ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF ECONOMICS

WORKING PAPER SERIES 10-2015

Stochastic Spanning

Stelios Arvanitis Mark Hallam Thierry Post
Nikolas Topaloglou

76 Patission Str., Athens 104 34, Greece
Tel. (++30) 210-8203911 - Fax: (++30) 210-8203301
www.econ.aueb.gr

Stochastic Spanning

Stelios Arvanitis
Mark Hallam
Thierry Post[1]
Nikolas Topaloglou

Abstract: This study develops and implements methods for analyzing whether introducing new securities or relaxing investment constraints improves the investment opportunity set for risk averse investors. We develop a statistical test procedure for 'stochastic spanning' for two nested polyhedral portfolio sets based on subsampling and Linear Programming. The test is statistically consistent and asymptotically exact for a class of weakly dependent processes. Using this test, we accept market portfolio efficiency but reject two-fund separation in standard data sets of historical stock market returns. The divergence between the test results for the two hypotheses illustrates the role for higher-order moment risk in portfolio choice and challenges representative-investor models of capital market equilibrium.

Key words: Portfolio choice, Stochastic Dominance, Spanning, Subsampling, Linear Programming, Asset Pricing.

JEL subject codes: C61, D81, G11.

October 04, 2015

[1] Post is professor of Finance at Koç University's Graduate School of Business; Rumelifeneri Yolu; 34450 Sariyer; Istanbul; Turkey; e-mail: thierrypost@hotmail.com; tel: +90 (212) 338 13 75; fax: +90 (212) 338 16 53.

1. INTRODUCTION

Stochastic Dominance (SD) is a mathematical order on prospects based on general regularity conditions for decision making under risk (Quirk and Saposnik (1962), Hadar and Russell (1969), Hanoch and Levy (1969)). It can be seen as a model-free alternative to mean-variance (M-V) dominance that applies also for non-normal probability distributions and non-quadratic utility functions. 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 pairwise SD.

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) and Linton, Post and Whang (2014) 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).

This study introduces the related concept of 'stochastic spanning' and develops methods for implementing this new concept. Spanning occurs if introducing new securities or relaxing investment constraints does not improve the investment opportunity set for a given class of investors. Stochastic spanning can be seen as a model-free alternative to M-V spanning (Huberman and Kandel (1987)) that accounts for higher-order moment risk in addition to variance.

Accounting for higher-order risk arguably is more relevant for analyzing spanning than for efficiency. Efficiency tests are typically applied to a given broad market index with limited higher-order moment risk (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 assets, and can therefore be more strongly affected by higher-order moment risk.

We propose a theoretical measure for stochastic spanning for two nested polyhedral investment opportunity sets 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).

Spanning involves the comparison of two choice sets. Pairwise dominance analysis and portfolio efficiency analysis are 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), the economic interpretation of which is well established 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.

We apply the stochastic efficiency and spanning tests to standard data sets of historical stock returns from the empirical asset pricing literature, for which we accept market portfolio efficiency but reject two-fund separation. By Tobin's (1958) separation theorem, these two concepts are equivalent under a multivariate normal distribution and therefore the divergence of our two sets of test results suggests an important role for higher-order moment risk in portfolio choice. The rejection of two-fund separation also casts doubt on representative investor models of capital market equilibrium.

Furthermore, the application also illustrates that the proposed resampling scheme and mathematical programs are computationally feasible with modern-day computer hardware and solver software for the typical problem dimensions. 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.

2. Stochastic Efficiency

The investment universe consists of M assets with random investment returns $X = (x_1 \cdots x_M)^{\mathrm{T}} \in \mathcal{X}^M$ with compact support $\mathcal{X} := [\underline{x}, \overline{x}]$, $-\infty < \underline{x} < \overline{x} < \infty$. The investment opportunity set is assumed to be an M-simplex $\Lambda := \{\lambda \in \mathbb{R}_+^M : 1_M^{\mathrm{T}}\lambda = 1\}$. We may deal with a more general polytope $\Lambda_0 \subset \mathbb{R}^M$ by replacing the convex hull of the assets with the convex hull of the vertices of Λ_0. To allow for dynamic intertemporal choice problems, the base assets could be periodically rebalanced portfolios of individual securities.

Let $F: \mathcal{X}^M \to [0,1]$ denote the continuous joint c.d.f. of X and $F(y,\lambda) := \int 1(X^{\mathrm{T}}\lambda \leq y)\partial F(X)$ the 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.:

$$F^{(2)}(x,\lambda) := \int_{-\infty}^{x} F(y,\lambda)\, dy = \int_{-\infty}^{x} (x-y)\, \partial F(y,\lambda)\, . \tag{2.1}$$

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

Definition 2.1 (Weak Stochastic Dominance): *Portfolio $\lambda \in \Lambda$ weakly second-order stochastically dominates portfolio* $\tau \in \Lambda$, or $\lambda \succcurlyeq_F \tau$, *if*

$$G(x,\lambda,\tau;F) \leq 0 \ \forall x \in \mathcal{X}; \tag{2.2}$$
$$G(x,\lambda,\tau;F) := F^{(2)}(x,\lambda) - F^{(2)}(x,\tau); \tag{2.3}$$

Weak stochastic dominance does not occur, or $\lambda \not\succcurlyeq_F \tau$, if $G(x,\lambda,\tau;F) > 0$ for some $x \in \mathcal{X}$.

DEFINITION 2.2 (STRICT STOCHASTIC DOMINANCE): *Portfolio* $\lambda \in \Lambda$ *strictly second-order stochastically dominates portfolio* $\tau \in \Lambda$, or $\lambda \succ_F \tau$, *if*

$$(\lambda \succcurlyeq_F \tau) \wedge (G(x, \lambda, \tau; F) < 0 \text{ for some } x \in \mathcal{X}). \tag{2.4}$$

Strict stochastic dominance does not occur, or $\lambda \nsucc_F \tau$, *if* $(\lambda \not\succcurlyeq_F \tau) \vee (G(x, \lambda, \tau; F) = 0\ \forall x \in \mathcal{X})$.

Stochastic dominance is a preorder rather than a partial order, because two distinct portfolios ($\tau \neq \lambda$) may be equivalent ($X^{\mathrm{T}}\tau = X^{\mathrm{T}}\lambda\ \forall X \in \mathcal{X}^M$), which violates the antisymmetric property; $(\lambda \succcurlyeq_F \tau) \wedge (\tau \succcurlyeq_F \lambda) \nRightarrow \kappa = \lambda$. Furthermore, dominance is not a total order, as a pair of portfolios may be incomparable, that is, $G(x, \lambda, \tau; F) < 0$ for some $x \in \mathcal{X}$ and $G(x, \lambda, \tau; F) > 0$ for some other $x \in \mathcal{X}$. Strict dominance is the irreflexive part of the preorder, as a given portfolio does not strictly dominate itself ($\tau \nsucc_F \tau$).

DEFINITION 2.3 (STOCHASTIC EFFICIENCY): *Portfolio* $\tau \in \Lambda$ *is second-order stochastically efficient if it is not strictly second-order stochastically dominated by any feasible portfolio*: $\lambda \nsucc_F \tau\ \ \forall \lambda \in \Lambda$. *Stochastic inefficiency occurs if* $\lambda \succ_F \tau\ \ \lambda \in \Lambda$.

We denote the set of all stochastically efficient portfolios by $E(\Lambda) \coloneqq \{\tau \in \Lambda : \lambda \nsucc_F \tau\ \ \forall \lambda \in \Lambda\}$. In mathematical order theory, $E(\Lambda)$ amounts to the set of maximal elements.

$E(\Lambda)$ is a model-free generalization of the M-V efficient set, which is based on the assumption of a normal probability distribution or a quadratic utility function. 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 economically irrational 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 $F^{(2)}(x, \lambda), x \in \mathcal{X}$.

The above definition of stochastic efficiency should not be confused with an alternative definition by Scaillet and Topaloglou (2010, henceforth ST2010), which we label here as 'stochastic super-efficiency':

DEFINITION 2.4 (STOCHASTIC SUPER-EFFICIENCY): *Portfolio* $\tau \in \Lambda$ *is second-order stochastically super-efficient if it weakly second-order stochastically dominates all feasible portfolios, or* $\tau \succcurlyeq_F \lambda\ \ \forall \lambda \in \Lambda$. *Stochastic super-efficiency does not occur if* $\tau \not\succcurlyeq_F \lambda\ \ \lambda \in \Lambda$.

We denote all super-efficient portfolios by $S(\Lambda) \coloneqq \{\tau \in \Lambda : \tau \succcurlyeq_F \lambda\ \ \forall \lambda \in \Lambda\}$. In order theory, $S(\Lambda)$ amounts to the set of greatest elements rather than the set of maximal elements. Clearly, stochastic super-efficiency gives a sufficient condition for stochastic efficiency; $(\tau \succcurlyeq_F \lambda\ \ \forall \lambda \in \Lambda) \Rightarrow (\lambda \nsucc_F \tau\ \ \forall \lambda \in \Lambda)$, 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 super-efficient set is either equal to the efficient set $\left(S(\Lambda) = E(\Lambda)\right)$ 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.

3. Stochastic Spanning

Despite its restrictiveness, the notion of stochastic super-efficiency can be generalized to a useful notion of stochastic spanning for comparing two nested choice sets:

Definition 3.1 (Stochastic Spanning): *Portfolio set* Λ *is second-order stochastically spanned by a non-empty polyhedral subset* $\mathrm{K} \subset \Lambda$ *if all portfolios* $\lambda \in \Lambda$ *are weakly second-order stochastically dominated by some portfolios* $\kappa \in \mathrm{K}$:

$$(\kappa \succcurlyeq_F \lambda \quad \kappa \in \mathrm{K})\ \forall \lambda \in \Lambda$$
$$\Leftrightarrow \left((G(x, \kappa, \lambda; F) \leq 0\ \forall x \in \mathcal{X}) \quad \kappa \in \mathrm{K}\right) \forall \lambda \in \Lambda. \tag{3.1}$$

Stochastic spanning does not occur if

$$(\kappa \not\succcurlyeq_F \lambda \quad \forall \kappa \in \mathrm{K}) \quad \lambda \in \Lambda$$
$$\Leftrightarrow \left((G(x, \kappa, \lambda; F) > 0\ x \in \mathcal{X}) \quad \forall \kappa \in \mathrm{K}\right) \quad \lambda \in \Lambda. \tag{3.2}$$

We can view the spanning relation as an SSD order-preserving reduction of the portfolio opportunity set. We will let $R(\Lambda) \coloneqq \{\mathrm{K} \subseteq \Lambda : (\kappa \succcurlyeq_F \lambda \quad \kappa \in \mathrm{K})\ \forall \lambda \in \Lambda\}$ denote all relevant subsets that span Λ. Spanning occurs if and only if $\mathrm{K} \in R(\Lambda)$. $R(\Lambda)$ is non-empty because it includes at least Λ; a span $\mathrm{K} \in R(\Lambda)$ may itself be spanned by another span $\mathrm{K}' \in R(\mathrm{K}) \subseteq R(\Lambda)$. This study analyzes a given subset $\mathrm{K} \subset \Lambda$. In other applications, it may be interesting to find an irreducible span $\mathrm{K} \subseteq R(\Lambda)$, so that $\mathrm{K} = R(\mathrm{K})$. However, there generally exist multiple irreducable spans due to the possibility that two distinct portfolios have equivalent returns.

