Spanning Tests for Markowitz Stochastic Dominance

Stelios Arvanitis

Athens University of Economics and Business *

Olivier Scaillet

University of Geneva and Swiss Finance Institute †

Nikolas Topaloglou

IPAG Business School and Athens University of Economics and Business ‡

First draft: January 2018

This draft: November 2018

*Department of Economics, 76, Patision Street, GR10434, Athens, Greece; e-mail: stelios@aueb.gr

†Corresponding author: GFRI, Bd du Pont d'Arve 40, 1211 Geneva, Switzerland; e-mail: olivier.scaillet@unige.ch

‡ IPAG Business School, 75006, France, and Department of International European & Economic Studies, 76, Patision Street, GR10434, Athens, Greece; e-mail: nikolas@aueb.gr

Abstract

We derive properties of the cdf of random variables defined as saddle-type points of real valued continuous stochastic processes. This facilitates the derivation of the first-order asymptotic properties of tests for stochastic spanning given some stochastic dominance relation. We define the concept of Markowitz stochastic dominance spanning, and develop an analytical representation of the spanning property. We construct a non-parametric test for spanning based on subsampling, and derive its asymptotic exactness and consistency. The spanning methodology determines whether introducing new securities or relaxing investment constraints improves the investment opportunity set of investors driven by Markowitz stochastic dominance. In an application to standard data sets of historical stock market returns, we reject market portfolio Markowitz efficiency as well as two-fund separation. Hence, we find evidence that equity management through base assets can outperform the market, for investors with Markowitz type preferences.

Key words and phrases: Saddle-Type Point, Markowitz Stochastic Dominance, Spanning Test, Linear and Mixed integer programming, reverse S-shaped utility.

JEL Classification: C12, C14, C44, C58, D81, G11.

1 Introduction

An essential feature of any model trying to understand asset prices or trading behavior is an assumption about investor preferences, or about how investors evaluate portfolios. The vast majority of models assume that investors evaluate portfolios according to the expected utility framework. Investors are assumed to act as non -atiable and risk averse agents, and their preferences are represented by increasing and globally concave utility functions.

Empirical evidence suggests that investors do not always act as risk averters. Instead, under certain circumstances, they behave in a much more complex fashion exhibiting characteristics of both risk loving and risk averting. They seem to evaluate wealth changes of assets w.r.t. benchmark cases rather than final wealth positions. They behave differently on gains and losses, and they are more sensitive to losses than to gains (loss aversion). The relevant utility function can be either concave for gains and convex for losses (S-Shaped) or convex for gains and concave for losses (reverse S-Shaped). They seem to transform the objective probability measures to subjective ones using transformations that potentially increase the probabilities of negligible (and possibly averted) events, which, in some cases, share similar analytical characteristics to the aforementioned utility functions. Examples of risk orderings that (partially) reflect such findings are the dominance rules of behavioral finance (see Friedman and Savage (1948), Baucells and Heukamp (2006), Edwards (1996), and the references therein).

Accordingly, stochastic dominance has been used over the last decades in this framework, having more generally evolved into an important concept in the fields of

economics, finance and statistics/econometrics (see inter alia Kroll and Levy (1980), McFadden (1989), Levy (1992), Mosler and Scarsini (1993), and Levy (2005)), since it enables inference on the issue of optimal choice in a non-parametric setting. Several statistical tools have been developed to test whether, given some fixed notion of stochastic dominance, a probability distribution of interest (or some random element that represents it) dominates any other similar distribution in a given set, i.e., the former is super-efficient over the latter set (see Arvanitis et al. (2018)). Analogous procedures have been developed to test whether this distribution is not dominated by any other member of the given set, i.e., whether it is an efficient element of it (see Linton Post and Wang (2014)). We can find some illustrative examples in the application sections of Horvath, Kokoszka, and Zitikis (2006), where interest lies in distributions of income, or Post and Levy (2005), Scaillet and Topaloglou (2010), Linton, Post and Whang (2014), where interest lies in financial portfolios.

There is a large evolving literature on the first (FSD) and on the second (SSD) order stochastic dominance. We can characterize FSD via the choice under uncertainty of every non-satiable investor, while we can characterize SSD by the analogous choice of every risk averse and non-satiable investor (see Hadar and Russell (1969), Hanoch and Levy (1969), and Rothschild and Stiglitz (1970)). Higher order stochastic dominance relations impose more restrictions on the underlying utilities of the set of investors while retaining non-satiety and risk aversion. Dropping global risk aversion, Levy and Levy (2002) formulate the notions of prospect stochastic dominance (PSD) (see also Levy and Wiener (1998), Levy and Levy (2004)) and Markowitz stochastic dominance (MSD). Those notions investigate choices by investors who have S-shaped utility functions and reverse S-shaped utility functions. Arvanitis and

Topaloglou (2017) accordingly develop consistent statistical tests for PSD and MSD super-efficiency.

Given a stochastic dominance relation, the concept of stochastic spanning subsumes the aforementioned notion of super-efficiency. It is an idea of Thierry Post, influenced by Mean-Variance spanning in Huberman and Kandell (1987), that was formulated in the context of second order stochastic dominance in Arvanitis et al. (2018). It is yet generalizable to arbitrary stochastic dominance relations. Given such a relation, and if the underlying set of efficient elements, i.e., the efficient set, is non-empty, a spanning set is simply any superset of the efficient set. As such, we can use a spanning set to provide an "outer approximation" of the underlying efficient set, and/or, when small enough, to provide with a desirable reduction of the initial set of distributions upon which the stochastic dominance ordering is defined, and which could be complicated. In such a case, we can reduce the examination of the optimal choice problem, to a potentially easier and more parsimonious one. Both issues are of interest to financial economics since the underlying distributions often represent the return behaviour of financial assets and the dominance orderings reflect classes of investor preferences (e.g. for the FSD and SSD, as well as the PSD and MSD rules and their relations to classes of utility functions, see Levy and Levy (2002)). Those notions could also be of potential interest in any field of economic theory or decision science that examines optimal choice under uncertainty.

For example, if a strict subset of a universe of available assets is known to be spanning w.r.t. a stochastic dominance relation that reflects all preferences with some sort of combination of local risk aversion with local risk seeking behavior (see for example the MSD preorder defined in Section 3.1), any investor with such a

disposition towards risk can safely restrict her choice to the spanning set. On the contrary, if it is not spanning, there must exist investors with suchlike preferences that benefit from the enlargement of the investment opportunities from the subset to the superset. This implies that stochastic spanning can be useful in extracting important properties of financial markets for investment decisions taylor made for particular shapes of utility functions.

Hence the following question naturally arises: for some fixed stochastic dominance relation, is a given set of assets spanned by a (possibly economically relevant) subset? When the two sets are not equal, spanning occurs if and only if a functional defined by a complex recursion of optimizations is zero (see for example the discussion in page 6 of Arvanitis et al. (2018) for the case of SSD, or Proposition 1 below for the case of MSD). Its empirical verification is usually analytically intractable due to the dependence of the functional on the generally unknown underlying distributions and/or due to the complexity of the optimizations involved. Hence, this is not of direct practical use. However, we can design non-parametric tests of the null hypothesis of spanning given the existence of empirical information. The limit theory of tests for stochastic spanning[1] usually involves null weak limits represented as a finite recursion of optimization functionals applied on some relevant Gaussian process that could have the form of a saddle-type functional. The possibility of the existence of atoms in their distribution affects the issue of asymptotic exactness of the aforementioned tests which are usually based on resampling procedures such as

[1]Spanning tests subsume as special cases tests of super-efficiency w.r.t. the underlying preorder. For example, procedures developed in Scaillet and Topaloglou (2010), Arvanitis and Topaloglou (2017), and Linton et al. (2014), can be considered as spanning tests for singleton spanning sets.

bootstrap and subsampling (Linton et al. (2005)). In order to obtain exactness, we cannot thus rely on standard probabilistic results used in the previous work on tests of super-efficiency, due to the complexity of the aforementioned functional.

Hence, our first contribution is the theoretical study of continuity properties of the cdf of random variables defined as saddle type points of real valued stochastic processes. Section 2 of the paper sets up the probabilistic framework, and derives new properties of the law of a random variable defined by a finite number of nested optimizations on a continuous process w.r.t. possibly interdependent parameter spaces. Beside its usefulness for the limit theory of spanning tests developed in this paper, this result is also a non-trivial extension to results concerning suprema of other stochastic processes and can be useful in other econometric settings (see Section 2 for references and examples).

Our second contribution is the following. The results in Arvanitis et al. (2018) concern the concept of stochastic spanning w.r.t. the SSD relation, which essentially represents all preferences with global risk aversion, and are derived in a context of bounded support for the underlying distributions. We expect that analogous, yet possibly more complex results, on the properties of spanning sets, their representation by relevant functionals, the construction of testing procedures, and the derivation of their limit theory hold if we extend to local risk aversion and general supports. Statistical tests concerning the issue of super-efficiency w.r.t. stochastic dominance rules representing local attitudes towards risk have already appeared in the literature (see for example Post and Levy (2005), or Arvanitis and Topaloglou (2017)), but to our knowledge the concept of spanning has not been studied yet for such dominance relations.

Section 3 investigates the concept of stochastic spanning w.r.t. the MSD preorder in the context of financial portfolios formation. We define the notion and provide with an original characterization of spanning by the zero of a functional. Using the principle of analogy, we define the non-parametric test statistic, derive its limit distribution under the null hypothesis, and define a subsampling algorithm for the approximation of the asymptotic critical values. Among others, we use the new probabilistic results of Section 2 and a novel combinatorial argument, for the derivation of asymptotic exactness when the relevant limit distribution is non-degenerate and a restriction on the significance level holds. In particular, we derive consistency of the subsampling procedure. In contrast to the results in Arvanitis et al. (2018), we allow for unbounded supports for the return distributions, and we suppose that the relevant parameter spaces are simplicial complexes. We explain in Section 3 why those extensions are useful and how we have to modify the theoretical arguments to accommodate them.

Section 4 provides with a numerical implementation consisting of a finite set of Linear Programming (LP) and Mixed Integer Programming (MIP) problems, the latter being highly non linear optimization problems to solve.

Inspired by Arvanitis and Topaloglou (2017), who show that the market portfolio is not MSD efficient, we test in an empirical application in Section 5, whether investors with MSD preferences could beat the market through equity management, according to Markowitz preferences. We use equity portfolios as base assets. We show that the market portfolio is not Markowitz efficient, and the two-fund separation theorem does not hold for MSD investors. Thus, combinations of the market and the riskless asset do not span the portfolios created according to the MSD cri-

terion. We also show that equity managers with MSD preferences could generate portfolios that yield 30 times higher cumulative return than the market over the last 50 years. Standard performance and risk measures show that the optimal MSD portfolios better suit the MSD investors that are risk averse for losses and risk lovers for gains. It achieves a transfer of probability mass from the left to the right tail of the return distribution when compared to the market portfolio. Its return distribution exhibits less negative skewness, less kurtosis, and less negative tail risk. Finally, using the four-factor model of Carhart (1997) and the five-factor model of Fama and French (2015), we investigate which factors explain these returns. We find that a defensive tilt explains part of the performance of the optimal MSD portfolios, while momentum and profitability do not.

In the final section, we conclude. We present the proofs of the main and the auxiliary results in the Appendix.

2 Probabilistic Results

Suppose that $\Lambda_1, \Lambda_2, \ldots, \Lambda_s$ are separable metric spaces, and let $\Lambda := \prod_{i=1}^{s} \Lambda_i$ be equipped with the product topology. Consider the functional oper $:= \mathrm{opt}_1 \circ \mathrm{opt}_2 \circ \cdots \circ \mathrm{opt}_s$ where $\mathrm{opt}_i = \sup$ or $\inf$ w.r.t. to some non-empty compact $\Lambda_i^\star \subseteq \Lambda_i$, for $i = 1, \ldots, s$. When $i > 1$, $\Lambda_i^\star$ is allowed to depend on the elements of $\prod_{j=1}^{i-1} \Lambda_{i-j}^\star$.

The probabilistic framework follows closely Chapter 2 of Nualart (2006). It consists of a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is generated by some isonormal Gaussian process $W = \{W(h), h \in H\}$ and H is an appropriate Hilbert space. X is some vector valued stochastic process on Λ with sample paths in the space of

continuous functions $\Lambda \to \mathbb{R}^q$ equipped with the uniform metric. In many applications, X is a Gaussian weak limit of some net of processes. We denote the Malliavin derivative operator (see Nualart (2006)) by D and by $\mathbb{D}^{1,2}$ the completion of the family of Malliavin differentiable random variables w.r.t. the norm $\sqrt{\mathbb{E}\left[z^2 + (Dz)^2\right]}$.

We are interested in the form of the support and the continuity properties of the cdf of the law of the random variable $\xi := \text{oper} X_\lambda$. The following assumption describes sufficient conditions for the aforementioned law to have a countable number of atoms while being absolutely continuous when restricted between their successive pairs. Given this, the result to be established below, allows first for the random variable at hand to be defined by saddle-type functionals,[2] and second for discontinuities of the resulting cdf. Hence, it generalizes known results concerning the absolute continuity of the distribution of suprema of stochastic processes. For an excellent treatment of those see inter alia, Propositions 2.1.7 and 2.1.10 of Nualart (2006), and for the discontinuities related literature on the fibering method and its probabilistic applications, see Lifshits (1983).

Assumption 1. *For the process X suppose that:*

1. $\mathbb{E}\left[\sup_\Lambda \left(X_\lambda^2\right)\right] < +\infty$.

2. *For all $\lambda \in \Lambda$, $X(\lambda) \in \mathbb{D}^{1,2}$, and the H -valued process DX has a continuous version and $\mathbb{E}\left[\sup_\Lambda \|DX_\lambda\|^2\right] < +\infty$.*

3. *For some countable $\mathcal{T} \subset \mathbb{R}$, $\mathbb{P}\left(\{\xi = \tau\} \cap \Omega_\tau\right) \geq 0$ holds if and only if $\tau \in \mathcal{T}$, where Ω_τ denotes $\{\omega \in \Omega : DX_\lambda(\omega) = 0$ for some λ such that $\tau = X_\lambda(\omega)\}$.*

[2]The term "saddle-type" is used here in a somewhat abusive manner, since commutativity between the successive optimization operations does not hold in general.

