The *magnificent ten* equity factor model

Argyro Kofina[1] Ioannis Psaradellis[2] Kuntara Pukthuanthong[3]

Nikolas Topaloglou[4]

March 2025

[1]Athens University of Economics and Business, 76, Patision Street, GR10434, Athens, Greece; e-mail: kofina.a@aueb.gr.

[2]University of Edinburgh, University of Edinburgh Business School; e-mail: ipsarade@ed.ac.uk.

[3]University of Missouri; e-mail: pukthuanthongk@missouri.edu.

[4]IPAG Business School & Athens University of Economics and Business, 76, Patision Street, GR10434, Athens, Greece; e-mail: nikolas@aueb.gr.

Abstract

We propose a novel asset pricing factor model employing a new estimation method based on sparse second-order stochastic dominance (SSD), implemented via a greedy algorithm combined with linear programming. Initially drawing from a prominent set of 24 candidate factors featured extensively in the asset pricing literature, we find statistically and economically that no incremental benefit emerges from including more than ten factors. To further ensure robustness and validate our results, we extend our factor universe to a comprehensive set of 177 factors—153 factors from Jensen et al. (2023) plus the original 24. Remarkably, our empirical results again demonstrate that a ten-factor SSD-based sparse model remains optimal and consistently outperforms all leading benchmarks. This SSD-based sparse factor model significantly generalizes the mean-variance Arbitrage Pricing Theory (APT) framework by explicitly accommodating broader classes of investor preferences, thus providing enhanced theoretical rigor and clear economic interpretation. Comprehensive parametric and non-parametric tests, conducted in-sample and out-of-sample on cross-sectional and time-series equity returns, reveal that our proposed model consistently dominates prominent benchmark models, including LASSO-based machine learning approaches. These robust findings emphasize the considerable advantages of explicitly integrating investor welfare considerations into factor selection, thereby delivering rigorous theoretical insights and practical asset pricing relevance.

Keywords: Equities; Asset pricing; Sparse Stochastic dominance; Stochastic Bounding; Machine learning.

1. Introduction

Over the past decades, hundreds of factors have been proposed to explain the cross-section of equity returns. However, Ross (1976) conjectures that only a limited number of economically meaningful factors drive most cross-sectional variation in returns. This view is further reinforced by recent concerns about the proliferation of the "factor zoo" and issues of statistical overfitting, such as "p-hacking" (e.g., Cochrane, 2011; Hou et al., 2018; Feng et al., 2020; Harvey and Liu, 2021). Consequently, the literature has converged on only a few widely accepted seminal models—such as the four-factor model of Carhart (1997), the five-factor model of Fama and French (2015), the Hou et al. (2020) q-factor model, and the Barillas and Shanken (2018) factor set. This naturally leads to an important question: Can we construct a sparse portfolio of fewer factors from these prominent candidate factor sets that more effectively explains equity returns?

This paper employs a Stochastic Dominance (SD) framework, specifically the Sparse Spanning methodology developed by Arvanitis et al. (2025), to identify the dominant subset of factors from an extensive pool of candidate factors. These factors include those in prominent factor models (e.g., Fama and French, 1993, 2015; Barillas and Shanken, 2018; Hou et al., 2020; Daniel et al., 2020) as well as additional factors based on stock characteristics documented to have explanatory power for the cross-section of equity returns (see, e.g., Moskowitz et al., 2012; Asness and Frazzini, 2013; Frazzini and Pedersen, 2014).

We explicitly adopt second-order stochastic dominance (SSD) because it precisely

captures the preferences of all risk-averse investors without requiring additional restrictive assumptions. This explicit choice contrasts clearly with first-order stochastic dominance (FSD), which is typically too stringent and unrealistic for practical portfolio selection, and third-order stochastic dominance (TSD), which explicitly incorporates assumptions about decreasing absolute risk aversion and investor preferences toward skewness. Thus, SSD offers a robust yet economically meaningful middle ground, broadly applicable to all concave (risk-averse) utilities, and implicitly accounting for investor concerns about higher-order risks such as skewness and kurtosis.

