



Working paper

2018

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

Spanning tests for markowitz stochastic dominance

Arvanitis, Stelios; Scaillet, Olivier; Topaloglou, Nikolas

How to cite

ARVANITIS, Stelios, SCAILLET, Olivier, TOPALOGLOU, Nikolas. Spanning tests for markowitz stochastic dominance. 2018

This publication URL: <https://archive-ouverte.unige.ch/unige:102836>

© The author(s). This work is licensed under a Creative Commons Public Domain (CC0)

<https://creativecommons.org/publicdomain/zero/1.0/>

Spanning Tests for Markowitz Stochastic Dominance

Stelios Arvanitis

Athens University of Economics and Business ¹

Olivier Scaillet

University of Geneva and Swiss Finance Institute ²

Nikolas Topaloglou

Athens University of Economics and Business ³

Current draft: January 2018

¹Department of Economics, 76, Patision Street, GR10434, Athens, Greece; e-mail: stelios@aueb.gr

²Corresponding author: GFRI, Bd du Pont d'Arve 40, 1211 Geneva, Switzerland; e-mail: olivier.scaillet@unige.ch

³Department of International European & Economic Studies, 76, Patision Street, GR10434, Athens, Greece; e-mail: nikolas@aueb.gr

Abstract

Using properties of the cdf of a random variable defined as a saddle-type point of a real valued continuous stochastic process, we derive first-order asymptotic properties of tests for stochastic spanning w.r.t. a stochastic dominance relation. First, we define the concept of Markowitz stochastic dominance spanning, and develop an analytical representation of the spanning property. Second, we construct a non-parametric test for spanning via the use of an empirical analogy. The method 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 there exists 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-satiable 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)), among others, since it enables inference on the issue of optimal choice in a non-parametric setting. Several statistical tools have been developed to test whether a particular random element (or probability distribution) of interest, w.r.t. some fixed notion of stochastic dominance, dominates any other similar random element (or probability

distribution) in a given set, i.e., the former is super-efficient w.r.t. the aforementioned notion over the latter set.¹ We can find some illustrative examples among others, in the relevant application sections of Horvath, Kokoszka, and Zitikis (2006), where interest lies in distributions of income, or Post and Levy (2005) and Scaillet and Topaloglou (2010), where interest lies in financial portfolios.

Inspired by previous work, 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 extend the well-known first degree stochastic dominance (FSD) and second degree stochastic dominance (SSD). FSD can be characterised by the preferred choice between two loteries by investors who prefer more than less, while SSD can be characterised by the preferred choice between two loteries by risk averse investors (see Hadar and Russell (1969), Hanoch and Levy (1969), and Rothschild and Stiglitz (1970)). PSD and MSD investigate choices by investors who have S-shaped utility functions and reverse S-shaped utility functions. Arvanitis and Topaloglou (2017) develop consistent tests for PSD and MSD efficiency which is an extension to the case where full diversication is allowed.

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 the notion of Mean-Variance spanning in Huberman and Kandell (1987), that was formulated in the context of second order stochastic dominance in Arvanitis et al. (2017). Yet, we can generalize it in the framework of an arbitrary preorder defined on some set of probability distributions. With some loss of generality, given such a preorder, and if the efficient set is non-empty, a spanning subset of the original set of distributions w.r.t. the relation is essentially any superset of the efficient

¹Equivalently, it refers to the property of being a greatest element of the preorder. The term super-efficiency is adopted in Arvanitis et al. (2017), so as to distinguish from the concept of an efficient element, i.e., a maximal element of the preorder, namely a member of the relevant efficient set.

set.² 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 which could be very large. In such a case, we can reduce the examination of the optimal choice problem, to a potentially easier one. Both issues can be of great interest to financial economics since, and in the light of the aforementioned applications, the underlying distributions represent returns of financial assets and the preorders 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)).³

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 also be useful in extracting important properties of financial markets.

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, in some cases 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. (2017) for the case of SSD, or Proposition 1 below for the case of MSD). The verification of the above is usually analytically

²The exact definition of spanning enables the existence of spanning sets even if the preorder has no greatest elements or more generally no maximal elements, see Section 3.1 below. Hence it is a generalization of the concept of super-efficiency, since a greatest element essentially defines a singleton spanning set.

³Obviously those notions could also be of potential interest in any field of economic theory or decision science that examines optimal choice under uncertainty.

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 use. However, given the existence of empirical information and using the principle of analogy, we can design non-parametric tests of the null hypothesis of spanning. The limit theory of tests for stochastic spanning⁴ 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 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 motivation for the present paper is the study of continuity properties of the cdf of random variables defined as saddle type points of real valued stochastic processes. In this respect, Section 2 of the paper sets up the probabilistic framework, and derives the relevant to the above, 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. More specifically, under weak conditions involving Malliavin differentiability, existence of moments for suprema as well as a countability property for the singular points of the derivative, we derive connectedness for the support of the law, a countable number of atoms, and absolute continuity when restricted between successive atoms. 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 (see Section 2 for references and examples).

Our second motivation is the following. The results in Arvanitis et al. (2017)

⁴Spanning 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.

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 results⁵ 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 allow for 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 preorders.

Given this motivation, in Section 3 of the paper, and in the context of financial portfolios formation, we are occupied with the concept of stochastic spanning w.r.t. the MSD preorder. We define the notion and provide with an original characterization of spanning by the zero of an analogous 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 aforementioned probabilistic results and a 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. (2017), we allow for unbounded supports for the involved distributions, we suppose that the relevant parameter spaces are simplicial complexes, as well as obtain results of asymptotic degeneracy in some cases. In this sense, we can also consider the present results as generalizations of the ones derived in the aforementioned paper. We provide 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.

⁵possibly also involving more complex derivations.

Inspired by Arvanitis and Topaloglou (2017), who show that the market portfolio is not MSD efficient, we test in an empirical application in Section 4, whether investors with MSD preferences could beat the market through equity management, according to Markowitz preferences.

For this purpose, we use proxies of the individual assets in the investment universe. 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, along with the market portfolio and the T-bill. These portfolios have been at the center of the empirical asset pricing literature over the past two decades (see for example Post (2003), Kuosmanen (2004), Post and Levy (2005), Post and Kopa (2013), Gonzalo and Olmo (2014) among others in the stochastic dominance framework).

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 criterion. 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. The optimal MSD portfolio better suits the MSD investors that are risk averse for losses and risk lover 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 $\Lambda_1, \Lambda_2, \dots, \Lambda_s$ are separable metric spaces, and consider $\Lambda = \prod_{i=1}^s \Lambda_i$ equipped with the product topology. Furthermore, consider the functional oper $\equiv \text{opt}_1 \circ \text{opt}_2 \circ \dots \circ \text{opt}_s$ where $\text{opt}_i = \text{sup}$ or inf w.r.t. to some non-empty compact $\Lambda_i^* \subseteq \Lambda_i$, for $i = 1, \dots, s$. When $i > 1$, Λ_i^* is allowed to depend on the elements of $\prod_{j=1}^{i-1} \Lambda_{i-j}^*$, $i > 1$.