In the usual case where X is zero-mean Gaussian, we can establish the first condition by strong results that imply the subexponentiality of the distribution of $\sup_\Lambda X_\lambda$, like Proposition A.2.7 of van der Vaart and Wellner (1996). Its validity follows from conditions that restrict the packing numbers of $\Lambda \times \mathbb{R}$ metrized as a totally bounded metric space by the use of the covariance function of X, to be polynomially bounded, something that is easily established if the Λ_i are subsets of Euclidean spaces for all i. In the same respect, the second condition is easily established as in Nualart (2006) (see page 110). More specifically, if $K(\lambda_1, \lambda_2)$ is the aforementioned covariance function, then H is the closed span of $\{h_\lambda(\cdot) = K(\lambda, \cdot), \lambda \in \Lambda\}$, with inner product $\langle h_{\lambda_1}, h_{\lambda_2} \rangle_H = K(\lambda_1, \lambda_2)$, whence $DX_\lambda = K(\lambda, \lambda)$. In this case, the previous along with dominated convergence implies the existence of $\mathbb{E}[\sup_\Lambda \|DX_\lambda\|^2]$. The third condition is the most difficult to establish. In the cases that we have in mind, we can derive "outer approximations" of $\mathcal{T}$ by analogous, as well as easier to establish, properties of random variables that are stochastically dominated by ξ, see for example the corollary below.

We are now able to state and prove the main probabilistic result.

Theorem 1. *Under Assumption 1, the law of ξ has connected support, say* $\operatorname{supp}(\xi)$*, that contains $\mathcal{T}$. If $\tau \in \mathcal{T}$, the cdf of the law evaluated at τ has a jump discontinuity of size at most $\mathbb{P}(\Omega_\tau)$. If τ_1, τ_2 are successive elements of $\mathcal{T}$, the law restricted to (τ_1, τ_2) is absolutely continuous w.r.t. the Lebesgue measure. If $\mathcal{T}$ is bounded from below then the law restricted to $(-\infty, \inf \mathcal{T})$ is absolutely continuous w.r.t. the Lebesgue measure. Dually, if $\mathcal{T}$ is bounded from above then the law restricted to $(\sup \mathcal{T}, +\infty)$ is absolutely continuous w.r.t. the Lebesgue measure.*

Theorem 1 encompasses the standard absolute continuity results in the aforemen-

tioned literature that hold when oper is a composition of suprema (or dually infima), the parameter spaces Λ are not dependent, and $\mathbb{P}(\Omega_\tau) = 0$, for all $\tau \in \mathcal{T}$. Furthermore, even in the special case where $\mathcal{T}$ is a singleton, the result is a generalization of Theorem 2 of Lifshits (1983) since it allows for non-Gaussianity, dependence between the domains of the optimization operators, as well as saddle-type optimizations. The following corollary focuses on this particular case and estimates the size of the potential jump discontinuity by assuming the existence of an auxiliary random variable that is stochastically dominated by ξ.

Corollary 1. *Suppose that Assumption 1 is satisfied. Furthermore, suppose that* $\mathcal{T} = \{c\}$, $\xi \geq \eta$, $\mathbb{P}$ *a.s., and that* $supp(\eta) = [c, +\infty)$. *Then,* $supp(\xi) = [c, +\infty)$, *its cdf is absolutely continuous on* $(c, +\infty)$, *and it may have a jump discontinuity of size at most* $\mathbb{P}(\eta = c)$ *at* c.

The results above, and especially the previous corollary, are useful for the derivation of the limit theory for our test of stochastic spanning (see Arvanitis et al. (2018) for the case of SSD based on other arguments). For a given pair of sets of probability distributions driven by sets of portfolio allocations, the null hypothesis of spanning posits that, for any distribution in the first set, there exists some in the other one that dominates it. Below, such a hypothesis is represented by a functional of the form $\sup\sup\inf$ of an appropriate set of moment conditions parameterized by such a Λ. We can obtain a test statistic through a scaled empirical version of this functional. Under the null limit theory for the test statistic, the results above are useful for the construction of an asymptotically exact decision procedure based on a resampling scheme. They do so by providing with restrictions on the asymptotic significance level that guarantee the convergence of the critical values to continuity points of the

null limiting cdf. In such frameworks, X is usually zero-mean Gaussian, while ξ is conveniently defined as a difference between infima of X defined on different regions of Λ with given properties (see the following sections for explicit derivations of those properties in the case of MSD).

We can meet similar probabilistic structures in other econometric applications. An example concerns the null hypothesis of nesting of a given statistical model by a set identified model represented by moment inequalities. More specifically, suppose that given a random matrix Y, a statistical model is comprised by a set of probability distributions conditional on Y and parameterized by a Euclidean parameter $\varphi \in \Phi$. A second statistical model is comprised by the set of probability distributions conditional on Y that satisfy the conditional moment inequalities $\mathbb{E}\left[\mathbf{g}\left(\theta\right)|Y\right] \leq \mathbf{0}_d$, for some $\theta \in \Theta$, where Θ is again a subset of some Euclidean space, the moment function $\mathbf{g} = (g_1, g_2, \ldots, g_d)$ is finite dimensional and the inequality sign is interpreted pointwisely. We are interested in testing the hypothesis that the first model is nested in the second model, i.e., that, for any $\varphi \in \Phi$, there exists some $\theta \in \Theta$ such that $\mathbb{E}_\varphi\left[\mathbf{g}\left(\theta\right)|Y\right] \leq \mathbf{0}_d$, where $\mathbb{E}_\varphi$ denotes expectation w.r.t. the distributions corresponding to φ. When Φ is a singleton, we obtain specification hypotheses similar to the ones in Guggenberger, Hahn and Kim (2008). Under some further conditions on the properties of Φ, Θ and $\mathbf{g}$, the null hypothesis of nesting is equivalent to that $\sup_{\varphi\in\Phi}\sup_{j=1,2,\ldots,d}\inf_{\theta\in\Theta}\mathbb{E}_\varphi\left[g_j\left(\theta\right)/Y\right] \leq 0$. If sampling is available for any $\varphi \in \Phi$ in the first model (this would be trivial in the specification related to the singleton case for Φ), we can form test statistics via empirical counterparts of the functional $\sup_{\varphi\in\Phi}\sup_{j=1,2,\ldots,d}\inf_{\theta\in\Theta}\mathbb{E}_\varphi\left[g_j\left(\theta\right)|Y\right]$. Then, the results above are also useful for the construction of asymptotically exact decision procedures in such a context.

3 A Spanning Test for MSD

We now introduce the concept of stochastic spanning for the MSD relation. We initially provide some order theoretical characterization of the concept, and derive an analytical representation using a functional defined by recursive optimizations. We then construct a testing procedure using a scaled empirical counterpart of that functional and subsampling. We derive its first order limit theory mainly thanks to Corollary 1.

3.1 MSD and Stochastic Spanning

Given $(\Omega, \mathcal{F}, \mathbb{P})$, suppose that F denotes the cdf of some probability measure on $\mathbb{R}^n$ with finite first moment.[3] Let $G(z, \lambda, F)$ be $\int_{\mathbb{R}^n} \mathbb{I}\{\lambda^{Tr} u \leq z\} dF(u)$, i.e., the cdf of the linear transformation $\mathbb{R}^n \ni x \to \lambda^{Tr} x$ where λ assumes its values in $\mathbb{L}$ which is a closed non-empty subset of $\mathbb{S} = \{\lambda \in \mathbb{R}^n_+ : \mathbf{1}^{Tr}\lambda = 1, \}$. Analogously, let $\mathbb{K}$ denote some distinguished subcollection of $\mathbb{L}$. In the context of financial econometrics, F usually represents the joint distribution of n base asset returns, and $\mathbb{S}$ the set of linear portfolios that can be constructed upon the previous.[4] The parameter set $\mathbb{L}$ represents the collection of feasible portfolios formed by economic, legal, and/or other investment restrictions. We denote generic elements of $\mathbb{L}$ by λ, κ, etc. In order

[3] In comparison to the spanning test for the SSD relation of Arvanitis et al. (2018), we do not assume that the random variables have compact supports.

[4] The base assets are not restricted to be individual securities but are defined simply as the extreme points of the maximal portfolio set $\mathbb{S}$.

to define the concepts of MSD and subsequently of spanning, we consider

$$\mathcal{J}(z_1, z_2, \lambda; F) := \int_{z_1}^{z_2} G(u, \lambda, F)\, du.$$

Definition 1. κ weakly Markowitz-dominates λ, denoted by $\kappa \succcurlyeq_M \lambda$, iff

$$\begin{aligned} \Delta_1(z, \lambda, \kappa, F) &:= \mathcal{J}(-\infty, z, \kappa, F) - \mathcal{J}(-\infty, z, \lambda, F) \leq 0,\ \forall z \in \mathbb{R}_-, \text{ and} \\ \Delta_2(z, \lambda, \kappa, F) &:= \mathcal{J}(z, +\infty, \kappa, F) - \mathcal{J}(z, +\infty, \lambda, F) \leq 0,\ \forall z \in \mathbb{R}_{++}. \end{aligned} \tag{1}$$

The existence of the mean of the underlying distribution implies that we can allow the limits of integration above to assume extended values, hence the integral differences Δ_1 and Δ_2 in (1) are well defined. Levy and Levy (2002) show that $\kappa \succcurlyeq_M \lambda$ iff the expected utility of κ is greater than or equal to the expected utility of λ for any utility function in the set of increasing and, concave on the negative part and convex on the positive part real functions (termed as reverse S-shaped (at zero) utility functions). Such utility functions represent preferences towards risk that are associated with risk aversion for losses and risk loving for gains. Hence, in financial economics, Markowitz-dominance is the case iff portfolio κ is weakly prefered to portfolio λ by every reverse S-shaped individual investor.

The uncountable system of inequalities in (1) defines an order on $\mathbb{L}$. If those are satisfied as equalities, the pair (κ, λ) belongs to the possibly non-trivial equivalence part of the order. Strict dominance $\kappa \succ_M \lambda$ corresponds to the irreflexive part of the order and it holds iff at least one of the previous inequalities holds strictly for some $z \in \mathbb{R}$, i.e., portfolio κ is strictly prefered to portfolio λ by some reverse S-shaped individual investor. Finally, given the possibility that Δ_1 and/or Δ_2 can change sign as functions of z, the relation is not generally total. When this is the case, we cannot compare κ and λ w.r.t. $\succcurlyeq_M$.

As in the Mean-Variance case, we can define the efficient set of $\mathbb{L}$ w.r.t. $\succcurlyeq_M$, as the set of maximal elements of the preorder. This means that κ lies in the efficient set iff for any $\lambda \in \mathbb{L}$, either $\kappa \succcurlyeq_M \lambda$ or κ is incomparable to λ. The efficient set has the property that, for any $\lambda \in \mathbb{L}$, there exists some κ in the former for which $\kappa \succcurlyeq_M \lambda$. Any superset of the efficient set has also the same property, but the efficient set is minimal (if we ignore equivalencies) w.r.t. this property. This observation motivates the definition of MSD spanning. This is analogous to the concept of Mean-Variance spanning introduced by Huberman and Kandel (1987), and extended to the SSD case by Arvanitis et al. (2018).

Definition 2. $\mathbb{K}$ Markowitz-spans $\mathbb{L}$ (say $\mathbb{K} \succcurlyeq_M \mathbb{L}$) iff for any $\lambda \in \mathbb{L}$, $\exists \kappa \in \mathbb{K} : \kappa \succcurlyeq_M \lambda$. If $\mathbb{K} = \{\kappa\}$, κ is termed as Markowitz super-efficient.

Spanning sets always exist since by construction $\mathbb{L} \succcurlyeq_M \mathbb{L}$. The efficient set minimally (ignoring equivalencies) spans $\mathbb{L}$, in the sense that any other spanning set must be a superset of it. Hence, we can view any spanning subset of $\mathbb{L}$ as an "outer approximation" of the efficient set. Due to the complexity of (1) w.r.t. the Mean-Variance case, the mathematical properties of the efficient set are generally difficult to derive, but fortunately, they are approximable by properties of sequences of spanning sets that converge to it (see below).

Furthermore, if $\mathbb{K} \succcurlyeq_M \mathbb{L}$, the optimal choice of every reverse S-shaped investor function lies necessarily inside $\mathbb{K}$. Hence, if $\mathbb{K} \subset \mathbb{L}$ and spanning occurs, we can reduce the problem of optimal choice within $\mathbb{L}$ to the analogous problem within $\mathbb{K}$, and the latter is more parsiminious than the former. Dually, if $\mathbb{K}$ does not span $\mathbb{L}$, there must exist optimal choices, and thereby investment opportunities, in the increment $\mathbb{L} - \mathbb{K}$ for some MSD investors. Therefore we can motivate the interest

in the verification of spanning by tractability reasons related to optimal portfolio choice, or by detection of new investment opportunities.

Super-efficiency (Arvanitis and Topaloglou (2017)) corresponds to the existence of a greatest element for $\succcurlyeq_M$, i.e., of a unique (excluding equivalencies) element that weakly Markowitz-dominates every element of $\mathbb{L}$. Given the complexity of (1), greatest elements do not generally exist. This implies that the notion of spanning not only encompasses that of super-efficiency but it is also a property of the order that will more often hold.

The above raise the following question: given $\mathbb{K}$, a non empty subset of $\mathbb{L}$,[5] is $\mathbb{K} \succcurlyeq_M \mathbb{L}$? The following proposition provides with an analytical characterization by means of nested optimizations.

Proposition 1. *Suppose that $\mathbb{K}$ is closed. Then $\mathbb{K} \succcurlyeq_M \mathbb{L}$ iff*

$$\xi(F) := \max_{i=1,2} \sup_{\lambda \in \mathbb{L}} \sup_{z \in A_i} \inf_{\kappa \in \mathbb{K}} \Delta_i(z, \lambda, \kappa, F) = 0, \tag{2}$$

where $A_1 = \mathbb{R}_-$, $A_2 = \mathbb{R}_{++}$. Spanning does not occur iff $\xi(F) > 0$.

The case of super-efficiency is then trivially obtained.

Corollary 2. *Under the scope of the previous lemma, κ is Markowitz super-efficient iff*

$$\max_{i=1,2} \sup_{\lambda \in \mathbb{L}} \sup_{z \in A_i} \Delta_i(z, \lambda, \kappa, F) = 0.$$

[5] We do not look at the issue of the selection of $\mathbb{K}$. Here, the latter is considered as given. In some cases, we can select $\mathbb{K}$ by economically relevant information, see for example the application in Arvanitis et al. (2018) for SSD. We leave the issue of the selection of a candidate spanning set, especially when this selection is related to the approximation of the efficient set, for future research.

Given $\mathbb{K}$, it is generally difficult to directly use the previous proposition since F is usually unknown and/or the optimizations involved are infeasible. However, given the availability of a sample containing information for F and in conjunction with the principle of analogy, it provides the backbone for the construction of inferential procedures that address MSD spanning.

