

Stochastic Spanning

Stelios Arvanitis*, Mark Hallam†, Thierry Post‡ and Nikolas Topaloglou§

April 29, 2017

Abstract

This study develops and implements methods for determining whether introducing new securities or relaxing investment constraints improves the investment opportunity set for all risk averse investors. We develop a test procedure for ‘stochastic spanning’ for two nested portfolio sets based on subsampling and Linear Programming. The test is statistically consistent and asymptotically exact for a class of weakly dependent processes. A Monte-Carlo simulation experiment shows good statistical size and power properties in finite samples of realistic dimensions. In an application to standard data sets of historical stock market returns, we accept market portfolio efficiency but reject two-fund separation, which suggests an important role for higher-order moment risk in portfolio theory and asset pricing.

Keywords: Portfolio choice, Stochastic Dominance, Spanning, Subsampling, Linear Programming.

JEL subject codes: C61, D81, G11.

Acknowledgement. We are grateful to two referees, the Associate Editor and the Editor for their very helpful comments, which have led to this improved version of this study.

*Department of Economics, AUEB, Athens, Greece.

†Essex Business School, University of Essex, UK.

‡Graduate School of Business, Nazarbayev University, Astana, Kazakhstan.

§Department of IEES, AUEB, Athens, Greece.

1 Introduction

Stochastic Dominance (SD) ranks prospects based on general regularity conditions for decision making under risk (Quirk and Saposnik (1962), Hadar and Russell (1969), Hanoch and Levy (1969) and Rothschild and Stiglitz (1970)). SD can be seen as a model-free alternative to mean-variance (M-V) dominance. The M-V criterion is consistent with Expected Utility for elliptical distributions such as the normal distribution (Chamberlain (1983), Owen and Rabinovitch (1983), Berk (1997)) but has limited economic meaning when the probability distribution cannot be characterized completely by its location and scale.

Simaan (1993), Athayde and Flores (2004) and Mencia and Sentana (2009) develop a mean-variance-skewness framework based on generalizations of elliptical distributions that are fully characterized by their first three moments. SD presents a further generalization that accounts for all moments of the return distributions without assuming a particular family of distributions.

SD is traditionally applied for comparing a pair of given prospects, for example, two income distributions or two medical treatments. Davidson and Duclos (2000), Barrett and Donald (2003) and Linton, Maasoumi and Whang (2005), among others, develop statistical tests for such pairwise comparisons.

A more general, multivariate problem is that of testing whether a given prospect is stochastically efficient relative to all mixtures of a discrete set of alternatives (Bawa et al. (1985), Shalit and Yitzhaki (1994), Post (2003), Kuosmanen (2004), Roman, Darby-Dowman and Mitra (2006)). This problem arises naturally in applications of portfolio theory and asset pricing theory, where the mixtures are portfolios of financial securities. Post and Versijp (2007), Scaillet and Topaloglou (2010), Linton, Post and Whang (2014) and Post and Poti (2017) address this problem using various statistical methods. Their stochastic efficiency tests can be seen as model-free alternatives to tests for M-V efficiency, such as Gibbons, Ross and Shanken (1989).

In an analogous manner, the current study introduces the concept of ‘stochastic spanning’, which can be viewed as a model-free alternative to M-V spanning (Huberman and Kandel (1987)). Spanning occurs if introducing new securities or relaxing investment constraints does not improve the investment possibility set for a given class of investors. We develop methods for implementing the concept of stochastic spanning, which, unlike M-V spanning, accounts for higher-order moment risk in addition to variance.

Higher-order moment risk is arguably more relevant for analyzing spanning than for efficiency. Efficiency tests are generally applied to a given broad market index with limited skewness and kurtosis (at the typical monthly to annual return

frequency), in which case the arguments of Levy and Markowitz (1979) for the mean-variance approximation are compelling. By contrast, a spanning test evaluates all feasible portfolios, including those concentrated in a small number of risky securities, for which the same arguments are unlikely to hold.

Unfortunately, the spanning question is analytically difficult to resolve for parametric families of non-normal distributions, among other things, because relevant distributions such as the log-normal are not stable and the statistical calculus is complicated. For the aforementioned three-moment model, simplifications arise for important families of parametric distributions, but this approach does not account for higher-order moments. This study attempts to circumvent the analytical challenges by developing a statistical inference methodology and computational strategy that are based on a nonparametric assumption framework.

We propose a theoretical measure for stochastic spanning and derive the exact limit distribution for the associated empirical test statistic for a general class of dynamic processes. In addition, we develop consistent and feasible test procedures based on subsampling and Linear Programming (LP). A Monte-Carlo simulation experiment shows good statistical size and power properties in finite samples of realistic dimensions.

Spanning involves the comparison of two choice sets, with pairwise dominance analysis and portfolio efficiency analysis arising as special cases that assume that one or two of the choice sets is a singleton. In this respect, we expect that our inference and optimization methods have a wider applicability for SD analysis.

Our focus is on the most common SD criterion of second-order stochastic dominance (SSD), which has a well-established economic interpretation in terms of expected utility theory and Yaari's (1987) dual theory of risk. Extensions to the first-order rule (FSD) and third-order rule (TSD) would require large-scale mixed-integer programs and quadratic programs, respectively, which are computationally demanding when embedded in re-sampling routines.

The proofs to our propositions are available in the separate Supplementary Appendix.

2 Stochastic Spanning

The investment universe consists of M base assets with random investment returns $X := (x_1, \dots, x_M)$ with support bounded by $\mathcal{X}^M := [\underline{x}, \bar{x}]^M$, $-\infty < \underline{x} < \bar{x} < +\infty$. \mathcal{X} can be chosen arbitrarily if it is a superset of the maximal support of the base

assets. It does not seem realistic to allow for unbounded investment opportunities, because of the risk of financial ruin and the associated negative spill-overs to counterparties. For any realistic investment problem, private contracts, law and regulation will limit the investment possibilities. These restrictions will, for example, prevent that a risk neutral investor will borrow an infinite amount of money and take an infinite and concentrated position in a single high-risk security.

In this study, the M -simplex $\Lambda := \{\lambda \in \mathbb{R}_+^M : 1_M^T \lambda = 1\}$ represents the investment opportunity set. Importantly, the base assets are not restricted to be individual securities. In general, the base assets are defined as the vertices of the opportunity set, or the most extreme feasible combinations of the individual securities.

For example, some of the base assets could include a short position in a riskless asset and a long position in risky assets, to allow for bounded riskless borrowing. Similarly, some of the base assets could include a short position in risky assets and a long position in the riskless asset, to allow for bounded short sales.

Since the portfolio set is represented in vertex form rather than halfspace form, the convexity constraint $1_M^T \lambda = 1$ should not be confused with the classic budget constraint. Relaxing the budget constraint would affect the number of and the composition of the base assets rather than the convexity constraint.

The analysis considers a myopic, single-period choice problem. However, dynamic intertemporal choice problems could be allowed for by considering base assets that are periodically rebalanced based on conditioning information.

Let $F : \mathbb{R}^M \rightarrow [0, 1]$ denote the continuous joint c.d.f. of X and $F(y, \lambda) := \int 1(X^T \lambda \leq y) dF(X)$ the marginal c.d.f. for portfolio $\lambda \in \Lambda$. In order to define stochastic dominance and stochastic efficiency, we use the following integrated c.d.f.:

$$F^{(2)}(x, \lambda) := \int_{-\infty}^x F(y, \lambda) dy = \int_{-\infty}^x (x - y) dF(y, \lambda). \quad (1)$$

This measure corresponds to Bawa's (1975) first-order lower-partial moment, or expected shortfall, for return threshold $x \in \mathcal{X}$.

Definition 1. (Weak Stochastic Dominance): *Portfolio $\lambda \in \Lambda$ weakly second-order stochastically dominates portfolio $\tau \in \Lambda$ or $\lambda \succeq_F \tau$, if*

$$G(x, \lambda, \tau; F) \leq 0 \quad \forall x \in \mathcal{X}; \quad (2)$$

$$G(x, \lambda, \tau; F) := F^{(2)}(x, \lambda) - F^{(2)}(x, \tau). \quad (3)$$

Definition 2. (Strict Stochastic Dominance): *Portfolio $\lambda \in \Lambda$ strictly second-order stochastically dominates portfolio $\tau \in \Lambda$ or $\lambda \succ_F \tau$, if*

$$(\lambda \succeq_F \tau) \wedge (G(x, \lambda, \tau; F) < 0 \text{ for some } x \in \mathcal{X}). \quad (4)$$

A well-known equivalent formulation says that stochastic dominance occurs if and only if $\lambda \in \Lambda$ is preferred to $\tau \in \Lambda$ by all risk averters; see Hadar and Russell (1969), Hanoch and Levy (1969) and Rothschild and Stiglitz (1970).

Definition 3. (Stochastic Efficiency): *Portfolio $\tau \in \Lambda$ is second-order stochastically efficient if there exists no other feasible portfolio that strictly second-order stochastically dominates it: $\lambda \not\succeq_F \tau \forall \lambda \in \Lambda$.*

Equivalently, stochastic efficiency occurs if and only if portfolio $\tau \in \Lambda$ is the optimum for some risk averters (Post (2003, Thm 1)). This result relies on convexity of the choice set Λ , which allows us to apply Sion's (1958) minimax theorem to the joint analysis of portfolio weights and risk preferences. By contrast, for discrete choice sets, non-dominance does not imply optimality (Fishburn, 1974). Stochastic efficiency is not a trivial property. Notably, Post (2003) shows that a broad stock market index is significantly stochastically inefficient relative to a set of actively managed stock portfolios.