The following result clarifies the relation between stochastic spanning and stochastic efficiency:

Proposition 3.1: *Stochastic spanning occurs if the enlargement* $(\Lambda - \mathrm{K})$ *does not change the efficient set, that is,*

$$\mathrm{K} \in R(\Lambda) \Longleftarrow E(\Lambda) \subseteq \mathrm{K}. \tag{3.3}$$

The reverse relation generally does not hold, because the weak dominance relation does not possess the antisymmetric property. Specifically, the condition $E(\Lambda) \subseteq \mathrm{K}$ does not allow that two portfolios $(\kappa, \lambda) \in E(\mathrm{K}) \times (\Lambda - \mathrm{K})$ are equivalent $(X^{\mathrm{T}}\kappa = X^{\mathrm{T}}\lambda)$, whereas the condition $\mathrm{K} \in R(\Lambda)$ does allow equivalence. In other words, $E(\Lambda)$ always spans Λ, but it may be reducable by excluding equivalent elements. Consequently, $E(\Lambda) \subseteq \mathrm{K}$ is a sufficient but not necessary condition for $\mathrm{K} \in R(\Lambda)$. In addition, the sufficient condition $E(\Lambda) \subseteq \mathrm{K}$ is not practical, because $E(\Lambda)$ generally is 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 $\mathrm{K} \in R(\Lambda)$ 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 degree measure for deviations from stochastic spanning:

$$\eta(F) \coloneqq \sup_{\lambda \in \Lambda} \inf_{\kappa \in \mathrm{K}} \sup_{x \in \mathcal{X}} G(x, \kappa, \lambda; F). \tag{3.4}$$

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

REMARK 3.1: *Stochastic super-efficiency* ($\tau \succcurlyeq_F \lambda \quad \forall \lambda \in \Lambda$) *occurs as the special case of stochastic spanning with* $\mathrm{K} = \{\tau\}$, $\tau \in \Lambda$. *In this case, our degree measure reduces to*

$$\eta(F) = \sup_{\lambda \in \Lambda} \sup_{x \in \mathcal{X}} G(x, \tau, \lambda; F). \tag{3.5}$$

REMARK 3.2: *Since* $G(\overline{x}, \kappa, \lambda; F) = \mathbb{E}_F[X^{\mathrm{T}}\lambda - X^{\mathrm{T}}\kappa]$, *we find the following lower bound for the stochastic spanning measure:*

$$\begin{aligned} \eta(F) &\geq \sup_{\lambda \in \Lambda} \inf_{\kappa \in \mathrm{K}} G(\overline{x}, \kappa, \lambda; F) \\ &= \sup_{\lambda \in \Lambda} \inf_{\kappa \in \mathrm{K}} \mathbb{E}_F[X^{\mathrm{T}}\lambda - X^{\mathrm{T}}\kappa]. \end{aligned} \tag{3.6}$$

To further clarify the economic meaning of the notion of stochastic spanning, it is useful to formulate it in terms of expected utility:

PROPOSITION 3.2: *The stochastic spanning measure* (3.4) *can be reformulated as follows:*

$$\eta(F) = \sup_{\substack{w \in \mathcal{W};\\ \lambda \in \Lambda}} \inf_{\kappa \in \mathrm{K}} H(w, \kappa, \lambda; F); \tag{3.7}$$

$$H(w, \kappa, \lambda; F) \coloneqq \int_{\underline{x}}^{\overline{x}} w(x) G(x, \kappa, \lambda; F)\, \partial x; \tag{3.8}$$

$$\mathcal{W} \coloneqq \left\{ w: \mathcal{X} \to [0,1]: \int_{\underline{x}}^{\overline{x}} w(x) \partial x = 1 \right\}. \tag{3.9}$$

Alternatively,

$$\eta(F) = \sup_{\substack{u \in \mathcal{U}_2;\\ \lambda \in \Lambda}} \inf_{\kappa \in \mathrm{K}} \mathbb{E}_F[u(X^{\mathrm{T}}\lambda) - u(X^{\mathrm{T}}\kappa)]; \tag{3.10}$$

$$\mathcal{U}_2 := \left\{ u \in \mathcal{C}^0 : u(y) = \int_{\underline{x}}^{\overline{x}} w(x) r(y;x) \partial x \quad w \in \mathcal{W} \right\}; \tag{3.11}$$

$$r(y;x) \coloneqq (y-x)1(y \leq x), \qquad (x,y) \in \mathcal{X}^2. \tag{3.12}$$

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 non-satiable and risk-averse investor $u \in \mathcal{U}_2$ benefits from the enlargement $(\Lambda - \mathrm{K})$. Put differently, the additional restrictions have a shadow price of zero for all relevant investors. The lower bound (3.6) represents the potential benefit of the enlargement to a risk-neutral investor with utility function $u(y) = (y - \overline{x})$. In addition to clarifying the economic meaning of stochastic spanning, Proposition 3.2 will also prove useful for developing a consistent and feasible test procedure below.

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)). Stochastic super-efficiency is the extreme (and generally impossible) case with a single mutual fund ($N = 1$). 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 $\mathrm{K} \subset \Lambda$ can be represented as the convex hull of its $V(\mathrm{K}) \in \mathbb{N}_1$ vertices. Hence, in case of stochastic spanning, rational investors can limit their attention to combining the $V(\mathrm{K})$ vertices of K, and $N \leq V(\mathrm{K})$.

4. Numerical example

Figure 1 illustrates the relevant concepts using a numerical example based on a discrete probability distribution with two equiprobable states ($s = 1,2$). We use both state-space diagrams and mean-standard deviation diagrams, in order to illustrate the difference between the SD and M-V criteria.

Panel A and B are based on a single risky asset ($M = 1$) with gross investment returns $(x_{1;1}, x_{1;2}) = (0.8, 1.2)$. The clear area contains all inefficient return vectors, which are stochastically dominated by $(0.8, 1.2)$, the grey areas contain the efficient vectors, which are not stochastically dominated by $(0.8, 1.2)$ and the dark grey area contains the stochastically super-efficient vectors, which stochastically dominate $(0.8, 1.2)$. It is clear that many of the super-efficient vectors do not M-V dominate (0.8, 1.2). Since the feasible set includes only a single asset, it obviously coincides with the set of efficient portfolios and the set of super-efficient portfolios in this example: $\Lambda = E(\Lambda) = S(\Lambda) = \{1\}$.

Panel C and D include two additional assets ($M = 3$) with gross investment returns $(x_{2;1}, x_{2;2}) = (0.8, 2)$ and $(x_{3;1}, x_{3;2}) = (1, 1.5)$. The set of efficient vectors (grey area) and the set of super-efficient vectors (dark grey area) shrink substantially. The black line represents the edges of the feasible portfolio set $\Lambda = \{\lambda \in \mathbb{R}_+^3 : \lambda_1 + \lambda_2 + \lambda_3 = 1\}$. The set of stochastically efficient portfolios is now given by $E(\Lambda) = \{\lambda \in \Lambda : \lambda_1 = 0\}$ and is a proper subset of the M-V efficient set $E_{MV}(\Lambda) = \{\lambda \in \Lambda : \lambda_1 = 0\} \cup \{\lambda \in \Lambda : \lambda_2 = 0\}$. Clearly, no feasible portfolio dominates all $\lambda \in \Lambda$ in this example and hence the set of super-efficient portfolios is empty; $S(\Lambda) = \emptyset$.

By definition, every superset of the efficient set $E(\Lambda) = \{\lambda \in \Lambda: \lambda_1 = 0\}$ spans the entire portfolio set Λ (see Proposition 3.1); for example, $\mathrm{K}_1 = \{\lambda \in \Lambda: \lambda_1 \leq 0.5\}$ spans Λ, or $\mathrm{K}_1 \in R(\Lambda)$. Furthermore, in this simple example without equivalent portfolios, no proper subset of $E(\Lambda)$ spans Λ. For example, $\mathrm{K}_2 = \{\lambda \in \Lambda: \lambda_1 = 0, \lambda_3 \leq 0.5\}$ does not span Λ; $\mathrm{K}_2 \notin R(\Lambda)$. If we set $\mathrm{K} = \mathrm{K}_2$, then it is easy to verify that the optimal solution for the spanning measure (3.4) is given by $\lambda^* = (0\ 0\ 1)^{\mathrm{T}}$, $\kappa^* = (0\ 0.5\ 0.5)^{\mathrm{T}}$ and $x^* = 1.00$. We therefore find expected shortfall levels of $F^{(2)}(x^*, \kappa^*) = 0.00$ and $F^{(2)}(x^*, \lambda^*) = 0.10$, and the spanning measure amounts to $\eta(F) = G(x^*, \kappa^*, \lambda^*; F) = (0.10 - 0.00) = 0.10$. Clearly, spanning does not occur. In this case, the optimal utility function in (3.10) is simply $u^*(y) = r(y; 1) = (y-1)1(y \leq 1)$.

[Insert Figure 1 about here.]

5. Spanning Test Statistic

In empirical applications, the c.d.f. F is latent and the analyst has access to a discrete time series of realized returns $s_T \coloneqq (X_t)_{t=1}^T$, $X_t \in \mathcal{X}$, $t = 1, \cdots, T$. We make the following general assumptions on the multivariate return process:

Assumption 5.1: *(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*

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

is positive definite.

Let $F_T(x) \coloneqq T^{-1}\sum_{t=1}^T 1(X_t \leq x)$ 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 $\mathcal{B}_F = \mathcal{B} \circ F$ with covariance kernel given by $Cov(\mathcal{B}_F(x), \mathcal{B}_F(y)) \coloneqq \sum_{t \in \mathbb{Z}} Cov(1(X_0 \leq x), 1(X_t \leq y))$ and almost surely uniformly continuous sample paths defined on $\mathbb{R}^M$ (see Thm 7.3 of Rio (2000)).