Applying this SSD-based Sparse Spanning criterion, we construct a sparse portfolio composed of a limited number of factors, demonstrating that no further welfare improvement for risk-averse investors can be gained by adding additional factors. Specifically, our analysis shows no incremental benefits from expanding the sparse factor set beyond ten factors. The SD-Sparse methodology is firmly anchored in investor welfare, ensuring the resulting factor model excludes no factor that any risk-averse investor would find beneficial.

Although our testing and factor-selection methodology explicitly relies only on second-order stochastic dominance (SSD), it provides greater economic depth and theoretical rigor compared to traditional mean-variance frameworks. This advantage arises because SSD accommodates a broad spectrum of risk-averse investor preferences and implicitly captures investor aversion to higher-order risks such as skewness and kurtosis, rather than relying solely on mean and variance.

Empirically, we observe that optimized portfolios constructed from abnormal re-

turns using our proposed SD-Sparse factor model closely mirror the behavior of a risk-free asset over the same period, achieving nearly identical average returns, standard deviations, and consequently, a Sharpe ratio close to zero. This result suggests our SD-Sparse factor model closely approximates the theoretical ideal of a "true factor model," as the diversified portfolio of abnormal returns—once adjusted for systematic risks—is indistinguishable from a riskless asset.

Our SD-Sparse model consistently outperforms several prominent benchmarks, including the Carhart (1997) four-factor model, the Fama and French (2015) five-factor model (with momentum), the extended Fama and French (2015) five-factor model incorporating both momentum and liquidity, the Barillas and Shanken (2018) factor set, the Hou et al. (2020) factor model, the Daniel et al. (2020) factor model, and a sparse factor-selection model using adaptive LASSO applied to the same extensive factor set. This superior performance holds robustly in both parametric and non-parametric asset pricing tests, as well as in-sample and out-of-sample analyses.

Most multi-factor models in finance rely on the mean-variance framework introduced by Huberman and Kandel (1987). Under this framework, if factors represent tradable assets, no portfolio formed from available investments should yield a Sharpe ratio exceeding that of the factor portfolios themselves. However, the mean-variance approach is economically meaningful only under restrictive assumptions about return distributions—specifically, when distributions are fully characterized by just two parameters: mean (location) and variance (scale), as in elliptical distributions like the normal distribution. When returns exhibit more complex features (such as skewness, kurtosis, or other higher-order moments), the mean-variance criterion

becomes economically less meaningful, potentially misrepresenting investors' actual risk preferences.[1]

In contrast, the Stochastic Spanning methodology significantly relaxes these restrictive assumptions. Rather than focusing exclusively on the Sharpe ratio under narrowly defined distributional assumptions, stochastic spanning explicitly addresses broader investor preferences by maximizing expected utility for investors with risk-averse (concave and increasing) utility functions. Consequently, stochastic spanning inherently captures deeper economic concerns, such as aversion to downside risk and sensitivity to higher-order moments, features typically neglected by mean-variance approaches. Additionally, because stochastic spanning does not rely on traditional regression-based methods typically used for selecting factors, it naturally mitigates issues such as selection bias—favoring factors that fit well merely due to statistical noise—and data snooping—unintentionally selecting factors that appear significant purely by chance.

Our proposed SD-Sparse criterion explicitly ensures that the selected factors span investors' marginal utilities, represented economically by the pricing kernel. Any factor omitted from our model is excluded precisely because it provides no incremental benefit for risk-averse investors—it offers no additional hedging value or improvement in consumption smoothing in states of high marginal utility (low wealth states).