3.2 A Consistent Non-parametric Test

Hypotheses Structure and Test Statistic

We employ Lemma 1 to construct a non-parametric test for MSD spanning. If $\mathbb{K} \succcurlyeq_M \mathbb{L}$ is chosen as the null hypothesis, the hypothesis structure takes the form:[6]

$$\mathbf{H_0} : \xi(F) = 0,$$
$$\mathbf{H_a} : \xi(F) > 0.$$

To design the decision rule, we extend our framework as follows. Consider a process $(Y_t)_{t\in\mathbb{Z}}$ taking values in $\mathbb{R}^n$. Y_{t_i} denotes the ith element of Y_t. The sample of size T is the random element $(Y_t)_{t=1,\ldots,T}$. In our portfolio framework, it represents the observable returns of the n financial base assets. We denote the unknown cdf of Y_t by F, and the empirical cdf by F_T. We consider the test statistic

$$\xi_T := \xi\left(\sqrt{T}F_T\right) = \max_{i=1,2} \sup_{\lambda\in\mathbb{L}} \sup_{z\in A_i} \inf_{\kappa\in\mathbb{K}} \Delta_i\left(z, \lambda, \kappa, \sqrt{T}F_T\right),$$

which is the $\sqrt{T}$-scaled empirical analog of $\xi(F)$. We can equivalently express ξ_T as

[6] Corollary 2 implies that the hypotheses are in the special case of super-efficiency as in Arvanitis and Topaloglou (2017).

a usual scaled empirical average:

$$\xi_T = \max_{i=1,2} \sup_{\lambda \in \mathbb{L}} \sup_{z \in A_i} \inf_{\kappa \in \mathbb{K}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} q_i \left(z, \lambda, \kappa, Y_t\right), \tag{3}$$

where

$$q_i \left(z, \lambda, \tau, Y_t\right) := \begin{cases} K \left(z, \lambda, \kappa, Y_t\right), & i = 1 \\ \left[\left(\lambda' Y_t\right)_+ - \left(\kappa' Y_t\right)_+ - v \left(z, \lambda, \kappa, Y\right)\right], & i = 2 \end{cases},$$

with $v \left(z, \lambda, \kappa, Y\right) := K \left(z, \lambda, \kappa, Y\right) - K \left(0, \lambda, \kappa, Y\right)$, and $K \left(z, \lambda, \kappa, Y\right) := \left(z - \kappa' Y\right)_+ - \left(z - \lambda' Y\right)_+$. This is instrumental in the numerical implementation of (3) in Section 4. When $\mathbb{K}$ is a singleton, the test statistic coincides with the one used in Arvanitis and Topaloglou (2017).

Null Limit Distribution

In order to show that our testing procedure is asymptotically meaningful, we need a limit theory for ξ_T under the null hypothesis. We derive it using the following assumption.

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

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

is positive definite.

The mixing rates condition is implied by stationarity and geometric ergodicity. The latter holds for many stationary models used in the context of financial econometrics, like ARMA, GARCH-type, and stochastic volatility models (see Francq and

Zakoian (2011) for several examples). The moment existence condition enables the validity of a mixing CLT. A CLT typically holds under stricter restrictions. The positive definiteness of the long run covariance matrix is for instance satisfied, if $(Y_t)_{t\in\mathbb{Z}}$ is a vector martingale difference process and the elements of Y_0 are linearly independent random variables. From the compactness of $\mathbb{L}$, the previous implies that $\sup_{\lambda\in\mathbb{L}}\int_{-\infty}^{+\infty}\sqrt{G(u,\lambda,F)(1-G(u,\lambda,F))}du<+\infty$, which is a uniform version of the analogous condition used in Horvath et al. (2006).

We establish the limit theory below via the use of the concept of Skorokhod representations along with an iterative consideration of the dual notions of epi/hypo-convergence. The result depends on the contact sets

$$\Gamma_i=\{\lambda\in\mathbb{L},\kappa\in\mathbb{K},z\in A_i:\Delta_i(z,\lambda,\kappa,F)=0\}.$$

For any i, Γ_i is non empty since $\Gamma_i^\star\equiv\{(\kappa,\kappa,z),\kappa\in\mathbb{K},z\in A_i\}\subseteq\Gamma_i$. Furthermore, if the support of F is bounded, for any $\lambda\in\mathbb{L},\kappa\in\mathbb{K}$, $\exists z\in A_i:(\lambda,z)\in\Gamma_i$, for all $i=1,2,$.[7] Hence, $\Gamma_i^\star\subset\Gamma_i$.

In what follows, we denote convergence in distribution by $\rightsquigarrow$.

Proposition 2. *Suppose that $\mathbb{K}$ is closed, Assumption 2 holds, and $\mathbf{H_0}$ is true. Then as $T\to\infty$, $\xi_T\rightsquigarrow\xi_\infty$, where $\xi_\infty:=\max_{i=1,2}\sup_{z\in A_i}\sup_\lambda\inf_\kappa\Delta_i(z,\lambda,\kappa,\mathcal{G}_F),(\lambda,z,\kappa)\in\Gamma_i$, and $\mathcal{G}_F$ is a centered Gaussian process with covariance kernel given by $Cov(\mathcal{G}_F(x),\mathcal{G}_F(y))=\sum_{t\in\mathbb{Z}}Cov(\mathbb{I}_{Y_0\le x},\mathbb{I}_{Y_t\le y})$ and $\mathbb{P}$ almost surely uniformly continuous sample paths defined on $\mathbb{R}^n$.*[8]

[7] For example, since the support is bounded, we can cover it by some hypercube of the form $[z_l,z_u]^n$ where we can choose z_l as negative. Obviously, $(\lambda,z_l)\in\Gamma_1$, for any $\lambda\in\mathbb{L}$.

[8] See Theorem 7.3 of Rio (2013).

The covariance kernel above, and thereby $\mathcal{G}_F$, are well defined due to the mixing condition and the existence of $\sup_{\lambda\in\mathbb{L}}\int_{-\infty}^{+\infty}\sqrt{G(u,\lambda,F)(1-G(u,\lambda,F))}du$ implied by Assumption 2 (see Remark 1 in Arvanitis and Topaloglou (2017)).

A Subsampling Based Testing Procedure: Limit Theory and Combinatorial Considerations

We cannot directly use the result in Proposition 2 for the construction of an asymptotic decision rule since the distribution of ξ_∞ depends on the unknown covariance kernel of $\mathcal{G}_F$. We can establish a feasible decision rule by the use of a resampling procedure. We consider subsampling, as in Linton et al. (2014)-see also Linton et al. (2005). This resampling is of a non-parametric nature since we do not want to specify parametric conditional distributions for the multivariate return dynamics.

Algorithm. *The testing procedure consists of the following steps:*

1. Evaluate ξ_T at the original sample value.

2. For $0 < b_T \leq T$, generate subsample values from the original observations $(Y_i)_{i=t,\ldots t+b_T-1}$ for all $t = 1, 2, \ldots, T - b_T + 1$.

3. Evaluate the test statistic on each subsample value, obtaining $\xi_{T,b_T,t}$ for all $t = 1, 2, \ldots, T - b_T + 1$.

4. Approximate the cdf of the asymptotic distribution of ξ_T by $s_{T,b}(y) = \frac{1}{T-b_T+1}\sum_{t=1}^{T-b_T+1} 1\left(\xi_{T,b_T,t} \leq y\right)$ and evaluate its $1-\alpha$ quantile $q_{T,b_T}(1-\alpha)$.

5. Reject $\mathbf{H_0}$ iff $\xi_T > q_{T,b_T}(1-\alpha)$.

We derive the first order limit theory via the use of Proposition 2 and of relevant results from the theory of subsampling. We first make the following standard assumption in the subsampling methodology.

Assumption 3. *Suppose that* (b_T)*, possibly depending on* $(Y_t)_{t=1,\dots,T}$*, satisfies*

$$\mathbb{P}\left(l_T \leq b_T \leq u_T\right) \to 1,$$

where (l_T) *and* (u_T) *are real sequences such that* $1 \leq l_T \leq u_T$ *for all* T*,* $l_T \to \infty$ *and* $\frac{u_T}{T} \to 0$ *as* $T \to \infty$*.*

The assumption does not provide with much information on the practical choice of the subsampling rate for fixed T. It is designed to handle issues like asymptotic exactness and consistency. In the following section, along with the numerical implementation for ξ_T, we discuss a method of fixed T correction for the algorithm above, in the spirit of Arvanitis et al. (2018), that involves the use of several subsampling rates.

Asymptotic exactness is derivable by results like Theorem 3.5.1 in Politis, Romano and Wolf (1999). The latter requires continuity of the limit cdf at the quantile corresponding to the significance level α. Even when the distribution of ξ_∞ is non-degenerate, it is possible that it has a cdf with a unique discontinuity at zero (see the proof of Lemma 2 in the Appendix). If there exists a lower bound for ξ_∞, Corollary 1 provides with an estimate for the cdf jump size at zero. Then the use of the aforementioned theorem becomes possible by properly restricting α. This is where the new probabilistic results of Section 2 become useful in our context. It turns out (see the proof of Lemma 2 in the Appendix) that we can obtain such a bound in the form of a non-negative random variable defined as the difference between the

suprema at $\mathbb{L}$ and $\mathbb{K}$ respectively, of a linear Gaussian process. Hence, we get the needed estimate of the jump size as the probability that the latter random variable attains the value zero.

In order to evaluate this, we essentially use some combinatorial notions that allow the estimation of the proportion of the linear functions for which their unique maximizer over $\mathbb{L}$ is a common extreme point of both the parameter spaces.

Definition 3. Suppose that $\mathbb{M}, \mathbb{N}$ are simplicial complexes inside $\mathbb{S}$ and $\mathbb{M} \supseteq \mathbb{N}$. The set of effective extreme points of $\mathbb{N}$ w.r.t. $\mathbb{M}$ is

$$e_{\mathbb{M}}(\mathbb{N}) := \left\{ \lambda \text{ is an extreme point of } \mathbb{N} : \exists \text{ extreme point } s \text{ of } \mathbb{S}\colon \|\lambda - s\| \leq \inf_{\kappa \in \mathbb{M}} \|\kappa - s\| \right\}.$$

Furthermore, if $\lambda \in e_{\mathbb{M}}(\mathbb{N})$ then the set of the adjoint to λ extreme points of $\mathbb{S}$ is

$$c(\lambda) := \left\{ s \text{ is an extreme point of } \mathbb{S} : \|\lambda - s\| \leq \inf_{\kappa \in \mathbb{M}} \|\kappa - s\| \right\}.$$

Given the non-linear simplicial complex forms of $\mathbb{M}, \mathbb{N}$, the notion of an effective extreme point essentially picks the extreme points of $\mathbb{N}$ that can be restricted to $\mathbb{M}$ maximizers of linear real functions defined on $\mathbb{S}$. Given any such extreme point, its adjoint set essentially picks up the extreme points of the incorporating simplex $\mathbb{S}$ that are closer to it than any other extreme point of $\mathbb{M}$.

Definition 4. The $\mathbb{M}$-character of $\lambda \in e_{\mathbb{M}}(\mathbb{N})$ w.r.t. $s \in c(\lambda)$ is

$$ch_{\mathbb{M}}(s, \lambda) := \#\left\{ \kappa \in e_{\mathbb{M}}(\mathbb{N}) : \|\lambda - s\| = \|\kappa - s\| \right\}.$$

Furthermore, the $\mathbb{M}$-character of $\mathbb{N}$ is

$$ch_{\mathbb{M}}(\mathbb{N}) := \sum_{\lambda \in e_{\mathbb{M}}(\mathbb{N})} \sum_{s \in c(\lambda)} \frac{(n - ch_{\mathbb{M}}(s, \lambda))!}{n!}. \tag{4}$$

The ratio $\frac{(n-ch_{\mathbb{M}}(s,\lambda))!}{n!}$ counts the proportion of linear real functions with unique maximizer s over $\mathbb{S}$, and unique maximizer λ over the restricted $\mathbb{M}$. Hence, $ch_{\mathbb{M}}(\mathbb{N})$ counts the proportion of such functions for which the maximizer over $\mathbb{M}$ is an extreme point of $\mathbb{N}$. Suppose now that Z follows a non-degenerate, zero mean, n-dimensional Normal distribution. The characterization (4) of the $\mathbb{M}$-character of $\mathbb{N}$ allows the bounding from above (see the proof of Lemma 2 in the Appendix) of the probability of the event $\sup_{\lambda\in\mathbb{M}} \lambda' Z = \sup_{\lambda\in\mathbb{N}} \lambda' Z$ by $ch_{\mathbb{M}}(\mathbb{N})$, and this is directly related to the estimation of the potential jump size discontinuity of the cdf of ξ_∞ at zero. Thereby, if we assume that $\mathbb{L}$ and $\mathbb{K}$ are simplicial complexes, and if $ch_{\mathbb{L}}(\mathbb{K})$ is easy to evaluate, the previous definitions greatly facilitate and are key for the derivations of the asymptotic exactness of our test.

Assumption 4. *$\mathbb{L}$ and $\mathbb{K}$ are simplicial complexes inside the standard simplex $\mathbb{S} = \{\lambda \in \mathbb{R}^n_+ : 1'\lambda = 1, \}$ and $e_{\mathbb{L}}(\mathbb{K}) \subset e_{\mathbb{L}}(\mathbb{L})$.*

The simplicial form of $\mathbb{L}$ and $\mathbb{K}$ generalizes considerably the setting of Arvanitis et al. (2018). There, those spaces are restricted as convex polytopes. Here, they need not be convex, and they can be disconnected, discrete, etc. This could be useful when the investment categories are constrained because of SRI screening, restrictions on foreign investment, restrictions on available type of shares, etc. This generalization allows for the establishment of the asymptotic validity and thereby the applicability of our test in more complicated scenarios. For example, suppose that $\mathbb{K} = \mathbb{K}_1 \cup \mathbb{K}_2$ which are disjoint simplices. If $\mathbb{K} \succcurlyeq_M \mathbb{L}$, but neither $\mathbb{K}_1 \succcurlyeq_M \mathbb{L}$, nor $\mathbb{K}_2 \succcurlyeq_M \mathbb{L}$, then this directly implies that the efficient set is disconnected. If $e_{\mathbb{L}}(\mathbb{K}) \subset e_{\mathbb{L}}(\mathbb{L})$ holds, Assumption 4 holds for the pairs $(\mathbb{L}, \mathbb{K})$, $(\mathbb{L}, \mathbb{K}_1)$, $(\mathbb{L}, \mathbb{K}_2)$. Then, we can use the test to determine the spanning relations inside each pair and thereby determine the

disconnectedness of the efficient set. If the convex polytope case is not generalized as in this paper, the determination of the spanning relation between the elements of the first pair is not feasible.