We consider the following scaled empirical analogue of (3.7) as a test statistic for stochastic spanning:

$$\eta_T \coloneqq \sqrt{T}\eta(F_T) = \sqrt{T} \sup_{\lambda \in \Lambda} \inf_{\kappa \in \mathrm{K}} \sup_{x \in \mathcal{X}} G(x, \kappa, \lambda; F_T) \tag{5.1}$$

$$= \sqrt{T} \sup_{\substack{w \in \mathcal{W};\\ \lambda \in \Lambda}} \inf_{\kappa \in \mathrm{K}} H(w, \kappa, \lambda; F_T). \tag{5.1'}$$

Remark 5.1 *If the portfolio set K is a singleton, or* $\mathrm{K} = \{\tau\}$, $\tau \in \Lambda$, *then we obtain the super-efficiency test of ST2010;* $\eta_T = \sqrt{T} \sup_{\lambda \in \Lambda} \sup_{x \in \mathcal{X}} G(x, \tau, \lambda; F_T)$. *Our results below thus also apply to the ST2010 test.*

REMARK 5.2 *If the portfolio set* Λ *is a singleton, or* $\Lambda = \{\lambda\}$, *then we obtain a test for stochastic efficiency that resembles the test of Linton et al. (henceforth LPW2014)*; $\eta_T = \sqrt{T} \inf_{\kappa \in \mathrm{K}} \sup_{x \in \mathcal{X}} G(x, \kappa, \lambda; F_T)$. *This is however not a proper spanning test statistic, as* $\mathrm{K} \not\subset \Lambda$.

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:

$$\Gamma^{=} := \left\{(w, \lambda) \in \Gamma: \inf_{\kappa \in \mathrm{K}} H(w, \kappa, \lambda; F) = 0\right\}; \tag{5.2}$$
$$\Gamma^{<} := \left\{(w, \lambda) \in \Gamma: \inf_{\kappa \in \mathrm{K}} H(w, \kappa, \lambda; F) < 0\right\}. \tag{5.3}$$

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

$$\mathrm{K}^{\leq}_{(w,\lambda)} := \{\kappa \in \mathrm{K}: H(w, \kappa, \lambda; F) \leq 0\} \quad (w, \lambda) \in \Gamma; \tag{5.4}$$
$$\mathrm{K}^{>}_{(w,\lambda)} := \{\kappa \in \mathrm{K}: H(w, \kappa, \lambda; F) > 0\} \quad (w, \lambda) \in \Gamma. \tag{5.5}$$

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

PROPOSITION 5.1: *Under Assumption 5.1,*

(i) $\quad H\big(w, \kappa, \lambda; \sqrt{T}(F_T - F)\big) \rightsquigarrow H(w, \kappa, \lambda; \mathcal{B} \circ F)$*;*

(ii) $\quad \operatorname*{oper}_{(w,\lambda) \in A_T} \operatorname*{oper}^*_{\kappa \in B_T} H\big(w, \kappa, \lambda; \sqrt{T}(F_T - F)\big) \rightsquigarrow \operatorname*{oper}_{(w,\lambda) \in A} \operatorname*{oper}^*_{\kappa \in B} H(w, \kappa, \lambda; \mathcal{B} \circ F)$,

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

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

PROPOSITION 5.2: *If Assumption 5.1 holds and* $\mathbf{H}_0$ *is true, then*

$$\eta_T \rightsquigarrow \eta_\infty := \sup_{(w,\lambda) \in \Gamma^{=}} \inf_{\kappa \in \mathrm{K}^{\leq}_{(w,\lambda)}} H(w, \kappa, \lambda; \mathcal{B}_F). \tag{5.6}$$

COROLLARY 5.1: *For the case of super-efficiency, or* $\mathrm{K} = \{\tau\}$, $\tau \in \Lambda$, *we obtain the exact limit distribution of the ST2010 test statistic as the law of* $\sup_{(w,\lambda) \in \Gamma^{=}} H(w, \tau, \lambda; \mathcal{B}_F)$.

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 in Proposition 5.2, we can develop a test procedure for stochastic spanning 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 $\alpha \in]0,1[$. Our decision rule is to reject $\mathbf{H}_0$ against $\mathbf{H}_1$ if and only if $\eta_T > q(\eta_\infty, 1-\alpha)$. Clearly this rule 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 re-sampling procedures to estimate $q(\eta_\infty, 1-\alpha)$ from the data. The next section develops a consistent subsampling procedure for this task.

6. Subsampling Procedure

This section develops a subsampling procedure to estimate the distribution of η_∞ similar to that proposed by LPW2014. 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 6.1: *Under Assumption 5.1, (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, $s_{b_T;T,t} := (X_s)_{s=t}^{t+b_T-1}$, $t = 1, \cdots, T - b_T + 1$, 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 $s_{b_T;T,t}$, $t = 1, \cdots, T - b_T + 1$. The distribution of subsample test scores can be described by the following c.d.f. and quantile function:

$$S_{T,b_T}(y) := \frac{1}{T - b_T + 1} \sum_{t=1}^{T-b_T+1} 1(\eta_{b_T;T,t} \leq y); \tag{6.1}$$

$$q_{T,b_T}(1-\alpha) := \inf_y \{y : S_{T,b_T}(y) \geq 1 - \alpha\}. \tag{6.2}$$

The 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, $1 - S_{T,b_T}(\eta_T) < \alpha$.

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

Assumption 6.1: *The positive sequence (b_T), possibly dependent on $(X_t)_{t=1}^T$, obeys*

$$\mathbb{P}(l_T \leq b_T \leq u_T) \to 1, \tag{6.3}$$

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

ASSUMPTION 6.2: *Let* $M^* \in \mathbb{N}_0$, $M^* < V(\Lambda)$, *denote the number of vertices of* Λ *that are also included in* K. *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 6.2.

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

PROPOSITION 6.2: *If Assumption 5.1, Assumption 6.1 and Assumption 6.2 hold, then we find the following asymptotic size and power properties:*

$$\lim_{T\to\infty} \mathbb{P}\big(\eta_T > q_{T,b_T}(1-\alpha)|\mathbf{H}_0\big) = \alpha; \tag{6.4}$$
$$\lim_{T\to\infty} \mathbb{P}\big(\eta_T > q_{T,b_T}(1-\alpha)|\mathbf{H}_1\big) = 1. \tag{6.5}$$

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 senstivity 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:

$$q_{T,b_T}(1-\alpha) = \gamma_{0;T,1-\alpha} + \gamma_{1;T,1-\alpha}\,(b_T)^{-1} + \upsilon_{T;1-\alpha,b_T}. \tag{6.6}$$

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

$$q_T^{BC}(1-\alpha) := \hat{\gamma}_{0;T,1-\alpha} + \hat{\gamma}_{1;T,1-\alpha}\,(T)^{-1}. \tag{6.7}$$

Since $q_{T,b_T}(1-\alpha)$ converges in probability to $q(\eta_\infty, 1-\alpha)$ and $(b_T)^{-1}$ converges to zero as $T \to 0$, $\hat{\gamma}_{0;T,1-\alpha}$ 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 5.2 is based on the properties of the partitions of Γ and K in (5.2)-(5.5) and the behavior of the degree measure $\eta(F)$ on these subsets. Our analysis does not directly apply to the bootstrap, because the bootstrap re-centering is not with respect to the degree measure $\eta(F)$, but instead its empirical counterpart $\eta(F_T)$, which does not have the same behavior on the relevant subsets. We expect that it is possible based on our Proposition 3.2 to construct a bootstrap critical value that allows for a consistent test procedure (in the spirit of Prop. 3.1 of ST2010), but that the procedure would be asymptotically conservative and less powerful than the subsampling approach under local alternative hypotheses. On the other hand, we also envisage cases where the bootstrap is more efficient in finite samples than subsampling, since each pseudo-sample utilizes the full sample information, rather than a subset of the observations. We leave the further development of a bootstrap test procedure for stochastic spanning for further research.

7. Computational Strategy

In general computing the test statistic ζ_T is a challenging global optimization problem, but depending on the application, there are various alternative computational strategies available. Below, we outline two possible strategies, one for a small enlargement $(\Lambda - \mathrm{K})$ and another for a limited return interval $(\overline{x} - \underline{x})$.

If the enlargement $(\Lambda - \mathrm{K})$ is small, we may perform a quasi-Monte Carlo simulation and solve an embedded LP problem for every simulated portfolio $\lambda \in (\Lambda - \mathrm{K})$. Specifically, we can use the following reformulation of (5.1):

$$\eta_T = -\sqrt{T} \inf_{\lambda \in (\Lambda - \mathrm{K})} \eta_T(\lambda); \tag{7.1}$$

$$\eta_T(\lambda) \coloneqq \sup_{\kappa \in \mathrm{K}} \inf_{x \in \mathcal{X}} G(x, \lambda, \kappa; F_T). \tag{7.2}$$

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

Proposition 7.1: *The embedded test statistic $\eta_T(\lambda)$ equals the optimal value of the objective function of the following LP problem in canonical form*:

$$\begin{aligned} & \max \sqrt{T}\gamma && (7.3) \\ s.t.\ & \gamma + T^{-1}\sum_{t=1}^{T} \theta_{s,t} \le F_T^{(2)}(X_s\lambda, \lambda), && s = 1, \cdots, T; \\ & -\theta_{s,t} - X_t^{\mathrm{T}}\kappa \le -X_s\lambda, && s,t = 1, \cdots, T; \\ & \sum_{i=1}^{M} \kappa_i = 1; && \\ & \theta_{s,t} \ge 0, && s,t = 1, \cdots, T; \\ & \kappa_i \ge 0, && i = 1, \cdots, M; \\ & \gamma \text{ free}. && \end{aligned}$$

The linear program is reminiscent of existing programs for testing whether a given portfolio $\lambda \in \Lambda$ is SSD efficient relative to portfolio set Λ; see Post (2003), Kuosmanen (2004) and Roman, Darby-Dowman and Mitra (2006), among others. However, we analyze whether a portfolio $\lambda \in \Lambda \backslash \mathrm{K}$ improves the investment possibilities relative to a portfolio set K that does not include that portfolio. In addition, we use a different objective function, a Kolmogorov-Smirnov type test statistic, and we derive and formulate our program in terms of the empirical shortfall measures that form the building blocks of that objective function.