Finally, the stochastic spanning framework readily accommodates realistic and economically practical constraints on portfolio weights, such as short-selling limi-

[1]While investors can explicitly incorporate skewness or kurtosis measures, SSD implicitly captures all higher-order moments simultaneously. Thus, SSD economically represents the entire set of risk-averse investor preferences without requiring arbitrary selection or parameterization of specific moments.

tations, reasonable leverage boundaries, and diversification requirements reflecting actual investment practice.[2] Incorporating these constraints ensures that resulting optimal portfolios are economically sensible, stable, and practically implementable, thereby enhancing the economic relevance and robustness of the sparse factor model constructed under the SSD framework.

We make several contributions to the existing literature. First, we propose a novel factor model for equity returns based on the Stochastic Dominance (SD) framework, extending the Arbitrage Pricing Theory (APT) tests introduced by Ross (1976). Our approach explicitly accommodates a broad class of investor preferences (general risk aversion) and captures higher-order moments beyond mean and variance. Using Sparse Spanning, we identify a parsimonious set of ten economically significant factors (denoted as SD-Sparse): the equity market factor, SMB, return on equity, time-series momentum of Moskowitz et al. (2012), the *High Minus Low* factor of Asness and Frazzini (2013), expected growth, *Quality Minus Junk* of Asness and Pedersen (2019), asset growth, gross profitability, and net operating assets-to-total assets.

Second, we perform the first comprehensive empirical exercise to systematically identify the dominant factors explaining the cross-section of equity returns, based on both established factor models and recent anomaly studies (e.g., Hou et al., 2020; Chen and Zimmermann, 2022). Unlike previous studies, which typically impose ex-

[2]While SSD itself does not inherently prevent extreme leverage, unrestricted short-selling, or excessive asset concentration, the optimization algorithms used to implement SSD factor selection (such as linear programming or greedy algorithms) explicitly allow for imposing practical constraints. Thus, optimal portfolios derived under the SSD criterion are ensured to remain economically realistic and practically implementable.

ante factor selections, we use Sparse Spanning on an extensive set of 24 candidate factors from prominent models and fundamental equity characteristics, thus allowing the data and economic theory to jointly determine the most informative factors.

Third, we revisit the effectiveness of machine learning methods commonly employed in big-data financial applications by explicitly benchmarking them against our economically-grounded SD approach. Specifically, we compare our SD-Sparse methodology to an adaptive LASSO approach similar to Chen et al. (2024), which uses shrinkage techniques for factor selection. We find that the Sparse Spanning approach consistently identifies superior explanatory factors for equity returns across both in-sample and out-of-sample tests, reflecting its theoretical robustness derived from explicitly optimizing expected utility. To our knowledge, we are the first study to propose a multifactor equity return model grounded explicitly in a stochastic dominance framework, and also the first to rigorously evaluate the robustness of Stochastic Spanning tests in- and out-of-sample.

We rigorously evaluate the SD-Sparse factor model through comprehensive in-sample and out-of-sample tests, employing both parametric and non-parametric asset pricing methodologies. Specifically, we implement the *Lucky Factors* approach of Harvey and Liu (2021) to rigorously control for statistical luck in factor selection from a large pool of candidate factors. The resulting scaled intercept statistic and multiple hypothesis testing demonstrate that the SD-Sparse factors significantly enhance explanatory power for equity returns relative to each competing benchmark model.

Further, cross-sectional tests—evaluating the factor risk premia and their corre-

sponding prices of risk using the misspecification methodology of Kan et al. (2013)—confirm the superior performance of the SD-Sparse model. Specifically, our model achieves the highest and statistically significant GLS and OLS cross-sectional values (91

We also conduct non-parametric, fully out-of-sample tests based on optimal SD portfolios constructed from abnormal returns generated by each competing factor model, using a rolling-window optimization framework. The optimized portfolio derived from the SD-Sparse abnormal returns exhibits minimal performance, closely mirroring the risk-free rate—consistent with the behavior expected from a correctly specified factor model.