The probabilistic framework follows closely Chapter 2 of Nualart (2006). In this respect, it consists of $(\Omega, \mathcal{F}, \mathbb{P})$, a complete probability space, where \mathcal{F} is generated by some isonormal Gaussian process $W = \{W(h), h \in H\}$, where H is an appropriate Hilbert space. X is some real valued stochastic process on Λ with bounded sample paths (i.e., its paths are \mathbb{P} a.s. elements of $C(\Lambda, \mathbb{R})$). In many applications, X is a Gaussian weak limit for some net of appropriate processes. We denote the Malliavin derivative operator by D and by $\mathbb{D}^{1,2}$ the completion of the family of Malliavin differentiable random variables w.r.t. the norm $\sqrt{\mathbb{E}[z^2 + (Dz)^2]}$.

We are interested in the form of the support and the continuity properties of the cdf of the law of the random variable $\xi \equiv \text{oper}X_\lambda$. 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,⁶ and second for discontinuities. Hence, it generalizes established results concerning the absolute continuity of the distribution of suprema of stochastic processes. For an excellent treatment of those results see, inter alia Propositions 2.1.7 and 2.1.10 of Nualart (2006), and for the literature on the fibering method and its probabilistic applications, see for example Lifshits (1983).

Assumption 1. *For the stochastic process $X : \Omega \rightarrow C(\Lambda, \mathbb{R}^q)$ suppose that:*

⁶The term "saddle-type" is used here slightly in an abusive manner, since commutativity between the successive optimization functionals does not hold in general.

1. $\mathbb{E} [\sup_{\Lambda} (X_{\lambda}^2)] < +\infty$.
2. For all $\lambda \in \Lambda$, $X(\lambda) \in \mathbb{D}^{1,2}$, and the H -valued process DX has a continuous version and $\mathbb{E} [\sup_{\Lambda} \|DX_{\lambda}\|^2] < +\infty$.
3. For some countable $\mathcal{T} \subset \mathbb{R}$, $\mathbb{P}(\{\xi = \tau\} \cap \Omega_{\tau}) \geq 0$ holds if and only if $\tau \in \mathcal{T}$, where Ω_{τ} denotes $\{\omega \in \Omega : DX_{\lambda}(\omega) = 0 \text{ for some } \lambda \text{ such that } \tau = X_{\lambda}(\omega)\}$.

In the usual case, i.e., when X is zero-mean Gaussian, we can establish the first condition by strong results that imply the subexponentiality of the distribution of $\sup_{\Lambda} X_{\lambda}$, such as Proposition A.2.7 of van der Vaart and Wellner (1996). This 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 would imply 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, "outer approximations" of \mathcal{T} can be derived 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 $\text{supp}(\xi)$, 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} has infimum, the law restricted to $(-\infty, \inf \mathcal{T})$ is absolutely continuous w.r.t. the Lebesgue measure. Dually, if \mathcal{T} has supremum, 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 aforementioned literature that hold when oper is a composition of suprema, the parameter spaces Λ are not dependent, and $\mathbb{P}(\Omega_\tau) = 0$, for all $\tau \in \mathcal{T}$. 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 functionals. The following corollary focuses on this particular case and estimates the size of the potential jump discontinuity by assuming the existence of an auxiliary first-order stochastically dominated random variable.

Corollary 1. *Suppose that Assumption 1 is satisfied. Furthermore, suppose that $\mathcal{T} = \{c\}$, $\xi \geq \eta$, \mathbb{P} a.s., and that $\text{supp}(\eta) = [c, +\infty)$. Then, $\text{supp}(\xi) = [c, +\infty)$, 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 .*

This corollary is to our view the most useful result for the establishment of the limit theory for tests of stochastic spanning. In such frameworks, it is usually the case that X is Gaussian, that it is derived as a weak limit of processes used in the definition of the test statistics while ξ can be conveniently defined as a difference between infima of X defined on different regions of Λ with easily derivable properties.

3 A Spanning Test for MSD

In this section, we introduce the concept of stochastic spanning for the MSD relation. First, we provide some order theoretical characterization of the concept, provide the non-emptiness of the efficient set, and derive an analytical representation using a functional defined by recursive optimizations. Then, we define a testing procedure using the principle of conditioning based on subsampling, and derive its first-order limit theory, among others via the use of 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. 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 $\mathbb{R}^n \ni x \rightarrow \lambda^{Tr} x$, where λ assumes its values in \mathbb{L} which is a closed non-empty subset of $\mathbb{S} = \{\lambda \in \mathbb{R}_+^n : \mathbf{1}^{Tr} \lambda = 1, \}$, and Tr denotes the transpose operator. 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 we can construct upon the previous.⁷ The parameter set \mathbb{L} represents the collection of feasible portfolios formed by economic, legal, and other potential restrictions. We denote generic elements of \mathbb{L} by λ, κ etc. In order to define the concepts of MSD and subsequently of spanning consider

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

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

$$\Delta_1(z, \lambda, \kappa, F) \equiv \mathcal{J}(-\infty, z, \kappa, F) - \mathcal{J}(-\infty, z, \lambda, F) \leq 0, \forall z \in \mathbb{R}_-,$$

and

$$\Delta_2(z, \lambda, \kappa, F) \equiv \mathcal{J}(z, +\infty, \kappa, F) - \mathcal{J}(z, +\infty, \lambda, F) \leq 0, \forall z \in \mathbb{R}_{++}.$$

The existence of the mean of the underlying distribution implies that we can allow the limits of integration above to assume extended values, whence the differences $\Delta_1(z, \lambda, \kappa, F)$ and $\Delta_2(z, \lambda, \kappa, F)$ are well defined. Furthermore, Levy and Levy (2002) show that $\kappa \succ_M \lambda$ 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.

⁷The base assets are not restricted to be individual securities but are defined simply as the extreme points of the largest portfolio set \mathbb{S} .

We have that \succsim_M is a preorder on \mathbb{L} since it does not generally satisfy antisymmetry since $G(u, \kappa, F) = G(u, \lambda, F)$, for any u , does not imply that $\kappa = \lambda$. If the inequalities above are satisfied as equalities, the pair (κ, λ) belongs to the possibly non-trivial equivalence part of the preorder. Strict dominance $\kappa \succ_M \lambda$ is the irreflexive part of the preorder and it holds iff at least one of the previous inequalities holds strictly for some $z \in \mathbb{R}$. Finally, since, for some $z \in \mathbb{R}$, at least one of the inequalities defining the preorder can change, the relation is not generally total. When this is the case, we cannot compare κ and λ w.r.t. \succsim_M . The following definition and the subsequent lemma clarify the concept of spanning and part of its structure.

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

If the set of maximal elements, i.e., the efficient set \mathcal{E}_M of the preorder, is non-empty, $\mathbb{K} \supseteq \mathcal{E}_M$ implies that $\mathbb{K} \succsim_M \mathbb{L}$.⁸ If equivalent (w.r.t. \succsim_M) elements of \mathbb{L} exist inside \mathcal{E}_M , there exist from the axiom of choice strict subsets of the efficient set that span \mathbb{L} .

