

$$\hat{\Lambda}^{=} \equiv \left\{\lambda \in \Lambda : \left|D\left(z, \kappa, \lambda, \hat{F}\right)\right| < k_T, \forall z \in \mathcal{Z}\right\}.$$

Then the bootstrap test statistic is defined by

$$\xi_T^\star \equiv \sup_{\lambda\in\hat{\Lambda}^{=}} \sup_{z\in\mathcal{Z}} \left(D\left(z, \kappa, \lambda, \sqrt{T}\left(\hat{F}^\star - \hat{F}\right)\right)\right),$$

if $\hat{\Lambda}^{=} \neq \{\kappa\}$ and by

$$\xi_T^\star \equiv \pi_T,$$

if $\hat{\Lambda}^{=} = \{\kappa\}$, where π_T is a sequence of negative numbers such that $\pi_T \to -\infty$ (Linton and Whang (2010)).

The bootstrap critical value of significance level α is then calculated as

$$h_T^\star(1-\alpha) = \inf_y \left\{\frac{1}{M}\sum_{i=1}^{M} \mathbb{I}\left(\xi_{\mathrm{T,i}}^\star \leq \mathrm{y}\right) \geq 1-\alpha\right\},$$

and we reject the null hypothesis iff $\xi_T < h_T^\star(1-\alpha)$.

However, the p-values $\mathbb{P}_T^\star(\xi_T^\star > \xi_T)$ are usually analytically intractable and approximated by an empirical frequency argument. Based on $R \geq 1$ bootstrap samples $\left\{\left(Y_{t,r}^\star\right)_{t=1,\ldots,T}\right\}_{r=1,\ldots,R}$, the approximated p-values are given by

$$\hat{p}_T^\star = \frac{1}{R}\sum_{r=1}^{R}\mathbb{I}\{\xi_{T,r}^\star > \xi_T\}.$$

The next proposition states the asymptotic size and power properties of the testing procedure. Suppose that Assumptions 1, 2 and 3 hold. Then for the boostrap test described we have that,

1. If $\mathbf{H_0}$ is true then,

$$\lim_{\mathbb{T}\to\infty}\mathbb{P}\left(\xi_T < h_T^\star(1-\alpha)\right) \leq \alpha.$$

   (a) If $\mathbf{H_0}$ is true, $\Lambda^= \neq \{\kappa\}$ and then for some $\delta > 0$, $\underline{\lim}_{T\to\infty}\mathbb{P}\left(q\left(\xi_T, 1-\alpha\right) < -\delta\right) = 1$ then,

$$\lim_{\mathbb{T}\to\infty}\mathbb{P}\left(\xi_T < h_T^\star(1-\alpha)\right) = \alpha.$$

   (b) If $\mathbf{H_a}$ is true then,

$$\lim_{\mathbb{T}\to\infty}\mathbb{P}\left(\xi_T < h_T^\star(1-\alpha)\right) = 1.$$

The first result establishes that the bootstrap tests have asymptotically correct sizes. The second result states that the bootstrap test has asymptotically exact size on the boundary under the stated restrictions that establish continuity of the limit cdf at the quantile corresponding to the relevant significance level. The third result states consistency under any fixed alternative.

3 Numerical implementation

In this section, we present one numerical method for testing the stochastic uniform inefficiency of a given portfolio $\kappa \in \Lambda$. The portfolio set is assumed to be a convex polytope, to allow for Linear Programming (LP).

For the test statistic, the algorithm can be described by the following steps.

1. Discretize $\hat{\mathcal{Z}}$ using J equally spaced grid points $\hat{z}_j := \hat{A} + (j-1)\left(\hat{B} - \hat{A}\right)(J-1)^{-1}$, $j = 1, .., J$ where $\hat{\mathcal{Z}} := \left[\hat{A}, \hat{B}\right]$, $\hat{A} := \min_{t,\lambda\in\Lambda} y_t^{\mathrm{T}}\lambda$ and $\hat{B} := \max_{t,\lambda\in\Lambda} y_t^{\mathrm{T}}\lambda$, where $J < \frac{\hat{B}-\hat{A}}{\eta}$ so that the condition of (2) holds.

2. Set $\hat{\mathcal{Z}}^{=} = \emptyset$.

3. For $j = 1, \ldots, J$ compute

$$\max_{\Lambda} \left\{ D\left(\hat{z}_j, \kappa, \lambda, \hat{F}\right) \geq c_T \right\} \tag{5}$$

   for $c_T = c_0\sqrt{\frac{\log T}{T}}$, where $c_0 > 0$. If (5) does not give a solution set $\hat{\mathcal{Z}}^{=} = \hat{\mathcal{Z}}^{=} \cup \{\hat{z}_j\}$.

4. Then:

   (a) If $\hat{\mathcal{Z}} \neq \hat{\mathcal{Z}}^{=}$ set $\xi_T = \max_{z\in\hat{\mathcal{Z}}\setminus\hat{\mathcal{Z}}^{=}} \max_{\Lambda} \left\{ D\left(\hat{z}_j, \kappa, \lambda, \hat{F}\right) \geq c_T \right\}$.

   (b) If $\hat{\mathcal{Z}} = \hat{\mathcal{Z}}^{=}$, for $j = 1, \ldots, J$ compute

$$\max_{\Lambda} \left\{ D\left(\hat{z}_j, \kappa, \lambda, \hat{F}\right) \leq -c_T \right\} \tag{6}$$

   using linear relaxations. Set $\xi_T = \max_{z\in\hat{\mathcal{Z}}^{=}} \max_{\Lambda} \left\{ D\left(\hat{z}_j, \kappa, \lambda, \hat{F}\right) \leq -c_T \right\}$.

For each discrete value of $z \in \hat{\mathcal{Z}}^{=}$ we solve the optimization problem for $\lambda \in \Lambda$ via the following LP:

$$\begin{aligned}
\max_{\lambda \in \Lambda} \quad & \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (L_t - W_t) \\
\text{s.t.} \quad & L_t = (z - \kappa' Y_t)_+, \qquad \forall t \in T, \\
& W_t \geq z - \lambda' Y_t, \qquad \forall t \in T, \\
& e' \lambda = 1, \\
& \lambda \geq 0, \quad W_t \geq 0, \qquad \forall t \in T.
\end{aligned} \tag{7a}$$

The test statistic ξ_T is the maximum objective value of the above LP for $z \in \hat{\mathcal{Z}}^=$ that maximizes $D\left(\hat{z}_j, \kappa, \lambda, \hat{F}\right) \geq c_T$, if $\hat{\mathcal{Z}} \neq \hat{\mathcal{Z}}^=$. However, if $\hat{\mathcal{Z}} = \hat{\mathcal{Z}}^=$, then ξ_T is the maximum objective value for $z \in \hat{\mathcal{Z}}^=$ that maximizes $D\left(\hat{z}_j, \kappa, \lambda, \hat{F}\right) \leq c_T$.

For the bootstrap test statistic, the computational strategy is similar. After the discretization step, we set $\hat{\Lambda}^= = \emptyset$ and for $j = 1, \ldots, J$ and we compute

$$\max_{\Lambda} \left\{ D\left(\hat{z}_j, \kappa, \lambda, \hat{F}\right) \geq k_T \right\} \tag{8}$$

using linear relaxations. If (8) gives a solution then we set $\hat{\Lambda}^= = \hat{\Lambda}^= \cup \{\lambda^*\}$ and

$$\xi_T^\star = max_{\lambda \in \hat{\Lambda}^=} max_{z \in \hat{\mathcal{Z}}} \left(D\left(\hat{z}_j, \kappa, \lambda, \sqrt{T}\left(\hat{F^\star} - \hat{F}\right)\right)\right).$$

If (8) does not give a solution then we set $\hat{\Lambda}^= = \emptyset$ and for $j = 1, \ldots, J$ we compute

$$\max_{\Lambda} \left\{ D\left(\hat{z}_j, \kappa, \lambda, \hat{F}\right) \leq -k_T \right\} \tag{9}$$

using linear relaxations. If (9) gives a solution then we set $\hat{\Lambda}^= = \hat{\Lambda}^= \cup \{\lambda^*\}$ and

$$\xi_T^\star = max_{\lambda \in \hat{\Lambda}^=} max_{z \in \hat{\mathcal{Z}}} \left(D\left(\hat{z}_j, \kappa, \lambda, \sqrt{T}\left(\hat{F^\star} - \hat{F}\right)\right)\right).$$

If (9) does not give a solution then we set

$$\xi_T^\star = \pi_T.$$

Again, to compute $\xi_T^\star$ we solve the linear optimization model (7). [6]

The p-value is approximated by $\tilde{p}_j = \frac{1}{R}\sum_{r=1}^{R}\{\xi_{T,r}^\star > \xi_T\}$, where the averaging is taken over R replications. Note that the replication number can be chosen to make the approximations as accurate as one desires given time and computer constraints. Generally the null hypothesis would be rejected if the p-value is lower that 5%.