Finally, we assess out-of-sample predictive performance by forecasting the returns of an equally weighted equity portfolio using linear regression models with expanding estimation windows. The SD-Sparse model consistently produces the lowest prediction errors. Additionally, Diebold-Mariano (Diebold and Mariano, 1995) t-statistics confirm that SD-Sparse forecasts significantly outperform all benchmark factor models in predicting future equity returns.

Our study contributes to several strands of existing literature. A rich body of research explores candidate factors for explaining the cross-sectional variation in equity portfolio returns (e.g., Carhart, 1997; Berk and van Binsbergen, 2016; Feng et al., 2020; Giglio and Xiu, 2021). For example, Berk and van Binsbergen (2016) propose a method to identify which factor model equity investors implicitly use from a universe of candidate models. Their findings suggest that investors' allocations are closest to the predictions of the CAPM. Our SD-Sparse approach expands on

such analyses by explicitly embedding investor welfare considerations—represented by second-order stochastic dominance—to rigorously identify the economically dominant factor set from a broad universe of candidates. This direct linkage to investor utility provides a theoretically robust and economically interpretable foundation that distinctly advances existing empirical asset pricing research.

Our work also connects with the recent literature that employs advanced methodological approaches—particularly nonparametric and machine learning techniques—to analyze large datasets of asset factors and characteristics in search of incremental explanatory power for expected returns (e.g., Freyberger et al., 2020; Kim et al., 2020; Chen et al., 2024). For instance, Freyberger et al. (2020) apply an adaptive group LASSO method to select from 62 stock characteristics, concluding that only 13 characteristics provide incremental explanatory power for stock returns. Similarly, Chen et al. (2024) utilize adaptive LASSO on 44 candidate factors, including return anomalies, hedge fund strategies, and macroeconomic factors, to propose a nine-factor model that effectively explains hedge fund returns. While machine learning methods excel at handling large datasets and capturing nonlinear relationships, they typically lack clear economic assumptions regarding investor preferences, making economic interpretations and justifications challenging. In contrast, our SD-Sparse methodology, grounded explicitly in second-order stochastic dominance (SSD), naturally aligns with economic theory through generalized utility maximization, facilitating clearer economic interpretation and robust theoretical justification.

More broadly, our paper contributes to the literature emphasizing the economic importance of stochastic dominance (SD) frameworks in finance, particularly within

portfolio choice and asset pricing contexts. Notable studies in this domain include Constantinides et al. (2011), Hodder et al. (2015), and Jackwerth et al. (2009), which build upon foundational work integrating SD principles into portfolio optimization and pricing anomalies. In asset pricing specifically, Chalamandaris et al. (2024) employ stochastic spanning tests to assess market anomalies compared to traditional alpha-based and mean-variance methods, finding the SD approach to exhibit superior robustness across multiple models and performance metrics, both in-sample and out-of-sample. Further, Arvanitis et al. (2024) adopt a prospect spanning approach to evaluate whether standard factors adequately span well-known market anomalies, identifying select strategies that genuinely enhance the investment opportunity set for prospect-type investors, particularly those without loss aversion. Our SD-Sparse factor model directly extends this line of inquiry by providing a theoretically rigorous, economically intuitive, and empirically robust method for factor selection in equity markets.

Finally, to examine the robustness and generalizability of our results, we significantly expand our analysis by incorporating an extensive set of 177 candidate factors. Specifically, we include the 153 factors documented in Jensen et al. (2023) along with our initial set of 24 factors. Applying our sparse SSD selection methodology to this broader factor universe, we reaffirm our core finding: a ten-factor model continues to optimally balance parsimony and explanatory power. Interestingly, while the optimal model still comprises ten factors, the factor composition differs slightly from that selected from the initial set, underscoring the nuanced nature of factor selection and

the robustness of the sparse SSD methodology.[3] These additional analyses strengthen our conclusion that a carefully selected, theoretically grounded sparse factor set consistently delivers superior explanatory power relative to alternative methodologies and factor specifications.