We use $E(\Lambda) := \{\tau \in \Lambda : \lambda \not\succeq_F \tau \forall \lambda \in \Lambda\}$ to denote the set of all stochastically efficient portfolios. $E(\Lambda)$ is a model-free generalization of the M-V efficient set. For important families of parametric distributions, $E(\Lambda)$ is a proper subset of the M-V efficient set (Ali (1975)). For these distributions, the M-V set is larger than $E(\Lambda)$ because the M-V rule can assign an irrationally high weight to variance. In general, however, the two efficient sets are not nested, because the mean and the variance do not capture all lower partial moments $F^{(2)}(x, \lambda)$, $x \in \mathcal{X}$.

This study focuses on the effects of changing the set of base assets or investment constraints. For this purpose, we introduce a non-empty polyhedral subset $K \subset \Lambda$. A polyhedral structure is analytically convenient and arises naturally if we remove some of the base assets or tighten the linear constraints which define Λ . The concluding section briefly discusses the case with convex subsets (which can be non-polyhedral) or simplicial complex subsets (which can be non-convex).

Definition 4. (Stochastic Spanning): *Portfolio set Λ is second-order stochastically spanned by subset $K \subset \Lambda$ if all portfolios $\lambda \in \Lambda$ are weakly second-order stochastically dominated by some portfolios $\kappa \in K$:*

$$\begin{aligned} & (\kappa \succeq_F \lambda \ \kappa \in K) \ \forall \lambda \in \Lambda \iff \\ & ((G(x, \kappa, \lambda; F) \leq 0 \ \forall x \in \mathcal{X}) \ \kappa \in K) \ \forall \lambda \in \Lambda. \end{aligned} \tag{5}$$

We will use $R(\Lambda) := \{K \subseteq \Lambda : (\kappa \succeq_F \lambda \ \kappa \in K) \ \forall \lambda \in \Lambda\}$ to denote all relevant subsets that span Λ . Spanning occurs if and only if $K \in R(\Lambda)$. $R(\Lambda)$ is non-empty because it includes at least Λ ; a span $K \in R(\Lambda)$ may itself be spanned by another span $K' \in R(K) \subseteq R(\Lambda)$.

This study analyzes a given subset $K \subset \Lambda$. In other applications, it may be interesting to find an irreducible span $K \subseteq R(\Lambda)$, so that $R(K) = K$. However, there generally exist multiple irreducible spans due to the possibility that two distinct portfolios have equivalent returns.

Below, we will discuss the relations between stochastic spanning, stochastic efficiency, expected utility and mutual fund separation, and introduce a measure for stochastic spanning.

Proposition 1. *Stochastic spanning occurs if the enlargement $(\Lambda - K)$ does not change the efficient set, that is,*

$$K \in R(\Lambda) \iff E(\Lambda) \subseteq K. \tag{6}$$

The reverse relation generally does not hold, because the weak dominance relation does not possess the antisymmetric property. In other words, $E(\Lambda)$ always spans Λ , but it may be reducible by excluding equivalent elements. Consequently, $E(\Lambda) \subseteq K$ is a sufficient but not necessary condition for $K \in R(\Lambda)$. In addition, the sufficient condition $E(\Lambda) \subseteq K$ is not practical, because $E(\Lambda)$ is generally non-convex and disconnected, which makes it difficult to identify all its elements and test the sufficient condition directly. On the contrary, a small polyhedral span $K \in R(\Lambda)$ could be used as a practical approximation to the intractable efficient set $E(\Lambda)$.

We use the following scalar-valued functional of the population c.d.f. as a measure for deviations from stochastic spanning:

$$\eta(F) := \sup_{\lambda \in \Lambda} \inf_{\kappa \in \mathbf{K}} \sup_{x \in \mathcal{X}} G(x, \kappa, \lambda; F). \quad (7)$$

The outer maximization searches for a feasible portfolio $\lambda \in \Lambda$ that is not weakly dominated by a portfolio $\kappa \in \mathbf{K}$. If $\eta(F) = 0$, then no such portfolio exists and \mathbf{K} spans Λ ; if $\eta(F) > 0$, then stochastic spanning does not occur.

Remark 1. Since $G(\bar{x}, \kappa, \lambda; F) = \mathbb{E}_F[X^T \lambda - X^T \kappa]$, we find the following lower bound for the stochastic spanning measure:

$$\begin{aligned} \eta(F) &\geq \sup_{\lambda \in \Lambda} \inf_{\kappa \in \mathbf{K}} G(\bar{x}, \kappa, \lambda; F) \\ &= \sup_{\lambda \in \Lambda} \inf_{\kappa \in \mathbf{K}} \mathbb{E}_F[X^T \lambda - X^T \kappa]. \end{aligned} \quad (8)$$

To further clarify the economic meaning of the notion of stochastic spanning, we can formulate it in terms of expected utility, by analogy to the aforementioned formulation of dominance and efficiency.

Proposition 2. *The stochastic spanning measure (7) can be reformulated as follows:*

$$\eta(F) = \sup_{\lambda \in \Lambda; w \in \mathcal{W}} \inf_{\kappa \in \mathbf{K}} H(w, \kappa, \lambda; F); \quad (9)$$

$$H(w, \kappa, \lambda; F) := \int_{\underline{x}}^{\bar{x}} w(x) G(x, \kappa, \lambda; F) dx; \quad (10)$$

$$\mathcal{W} := \left\{ w : \mathcal{X} \rightarrow [0, 1] : \int_{\underline{x}}^{\bar{x}} w(x) dx = 1 \right\}. \quad (11)$$

Alternatively,

$$\eta(F) = \sup_{\lambda \in \Lambda; u \in \mathcal{U}_2} \inf_{\kappa \in \mathbf{K}} \mathbb{E}_F [u(X^\top \lambda) - u(X^\top \kappa)]; \quad (12)$$

$$\mathcal{U}_2 := \left\{ u \in \mathcal{C}^0 : u(y) = \int_{\bar{x}}^y w(x) r(y; x) dx \ w \in \mathcal{W} \right\}; \quad (13)$$

$$r(y; x) := (y - x)1(y \leq x), \ (x, y) \in \mathcal{X}^2. \quad (14)$$

In this formulation, \mathcal{U}_2 is a set of normalized, increasing and concave utility functions that are constructed as convex mixtures of elementary Russell and Seo (1989) ramp functions $r(y; x)$, $x \in \mathcal{X}$. Stochastic spanning ($\eta(F) = 0$) occurs if no risk averter ($u \in \mathcal{U}_2$) benefits from the enlargement ($\Lambda - \mathbf{K}$). The lower bound (8) represents the potential benefit of the enlargement to a risk-neutral investor with utility function $u(y) = (y - \bar{x})$.

Stochastic spanning can also be formulated in terms of mutual fund separation; in portfolio theory, N -fund separation occurs if all rational risk averters combine at most $N \in \mathbb{N}_1$ distinct mutual funds (see, for example, Ross (1978)). If we assume a multivariate normal distribution and free portfolio formation, then two-fund separation arises ($N \leq 2$). Our definition of stochastic spanning however allows for non-normality and investment restrictions. Using the Minkowski-Weyl Theorem, the nested portfolio set $\mathbf{K} \subset \Lambda$ can be represented as the convex hull of its $V(\mathbf{K}) \in \mathbb{N}_1$ vertices. Hence, in case of stochastic spanning, rational investors can limit their attention to combining the $V(\mathbf{K})$ vertices of \mathbf{K} , and thus $N \leq V(\mathbf{K})$.

Appendix A discusses the relation between our analysis of stochastic spanning and the study of Scaillet and Topaloglou (2010), which was an important source of inspiration for our analysis.

3 Statistical Theory

In empirical applications, the c.d.f. F is latent and the analyst has access to a discrete time series of realized returns $s_T := (X_t)_{t=1}^T$, $X_t \in \mathcal{X}$, $t = 1, \dots, T$. This section analyses the asymptotic behavior of a test statistic for stochastic spanning in situations in which the number of assets M is fixed and the number of time series observations T goes to infinity, which in practice means that M is much smaller than T .

We make the following general assumptions on the multivariate return process:

Assumption 1. (i) The return sequence $(X_t)_{t \in \mathbb{N}_0}$ is α -mixing with mixing coefficients $(a_t)_{t \in \mathbb{N}_0}$ such that $a_t = \mathcal{O}(t^{-\delta})$ for some $\delta > 1$. (ii) Furthermore, the covariance matrix

$$\mathbb{E}_F [(X_0 - \mathbb{E}_F[X_0])(X_0 - \mathbb{E}_F[X_0])^\top] + 2 \sum_{t=1}^{\infty} \mathbb{E}_F [(X_0 - \mathbb{E}_F[X_0])(X_t - \mathbb{E}_F[X_t])^\top]$$

is positive definite.

These assumptions allow for various stationary ARMA, GARCH and stochastic volatility processes based on innovations with appropriately bounded supports (Carrasco and Chen (2002)).

Let $F_T(x) := T^{-1} \sum_{t=1}^T 1(X_t \leq x)$ denote the empirical joint c.d.f. constructed from the sample s_T . The multivariate empirical process CLT for strongly mixing sequences implies that $\sqrt{T}(F_T - F)$ weakly tends to the Gaussian process \mathcal{B}_F with covariance kernel given by $\text{Cov}(\mathcal{B}_F(x), \mathcal{B}_F(y)) = \sum_{t \in \mathbb{Z}} \text{Cov}(1(X_0 \leq x), 1(X_t \leq y))$ and almost surely uniformly continuous sample paths defined on \mathbb{R}^M (see Thm 7.3 of Rio (2013)).