4 Monte Carlo simulations

In this section, we design a set of Monte Carlo experiments to evaluate the size and power of the proposed test in finite samples. We follow Arvanitis and Topaloglou (2017) and Arvanitis (2015) and use a framework of conditional heteroskedasticity where $(Y_t)_{t=1,\ldots,T}$ process is constructed as GARCH(1,1) process.

Suppose that

$$z_t \overset{\text{iid}}{\sim} N\left(0,1\right), t \in \mathbb{Z}.$$

Furthermore for all $t \in \mathbb{Z}$, for $i = 1,2,3$, $\omega_i, a_i, \beta_i \in \mathbb{R}_{++}$, such that $\mathbb{E}\left[\left(a_i z_0^2 + \beta_i\right)^{1+\delta}\right] < 1$ for some $\delta > 0$, $\mu_i \in \mathbb{R}_+$ define

$$\begin{aligned} y_{it} &= \mu_i + z_t h_{it}^{1/2}, \\ h_{it} &= \omega_i + \left(a_i z_{t-1}^2 + \beta_i\right) h_{it-1}, \end{aligned}$$

[6]We note here that the total run time of this LP is less than a minute on a standard desktop PC with a 2.93 GHz quad-core Intel i7 processor, 16GB of RAM and using GAMS and MATLAB with the Gurobi Optimizer solver.

while for $i = 4$ and $v \in \mathbb{R}$ define

$$y_{4t} = v h_{4t}^{1/2} z_t.$$

Let $Y_t = (y_{1t}, y_{2t}, y_{3t}, y_{4t})'$. Arvanitis and Topaloglou (2017) establish that the vector process above satisfies our assumption framework. Let $\tau = (0,0,1,0)$, $\tau^\star = (0,0,0,1)$ and $\Lambda = \{(\lambda, 1-\lambda, 0, 0), \lambda \in [0,1], \tau, \tau^\star\}$. For this choice of Λ we can easily specify portfolios for which the relevant null and alternative hypotheses are valid. These are verified by Propositions 4 and 5 in Arvanitis and Topaloglou (2017). If $\mu_i = 0$ for $i = 1,2,3$, $\omega_1 < \omega_3$, $a_1 < a_3$ and $\beta_1 < \beta_3$ then τ is inefficient w.r.t. Λ. We use instances of the GARCH processes conforming to Proposition 4 in order to evaluate the empirical size. If $\mu_i = 0$ for $i = 1,2,3$, $\sqrt{\frac{\max\{\omega_i, a_i, \beta_i,\ i=1,2,3\}}{\min\{\omega_i, a_i, \beta_i,\ i=1,2,3\}}} < v < \sqrt{\frac{\min\{\omega_i, a_i, \beta_i,\ i=1,2,3\}}{\max\{\omega_i, a_i, \beta_i,\ i=1,2,3\}}}$ and then $\tau^\star$ is efficient w.r.t. Λ. We use instances of the GARCH processes conforming to Proposition 4 in order to evaluate the empirical power.

Scenarios: We use as DGPs instances of the GARCH processes described above, by choosing the parameters according to Propositions 4 and 4, to approximate the fixed T size and power. For $M = 300$, we generate independent across $i = 1, \ldots, M$ samples $\left(Y_t^{(b)}\right)_{t=1,\ldots,T}$ for several values of T. For each b, we use the non-overlapping block bootstrap methodology described above to evaluate $\hat{p}_T^{\star(b)}$ and decide by choosing $\alpha = 0.05$, and several values of R and l_T. We choose $k_T = kT^{-1/2}$ and $k \in \{3.0, 3.2, ..., 4.0\}$.

Size Evaluation Parameter Selection: To approximate the fixed T size, we test for SSD inefficiency of portfolio τ by setting $\mu_i = 0$ for $i = 1,2,3$, $\omega_1 = 0.5$, $\omega_2 = 0.5, \text{and } \omega_3 = 0.8$, $a_1 = 0.3$, $a_2 = 0.4, \text{and } a_3 = 0.45$ and $\beta_1 = 0.3$, $\beta_2 = 0.4$, $\beta_3 = 0.45$, $v_1 = 2$ and $v_2 = 0.2$. In this case, we have that $\omega_1 < \omega_3$, $a_1 < a_3$ and $\beta_1 < \beta_3$.

Power Evaluation Parameter Selection: To approximate the fixed T power, we test for SSD inefficiency of portfolio τ^* by setting $\mu_i = 0$ for $i = 1,2,3$, $\omega_1 = 0.5$, $\omega_2 = 0.5, \text{and } \omega_3 = 0.5$, $a_1 = 0.4$, $a_2 = 0.45, \text{and } a_3 = 0.5$ and $\beta_1 = 0.5$, $\beta_2 = 0.45$, $\beta_3 = 0.4$, $v_1 =$

1.5 and $v_2 = 0.5$. In this case, we have that $\sqrt{\frac{\max\{\omega_i, a_i, \beta_i,\ i=1,2,3\}}{\min\{\omega_i, a_i, \beta_i,\ i=1,2,3\}}} < v < \sqrt{\frac{\min\{\omega_i, a_i, \beta_i,\ i=1,2,3\}}{\max\{\omega_i, a_i, \beta_i,\ i=1,2,3\}}}$.

Results: We present our Monte Carlo results in Table 1. We investigate cases where l_T ranges from 4 to 12 by a step size of 4, choices motivated by Hall et al. (1995), who suggest an optimal block size of $T^{1/3}$. Our experiments show that the choice of the block size does not seem to dramatically alter the performance of our methodology even for moderately smaller and larger values of T. We also investigate the sensitivity of the tests to the choice of the number of bootstrap samples and sample size by allowing for $(R, T) = (100, 200), (300, 500), (500, 1000)$. The tests seem to perform extremely well in every case.

[Insert Table 1]

5 Widespread portfolio inefficiency

Despite the large size of the mutual fund industry, there has been a debate about whether actively managed mutual fund managers outperform passive benchmarks. Numerous studies have shown that, post fees, mutual funds provide investors with average returns significantly below those of passive benchmarks. In one of the most influential papers on the subject of managers' skill, Carhart (1997) uses the net alpha earned by investors to measure managerial skill, and concludes that there is no evidence of "skilled" mutual fund managers. Notwithstanding these findings on actively managed mutual funds, further refined search for skilled managers has continued. Kacperczyk et al. (2005) find that funds concentrated in industries perform better. Kacperczyk et al. (2008) compare the actual performance of funds with the performance of the funds' beginning of quarter holdings. They conclude that the average fund manager adds value during the quarter. Cremers and Petajisto (2009) find better performance for managers with higher Active Share, which represents the share of portfolio holdings that differ from the benchmark index holdings. Amihud and Goyenko (2013) document that mutual funds with lower R^2 perform better. Kacperczyk et al. (2014)

provide evidence that managers successfully market time in bad times and select stocks in good times. Kosowski et al. (2006), Fama and French (2010), and Barras et al. (2010) try to disentangle skill and luck by using the cross-sectional distribution of active fund performance.