If $\mathbb{K} \succsim_M \mathbb{L}$, the optimal choice of every agent with preferences represented by a reverse S-shaped utility 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 often less complex than the former. Therefore we can motivate the interest in the verification of spanning by reasons of tractability to the problem of optimal choice in such frameworks.

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.

Furthermore, and given that questions about the mathematical structure of \mathcal{E}_M are important in decision theory, any spanning set can be perceived as an outer approximation of the former. Hence the notion becomes relevant to the problem of the examination of the properties of the efficient set, which is generally complex. We

⁸ Super-efficiency occurs when \mathcal{E}_M is non empty and "inhabited" by greatest elements.

expect that we can approximate the properties of the efficient set by the properties of sequences of spanning sets that appropriately converge to it.

Naturally, those raise the following question: given \mathbb{K} , a non empty subset of \mathbb{L} ,⁹ is $\mathbb{K} \succ_M \mathbb{L}$? The following proposition provides with an analytical characterization by means of nested optimizations.

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

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

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

The case of super-efficiency is trivially implied by the previous result.

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

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

We cannot directly use the previous proposition if F is unknown and/or the optimizations involved are infeasible as it is usually the case. 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 statistical inference procedures that address the question above.

3.2 A Consistent Non-parametric Test

We employ Lemma 1 in order to construct a non-parametric test for the question above. If $\mathbb{K} \succ_M \mathbb{L}$ is chosen as the null hypothesis, the hypothesis structure takes

⁹We are not occupied here with the issue of the selection of \mathbb{K} , and the latter is considered as somehow given. In some cases, we can select \mathbb{K} by economically relevant information, see for example the application in Arvanitis et al. (2017) for SSD. We leave the crucial 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.

the following form:¹⁰

$$\begin{aligned}\mathbf{H}_0 &: \xi(F) = 0, \\ \mathbf{H}_a &: \xi(F) > 0.\end{aligned}$$

In order to proceed with the development of the decision process, 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 i^{th} element of Y_t . The sample of size T is the random element $(Y_t)_{t=1, \dots, T}$. In a financial framework it usually represents returns of n financial base assets upon which we can construct portfolios via convex combinations. We denote the cdf of Y_0 by F , and the empirical cdf of the random element $(Y_t)_{t=1, \dots, T}$ by \hat{F}_T . Given the previous and using the principle of analogy, we consider the following random variable that takes the role of the test statistic

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

and is the appropriately scaled empirical analog of $\xi(F)$. When \mathbb{K} is a singleton, the test statistic coincides with the one used in Arvanitis and Topaloglou (2017). Now, the following assumption enables the derivation of the limit distribution of ξ_T under \mathbf{H}_0 .

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

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

is positive definite.

The mixing part of the previous assumption is readily implied by concepts such as geometric ergodicity which holds for many stationary models used in the context of financial econometrics under parameter restrictions and restrictions on the properties

¹⁰Corollary 2 implies that the hypotheses are, in the special case of super-efficiency, as in Arvanitis and Topaloglou (2017).

of the innovation processes involved. Prominent examples are the strictly stationary versions of (possibly multivariate) ARMA or several GARCH and stochastic volatility type of models (see Francq and Zakoian (2011) for several examples). Counter-examples are stationary models that exhibit long memory, etc. The moment existence condition enables the validity of a mixing CLT. It is readily established in models such as the ones mentioned above usually in the form of stricter restrictions on the properties of building blocks and the parameters of the processes involved. 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, Kokoszka, and Zitikis (2006).

In order for an asymptotically meaningful testing procedure to be established, we need the limit theory for ξ_T under the null hypothesis. We derive it via among others the use of the concept of Skorokhod representations along with an iterative consideration of the dual notions of epi/hypo-convergence. In this respect, consider the following parameter space

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

For any i , it is non empty since $\Gamma_i^* \equiv \{(\kappa, \kappa, z), \kappa \in \mathbb{K}, z \in A_i\} \subseteq \Gamma_i$. 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 $i = 1, 2$.¹¹ Hence, $\Gamma_i^* \subset \Gamma_i$. 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 $T \rightarrow \infty$, $\xi_T \rightsquigarrow \xi_\infty$, where $\xi_\infty \equiv \max_{i=1,2} \sup_{z \in A_i} \sup_{\lambda} \inf_{\kappa} \Delta_i(z, \lambda, \kappa, \mathcal{G}_F)$, $(\lambda, z, \kappa) \in \Gamma_i$, and \mathcal{G}_F is a centered Gaussian process with covariance kernel given by*

¹¹For 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 $\lambda \in \mathbb{L}$.

$Cov(\mathcal{G}_F(x), \mathcal{G}_F(y)) = \sum_{t \in \mathbb{Z}} Cov(\mathbb{I}_{Y_0 \leq x}, \mathbb{I}_{Y_t \leq y})$ and \mathbb{P} almost surely uniformly continuous sample paths defined on \mathbb{R}^n .¹²

The covariance kernel above is well defined since

$$\int_0^{+\infty} \sum_{t \in \mathbb{Z}} Cov(\mathbb{I}_{\lambda^{Tr} Y_0 \leq u}, \mathbb{I}_{\lambda^{Tr} Y_t \leq u}) du \leq 2 \sum_{t=0}^{\infty} \sqrt{a_T} \int_0^{+\infty} \sqrt{1 - G(u, \lambda, F)} du < +\infty,$$

and

$$\int_{-\infty}^0 \sum_{t \in \mathbb{Z}} Cov(\mathbb{I}_{\lambda^{Tr} Y_0 \leq u}, \mathbb{I}_{\lambda^{Tr} Y_t \leq u}) du \leq 2 \sum_{t=0}^{\infty} \sqrt{a_T} \int_{-\infty}^0 \sqrt{G(u, \lambda, F)} du < +\infty,$$

where the first inequalities in each of the previous displays come from Inequality 1.12b in Rio (2013), and the second ones come from Assumption 2 (see also p. 196 of Horvath et al. (2006)). Furthermore, if $\Gamma_i = \Gamma_i^*$, for all $i = 1, 2$, the distribution of ξ_∞ is degenerate at zero.

We cannot directly use the results of the previous lemma, in order to construct an asymptotic decision procedure since ξ_∞ depends on the generally unknown F and is highly non linear. However, we can establish a feasible decision rule by the use of a resampling procedure. We consider the method of subsampling, as in the relevant framework of Linton, Post and Whang (2014) (see also Linton, Maasoumi and Whang (2005)).

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 $(Y_i)_{i=t, \dots, t+b_T-1}$ for all $t = 1, 2, \dots, T - b_T + 1$.*
3. *Evaluate the test statistic on each subsample value thereby obtaining $\xi_{T, b_T, t}$ for all $t = 1, 2, \dots, T - b_T + 1$.*

¹²See Theorem 7.3 of Rio (2013).

4. Approximate the cdf of the asymptotic distribution under the null of ξ_T by $s_{T,b}(y) = \frac{1}{T-b_T+1} \sum_{t=1}^{T-b_T+1} 1(\xi_{T,b_T,t} \leq y)$ 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 can derive the first-order limit theory of the procedure, using the results of Proposition 2 and standard asymptotic results from the theory of subsampling. In order to do so, we first use the following usual assumption in the subsampling methodology. It restricts the asymptotic behaviour of the subsampling rate.