The rest of the paper is organised as follows. Section 2 discusses the Sparse Spanning factor selection frameworks used in the paper, along with the proposed equities factor model. Section 3 presents the benchmark factor models (i.e., the leading factor and the adaptive LASSO model). Section 4 describes the equity candidate factors datasets. Section 5 reports in-sample and out-of-sample empirical asset pricing tests, respectively. 6 concludes. Further details regarding the computational strategies for Stochastic Spanning, are provided in the Appendix.

2. Literature Review

Our work relates to several strands of literature examining factor selection methodologies, especially in the context of extensive candidate factor sets often referred to as the *factor zoo*. Feng et al. (2020) address this proliferation issue by introducing statistical dimensionality reduction methods, specifically principal component analysis (PCA), to systematically evaluate and control for multiple testing problems inherent in large-scale factor analysis. While their approach significantly reduces false discoveries through robust statistical frameworks, it does not explicitly incorporate economic theory or investor preferences.

Similarly, Bryzgalova et al. (2023) utilize a Bayesian methodology, examining an

[3]We provide detailed empirical results based on this expanded set in an ongoing analysis, which will be incorporated in subsequent drafts.

unprecedentedly large number of models (two quadrillion), to identify robust equity return predictors from a vast set of candidate factors. Their Bayesian approach provides a powerful statistical framework to identify credible factors by evaluating posterior probabilities across competing models. However, their approach remains fundamentally statistical and does not explicitly embed investor welfare or economic interpretation into the factor selection process.

Complementing and extending these contributions, our study employs a Stochastic Dominance (SD)-based factor selection criterion—specifically second-order stochastic dominance (SSD)—to explicitly ground factor selection within investor welfare and economic theory. Unlike purely statistical methodologies, SD methods inherently embed investor preferences, capturing general risk aversion and implicitly accommodating higher-order moments of return distributions.

Our approach is also connected to recent empirical finance literature employing advanced machine learning techniques. For instance, Freyberger et al. (2020) implement an adaptive group LASSO method on a comprehensive set of 62 stock characteristics, identifying only 13 as significantly predictive. Similarly, Chen et al. (2024) apply adaptive LASSO to a broad set of hedge fund return factors, identifying a parsimonious subset that effectively explains hedge fund performance. Although machine learning methods offer powerful tools for managing extensive datasets and capturing nonlinear relationships, they often lack transparent economic intuition or clear ties to investor decision-making frameworks.

Furthermore, our paper contributes to the literature on applications of stochastic dominance within finance. Key contributions include Constantinides et al. (2011),

Hodder et al. (2015), and Jackwerth et al. (2009), who employ SD principles for asset pricing and portfolio choice problems. Chalamandaris et al. (2024) extend this line of inquiry by demonstrating that SD-based methodologies for evaluating market anomalies outperform traditional alpha-based and mean-variance methods in robustness tests. Additionally, Arvanitis et al. (2024) adopt a prospect spanning approach to determine which market anomalies expand investors' opportunities, highlighting strategies that significantly enhance investor welfare.

Our study bridges these strands by proposing a theoretically grounded yet practically implementable SD-based sparse factor model. By explicitly linking factor selection to investor welfare through SSD, we provide both rigorous economic interpretation and superior empirical robustness, contributing significantly to existing factor model literature.

3. Stochastic dominance

Stochastic dominance (SD) is a generalization of the mean-variance (MV) dominance criterion, without assuming a particular family of distributions, thereby allowing inferences accounting for the higher-order moments of returns. SD is a model-free alternative to MV dominance (see, Levy, 2016; Perrakis, 2019; Whang, 2019). Second-order stochastic dominance (SSD) ranks investments based on mild non-parametric regularity conditions on the distributions involved and assumes decision-making under risk for the investors with utilities that exhibit non-satiation and risk aversion. SSD is represented by sets of conditions in the form of lower partial moment inequalities between the compared distributions. The non-parametric nature of SSD makes

it particularly appealing for the asymmetric distributions of equities and factors.