In this section we present robust empirical evidence that many well-known mutual funds are stochastically uniformly inefficient relative to various benchmarks. We consider mutual funds as the opportunity cost of active managements and examine whether a particular fund could stochastically dominate a portfolio of mutual funds. We benchmark fund managers against three types of opportunity sets. First, we include 248 equity mutual funds in three well-known mutual fund families (Vanguard, Fidelity, and American) in the opportunity set. Second, we use various market indexes as the opportunity sets. Lastly, following Berk and Van Binsbergen (2015), we use seven Vanguard index funds as the benchmark funds. We test for stochastic uniform inefficiency of a specific fund with respect to all possible portfolios constructed within these three opportunity sets.

5.1 The dataset

The CRSP survivor-bias-free US mutual fund dataset provides information about fund names, returns, total assets under management (AUM), inception dates, expense ratios and other fund characteristics. Morningstar provides the Morningstar Box, a nine-square grid that summarizes the investment style of a mutual fund. In this study, we limit our analysis to domestic equity mutual funds between March 1980 and March 2018.

Some mutual funds have multiple share classes. The CRSP data lists each share class as a separate fund. Different share classes have the same holding compositions and typically differ only in fee structure. For funds with multiple share classes, we use the identification code in MFLINKS to combine different classes of the same fund into a single value-weighted fund. Wermers (2012) provides a description of how MFLINKS are created. Each monthly fund return is computed by weighting the return of its component share classes by their beginning-of-month total net asset values.

5.2 In-sample analysis

In Table 2, we test for stochastic uniform inefficiency of a specific fund with respect to all possible portfolios constructed by 248 equity funds from Vanguard, Fidelity, and American fund familes. We consider funds in different size groups and style categories and use net return for all the analysis. Table 2 reports the test results for test funds in three different categories (Large Index Fund, Small Actively Managed Fund, and Large Actively Managed Fund). The results in Table 2 show that we cannot reject the null hypothesis that the test funds are stochastically uniformly inefficient. The p-values are far beyond the 5%. These findings are significant with strong implications. Investors and managers could potentially diversify across various funds and outperform the dominated fund, whatever their risk preferences and utility shapes, and whatever the true probability distribution of the underlying returns.

[Insert Table 2]

Next, we test whether some well-known funds are stochastically uniformly inefficient with respect to the market. Market portfolios are represented by the value-weighted portfolio of all CRSP listed stocks, the equally-weighted portfolio of all CRSP listed stocks, as well as the S&P 500 index. The results from Table 3 - Table 5 suggest that we cannot reject the null hypothesis that the test funds are stochastically uniformly inefficient relevant to any of the proxy for the market portfolios. Some investors will benefit from simply holding the market portfolio.

[Insert Table 3]

[Insert Table 4]

[Insert Table 5]

Finally, we test whether each of the fifteen funds are stochastically uniformly inefficient with respect to seven Vanguard index funds. The question here is whether a specific fund is inefficient compared to "any portfolio" that can be constructed from the seven Vanguard index funds. The seven Vanguard indexes we use are the Vanguard S&P500 Index Fund, the Vanguard Extended Market Index Fund, the Vanguard Small-Cap Index Fund, the Vanguard Value Index Fund, the Vanguard Mid-Cap Index Fund, the Vanguard Small-Cap Growth Index Fund and the Vanguard Small-Cap Value Index Fund. The results in Tables 6 suggest that we cannot reject the null hypothesis that the test funds are stochastically uniformly inefficient.[7]

[Insert Table 6]

5.3 Out-of-sample analysis

In this section, we examine whether the 5 Largest Index Funds, the 5 Smallest Active Funds, as well as the 5 Largest Active Funds are dominated by optimal portfolios of the seven Vanguard index funds out-of-sample. Although we do not reject the null hypothesis of stochastic inefficiency of the test funds compared to the seven Vanguard index funds in the in-sample tests, it is not known *a priori* whether the Vanguard index will outperform the test funds in an out-of-sample setting. This is because by construction these portfolios are formed at time t, based on the information prevailing at time t, while the portfolio returns are calculated over $[t, t+1]$ (next month). The out-of-sample test is a real-time test mimicking the way that a real-time investor acts. We conduct backtesting experiments on a rolling window basis within a period of 118 months (from June, 2008 to March, 2018), using the 120 most recent observations. We calculate and record, for each month, the weights for the optimal index portfolios. Table 7 summarizes the weights for the optimal index portfolios for each test fund.

[Insert Table 7]

[7] In untabulated results (available upon request) we also analyze whether results in this Section are robust to the time period. Generally, we cannot reject the null that the test funds are stochastically uniformly inefficient.

We compute a number of commonly used performance measures, namely the mean and standard deviation (SD) of returns, the Sharpe ratio, the downside Sharpe ratio of Ziemba (2005), which uses the downside variance, and the UP (upside potential) ratio, which compares the upside potential to the shortfall risk over a specific target of Sortino and Van Der Meer (1991). The downside Sharpe and UP ratios are considered to be more appropriate measures of performance than the standard Sharpe ratio given the asymmetric distribution of returns.

Table 8 reports performance measures of optimal index portfolios, with respect to each of the test fund. The higher the performance measure of the optimal index portfolio relative to the fund, the greater investment opportunity created by these portfolios. We observe that in all cases, and regardless of the performance measure, the optimal index portfolios outperform the performance of the test fund. The out-of-sample results supplement our in-sample results and confirm that the mutual funds are inefficient.

[Insert Table 8]

6 Conclusions

This paper's evidence of widespread, uniform, stochastic inefficiency of many important and active funds is rigorous, and remarkable in its robustness relative to any decision maker utility function, and not based on any arbitrary assumptions on the distribution of returns. Each of these latter two choices represents considerable challenges to researchers and decision makers, and can cast doubt on generality of prior inferences in this area. The challenge of mounting a rigorous test that is capable of assessing such uniform and robust inferences is also met with careful technical modifications, and algorithmic steps for implementation, of state of art tests for stochastic dominance and portfolio choice. Uniformly dominant portfolios of index funds and other similar passive portfolios may be constructed that expose the inefficiency of many of the exisiting specialized and active funds over our sample period.

References

Amihud, Y. and Goyenko, R. (2013). Mutual fund's r 2 as predictor of performance. *The Review of Financial Studies*, 26(3):667–694.

Arvanitis, S. (2015). Saddle-type functionals for continuous processes with applications to tests for stochastic spanning. *Available at SSRN 2693562*.

Arvanitis, S., Scaillet, O., and Topaloglou, N. (2019). Spanning tests for markowitz stochastic dominance. *Journal of Econometrics*.

Arvanitis, S. and Topaloglou, N. (2017). Testing for prospect and markowitz stochastic dominance efficiency. *Journal of Econometrics*, 198(2):253 – 270.

Barras, L., Scaillet, O., and Wermers, R. (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The journal of finance*, 65(1):179–216.

Berk, J. B. and Van Binsbergen, J. H. (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics*, 118(1):1–20.

Cao, C., Iliev, P., and Velthuis, R. (2017). Style drift: Evidence from small-cap mutual funds. *Journal of Banking & Finance*, 78:42–57.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82.

Chernozhukov, V., Hong, H., and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75(5):1243–1284.

Cortissoz, J. (2007). On the skorokhod representation theorem. *Proceedings of the American Mathematical Society*, 135(12):3995–4007.

Crane, A. D. and Crotty, K. (2018). Passive versus active fund performance: do index funds have skill? *Journal of Financial and Quantitative Analysis*, 53(1):33–64.

Cremers, K. M. and Petajisto, A. (2009). How active is your fund manager? a new measure that predicts performance. *The review of financial studies*, 22(9):3329–3365.

Daniel, K., Grinblatt, M., Titman, S., and Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of finance*, 52(3):1035–1058.

Daskalaki, C., Skiadopoulos, G., and Topaloglou, N. (2017). Diversification benefits of commodities: A stochastic dominance efficiency approach. *Journal of Empirical Finance*.

Davidson, R. (2009). Testing for restricted stochastic dominance: some further results. *Review of Economic Analysis*, 1:34–59.

Davidson, R. and Duclos, J.-Y. (2013). Testing for restricted stochastic dominance. *Econometric Reviews*, 32(1):84–125.