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

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

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

In subsampling theory, results such as Theorem 3.5.1 in Politis, Romano and Wolf (1999) require continuity of the limit cdf at the quantile corresponding to the relevant significance level. When ξ_∞ is constant, this holds for any $\alpha \in (0, 1)$. When the distribution of ξ_∞ is non-degenerate, it is possible that it has a discontinuous quantile function. Our current framework implies (see the proof of Lemma 2 in the Appendix) that such a discontinuity can only occur at zero. Given this, and if we can construct an appropriate lower bound for ξ_∞ in the spirit of Corollary 1, it is possible to use it in order to obtain an estimate for the possible jump size of the limiting cdf at zero. Then, it is possible to use the aforementioned theorem by restricting α .

It turns out (again, 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 through the probability that the latter random variable attains the value zero. In order to proceed with this, we essentially use a combinatorial argument facilitated by the following definitions which allows us to

estimate the proportion of the linear functions that when maximized over \mathbb{L} , their unique maximizer is a common extreme point of both the parameter spaces. We also use a final assumption that restricts the form of \mathbb{L} and \mathbb{K} inside \mathbb{S} .

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

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

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

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

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

Given the simplicial complex form of the parameter spaces, which are non linear, the notion of an effective extreme point essentially picks the extreme points that can be maximizers over those spaces of linear real functions defined on \mathbb{S} . Obviously, Assumption 4 implies that $e_{\mathbb{L}}(\mathbb{L})$ is finite. Furthermore, if \mathbb{K} lies in the interior of \mathbb{L} , $e_{\mathbb{L}}(\mathbb{K}) = \emptyset$.

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

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

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

$$ch_{\mathbb{M}}(\mathbb{N}) \equiv \sum_{\lambda \in e_{\mathbb{M}}(\mathbb{N})} \sum_{s \in ch(\lambda)} \frac{(n - c_{\mathbb{M}}(s, \lambda))!}{n!}.$$

Essentially, $\frac{(n - c_{\mathbb{M}}(s, \lambda))!}{n!}$ counts the proportion of linear real functions such that, when maximized over \mathbb{S} , have a unique maximizer at s , and when the maximization is restricted at \mathbb{M} , their unique maximizer becomes λ . Hence, $ch_{\mathbb{M}}(\mathbb{N})$ counts the

proportion of such functions for which, when maximized over \mathbb{M} , this occurs at an extreme point of \mathbb{N} . This directly implies that $ch_{\mathbb{M}}(\mathbb{M}) = 1$, and that, from Assumption 4, $ch_{\mathbb{L}}(\mathbb{K}) \leq 1$. 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 those combinatorial considerations and our assumptions, 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 $ch_{\mathbb{L}}(\mathbb{K})$, and via the subsequent 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 through standard arguments from subsampling theory (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,*

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

2. *If \mathbf{H}_0 is true, ξ_{∞} is non-constant, and $\alpha < 1 - ch_{\mathbb{L}}(\mathbb{K})$, then,*

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

3. If \mathbf{H}_a is true, then,

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

When the distribution of ξ_∞ is degenerate, then the procedure is asymptotically conservative even if the restriction $\alpha < 1 - ch_{\mathbb{L}}(\mathbb{K})$ does not hold. This is reminiscent of the results in Linton, Maasoumi and Whang (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 $\alpha < 1 - ch_{\mathbb{L}}(\mathbb{K})$ 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 spanning null hypothesis. We could handle that by breaking up \mathbb{K} in "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 asymptotically conservative and consistent.

4 A Numerical Implementation

In this section, we 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 MIP and LP. For each T , let $A_i^{(T)}$ denote a finite subset of A_i for each i . Then consider the test statistic defined by

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

and modify the algorithm of Subsection 3.2 by using ξ_T^* in place of ξ_T . Under the previous assumption framework if, as $T \rightarrow +\infty$, $A_i^{(T)}$ 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 $T \rightarrow +\infty$, $A_i^{(T)}$ 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

$$\xi_T^* = \max_{i=1,2} \sup_{z \in A_i^{(T)}} \sup_{\lambda \in \mathbb{L}} \inf_{\kappa \in \mathbb{K}} \frac{1}{\sqrt{T}} \sum_{t=1}^T d_i(z, \lambda, \kappa, Y_t), \quad (2)$$

where

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

$v(z, \lambda, \kappa, Y) \equiv K(z, \lambda, \kappa, Y) - K(0, \lambda, \kappa, Y)$, and $K(z, \lambda, \kappa, Y) \equiv (z - \kappa^{Tr} Y)_+ - (z - \lambda^{Tr} Y)_+$. From the finiteness of $A_i^{(T)}$, $i = 1, 2$, the non trivial parts of the

optimizations involved concern the $n_{i,T} \equiv \sup_{\lambda \in \mathbb{L}} \inf_{\kappa \in \mathbb{K}} \frac{1}{\sqrt{T}} \sum_{t=1}^T d_i(z, \lambda, \kappa, Y_t)$. Furthermore,

$$n_{1,T} = \inf_{\kappa \in \mathbb{K}} \frac{1}{\sqrt{T}} \sum_{t=1}^T (z - \kappa^{Tr} Y)_+ - \inf_{\lambda \in \mathbb{L}} \frac{1}{\sqrt{T}} \sum_{t=1}^T (z - \lambda^{Tr} Y)_+,$$

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

There is a set of at most T values, say $\mathcal{R} = \{r_1, r_2, \dots, r_T\}$, 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:

$$\begin{aligned} \min \quad & \sum_{t=1}^T W_t \\ \text{s.t.} \quad & W_t \geq r - \kappa^{Tr} Y_t, \quad \forall t \in T \\ & e^{Tr} \kappa = 1, \\ & \kappa \geq 0, \\ & W_t \geq 0, \quad \forall t \in T. \end{aligned}$$

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

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

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 $\mathcal{R}' = \{r'_1, r'_2, \dots, r'_T\}$, 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:

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

Hence, the computational cost of the implementation above consists of $\text{card}A_1$ LP problems, $\text{card}A_2$ MIP problems, and three trivial optimizations.

Finally and although the tests above have asymptotically correct size, we expect 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. (2017). 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 . Then, we estimate the intercept and slope of the following regression line using OLS regression analysis:

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

Finally, we estimate the bias-corrected $(1 - \alpha)$ -quantile as the OLS predicted value for $b_T = T$: $q_T^{BC}(1 - \alpha) := \hat{\gamma}_{0;T,1-\alpha} + \hat{\gamma}_{1;T,1-\alpha}(T)^{-1}$. Since $q_{T,b_T}(1 - \alpha)$ converges in probability to $q(\xi_\infty, 1 - \alpha)$ and $(b_T)^{-1}$ converges to zero as $T \rightarrow \infty$, $\hat{\gamma}_{0;T,1-\alpha}$ converges in probability to $q(\xi_\infty, 1 - \alpha)$, and the asymptotic properties are not affected.