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 $\lambda \in (\Lambda - \mathrm{K})$ and the computational burden will therefore explode if the enlargement $(\Lambda - \mathrm{K})$ is large. For example, in our application in Section 8, 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 approporiate when the enlargement $(\Lambda - \mathrm{K})$ is large but the return range $(\overline{x} - \underline{x})$ is limited. Using (3.10) and (5.1'), we find

$$\eta_T = \sqrt{T} \sup_{u \in \mathcal{U}_2} \left(\sup_{\lambda \in \Lambda} \mathbb{E}_{F_T}[u(X^{\mathrm{T}}\lambda)] - \sup_{\kappa \in \mathrm{K}} \mathbb{E}_{F_T}[u(X^{\mathrm{T}}\kappa)] \right). \tag{7.4}$$

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 analytical 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 $\underline{x} = z_1 < \cdots < z_{N_1} = \overline{x}$, where $z_n \coloneqq \underline{x} + \frac{n-1}{N_1-1}(\overline{x} - \underline{x})$, $n = 1, \cdots, N_1$; $N_1 \geq 2$. 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 $0 < \frac{1}{N_2-1} < \cdots < \frac{N_2-2}{N_2-1} < 1$, $N_2 \geq 2$. Using this partition, let

$$\underline{\eta}_T \coloneqq \sqrt{T} \sup_{u \in \underline{\mathcal{U}}_2} \left(\sup_{\lambda \in \Lambda} \mathbb{E}_{F_T}[u(X^{\mathrm{T}}\lambda)] - \sup_{\kappa \in \mathrm{K}} \mathbb{E}_{F_T}[u(X^{\mathrm{T}}\kappa)] \right); \tag{7.5}$$

$$\underline{\mathcal{U}}_2 \coloneqq \left\{ u \in \mathcal{C}^0 : u(y) = \sum_{n=1}^{N_1} w_n r(y; z_n) \quad \boldsymbol{w} \in \mathrm{W} \right\}; \tag{7.6}$$

$$\mathrm{W} := \left\{ \boldsymbol{w} \in \left\{0, \frac{1}{N_2-1}, \cdots, \frac{N_2-2}{N_2-1}, 1\right\}^{N_1} : \sum_{n=1}^{N_1} w_n = 1 \right\}. \tag{7.7}$$

Every element of $u \in \underline{\mathcal{U}}_2$ 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 (7.5) using LP:

PROPOSITION 7.2: *Let*

$$c_{0,n} := \sum_{m=n}^{N_1} (c_{1,m+1} - c_{1,m}) z_m \, ; \tag{7.8}$$

$$c_{1,n} := \sum_{m=n}^{N_1} w_m \, ; \tag{7.9}$$

$$\mathcal{N} := \{n = 1, \cdots, N_1 : w_n > 0\} \cup \{N_1\}. \tag{7.10}$$

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

$$\max T^{-1} \sum_{t=1}^{T} y_t \tag{7.11}$$

$$y_t - c_{1,n} X_t^{\mathrm{T}} \lambda \leq c_{0,n}, \qquad t = 1, \cdots, T; n \in \mathcal{N};$$

$$\sum_{i=1}^{M} \lambda_i = 1;$$

$$\lambda_i \geq 0, \qquad i = 1, \cdots, M;$$

$$y_t \text{ free}, \qquad t = 1, \cdots, T.$$

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 8 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.

An alternative computational approach could build on the sufficient condition $E(\Lambda) \subseteq \mathrm{K}$ of Proposition 3.1 and the linear Karush-Kuhn Tucker (KKT) portfolio optimality conditions of Post (2003). Specifically, we could enumerate all feasible rankings of investment returns to efficient portfolios in K and, for every feasible ranking, solve an LP problem that searches for an anti-monotonic utility gradient vector that obeys the KKT conditions for K but violates the KKT conditions for Λ. This approach would however require an adjustment of the spanning measure (based on violations of the KKT condition rather than improvements in expected utility) and the statistical theory for the spanning test statistic. We leave this route for future research.

8. EMPIRICAL APPLICATION

This section applies efficiency and spanning tests to a standard data set of historical investment returns from the online data library of Kenneth French.

The relevant investment universe consists of $M = 11$ distinct base assets: the one-month T-bill and ten stock portfolios that are formed by classifying stocks based on the four-digit SIC industry code. Our market portfolio is the CRSP all-share index. We use monthly value-weighted total returns from July 1926 to December 2014 ($T = 1{,}062$).

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. The dimensions of the data set $(N = 11, T = 1{,}062)$ 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 market capitalization of equity (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 our hypotheses and make the test results uninformative. For the same reason, we do not consider data sets of portfolios that are formed on price reversal and momentum patterns.

8.1 Market portfolio efficiency

We first analyze whether the market portfolio is stochastically efficient. 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 aggregation across sufficiently homogenous investors in incomplete markets. Alternatively, a market portfolio efficiency test can 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 11 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.

To test market portfolio efficiency, we use the LPW2014 test and for the sake of comparability, we embed the LPW2014 test in the same subsampling procedure as our spanning test. The four panels of Figure 2 illustrate our results.

The optimal solution $\lambda^* \in \Lambda$ consists of large positions in the non-durables 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 2, 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 $x \in \overline{\mathcal{X}}_T = [-25.15, 42.07]$, from which it is clear that the market portfolio has a strictly higher expected shortfall than the solution portfolio for every return level $x \in \overline{\mathcal{X}}_T$; it follows that $\lambda^* \succ_{F_T} \tau$. The value of the LPW2014 test statistic is $\zeta_T = \sqrt{T} \min_{x \in \overline{\mathcal{X}}_T} G(x, \tau, \lambda^*; F_T) = 0.114$.

Panel C shows the de-cumulative 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 reiterates the need to correct the quantile estimates for bias. Panel D shows the estimated OLS regression line (6.6) 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 (6.7), 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 2 about here.]

8.2 Two-fund separation

Our second research hypothesis is two-fund separation: do all rational risk averters combine the T-bill and the market portfolio? For a multivariate normal distribution, two-fund separation is equivalent to market portfolio efficiency, as a result of Tobin's (1958) separation theorem. Without normality, one generally needs to assume that preferences are sufficiently similar across investors in order to justify two-fund separation (see, for example, Cass and Stiglitz (1970)). Our stochastic spanning test can analyze two-fund separation without assuming a particular shape for the return distribution or utility function.

We include a synthetic index futures contract as the 12th base asset to allow risk-tolerant investors to take leveraged equity positions. The futures contract is built using a short position of 100% in the T-bill and a long position of 200% in the market portfolio. In this application, Λ consists of all convex combinations of the T-bill, ten stock portfolios and the index futures contract and K consists of all convex combinations of the T-bill and the index futures contract. For the computational strategy outlined in Section 7, our partition is based on $\underline{x} = \min_{i,t}(x_{i;t})$, $\overline{x} = \max_{i,t}(x_{i;t})$, $N_1 = 10$ and $N_2 = 5$.

Figure 3 illustrates the estimation results for the industry data set. The optimal solution $\kappa^* \in \mathrm{K}$ 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 non-durables 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 $\kappa^* \nsucceq_{F_T} \lambda^*$; 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 de-cumulative 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 (6.6) for significance levels of $\alpha = 0.01$ and $\alpha = 0.10$ using various subsample sizes $b_T \in [120,480]$. Using (6.7), 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 3 about here.]

8.3 Mean-Variance Analysis

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. Figure 4 summarizes the test results. In contrast to the stochastic spanning, we cannot reject M-V spanning at conventional significance levels.

[Insert Figure 4 about here.]

8.4 Conclusion

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 results must reflect 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.

Appendix

Proof of Proposition 3.1: Our proof consists of the following arguments: $E(\Lambda) \subseteq \mathrm{K} \Leftrightarrow \big((\kappa \succ_F \lambda \quad \kappa \in \mathrm{K}) \quad \forall \lambda \in (\Lambda - \mathrm{K})\big) \Rightarrow (\kappa \succcurlyeq_F \lambda \quad \kappa \in \mathrm{K})\ \forall \lambda \in \Lambda \Leftrightarrow \mathrm{K} \in R(\Lambda)$. ■

Proof of Proposition 3.2: We use of the following chain of arguments:

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

$$= \sup_{\lambda \in \Lambda} \inf_{\kappa \in \mathrm{K}} \sup_{w \in \mathcal{W}} \int_{\underline{x}}^{\overline{x}} w(x) G(x, \kappa, \lambda; F)\, \partial x \tag{3.A}$$

$$= \sup_{\substack{w \in \mathcal{W}; \\ \lambda \in \Lambda}} \inf_{\kappa \in \mathrm{K}} \int_{\underline{x}}^{\overline{x}} w(x) G(x, \kappa, \lambda; F)\, \partial x \tag{3.B}$$

$$= \sup_{\substack{w \in \mathcal{W};\\ \lambda \in \Lambda}} \inf_{\kappa \in \mathrm{K}} \int_{\underline{x}}^{\overline{x}} w(x) \left(F^{(2)}(x, \kappa) - F^{(2)}(x, \lambda) \right) \partial x$$

$$= \sup_{\substack{w \in \mathcal{W};\\ \lambda \in \Lambda}} \inf_{\kappa \in \mathrm{K}} \int_{\underline{x}}^{\overline{x}} w(x) \left(\int_{\underline{x}}^{x} (x - y)\, \partial F(y, \kappa) - \int_{\underline{x}}^{\overline{x}} (x - y)\, \partial F(y, \lambda) \right) \partial x$$

$$= \sup_{\substack{w \in \mathcal{W};\\ \lambda \in \Lambda}} \inf_{\kappa \in \mathrm{K}} \int_{\underline{x}}^{\overline{x}} w(x) \left(\int_{\underline{x}}^{x} (y - x)\, \partial F(y, \lambda) - \int_{\underline{x}}^{x} (y - x)\, \partial F(y, \kappa) \right) \partial x$$

$$= \sup_{\substack{w \in \mathcal{W};\\ \lambda \in \Lambda}} \inf_{\kappa \in \mathrm{K}} \left(\int_{\underline{x}}^{\overline{x}} \int_{\underline{x}}^{\overline{x}} w(x)\, r(y; x) \partial x\, \partial F(y, \lambda) - \int_{\underline{x}}^{\overline{x}} \int_{\underline{x}}^{\overline{x}} w(x)\, r(y; x) \partial x\, \partial F(y, \kappa) \right)$$

$$= \sup_{\substack{u \in \mathcal{U}_2;\\ \lambda \in \Lambda}} \inf_{\kappa \in \mathrm{K}} \left(\int_{\underline{x}}^{\overline{x}} u(y)\, \partial F(y, \lambda) - \int_{\underline{x}}^{\overline{x}} u(y)\, \partial F(y, \kappa) \right)$$

$$= \sup_{\substack{u \in \mathcal{U}_2;\\ \lambda \in \Lambda}} \inf_{\kappa \in \mathrm{K}} (\mathbb{E}_F[u(X^{\mathrm{T}}\lambda) - u(X^{\mathrm{T}}\kappa)]).$$

Equality (3.A) makes the objective function upper semicontinuous and quasiconcave for every $\lambda \in \Lambda$ and $\kappa \in \mathrm{K}$, allowing us to invoke Sion's (1958) Minimax Theorem to change the order of the optimization operators in Equality (3.B). ■