Fama, E. F. and French, K. R. (2010). Luck versus skill in the cross-section of mutual fund returns. *The journal of finance*, 65(5):1915–1947.

Gibbons, M., Ross, S., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, pages 1121–1152.

Hall, P., Horowitz, J. L., and Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574.

Hodder, J., Jackwerth, J., and Kolokolova, O. (2015). Improved portfolio choice using second-order stochastic dominance. *Review of Finance*, 19:1623–1647.

Howes, S. (1993). Asymptotic properties of four fundamental curves of distributional analysis. *Unpublished paper, STICERD, London School of Economics.*

Kacperczyk, M., Nieuwerburgh, S. V., and Veldkamp, L. (2014). Time-varying fund manager skill. *The Journal of Finance*, 69(4):1455–1484.

Kacperczyk, M., Sialm, C., and Zheng, L. (2005). On the industry concentration of actively managed equity mutual funds. *The Journal of Finance*, 60(4):1983–2011.

Kacperczyk, M., Sialm, C., and Zheng, L. (2008). Unobserved actions of mutual funds. *The Review of Financial Studies*, 21(6):2379–2416.

Kaur, A., Rao, B. P., and Singh, H. (1994). Testing for second-order stochastic dominance of two distributions. *Econometric theory*, 10(5):849–866.

Knight, K. (1999). Epi-convergence in distribution and stochastic equi-semicontinuity.

Kosowski, R., Timmermann, A., Wermers, R., and White, H. (2006). Can mutual fund âstarsâ really pick stocks? new evidence from a bootstrap analysis. *The Journal of finance*, 61(6):2551–2595.

Kuosmanen, T. (2004). Efficient diversification according to stochastic dominance criteria. *Management Science*, 50(10):1390–1406.

Levy, H. (1992). Stochastic dominance and expected utility: survey and analysis. *Management Science*, 38(4):555–593.

Linton, O., Maasoumi, E., and Whang, Y.-J. (2005). Consistent testing for stochastic dominance under general sampling schemes. *The Review of Economic Studies*, 72(3):735–765.

Linton, O., Post, T., and Whang, Y.-J. (2014). Testing for the stochastic dominance efficiency of a given portfolio. *The Econometrics Journal*, 17(2).

Linton, O., S. K. and Whang, Y.-J. (2010). An improved bootstrap test of stochastic dominance. *Journal of Econometrics*, 154(2):186–202.

Malkiel, B. G. (1995). Returns from investing in equity mutual funds 1971 to 1991. *The Journal of finance*, 50(2):549–572.

McFadden, D. (1989). Testing for stochastic dominance. *Studies in the Economics of Uncertainty*, 2.

Molchanov, I. (2006). *Theory of random sets*. Springer Science & Business Media.

Peligrad, M. (1998). On the blockwise bootstrap for empirical processes for stationary sequences. *Ann. Probab.*, 26(2):877–901.

Post, T. (2003). Empirical tests for stochastic dominance efficiency. *The Journal of Finance*, 58(5):1905–1931.

Post, T. (2016). Empirical tests for stochastic dominance optimality. *Review of Finance*, 21(2):793–810.

Post, T. and Poti, V. (2016). Portfolio analysis using stochastic dominance, relative entropy, and empirical likelihood. *Management Science*, 63(1):153–165.

Post, T. and Versijp, P. (2007). Multivariate tests for stochastic dominance efficiency of a given portfolio. *Journal of Financial and Quantitative Analysis*, 42(2):489–515.

Scaillet, O. and Topaloglou, N. (2010). Testing for stochastic dominance efficiency. *Journal of Business & Economic Statistics*, 28(1):169–180.

Sortino, F. A. and Van Der Meer, R. (1991). Downside risk. *Journal of portfolio Management*, 17(4):27.

Wermers, R. (2012). Matter of style: The causes and consequences of style drift in institutional portfolios. *Available at SSRN 2024259*.

Ziemba, W. T. (2005). The symmetric downside-risk sharpe ratio. *The Journal of Portfolio Management*, 32(1):108–122.

Table 1: Monte Carlo results

Block size l_T :	4	8	10	12
Case 1:	R=100	T=200		
Empirical size	5.6%	4.8%	5.7%	6.6%
Empirical power	91.2%	92.2%	90.6%	91.1%
Case 2:	R=300	T=500		
Empirical size	4.9%	5.2%	4.7%	5.3%
Empirical power	93.1%	92.8%	92.2%	91.6%
Case 3:	R=500	T=1000		
Empirical size	4.8%	4.4%	5.0%	5.2%
Empirical power	95.2%	94.8%	96.1%	93.9%

This table reports the empirical size and power based on 1000 replications and a nominal size $\alpha = 5\%$

Table 2: **In-sample tests of stochastic uniform inefficiency with respect to all funds**

	test stats.	p-value
5 Largest AUM Index Funds		
Vanguard 500 Index Fund	0.0059	98.00%
Vanguard Institutional Index Fund	0.0062	96.79%
Vanguard Total Stock Market Index	0.0062	98.40%
Fidelity 500 Index Fund	0.0034	98.60%
Vanguard Mid-Cap Index Fund	0.0063	98.60%
5 Smallest AUM Active Funds		
Fidelity Advisor Tax Managed Stock	0.0166	59.72%
Fidelity Select Paper & Forest Products	0.0199	65.33%
Fidelity Advisor Value Leaders Fund	0.0070	92.99%
Fidelity Advisor Semiconductors Fund	0.0027	87.78%
Fidelity Select Env. & Alter. Energy		
5 Largest AUM Active Funds		
Growth Fund of America	0.0039	99.40%
Investment Company of America	0.0058	96.59%
Fidelity Contrafund Fund	0.0041	97.60%
Fidelity Magellan Fund	0.0069	97.80%
Vanguard Windsor II Fund	0.0063	93.79%

This table exhibits the test and p-values for stochastic uniform inefficiency of a specific fund with respect to all possible portfolios constructed with 248 equity funds from Vanguard, Fidelity, and American fund family. The sample period is March 1980 to March 2018.

Table 3: **In-sample tests of stochastic uniform inefficiency with respect to Value-Weighted Index**

	test stats.	p-value
5 Largest AUM Index Funds		
Vanguard 500 Index Fund	0.00017	46.09%
Vanguard Institutional Index Fund	-0.00019	34.67%
Vanguard Total Stock Market Index	-0.00014	38.28%
Fidelity 500 Index Fund	0.00289	59.32%
Vanguard Mid-Cap Index Fund	-0.00022	43.49%
5 Smallest AUM Active Funds		
Fidelity Advisor Tax Managed Stock	-0.00046	63.73%
Fidelity Select Paper & Forest Products	0.00324	21.44%
Fidelity Advisor Value Leaders Fund	-0.00215	60.52%
Fidelity Advisor Semiconductors Fund	-0.00038	34.27%
Fidelity Select Env. & Alter. Energy	-0.00200	49.10%
5 Largest AUM Active Funds		
Growth Fund of America	0.00239	41.68%
Investment Company of America	0.00016	44.09%
Fidelity Contrafund Fund	0.00201	40.28%
Fidelity Magellan Fund	-0.00059	42.69%
Vanguard Windsor II Fund	-0.00016	54.51%

This table exhibits the test and p-values for stochastic uniform inefficiency of a specific fund with respect to the value-weighted index of all CRSP listed stocks. The sample period is March 1980 to March 2018.

Table 4: **In-sample Tests of Stochastic Uniform Inefficiency with Respect to Equal-Weighted Index**

	test stats.	p-value
5 Largest AUM Index Funds		
Vanguard 500 Index Fund	-0.00251	36.47%
Vanguard Institutional Index Fund	-0.00287	38.48%
Vanguard Total Stock Market Index	-0.00282	38.88%
Fidelity 500 Index Fund	0.00021	35.07%
Vanguard Mid-Cap Index Fund	-0.00290	37.07%
5 Smallest AUM Active Funds		
Fidelity Advisor Tax Managed Stock	-0.00306	68.94%
Fidelity Select Paper & Forest Products	0.00064	49.70%
Fidelity Advisor Value Leaders Fund	-0.00303	68.14%
Fidelity Advisor Semiconductors Fund	-0.00679	33.67%
Fidelity Select Env. & Alter. Energy	-0.00724	55.51%
5 Largest AUM Active Funds		
Growth Fund of America	-0.00029	37.07%
Investment Company of America	-0.00251	38.08%
Fidelity Contrafund Fund	-0.00066	39.88%
Fidelity Magellan Fund	-0.00327	40.08%
Vanguard Windsor II Fund	-0.00283	44.69%

This table exhibits the test statistics and p-value for stochastic uniform inefficiency test of a specific fund with respect to equal-weighted index of all CRSP listed stocks. The sample period is from March 1980 to March 2018.