5 Monte Carlo Study

We now design and perform a set of Monte Carlo experiments to evaluate the size and power of the proposed tests in finite samples. We do so in a framework of conditional heteroskedasticity that is often consistent with empirical findings on returns

of financial data and relevant to the empirical application that follows. The $(Y_t)_{t \in \mathbb{Z}}$ process is constructed as a vector GARCH(1,1) process that also contains an appropriately transformed element. Under the relevant restrictions, this allows for both temporal as well as cross sectional dependence between the random variables stacked in the vector process.

Suppose that $z_t \stackrel{\text{iid}}{\sim} N(0, 1)$, $t \in \mathbb{Z}$. Furthermore, for all $t \in \mathbb{Z}$, for $i = 1, 2, 3$, $\omega_i, \alpha_i, \beta_i \in \mathbb{R}_{++}$, $\mu_i \in \mathbb{R}_+$, define

$$\begin{aligned} y_{it} &= \mu_i + z_t h_{it}^{1/2}, \\ h_{it} &= \omega_i + (\alpha_i z_{t-1}^2 + \beta_i) h_{i,t-1}, \mathbb{E}(\alpha_i z_0^2 + \beta_i)^{1+\epsilon} < 1, \end{aligned}$$

for some $\epsilon > 0$, while, for $i = 4$ and $v_1, v_2 \in \mathbb{R}$, define

$$y_{4t} = v_1 \left(z_t h_{3t}^{1/2} \right)_+ + v_2 \left(z_t h_{3t}^{1/2} \right)_-.$$

Suppose that $\mathbf{Y}_t = (y_{1t}, y_{2t}, y_{3t}, y_{4t})^{Tr}$. Arvanitis and Topaloglou (2017) establish that the vector process above satisfies our assumption framework. Let $\tau = (0, 0, 1, 0)$, $\tau^* = (0, 0, 0, 1)$, and $\mathbb{L} = \{(\lambda, 1 - \lambda, 0, 0), \lambda \in [0, 1], \tau, \tau^*\}$. Using this portfolio space, we obtain the following result on Markowitz-spanning. Its proof follows directly from Proposition 4 of Arvanitis and Topaloglou (2017), and it essentially depends on the fact that τ^* is a Markowitz super-efficient portfolio w.r.t. the portfolio space.

Proposition 3. *If $\mu_i = 0$ for $i = 1, 2, 3$, $|v_1| > \sqrt{\frac{\max\{\omega_i, \alpha_i, \beta_i, i=1,2,3\}}{\min\{\omega_i, \alpha_i, \beta_i, i=1,2,3\}}}$ and $|v_2| < \sqrt{\frac{\min\{\omega_i, \alpha_i, \beta_i, i=1,2,3\}}{\max\{\omega_i, \alpha_i, \beta_i, i=1,2,3\}}}$, then $\mathbb{K} := \{(\lambda, 1 - \lambda, 0, 0), \lambda \in [0, 1]\} \cup \{\tau^*\}$ Markowitz-spans \mathbb{L} , while $\mathbb{K} - \{\tau^*\}$ does not Markowitz-span \mathbb{L} .*

We use instances of the GARCH processes conforming to Proposition 3 above in order to evaluate the empirical size and power for given T s. For a fixed sample size T , we test for MSD spanning under the null hypothesis to gauge the empirical size. We set $\mu_i = 0$ for $i = 1, 2, 3$, $\omega_1 = 0.5$, $\omega_2 = 0.5$, and $\omega_3 = 0.5$, $a_1 = 0.4$, $a_2 = 0.45$, and $a_3 = 0.5$ and $\beta_1 = 0.5$, $\beta_2 = 0.45$, $\beta_3 = 0.4$, $v_1 = 1.5$ and $v_2 = 0.5$. In this case, we have that $|v_1| > \sqrt{\frac{\max\{\omega_i, a_i, \beta_i, i=1,2,3\}}{\min\{\omega_i, a_i, \beta_i, i=1,2,3\}}}$ and $|v_2| < \sqrt{\frac{\min\{\omega_i, a_i, \beta_i, i=1,2,3\}}{\max\{\omega_i, a_i, \beta_i, i=1,2,3\}}}$. To

get the empirical power, we test for MSD spanning under the alternative hypothesis by setting $\mu_i = 0$ for $i = 1, 2, 3$, $\omega_1 = 0.5$, $\omega_2 = 0.5$, and $\omega_3 = 0.8$, $a_1 = 0.3$, $a_2 = 0.4$, and $a_3 = 0.45$ and $\beta_1 = 0.3$, $\beta_2 = 0.4$, $\beta_3 = 0.45$, $v_1 = 2$ and $v_2 = 0.2$. In this case, we have that $\omega_1 < \omega_3$, $a_1 < a_3$ and $\beta_1 < \beta_3$.

We present our Monte Carlo results in Table 1. The number of Monte Carlo replications to compute the empirical size and power is 1000 runs. We use three cases. In the first case $T = 300$, we get the subsampling distribution of the test statistic for subsample sizes $b_T \in [50, 100, 150, 200]$. In the second case, $T = 500$ and $b_T \in [100, 200, 300, 400]$. Finally, in the third case, $T = 1000$ and $b_T \in [120, 240, 360, 480]$. The test seems to perform well in all cases with an empirical size close to 5% and an empirical power above 90% for a nominal size $\alpha = 5\%$.

Cases	$T=300$	$T=500$	$T=1000$
Empirical size	4.4%	3.6%	4.8%
Empirical power	93.7%	92.5%	96.2%

Table 1: Monte Carlo Results. Entries report the empirical size and empirical power based on 1000 replications and a nominal size $\alpha = 5\%$.

6 Empirical Applications

In the empirical applications, we first analyze whether the market portfolio is MSD efficient. Representative investor models of capital market equilibrium predict that the market portfolio is efficient as a result of risk sharing in sufficiently complete markets, or aggregation across sufficiently homogeneous investors in incomplete markets. Alternatively, we can interpret a market portfolio efficiency test as a revealed preference analysis of those individual investors who adopt a passive strategy of broad diversification.

In this application, \mathbb{L} consists of all convex combinations of the market portfolio, the T-bill, and the set of the 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. To test market portfolio MSD efficiency, we use the Arvanitis and Topaloglou (2017) test.

We also test 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 third 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 $\{(0, \dots, 1)\}$, 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 homepage¹³ of Kenneth French. The test portfolio is the Fama and French market portfolio, which is the value-weighted average of all non-financial 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

¹³ <http://mba.turc.dartmouth.edu/pages/faculty/ken.french>

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).

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 Efficiency Test

We first analyze whether the market portfolio is MSD efficient. This hypothesis seems interesting because representative investor models of capital market equilibrium predict that the market portfolio is efficient as a result of risk sharing in sufficiently complete markets or, alternatively, aggregation across sufficiently homogeneous investors in incomplete markets. We can interpret a market portfolio efficiency test as a revealed preference analysis of those individual investors who adopt a passive strategy of broad diversification. If the market is not efficient with respect to the MSD, investors with MSD preferences could outperform the market using equity management strategies. To test market portfolio MSD efficiency, we use the Arvanitis and Topaloglou (2017) test, using the same subsampling procedure as our spanning test. We divide the full period into 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. We test for MSD of the market portfolio in the whole period, as well as in each subperiod, where the superscripts 1 and 2 indicate the subperiod.