We consider the following scaled empirical analogue of (7) as a test statistic for stochastic spanning:

$$\eta_T := \sqrt{T} \eta(F_T) = \sqrt{T} \sup_{\lambda \in \Lambda} \inf_{\kappa \in \mathbf{K}} \sup_{x \in \mathcal{X}} G(x, \kappa, \lambda; F_T) \quad (15)$$

$$= \sqrt{T} \sup_{\lambda \in \Lambda; w \in \mathcal{W}} \inf_{\kappa \in \mathbf{K}} H(w, \kappa, \lambda; F_T). \quad (16)$$

In general, computing the test statistic η_T is a challenging global optimization problem. Appendix C forwards two alternative computational strategies based on simulation or enumeration of a large number of small LP problems.

We use the test statistic η_T to test the null hypothesis of stochastic spanning, $\mathbf{H}_0 : \eta(F) = 0$, against the alternative hypothesis of no stochastic spanning, $\mathbf{H}_1 : \eta(F) > 0$. To derive the limit distribution of the test statistic under the null, we first introduce some additional notation.

Under the null, the set $\Gamma := \mathcal{W} \times \Lambda$ can be partitioned into the following two subsets:

$$\Gamma^= := \left\{ (w, \lambda) \in \Gamma : \inf_{\kappa \in \mathbf{K}} H(w, \kappa, \lambda; F) = 0 \right\}; \quad (17)$$

$$\Gamma^< := \left\{ (w, \lambda) \in \Gamma : \inf_{\kappa \in \mathbf{K}} H(w, \kappa, \lambda; F) < 0 \right\}. \quad (18)$$

Since $\mathbf{K} \subseteq \Lambda$, we find $\Gamma^= \neq \emptyset$. In addition, for any $(w, \lambda) \in \Gamma$, \mathbf{K} can be decomposed into the following two subsets:

$$\mathbf{K}^{\leq}_{(w, \lambda)} := \{ \kappa \in \mathbf{K} : H(w, \kappa, \lambda; F) \leq 0 \mid (w, \lambda) \in \Gamma \}; \quad (19)$$

$$\mathbf{K}^{>}_{(w, \lambda)} := \{ \kappa \in \mathbf{K} : H(w, \kappa, \lambda; F) > 0 \mid (w, \lambda) \in \Gamma \}. \quad (20)$$

Under the null, we have that $(H(w, \kappa, \lambda; F) \leq 0 \mid \forall w \in \mathcal{W} \mid \kappa \in \mathbf{K})$ for all $\lambda \in \Lambda$, and hence $\mathbf{K}^{\leq}_{(w, \lambda)} \neq \emptyset$ for all $(w, \lambda) \in \Gamma$.

Proposition 3. *Under Assumption 1,*

$$H \left(w, \kappa, \lambda; \sqrt{T}(F_T - F) \right) \rightsquigarrow H(w, \kappa, \lambda; \mathcal{B}_F); \quad (21)$$

$$\begin{aligned} \operatorname{oper}_{(w, \lambda) \in A_T} \operatorname{oper}_{\kappa \in B_T} H \left(w, \kappa, \lambda; \sqrt{T}(F_T - F) \right) &\rightsquigarrow \\ \operatorname{oper}_{(w, \lambda) \in A} \operatorname{oper}_{\kappa \in B} H(w, \kappa, \lambda; \mathcal{B}_F). &\quad (22) \end{aligned}$$

where \rightsquigarrow denotes weak convergence; oper and oper^* are sup or inf; A_T and A are measurable subsets of Γ such that $A_T \rightarrow A$; B_T and B are measurable subsets of \mathbf{K} such that $B_T \rightarrow B$.

The following proposition establishes the asymptotic distribution of the test statistic η_T under the null:

Proposition 4. *If Assumption 1 holds and \mathbf{H}_0 is true, then*

$$\eta_T \rightsquigarrow \eta_\infty := \sup_{(w, \lambda) \in \Gamma^=} \inf_{\kappa \in \mathbf{K}^{\leq}_{(w, \lambda)}} H(w, \kappa, \lambda; \mathcal{B}_F). \quad (23)$$

Notice that $H(\cdot, \cdot, \cdot; \mathcal{B}_F)$ is a well defined zero-mean Gaussian process due to the moment existence condition and the rate of convergence of the mixing coefficients in Assumption 1 (see for example inequality 1.12b in Rio (2013)). We were able to also derive asymptotic unbiasedness for a class of non-trivial local alternative hypotheses. For the sake of compactness, we do not report these additional results here and we focus on testing the null hypothesis of stochastic spanning ($\mathbf{H}_0 : \eta(F) = 0$).

Given the asymptotic null distribution, we can develop a test procedure based on η_T and η_∞ . Let $q(\eta_\infty, 1 - \alpha)$ denote the $(1 - \alpha)$ quantile of the distribution of η_∞ for any significance level $\alpha \in]0, 1[$. The basic decision rule to reject \mathbf{H}_0 against \mathbf{H}_1 if and only if $\eta_T > q(\eta_\infty, 1 - \alpha)$ is infeasible due to the dependence of $q(\eta_\infty, 1 - \alpha)$ on the latent c.d.f. F . However, feasible decision rules can be obtained by using a subsampling procedure to estimate $q(\eta_\infty, 1 - \alpha)$ from the data.

To implement the subsampling procedure, we begin by generating $(T - b_T + 1)$ maximally overlapping subsamples of $b_T \in \mathbb{N}_1$ consecutive observations, $s_{b_T;T,t} := (X_s)_{s=t}^{t+b_T-1}$, $t = 1, \dots, T - b_T + 1$, and compute test scores $\eta_{b_T;T,t} = \sqrt{b_T} \eta(F_{b_T;T,t})$ for each subsample, where $F_{b_T;T,t}$ denotes the empirical joint c.d.f. constructed from $s_{b_T;T,t}$, $t = 1, \dots, T - b_T + 1$. The distribution of subsample test scores can be described by the following c.d.f. and quantile function:

$$S_{T,b_T}(y) := \frac{1}{T - b_T + 1} \sum_{t=1}^{T-b_T+1} 1(\eta_{b_T;T,t} \leq y); \quad (24)$$

$$q_{T,b_T}(1 - \alpha) := \inf_y \{y : S_{T,b_T}(y) \geq 1 - \alpha\}. \quad (25)$$

Our decision rule is to reject the null $\mathbf{H}_0 : \eta(F) = 0$ against the alternative $\mathbf{H}_1 : \eta(F) > 0$ at a significance level of $\alpha \in]0, 1[$ if and only if $\eta_T > q_{T,b_T}(1 - \alpha)$, or, equivalently, $1 - S_{T,b_T}(\eta_T) < \alpha$. As shown in Appendix B, this subsampling routine is asymptotically exact and consistent under reasonable assumptions on the subsample length and significance level.

Although the test has asymptotically correct size, simulation exercises show that the quantile estimates $q_{T,b_T}(1 - \alpha)$ may be biased and sensitive to the subsample size b_T in finite samples of realistic dimensions (M and T). To correct for small-sample bias and reduce the sensitivity to the choice of b_T , we propose a regression-based bias-correction method that is motivated by our observations from simulation exercises. For a given significance level α , we compute the quantiles $q_{T,b_T}(1 - \alpha)$ for a ‘reasonable’ range of the subsample size b_T . Next, we estimate the intercept and

slope of the following regression line using OLS regression analysis:

$$q_{T,b_T}(1 - \alpha) = \gamma_{0;T,1-\alpha} + \gamma_{1;T,1-\alpha}(b_T)^{-1} + \nu_{T;1-\alpha,b_T}. \quad (26)$$

Finally, we estimate the bias-corrected $(1 - \alpha)$ -quantile as the OLS predicted value for $b_T = T$:

$$q_T^{BC}(1 - \alpha) := \hat{\gamma}_{0;T,1-\alpha} + \hat{\gamma}_{1;T,1-\alpha}(T)^{-1}. \quad (27)$$

Since $q_{T,b_T}(1 - \alpha)$ converges in probability to $q(\eta_\infty, 1 - \alpha)$ and $(b_T)^{-1}$ converges to zero as $T \rightarrow \infty$, $\hat{\gamma}_{0;T,1-\alpha}$ converges in probability to $q(\eta_\infty, 1 - \alpha)$ and the asymptotic properties are not affected. However, computational experiments show that the bias-corrected method is more efficient and more powerful in small samples.

The (block) bootstrap is an obvious alternative to subsampling. Proposition 4 is based on the properties of the partitions of Γ and K in (17) to (20) and the behavior of the measure $\eta(F)$ on these subsets. Given the relevant discussion on page S67 of LPW2014, we expect that the use of a bootstrap critical value based on the appropriately centered H would lead to a test which is consistent but asymptotically conservative, and hence, less powerful than the subsampling approach under particular local alternatives. However, we believe that we can obtain an asymptotically exact bootstrap procedure without centering, if we strengthen the null hypothesis to hold for any element of some weak neighborhood of F . In any case, we expect that the bootstrap is more powerful in finite samples than subsampling, since each pseudo-sample utilizes the full sample information, rather than a subset of the observations. We leave the development of a bootstrap procedure for stochastic spanning for further research.

4 Simulation Experiment

We use a Monte Carlo simulation experiment to analyze the small-sample properties of our test procedure for stochastic spanning and compare those properties with similar results for M-V spanning tests.