Table 5: **In-sample tests of stochastic uniform inefficiency with respect to S&P 500**

	test stats.	p-value
5 Largest AUM Index Funds		
Vanguard 500 Index Fund	0.00190	54.71%
Vanguard Institutional Index Fund	0.00154	45.49%
Vanguard Total Stock Market Index	0.00158	44.69%
Fidelity 500 Index Fund	0.00462	56.51%
Vanguard Mid-Cap Index Fund	0.00150	47.29%
5 Smallest AUM Active Funds		
Fidelity Advisor Tax Managed Stock	0.00155	55.11%
Fidelity Select Paper & Forest Products	0.00524	23.85%
Fidelity Advisor Value Leaders Fund	-0.00036	52.71%
Fidelity Advisor Semiconductors Fund	0.00296	33.47%
Fidelity Select Env. & Alter. Energy		
5 Largest AUM Active Funds		
Growth Fund of America	0.00412	44.29%
Investment Company of America	0.00189	50.70%
Fidelity Contrafund Fund	0.00374	43.49%
Fidelity Magellan Fund	0.00113	48.70%
Vanguard Windsor II Fund	0.00157	56.51%

This table exhibits the test statistics and p-value for stochastic uniform inefficiency test of a specific fund with respect to S&P 500. The sample period is from March 1980 to March 2018.

Table 6: **In-sample tests of stochastic uniform inefficiency with respect to 7 Vanguard Index Funds**

	test stats.	p-value
5 Largest AUM Index Funds		
Vanguard 500 Index Fund	0.00272	73.24%
Vanguard Institutional Index Fund	0.00308	68.56%
Vanguard Total Stock Market Index	0.00303	69.57%
Fidelity 500 Index Fund	-0.0003	99.60%
Vanguard Mid-Cap Index Fund	0.00311	70.90%
5 Smallest AUM Active Funds		
Fidelity Advisor Tax Managed Stock	0.00287	66.13%
Fidelity Select Paper & Forest Products	0.00316	91.27%
Fidelity Advisor Value Leaders Fund	0.00433	88.14%
Fidelity Advisor Semiconductors Fund	0.00211	70.30%
Fidelity Select Env. & Alter. Energy	0.00834	82.13%
5 Largest AUM Active Funds		
Growth Fund of America	0.00050	70.80%
Investment Company of America	0.00273	74.58%
Fidelity Contrafund Fund	0.00088	69.23%
Fidelity Magellan Fund	0.00348	68.23%
Vanguard Windsor II Fund	0.00305	62.88%

This table exhibits the test statistics and p-value for stochastic uniform inefficiency test of a specific fund with respect to seven Vanguard index funds. The sample period is from March 1980 to March 2018.

Table 7: **Weight allocation of the optimal portfolios**

	Vanguard 500 Index Fund				Vanguard Institutional Index Fund			
	Mean	Std. Dev.	Skewness	Kurtosis	Mean	Std. Dev.	Skewness	Kurtosis
S&P 500 Index	0.3902	0.3470	0.0911	-1.6439	0.3117	0.3351	0.5673	-1.1458
Extended Market Index	0.0482	0.1098	2.7088	8.5311	0.0323	0.1416	5.4465	31.0915
Small-Cap Index	0.0091	0.0660	8.0834	67.9509	0.0164	0.0982	7.1165	55.5122
Value Index	0.1211	0.2227	2.1188	4.3669	0.0875	0.2010	2.7524	7.7060
Mid-Cap Index	0.1670	0.2003	1.7479	4.0102	0.1411	0.2487	1.9125	2.8029
Small-Cap Growth Index	0.1066	0.2680	2.5411	5.0802	0.2656	0.3898	1.0603	-0.5661
Small-Cap Value Index	0.1578	0.3253	1.8769	1.9660	0.1454	0.3251	2.0361	2.4832

	Vanguard Total Stock Market Index				Fidelity 500 Index Fund			
	Mean	Std. Dev.	Skewness	Kurtosis	Mean	Std. Dev.	Skewness	Kurtosis
S&P 500 Index	0.6183	0.4428	-0.5464	-1.5685	0.1201	0.2648	2.2252	3.7582
Extended Market Index	0.0051	0.0322	6.4517	41.7223	0.0061	0.0404	7.8446	64.9956
Small-Cap Index	0.0025	0.0268	10.8628	118.00	0.0364	0.1825	5.0725	24.4714
Value Index	0.0532	0.1647	4.1345	18.7969	0.0588	0.1745	3.6672	14.5656
Mid-Cap Index	0.0724	0.1745	3.1746	10.7958	0.1833	0.2932	1.4010	0.6721
Small-Cap Growth Index	0.1031	0.2918	2.6287	5.1869	0.4499	0.4478	0.1986	-1.7890
Small-Cap Value Index	0.1454	0.3251	2.0361	2.4832	0.1454	0.3251	2.0361	2.4832

	Vanguard Mid-Cap Index Fund				Fidelity Advisor Tax Managed Stock			
	Mean	Std. Dev.	Skewness	Kurtosis	Mean	Std. Dev.	Skewness	Kurtosis
S&P 500 Index	0.5918	0.4677	-0.3925	-1.8011	0.0259	0.0758	3.0747	8.8953
Extended Market Index	0.0047	0.0443	10.4673	111.6341	0.1369	0.3146	2.2235	3.4198
Small-Cap Index	0.0025	0.0268	10.8628	118.00	0.0257	0.1020	4.4052	19.8042
Value Index	0.0500	0.1670	4.1022	18.1371	0.3050	0.3711	0.6439	-1.4050
Mid-Cap Index	0.0613	0.1810	3.3612	11.0234	0.3704	0.4179	0.5616	-1.5038
Small-Cap Growth Index	0.1443	0.3345	2.0505	2.4210	0.0892	0.2469	2.7545	6.6164
Small-Cap Value Index	0.1454	0.3251	2.0361	2.4832	0.0468	0.1083	2.6691	7.2165

Table 7 Continued

	Fidelity Select Paper & Forest Products				Fidelity Advisor Value Leaders Fund			
	Mean	Std. Dev.	Skewness	Kurtosis	Mean	Std. Dev.	Skewness	Kurtosis
S&P 500 Index	0.0112	0.0501	4.4075	18.2963	0.2023	0.3541	1.3897	0.1516
Extended Market Index	0.1033	0.2912	2.8058	6.2765	0.0199	0.1322	7.3583	54.8703
Small-Cap Index	0.0370	0.1724	4.9594	25.1479	0.0930	0.2840	2.9097	6.8272
Value Index	0.2040	0.3380	1.2753	-0.1578	0.0781	0.2066	3.0426	9.1832
Mid-Cap Index	0.3402	0.4265	0.7050	-1.3334	0.0945	0.1910	2.1206	3.6279
Small-Cap Growth Index	0.2531	0.3831	1.1061	-0.4513	0.4840	0.4588	0.0571	-1.9275
Small-Cap Value Index	0.0513	0.1727	4.5585	22.5655	0.0282	0.1305	5.0527	26.0840