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 market efficiency if the Arvanitis and Topaloglou (2017) test statistic p_T is higher than the regression estimate q_T .

- **The 6 FF benchmark portfolios:** The regression estimate $q_T = 0.17$ is lower than the full-sample value of the test statistic $p_T = 0.35$. Moreover, the regression estimate in the first subperiod $q_T^1 = 0.16$ is lower than the sub-sample test statistic $p_T^1 = 0.49$, and the regression estimate in the second subperiod $q_T^2 = 0.13$ is lower than the sub-sample test statistic $p_T^2 = 0.46$.
- **The 10 momentum portfolios:** The regression estimate $q_T = 0.24$ is lower than the full-sample value of the test statistic $p_T = 0.51$. Moreover, $q_T^1 = 0.29$ is lower than the sub-sample test statistic $p_T^1 = 0.46$, and $q_T^2 = 0.21$ is lower than the sub-sample test statistic $p_T^2 = 0.42$.
- **The 10 industry portfolios:** The regression estimate $q_T = 0.30$ is lower than the full-sample value of the test statistic $p_T = 0.53$. Moreover, $q_T^1 = 0.25$ is lower than the sub-sample test statistic $p_T^1 = 0.57$, and $q_T^2 = 0.28$ is lower than the sub-sample test statistic $p_T^2 = 0.59$.
- **The 10 size decile portfolios:** The regression estimate $q_T = 0.25$ is lower than the full-sample value of the test statistic $p_T = 0.63$. Moreover, $q_T^1 = 0.24$ is lower than the sub-sample test statistic $p_T^1 = 0.54$, and $q_T^2 = 0.29$ is lower than the sub-sample test statistic $p_T^2 = 0.61$.
- **The 10 beta decile portfolios:** The regression estimate $q_T = 0.17$ is lower than the full-sample value of the test statistic $p_T = 0.45$. Moreover, $q_T^1 = 0.11$ is lower than the sub-sample test statistic $p_T^1 = 0.43$, and $q_T^2 = 0.15$ is lower than the sub-sample test statistic $p_T^2 = 0.48$.

The results provide evidence in favor of the claim that the market portfolio is MSD inefficient. 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.

6.2 Results of the MSD Spanning Test

Since we find that the market is MSD inefficient, our second 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.

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 ρ_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 $\rho_T = 26.78$.
- **The 10 momentum portfolios:** The regression estimate $q_T = 19.42$ is lower than the value of the test statistic $\rho_T = 41.55$.
- **The 10 industry portfolios:** The regression estimate $q_T = 22.46$ is lower than the value of the test statistic $\rho_T = 31.74$.

- **The 10 size decile portfolios:** The regression estimate $q_T = 19.62$ is lower than the value of the test statistic $\rho_T = 32.34$.
- **The 10 beta decile portfolios:** The regression estimate $q_T = 31.48$ is lower than the value of the test statistic $\rho_T = 44.76$.

The results suggest the rejection of MSD spanning and of the two-fund separation theorem for MSD investors.

As a final step in our 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.

6.3 Performance Summary of the MSD portfolio

Finally, we analyze the performance of the optimal MSD portfolios through time, compared to the performance of the market portfolio.

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. The clock is advanced and we determine the realized return of the MSD portfolio from the actual returns of the assets picked by the optimizer. Then we repeat the same procedure for the next time period and compute the ex post realized returns over the period from 07/1963 to 12/2016.



Figure 1: Cumulative Returns.

Figure 1 illustrates the cumulative performance of the MSD optimal portfolio as well as the market portfolio for the entire sample period from July 1963 to December 2016. We observe that the value of the MSD optimal portfolio is 426 times higher at the end of the holding period compared to the initial value, while, for the market

portfolio, is it only higher 13.9 times. Not surprisingly, the relevant performance of MSD type investors is 30 times higher than the performance of the market. The optimal MSD portfolio includes industry portfolios (telecommunications, health, energy and utilities), size portfolios (small caps), beta sorted portfolios (low and medium beta), and momentum portfolios, in addition to the market portfolio and the T-Bill.

	MSD optimal portfolio	market portfolio
Performance Measures		
Mean	0.01035	0.00510
Standard deviation	0.04290	0.04420
Skewness	-0.27730	-0.52629
Excess Kurtosis	1.18535	1.96705
Sharpe ratio	0.17495	0.04697
VaR 5%	0.06133	0.0718

Table 2: Parametric portfolio measures. Entries report the performance and risk measures (mean, volatility, skewness, excess kurtosis, sharpe ratio, empirical VaR 5% (positive sign for a loss)) for the MSD optimal portfolio and the market portfolio. The dataset spans the period from July 31, 1963 to December 31, 2016.

Table 2 reports the performance and risk measures for the MSD optimal portfolio and the market portfolio. These measures supplement the evidence obtained from the previous graph. 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 wants 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

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.

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

$$R_{it} - R_{Ft} = a_i + b_i(R_{Mt} - R_{Ft}) + s_iSMB_t + h_iHML_t + r_iMOM_t + e_{it},$$

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.

The four-factor model						
	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	
Statistics						
	Adj. R^2	F-statistic	p -value			
	0.948	2.953	0			

Table 3: Carhart four-factors. Entries report the coefficients and their respective t -statistics, Adjusted R2, F-statistic, and p -values. The dataset spans 1963-2016.

Table 3 reports the coefficient estimates of the four factors, as well as their respec-

tive 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 portfolio. The intercept is not zero, which indicates that perhaps other factors drive the performance of the MSD portfolio as well.

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

$$R_{it} - R_{Ft} = a_i + b_i(R_{Mt} - R_{Ft}) + s_iSMB_t + h_iHML_t + r_iRMW_t + c_iCMA_t + e_{it},$$

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.

The five-factor model						
	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
Statistics						
	Adj. R^2	F-statistic	p -value			
	0.988	5361.6	0			

Table 4: Fama-French five factors. Entries report the coefficient estimates, their respective t -statistics, Adjusted R2, F-statistic, and p -values. The dataset spans 1963-2016.

Table 4 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 portfolio as well.

In both factor models, we observe that the beta market is slightly smaller than one (defensive) for the MSD portfolio 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 thinner 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.

References

- [1] Arvanitis, S., Hallam, M. S., Post, T. and N. Topaloglou. 2017. Stochastic spanning. Forthcoming in the *Journal of Business and Economic Statistics* (<http://dx.doi.org/10.1080/07350015.2017.1391099>).
- [2] Arvanitis, S., and N. Topaloglou. 2017. Testing for prospect and Markowitz stochastic dominance efficiency. *Journal of Econometrics* 198(2), 253-270.
- [3] Banz, Rolf W. 1981. The relationship between return and market value of common stocks. *Journal of Financial Economics* 9(1), 3-18.
- [4] Baucells, M., and F.H. Heukamp. 2006. Stochastic dominance and cumulative prospect theory. *Management Science* 52, 1409-1423.
- [5] 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.