Typical data sets in empirical asset pricing consist of low-frequency returns to diversified benchmark portfolios for multiple asset classes, market segments or investment styles. In this context, the primary factors that determine the statistical

performance of the test procedure seem to be the number of base assets (M), the number of time-series observations (T), the mutual covariance structure and the risk premiums of the risky assets. A serially i.i.d. multivariate normal return distribution is a convenient way to capture these factors, although it cannot capture the effects of higher-moment risk and weak dependence.

Our experiment is based on an investment problem with a riskless asset, a stock index futures contract and $(M - 2)$ other risky assets. We focus on testing the hypothesis that all convex combinations of the M base assets (Λ) are spanned by all convex combinations of the riskless asset and the futures contract (K). In this setup, spanning amounts to 'two-fund separation', where the riskless asset and the futures contract are the two relevant funds.

The joint return distribution is serially i.i.d. normal with a mutual covariance matrix that is fitted to the empirical distribution of monthly returns to the CRSP all-share index and active stock portfolios from July 1963 to December 2015 from the data library of Kenneth French. We set $\underline{x} = \min_{i,t}(x_{i,t})$ and $\bar{x} = \max_{i,t}(x_{i,t})$.

The normal distribution is unbounded, which is unrealistic and violates our assumption framework. However, truncating the normal distributions in the tails has no material effect on our simulated size and power properties.

We consider sets of $(M - 2) = 6, 25$ value-weighted portfolios that are formed by classifying stocks based on their market capitalization of equity (ME) and book-to-market-equity ratio (B/M). The risk-free return $x_1 = r$ is the one-month T-bill rate and the futures contract is built using a short position of 100% in the T-bill and a long position of 200% in the market index, so that its return is given by $x_2 = 2y - r$, where y is the index return. This futures contract obeys the spot-futures parity and requires that 50 percent margin is deposited in an interest-bearing account.

We equate the risk-free return r and the expected return to the market index $\mathbb{E}_F[y]$ with the corresponding historical averages. For every risky asset, we set the expected return using the following linear mean-beta relation:

$$\mathbb{E}_F[x_i] = \mathbb{E}_F[y] + \xi(\beta_i - 1)(\mathbb{E}_F[y] - r), \quad i = 3, \dots, M. \quad (28)$$

In this expression, β_i is the market beta and $0 \leq \xi < 1$ is a parameter which controls the deviations from the null. To measure the statistical size, we set $\xi = 1$, which yields the Security Market Line equation of the Capital Asset Pricing Model and which is consistent with the spanning hypothesis:

$$\mathbb{E}_F[x_i] = r + \beta_i(\mathbb{E}_F[y] - r), \quad i = 3, \dots, M. \quad (29)$$

To measure the statistical power, we set ξ such that the expected return to the lowest-beta asset equals the risk-free return:

$$\min_{i=3,\dots,M} \mathbb{E}_F[x_i] = r, \quad (30)$$

or, put differently,

$$\xi = \frac{1}{1 - \min_{i=3,\dots,M} \beta_i}. \quad (31)$$

This specification violates the null of spanning, because mixtures of the T-bill and high-beta assets dominate mixtures of the T-bill and the index futures contract.

We generate random samples of size $T = 240, 480, 960$, which corresponds to 20, 40 or 80 years of monthly observations, and apply our tests for stochastic spanning and M-V spanning to every random sample. For the stochastic spanning test, the subsample lengths employed are $\{50, 55, 60, 65, 70, 75\}$ for $T = 240$, $\{60, 75, 90, 105, 120, 135\}$ for $T = 480$, and $\{80, 120, 160, 200, 240, 280\}$ for $T = 960$.

Since the test statistic is computed using hundreds of LP problems for every sample and subsample, simulating the performance of the subsampling procedure involves solving tens of millions of LP problems. The computational costs limit the possibilities to calibrate the subsample lengths and the optimization procedure, which may adversely affect the reported results.

Under the multivariate normal distribution, two-fund separation is equivalent to M-V efficiency of the market portfolio, by Tobin's (1958) separation theorem. We may therefore use tests for M-V efficiency to test for M-V spanning. We employ two tests for M-V efficiency: a classical one and another one based on subsampling.

The first test for M-V efficiency is the classical Gibbons, Ross and Shanken (GRS; 1989) test, which is based on Seemingly Unrelated Regression. In this experiment, our test procedure cannot rival the GRS test, which correctly assumes a serially i.i.d. normal distribution. In this respect, the GRS test functions as an ideal benchmark and our objective is not to outperform the GRS test but to measure the divergence between the performance of our procedure and that of the benchmark in small samples.

However, the reported performance for the GRS test is clearly not representative for dynamic and non-normal distributions. In order to separate the effect of the M-V criterion and the effect of assuming i.i.d. normality, we also include a subsampling

test for M-V efficiency which embeds the computation of the standard GRS test statistic in the same subsampling and bias correction methodology that is employed for the SSD spanning test.

Table I shows the size and power properties of the three tests as a function of the data dimensions (M and T). In small samples, important size distortions occur for the subsampling method, which are attributable to imperfect calibration of the grid points for the subsample lengths used in the bias correction method. To better compare the power, we also report 'size-adjusted power' as the difference between the unadjusted power and the size.

The GRS test performs very well, as expected, in this experiment based on a serially i.i.d. normal distribution. The statistical size is under control and the power approximates 100% in all relevant cases. The use of the SSD criterion and subsampling leads to a loss of power in small samples, which is a price to pay for the ability to deal with dynamic and non-normal distributions.

Clearly, a narrow cross-section and long time-series are more important for the stochastic spanning test than for the GRS test. For a broad cross-section ($M - 2 = 25$) and short time series $T = 240$, the stochastic spanning tests is rather under-powered, with a rejection rate of 55.1 percent at a nominal significance level of 5 percent under the alternative. Encouragingly, the power increases quickly as we reduce the number of assets or increase the number of observations and high power levels are achieved for realistic data dimensions.

A more detailed simulation study would also analyze the effects of higher-moment risk and serial dependence. However, for typical data sets of low-frequency returns to diversified benchmark portfolios, the empirical deviations from i.i.d. multivariate normality seem of secondary importance for the statistical properties of the test procedure, despite their importance for investors more generally.

There are three obstacles to verifying this conjecture in this study. First, parametric specifications of dynamic and non-normal multivariate distributions tend to be intractable for portfolio analysis. Second, the methods developed in the present study do not allow for constructing a span $K \in R(\Lambda)$ for a given portfolio set Λ , which complicates the design of an experiment for simulating the statistical size of the spanning test. Third, the computational burden prohibits experimentation with the design of the simulation process and calibration of the methods.

Nevertheless, indirect evidence from earlier experiments by Post and Versijp (2007) and Post and Poti (2017) supports the notion that higher-order risk and serial dependence are of secondary importance here. In those experiments, the simulated properties of stochastic efficiency tests are robust to independent sampling from

the empirical distribution function instead of the normal distribution. Unreported results show that the simulated properties are also not affected by sampling random blocks of 12 consecutive monthly observations instead of individual months. Given the similarities between the relevant stochastic orders and simulation conditions, the robustness in the earlier experiments can be expected to carry over to the present experiment.

[Insert Table I about here]

5 Empirical Application

This section applies efficiency and spanning tests to empirical data rather than simulated data. Motivated by the above simulation experiment, we use a relatively narrow cross-section and long time-series. Our investment universe consists of $M = 12$ distinct base assets: the one-month T-bill, an index futures contract based on the CRSP all-share index and ten equity industry portfolios. We analyze monthly excess returns from July 1926 to December 2014 ($T = 1,062$) from Kenneth French online data library. Returns are computed in excess of the monthly T-bill rate, which means that the bill is treated as a riskless asset and has an excess return of zero in every month.

Several features of these data justify our model-free approach to account for higher-order moment risk and time-series dynamics. Firstly, the return distribution appears non-normal, witness, for example, the skewness of -0.511 and excess kurtosis of 1.813 of the market returns. In addition, the data show clear dynamic patterns; for example, the first-order auto-correlation coefficient for the market returns is 8.52 percent (t-stat.: **X.XX**). The dimensions of the data set ($M = 12$, $T = 1,062$) also seem favorable for our model-free approach.

We find similar results as reported below in two sub-periods of roughly equal length, as well as for a second data set of ten portfolios formed on estimated market beta and a third data set of ten portfolios formed on ME.

We deliberately do not consider data sets of equal-weighted returns and/or double-sorted portfolios that are formed on ME and a second stock characteristic in order to avoid a bias towards micro-cap stocks that would lead to a predictable rejection of all our hypotheses and make the test results uninformative. This consideration does not play a role in the above simulation experiment, because the

simulation process was based on the theoretical mean-beta relation (28) rather than the historical means.

We first analyze whether the market portfolio is stochastically efficient. This hypothesis seems interesting because representative investor models of capital market equilibrium predict that the market portfolio is efficient as a result of risk sharing in sufficiently complete markets or, alternatively, aggregation across sufficiently homogeneous investors in incomplete markets. A market portfolio efficiency test can also be interpreted as a revealed preference analysis of those individual investors who adopt a passive strategy of broad diversification.

In this application, Λ consists of all convex combinations of the 12 base assets. There is no need to explicitly allow for short selling in this application, because the market portfolio has no binding short-sales restrictions; non-binding constraints do not affect the efficiency classification. All risky assets have strictly positive market capitalization weights. If some investor would benefit from short-selling some risky asset, then she would also benefit from underweighting that asset without using a negative weight. In other words, the short-sales constraints are not binding and hence do not affect the efficiency classification.