	Fidelity Advisor Semiconductors Fund				Fidelity Select Env. & Alter. Energy			
	Mean	Std. Dev.	Skewness	Kurtosis	Mean	Std. Dev.	Skewness	Kurtosis
S&P 500 Index	0.1566	0.3229	1.7930	1.5404	0.1669	0.3203	1.7257	1.3605
Extended Market Index	0.0642	0.2303	3.6753	12.1716	0.0748	0.2444	3.4833	10.7230
Small-Cap Index	0.0584	0.2276	3.8224	13.2313	0.0750	0.2521	3.3029	9.4724
Value Index	0.0807	0.1991	2.7968	7.4811	0.1090	0.2172	2.0366	3.3840
Mid-Cap Index	0.1175	0.2331	1.9830	2.9382	0.1363	0.2393	1.6765	1.7893
Small-Cap Growth Index	0.4644	0.4744	0.1324	-1.9394	0.4240	0.4525	0.2897	-1.7984
Small-Cap Value Index	0.0582	0.2013	3.8042	13.9176	0.0139	0.0663	6.0016	39.1144

	Growth Fund of America				Investment Company of America			
	Mean	Std. Dev.	Skewness	Kurtosis	Mean	Std. Dev.	Skewness	Kurtosis
S&P 500 Index	0.4394	0.4040	0.2457	-1.6009	0.6604	0.4627	-0.6696	-1.5355
Extended Market Index	0.0181	0.0930	6.7982	49.5336	0.0	0.0	-	-
Small-Cap Index	0.0025	0.0268	10.8628	118.00	0.0025	0.0268	10.8628	118.00
Value Index	0.0909	0.1960	2.5740	7.1059	0.0405	0.1608	4.6103	22.3704
Mid-Cap Index	0.1919	0.2739	1.2739	0.3924	0.0502	0.1706	3.7442	13.9732
Small-Cap Growth Index	0.1117	0.2950	2.4888	4.5711	0.1010	0.2918	2.6484	5.2611
Small-Cap Value Index	0.1454	0.03251	2.0361	2.4832	0.1454	0.3251	2.0361	2.4832

Table 7 Continued

	Fidelity Contrafund Fund				Fidelity Magellan Fund			
	Mean	Std. Dev.	Skewness	Kurtosis	Mean	Std. Dev.	Skewness	Kurtosis
S&P 500 Index	0.3241	0.3626	0.6380	-1.1540	0.1522	0.3005	1.8168	1.8369
Extended Market Index	0.0235	0.0744	3.8610	16.6942	0.0270	0.1380	6.2883	41.3647
Small-Cap Index	0.0083	0.0459	5.6849	32.3443	0.0474	0.2040	4.3894	17.9898
Value Index	0.1195	0.2272	2.2194	4.7171	0.0788	0.2004	2.9219	8.5332
Mid-Cap Index	0.2012	0.2929	1.2354	0.1198	0.1437	0.2588	1.7409	1.8699
Small-Cap Growth Index	0.1454	0.3020	2.0955	3.0348	0.4055	0.4616	0.3865	-1.7724
Small-Cap Value Index	0.1779	0.3481	1.6679	1.0752	0.1454	0.3251	2.0361	2.4832

	Vanguard Windsor II Fund			
	Mean	Std. Dev.	Skewness	Kurtosis
S&P 500 Index	0.2393	0.3386	1.0481	-0.5178
Extended Market Index	0.0144	0.0946	9.7805	101.3324
Small-Cap Index	0.0474	0.2040	4.3894	17.9898
Value Index	0.1934	0.2981	1.3338	0.2579
Mid-Cap Index	0.1175	0.2022	2.2496	5.2876
Small-Cap Growth Index	0.2426	0.4031	1.1953	-0.4338
Small-Cap Value Index	0.1454	0.3251	2.0361	2.4832

This table exhibits the descriptive statistics of the weight allocation of the optimal Index portfolio for each fund. The dataset spans the period from June 2008 to March 2018.

Table 8: Out-of-sample performance

	Mutual Fund					Optimal Index Portfolios				
	Mean	SD	S	DS	US	Mean	SD	S	DS	US
Vanguard 500 Index Fund	0.0083	0.0448	0.1809	0.1851	0.6879	0.0097	0.053	0.1786	0.1886	0.6985
Vanguard Institutional Index Fund	0.0081	0.0435	0.1827	0.1874	0.691	0.0105	0.0569	0.1834	0.1883	0.6917
Vanguard Total Stock Market Index	0.0082	0.0434	0.1836	0.1884	0.6924	0.011	0.0563	0.1914	0.2003	0.6935
Fidelity 500 Index Fund	0.0086	0.0506	0.1658	0.1679	0.6569	0.0106	0.058	0.18	0.1868	0.6779
Vanguard Mid-Cap Index Fund	0.0081	0.0435	0.1826	0.1873	0.6908	0.0106	0.0564	0.1835	0.1916	0.6952
Fidelity Advisor Tax Managed Stock	-0.0095	0.0545	-0.1783	-0.1382	0.2664	-0.0058	0.0538	-0.1116	-0.0924	0.3175
Fidelity Select Paper & Forest Products	-0.001	0.0992	-0.0121	-0.0125	0.409	-0.0004	0.0606	-0.0101	-0.0093	0.4591
Fidelity Advisor Value Leaders Fund	0.0081	0.0305	0.26	0.3075	0.9531	0.012	0.0345	0.3426	0.3911	0.9653
Fidelity Advisor Semiconductors Fund	0.0187	0.0491	0.3779	0.5093	1.1963	0.0159	0.0343	0.4596	0.6118	1.2389
Fidelity Select Env. & Alter. Energy	0.0117	0.0359	0.3215	0.4228	1.086	0.0134	0.0357	0.3686	0.4596	1.0947
Growth Fund of America	0.008	0.0445	0.1765	0.1814	0.694	0.0113	0.0564	0.1968	0.2066	0.6975
Investment Company of America	0.0074	0.0412	0.1741	0.1807	0.703	0.011	0.0563	0.1917	0.2008	0.6942
Fidelity Contrafund Fund	0.0087	0.041	0.208	0.2235	0.7598	0.012	0.0568	0.2087	0.2377	0.7686
Fidelity Magellan Fund	0.0069	0.0534	0.125	0.1227	0.6048	0.0104	0.0582	0.1756	0.1819	0.6756
Vanguard Windsor II Fund	0.0071	0.0448	0.1539	0.1544	0.6531	0.0101	0.0574	0.1731	0.1788	0.6643

This table reports the performance measures (Mean, Standard Deviation, Sharpe ratio, Downside Sharpe ratio, UP ratio) for the mutual funds and the corresponding optimal Index portfolios. The dataset spans the period from June 2008 to March 2018. In the second half, the table exhibits the descriptive statistics of the weight allocation of the optimal Index portfolio.

Appendix

Proofs of main results

The results in the auxiliary Lemma 6 imply that $D\left(z,\kappa,\lambda,\sqrt{T}\left(\hat{F}-F\right)\right)$ weakly converges to $D\left(z,\kappa,\lambda,\mathcal{B}_F\right)$ w.r.t. to the product topology of hypo-convergence on the product of the relevant spaces of usc real valued functions (see e.g. Knight (1999) for the dual notion of epi-convergence). This product space is metrizable as complete and separable (see again Knight (1999)). Hence, Skorokhod representations are applicable (as above, see for example Theorem 1 in Cortissoz (2007)) and thereby there exists an enhanced probability space and processes $D_T\left(z,\lambda\right) \overset{d}{=} D\left(z,\kappa,\lambda,\sqrt{T}\left(\hat{F}-F\right)\right)$, where z is restricted to $\hat{A}_{\lambda}^{\delta_T}$, $D\left(z,\lambda\right) \overset{d}{=} D\left(z,\kappa,\lambda,\mathcal{B}_F\right)$,defined on it such that due to Lemma 6 $D_T \to D$ almost surely, w.r.t. to the product topology of hypo-convergence, where $\overset{d}{=}$ denotes equality in distribution. Notice that for z restricted to $\hat{A}_{\lambda}^{\delta_T}$,

$$D\left(z,\kappa,\lambda,\sqrt{T}\hat{F}\right) \equiv D_T\left(z,\lambda\right) + \sqrt{T}D\left(z,\kappa,\lambda,F\right),$$

and that under $\mathbf{H_0}$, almost surely, for any z,λ and any $z_T \to z$ and $\lambda_T \to \lambda$, due to that $\Lambda^{=}$ is closed, almost surely