Assumption 4 implies that $e_{\mathbb{L}}(\mathbb{L})$ is finite. The $e_{\mathbb{L}}(\mathbb{K}) \subset e_{\mathbb{L}}(\mathbb{L})$ part implies $ch_{\mathbb{L}}(\mathbb{K}) \leq 1$. Indicative examples are the following. First, consider the trivial case where $\mathbb{K}$ is interior to $\mathbb{L}$. Then, it is obvious that $ch_{\mathbb{L}}(\mathbb{K}) = 0$. Second, consider the case where $\mathbb{L} = \mathbb{S}$, and $e_{\mathbb{L}}(\mathbb{K}) \neq \emptyset$ and Assumption 4 holds. Then, $ch_{\mathbb{L}}(\mathbb{K}) = \frac{\#e_{\mathbb{L}}(\mathbb{K})}{n} < 1$. Finally, suppose that $n = 3$, $\mathbb{L}$ is a line in the interior of the triangle, such that each boundary point of the line has a minimal distance from a unique triangle vertex and that both boundary points have the same distance from the remaining vertex. Furthermore, suppose that $\mathbb{K}$ is some half of that line. Then, $e_{\mathbb{L}}(\mathbb{L})$ consists of both the line boundary points, and $e_{\mathbb{L}}(\mathbb{K})$ consists of the boundary point that lies in the chosen half. If λ is an effective extreme point in either set, the cardinality of $c(\lambda)$ equals two. Moreover, $ch_{\mathbb{L}}(s, \lambda)$ equals 1 if s lies closer to λ than to the other boundary point of the line, and equals 2 in the other case. Hence, $ch_{\mathbb{L}}(\mathbb{K}) = \frac{1}{2}$.

Given our assumptions and the new combinatorial arguments not used previously in the literature, we prove in the Appendix (see Lemma 2) that the probability that the aforementioned bounding random variable attains the value zero is less than or equal to $ch_{\mathbb{L}}(\mathbb{K})$. Then, via the use of Corollary 1, we establish that, when ξ_∞ is non degenerate, the $1 - \alpha$ quantile is a continuity point for its cdf when $\alpha < 1 - ch_{\mathbb{L}}(\mathbb{K})$. Hence, we immediately obtain the following first order limit theory for the subsampling testing procedure described above via Theorem 3.5.1 in Politis, Romano and Wolf (1999).

Theorem 2. *Suppose that Assumptions 2, 3 and 4 hold. For the testing procedure*

described in Algorithm 3.2, we have that

1. *If* $\mathbf{H_0}$ *is true and* ξ_∞ *is constant, then,*

$$\lim_{T\to\infty} \mathbb{P}\left(\xi_T > q_{T,b_T}\left(1-\alpha\right)\right) = 0.$$

2. *If* $\mathbf{H_0}$ *is true,* ξ_∞ *is non-constant, and* $\alpha < 1 - ch_{\mathbb{L}}\left(\mathbb{K}\right)$*, then,*

$$\lim_{T\to\infty} \mathbb{P}\left(\xi_T > q_{T,b_T}\left(1-\alpha\right)\right) = \alpha.$$

3. *If* $\mathbf{H_a}$ *is true, then,*

$$\lim_{T\to\infty} \mathbb{P}\left(\xi_T > q_{T,b_T}\left(1-\alpha\right)\right) = 1.$$

When the distribution of ξ_∞ is degenerate, the procedure is asymptotically conservative even if the restriction $\alpha < 1 - ch_{\mathbb{L}}\left(\mathbb{K}\right)$ does not hold. This is reminiscent of the results in Linton et al. (2005) concerning testing procedures for superefficiency w.r.t. several stochastic dominance relations. The non-degeneracy of the aforementioned limit distribution is not easy to establish except for cases such as the one about bounded supports which was discussed above.

When the distribution of ξ_∞ is non-degenerate, the procedure is asymptotically exact if the restriction $\alpha < 1 - ch_{\mathbb{L}}\left(\mathbb{K}\right)$ holds. The restriction on the significance level is non-binding in usual applications. For example, when $\mathbb{L} = \mathbb{S}$ and $\mathbb{K}$ is a singleton, i.e., when the test is applied for super-efficiency, it implies at worst that $\alpha < 1/2$, something that is usually satisfied. The closer to binding the restriction becomes, the more extreme points of $\mathbb{L} = \mathbb{S}$ exist inside $\mathbb{K}$. An extreme case is when n is large, $\mathbb{K}$ is finite, and contains $n-1$ extreme points. In such a case, the result

leads to subsampling tests that tend to asymptotically favor the null hypothesis of spanning. We could handle that by breaking up $\mathbb{K}$ is "smaller pieces" and iterating the testing procedure w.r.t. them. For example, we can apply the procedure for any subset of $\mathbb{K}$ that contains m points, for m sufficiently small in order to obtain a meaningful significance level. If for some subset, we cannot reject spanning, we can infer that we cannot reject spanning for the initial $\mathbb{K}$, since supersets of spanning sets are spanning sets from Definition 2. It is also possible that the structure of the efficient set prohibits such a $\mathbb{K}$ to be a spanning set. We leave the study of such questions for future work. In any case, the testing procedure is consistent.

Under some assumptions, we can prove, using again among others the main result, that an analogous testing procedure based on block bootstrap is generally asymptotically conservative and consistent.

4 A Numerical Implementation and Bias Correction

We first describe a potential numerical implementation via the use of a testing procedure asymptotically equivalent to the one of Subsection 3.2, and obtained by finite approximations of the A_i, $i = 1, 2$, as well as applications of mixed integer and linear programming. For each T, let $A_i^{(T)}$ denote a finite subset of A_i for each i. Then consider the test statistic defined by

$$\xi_T^\star := \max_{i=1,2} \sup_{\lambda \in \mathbb{L}} \sup_{z \in A_i^{(T)}} \inf_{\kappa \in \mathbb{K}} \Delta_i \left(z, \lambda, \kappa, \sqrt{T} F_T \right),$$

and modify the algorithm of Subsection 3.2 by using $\xi_T^\star$ in place of ξ_T. Under the previous assumption framework if, as $T \to +\infty$, $A_i^{(T)}$ appropriately approximates A_i, the modified procedure has the same first order limit theory with the original one.

Theorem 3. *Suppose that Assumptions 2, 3 and 4 hold. If, as $T \to +\infty$, $A_i^{(T)}$ converges to some dense subset of A_i in Painleve-Kuratowski sense for all $i = 1, 2$, the results of Theorem 2 hold also for the modified procedure.*

Now, the integration by parts formula for Lebesgue-Stieljes integrals and the commutativity of suprema imply that

$$\xi_T^\star = \max_{i=1,2} \sup_{z \in A_i^{(T)}} \sup_{\lambda \in \mathbb{L}} \inf_{\kappa \in \mathbb{K}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} q_i \left(z, \lambda, \kappa, Y_t\right), \tag{5}$$

where the q_i are defined in 3. From the finiteness of $A_i^{(T)}$, $i = 1, 2$, the non trivial parts of the optimizations involved concern the $n_{i,T} := \sup_{\lambda \in \mathbb{L}} \inf_{\kappa \in \mathbb{K}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} q_i \left(z, \lambda, \kappa, Y_t\right)$. Furthermore,

$$n_{1,T} = \inf_{\kappa \in \mathbb{K}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(z - \kappa' Y\right)_+ - \inf_{\lambda \in \mathbb{L}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(z - \lambda' Y\right)_+ ,$$

and we can reduce each of the minimizations involved to the solution of linear programming problems.

There is a set of at most T values, say $\mathcal{R} = \{r_1, r_2, ..., r_T\}$, containing the optimal value of the variable z (see Scaillet and Topaloglou (2010) for the proof). Thus, we solve smaller problems $P(r)$, $r \in \mathcal{R}$, in which z is fixed to r. Now, each of the above minimization problems boils down to a linear problem. Without loss of generality, the first optimization problem is the following:

$$
\begin{aligned}
\min \quad & \sum_{t=1}^{T} W_t \\
\text{s.t.} \quad & W_t \geq r - \kappa' Y_t, \quad \forall t \in T \\
& e'\kappa = 1, \\
& \kappa \geq 0, \\
& W_t \geq 0, \quad \forall t \in T.
\end{aligned} \tag{6a}
$$

Furthermore, and via the results in the first Appendix of Arvanitis and Topaloglou (2017), we have that

$$
n_{2,T} = \sup_{\lambda \in \mathbb{L}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \max\left(\lambda' Y_t, z\right) - \sup_{\kappa \in \mathbb{K}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \max\left(\kappa' Y_t, z\right).
$$

Hence, we need to solve both optimization problems appearing above. We do so via representing them as MIP programs. Again, there is a set of T values, say $\mathcal{R}' = \{r'_1, r'_2, ..., r'_T\}$, containing the optimal value of the variable z (see Arvanitis and Topaloglou (2017) for the proof). Thus, we solve smaller problems $P(r)$, $r \in \mathcal{R}'$, in which z is fixed to r. Consider without loss of generality the first optimization problem:

$$
\begin{aligned}
\max_{\lambda \in \mathbb{L}} \quad & \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (X_t - c b_t) \\
\text{s.t.} \quad & X_t = \lambda' Y_t b_t + r(1 - b_t) \quad \forall\, t \in T, && (7) \\
& r - \lambda' Y_t + M b_t > 0 \quad \forall\, t \in T, && (8) \\
& \lambda' \mathbf{1} = 1, && (9) \\
& \lambda \geq 0, && (10) \\
& b_t \in \{0, 1\} \quad \forall\, t \in T. && (11)
\end{aligned}
$$

Hence, the computational cost of the implementation above consists of cardA_1 linear programming problems, cardA_2 mixed integer programming problems, and three trivial optimizations.

Secondly, and although the tests above have asymptotically correct size, it is expected 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 for n and T. To correct for small-sample bias and reduce the sensitivity to the choice of b_T, we follow Arvanitis et al. (2018). For a given significance level α, we compute the quantiles $q_{T,b_T}(1-\alpha)$ for a range of values for the subsample size b_T. We then estimate the intercept and slope of the following regression line using OLS regression analysis:

$$
q_{T,b_T}(1-\alpha) = \gamma_{0;T,1-\alpha} + \gamma_{1;T,1-\alpha}(b_T)^{-1} + \nu_{T;1-\alpha,b_T}. \tag{12}
$$

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

$$
q_T^{BC}(1-\alpha) := \hat{\gamma}_{0;T,1-\alpha} + \hat{\gamma}_{1;T,1-\alpha}(T)^{-1}. \tag{13}
$$

Since $q_{T,b_T}(1-\alpha)$ converges in probability to $q(\xi_\infty, 1-\alpha)$ and $(b_T)^{-1}$ converges to zero as $T \to 0$, $\hat{\gamma}_{0;T,1-\alpha}$ converges in probability to $q(\xi_\infty, 1-\alpha)$, and the asymptotic properties are not affected.

5 Monte Carlo Study

We now design Monte Carlo experiments to evaluate the size and power of our testing procedure in finite samples. We allow for conditional heteroskedasticity consistent with empirical findings on returns of financial data as observed in the empirical application below. The multivariate return process $(Y_t)_{t\in\mathbb{Z}}$ is a vector GARCH(1,1) process, which is transformed to accommodate both spanning (size) and non spanning cases (power) for K given assets. Such a process permits both temporal and cross sectional dependence between the random variables stacked in the vector process.

Suppose that $(z_t), t \in \mathbb{Z}$, are i.i.d. with mean zero, unit variance, and $\mathbb{E}\left[|z_t|^{2+\epsilon}\right] < \infty$, for some $\epsilon > 0$. We assume that the cdf of z_t is strictly increasing. Furthermore, we define the components of the return process for $i = 1, ..., K-1$ as

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

with $\mathbb{E}\left[a_i z_t^2 + \beta_i\right]^{1+\epsilon} < 1$, for some $\epsilon > 0$, and $\omega_i, a_i, \beta_i \in \mathbb{R}_{++}$, $\mu_i \in \mathbb{R}_+$. For asset $i = K$, we define

$$y_{K,t} = v_1 \left(z_t h_{K-1,t}^{1/2}\right)_+ + v_2 \left(z_t h_{K-1,t}^{1/2}\right)_- ,$$

with $v_1, v_2 \in \mathbb{R}$.

Let $\tau = (0, 0, ..., 1, 0)$, $\tau^\star = (0, 0, 0, ..., 1)$, and $\mathbb{L} := \left\{(\lambda, 0, 0,)^{Tr}, \tau, \tau^\star\right\}$, with $\lambda \in$

$\mathbb{R}_+^{K-2}$ and $1^{Tr}\lambda = 1$. Using this portfolio space, we obtain the following result on Markowitz-spanning.

Proposition 3. *If $\mu_i = 0$ for $i = 1, ..., K-1$, $|v_1| > \sqrt{\frac{\max\{\omega_i, a_i, \beta_i,\ i=1,...,K-1\}}{\min\{\omega_i, a_i, \beta_i,\ i=1,...,K-1\}}}$ and $|v_2| < \sqrt{\frac{\min\{\omega_i, a_i, \beta_i,\ i=1,...,K-1\}}{\max\{\omega_i, a_i, \beta_i,\ i=1,...,K-1\}}}$, then for $M \leq K-2$, the subset $\mathbb{K} := \left\{(\lambda, 0, 0)^{Tr}, \tau^\star\right\}$ with $\lambda \in \mathbb{R}_+^M$ and $1^{Tr}\lambda = 1$, Markowitz-spans $\mathbb{L}$, while $\mathbb{K} \setminus \{\tau^\star\}$ does not Markowitz-span $\mathbb{L}$.*

The statement of Proposition 3 extends Proposition 4 of Arvanitis and Topaloglou (2017) to allow for K assets with any subset of M spanning assets, as well as non Gaussian innovations. Its proof follows the same arguments as in Arvanitis and Topaloglou (2017), and is thus omitted. It depends on $\tau^\star$ being a Markowitz super-efficient portfolio w.r.t. the portfolio space. The design of Monte Carlo experiments in a dynamic setting is not easy for our testing procedure since we need to work with stationary distributions and different assets. The properties of those distributions required to show spanning and no spanning results are often difficult to characterize.[9]

We present our Monte Carlo results in Table 1. The number of replications to compute the empirical size and power is 1000 runs. We use either a combination of

[9] Another example is a process with different positive means and no serial dependence such that

$$\begin{aligned} y_{i,t} &= \mu_i + z_t, \qquad i = 1, .., K-1, \\ y_{K,t} &= v_1 \mu_{K-1} 1_{z_t > 0} + v_2 \mu_{K-1} 1_{z_t < 0} + z_t, \end{aligned}$$

with $\mu_i > 0$, $i = 1, ..., K-1$. Then, if $v_2 > \dfrac{\max\{\mu_i, i = 1, ..., K-1\}}{\min\{\mu_i, i = 1, ..., K-1\}}$ and if $0 < v_1 < \dfrac{\min\{\mu_i, i = 1, ..., K-1\}}{\max\{\mu_i, i = 1, ..., K-1\}}$, the spanning results stated in Proposition 3 also hold. We have checked in unreported simulation results that the spanning test behaves also well in such an example including the case of student innovations with infinite variance.