- [6] M.E. Blume and I. Friend. 1973. A new look at the Capital Asset Pricing Model. *Journal of Finance* 28(1), 19-34.
- [7] Carhart, M. 1997. On persistence in Mutual Fund Performance. *Journal of Finance* 52(1), 57-82.
- [8] 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.
- [9] Cortissoz, J. 2007. On the Skorokhod representation theorem. *Proceedings of the American Mathematical Society* 135(12), 3995-4007.
- [10] Edwards, K.D. 1996. Prospect theory: A literature review, *International Review of Financial Analysis* 5, 18-38
- [11] Fama, E. and K. French. 1992. The Cross-Section of Expected Stock Returns. *Journal of Finance* 47(2), 427-465.
- [12] Fama, E. and K. French. 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33, 3-56.
- [13] Fama, E. and K. French. 1997. Industry costs of equity. *Journal of Financial Economics* 43, 153-193.
- [14] Fama, E. and K. French. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116, 1-22.
- [15] Fama, E.F. and J.D. MacBeth. 1973. Risk, return and equilibrium: empirical tests. *The Journal of Political Economy* 81, 607-636.
- [16] Francq, C., and J. M. Zakoian. 2011. GARCH models: structure, statistical inference and financial applications. John Wiley & Sons.

- [17] Friedman, M., and L. J. Savage. 1948. The utility analysis of choices involving risk. *Journal of Political Economy* 56, 279-304.
- [18] Gonzalo, J. and J. Olmo. 2014. Conditional Stochastic Ddominance Tests in Dynamic Settings. *International Economic Review* 55(3), 819-838.
- [19] Hadar, J. and W.R. Russell. 1969. Rules for ordering uncertain prospects. *American Economic Review* 59, 2-34.
- [20] Hanoch, G., and H. Levy. 1969. The efficiency analysis of choices involving risk. *Review of Economic Studies* 36, 335-346.
- [21] 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.
- [22] Huberman, G. and S. Kandel. 1987. Mean-Variance Spanning. *Journal of Finance* 42, 873-888.
- [23] Knight, K. 1999. Epi-convergence in distribution and stochastic equi-semicontinuity. Working Paper, Department of Statistics, University of Toronto.
- [24] 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
- [25] Kuosmanen, T. (2004). Efficient diversification according to stochastic dominance criteria. *Management Science* 50(10), 1390-1406.
- [26] Levy, H. 1992. Stochastic Dominance and Expected Utility: Survey and Analysis. *Management Science* 38, 555-593.
- [27] Levy, H. 2015. *Stochastic dominance: Investment decision making under uncertainty*. Springer.

- [28] M. Levy and H. Levy. 2002. Prospect Theory: Much Ado about Nothing?. *Management Science* 48, 1334-1349.
- [29] Levy, H., M. Levy. 2004. Prospect theory and mean-variance analysis. *Review of Financial Studies* 17(4), 1015-1041.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] Molchanov, I. 2006. *Theory of random sets*. Springer Science and Business Media.
- [35] Mosler, K., and M. Scarsini. 1993. *Stochastic Orders and Applications, a Classified Bibliography*, Berlin: Springer-Verlag.
- [36] Narici, L., and E. Beckenstein. 2010. *Topological vector spaces*. CRC Press.
- [37] Novy-Marx, R. 2016. Understanding defensive equity. Working Paper, Simon Graduate School of Business, University of Rochester and NBER.
- [38] Nualart, D. 2006. *The Malliavin calculus and related topics*. Berlin: Springer.
- [39] Politis, D. N., J. P. Romano and M. Wolf. 1999. *Subsampling*. Springer New York.

- [40] Post, T. 2003. Empirical Tests for Stochastic Dominance Efficiency. *The Journal of Finance* 58: 1905-1931.
- [41] Post, T., and M. Kopa. 2013. General linear formulations of stochastic dominance criteria. *European Journal of Operational Research* 230(2), 321-332.
- [42] 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.
- [43] M.R. Reinganum. 1981. A new empirical perspective on the CAPM. *Journal of Financial and Quantitative Analysis* 16(4), 439-462.
- [44] Rio, E. 2013. Inequalities and limit theorems for weakly dependent sequences. 3^eme cycle. pp.170. <cel-00867106>.
- [45] Rothschild, M. and J.E. Stiglitz. 1970. Increasing Risk: I. A definition. *Journal of Economic Theory* 2(3), 225-243.
- [46] Scaillet, O., and N. Topaloglou. 2010. Testing for stochastic dominance efficiency. *Journal of Business and Economic Statistics* 28(1), 169-180.
- [47] Sidak, Z., P. K. Sen, and J. Hajek. 1999. *Theory of rank tests*. Academic Press.
- [48] Tobin, J. 1958. Liquidity Preference as Behavior Towards Risk. *Review of Economic Studies* 25, 65-86.
- [49] van der Vaart, A. W., and J. A. Wellner. 1996. *Weak Convergence*. Springer New York.

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 $\xi_n \equiv \text{oper}X_\lambda$, where opt_i is considered w.r.t. $\Lambda_{i,n}^*(\lambda_{i-1}) = \{\text{the first } n \text{ elements of } \Lambda_i^*(\lambda_{i-1}) \cap \text{pr}_i\Lambda_\infty\}$ and $\lambda_{i-1} \in \Lambda_{i-1,n}^*$ when $i > 1$. The function $\text{oper} : C(\Lambda, \mathbb{R}) \rightarrow \mathbb{R}$ 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, $\xi_n \rightarrow \xi$ 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

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

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

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

define $\xi'_n = \text{oper}X_\lambda$ analogously. We have that as $n \rightarrow \infty$ $\xi'_n \rightarrow \xi'$ 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 $D\xi'_n \rightarrow D\xi'$ 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 $\langle D\xi'_n - D\xi', h \rangle > \frac{1}{k}$ 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 $X^\star \equiv X |_{\Omega - \cup_{\tau \in \mathcal{T}} \Omega_\tau^c}$ that obviously satisfies Assumption 1 with $\mathcal{T}^\star = \emptyset$ along with the obvious change of notation. Hence, ξ^\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 $X_\star \equiv X |_{\Omega_{\tau_1, \tau_2}}$ satisfies Assumption 1 with $\mathcal{T}_\star = \emptyset$, and thereby ξ_\star has an absolutely continuous law. The other cases follow analogously when the intersections appearing in the theorem are non empty. When empty the results are trivial. \square

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

Proof of Proposition 1. (\Leftarrow) If $\mathbb{K} \succ_M \mathbb{L}$, 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

$$\max_{i=1,2} \sup_{z \in A_i} \inf_{\kappa \in \mathbb{K}} \Delta_i(z, \lambda, \kappa, F) \leq 0. \quad (4)$$