$$\limsup_{T\to\infty} D\left(z_T,\kappa,\lambda_T,\sqrt{T}\hat{F}\right)$$

$$= \begin{cases} D\left(z,\lambda\right), & (z,\lambda) \in \mathcal{Z}\times\Lambda^{=}, (z_T,\lambda_T) \in_{T^{\star}} \hat{A}_{\lambda}^{\delta_T}\times\Lambda^{=}, \\ D\left(z,\lambda\right), & (z,\lambda) \in \mathcal{Z}\times\Lambda^{=}, (z_T,\lambda_T) \in \hat{A}_{\lambda}^{\delta_T}\times\Lambda^{=} \\ +\infty, & \lambda \notin \Lambda^{=}, \lambda_T \notin \Lambda^{=}, (z,z_T) \in \mathcal{Z}\times\hat{A}_{\lambda}^{\delta_T} \end{cases},$$

where $\in_{T^{\star}}$ denotes "eventually belongs", and,

$$\liminf_{T\to\infty} D\left(z_T,\kappa,\lambda_T,\sqrt{T}\hat{F}\right)$$

$$= \begin{cases} D(z,\lambda), & (z,\lambda) \in \mathcal{Z} \times \Lambda^{=}, (z_T, \lambda_T) \in \hat{A}_{\lambda}^{\delta_T} \times \Lambda^{=}, \\ +\infty, & \lambda \notin \Lambda^{=}, \lambda_T \notin \Lambda^{=}, (z, z_T) \in \mathcal{Z} \times \hat{A}_{\lambda}^{\delta_T} \end{cases}.$$

Hence, $D\left(z_T, \kappa, \lambda_T, \sqrt{T}\hat{F}\right)$ converges almost surely w.r.t. to the product topology of hypo-convergence to the limit $K(z,\lambda)$, with

$$K(z,\lambda) = \begin{cases} D(z,\lambda), & (z,\lambda) \in \mathcal{Z} \times \Lambda^{=}, (z_T, \lambda_T) \in \hat{A}_{\lambda}^{\delta_T} \times \Lambda^{=}, \\ +\infty, & \lambda \notin \Lambda^{=}, \lambda_T \notin \Lambda^{=}, (z, z_T) \in \mathcal{Z} \times \hat{A}_{\lambda}^{\delta_T} \end{cases},$$

due to the Proposition 3.2 (ch. 5, p. 337) of Molchanov (2006). Furthermore, since almost surely $\sup_T \left|\sup_{z,\lambda} D_T(z,\lambda)\right| < +\infty$, and due to the form of $\mathbf{H_0}$, we have that,

$$\lim_{T\to\infty} \sup_{\lambda} \sup_{z \in \hat{A}_{\lambda}^{\delta_T}} D\left(z_T, \kappa, \lambda_T, \sqrt{T}\hat{F}\right)$$

$$\leq \sup_{\lambda} \sup_{z\in\mathcal{Z}} K(z,\lambda) = \begin{cases} \sup_{\lambda} \sup_{z\in\mathcal{Z}} D(z,\lambda), & \lambda \in \Lambda^{=} \\ +\infty, & \lambda \notin \Lambda^{=} \end{cases}.$$

Hence, due to Theorem 3.4 (ch. 5, p. 338) of Molchanov (2006),

$$\sup_{\lambda} \sup_{z \in \hat{A}_{\lambda}^{\delta_T}} K_T(z,\lambda) \to \sup_{\lambda \in \Lambda^{=}} \sup_{z\in\mathcal{Z}} K(z,\lambda), \text{ almost surely.}$$

and the result follows. Lemma 6 and $\underline{\lim}_{T\to\infty} \mathbb{P}(q_T(1-\alpha) < -\delta) = 1$ imply the continuity of the quantile function of the law of ξ_∞at $(1-\alpha)$. Assumptions 2 and 3 and Theorem 2.3 of Peligrad (1998) imply that conditionally on the sample,

$$\sqrt{T}\left(\hat{F}^{\star} - \hat{F}\right) \rightsquigarrow^{p} \mathcal{B}_F^{*}$$

where $\mathcal{B}_F^{*}$ is an independent version of the Gaussian process in Proposition 2.3. Also analogously to the proofs of Proposition 2.3and Lemma 6 we have that conditionally on the sample

$$\sup_{\lambda\in\Lambda^{=}}\sup_{z\in\mathcal{Z}} D\left(z,\kappa,\lambda,\sqrt{T}\left(\hat{F}^{\star}-\hat{F}\right)\right) \rightsquigarrow^{p} \sup_{\lambda\in\Lambda^{=}}\sup_{z\in\mathcal{Z}} D\left(z,\kappa,\lambda,\mathcal{B}_{F}^{*}\right).$$

Since $\mathbb{P}\left(\Lambda^{=}\subseteq\hat{\Lambda}^{=}\subseteq\Lambda^{=2\mathrm{k_T}}\right)\to 1$ from Lemma 6 and $k_T\to 0$ we have the following with probability approaching to one

$$\sup_{\lambda\in\Lambda^{=}}\sup_{z\in\mathcal{Z}} D\left(z,\kappa,\lambda,\sqrt{T}\left(\hat{F}^{\star}-\hat{F}\right)\right)1_{\Lambda}+\pi_{\mathrm{T}}\left(1-1_{\Lambda}^{2\mathrm{k_T}}\right)$$

$$\leq \sup_{\lambda\in\hat{\Lambda}^{=}}\sup_{z\in\mathcal{Z}} D\left(z,\kappa,\lambda,\sqrt{T}\left(\hat{F}^{\star}-\hat{F}\right)\right)\hat{1}_{\Lambda}+\pi_{\mathrm{T}}\left(1-\hat{1}_{\Lambda}\right)$$

$$\leq \sup_{\lambda\in\Lambda^{=2\mathrm{k_T}}}\sup_{z\in\mathcal{Z}} D\left(z,\kappa,\lambda,\sqrt{T}\left(\hat{F}^{\star}-\hat{F}\right)\right)1_{\Lambda}^{2\mathrm{k_T}}+\pi_{\mathrm{T}}\left(1-1_{\Lambda}\right),$$

where

$$1_{\Lambda}=\mathbb{I}\left\{\Lambda^{=}\neq\{\kappa\}\right\}$$

$$\hat{1}_{\Lambda}=\mathbb{I}\left\{\hat{\Lambda}^{=}\neq\emptyset\right\}$$

$$1_{\Lambda}^{2k_T}=\mathbb{I}\left\{\Lambda^{=2k_T}\neq\emptyset\right\}.$$

Suppose that $\Lambda^{=}\neq\{\kappa\}$ then with probability approaching to one

$$\sup_{\lambda\in\Lambda^{=}}\sup_{z\in\mathcal{Z}} D\left(z,\kappa,\lambda,\sqrt{T}\left(\hat{F}^{\star}-\hat{F}\right)\right)$$

$$\leq \sup_{\lambda\in\hat{\Lambda}^{=}}\sup_{z\in\mathcal{Z}} D\left(z,\kappa,\lambda,\sqrt{T}\left(\hat{F}^{\star}-\hat{F}\right)\right)\hat{1}_{\Lambda}+\pi_{\mathrm{T}}\left(1-\hat{1}_{\Lambda}\right)$$

$$\leq \sup_{\lambda\in\Lambda^{=2\mathrm{k_T}}} \sup_{z\in\mathcal{Z}} D\left(z,\kappa,\lambda,\sqrt{T}\left(\hat{F^{\star}}-\hat{F}\right)\right).$$

Therefore with probability approaching to one

$$\left|\sup_{\lambda\in\hat{\Lambda}^{=}} \sup_{z\in\mathcal{Z}} D\left(z,\kappa,\lambda,\sqrt{T}\left(\hat{F^{\star}}-\hat{F}\right)\right)\hat{1}_{\Lambda}+\pi_{\mathrm{T}}\left(1-\hat{1}_{\Lambda}\right)-\sup_{\lambda\in\Lambda^{=}} \sup_{z\in\mathcal{Z}} D\left(z,\kappa,\lambda,\sqrt{T}\left(\hat{F^{\star}}-\hat{F}\right)\right)\right|$$

$$\leq \sup_{\lambda\in\Lambda^{=2\mathrm{k_T}}} \sup_{z\in\mathcal{Z}} D\left(z,\kappa,\lambda,\sqrt{T}\left(\hat{F^{\star}}-\hat{F}\right)\right)-\sup_{\lambda\in\Lambda^{=}} \sup_{z\in\mathcal{Z}} D\left(z,\kappa,\lambda,\sqrt{T}\left(\hat{F^{\star}}-\hat{F}\right)\right).$$