2 assets ($M = 2$) plus portfolios τ and $\tau^\star$ (Panel A for $K = 4$), or a combination of 10 assets ($M = 10$) plus portfolios τ and $\tau^\star$ (Panel B for $K = 12$). We do so to gauge the testing performance both in a small and a larger number of assets to accommodate the empirical setting where we investigate spanning with up to 10 base assets. To meet the conditions of Proposition 3, we set the parameters of the multivariate GARCH process as $\mu_i = 0$, for $i = 1, ...K - 1$, while we choose $(a_i) = (0.4, 0.45, 0.5)$, $(\beta_i) = (0.5, 0.45, 0.4)$, $(\omega_i) = (0.5, 0.5, 0.5)$, for $i = 1, 2, 3$ (Panel A), and similarly $(a_i) = (0.4, 0.41..., 0.5)$, $(\beta_i) = (0.5, 0.49, ..., 0.4)$, $(\omega_i) = (0.5, ..., 0.5)$, for $i = 1, .., 11$ (Panel B). We set $v_1 = 1.5$ and $v_2 = 0.5$, so that $|v_1| > \sqrt{\frac{\max\{\omega_i, a_i, \beta_i,\ i=1,...,K-1\}}{\min\{\omega_i, a_i, \beta_i,\ i=1,...,K-1\}}}$ and $|v_2| < \sqrt{\frac{\min\{\omega_i, a_i, \beta_i,\ i=1,...,K-1\}}{\max\{\omega_i, a_i, \beta_i,\ i=1,...,K-1\}}}$. We use innovations generated by a Student distribution with 5 degrees of freedom.[10]

We use three different sample sizes. For $T = 300$, we get the subsampling distribution of the test statistic for subsample sizes $b_T \in \{50, 100, 150, 200\}$. We set $b_T \in \{100, 200, 300, 400\}$ for $T = 500$, and $b_T \in \{120, 240, 360, 480\}$ for $T = 1000$. We present the results using the original subsampling critical values (without bias correction) as well as the ones obtained using the bias correction method. The comparison shows that the bias correction improves a lot the inference in finite samples. The bias correction method eliminates the size distortion and delivers excellent properties under the alternative hypothesis with empirical powers above 90% for a nominal size of 5%.

In our simulations, the computational time is only marginally increasing with the number of assets, and is mainly increasing with the number of observations. For example, we have roughly 5 minutes for $T = 300$ and the double for $T = 500$ per

[10] Unreported simulation results for Gaussian innovations are similar.

run. Therefore we believe that the procedure can scale up to a couple of hundred assets.

Panel A: $M = 2, K = 4$						
	Without bias correction			With bias correction		
	T=300	T=500	T=1000	T=300	T=500	T=1000
Size	12.6%	10.7%	8.2%	4.4%	3.6%	4.8%
Power	85.1%	87.4%	91.7%	93.7%	92.5%	96.2%
Panel B: $M = 10, K = 12$						
	Without bias correction			With bias correction		
	T=300	T=500	T=1000	T=300	T=500	T=1000
Size	12.4%	10.5%	9.8%	5.6%	5.1%	4.4%
Power	85.7%	87.5%	89.1%	92.1%	93.2%	95.8%

Table 1: Monte Carlo Results. Entries report the empirical size and empirical power based on 1000 replications, $T = 300, 500, 1000$, and a nominal size $\alpha = 5\%$. Panel A reports the rejection probabilities for $M = 2$ and $K = 4$, while Panel B reports the rejection probabilities for $M = 10$ and $K = 12$. In both panels, we use a multivariate GARCH process to generate returns and compute the rejection probabilities without and with the bias correction method for the subsampling critical values.

6 Empirical Applications

In this application, $\mathbb{L}$ consists of all convex combinations of the market portfolio, the T-bill, and a set of 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.

Thanks to our spanning testing procedure, we want to check whether the two-fund separation theorem holds: can all MSD investors combine the T-bill and the market portfolio to span the whole set of their efficient portfolios?

If not, there is indication that active management for MSD investors according to their preferences could outperform any combination of the market portfolio and the riskless asset. This is studied in our second empirical application.

We use as base assets either the Fama and French (FF) size and book to market portfolios, a set of momentum portfolios, a set of industry portfolios, or a set of beta or size decile portfolios as described below, along with the market portfolio and the T-bill. If the number of base assets equals n, $\mathbb{L}$ is essentially the union of the relevant $n-2$ subsimplex of the standard $n-1$ simplex with $\{(0, \cdots, 1)\}$, where the latter signifies the market portfolio. The base assets, aside the market portfolio and the T-bill are the following portfolios:

- **The 6 FF benchmark portfolios**: They are constructed at the end of each June, and correspond to the intersections of 2 portfolios formed on size (market equity, ME) and 3 portfolios formed on the ratio of book equity to market equity (BE/ME).

- **The 10 momentum portfolios**: They are constructed monthly using NYSE

prior (2-12) return decile breakpoints. The portfolios include NYSE, AMEX, and NASDAQ stocks with prior return data. To be included in a portfolio for month t (formed at the end of month $t-1$), a stock must have a price for the end of month $t-13$ and a good return for $t-2$.

- **The 10 industry portfolios**: They are constructed by assigning each NYSE, AMEX, and NASDAQ stock to an industry portfolio at the end of June of year t based on its four-digit SIC code at that time. The industries are defined with the goal of having a manageable number of distinct industries that cover all NYSE, AMEX, and NASDAQ stocks.

- **The 10 size decile portfolios**: We use a standard set of ten active US stock portfolios that are formed, and annually rebalanced, based on individual stock market capitalization of equity (ME or size), each representing a decile of the cross-section of NYSE, AMEX and NASDAQ stocks in a given year.

- **The 10 beta decile portfolios**: We use a set of ten active US stock portfolios that are formed, and annually rebalanced, based on individual stock beta, each representing a decile of the cross-section of NYSE, AMEX and NASDAQ stocks in a given year.

For each dataset, we use data on monthly returns (month-end to month-end) from January 1930 to December 2016 (1044 monthly observations) obtained from the data library on the homepage[11] of Kenneth French. The test portfolio is the Fama and French market portfolio, which is the value-weighted average of all non-financial

[11] http://mba.turc.dartmouth.edu/pages/faculty/ken.french

common stocks listed on NYSE, AMEX, and Nasdaq, and covered by CRSP and COMPUSTAT.

The portfolios used as base assets are of particular interest, because a wealth of empirical research, starting with Banz (1981), Basu (1983), and Fama and French (1993, 1997), suggests that the historical return spread between small value stocks and small growth stocks defies rational explanations based on investment risk. Moreover, book-to-market based sorts are the basis for the factor model examined in Fama and French (1993). Additionally, academics and practitioners show strong interest in momentum portfolios. Empirical evidence indicates that common stocks exhibit high returns on a period of 3-12 months will overperform on subsequent periods. This momentum phenomenon is an important challenge for the concept of market efficiency. Finally, industry sorted portfolios have posed a particularly challenging feature from the perspective of systematic risk measurement (see Fama and French (1997)). Beta-sorted portfolios have been used extensively to test the Sharpe-Lintner Mossin Capital Asset Pricing Model (CAPM) (see Black, Jensen, and Scholes (1972), Blume and Friend (1973), Fama and MacBeth (1973), Reinganum (1981), and Fama and French (1992), among others). Equity portfolios have also been at the center of the empirical literature in the stochastic dominance framework, see for example Post (2003), Kuosmanen (2004), Post and Levy (2005), Scaillet and Topaloglou (2010), Post and Kopa (2013), Gonzalo and Olmo (2014), among others.

To focus on the role of preferences and beliefs, we adhere to the assumptions of a single-period, portfolio-oriented model of a competitive capital market. The model-free nature of SD tests seems an advantage in this application area, because financial economists disagree about the relevant shape of utility functions of investors and the

probability distribution of stock returns.

6.1 Results of the MSD Spanning Test

Arvanitis and Topaloglou (2017) report evidence against the market portfolio being MSD efficient for data up to December 2012. We corroborate their findings in unreported results for our whole period up to December 2016 as well as two sub-periods, the first one from January 1930 to June 1975, a total of 522 monthly observations, and the second one from July 1975 to December 2016, 522 monthly observations. Thus, we find evidence that passive investment is suboptimal for investors with MSD preferences. Equity management, instead of a standard buy-and-hold strategy on the market portfolio, seems more appealing for investors with reverse S-shaped utility functions. The MSD inefficiency of the market portfolio is not affected by transformations that are increasing and convex over gains and increasing and concave over losses, i.e., reverse S-shaped transformations.

Since the market is MSD inefficient, our next research hypothesis is whether two-fund separation holds, i.e., whether all MSD investors can satisfy themselves with combining the T-bill and the market portfolio only. The test of MSD efficiency for a given portfolio developped by Arvanitis and Topaloglou (2017) cannot answer that question since their approach is limited to the simple case of a spanning test for $\mathbb{K}$ being a singleton, and not any linear combination of two assets.

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 MSD spanning test can analyze two-fund separation without assuming a particular form for the return distribution or utility

functions.

We get the subsampling distribution of the test statistic for subsample size $b_T \in [120, 240, 360, 480]$. Using OLS regression on the empirical quantiles $q_{T,b_T}(1-\alpha)$, for significance level $\alpha = 0.05$, we get the estimate q_T for the critical value. We reject the MSD spanning if the test statistic ξ_T is higher than the regression estimate q_T. In all the considered cases, $\mathbb{L} = \mathbb{S}$, and $\alpha < \frac{3}{4} \leq 1 - ch_{\mathbb{L}}(\mathbb{K})$ holds. Hence, if our assumption framework is valid, we expect that asymptotic exactness holds. We find that:

- **The 6 FF benchmark portfolios**: The regression estimate $q_T = 15.74$ is lower than the value of the test statistic $\xi_T = 26.78$.
- **The 10 momentum portfolios**: The regression estimate $q_T = 19.42$ is lower than the value of the test statistic $\xi_T = 41.55$.
- **The 10 industry portfolios**: The regression estimate $q_T = 22.46$ is lower than the value of the test statistic $\xi_T = 31.74$.
- **The 10 size decile portfolios**: The regression estimate $q_T = 19.62$ is lower than the value of the test statistic $\xi_T = 32.34$.
- **The 10 beta decile portfolios**: The regression estimate $q_T = 31.48$ is lower than the value of the test statistic $\xi_T = 44.76$.

The results suggest the rejection of MSD spanning and thus of the two-fund separation theorem for MSD investors. We get similar findings (unreported results) for the two subperiods 01/1930-06/1975 and 07/1975-12/ 2016.

As a final step in this analysis, we test for two-fund separation using the Mean-Variance criterion rather than the MSD criterion. We use the same methodology as for the above prospect spanning test, but we restrict the utility functions to take a quadratic shape. We solve the embedded expected-utility optimization problems (for every given quadratic utility function) using quadratic programming. In contrast to MSD spanning, we cannot reject the Mean-Variance spanning at conventional significant levels.

The combined results of the market MSD efficiency and market MSD spanning tests suggest that combining the T-bill and market portfolio is not optimal for some MSD investors. Investors with reverse S-shaped utility functions are investors that could outperform the market by staying away from a buy-and-hold strategy on the market. Active investors often take concentrated positions in assets with high upside potential or follow dynamic strategies like momentum. They can also prefer looking at defensive strategies. That can produce opportunities with positively skewed returns, or at least less negatively skewed, which are attractive for MSD investors.

6.2 Performance Summary of the MSD portfolios

The rejection of the spanning hypothesis implies that there exists at least one portfolio in $\mathbb{L}$ which is weakly prefered to every portfolio in $\mathbb{K}$ by at least one reverse S-shaped utility function (see Definition 2). Such a portfolio is by construction efficient w.r.t. $\mathbb{K}$ (see Definition 2.1 in Linton et al. (2014) for the SSD case which can be easily generalized to our MSD case). The empirical version of such a portfolio is the optimal portfolio λ that maximizes ξ_T for the particular sample value. In what follows, and given this characterization, we analyze the performance of such

empirically optimal MSD portfolios through time, compared to the performance of the market portfolio (buy-and-hold strategy).

We resort to backtesting experiments on a rolling window basis. The rolling horizon computations cover the 642-month period from 07/1963 to 12/2016. At each month, we use the data from the previous 30 years (360 monthly observations) to calibrate the procedure. We solve the resulting optimization model for the MSD spanning test and record the optimal portfolio made of the base assets as well as the market portfolio and the T-bill. We determine the realized return of the chosen MSD optimal portfolio from the actual returns of the asset weight allocation picked by the optimizer for that month. Then, we repeat the same procedure for the next one-month rolling window and compute the ex-post realized returns for the period from 07/1963 to 12/2016. Therefore, the MSD optimal portfolios are outcomes of the testing procedure based on an unconditional distribution updated for each rolling window and performance is realized out of the optimization sample (no look-ahead bias).

Let us first compute the cumulative performance of the MSD optimal portfolios as well as the market portfolio for the entire sample period from July 1963 to December 2016 based on the optimal portfolio weights obtained for each one-month rolling window. The value for the MSD optimal portfolios is 426 times higher at the end of the holding period compared to the initial value, while the market portfolio is only 13.9 times higher. Hence, the relative performance of MSD type investors is 30 times higher than the performance of the market in the evaluated period. Such an increase of 3000% is significant at any significance level (unreported results).

To get further insights of the differences between two investment strategies, we

report the first four moments of the realized returns and the Value-at-Risk in Table 2. We further compute a number of commonly used performance measures: the Sharpe ratio, the downside Sharpe ratio, the return loss and the opportunity cost.

The downside Sharpe ratio based on the semi-variance, also known as the Sortino ratio, is considered to be a more appropriate measure of performance than the typical Sharpe ratio given the asymmetric return distribution of the assets (see for example Ziemba (2005))

To account for transaction costs, we use the proposal of DeMiguel et al. (2009). This indicates the way that the proportional transaction costs, generated by the portfolio turnover, affect the portfolio returns. Let trc be the proportional transaction cost, and $R_{P,t+1}$ the realized return of portfolio P at time $t+1$. The change in the net of transaction cost wealth NW_P of portfolio P through time is,

$$NW_{P,t+1} = NW_{P,t}(1 + R_{P,t+1})[1 - trc \times \sum_{i=1}^{N}(|w_{P,i,t+1} - w_{P,i,t}|). \tag{14}$$

The portfolio return, net of transaction costs is defined as

$$RTC_{P,t+1} = \frac{NW_{P,t+1}}{NW_{P,t}} - 1. \tag{15}$$

Let μ_M and μ_{MSD} be the out-of-sample mean of (15) for the market portfolio and the MSD optimal portfolios, and σ_M and σ_{MSD} be the corresponding standard deviations. Then, the return-loss measure is,

$$R_{Loss} = \frac{\mu_{MSD}}{\sigma_{MSD}} \times \sigma_M - \mu_M, \tag{16}$$

i.e., the additional return needed so that the market performs equally well with the MSD optimal portfolios. We follow the literature and use 35 bps for the transaction costs of stocks and bonds.

Finally, the opportunity cost presented in Simaan (2013) gauges the economic significance of the performance difference between two portfolios. Let R_{MSD} and R_M be the realized returns of the MSD optimal portfolios and the market portfolio, respectively. Then, the opportunity cost θ is defined as the return that needs to be added to (or subtracted from) the market return R_M, so that the investor is indifferent (in utility terms) between the strategies imposed by the two different investment opportunity sets, i.e.,

$$E[U(1 + R_M + \theta)] = E[U(1 + R_{MSD})]. \tag{17}$$

The opportunity cost is not the certainty equivalent return R_C defined by $U(1 + R_C) = E[U(1 + R_{MSD})]$.