3.1. Preliminaries and definitions

We describe our limiting economy for a number of factors. We denote the financial returns by a process X living in $\ell(\mathbb{N}, \mathbb{R})$, which is the space of bounded real valued sequences equipped with the uniform metric. $X_0^{(i)}$ denotes the i^{th}, $i \in \mathbb{N}$ coordinate, and $\mathbb{P}$ denotes the distribution of X. The set Λ_∞, a non-empty subset of the $\mathbb{N}$-simplex, includes all portfolios we can construct from the factors $\left\{\lambda \in \mathbb{R}^{\mathbb{N}} : \lambda_i \geq 0, i \in \mathbb{N}, \sum_{i=0}^{\infty} \lambda_i = 1\right\}$, and for $p \in \mathbb{N}$, $\Lambda = \left\{\lambda \in \Lambda_\infty, \sum_{i=0}^{p-1} \lambda_i = 1\right\}$ denotes the $p-1$ dimensional unit sub-simplex of Λ_∞ and K is a non-empty closed subset of Λ. Factor combinations λ and κ denote generic elements of Λ. In the present context, X is a random vector of factor returns for p base factors, while Λ represents a set of combinations formed on X.

We define the Lower Partial Moment Differential (LPMD) by

$$D(z, \kappa, \lambda, \mathbb{P}) := \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty} \kappa_i X_0^{(i)}\right)_+\right] - \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty} \lambda_i X_0^{(i)}\right)_+\right], \tag{1}$$

The LPMD (1) takes the interpretation of the difference between two coherent risk measures evaluated on the portfolios at hand. We define the second stochastic dominance (SSD) of a factor portfolio κ over a factor portfolio λ, as follows:

Definition 1. (Stochastic Dominance) A factor portfolio κ SSD dominates a factor portfolio λ, written $\kappa \underset{\text{SSD}}{\succeq} \lambda$, iff $D(z, \kappa, \lambda, \mathbb{P}) \leq 0$, for all $z \in Z$.

3.2. Stochastic spanning

We use, specifically, the *stochastic spanning* approach (Arvanitis et al., 2019), which is the model-free alternative to MV spanning (De Roon et al., 2003; Huberman and Kandel, 1987).

Definition 2. (Stochastic Spanning) A factor set K spans a factor set Λ by SSD iff for every factor portfolio $\boldsymbol{\lambda} \in \Lambda$, there exists a factor portfolio $\boldsymbol{\kappa} \in$ K that dominates it by SSD, i.e., $\forall z \in Z$,

$$D\left(z, \kappa, \lambda, \mathbb{P}\right) \leq 0.$$

Spanning occurs if introducing factors does not improve the efficiency set of basic factors over the broad class of investor preferences. Hence, stochastic spanning is suitable for checking whether the testing factor model of dimension q is not spanned by the set of factors of dimension p. We test empirically the null hypothesis H_0 vis-á-vis the alternative H_1:

H_0**:** The factor model of dimension q is spanned by the set of 24 factors.

H_1**:** There exist a factor model of dimension q, that is not spanned by the set of 24 factors.

Using the continuity properties of $D(\cdot, \cdot, \cdot, \mathbb{P})$ and the compactness of sets Λ, K, $\mathcal{X}$, we can characterize spanning (Arvanitis et al., 2019), by the following:

$$\eta(\mathbb{P}) := \sup_{\Lambda} \inf_{K} \sup_{\mathcal{Z}} D(z, \boldsymbol{\kappa}, \boldsymbol{\lambda}, \mathbb{P}). \tag{2}$$

Spanning occurs iff $\eta(\mathbb{P}) = 0$, while some $\boldsymbol{\lambda} \in \Lambda$ exist that are not second order stochastically dominated by any portfolio $\boldsymbol{\kappa} \in K$ (i.e., reject spanning), iff $\eta(\mathbb{P}) > 0$.