To test market portfolio efficiency, we use the Linton, Post and Whang (2014) test, using the same subsampling procedure as our spanning test. The four panels of Figure 1 illustrate our results.

The optimal solution $\lambda^* \in \Lambda$ consists of large positions in the nondurables industry (46%) and energy industry (42%) and small positions in the health industry (6%), telecom industry (5%) and T-bill (1%). In Panel A of Figure 1, the return PDF of λ^* appears less risky than that of the market portfolio. Panel B shows the difference function $G(x, \tau, \lambda^*; F_T)$ for every return level $x \in \bar{\mathcal{X}}_T = [-25.15, 42.07]$, from which it is clear that the market portfolio has a strictly higher expected short-fall than the solution portfolio for every return level $x \in \bar{\mathcal{X}}_T$; it follows that $\lambda^* \succ_{F_T} \tau$. The value of the Linton, Post and Whang (2014) test statistic is $\zeta_T = \sqrt{T} \min_{x \in \bar{\mathcal{X}}_T} G(x, \tau, \lambda^*; F_T) = 0.114$.

Panel C shows the decumulative subsampling distribution of the test statistic for subsample sizes $b_T = 120$ and $b_T = 480$. Clearly, large values of the test statistic occur more frequently in smaller subsamples, which underlines the need to correct the quantile estimates for bias. Panel D shows the estimated OLS regression line (26) based on the empirical quantiles $q_{T, b_T}(1 - \alpha)$ for significance levels of $\alpha = 0.01$ and $\alpha = 0.10$ using various subsample sizes $b_T \in [120, 480]$. Using (27), the regression estimate for the critical value for ζ_T is $q_T^{BC}(0.90) = 0.370$, more than three times the full-sample value $\zeta_T = 0.114$. Hence, we cannot reject market portfolio efficiency at

conventional significance levels.

[Insert Figure 1 about here.]

Our second research hypothesis is two-fund separation: do all rational risk averters combine the T-bill and the index futures contract?

In the simulation experiment, which was based on a multivariate normal DGP, this hypothesis was equivalent to market portfolio efficiency. For non-normal distributions, two-fund separation generally does not occur, unless one assumes that preferences are sufficiently similar across investors (see, for example, Cass and Stiglitz (1970)). Our stochastic spanning test can analyze two-fund separation without assuming a particular form for the return distribution or utility functions.

Figure 2 illustrates the estimation results for the industry data set. The optimal solution $\kappa^* \in \mathbf{K}$ consists of the T-bill (56%) and the index futures contract (44%). The optimal solution $\lambda^* \in \Lambda$ consists of a large position in the nondurables industry (42%) and smaller positions in the health industry (26%), energy industry (20%) and telecom industry (12%). Panel B shows the difference function $G(x, \kappa^*, \lambda^*; F_T)$ for every relevant return level $x \in \mathcal{X}$. Clearly, we find a strictly positive difference for large positive return levels and hence $\kappa^* \not\prec_{F_T} \lambda^*$; stochastic spanning does not occur. We find $\max_{x \in \mathcal{X}} G(x, \kappa^*, \lambda^*; F_T) = 0.138$ and the test statistic amounts to $\eta_T = 4.480$.

Panel C shows the decumulative subsampling distribution of the test statistic for $b_T = 120$ and $b_T = 480$ months, with large values of the test statistic again occurring more frequently in smaller subsamples. Panel D shows the estimated OLS regression line (26) for significance levels of $\alpha = 0.01$ and $\alpha = 0.10$ using various subsample sizes $b_T \in [120, 480]$. Using (27), the regression estimate for the critical value for η_T at $\alpha = 0.01$ is $q_T^{BC}(0.99) = 4.354$, below the full-sample value $\eta_T = 4.480$. Hence, we can reject two-fund separation with at least 99% confidence.

[Insert Figure 2 about here.]

As a final step in our analysis, we test for two-fund separation using the M-V criterion rather than the SSD criterion. Clearly, our rejection of stochastic spanning is less informative if we can also reject M-V spanning.

We use the same methodology as for the above stochastic spanning test, but we restrict the utility functions to take a quadratic (rather than piecewise linear) shape. We solve the embedded expected-utility optimization problems (for every

given quadratic utility function) using quadratic programming. This nested model specification isolates the effect of the choice criterion (SSD vs. M-V).

Figure 3 summarizes the test results. In contrast to stochastic spanning, we cannot reject M-V spanning at conventional significance levels.

[Insert Figure 3 about here.]

The combined results of the efficiency and spanning tests suggest that combining the T-bill and market portfolio is optimal for some risk averters (market portfolio efficiency) but suboptimal for other risk averters (no two-fund separation). Since market portfolio efficiency and two-fund separation are equivalent under a multivariate normal distribution, the divergence of our two sets of test results points at economically significant deviations from normality.

Harvey and Siddique (2000) and Dittmar (2002) analyze the empirical explanatory power of skewness and kurtosis in cross-sectional regression tests for market portfolio efficiency. Their results, as the results of our structural efficiency test, seem consistent with the notion that the market portfolio is optimal for some utility functions with higher-order moment risk preferences. We caution however against interpreting these results as evidence for representative-investor models of capital market equilibrium.

If returns are not normally distributed, then aggregation across individual efficient risky portfolios may not produce an efficient market portfolio. Our spanning test results suggest that distinct risk averters will hold distinct risky portfolios. Since the SSD efficient set is generally non-convex, aggregation across distinct efficient risky portfolios unfortunately does not produce an efficient market portfolio. Hence, we caution against confusing market portfolio efficiency and market equilibrium models if two-fund separation is rejected.

We should mention that the GRS test, in contrast to the subsampling test, rejects M-V efficiency of the market portfolio at every conventional significance level. Similar results are found using the MacKinlay and Richardson (1991) test which accounts for serial dependence and non-normality. However, our analysis aims to isolate the effect of the choice criterion (SSD vs. M-V) and the effect of the stochastic order (efficiency vs. spanning), using a nested model specification and a single statistical methodology (subsampling). The point here is that, keeping all else equal, we cannot reject market portfolio efficiency for either the SSD or M-V criteria, but we can reject two-fund separation using a stochastic spanning test.

6 Concluding Remarks

We have introduced the model-free concept of stochastic spanning together with a consistent and feasible framework for implementation based on subsampling and LP. Our simulation experiment shows good statistical size and power properties in finite samples of realistic dimensions. The empirical application illustrates our methodology and points at new evidence for the relevance of higher-order moment risk in portfolio theory and asset pricing.

We conclude with a brief discussion of various extensions and generalizations that appear non-essential for our empirical application but that may be of interest in other applications.

First, although this study has worked with a polyhedral spanning set, the results would go through with minor modifications if we allow K to be a non-polyhedral convex set or a simplicial complex. Notably, in the non-polyhedral convex case, the LP strategy in Proposition 8 could be substituted by some convex optimization method, whereas in the simplicial complex case, the strategy would be implemented in each one of the simplices that comprise the complex.

One interesting line of further research which builds on this generalization is to construct an ‘outer approximation’ of the efficient set by considering decreasing sequences of simplicial spanning sets.

Second, our statistical theory can be extended to the case of unbounded support for the base assets, with some minor modifications of the definition and assumption framework. If in such a case, \mathcal{W} is defined by the additional condition that $\int_{-\infty}^{+\infty} w(x) |x| dx \leq +\infty$, and Assumption 1 includes the conditions that $\delta > 2$, and $\mathbb{E} \|X\|^{2+\epsilon} < +\infty$, for some $\epsilon > 0$, then results partially analogous to Propositions 2-4 would hold with the relevant modifications.

Third, if Assumption 1 is strengthened according to Theorem 2.3 of Andrews and Pollard (1994) and via the use of Theorem B.0.1 of Politis, Romano and Wolf (1999), the testing procedure can be shown to be asymptotically unbiased under classes of sequences of local alternatives.

The authors are working to extend and generalize the present framework along these lines.

References

- [1] Ali, M.M., 1975, Stochastic dominance and portfolio analysis, *Journal of Financial Economics* 2, 205-229.
- [2] Andrews, D.W. and D. Pollard, 1994, An introduction to functional central limit theorems for dependent stochastic processes, *International Statistical Review* 62, 119-132.
- [3] Athayde G.M. and R.G. Flores, 2004, Finding a maximum skewness portfolio: A general solution to three-moments portfolio choice, *Journal of Economic Dynamics and Control* 28, 1335-1352.
- [4] Barrett, G. and S. Donald, 2003, Consistent tests for stochastic dominance, *Econometrica* 71, 71-104.
- [5] Bawa, V.S., 1975, Optimal Rules for Ordering Uncertain Prospects, *Journal of Financial Economics* 2, 95-121.
- [6] Bawa, V.S., J.N. Bodurtha Jr., M.R. Rao and H. L. Suri, 1985, On Determination of Stochastic Dominance Optimal Sets, *Journal of Finance* 40, 417-431.
- [7] Berk, J., 1997, Necessary conditions for the CAPM, *Journal of Economic Theory* 73, 245-257.
- [8] Carrasco, M. and X. Chen, 2002, Mixing and moment properties of various GARCH and stochastic volatility models, *Econometric Theory* 18, 17-39.
- [9] Cass, D., and J. Stiglitz, 1970, The structure of investor preferences and asset returns, and separability in portfolio allocation, *Journal of Economic Theory* 2, 122-160.
- [10] Chamberlain, G., 1983, A characterization of the distributions that imply mean-variance utility functions, *Journal of Economic Theory* 29, 185-201.
- [11] Davidson, R. and J.-Y. Duclos, 2000, Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality, *Econometrica* 68, 1435-1464.
- [12] Dittmar, R.F., 2002, Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns, *Journal of Finance* 57, 369-403.
- [13] Fishburn, P.C., 1974, Convex Stochastic Dominance with Continuous Distribution Functions, *Journal of Economic Theory* 7, 143-58.
- [14] Gibbons, M. Ross, S., and J. Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121-1152.
- [15] Hadar, J. and W.R. Russell, 1969, Rules for Ordering Uncertain Prospects, *American Economic Review* 59, 2-34.