A positive (negative) opportunity cost implies that the investor is better (worse) off if the investment opportunity set allows for MSD type investing. The opportunity cost takes into account the entire probability density function of asset returns and hence it is suitable to evaluate strategies even when the distribution is not normal. For the calculation of the opportunity cost, we use the following utility function which satisfies the curvature of Markowitz theory (reverse-S-shaped):

$$U(R) = \begin{cases} R^a, & if\ R \geq 0, \\ -c(-R)^b, & if\ R < 0, \end{cases} \tag{18}$$

where c is the coefficient of loss aversion (usually $c = 2.25$) and $a, b > 1$. We use several values of a, b in Table 2 to drive the curvature of the utility functions.

Table 2 reports the performance and risk measures for the MSD optimal portfolios and the market portfolio. These measures allow us to better figure out the differences between the market portfolio and the MSD strategy. The mean is higher for the MSD

optimal portfolio and the variance is lower, which results in a higher Sharpe ratio. The skewness is less negative as expected for a portfolio built for investors with preferences towards risk that are associated with risk aversion for losses and risk loving for gains. The kurtosis and VaR are lower as expected when investors want to mitigate the impact of large losses. The MSD portfolio targets and achieves a transfer of probability mass from the left to the right tail of the return distribution when compared to the market portfolio. The opportunity cost is above 70 bps and increases with the curvatures of the gain and loss parts of the utility function.

Table 3 reports the descriptive statistics regarding the weight allocation of the MSD optimal portfolios. They load mainly on big size FF portfolios (FF portfolios), several momentum portfolios (momentum portfolios), telecommunications, health, energy and utilities (industry portfolios), small caps (size portfolios), and low and medium beta (beta sorted portfolios), in addition to the market portfolio and the T-Bill.

We also investigate which factors explain the returns of the active investors with MSD preferences. To do so, we use the four-factor model of Carhart (1997) which adds momentum in the three-factor model of Fama and French (1992, 1993), as well as the Fama and French five-factor model (2015). Our empirical test examines whether these models explain the returns on MSD portfolios that dominate any combination of the market and the riskless asset, namely whether standard factors used in the empirical asset pricing literature are potential drivers of returns of MSD optimal portfolios.

First, we consider the following linear regression (Carhart four-factor model):

$$R_{it} - R_{Ft} = a_i + b_i(R_{Mt} - R_{Ft}) + s_i SMB_t + h_i HML_t + r_i MOM_t + e_{it},$$

where R_{it} is the return of the MSD optimal portfolio at period t, R_{Ft} is the riskless rate, R_{Mt} is the return on the value-weight (VW) market portfolio, SMB_t is the return on a diversified portfolio of small stocks minus the return on a diversified portfolio of big stocks, HML_t is the difference between the returns on diversified portfolios of high and low BE/ME stocks, MOM_t is the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios, and e_{it} is a zero-mean residual. If the exposures b_i, s_i, h_i, and r_i to the market, size, value, and momentum factors capture all variation in expected returns, the intercept a_i is zero.

Table 4 reports the coefficient estimates of the four factors, as well as their respective t-statistics and p-values. The results indicate that apart from the momentum (MOM), all the other three factors explain part of the performance of the optimal MSD portfolios. The intercept is not zero, which indicates that perhaps other factors drive the performance of the MSD portfolios as well.

We additionally consider the following linear regression (five-factor model):

$$R_{it} - R_{Ft} = a_i + b_i(R_{Mt} - R_{Ft}) + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + e_{it},$$

where R_{it} is the return of the MSD optimal portfolio at period t, R_{Ft}, R_{Mt}, SMB_t and HML_t as before, RMW_t is the difference between the returns on diversified portfolios of stocks with robust and weak profitability, CMA_t is the difference between the returns on diversified portfolios of the stocks of low and high investment firms, which are called conservative and aggressive, and e_{it} is a zero-mean residual. If the exposures b_i, s_i, h_i, r_i, and c_i to the market, size, value, profitability and investment factors capture all variation in expected returns, the intercept a_i is zero.

Table 5 reports the coefficient estimates of the five-factor model, as well as their respective t-statistics and p-values. The results indicate that, apart from the profitability (RMW), all the other four factors explain part of the performance of the optimal MSD portfolios. The intercept clearly not being zero indicates that other factors possibly drive the performance of the MSD portfolios as well.

In both factor models, we observe that the beta market is slightly smaller than one (defensive) for the MSD portfolios as expected. The negative sign for the SMB factor loading and positive sign for the HML factor loading correspond to an additional defensive tilt. Defensive strategies overweight large value stocks and underweight small growth stocks (see Novy-Marx (2016)).

7 Conclusions

We have derived properties of the cdf of a random variable defined by recursive optimizations applied on a continuous stochastic process w.r.t. possibly dependent parameter spaces. Those properties extend previous results and can be useful for the derivation of the limit theory of tests for stochastic spanning w.r.t. stochastic dominance relations.

As a theoretical application, we have defined the concept of spanning, constructed an analogous test based on subsampling, and derived the first-order limit theory and a numerical implementation for the case of the MSD relation.

We have used the non-parametric test in an empirical application, inspired by Arvanitis and Topaloglou (2017), who show that the market portfolio is not MSD efficient. The spanning test enables us to explore whether MSD equity managers

could outperform the market portfolio. First, we test whether the market portfolio is MSD efficient, and then whether the two-fund separation theorem holds for investors with MSD preferences. We use as base assets either the FF size and book to market portfolios, a set of momentum portfolios, a set of industry portfolios, or a set of beta or size decile portfolios. Empirical results indicate that the market portfolio is not MSD efficient, and the two-fund separation theorem does not hold for MSD investors. Thus, the combination of the market and the riskless asset do not span the portfolios created according to the MSD criterion. Hence, there exist MSD investors that could benefit from investment opportunities that involve assets beyond portfolios constructed solely by the market portfolio and the safe asset. We verify this by showing that equity managers with MSD preferences could generate portfolios that yield 30 times higher cumulative return than the market over the last 50 years. The return distribution of the MSD optimal portfolio is less negatively skewed, less leptokurtic, and thiner left-tailed, when compared to the market portfolio. Finally, using the four-factor model of Carhart (1997) and the five-factor model of Fama and French (2015), we investigate which factors explain these returns. We find that a defensive tilt explains part of the performance of the optimal MSD portfolios, while momentum and profitability do not.

The derivations and methodology used above can also be explored for other forms of stochastic dominance relations, such as the first- or the third-order, or Prospect stochastic dominance. We leave such issues for future research.

Acknowledgements: We would like to thank the Editor, Co-Editors and the two referees for constructive criticism and numerous suggestions which have led to

substantial improvements over the previous version. We thank the participants at the SFI research days 2018 for helpful comments.

References

Arvanitis, S., Hallam, M. S., Post, T. and N. Topaloglou. 2018. Stochastic spanning. Forthcoming in the Journal of Business and Economic Statistics (http://dx.doi.org/10.1080/07350015.2017.1391099).

Arvanitis, S., and N. Topaloglou. 2017. Testing for prospect and Markowitz stochastic dominance efficiency. Journal of Econometrics 198(2), 253-270.

Banz, Rolf W. 1981. The relationship between return and market value of common stocks. Journal of Financial Economics 9(1), 3-18.

Baucells, M., and F.H. Heukamp. 2006. Stochastic dominance and cumulative prospect theory. Management Science 52, 1409-1423.

Black, F., Jensen, M. and M. Scholes. 1972. The capital asset pricing model: some empirical tests, in M.C. Jensen (ed.). Studies in the Theory of Capital Markets, Praeger: New York, 79-124.

M.E. Blume and I. Friend. 1973. A new look at the Capital Asset Pricing Model. Journal of Finance 28(1), 19-34.

Carhart, M. 1997. On persistence in Mutual Fund Performance. Journal of Finance 52(1), 57-82.

Cass, D., and J. E. Stiglitz. 1970. The structure of investor preferences and asset returns, and separability in portfolio allocation: A contribution to the pure theory of mutual funds. Journal of Economic Theory, 2(2), 122-160.

Cortissoz, J. 2007. On the Skorokhod representation theorem. Proceedings of the American Mathematical Society 135(12), 3995-4007.

DeMiguel, V., L. Garlappi and R. Uppal, 2009, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?. Review of Financial Studies 22, 1915-1953.

Edwards, K.D. 1996. Prospect theory: A literature review, International Review of Financial Analysis 5, 18-38

Fama, E. and K. French. 1992. The Cross-Section of Expected Stock Returns. Journal of Finance 47(2), 427-465.

Fama, E. and K. French. 1993. Common Risk Factors in the Returns on Stocks and Bonds. Journal of Financial Economics 33, 3-56.

Fama, E. and K. French. 1997. Industry costs of equity. Journal of Financial Economics 43, 153-193.

Fama, E. and K. French. 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1-22.

Fama, E.F. and J.D. MacBeth. 1973. Risk, return and equilibrium: empirical tests. The Journal of Political Economy 81, 607-636.

Francq, C., and J. M. Zakoian. 2011. GARCH models: structure, statistical inference and financial applications. John Wiley & Sons.

Friedman, M., and L. J. Savage. 1948. The utility analysis of choices involving risk. Journal of Political Economy 56, 279-304.

Gonzalo, J. and J. Olmo. 2014. Conditional Stochastic Ddominance Tests in Dynamic Settings. International Economic Review 55(3), 819-838.

Guggenberger, P., Hahn, J., & Kim, K. (2008). Specification testing under moment inequalities. Economics Letters, 99(2), 375-378.

Hadar, J. and W.R. Russell. 1969. Rules for ordering uncertain prospects. American Economic Review 59, 2-34.

Hanoch, G., and H. Levy. 1969. The efficiency analysis of choices involving risk. Review of Economic Studies 36, 335-346.

Horvath, L. and Kokoszka, P. and R. Zitikis. 2006. Testing for Stochastic Dominance Using the Weighted McFadden-type Statistic. Journal of Econometrics 133, 191-205.

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

Knight, K. 1999. Epi-convergence in distribution and stochastic equi-semicontinuity. Working Paper, Department of Statistics, University of Toronto.

Kroll, Y., and H. Levy. 1980. Stochastic Dominance Criteria: A Review and Some New Evidence, in Research in Finance, Vol. II, Greenwich: JAI Press, pp. 263-277

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

Levy, H. 1992. Stochastic Dominance and Expected Utility: Survey and Analysis. Management Science 38, 555-593.

Levy, H. 2015. Stochastic dominance: Investment decision making under uncertainty. Springer.

M. Levy and H. Levy. 2002. Prospect Theory: Much Ado about Nothing?. Management Science 48, 1334-1349.

Levy, H., M. Levy. 2004. Prospect theory and mean-variance analysis. Review of Financial Studies 17(4), 1015-1041.

Lifshits, M. A. 1983. On the absolute continuity of distributions of functionals of random processes. Theory of Probability and Its Applications 27(3), 600-607.

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

Linton, O., Post, T., and Whang, Y. J. 2014. Testing for the stochastic dominance efficiency of a given portfolio. The Econometrics Journal 17(2), 59-74.

McFadden, D. 1989. Testing for Stochastic Dominance, in Studies in the Economics of Uncertainty, eds. T. Fomby and T. Seo, New York: Springer- Verlag, pp. 113–134.

Molchanov, I. 2006. Theory of random sets. Springer Science and Business Media.

Mosler, K., and M. Scarsini. 1993. Stochastic Orders and Applications, a Classified Bibliography, Berlin: Springer-Verlag.

Narici, L., and E. Beckenstein. 2010. Topological vector spaces. CRC Press.

Novy-Marx, R. 2016. Understanding defensive equity. Working Paper, Simon Graduate School of Business, University of Rochester and NBER.

Nualart, D. 2006. The Malliavin calculus and related topics. Berlin: Springer.

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

Post, T. 2003. Empirical Tests for Stochastic Dominance Efficiency. The Journal of Finance 58: 1905-1931.

Post, T., and M. Kopa. 2013. General linear formulations of stochastic dominance criteria. European Journal of Operational Research 230(2), 321-332.

Post, T., and H. Levy. 2005. Does risk seeking drive stock prices? A stochastic dominance analysis of aggregate investor preferences and beliefs. Review of Financial Studies 18(3), 925-953.

M.R. Reinganum. 1981. A new empirical perspective on the CAPM. Journal of Financial and Quantitative Analysis 16(4), 439-462.

Rio, E. 2013. Inequalities and limit theorems for weakly dependent sequences. 3‘eme cycle. pp.170. <cel-00867106>.

Rothschild, M. and J.E. Stiglitz. 1970. Increasing Risk: I. A definition. Journal of Economic Theory 2(3), 225-243.

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

Simaan, Y., 1993, Portfolio selection and asset pricing-three-parameter framework. Management Science 39, 568-577.

Sidak, Z., P. K. Sen, and J. Hajek. 1999. Theory of rank tests. Academic Press.

Tobin, J. 1958. Liquidity Preference as Behavior Towards Risk. Review of Economic Studies 25, 65-86.

van der Vaart, A. W., and J. A. Wellner. 1996. Weak Convergence. Springer New York.

Ziemba, W., 2005, The symmetric downside risk sharpe ratio. Journal of Portfolio Management 32, 108-122.