Proof to Proposition 5.1: We endow $\mathcal{W} \times \mathrm{K} \times \Lambda$ with the metric $\delta(v, v^*) \coloneqq \rho \sup_{x \in [\overline{x}, \underline{x}]} |w(x) - w^*(x)| + |\kappa - \kappa^*| + |\lambda - \lambda^*|$, with $0 < \rho \coloneqq \sup_{\kappa \in \mathrm{K}, \lambda \in \Lambda} |\kappa - \lambda| \leq \mathrm{diam}(\Lambda)$, $v \coloneqq (w, \kappa, \lambda)$ and $v^* \coloneqq (w^*, \kappa^*, \lambda^*)$. For any (w, κ, λ),

$$\mathbb{E}_F \left| H\left(w, \kappa, \lambda; \sqrt{T}(F_T - F)\right) \right|^2$$
$$\leq \mathbb{E}_F \left(\int_{\underline{x}}^{\overline{x}} \int_{\mathbb{R}^M} |(x - \kappa^{\mathrm{T}} X)^+ - (x - \kappa\lambda^{\mathrm{T}} X)^+| d\sqrt{T}(F_T - F) dx \right)^2$$
$$\leq (\overline{x} - \underline{x})^2 (\kappa - \lambda)^2 \mathbb{E}_F \left(\int_{\mathbb{R}^M} \|X\| d\sqrt{T}(F_T - F) \right)^2. \tag{5.A}$$

The latter r.h.s. is bounded w.r.t. T due to Assumption 5.1. This result, along with Prop. 3.2 (a) of Jakubowski, Memin and Pages (1989) and Lemma 2.1 of ST2010, implies the fidi convergence of $H\left(w, \kappa, \lambda; \sqrt{T}(F_T - F)\right)$ to $H(w, \kappa, \lambda; \mathcal{B} \circ F)$. Furthermore,

$$\left| H\left(w, \kappa, \lambda; \sqrt{T}(F_T - F)\right) - H\left(w^*, \kappa^*, \lambda^*; \sqrt{T}(F_T - F)\right) \right|$$
$$\leq (\overline{x} - \underline{x}) \delta(v, v^*) \int_{\mathbb{R}^M} \|X\| d\sqrt{T}(F_T - F), \tag{5.B}$$

where the r.h.s. follows from the Lipschitz continuity of $(\cdot)^+$. Notice that $\int_{\mathbb{R}^M} \|X\| d\sqrt{T}(F_T - F) = T^{-1/2} \sum_{t=1}^{T} (\|X_t\| - \mathbb{E}_F \|X_1\|)$, which converges in distribution to a normal random variable due to Assumption 5.1 and the CLT for α-mixing processes (see Corllary 4.1 of Rio (2000)). Hence, the above integral is uniformly (w.r.t. T) tight due to Prokhorov's Theorem (see Thm 18.12 in Van Der Vaart (1997)). This result, along with the total boundedness of $\mathcal{W} \times \mathrm{K} \times \Lambda$ w.r.t. δ, implies the second condition in Thm 18.14 of Van Der Vaart (1997), which establishes part (i) of our proposition. Part (ii) follows from the Continuous Mapping Theorem. ■

PROOF OF PROPOSITION 5.2: Our proof uses a sequence of weak approximations of η_T under the null hypothesis. For $\epsilon_T \to 0, \sqrt{T}\epsilon_T \to \infty$ as $T \to \infty$, and $0 < \delta < 1$, let

$$\eta_T^{=} := \sup_{(w,\lambda)\in\Gamma^=} \inf_{\kappa\in\mathrm{K}} H\big(w,\kappa,\lambda;\sqrt{T}F_T\big). \tag{5.C}$$

$$\eta_T^{\epsilon} := \sup_{(w,\lambda)\in\Gamma^= \cup \Gamma_{\epsilon_T}^<} \inf_{\kappa\in\mathrm{K}} \sqrt{T}H(w,\kappa,\lambda;F_T)\,; \tag{5.D}$$

$$\Gamma_{\epsilon_T}^< := \Big\{(w,\lambda) \in \Gamma\colon \inf_{\kappa\in\mathrm{K}} H(w,\kappa,\lambda;F) < -\epsilon_T\Big\}. \tag{5.E}$$

$$\mathrm{K}_{(w,\lambda)}^{\delta\epsilon_T} := \{\kappa \in \mathrm{K}\colon H(w,\kappa,\lambda;F) \le -\delta\epsilon_T\} \quad (w,\lambda) \in \Gamma. \tag{5.F}$$

Our strategy weakly approximates η_T by η_T^{ϵ}, weakly approximates η_T^{ϵ} by $\eta_T^{=}$ and uses η_∞ as the weak limit of $\eta_T^{=}$.

For any $(w,\lambda) \in \Gamma^=$ for which $\mathrm{K}_{(w,\lambda)}^{>} \neq \emptyset$,

$$\inf_{\kappa\in\mathrm{K}} H\big(w,\kappa,\lambda;\sqrt{T}F_T\big) = \min\left\{\inf_{\kappa\in\mathrm{K}_{(w,\lambda)}^{\le}} H\Big(w,\kappa,\lambda;\sqrt{T}(F_T - F)\Big), R_T(w,\lambda)\right\}, \tag{5.G}$$

where $R_T(w,\lambda) := \inf_{\kappa\in\mathrm{K}_{(w,\lambda)}^{>}} H\big(w,\kappa,\lambda;\sqrt{T}F_T\big)$. If the infimum is achieved on the boundary of $\mathrm{K}_{(w,\lambda)}^{\le}$, then $R_T(w,\lambda) = H\Big(w,\kappa,\lambda;\sqrt{T}(F_T - F)\Big)$. In any case, using a Skorokhod representation argument (see Thm 1.10.4 of van der Vaart and Wellner (1996)), the sequence $\big(R_T(w,\lambda)\big)$ can be partitioned to subsequences which (if any) diverge to $+\infty$, and to subsequences which (if any) converge to the limit of $H\Big(w,\kappa,\lambda;\sqrt{T}(F_T - F)\Big)$ evaluated on the boundary of $\mathrm{K}_{(w,\lambda)}^{\le}$. Consequently, the above minimum weakly converges to $\inf_{\kappa\in\mathrm{K}_{(w,\lambda)}^{\le}} H(w,\kappa,\lambda;\mathcal{B}_F)$. The CMT and Proposition 5.1 then imply

$$\eta_T^{=} \rightsquigarrow \sup_{(w,\lambda)\in\Gamma^=} \inf_{\kappa\in\mathrm{K}_{(w,\lambda)}^{\le}} H(w,\kappa,\lambda;\mathcal{B}_F) =: \eta_\infty. \tag{5.H}$$

Furthermore, we can derive the following results:

$$\mathbb{P}(\eta_T - \eta_T^{\epsilon} > \varepsilon) = \mathbb{P}\left(\max\left\{\eta_T^{\epsilon}, \sup_{\Gamma^< - \Gamma_{\epsilon_T}^<} \inf_{\mathrm{K}} \sqrt{T}H(w,\kappa,\lambda;F_T)\right\} - \eta_T^{\epsilon} > \varepsilon\right)$$

$$\le \mathbb{P}\left(\max\left\{\sup_{\Gamma^= \cup \Gamma_{\epsilon_T}^<} \inf_{\mathrm{K}} \sqrt{T}H(w,\kappa,\lambda;F_T), \sup_{\Gamma^< - \Gamma_{\epsilon_T}^<} \inf_{\mathrm{K}} \sqrt{T}H(w,\kappa,\lambda;F_T)\right\} - \eta_T^{\epsilon} > \varepsilon\right)$$

$$\le \mathbb{P}\left(\sup_{\Gamma^= \cup \Gamma_{\epsilon_T}^<} \inf_{\mathrm{K}} \sqrt{T}H(w,\kappa,\lambda;F_T) < \sup_{\Gamma^< - \Gamma_{\epsilon_T}^<} \inf_{\mathrm{K}} \sqrt{T}H(w,\kappa,\lambda;F_T)\right)$$

$$\leq \mathbb{P}\left(\sup_{\Gamma^{=}} \inf_{\mathrm{K}} \sqrt{T} H(w,\kappa,\lambda;F_T) < \sup_{\Gamma^{<}} \sup_{\mathrm{K}_{(w,\lambda)}^{\delta\epsilon_T}} \sqrt{T} H(w,\kappa,\lambda;F_T)\right). \tag{5.I}$$

The probability (5.I) can be shown to converge to zero. From (5.H), the l.h.s. of the inequality inside the probability (5.I) weakly converges to η_∞. For the r.h.s., we obtain

$$\sup_{\Gamma^{<}} \sup_{\mathrm{K}_{(w,\lambda)}^{\delta\epsilon_T}} \sqrt{T} H(w,\kappa,\lambda;F_T) \leq$$

$$\sup_{\Gamma^{<}} \sup_{\mathrm{K}_{(w,\lambda)}^{\delta\epsilon_T}} H\big(w,\kappa,\lambda;\sqrt{T}(F_T - F)\big) + \sup_{\Gamma^{<}} \sup_{\mathrm{K}_{(w,\lambda)}^{\delta\epsilon_T}} \sqrt{T} H(w,\kappa,\lambda;T). \tag{5.J}$$

Due to Proposition 5.1, the first term on the r.h.s. of the last display weakly converges to $\sup_{\Gamma^{<}} \sup_{\mathrm{K}_{(w,\lambda)}^{\leq}} H(w,\kappa,\lambda;\mathcal{B}_F)$ and the second terms diverges to $-\infty$ due to the construction of $\mathrm{K}_{(w,\lambda)}^{\delta\epsilon_T}$. It follows that

$$\lim_{T\to\infty} \mathbb{P}(\eta_T - \eta_T^{\epsilon} > \varepsilon) = 0 \ \forall \varepsilon > 0. \tag{5.K}$$

For any $y \in \mathbb{R}$, we can use the following arguments:

$$\begin{aligned}
&\limsup_{T\to\infty} |\mathbb{P}(\eta_T^{\epsilon} \leq y) - \mathbb{P}(\eta_T^{=} \leq y)| \\
&\qquad = \limsup_{T\to\infty} \left|\mathbb{P}(\eta_T^{\epsilon} \leq y) - \mathbb{P}\left(\sup_{\Gamma^{=}} \inf_{\mathrm{K}} H\big(w,\kappa,\lambda;\sqrt{T}F_T\big) \leq y\right)\right| \\
&\qquad \leq \limsup_{T\to\infty} \mathbb{P}\left(\sup_{\Gamma_{\epsilon_T}^{<}} \inf_{\mathrm{K}} H\big(w,\kappa,\lambda;\sqrt{T}F_T\big) > y\right)
\end{aligned}$$

$$\leq \limsup_{T\to\infty} \mathbb{P}\left(\sup_{\Gamma} H\left(w,\kappa,\lambda;\sqrt{T}(F_T - F)\right) > y + \sqrt{T}\epsilon_T\right) = 0 \ \ \forall \kappa \in \mathrm{K}. \tag{5.L}$$