- [16] Hanoch, G., and H. Levy, 1969, The Efficiency Analysis of Choices Involving Risk, *Review of Economic Studies* 36, 335-346.
- [17] Harvey, C. R., and A. Siddique, 2000, Conditional Skewness in Asset Pricing Tests, *Journal of Finance* 55, 1263–1295.
- [18] Huberman, G. and S. Kandel, 1987, Mean-Variance Spanning, *Journal of Finance* 42, 873-888.
- [19] Kuosmanen, T., 2004, Efficient diversification according to stochastic dominance criteria, *Management Science* 50, 1390-1406.
- [20] Levy, H., and H. Markowitz, 1979, Approximating Expected Utility by a Function of Mean and Variance, *American Economic Review* 69 (3), 308-317.
- [21] Linton, O.B., E. Maasoumi and Y.-J. Whang, 2005, Consistent Testing for Stochastic Dominance under General Sampling Schemes, *Review of Economic Studies* 72, 735-765.
- [22] Linton, O.B., Th. Post, and Y.-J. Whang, 2014, Testing for the Stochastic Dominance Efficiency of a Given Portfolio, *Econometrics Journal* 17, 59-74.
- [23] MacKinlay, A.C. and M. Richardson, 1991, Using generalized method of moments to test mean-variance efficiency, *Journal of Finance* 46, 511-527.
- [24] Mencia, J. and Sentana, E., 2009, Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation, *Journal of Econometrics* 153, 105-121.
- [25] Owen, J. and Rabinovitch, R., 1983, On the class of elliptical distributions and their applications to the theory of portfolio choice, *Journal of Finance* 38, 745-752.
- [26] Politis, D.N., J.P. Romano and M. Wolf, 1999, *Subsampling*. Springer New York.
- [27] Post, Th., 2003, Empirical Tests for Stochastic Dominance Efficiency, *Journal of Finance* 58, 1905-1932.
- [28] Post, Th., and Ph. Versijp, 2007, Multivariate tests for stochastic dominance efficiency of a given portfolio, *Journal of Financial and Quantitative Analysis* 42, 489-516.
- [29] Post, Th. and V. Potì, 2017, Portfolio Analysis using Stochastic Dominance, Relative Entropy and Empirical Likelihood, *Management Science* 63, 153-165.
- [30] Quirk, J.P. and R. Saposnik, 1962, Admissibility and Measurable Utility Functions, *Review of Economics Studies* 29, 140-146.
- [31] Rio, E., 2013, Inequalities and limit theorems for weakly dependent sequences. 3^eme cycle. pp.170.

- [32] Roman, D., K. Darby-Dowman and G. Mitra, 2006, Portfolio construction based on stochastic dominance and target return distributions, *Mathematical Programming* 108, 541-569.
- [33] Ross, S.A., 1978, Mutual Fund Separation in Financial Theory—The Separating Distributions, *Journal of Economic Theory* 17, 254-83.
- [34] Rothschild, M. and J.E. Stiglitz, 1970, Increasing Risk: I. A Definition, *Journal of Economic Theory*, 225-243.
- [35] Russell, W.R. and T.K. Seo, 1989, Representative sets for stochastic dominance rules, in: T.B. Fomby and T.K. Seo (eds.), *Studies in the economics of uncertainty: In honor of Josef Hadar*, New York: Springer Verlag, 59-76.
- [36] Scaillet, O., and N. Topaloglou, 2010, Testing for stochastic dominance efficiency, *Journal of Business and Economic Statistics* 28, 169-180.
- [37] Shalit, H., and S. Yitzhaki, 1994, Marginal Conditional Stochastic Dominance, *Management Science* 40, 670-684.
- [38] Simaan, Y., 1993, Portfolio selection and asset pricing-three-parameter framework, *Management Science* 39, 568-577.
- [39] Sion, M., 1958, On general minimax theorems, *Pacific Journal of Mathematics* 8, 171-176.
- [40] Tobin, J., 1958, Liquidity Preference as Behavior Towards Risk, *Review of Economic Studies* 25, 65-86.
- [41] Yaari, M.E., 1987, The dual theory of choice under risk, *Econometrica* 55, 95-115.

Appendix A: Superefficiency

Our definition of stochastic efficiency ([Definition 3](#)) should not be confused with an alternative definition by Scaillet and Topaloglou (2010, henceforth ST2010), which we label here as ‘stochastic superefficiency’:

Definition 5. (Stochastic superefficiency): *Portfolio $\tau \in \Lambda$ is second-order stochastically superefficient if it weakly second-order stochastically dominates all feasible portfolios, or $\tau \succeq_F \lambda \forall \lambda \in \Lambda$.*

Let $S(\Lambda) := \{\tau \in \Lambda : \tau \succeq_F \lambda \forall \lambda \in \Lambda\}$ denote the set of all superefficient portfolios. In order theory, $S(\Lambda)$ amounts to the set of greatest elements, whereas $E(\Lambda)$ is the set of maximal elements. Clearly, stochastic superefficiency gives a sufficient condition for stochastic efficiency; $(\tau \succeq_F \lambda \forall \lambda \in \Lambda) \Rightarrow (\lambda \not\prec_F \tau \forall \lambda \in \Lambda)$, or $S(\Lambda) \subseteq E(\Lambda)$. The reverse is not true, as all superefficient portfolios must be equivalent and comparable, whereas efficient portfolios may be non-equivalent or incomparable.

The superefficient set is either equal to the efficient set ($S(\Lambda) = E(\Lambda)$) or empty ($S(\Lambda) = \emptyset$). In our applications, the efficient set generally has non-equivalent and incomparable elements, and therefore $S(\Lambda) = \emptyset$. For example, an efficient portfolio that maximizes expected return generally takes a concentrated position in the individual asset with the highest mean. By contrast, an efficient portfolio that minimizes semi-variance generally takes a diversified position in multiple risky assets or a position in a risk-free asset.

Stochastic super-efficiency ($\tau \succeq_F \lambda \forall \lambda \in \Lambda$) occurs as the special case of stochastic spanning if the portfolio set K is a singleton, or $K = \{\tau\}$, $\tau \in \Lambda$. In this case, our measure reduces to

$$\eta(F) = \sup_{\lambda \in \Lambda} \sup_{x \in \mathcal{X}} G(x, \kappa, \lambda; F). \quad (32)$$

Furthermore, our test statistic (15) in this case equals the superefficiency test statistic of ST2010:

$$\eta_T = \sqrt{T} \sup_{\lambda \in \Lambda} \sup_{x \in \mathcal{X}} G(x, \tau, \lambda; F_T). \quad (33)$$

Our statistical theory thus also applies to the ST2010 test statistic. Notably, using Proposition 4, we obtain the exact limit distribution of the ST2010 test statistic as the law of

$$\eta_\infty = \sup_{(w, \lambda) \in \Gamma^\infty} H(w, \tau, \lambda; \mathcal{B}_F). \quad (34)$$

Appendix B: Subsampling Estimation

This section analyzes the asymptotic properties of the subsampling procedure that is described in [Section 3](#). Our analysis uses $M^* \in \mathbb{N}_0$, $M^* < V(\Lambda)$ for the number of vertices of Λ that are also included in K .

The following (non-trivial) properties of the limit distribution are essential to motivate our use of subsampling, by allowing us to invoke established results of Politis *et al.* (1999):

Proposition 5. *Under [Assumption 1](#), (i) the distribution of η_∞ has support $[0, +\infty[$; (ii) the c.d.f. of η_∞ may have a jump discontinuity with a size of at most (M^*/M) at zero; (iii) the c.d.f. of η_∞ is continuous on $]0, +\infty[$.*

To implement the subsampling procedure we begin by generating $(T - b_T + 1)$ maximally overlapping subsamples of $b_T \in \mathbb{N}_1$ consecutive observations, $s_{b_T;T,t} := (X_s)_{s=t}^{t+b_T-1}$, $t = 1, \dots, T - b_T + 1$, and compute test scores $\eta_{b_T;T,t} = \sqrt{b_T} \eta(F_{b_T;T,t})$ for each subsample, where $F_{b_T;T,t}$ denotes the empirical joint c.d.f. constructed from $s_{b_T;T,t}$, $t = 1, \dots, T - b_T + 1$. The distribution of subsample test scores can be described by the following c.d.f. and quantile function:

$$S_{T,b_T}(y) := \frac{1}{T - b_T + 1} \sum_{t=1}^{T-b_T+1} \mathbf{1}(\eta_{b_T;T,t} \leq y); \quad (35)$$

$$q_{T,b_T}(1 - \alpha) := \inf_y \{y : S_{T,b_T}(y) \geq 1 - \alpha\}. \quad (36)$$

To establish the statistical properties of this subsampling procedure, we assume that the subsample size b_T and significance level α are selected appropriately:

Assumption 2. *The positive sequence (b_T) , possibly dependent on $(X_t)_{t=1}^T$, obeys*

$$\mathbb{P}(l_T \leq b_T \leq u_T) \rightarrow 1, \quad (37)$$

where (l_T) and (u_T) are deterministic sequences of natural numbers such that $1 \leq l_T \leq u_T$ for all T , $l_T \rightarrow \infty$ and $u_T/T \rightarrow 0$ as $T \rightarrow \infty$.