In this paper, we investigate such a problem for investors with risk-averse preferences which are interested in the whole return distributions generated by two sets of factors, a sparse subset K of dimension q with a limited number of factors coming from a larger set Λ of dimension p (24 factors).

3.3. Sparse Spanning

We use a discrete combinatorial algorithm for maximizing a function subject to a cardinality constraint. It starts with the empty set, and then adds factors to it in r iterations. In each iteration, the algorithm adds to its current solution the single factor increasing the value of this solution by the most, i.e., the factor with the largest marginal value with respect to the current solution.

To check sparse SSD spanning on factors data, we use consistent and feasible estimation procedures based on Linear Programming (LP) and a greedy algorithm, namely the Forward Stepwise Selection (FSS) algorithm. We use a finite set of increasing piecewise-linear functions, restricted to the bounded empirical supports, that are constructed as convex mixtures of appropriate "ramp functions" (in the spirit of (Russel and Seo, 1989)) in our representation as in (Arvanitis et al., 2024)). For every such utility function, we solve two embedded linear maximization problems.

Sparse Spanning SSD Characterization:

$$K \underset{\text{SSD}}{\succeq} \Lambda \text{ iff } \inf_{\mathcal{L}_{p,q}} \sup_{\Lambda} \inf_{K} \sup_{z \in Z} D\left(z, \kappa, \lambda, \mathbb{P}\right) \leq 0,$$

with $\mathcal{L}_{p,q} := \{K \subset \Lambda : K \text{ closed}, 0 < \text{csupp}\,(K) \leq q\}$.

3.4. Sparse Spanning SSD

Definition 3 (Sparse Spanning SSD)**.** For some fixed q, there exists a $K \subset \Lambda$ with $\text{csupp}\,(K) \leq q$ and such that $K \underset{\text{SSD}}{\succeq} \Lambda$, i.e. $\forall \lambda \in \Lambda$, $\exists \kappa \in K : \kappa \underset{\text{SSD}}{\succeq} \lambda$,
where $\text{csupp}\,(K) := \#\{i : \kappa_i \neq 0, \kappa \in K\}$ is the support of a portfolio set.

$$\kappa \underset{\text{SSD}}{\succeq} \lambda, \text{ iff } D\,(z, \kappa, \lambda, \mathbb{P}) \leq 0 \text{ for all } z \in Z :.$$

The Sparse Spanning Definition generalizes (Arvanitis et al., 2019) since i) it allows for a **high dimensional** setting, ii) it only prescribes the existence of a "low-dimensional" spanning subset of Λ. Any procedure designed to test whether Sparse Spanning holds, would have to search for a spanning set inside the collection of "low-dimensional" subsets of Λ.

3.5. Numerical Implementation

For $q < p$, we consider the following empirical optimization problem:

$$\sup_{u \in \mathcal{U}} \inf_{\mathcal{L}_{p,q}} \left(\sup_{\boldsymbol{\lambda} \in \Lambda} \mathbb{E}_{\mathbb{P}_T} \left[u \left(X^{\mathrm{T}} \boldsymbol{\lambda} \right) \right] - \sup_{\boldsymbol{\kappa} \in \mathrm{K}} \mathbb{E}_{\mathbb{P}_T} \left[u \left(X^{\mathrm{T}} \boldsymbol{\kappa} \right) \right] \right). \tag{3}$$

We approximate every element of $\mathcal{U}$ with arbitrary prescribed accuracy using a finite set of increasing and concave piecewise-linear functions. Thus, the produced factor model is founded on the expected utility maximisation framework.

3.5.1. Greedy algorithm

We start with an empty set, and then we gradually increase the number of factors w in K (we add 1 factor at a time) until we find a set $K \subset \Lambda$ with $K \underset{\text{SSD}}{\succeq} \Lambda$.

In each iteration, we search for the factor that increases (3) the most.