The first inequality uses $|\mathbb{P}(\max(Y,Z) \leq y) - \mathbb{P}(Y \leq y)| \leq \mathbb{P}(Z > y)$. The final equality follows from Proposition 5.1 and the assumed properties of ϵ_T. Clearly,

$$|\mathbb{P}(\eta_T^{\epsilon} \leq y) - \mathbb{P}(\eta_T^{=} \leq y)| \to 0. \tag{5.M}$$

Combining (5.H), (5.K) and (5.M) completes the proof. ■

Proof of Proposition 6.1: To prove part (i), notice that zero lies in the support of the distribution of η_∞, witness, for example, the possible yet negligible event $\mathcal{B}_F(X) = 0, \forall X \in \mathcal{X}^M$. Furthermore, due to the convexity of the sets $\{X \in \mathcal{X}^M : \lambda^{\mathrm{T}} X \geq Y\}$, $\{X \in \mathcal{X}^M : \lambda^{\mathrm{T}} X < Y\}$ for all $\lambda \in \Lambda$ and $Y \in \mathcal{X}^M$ and Assumption 5.1.ii, we have that, excluding negligible events, $H(w,\kappa,\lambda;\mathcal{B}_F)$ equals zero, and thus has a degenerate variance, only if $\kappa = \lambda$. By generalizing Prop. 2.1.10 and 2.1.11 of Nualart (2006) to our case (with his sup replaced by our sup inf), we can derive that the process $\mathcal{H} := H(w,\kappa,\lambda;\mathcal{B}_F)$ has a square integrable Malliavin derivative. By Prop. 2.1.7 of Nualart (2006), we obtain that the support of η_∞ is connected and thereby via our earlier results, we obtain that the support is $[0,+\infty[$.

To prove part (ii), we use that, by analogy to (3.6),

$$\eta_\infty \geq \eta_\infty^* := \sup_{\lambda\in\Lambda} \lambda^{\mathrm{T}} Z - \sup_{\kappa\in\mathrm{K}} \lambda^{\mathrm{T}} Z, \tag{6.A}$$

where Z follows a zero-mean M-variate normal distribution with non-singular variance matrix. The support of η_∞^* is $[0,+\infty[$, which implies that the latter interval also includes the support of η_∞. Inequality (6.A) implies that $\mathbb{P}(\eta_\infty = 0) \leq \mathbb{P}(\eta_\infty^* = 0)$ and due to closeness and convexity of Λ , K and the non-degeneracy of the distribution of Z, $\mathbb{P}(\eta_\infty^* = 0)$ equals the probability that the maximum of $\mathcal{H}$ occurs at a coordinate that corresponds to a common extreme point of Λ and K . Using Thm 2 of Sidak *et al.* (1999, p. 37), we find that $\mathbb{P}(\eta_\infty^* = 0) \leq (M^*/M)$. Hence, the distribution of η_∞ may have an atom at zero of probability at most (M^*/M).

To prove part (iii), consider a restriction of the $\mathcal{H}$ process. $\mathcal{H}$ induces a Gaussian measure on the subspace of the continuous functions on $\Gamma \times \mathrm{K}$ equipped with the sup inf norm that attain the value zero if $\kappa = \lambda$. Let $\mathcal{H}^*$ denote the restriction of $\mathcal{H}$ to the elements of this function space for which the sup inf is strictly positive. The original Gaussian measure assigns a strictly positive probability to this set of functions, because $\mathbb{P}(\eta_\infty^* = 0) < 1$. The (generalized) Nualart propositions apply also to $\mathcal{H}^*$. In addition, the Malliavin derivative of $\mathcal{H}^*$, in contrast to that of $\mathcal{H}$, has a non-zero norm on the set

$$\left\{t \in \Gamma \times \mathrm{K} : \mathcal{H}^*(t) = \sup_{(w,\lambda)\in\Gamma^=} \inf_{\kappa\in \mathrm{K}^{\leq}_{(w,\lambda)}} \mathcal{H}^*\right\}. \tag{6.B}$$

Hence, the law of $\sup_{(w,\lambda)\in\Gamma^=} \inf_{\kappa\in \mathrm{K}^{\leq}_{(w,\lambda)}} \mathcal{H}^*$, whose support is $[0,+\infty[$, is absolutely continuous w.r.t. the Lebesgue measure. Combining this results with the possibility of an atom at zero, we obtain the differentiability and hence continuity of the relevant c.d.f. on $]0,+\infty[$, as in Thm 3 of Lifshits (1983). ■

PROOF OF PROPOSITION 6.2: The behavior under $\mathbf{H}_0$ follows by a direct application of Thm 3.5.1.i of Politis *et al.* (1999), which Proposition 6.1 allows us to use. The behavior under $\mathbf{H}_1$ follows from the following considerations. Proposition 5.1 along with (3.A) - (3.B) imply that

$$\sup_{\mathcal{X},\mathrm{K},\Lambda} |G(x,\kappa,\lambda;F_T) - G(x,\kappa,\lambda;F)| \xrightarrow{p} 0, \tag{6.C}$$

and thereby that

$$\eta_T \xrightarrow{p} \sup_{\lambda\in\Lambda} \inf_{\kappa\in\mathrm{K}} \sup_{x\in\mathcal{X}} G(x,\kappa,\lambda;F) =: \eta(F). \tag{6.D}$$

Using the Skorokhod Representation Theorem (see inter alia Thm 1.10.4 of van der Vaart and Wellner (1996), we may assume the existence of an enriched probability space that supports random variables $l_T \overset{d}{\Rightarrow} c(F_T) - \eta(F)$ that a.s. converge to zero. Hence, for some $\varepsilon > 0$ small enough, and any elementary event ω in the enlarged space we can choose a $T^*(\omega)$ large enough so that, for any $T \geq T^*(\omega)$, under $\mathbf{H}_1$,

$$0 < \sqrt{T}(\eta(F) - \varepsilon) \leq \sqrt{T}\big(\eta(F) + l_T(\omega)\big) \leq \sqrt{T}(\eta(F) + \varepsilon), \tag{6.E}$$

which implies that $\sqrt{T}\big(\eta(F) + l_T(\omega)\big)$ diverges to $+\infty$ a.s. Hence, for any $\delta > 0$,

$$\mathbb{P}(\eta_T > q(\eta_\infty, 1-\alpha) + \delta) = \mathbb{P}\Big(\sqrt{T}\big(\eta(F) + l_T(\omega)\big) > q(\eta_\infty, 1-\alpha) + \delta\Big) \to 1. \tag{6.F}$$

The behavior under $\mathbf{H}_1$ now follows from $q_{T,b_T}(1-\alpha) \xrightarrow{p} q(\eta_\infty, 1-\alpha)$, Proposition 5.2 and Thm 3.5.1.i of Politis *et al.* (1999). ■

PROOF OF PROPOSITION 7.1: We may linearize the empirical shortfall measures in the spirit of the analysis of Conditional Value-at-Risk by Rockafellar and Uryasev (2000):

$$F_T^{(2)}(x,\kappa) = \min_{\theta} \left\{ T^{-1} \sum_{t=1}^{T} \theta_t \,\middle|\, \theta_t \geq (x - X_t^{\mathrm{T}}\kappa);\ \theta_t \geq 0, \qquad t = 1, \cdots, T \right\}. \tag{7.A}$$

The functions $F_T^{(2)}(x,\kappa) = T^{-1}\sum_{t=1}^{T}(x - X_t^{\mathrm{T}}\kappa)1(X_t^{\mathrm{T}}\kappa \leq x)$ and $F_T^{(2)}(x,\lambda) = T^{-1}\sum_{t=1}^{T}(x - X_t^{\mathrm{T}}\lambda)1(X_t^{\mathrm{T}}\lambda \leq x)$ have an increasing and convex piece-wise linear shape with kinks at $x = X_s^{\mathrm{T}}\kappa$ and $x = X_s\lambda$, $s = 1, \cdots, T$, respectively. It follows that the minimization of $G(x,\lambda,\kappa;F_T) = F_T^{(2)}(x,\lambda) - F_T^{(2)}(x,\kappa)$ over $x \in \mathcal{X}$ always achieves an optimal solution at a sub-interval boundary point $x = X_s^{\mathrm{T}}\lambda$, $s = 1, \cdots, T$. Therefore,

$$\begin{aligned} \eta_T(\lambda;\kappa) &:= \sqrt{T} \min_{x \in \mathcal{X}} G(x,\lambda,\kappa;F_T) \\ &= \sqrt{T} \min_{s=1,\cdots,T} G(X_s\lambda,\lambda,\kappa;F_T) \\ &= \sqrt{T} \max_{\gamma} \{\gamma | \gamma \leq G(X_s\lambda,\lambda,\kappa;F_T) \quad s = 1, \cdots, T\}. \end{aligned} \tag{7.B}$$

Combining (7.A) and (7.B), we find the following linear problem in canonical form for pairwise comparison of two given portfolios:

$$\begin{aligned} \eta_T(\lambda;\kappa) = {} & \max \sqrt{T}\gamma \\ s.t.\ \gamma + T^{-1}\sum_{t=1}^{T}\theta_{s,t} &\leq F_T^{(2)}(X_s\lambda,\lambda), \qquad s = 1, \cdots, T; \\ -\theta_{s,t} &\leq -X_s\lambda + X_t^{\mathrm{T}}\kappa, \qquad s,t = 1, \cdots, T; \\ \theta_{s,t} &\geq 0, \qquad s,t = 1, \cdots, T; \\ &\gamma \text{ free.} \end{aligned} \tag{7.C}$$

This linear maximization problem can be embedded directly in the maximization over the portfolio weights $\kappa \in \mathrm{K}$ to yield $\eta_T(\lambda) = \max_{\kappa \in \mathrm{K}} \eta_T(\lambda;\kappa)$, or the optimal value of the objective function of LP problem in canonical form (7.3).■

PROOF OF PROPOSITION 7.2: Any piecewise-linear function $u \in \underline{\mathcal{U}_2}$ consists of segments of $card\{\mathcal{N}\}$ linear lines $g_n(y) := c_{0,n} + c_{1,n}y$ that connect knots z_n, $n \in \mathcal{N}$. Since the piecewise-linear function is concave, it can equivalently be formulated as $u(y) = \min_{n \in \mathcal{N}} g_n(y)$. Equipped with this result, our proof consists of the following arguments:

$$\begin{aligned} \sup_{\lambda \in \Lambda} \mathbb{E}_{F_T}[u(X^{\mathrm{T}}\lambda)] &= \max_{\lambda \in \Lambda} T^{-1} \sum_{t=1}^{T} u(X_t^{\mathrm{T}}\lambda) \\ &= \max_{\lambda \in \Lambda} T^{-1} \sum_{t=1}^{T} \min_{n \in \mathcal{N}} \left(c_{0,n} + c_{1,n}X_t^{\mathrm{T}}\lambda\right) \\ &= \max_{\lambda \in \Lambda} T^{-1} \sum_{t=1}^{T} \max_{y_t} \{y_t | y_t \leq c_{0,n} + c_{1,n}X_t^{\mathrm{T}}\lambda, n \in \mathcal{N}\} \end{aligned}$$

$$= \max_{\substack{\lambda \in \Lambda, \\ \{y_t\}}} T^{-1} \left\{ \sum_{t=1}^{T} y_t \middle| y_t \le c_{0,n} + c_{1,n} X_t^{\mathrm{T}} \lambda, n \in \mathcal{N} \right\}.$$

Bringing all model variables to the l.h.s. and coefficients to the r.h.s. gives the canonical form (7.11). ■

REFERENCES

[1] Ali, Mukhtar M., 1975, Stochastic dominance and portfolio analysis, *Journal of Financial Economics* 2, 205-229.