Assumption 3. *The significance level obeys $\alpha < 1 - (M^*/M)$.*

Since K is a proper subset of Λ , we can safely assume that $M^* < V(\Lambda)$. The smaller the overlap between K and Λ , the higher the significance level that we can employ under [Assumption 3](#).

The following proposition shows that our test based on the subsample critical value is asymptotically exact and consistent:

Proposition 6. *If [Assumptions 1-3](#) hold, then we find the following asymptotic size and power properties:*

$$\lim_{T \rightarrow \infty} \mathbb{P}(\eta_T > q_{T,b_T}(1 - \alpha) | \mathbf{H}_0) = \alpha; \quad (38)$$

$$\lim_{T \rightarrow \infty} \mathbb{P}(\eta_T > q_{T,b_T}(1 - \alpha) | \mathbf{H}_1) = 1. \quad (39)$$

Appendix C: Computational Strategy

This section outlines two possible strategies for computing the test statistic η_T using LP.

If the enlargement $(\Lambda - K)$ is small, then we may perform a quasi-Monte Carlo simulation and solve an embedded LP problem for every simulated portfolio $\lambda \in (\Lambda - K)$. Specifically, we can use the following reformulation of [\(15\)](#):

$$\eta_T = -\sqrt{T} \inf_{\lambda \in (\Lambda - K)} \eta_T(\lambda); \quad (40)$$

$$\eta_T(\lambda) := \sup_{\kappa \in K} \inf_{x \in \mathcal{X}} G(x, \lambda, \kappa; F_T). \quad (41)$$

The embedded statistic $\eta_T(\lambda)$ can be computed by solving an LP problem:

Proposition 7. *The embedded test statistic $\eta_T(\lambda)$ equals the optimal value of the objective function of the following LP problem in canonical form:*

$$\begin{aligned}
& \max \sqrt{T}\gamma & (42) \\
\text{s.t. } & \gamma + T^{-1} \sum_{t=1}^T \theta_{s,t} \leq F_T^{(2)}(X_s \lambda, \lambda), \quad s = 1, \dots, T; \\
& -\theta_{s,t} - X_t^T \kappa \leq -X_s \lambda, \quad s, t = 1, \dots, T; \\
& \sum_{i=1}^M \kappa_i = 1; \\
& \theta_{s,t} \geq 0, \quad s, t = 1, \dots, T; \\
& \kappa_i \geq 0, \quad i = 1, \dots, M; \\
& \gamma \text{ free.}
\end{aligned}$$

Although the problem has $\mathcal{O}(T^2 + M)$ variables and constraints, for a specific portfolio λ the computational burden is perfectly manageable with modern-day computer hardware and solver software for the typical data dimensions in empirical asset pricing research. Nevertheless, we need to solve the LP problem for a sufficiently large number of portfolios $\lambda \in (\Lambda - K)$ and the computational burden will therefore explode if the enlargement $(\Lambda - K)$ is large. For example, in our application in [Section 5](#), K is a 2-simplex and Λ is a 11-simplex; this enlargement is too large to allow for an accurate and manageable discrete approximation.

An alternative strategy seems more appropriate when the enlargement $(\Lambda - K)$ is large but the return range $(\bar{x} - \underline{x})$ is limited. Using [\(12\)](#) and [\(15\)](#), we find

$$\eta_T = \sqrt{T} \sup_{u \in \mathcal{U}_2} \left(\sup_{\lambda \in \Lambda} \mathbb{E}_{F_T} [u(X^T \lambda)] - \sup_{\kappa \in K} \mathbb{E}_{F_T} [u(X^T \kappa)] \right). \quad (43)$$

The term in parentheses is the difference between the solutions to two standard convex optimization problems of maximizing a quasi-concave objective function over a polyhedral feasible set. The analytic complexity of computing η_T stems from the search over all admissible utility functions (\mathcal{U}_2) . However, the utility functions are univariate, normalized, and have a bounded domain (\mathcal{X}) . As a result, we can approximate \mathcal{U}_2 with arbitrary accuracy using a finite set of increasing and concave piecewise-linear functions in the following way.

The term in parentheses is the difference between the solutions to two standard convex optimization problems of maximizing a quasi-concave objective function over a polyhedral feasible set. The analytic complexity of computing η_T stems from the search over all admissible utility functions (\mathcal{U}_2) . However, the utility functions are univariate, normalized, and have a bounded domain (\mathcal{X}) . As a result, we can

approximate \mathcal{U}_2 with arbitrary accuracy using a finite set of increasing and concave piecewise-linear functions in the following way.

We partition \mathcal{X} into N_1 equally spaced values as $\underline{x} = z_1 < \dots < z_{N_1} = \bar{x}$, where $z_n := \underline{x} + \frac{n-1}{N_1-1}(\bar{x} - \underline{x})$, $n = 1, \dots, N_1$; $N_1 \geq 2$. Instead of an equal spacing, the partition could also be based on percentiles of the return distribution. Similarly, we partition the interval $[0, 1]$, as $0 < \frac{1}{N_2-1} < \dots < \frac{N_2-2}{N_2-1} < 1$, $N_2 \geq 2$. Using this partition, let

$$\underline{\eta}_T := \sqrt{T} \sup_{u \in \underline{\mathcal{U}}_2} \left(\sup_{\lambda \in \mathcal{K}} \mathbb{E}_{F_T} [u(X^T \lambda)] - \sup_{\kappa \in \mathcal{K}} \mathbb{E}_{F_T} [u(X^T \kappa)] \right); \quad (44)$$

$$\underline{\mathcal{U}}_2 := \left\{ u \in \mathcal{C}^0 : u(y) = \sum_{n=1}^{N_1} w_n r(y; z_n) \mathbf{w} \in \mathbf{W} \right\}; \quad (45)$$

$$\mathbf{W} := \left\{ \mathbf{w} \in \left\{ 0, \frac{1}{N_2-1}, \dots, \frac{N_2-2}{N_2-1}, 1 \right\}^{N_1} : \sum_{n=1}^{N_1} w_n = 1 \right\}. \quad (46)$$

Every element $u \in \underline{\mathcal{U}}_2$ consists of at most N_2 linear line segments with knots at N_1 possible outcome levels. Clearly, $\underline{\mathcal{U}}_2 \subset \mathcal{U}_2$ and $\underline{\eta}_T$ approximates η_T from below as we refine the partition ($N_1, N_2 \rightarrow \infty$). The appealing feature of $\underline{\eta}_T$ is that we can enumerate all $N_3 := \frac{1}{(N_1-1)!} \prod_{i=1}^{N_1-1} (N_2 + i - 1)$ elements of $\underline{\mathcal{U}}_2$ for a given partition, and, for every $u \in \underline{\mathcal{U}}_2$, solve the two embedded maximization problems in (44) using LP:

Proposition 8. *Let*

$$c_{0,n} := \sum_{m=n}^{N_1} (c_{1,m+1} - c_{1,m}) z_m; \quad (47)$$

$$c_{1,n} := \sum_{m=n}^{N_1} w_m; \quad (48)$$

$$\mathcal{N} := \{n = 1, \dots, N_1 : w_n > 0\} \cup \{N_1\}. \quad (49)$$

For any given $u \in \underline{\mathcal{U}}_2$, $\sup_{\lambda \in \mathcal{K}} \mathbb{E}_{F_T} [u(X^T \lambda)]$ is the optimal value of the objective function of the following LP problem in canonical form:

$$\begin{aligned}
& \max T^{-1} \sum_{t=1}^T y_t & (50) \\
\text{s.t. } & y_t - c_{1,n} X_t^T \lambda \leq c_{0,n}, \quad t = 1, \dots, T; n \in \mathcal{N}; \\
& \sum_{i=1}^M \lambda_i = 1; \\
& \lambda_i \geq 0, \quad i = 1, \dots, M; \\
& y_t \text{ free}, \quad t = 1, \dots, T.
\end{aligned}$$

The LP problem always has a feasible and finite solution and has $\mathcal{O}(T + M)$ variables and constraints, making it small for typical data dimensions. Our application in [Section 5](#) is based on the entire available history of monthly investment returns to a standard set of benchmark assets ($M = 11$, $T = 1,062$), and uses $N_1 = 10$ and $N_2 = 5$. This gives $N_3 = \frac{1}{9!} \prod_{i=1}^9 (4 + i) = 715$ distinct utility functions and $2N_3 = 1,430$ small LP problems, which is perfectly manageable with modern-day computer hardware and solver software.

The total run time of all computations for our application amounts to several working days on a standard desktop PC with a 2.93 GHz quad-core Intel i7 processor, 16GB of RAM and using MATLAB with the external Gurobi Optimizer solver.