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

κ^* be an element of the latter. Then, the first equality follows from

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

If $\mathbb{K} \not\subseteq_M \mathbb{L}$ for some $\lambda^* \in \mathbb{L}$, and any $\kappa \in \mathbb{K}$, there exists some $i(\lambda^*, \kappa)$, $z^*(\lambda^*, \kappa) \in A_i$ 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

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

(\Rightarrow) Suppose now that $\xi(F) = 0$ and consider an arbitrary λ . This implies that (4) 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 $\lambda^* \in \mathbb{L}$, and some $i = 1, 2$, $\inf_{\kappa \in \mathbb{K}} \sup_{z \in A_i} \Delta_i(z, \lambda^*, \kappa, F) > 0$. It implies that for any $\kappa \in \mathbb{K}$, $\sup_{z \in A_i} \Delta_i(z, \lambda^*, \kappa, F) > 0$ and the result follows. \square

Proof of Proposition 2. The results in the auxiliary Lemma 1 imply that $\left(\begin{array}{c} \Delta_1 \left(z_1, \lambda, \kappa, \sqrt{T} (F_T - F) \right) \\ \Delta_2 \left(z_2, \lambda, \kappa, \sqrt{T} (F_T - F) \right) \end{array} \right)$ weakly converges to $\left(\begin{array}{c} \Delta_1(z_1, \lambda, \kappa, \mathcal{G}_F) \\ \Delta_2(z_2, \lambda, \kappa, \mathcal{G}_F) \end{array} \right)$ 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 $(z_{1,T}, z_{2,T}, \lambda_T) \rightarrow (z_1, z_2, \lambda)$, there exists an enhanced probability space and processes

$$\left(\begin{array}{c} \Delta_{1,T}(\kappa) \\ \Delta_{2,T}(\kappa) \end{array} \right) \stackrel{d}{=} \left(\begin{array}{c} \Delta_1 \left(z_{1,T}, \lambda_T, \kappa, \sqrt{T} (F_T - F) \right) \\ \Delta_2 \left(z_{2,T}, \lambda_T, \kappa, \sqrt{T} (F_T - F) \right) \end{array} \right), \left(\begin{array}{c} \Delta_1^*(\kappa) \\ \Delta_2^*(\kappa) \end{array} \right) \stackrel{d}{=} \left(\begin{array}{c} \Delta_1(z_1, \lambda, \kappa, \mathcal{G}_F) \\ \Delta_2(z_2, \lambda, \kappa, \mathcal{G}_F) \end{array} \right),$$

defined on it such that $\begin{pmatrix} \Delta_{1,T} \\ \Delta_{2,T} \end{pmatrix} \rightarrow \begin{pmatrix} \Delta_1^* \\ \Delta_2^* \end{pmatrix}$ almost surely, w.r.t. to the product topology of epi-convergence, where $\stackrel{d}{=}$ denotes equality in distribution. Notice that,

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

Under \mathbf{H}_0 , due to the previous, we have that for any $i = 1, 2$, $\kappa, \kappa_T \in \mathbb{K}$, and $\kappa_T \rightarrow \kappa$, $\lim_{T \rightarrow \infty} K_{i,T}(\kappa_T)$ is almost surely equal to

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

Furthermore, for any compact \mathcal{K}_i that contains $\kappa \in \mathbb{K}$ such that $(z_{i,T}, \lambda_T, \kappa)$ eventually belongs to the boundary of Γ_i , we have that almost surely,

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

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 $K_i(\kappa) = \begin{cases} \Delta_i^*(\kappa), & (z_i, \lambda, \kappa) \in \Gamma_i \\ -\infty, & (z_i, \lambda, \kappa) \notin \Gamma_i \end{cases}$.

Since \mathbb{K} is compact, Theorem 3.4 (ch. 5, p. 338) of Molchanov (2006) implies that almost surely,

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

jointly over $i = 1, 2$. When Γ_i is not empty, by Theorem 7.11 of Rockafellar and Wets (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,

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

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. \square

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 $\lambda^* \in \mathbb{L} - \mathbb{K}$, and any $\kappa \in \mathbb{K}$, there exists some $i, z^* \in A_i$ such that $\Delta_i(z, \lambda, \kappa, F) > 0$. Then, we have that

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

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 . \square

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). \square

Proof of Proposition 3. From Proposition 4 of Arvanitis and Topaloglou (2017), we have that τ^* strictly Markowitz dominates every portfolio in \mathbb{L} . Hence $\mathbb{K} \succ_M \mathbb{L}$ and thereby $\mathbb{K} \supset \{\tau^*\} \succ_M \mathbb{L}$. By the same reasoning, there is no element in $\mathbb{K} - \{\tau^*\}$ that Markowitz dominates τ^* . Hence $\mathbb{K} - \{\tau^*\}$ cannot Markowitz-span \mathbb{L} . \square

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*

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

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

Proof. Let $\theta \equiv (\lambda, \kappa, z_1, z_2) \in \Theta \equiv \mathbb{L} \times \mathbb{K} \times \mathbb{R}_- \times \mathbb{R}_{++}$, ρ any non zero element of \mathbb{R}^2 , and consider $\Delta(\theta, \cdot) \equiv \rho_1 \Delta_1(z_1, \lambda, \kappa, \cdot) + \rho_2 \Delta_2(z_2, \lambda, \kappa, \cdot)$. 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 $f_T^* \stackrel{d}{=} \sqrt{T}(F_T - F)$, $f^* \stackrel{d}{=} \mathcal{G}_F$, and $f_T^* \rightarrow f^*$ a.s.. Let $J \equiv \overline{\text{span}} \{f_T^*, f^*, T = 1, 2, \dots\}$ equipped with the metrizable topology of weak convergence.¹⁴ Consider $\Delta(\cdot, \cdot)$ restricted to J with values in the linear space of stochastic processes, equipped with

¹⁴Here $\overline{\text{span}}$ denotes the closure w.r.t. the particular topology of the linear span.

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, Remark 3.2, Corollary 4.1, and Theorem 7.3 of Rio (2013), we also have that

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

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 $\sup_{\theta} \mathbb{E} [(x - y)^2]$, 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 $\Delta(\theta, \sqrt{T}(F_T - F)) \rightsquigarrow \Delta(\theta, \mathcal{G}_F)$. This and the Cramer-Wold Theorem imply the needed result. The final assertion follows from $\sup_{\theta \in \Theta} \mathbb{E} [(\Delta(\theta, \mathcal{G}_F))^2] < +\infty$, the discussion in Example 1.5.10 of van der Vaart and Wellner (1996), and the continuity of $\mathbb{E} [(\Delta(\theta, \mathcal{G}_F))^2]$ w.r.t. θ . \square

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

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

From the zero mean Gaussianity of the processes involved, Remark 3.2, 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 $\Delta_i(z, \lambda, \kappa, \mathcal{G}_F)$ 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

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

where $Z \sim N(0_{n \times 1}, \mathbb{V})$. 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 $\mathbb{P}(\eta_\infty = 0)$. 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 $N(0_{n \times 1}, \mathbb{V})$, along with Definition 4, we get $\mathbb{P}(\eta_\infty = 0) \leq ch_{\mathbb{L}}(\mathbb{K})$. \square