8 Appendix

8.1 Proofs of Main Results

Proof of Theorem 1. First, we know that $\xi \in \mathbb{D}^{1,2}$, from similar arguments to the ones in the proof of Proposition 2.1.10 of Nualart (2006). Precisely, consider a countable dense subset of Λ , say Λ_∞ as well as $\xi_n := \text{oper} X_\lambda$, where opt_i is considered w.r.t. $\Lambda^\star_{i,n}(\lambda_{i-1}) = \{\text{the first } n \text{ elements of } \Lambda^\star_i(\lambda_{i-1}) \cap \text{pr}_i \Lambda_\infty\}$ and $\lambda_{i-1} \in \Lambda^\star_{i-1,n}$

when $i > 1$. The function oper : $C(\Lambda, \mathbb{R}) \to \mathbb{R}$ is Lipschitz, and from Proposition 1.2.4 of Nualart (2006), we get $\eta_n \in \mathbb{D}^{1,2}$. Furthermore, from Assumption 1.1, $\xi_n \to \xi$ in $L^2(\Omega)$, and therefore the preliminary result follows if $(D\xi_n)_{n\in\mathbb{N}}$ is $L^2(\Omega)$ bounded. Define

$$A_n = \{\omega \in \Omega : \xi_n = X_{\lambda_n}, \xi_n \neq X_{\lambda_k}, \forall k < n\}.$$

Using the local property of D, we have that $D\xi_n = \sum_{n\in\mathbb{N}} 1_{A_n} DX_{\lambda_n}$, and thereby $\mathbb{E}[\|D\xi_n\|_H^2] < +\infty$ from Assumption 1.2. Then Assumption 1.3 as well as Proposition 2.1.7 of Nualart (2006) imply the first part of the theorem. For the following, assume first that T is empty. Then the result will follow from a series of arguments almost identical to the ones in the proof of Proposition 2.1.11 of Nualart (2006). Specifically, consider the set

$$G = \{\omega \in \Omega : \text{ there exists } \lambda \in \Lambda \text{ such that } DX_\lambda \neq D\xi \text{ and } X_\lambda = \xi\},$$

and using Λ_∞ above H_∞ a countable dense subset of the unit ball of H, and $B_r(\lambda)$ the ball in Λ with center λ and radius $r > 0$ we have that $G \subseteq \cup_{\lambda\in\Lambda_\infty, r\in\mathbb{Q}_{++}, k\in\mathbb{N}_0, h\in H_\infty} G_{\lambda,r,k,h}$ i.e., a countable union, where

$$G_{\lambda,r,k,h} := \left\{\omega \in \Omega : \langle DX_{\lambda'} - D\eta, h\rangle > \frac{1}{k} \text{ for all } \lambda' \in B_r(\lambda)\right\} \cap \{\text{oper}X_{\lambda'} = \xi\}.$$

For some λ, r, k, h as above, define $\xi' = \text{oper}X_{\lambda'}$, where now opt_i is considered w.r.t. $\Lambda_i^\star(\lambda_{i-1}) \cap \text{pr}_1 B_r(\lambda)$ choose a countable dense subset of $B_r(\lambda)$, say $B_r^\infty(\lambda)$ and using

$$\Lambda_{i,n}^\infty(\lambda_{i-1}) = \{\text{the first } n \text{ elements of } \Lambda_i^\star(\lambda_{i-1}) \cap \text{pr}_i B_r^\infty(\lambda)\},$$

define $\xi_n^{'} = \text{oper}X_\lambda$ analogously. We have that as $n \to \infty$ $\xi_n^{'} \to \xi'$ in $L^2(\Omega)$ norm due to Assumption 1.1. From Lemma 1.2.3 of Nualart (2006) and Assumption 1.2

we also have that $D\xi_n^{'} \to D\xi'$ in the weak topology of $L^2(\Omega, H)$. Using again the local property argument as above, we have that for any $\omega \in G_{\lambda,r,k,h}$, $D\xi_n^{'} = DX_{\lambda'}$, for some $\lambda' \in B_r^{\infty}(\lambda)$. But, for such ω, we have that $\left\langle D\xi_n^{'} - D\xi^{'}, h \right\rangle > \frac{1}{k}$ for all n. This directly implies that $\mathbb{P}(G_{\lambda,r,k,h}) = 0$ which, due to countability, implies that $\mathbb{P}(G) = 0$. Then the result follows from Theorem 2.1.3 of Nualart (2006). Now, suppose that $\tau \in \mathcal{T}$, and consider

$$\mathbb{P}(\xi = \tau) = \mathbb{P}(\{\xi = \tau\} \cap \Omega_\tau) + \mathbb{P}(\{\xi = \tau\} \cap \Omega_\tau^c)$$

If, for some $\tau \in \mathcal{T}$, $\mathbb{P}(\Omega_\tau^c) > 0$, we get

$$\mathbb{P}(\{\xi = \tau\} \cap \Omega_\tau^c) = \mathbb{P}(\xi = \tau / \Omega_\tau^c)\, \mathbb{P}(\Omega_\tau^c),$$

and we can consider the process $X^\star := X\left|_{\Omega - \cup_{\tau \in \mathcal{T}} \Omega_\tau^c}\right.$ that obviously satisfies Assumption 1 with $\mathcal{T}^\star = \emptyset$ along with the obvious change of notation. Hence, $\xi^\star$ has an absolutely continuous law something that implies that $\mathbb{P}(\xi = \tau / \Omega_\tau^c) = \mathbb{P}(\xi^\star = \tau) = 0$. If $\mathbb{P}(\Omega_\tau^c) = 0$ trivially $\mathbb{P}(\{\xi = \tau\} \cap \Omega_\tau^c) = 0$ establishing that $\mathbb{P}(\xi = \tau) = \mathbb{P}(\{\xi = \tau\} \cap \Omega_\tau^c)$ in any case. Now, suppose that τ_1, τ_2 are successive elements of $\mathcal{T}$ and consider $\Omega_{\tau_1,\tau_2} = \{\omega \in \Omega : \xi \in (\tau_1, \tau_2)\}$. The previous imply that $\mathbb{P}(\Omega_{\tau_1,\tau_2}) > 0$, hence the process $X_\star := X\left|_{\Omega_{\tau_1,\tau_2}}\right.$ satisfies Assumption 1 with $\mathcal{T}_\star = \emptyset$, and thereby $\xi_\star$ has an absolutely continuous law. The other cases follow analogously when the intersections apperaring in the theorem are non empty. When empty the results are trivial. □

Proof of Corollary 1. It follows simply by Theorem 1 since the relation between ξ and η implies that $\operatorname{supp}(\xi)$ is the closure of $(c, +\infty)$ and also that $\mathbb{P}(\xi = c) \leq \mathbb{P}(\eta = c)$. □

Proof of Proposition 1. ($\Leftarrow$) If $\mathbb{K} \succcurlyeq_M \mathbb{L}$, for any λ, there exists some κ such that $\sup_{z\leq 0} \Delta_1(z,\lambda,\kappa,F) \leq 0$ and $\sup_{z>0} \Delta_2(z,\lambda,\kappa,F) \leq 0$. This implies that

$$\max_{i=1,2} \sup_{z\in A_i} \inf_{\kappa\in\mathbb{K}} \Delta_i(z,\lambda,\kappa,F) \leq 0. \tag{19}$$

Since $\mathbb{K}$ is closed, hence compact, and F has a finite first moment, the Dominated Convergence Theorem implies that $\mathcal{J}(-\infty,0,\kappa,F)$ is continuous w.r.t. κ. This along with the compactness of $\mathbb{K}$ imply that $\arg\min_{\kappa\in\mathbb{K}} \mathcal{J}(-\infty,0,\kappa,F)$ is non empty. Let $\kappa^\star$ be an element of the latter. Then, the first equality follows from

$$\xi(F) \geq \inf_{\kappa\in\mathbb{K}} \mathcal{J}(-\infty,0,\kappa,F) - \mathcal{J}(-\infty,0,\kappa^\star,F) = 0.$$

If $\mathbb{K} \not\succcurlyeq_M \mathbb{L}$ for some $\lambda^\star \in \mathbb{L}$, and any $\kappa \in \mathbb{K}$, there exists some $i(\lambda^\star,\kappa), z^\star(\lambda^\star,\kappa) \in A_i$ such that $\Delta_i(z,\lambda,\kappa,F) > 0$. Then the continuity of $\mathcal{J}(-\infty,z,\kappa,F)$ and $\mathcal{J}(z,+\infty,\kappa,F)$ w.r.t. κ, and the compactness of $\mathbb{K}$, imply that, for any $\lambda \notin \mathbb{K}, z \in A_1, \exists \kappa_{\lambda,z} \in \mathbb{K}$ such that

$$\inf_{\kappa\in\mathbb{K}} \Delta_1(z,\lambda,\kappa,F) = \Delta_1(z,\lambda,\kappa_{\lambda,z},F),$$

and thereby

$$\xi(F) \geq \Delta_{1\left(\lambda^\star,\kappa_{\lambda^\star,z^\star}\right)}(z^\star,\lambda^\star,\kappa_{\lambda^\star,z^\star},F) > 0.$$

($\Rightarrow$) Suppose now that $\xi(F) = 0$ and consider an arbitrary λ. This implies that (19) holds and thereby there exists some element of $\mathbb{K}$ for which $\Delta_i(z,\lambda,\kappa,F) \leq 0$, for every $z \in A_i$, $i = 1,2$. If $\xi(F) > 0$, for some $\lambda^\star \in \mathbb{L}$, and some $i = 1,2$, $\inf_{\kappa\in\mathbb{K}} \sup_{z\in A_i} \Delta_i(z,\lambda^\star,\kappa,F) > 0$. It implies that for any $\kappa \in \mathbb{K}$, $\sup_{z\in A_i} \Delta_i(z,\lambda^\star,\kappa,F) > 0$ and the result follows. $\square$

Proof of Proposition 2. The results in the auxiliary Lemma 1 imply that

$\begin{pmatrix} \Delta_1\left(z_1,\lambda,\kappa,\sqrt{T}\left(F_T-F\right)\right) \\ \Delta_2\left(z_2,\lambda,\kappa,\sqrt{T}\left(F_T-F\right)\right) \end{pmatrix}$ weakly converges to $\begin{pmatrix} \Delta_1\left(z_1,\lambda,\kappa,\mathcal{G}_F\right) \\ \Delta_2\left(z_2,\lambda,\kappa,\mathcal{G}_F\right) \end{pmatrix}$ w.r.t. to the product topology of continuous (w.r.t. (z_1,z_2,λ)) epi-convergence (w.r.t. κ) on the product of the relevant spaces of lsc real valued functions (see e.g. Knight (1999) for the dual notion of epi-convergence). This product space is metrizable as complete and separable (see again Knight (1999)). Hence, Skorokhod representations are applicable (as above, see for example Theorem 1 in Cortissoz (2007)) and thereby for any (z_1,z_2,λ) and any sequence $(z_{1,T},z_{2,T},\lambda_T)\to(z_1,z_2,\lambda)$, there exists an enhanced probability space and processes $\begin{pmatrix} \Delta_{1,T}(\kappa) \\ \Delta_{2,T}(\kappa) \end{pmatrix} \overset{d}{=} \begin{pmatrix} \Delta_1\left(z_{1,T},\lambda_T,\kappa,\sqrt{T}\left(F_T-F\right)\right) \\ \Delta_2\left(z_{2,T},\lambda_T,\kappa,\sqrt{T}\left(F_T-F\right)\right) \end{pmatrix}$, $\begin{pmatrix} \Delta_1^\star(\kappa) \\ \Delta_2^\star(\kappa) \end{pmatrix} \overset{d}{=} \begin{pmatrix} \Delta_1\left(z_1,\lambda,\kappa,\mathcal{G}_F\right) \\ \Delta_2\left(z_2,\lambda,\kappa,\mathcal{G}_F\right) \end{pmatrix}$, defined on it such that $\begin{pmatrix} \Delta_{1,T} \\ \Delta_{2,T} \end{pmatrix} \to \begin{pmatrix} \Delta_1^\star \\ \Delta_2^\star \end{pmatrix}$ almost surely, w.r.t. to the product topology of epi-convergence, where $\overset{d}{=}$ denotes equality in distribution. Notice that,

$$\begin{pmatrix} \Delta_1\left(z_{1,T},\lambda_T,\kappa,\sqrt{T}F_T\right) \\ \Delta_2\left(z_{2,T},\lambda_T,\kappa,\sqrt{T}F_T\right) \end{pmatrix} \overset{d}{=} \begin{pmatrix} K_{1,T}(\kappa) \\ K_{2,T}(\kappa) \end{pmatrix} := \begin{pmatrix} \Delta_{1,T}(\kappa) \\ \Delta_{2,T}(\kappa) \end{pmatrix} + \sqrt{T}\begin{pmatrix} \Delta_1\left(z_{1,T},\lambda_T,\kappa,F\right) \\ \Delta_2\left(z_{2,T},\lambda_T,\kappa,F\right) \end{pmatrix}.$$

Under $\mathbf{H_0}$, due to the previous, we have that for any $i=1,2$, $\kappa,\kappa_T\in\mathbb{K}$, and $\kappa_T\to\kappa$, $\lim_{T\to\infty}K_{i,T}(\kappa_T)$ is almost surely equal to

$$\begin{cases} \Delta_i^\star(\kappa), & (z_i,\lambda,\kappa,)\in \mathrm{Int}\Gamma_i \\ +\infty, & (z_i,\lambda,\kappa,)\notin\Gamma_i, \Delta_i(z_i,\lambda,\kappa,F)>0 \\ -\infty, & (z_i,\lambda,\kappa,F)\notin\Gamma_i, \Delta_i(z_i,\lambda,\kappa,F)<0 \end{cases}.$$

Furthermore, for any compact $\mathcal{K}_i$ that contains $\kappa\in\mathbb{K}$ such that $(z_{i,T},\lambda_T,\kappa,)$ even-

tually belongs to the boundary of Γ_i we have that almost surely,

$$\lim \inf_{T\to\infty} \inf_{\kappa\in\mathcal{K}_i} K_{i,T}(\kappa) \geq \inf_{\kappa\in\mathcal{K}_i} \Delta_i^\star(\kappa) + \lim \inf_{T\to\infty} \inf_{\kappa\in\mathcal{K}_i} \sqrt{T}\Delta_i(z_{i,T}, \lambda_T, \kappa, F) \geq \inf_{\kappa\in\mathcal{K}_i} \Delta_i^\star(\kappa).$$

Hence due to Proposition 3.2.(ii)-(iii) (ch. 5, p. 337) of Molchanov (2006), $\begin{pmatrix} K_{1,T}(\kappa) \\ K_{2,T}(\kappa) \end{pmatrix}$ almost surely converges w.r.t. to the product topology of epi-convergence over $\mathbb{K}$, and continuously over $A_i \times \mathbb{L}$ to $K(\kappa) = \begin{pmatrix} K_1(\kappa) \\ K_2(\kappa) \end{pmatrix}$, with $K_i(\kappa) = \begin{cases} \Delta_i^\star(\kappa), & (z_i, \lambda, \kappa) \in \Gamma_i \\ -\infty, & (z_i, \lambda, \kappa) \notin \Gamma_i \end{cases}$. Since $\mathbb{K}$ is compact, Theorem 3.4 (ch. 5, p. 338) of Molchanov (2006) implies that almost surely,

$$\inf_{\kappa\in\mathbb{K}} K_{i,T}(\kappa) \to \begin{cases} \inf_{\kappa:(z_i,\lambda,\kappa)\in\Gamma_i} \Delta_i^\star(\kappa), & \exists\kappa : (z_i, \lambda, \kappa) \in \Gamma_i \\ -\infty, & \nexists\kappa : (z_i, \lambda, \kappa) \in \Gamma_i \end{cases},$$

jointly over $i = 1, 2$. When Γ_i is not empty, by Theorem 7.11 of Rockafellar and Wetts (2009), and using the same notations (to streamline the proof) for the random elements defined in the relevant enhanced probability space, the sequence $\left(\inf_\kappa \Delta_i\left(z_i, \lambda, \kappa, \sqrt{T}F_T\right)\right)_T$ is also equi-upper semi-continuous. Due to the proof of Lemma 2 below and the form of $\mathbf{H_0}$, we have that the above sequence is almost surely bounded, and thereby Theorem 3.4 (ch. 5, p. 338) of Molchanov (2006) implies that almost surely,

$$\sup_{z_i,\lambda} \inf_\kappa \Delta_i\left(z_i, \lambda, \kappa, \sqrt{T}F_T\right) \to \sup_{z_i,\lambda} \inf_{\kappa\in\Gamma_i} \Delta_i(z_i, \lambda, \kappa, \mathcal{G}_F).$$