Hence from Lemma 2.3, the choice of k_T and the fact that $\Lambda^{=}$ is closed, as $T\to\infty$ we have that conditionally on the sample

$$\sup_{\lambda\in\hat{\Lambda}^{=}} \sup_{z\in\mathcal{Z}} D\left(z,\kappa,\lambda,\sqrt{T}\left(\hat{F^{\star}}-\hat{F}\right)\right)\hat{1}_{\Lambda}+\pi_{\mathrm{T}}\left(1-\hat{1}_{\Lambda}\right)\rightsquigarrow^{p} \sup_{\lambda\in\Lambda^{=}} \sup_{z\in\mathcal{Z}} D\left(z,\kappa,\lambda,\sqrt{T}\left(\hat{F^{\star}}-\hat{F}\right)\right).$$

This yields the second result.

When $\Lambda^{=}=\{\kappa\}$ by using the fact that the bootstrap test statstic is bounded by $\pi_T\to -\infty$, the form of the null hypothesis proves the result.

Analogously to the proofs of Lemma 6, Proposition 2.3 and the proof of Proposition 3 in Arvanitis and Topaloglou (2017) we establish the asymptotic conservatism for the test. Notice that if $\mathrm{H_A}$ is true let $\lambda^{*}\in\Lambda$. Then we have that

$$\xi_T \leq \sup_{z\epsilon\hat{A}_{\lambda}^{\delta_T}} D\left(z\ ,\kappa,\lambda^{*},\sqrt{T}\left(\hat{F}-F\right)\right)+\sqrt{T}\sup_{z\epsilon\hat{A}_{\lambda}^{\delta_T}} D\left(z\ ,\kappa,\lambda^{*},F\right),$$

and again due to arguments analogous to the ones used in the proof of Proposition 2.3, we have that the first term in the rhs of the last display is asymptotically tight, while using the definition of δ_T the second term in the rhs of the last display diverges to $-\infty$. In our case, under the null hypothesis there is some $\lambda\in\Lambda$ such that λ is not strictly dominated by κ.

Let $\lambda = 1$ whence if $\mu_i = 0$ for $i = 1, 2, 3$, $\omega_1 < \omega_3$, $a_1 < a_3$ and $\beta_1 < \beta_3$ then $h_{3_t} > h_{1_t}$ $\mathbb{P}$ a.s. Using similar arguments as in the proof of Proposition 4 in Arvanitis and Topaloglou (2017)we have that for $z \leq 0$,

$$D(z, \kappa, \lambda, F) = \int_{-\infty}^{z} \left(\mathbb{E}\left[\Phi\left(\frac{x}{\sqrt{h_{3t}}} \right) \right] - \mathbb{E}\left[\Phi\left(\frac{x}{\sqrt{h_{1t}}} \right) \right] \right) dx > 0,$$

which implies that the first part of Definition 2.1 is valid. Define $\mathcal{F}_t = \sigma\{z_{t-1}, z_{t-2}, \ldots\}$ and notice that due to the definition of λ, the almost sure positivity of h_{it} for all i and Jensen's inequality,

$$\min\{h_{1t}, h_{2t}, h_{3t}\} \leq v_{\lambda_t} \leq \max\{h_{1t}, h_{2t}, h_{3t}\} \;\; \mathbb{P} \text{ a.s.},$$

where $v_{\lambda_t} := \operatorname{Var}(\lambda y_{1t} + (1-\lambda) y_{2t}/\mathcal{F}_t)$ or $v_{\lambda_t} := h_{3t}$. Define the auxiliary processes by

$$\begin{aligned} h_{*t} &= a_* \left(1 + \left(z_{t-1}^2 + 1\right) h_{*t-1}\right), \\ h_t^* &= a^* \left(1 + \left(z_{t-1}^2 + 1\right) h_{t-1}^*\right), \end{aligned}$$

for $a_* = \min\{\omega_i, a_i, \beta_i,\ i = 1, 2, 3\}$, $a^* = \max\{\omega_i, a_i, \beta_i,\ i = 1, 2, 3\}$ and notice that

$$h_{*t} \leq \min\{h_{1t}, h_{2t}, h_{3t}\} \leq \max\{h_{1t}, h_{2t}, h_{3t}\} \leq h_t^*, \; \mathbb{P} \text{ a.s.}$$

Following the same arguments as in the proof of Proposition 4 in Arvanitis and Topaloglou (2017)we have that for any $z \leq 0$ the condition $\sqrt{\frac{\max\{\omega_i, a_i, \beta_i,\ i=1,2,3\}}{\min\{\omega_i, a_i, \beta_i,\ i=1,2,3\}}} < v$ ensures that $D(z, \kappa^\star, \lambda, F) =< 0$ while for any $z > 0$ the condition $v < \sqrt{\frac{\min\{\omega_i, a_i, \beta_i,\ i=1,2,3\}}{\max\{\omega_i, a_i, \beta_i,\ i=1,2,3\}}}$ ensures that $D(z, \kappa^\star, \lambda, F) =< 0$.

Auxiliary lemmata

The following are auxiliary lemmata used for the derivation of the proofs of the propositions. Under Assumption 2

$$D\left(z, \kappa, \lambda, \sqrt{T}\left(\hat{F}-F\right)\right) \rightsquigarrow D\left(z, \kappa, \lambda, \mathcal{B}_F\right)$$

as random elements with values on the space of $\mathbb{R}^2$-valued bounded functions on $\Lambda \times \mathcal{Z}$ equiped with the sup-norm. The limiting process has continuous sample paths. The result follows from arguments similar to the proof of Lemma 1 of Arvanitis et al. (2019) where $\theta \equiv (\lambda, z) \in \Theta \equiv \Lambda \times \mathcal{Z}$. If $\Lambda^{=} \neq \{\kappa\}$, and under Assumption 2, the distribution of ξ_∞ may have a jump discontinuity at zero and is everywhere else continuous. The result follows from arguments similiar to the proof of Lemma 2 of Arvanitis et al. (2019). This follows from the verification of Assumption 1 in the aforementioned paper and via the subsequent use of Collorary 1. Suppose Assumptions 1 and 2 hold. Then, for any $\lambda \in \mathit{\Lambda}$,

$$Pr\left(A_\lambda^{2\delta_T} \subset \hat{A}_\lambda^{\delta_T} \subset A_\lambda^{\delta_T}\right) \to 1$$

and for $w \in \mathbb{R}$ we have that,

$$\mid Pr\left(\sup_{\lambda\in\mathit{\Lambda}^{=}} \sup_{z\in\hat{A}_\lambda^{\delta_T}} D\left(z, \kappa, \lambda, \sqrt{T}\hat{F}\right) \leq w\right) - Pr\left(\sup_{\lambda\in\mathit{\Lambda}^{=}} \sup_{z\in\mathcal{Z}} D\left(z, \kappa, \lambda, \sqrt{T}\hat{F}\right) \leq w\right) \mid$$

$$\leq Pr\left(\hat{A}_\lambda^{\delta_T} \neq \mathcal{Z} \; for \; \lambda \in \mathit{\Lambda}^{=}\right) \to 0$$

as $T \to \infty$. This follows directly from Lemma 2 of Linton et al. (2014) and Assumption 1. Suppose Assumptions 1 and 2 hold. Then, for any $\lambda \in \mathit{\Lambda}$,

$$Pr\left(\mathit{\Lambda}^{=} \subset \hat{\mathit{\Lambda}}^{=} \subset \mathit{\Lambda}^{=2k_T}\right) \to 1.$$

This follows directly from the arguments of the proof of Theorem 2 of Linton and Whang (2010) and the choice of k_T .