Table I: Simulation Experiment

Panel A: $(M - 2) = 2 \times 3$											
		Size			Power			'Size-adj. power'			
		$\alpha \rightarrow$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
$T \downarrow$											
Stochastic Spanning	240	0.110	0.132	0.157	0.915	0.947	0.963	0.805	0.815	0.806	
	480	0.046	0.061	0.093	0.998	1.000	1.000	0.952	0.939	0.907	
	960	0.020	0.042	0.071	1.000	1.000	1.000	0.980	0.958	0.929	
M-V spanning (subsampling)	240	0.019	0.028	0.061	0.981	0.989	0.997	0.962	0.961	0.936	
	480	0.007	0.018	0.057	1.000	1.000	1.000	0.993	0.982	0.943	
	960	0.022	0.045	0.087	1.000	1.000	1.000	0.978	0.955	0.913	
GRS	240	0.007	0.049	0.098	1.000	1.000	1.000	0.993	0.951	0.902	
	480	0.011	0.048	0.103	1.000	1.000	1.000	0.989	0.952	0.897	
	960	0.010	0.063	0.128	1.000	1.000	1.000	0.990	0.937	0.872	

Panel B: $(M - 2) = 5 \times 5$											
		Size			Power			'Size-adj. power'			
		$\alpha \rightarrow$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
$T \downarrow$											
Stochastic Spanning	240	0.092	0.113	0.132	0.435	0.551	0.643	0.343	0.438	0.511	
	480	0.034	0.058	0.075	0.687	0.788	0.855	0.653	0.730	0.780	
	960	0.012	0.017	0.026	0.947	0.963	0.981	0.935	0.946	0.955	
M-V spanning (subsampling)	240	0.044	0.057	0.071	0.651	0.781	0.872	0.607	0.724	0.801	
	480	0.020	0.083	0.155	0.987	0.997	0.999	0.967	0.914	0.844	
	960	0.036	0.057	0.087	1.000	1.000	1.000	0.964	0.943	0.913	
GRS	240	0.005	0.035	0.087	0.979	0.993	0.997	0.974	0.958	0.910	
	480	0.009	0.041	0.097	1.000	1.000	1.000	0.991	0.959	0.903	
	960	0.014	0.043	0.091	1.000	1.000	1.000	0.986	0.957	0.909	

Figure 1: Empirical Results for the hypothesis of Stochastic Efficiency

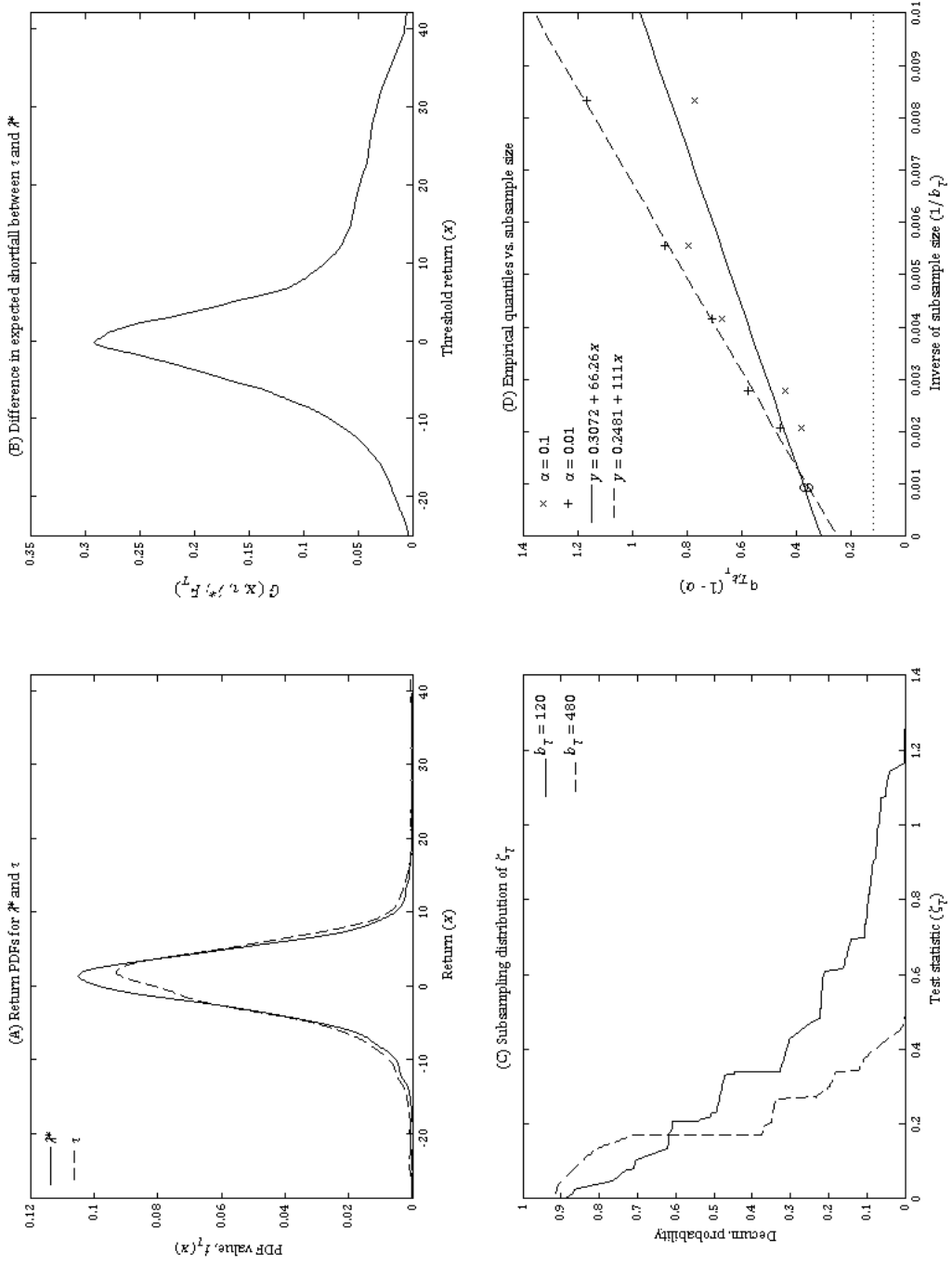


Figure 2: Empirical Results for the hypothesis of Stochastic Spanning

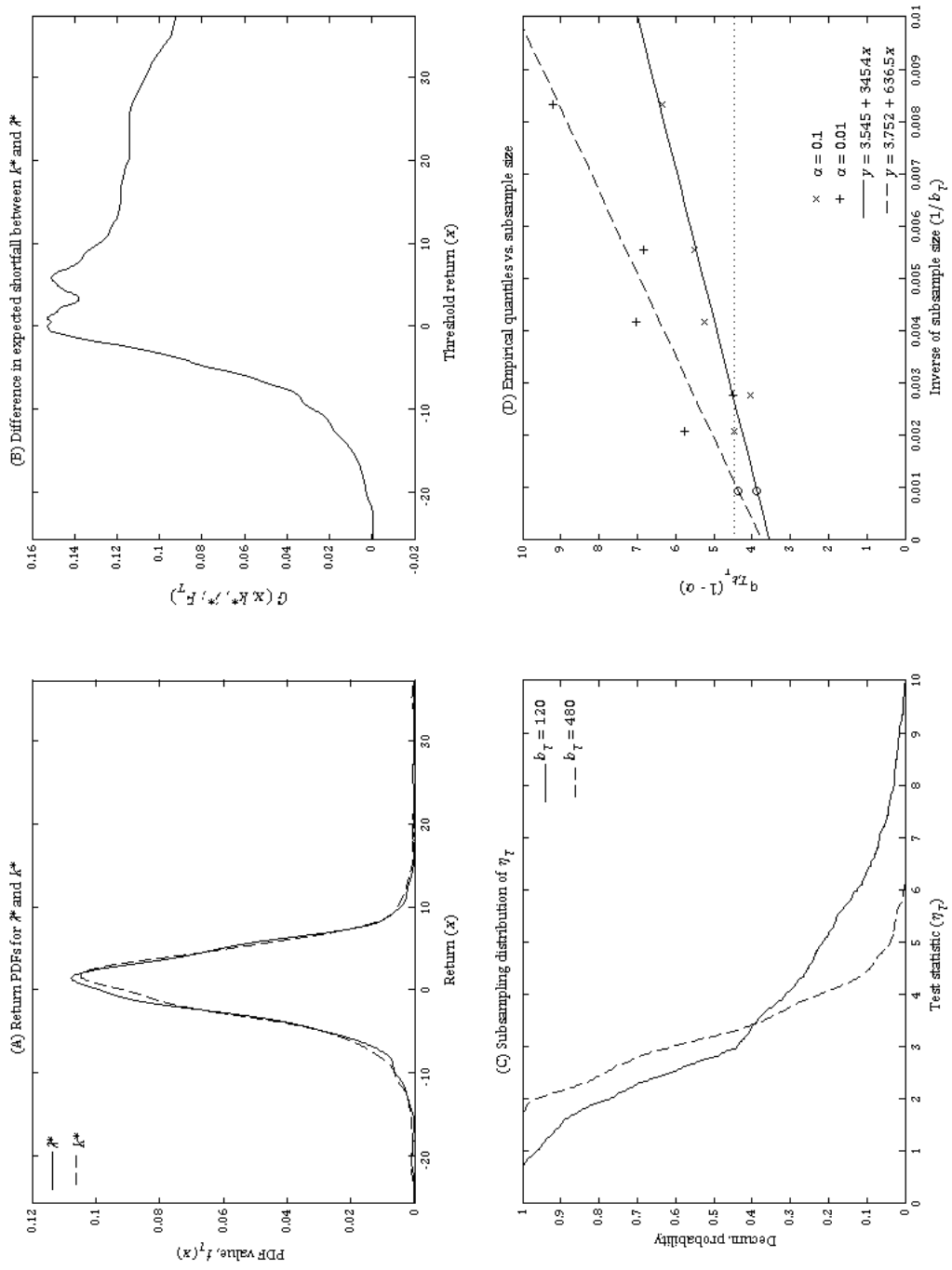


Figure 3: Empirical Results for the hypothesis of Mean-Variance Efficiency