When Γ_i is empty the limit is trivially $-\infty$. Reverting from the Skorokhod representations to the original sequences and employing the continuous mapping theorem we get the result. □

Proof of Theorem 2. The first result follows by a direct application of Theorem 3.5.1.i of Politis et al. (1999) from the results of Proposition 2, and the limiting quantile function being continuous for all $\alpha \in (0,1)$. The second result follows similarly, by also considering the results of the auxiliary Lemma 2. For the second result, if $\mathbf{H_a}$ is true, for some $\lambda^\star \in \mathbb{L} - \mathbb{K}$, and any $\kappa \in \mathbb{K}$, there exists some $i, z^\star \in A_i$ such that $\Delta_i(z, \lambda, \kappa, F) > 0$. Then, we have that

$$\xi_T \geq \inf_{\kappa \in \mathbb{K}} \Delta_i\left(z^\star, \lambda^\star, \kappa, \sqrt{T}\left(F_T - F\right)\right) + \sqrt{T} \inf_{\kappa \in \mathbb{K}} \Delta_i\left(z^\star, \lambda^\star, \kappa, F\right),$$

and from arguments analogous to the ones used in the proof of Proposition 2, we have that the first term in the rhs of the last display is asymptotically tight, while from the arguments used in the proof of Proposition 1, the second term in the rhs of the last display diverges to $+\infty$. The result follows from the properties of b_T. □

Proof of Theorem 3. The result follows exactly as in the proofs of Proposition 2 and Theorem 2 by noting first that the relevant hypo-epi convergence concepts in the aforementioned proposition also hold for the relevant function restricted to $A_i^{(T)}$ from the results there and the definition of the Painleve-Kuratowski set convergence, and that $\sup_\lambda \inf_\kappa \Delta_i(z, \lambda, \kappa, \mathcal{G}_F)$ has the same sup w.r.t. z with its restriction to any dense subset of A_i due to the compactness of $\mathbb{L}$ and $\mathbb{K}$ and Theorem 3.4 (ch. 5, p. 338) of Molchanov (1999). □

Auxiliary Lemmata

The following are auxiliary lemmata used for the derivation of the proofs of Proposition 2 and Theorem 2.

Lemma 1. *Under Assumption 2*

$$\begin{pmatrix} \Delta_1\left(z_1, \lambda, \kappa, \sqrt{T}\left(F_T - F\right)\right) \\ \Delta_2\left(z_2, \lambda, \kappa, \sqrt{T}\left(F_T - F\right)\right) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \Delta_1\left(z_1, \lambda, \kappa, \mathcal{G}_F\right) \\ \Delta_2\left(z_2, \lambda, \kappa, \mathcal{G}_F\right) \end{pmatrix}$$

as random elements with values on the space of $\mathbb{R}^2$-valued bounded functions on $\mathbb{L} \times \mathbb{K} \times \mathbb{R}_- \times \mathbb{R}_{++}$ equiped with the sup-norm. The limiting process has continuous sample paths.

Proof. Let $\theta := (\lambda, \kappa, z_1, z_2) \in \Theta := \mathbb{L} \times \mathbb{K} \times \mathbb{R}_- \times \mathbb{R}_{++}$, ρ any non zero element of $\mathbb{R}^2$, and consider $\Delta(\theta, \cdot) := \rho_1 \Delta_1(z_1, \lambda, \kappa, \cdot) + \rho_2 \Delta_2(z_1, \lambda, \kappa, \cdot)$. Notice that Theorem 7.3 of Rio (2013), due to Assumption 2, implies that $\sqrt{T}(F_T - F) \rightsquigarrow \mathcal{G}_F$. This implies that $\sqrt{T}(F_T - F)$ also weakly hypo-converges to $\mathcal{G}_F$ (see for example Knight (1999)). Both are upper semi-continuous (usc) $\mathbb{P}$ a.s. and the space of usc functions with the topology of epiconvergence can be metrized as complete and separable (see again Knight (1999)). Due to separability and the Skorokhod Representation Theorem (see for example Theorem 1 in Cortissoz (2007)) there exists a suitable probability space and random elements with values in the aforementioned function space such that $f_T^* \overset{d}{=} \sqrt{T}(F_T - F)$, $f^* \overset{d}{=} \mathcal{G}_F$, and $f_T^* \to f^*$ a.s.. Let $J \equiv \overline{\text{span}}\{f_T^*, f^*, T = 1, 2, \cdots\}$ equipped with the metrizable topology of weak convergence.[12] Consider $\Delta(\cdot, \cdot)$ restricted to J with values in the linear space of stochastic processes, equipped with the topology of convergence in distribution, with values in the space of bounded real functions defined on Θ equipped with the sup-norm. From Assumption 2, Corollary 4.1, and Theorem 7.3 of Rio (2013), we also have that

$$\sup_{\theta \in \Theta} \sup_T \mathbb{E}\left[\left(\Delta\left(\theta, \sqrt{T}\left(F_T - F\right)\right)\right)^2\right] + \sup_{\theta \in \Theta} \mathbb{E}\left[\left(\Delta\left(\theta, \mathcal{G}_F\right)\right)^2\right] < +\infty.$$

[12] Here $\overline{\text{span}}$ denotes the closure w.r.t the particular topology of the linear span.

The latter inequality along with Theorem 6.5.2 in Narici and Beckenstein (2010), the metrization of convergence in distribution by the bounded Lipschitz metric (see for example p. 73, van der Vaart and Wellner (1996)) which is bounded from above by $\sup_\theta \mathbb{E}\left[(x-y)^2\right]$, for x, y members of the aforementioned space of processes, imply that $\Delta(\cdot,\cdot)$ as restricted above is continuous. Hence the CMT implies that $\Delta(\theta, f_T^*) \rightsquigarrow \Delta(\theta, f^*)$ which means that $\Delta\left(\theta, \sqrt{T}(F_T - F)\right) \rightsquigarrow \Delta(\theta, \mathcal{G}_F)$. This and the Cramer-Wold Theorem imply the needed result. The final assertion follows from $\sup_{\theta\in\Theta} \mathbb{E}\left[(\Delta(\theta, \mathcal{G}_F))^2\right] < +\infty$, the discussion in Example 1.5.10 of van der Vaart and Wellner (1996), and the continuity of $\mathbb{E}\left[(\Delta(\theta, \mathcal{G}_F))^2\right]$ w.r.t. θ. □

Lemma 2. *If ξ_∞ is non-constant, and under Assumptions 2 and 4, the distribution of ξ_∞ has support $[0, +\infty)$, its cdf is absolutely continuous on $(0, +\infty)$, and it may have a jump discontinuity at zero, of size at most $ch_{\mathbb{L}}(\mathbb{K})$.*

Proof. The result stems from Corollary 1 as long as the requirements of Assumption 1 are satisfied and an appropriately bounding η is found. For $\Lambda = \mathbb{L} \times \mathbb{K} \times \{1,2\} \times \mathbb{R}_- \times \mathbb{R}_{++}$ where $\{1,2\}$ is considered equipped with the discrete metric, we have that $X_\lambda = 1_1(i)\,\Delta_1(z_1, \lambda, \kappa, \mathcal{G}_F) + 1_2(i)\,\Delta_2(z_2, \lambda, \kappa, \mathcal{G}_F)$, for $\lambda = (\lambda, \kappa, i, z_1, z_2)$, has continuous sample paths from the final assertion of Lemma 1. Then notice that

$$\mathbb{E}\left[\sup_\Lambda \left(X_\lambda^2\right)\right] \leq \sum_{i=1,2} \mathbb{E}\left[\sup_{\lambda\in\mathbb{L}} \sup_{\kappa\in\mathbb{K}} \sup_{z\in A_i} \Delta_i^2(z, \lambda, \kappa, \mathcal{G}_F)\right].$$

From the zero mean Gaussianity of the processes involved, the packing numbers of $\Lambda \times \mathbb{R}$ being bounded by a polynomial w.r.t. the inverted radii, Proposition A.2.7 of Van Der Vaart and Wellner (1996) implies the subexponentiality of the distributions of the suprema above, and thereby the existence of their second moments. Hence

Hypothesis 1 of Assumption 1 holds. Using the discussion in Nualart (2006), immediately after the proof of Proposition 2.1.11 (p. 109) we have that Hypothesis 2 of Assumption 1 also holds due to Assumption 2. Due to zero mean Gaussianity and excluding $\mathbb{P}$-negligible events $\Delta_i\left(z,\lambda,\kappa,\mathcal{G}_F\right)$ is zero only when $\kappa=\lambda$ and it is at most only then that ξ_∞ has degenerate variance. Thereby, $\mathcal{T}=\{0\}$ and we can try to obtain a lower bound for ξ_∞. From the integration by parts formula for the Lebesgue-Stieljes integral and Assumption 2, we get

$$\begin{aligned}\xi_T &\geq \max_i \sup_{\lambda\in\mathbb{L}} \inf_{\kappa\in\mathbb{K}} \Delta_i\left(0,\lambda,\kappa,\sqrt{T}F_T\right)\\ &\geq \eta_T := \frac{1}{2}\frac{1}{\sqrt{T}}\left(\sup_{\lambda\in\mathbb{L}}\lambda^{Tr}-\sup_{\kappa\in\mathbb{K}}\kappa^{Tr}\right)\sum_{i=1}^{T}\left(Y_i-\mathbb{E}\left(Y_0\right)\right)\\ &\rightsquigarrow \eta_\infty := \frac{1}{2}\sup_{\lambda\in\mathbb{L}}\lambda^{Tr}Z-\frac{1}{2}\sup_{\kappa\in\mathbb{K}}\kappa^{Tr}Z,\end{aligned}$$

where $Z\sim N\left(0_{n\times 1},\mathbb{V}\right)$. Hence, $\xi_\infty\geq\eta_\infty\geq 0$.

The previous inequality implies the applicability of Corollary 1 for $c=0$. We obtain the result by estimating an upper bound for $\mathbb{P}\left(\eta_\infty=0\right)$. From Assumption 2 and the non-degeneracy of $\mathbb{V}$ the latter probability equals exactly the probability that the maximum of the random vector Z occurs at a coordinate that represents an extreme point of $\mathbb{S}$ to which corresponds a common effective extreme point for $\mathbb{L}$ and $\mathbb{K}$ (w.r.t. $\mathbb{L}$), say λ, evaluated at which $\lambda^{Tr}Z$ is maximal. Using Theorem 2 in chapter 3 (p. 37) of Sidak et al. (1999) by (in their notation) letting p be the density of the n-variate standard normal distribution and q the density of $N\left(0_{n\times 1},\mathbb{V}\right)$, along with Definition 4, we get:

$$\mathbb{P}\left(\eta_\infty=0\right)\leq ch_{\mathbb{L}}\left(\mathbb{K}\right).$$

□

	MSD optimal portfolio	market portfolio
Mean	0.01035	0.00510
Standard Deviation	0.04290	0.04420
Skewness	-0.27730	-0.52629
Excess Kurtosis	1.18535	1.96705
VaR 5%	0.06133	0.0718
Sharpe Ratio	0.17495	0.04697
Downside Sharpe Ratio	0.12986	0.39570
Return Loss	0.7856%	
Opportunity Cost ($c = 2.25$)		
$a = b = 2$	0.704%	
$a = b = 3$	0.990%	
$a = b = 4$	1.565%	

Table 2: Performance and risk measures. Entries report performance and risk measures for the MSD optimal portfolios and the market portfolio computed with one-month rolling windows. We list mean, volatility, skewness, excess kurtosis, empirical VaR 5% (positive sign for a loss), Sharpe ratio, downside Sharpe ratio, return loss, and opportunity cost. The dataset spans the period from July 31, 1963 to December 31, 2016.

Descriptive statistics of the weight allocation of the MSD portfolio					
Base Assets	Portfolio	Mean	Std. Dev.	Skewness	Kurtosis
	Market	0.0933	0.0461	-0.5990	0.0279
	T-Bill	0.0265	0.0678	2.1743	2.7361
6 FF	Big LoBM	0.0282	0.0472	1.7689	2.4370
	Big AvBM	0.0203	0.0529	2.2407	3.0765
	Big Hi BM	0.1515	0.06231	-0.9441	-0.6088
10 momentum	Prior 3	0.0539	0.0397	-0.5491	-1.6188
	Prior 4	0.0178	0.0329	2.0128	3.2530
	Prior 5	0.0249	0.0554	1.7780	1.1651
	Prior 7	0.023	0.0099	4.0016	14.056
	Prior 8	0.0097	0.0165	1.1428	-0.6570
	Prior 9	0.0027	0.0101	3.5233	10.446
	Hi Prior	0.0886	0.0525	-0.7163	-0.3800
10 industry	Telcm	0.0301	0.0464	1.2129	0.1298
	Hith	0.0327	0.0680	1.8638	1.8056
	Utils	0.0124	0.0349	4.3510	18.234
	Energy	0.0984	0.0717	-0.2824	-1.4674
10 size	ME1	0.1911	0.0156	-1.1916	-0.5817
10 beta	Lo 10	0.0671	0.0519	-0.3527	-1.7449
	Quant. 20	0.0318	0.0627	1.9163	2.1448
	Quant. 30	0.0052	0.0198	3.5233	10.446
	Quant. 40	0.0025	0.0106	4.0016	14.056

Table 3: Descriptive statistics of the weight allocation of the MSD optimal portfolios over the period 07/1963-12/2016 computed with one-month rolling windows.

	a_i	$R_M - R_F$	SMB	HML	MOM
Coef.	0.508	0.948	-0.031	0.133	0.004
t-stat	1.294	1.021	-2.484	9,380	0.441
p-values	0	0	0.013	0	0.659

Adj. R^2	F-statistic	p-value
0.948	2.953	0

Table 4: Carhart four-factor model. Entries report the coefficients and their respective t-statistics,as well as Adjusted R2, F-statistic, and p-values. The dataset spans 07/1963-12/2016for MSD optimal portfolios computed with one-month rolling windows.

	a_i	$R_M - R_F$	SMB	HML	RMW	CMA
Coef.	0.419	0.981	-0.019	0.201	0.021	-0.06
t-stat	15.30	146.3	-2.075	15.51	1.597	-3.327
p-values	0	0	0.038	0	0.111	0.009

Adj. R^2	F-statistic	p-value
0.988	5361.6	0

Table 5: Fama-French five factor model. Entries report the coefficient estimates, their respective t-statistics, as well as Adjusted R2, F-statistic, and p-values. The dataset spans 07/1963-12/2016 for MSD optimal portfolios computed with one-month rolling windows.