[2] Barrett, Garry and Stephen Donald, 2003, Consistent tests for stochastic dominance, *Econometrica* 71, 71-104.

[3] Bawa, Vijay S., 1975, Optimal Rules for Ordering Uncertain Prospects, *Journal of Financial Economics* 2, 95-121.

[4] Bawa, Vijay S., James N. Bodurtha Jr., M. R. Rao and Hira L. Suri, 1985, On Determination of Stochastic Dominance Optimal Sets, *Journal of Finance* 40, 417–431.

[5] Cass, David, and Joseph Stiglitz, 1970, The structure of investor preferences and asset returns, and separability in portfolio allocation, *Journal of Economic Theory* 2, 122–160.

[6] Davidson, Russell and Jean-Yves Duclos, 2000, Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality, *Econometrica* 68, 1435-1464.

[7] Dittmar, R.F., 2002, Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns, *Journal of Finance* 57, 369-403.

[8] Gibbons, M. Ross, S., and J. Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

[9] Hadar, J. and W. R. Russell, 1969, Rules for Ordering Uncertain Prospects, *American Economic Review* 59, 2–34.

[10] Hanoch, G., and H. Levy, 1969, The Efficiency Analysis of Choices Involving Risk, *Review of Economic Studies* 36, 335–346.

[11] Harvey, C. R., and A. Siddique, 2000, Conditional Skewness in Asset Pricing Tests, *Journal of Finance* 55, 1263–1295.

[12] Huberman, G. and S. Kandel, 1987, Mean-Variance Spanning, *Journal of Finance* 42, 873-888.

[13] Jakubowski, A., Memin, J. and Pages, G., 1989, Convergence en loi des suites d'integrales stochastiques sur l'espace $\mathbb{D}^1$ de Skorokhod. *Probability Theory and Related Fields* 81, 111-137.

[14] Kuosmanen, Timo, 2004, Efficient diversification according to stochastic dominance criteria, *Management Science* 50, 1390–1406.

[15] Levy, Haim, and Harry Markowitz, 1979, Approximating Expected Utility by a Function of Mean and Variance, *American Economic Review* 69 (3), 308-317.

[16] Lifshits, M. A., 1983, On the absolute continuity of distributions of functionals of random processes, *Theory of Probability & Its Applications* 27.3, 600-607.

[17] Linton, O., E. Maasoumi and Y.-J. Whang, 2005, Consistent Testing for Stochastic Dominance under General Sampling Schemes, *Review of Economic Studies* 72, 735–765.

[18] Linton, Oliver B., Thierry Post, and Yoon-Jae Whang, 2014, Testing for the Stochastic Dominance Efficiency of a Given Portfolio, *Econometrics Journal* 17, 59–74.

[19] Nualart, David, 2006, *The Malliavin calculus and related topics*, Volume 1995. Berlin: Springer.

[20] Politis, D. N., J. P. Romano and M. Wolf. *Subsampling*. Springer New York, 1999.

[21] Post, Thierry, 2003, Empirical Tests for Stochastic Dominance Efficiency, *Journal of Finance* 58, 1905–1932.

[22] Post, Thierry, and P. Versijp, 2007, Multivariate tests for stochastic dominance efficiency of a given portfolio, *Journal of Financial and Quantitative Analysis* 42, 489-516.

[23] Quirk, J. P. and R. Saposnik, 1962, Admissibility and Measurable Utility Functions, *Review of Economics Studies* 29, 140-146.

[24] Rio, E., 2000, *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants*, Collection: Mathématiques et Applications, Vol. 31, Berlin: Springer-Verlag.
[25] Rockafellar, R. T. and S. Uryasev, 2000, Optimization of Conditional Value-at-Risk, *Journal of Risk* 2, 21–41.
[26] Roman, D., Darby-Dowman, K. and Mitra, G., 2006, Portfolio construction based on stochastic dominance and target return distributions, *Mathematical Programming* 108, 541-569.
[27] Ross, S. A., 1978, Mutual Fund Separation in Financial Theory—The Separating Distributions, *Journal of Economic Theory* 17, 254-83.
[28] 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.
[29] Scaillet, O., and N. Topaloglou, 2010, Testing for stochastic dominance efficiency, *Journal of Business and Economic Statistics* 28, 169–180.
[30] Shalit, Haim, and Shlomo Yitzhaki, 1994, Marginal Conditional Stochastic Dominance, *Management Science* 40, 670–684.
[31] Sidak, Zbynek, Pranab K. Sen, and Jaroslav Hajek, 1999, *Theory of rank tests*. Academic Press.
[32] Sion, M, 1958, On general minimax theorems, *Pacific Journal of Mathematics* 8, 171—176.
[33] Tobin, James, 1958, Liquidity Preference as Behavior Towards Risk, *Review of Economic Studies* 25, 65–86.
[34] Van Der Vaart, Aad W, 1997, *Asymptotic Statistics*. CUP.
[35] Van Der Vaart, Aad W. and Jon A. Wellner, 1996, *Weak Convergence*, Springer New York.
[36] Yaari, M.E., 1987, The dual theory of choice under risk*, Econometrica* 55, 95–115.

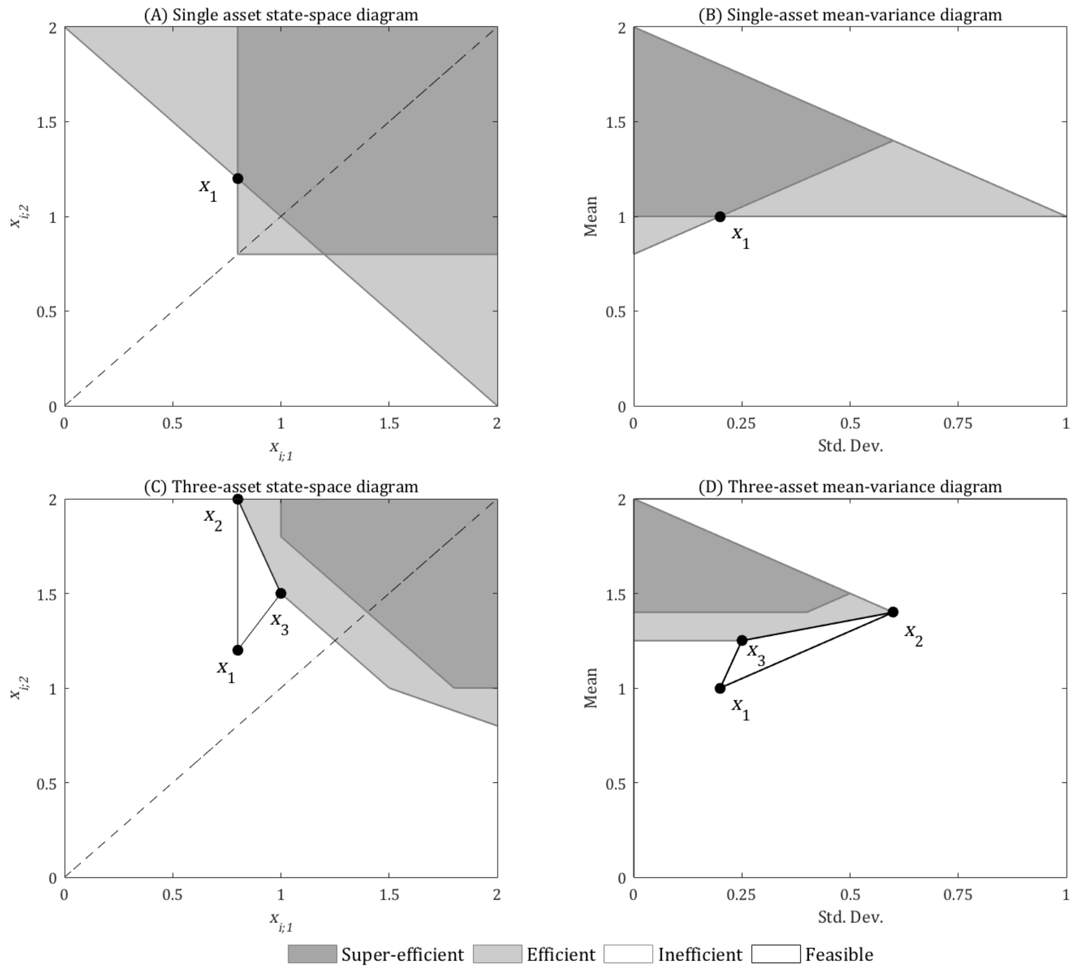

Figure 1: Numerical Example

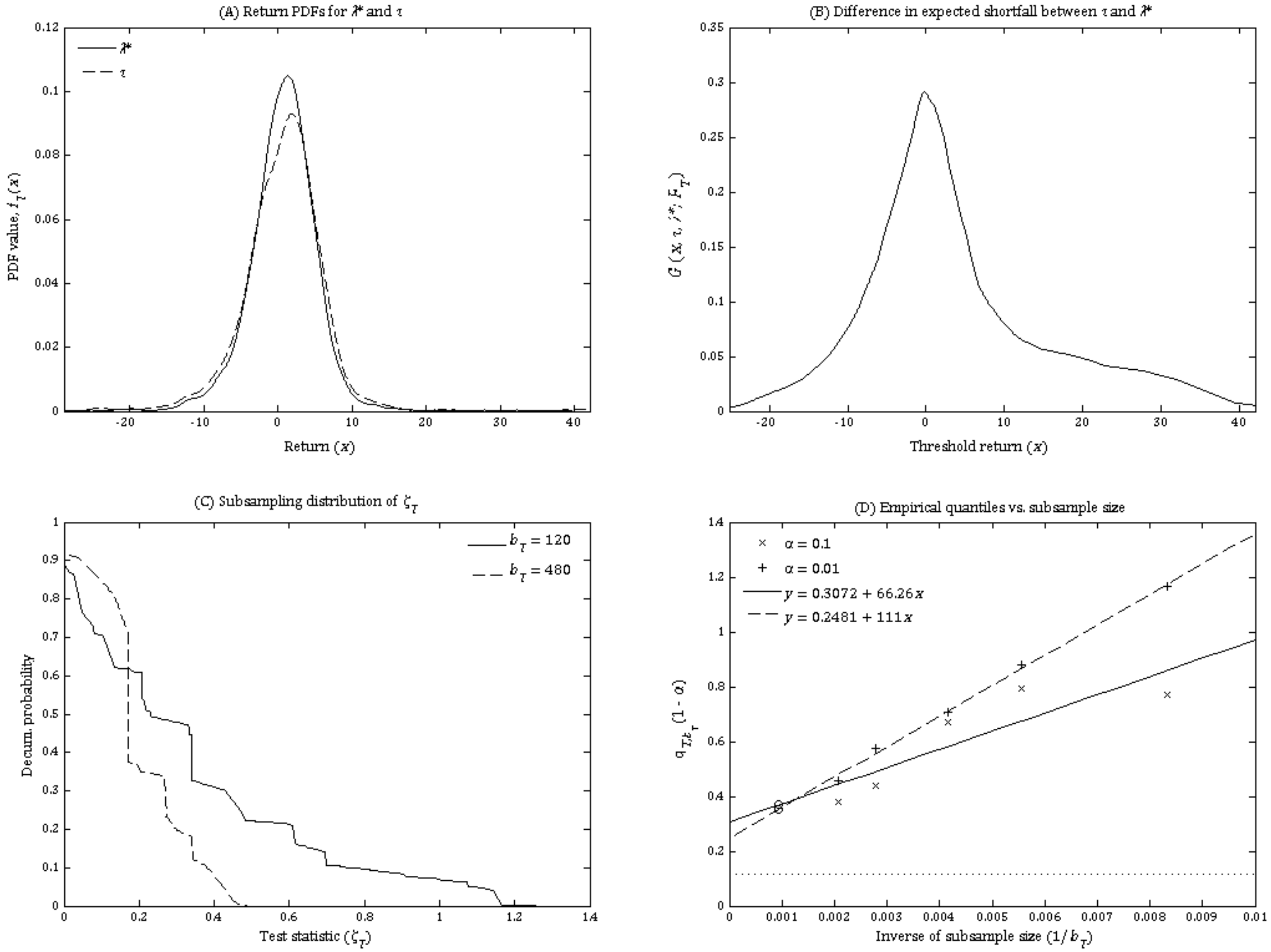

Figure 2: Empirical test for the hypothesis of stochastic efficiency

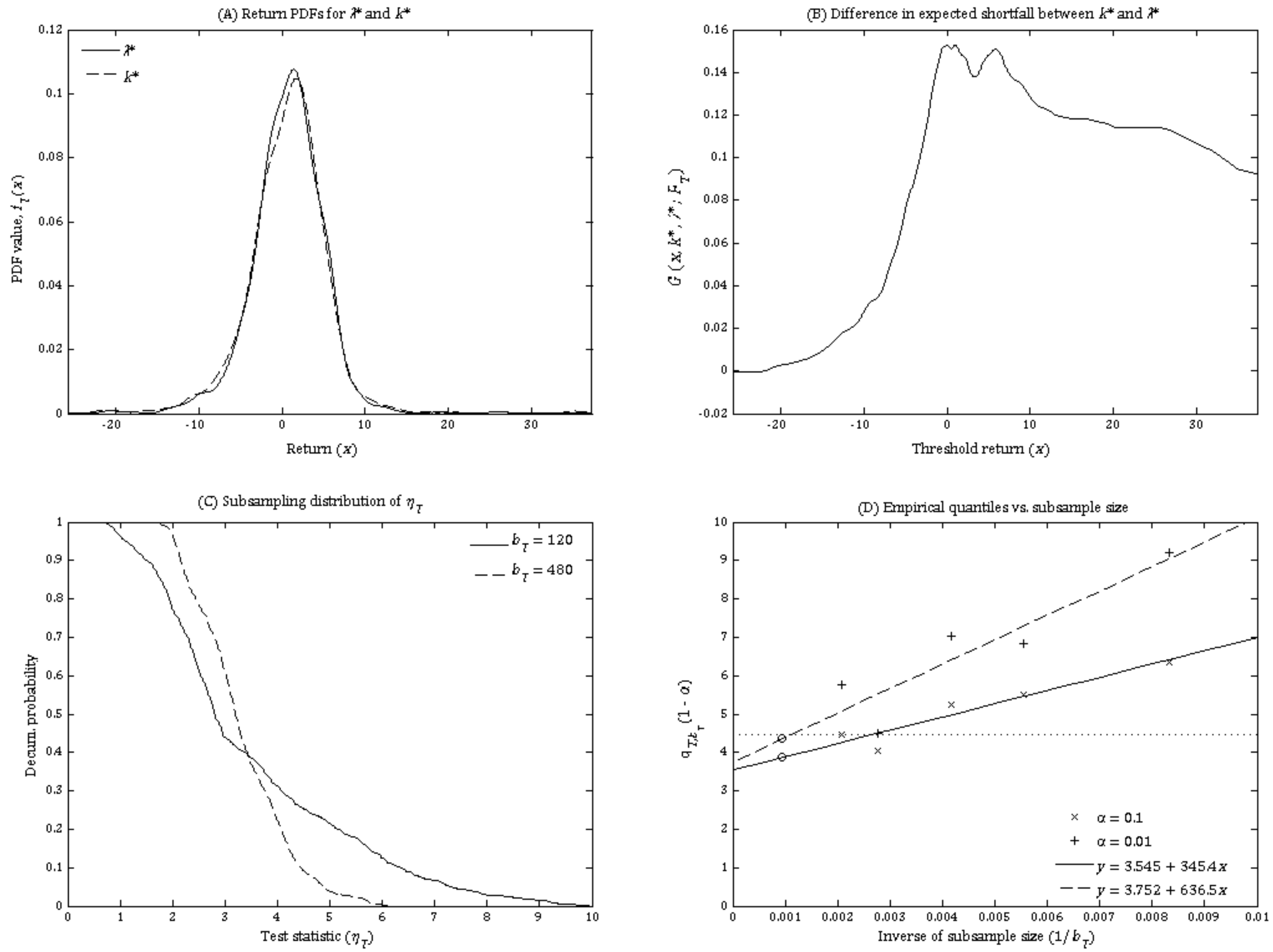

Figure 3: Empirical test for the hypothesis of stochastic spanning

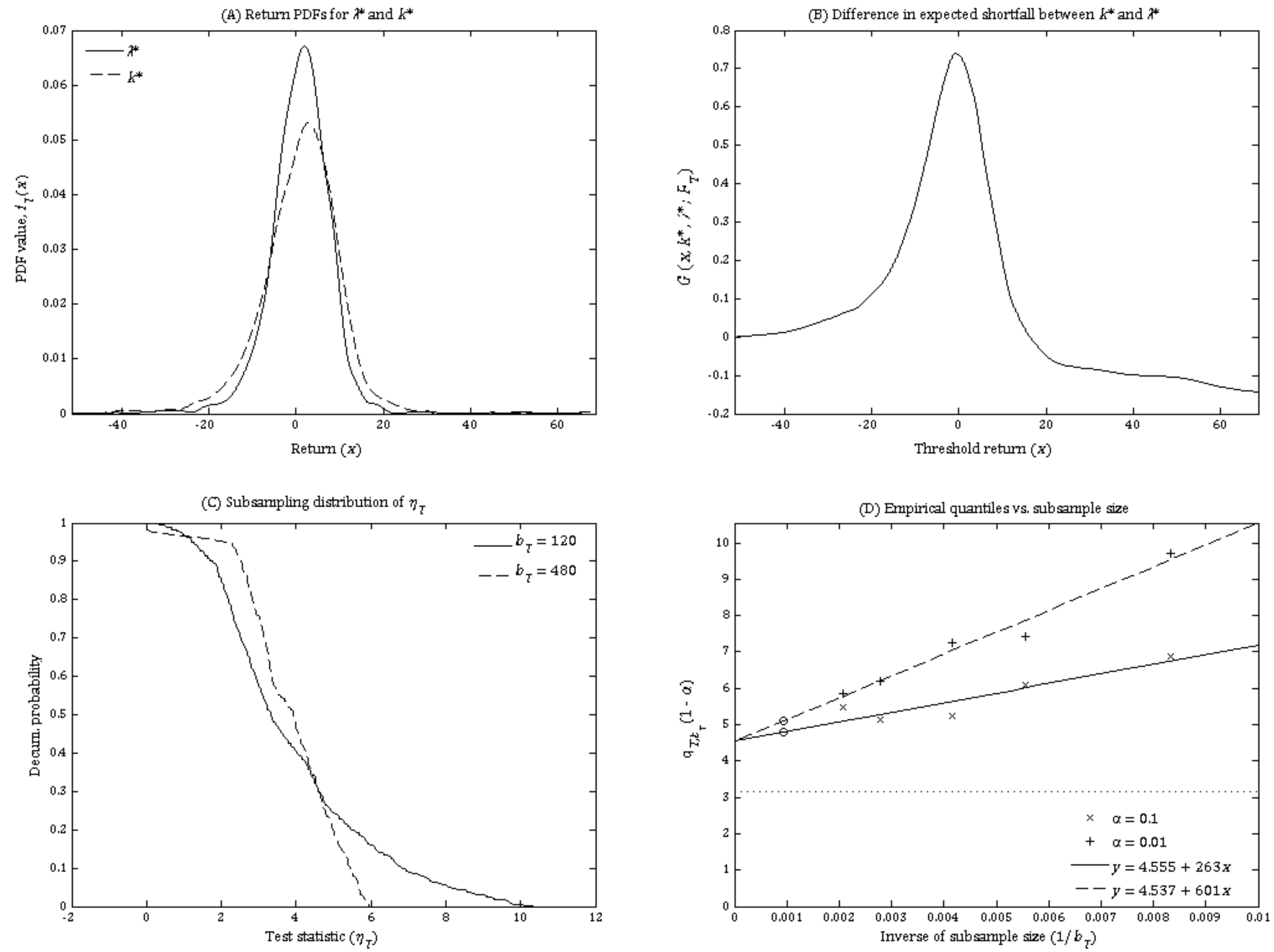

Figure 4: Empirical test for the hypothesis of M-V spanning