Steps of Greedy algorithm (Forward Stepwise Selection):

For $w = 1$ to q:

1. If $w = 1$, we search for the single factor that increases the value of (3) the most, thus $\mathcal{L}_{p,q}$ is a singleton.

2. For $1 < w < q$, we solve (3) for each additional factor, and we keep the subset K with dimension w, that maximizes (3) the most.

3. If we find a spanning set K inside the collection of all possible subsets of Λ with dimension w, then the algorithm stops.

4. Else, if $w = q$, we end up with a sparse portfolio set K that "comes as close as possible" to SSD spanning its high dimensional universe of portfolios, and we evaluate the utility loss. If non zero, we may continue for $w > q$, up to p.

Using the sparse spanning methodology described above, we find a sparse subset $K \subset \Lambda$ of q=10 factors. Our proposed factor model includes:

- **SD-Sparse**: including the market factor (MKT), the size (SMB), the return-on-equity (ROE), the expected growth factor (EG), time-series momentum (TSMOM), the *Devil in HML's Details* (HML^m), the quality-minus-junk (QMJ),

the asset growth (AG), the gross-profitability (GP) and net operating assets (NOA) factors.

4. Benchmark factor models

In this section, we describe the factor models we use as benchmarks for comparing the performance of our proposed model. We start by describing leading factor models for evaluating stock and equity portfolio performance in the relevant literature. Then, we present the Lasso framework used as an extra benchmark from machine learning.

4.1. Leading factor models

First, we evaluate the performance of our Stochastic Bounding factor model against various models that combine many prominent factors from the literature. Those include the Carhart (1997), the Fama and French (2015) with momentum (FF6), the Fama and French (2015) with liquidity (FF6Liq), the Stambaugh and Yuan (2017) (SY), the Barillas and Shanken (2018) (BS), the Hou et al. (2020) (HXZ) and the Daniel et al. (2020) (DHS) factor models.

- **Carhart**: the Fama and French (1993) four-factor model, including the market factor (MKT), the size factor (SMB), the high-minus-low factor (HML), and the momentum factor based on prior two to 12-month returns (MOM).
- **FF6**: the Fama and French (2015) five-factor model, including the market factor (MKT), the size factor (SMB), the high-minus-low factor (HML), the robust-minus-weak factor (RMW), the conservative-minus-aggressive factor (CMA), and the momentum factor based on prior two to 12-month returns (MOM).

- **FF6Liq**: the Fama and French (2015) five-factor model, including the momentum factor (MOM) and the liquidity factor (LIQ) of Pastor and Stambaugh (2003).

- **BS**: the Barillas and Shanken (2018) factor model, including the market factor (MKT), the size factor (SMB), the high-minus-low factor of Asness and Frazzini (2013) (HML^m), the investment factor (IA), the profitability factor (ROE), and the momentum factor (UMD).

- **HXZ**: the Hou et al. (2020) factor model, including the market factor (MKT), the investment factor (IA), the profitability factor (ROE), the market-equity (ME), and the expected growth factor (EG).

- **DHS**: the Daniel et al. (2020) factor model, including the market factor (MKT), the short-horizon behavioural factor (PEAD), and the long-horizon behavioural factor (FIN).

4.2. Adaptive LASSO model

We use the adaptive LASSO method as our machine learning benchmark for selecting factors following Chen et al. (2024), showing the ability of the method to identify informative factors for hedge fund returns. Contrary to the LASSO regression of Tibshirani (1996), which uses a constant shrinkage to the regression coefficients and underestimates the magnitude of non-zero coefficients in sparse frameworks, the adaptive Lasso shrinks more significant coefficients less severely. This property of adaptive Lasso is achieved by using "adaptive weights" for penalizing different coef-

ficients with the L_1 penalty to the ordinary least squares regression (see, Zou, 2006). At the same time, the adaptive LASSO retains the simple Lasso properties by setting less significant coefficients to zero, achieving the goal of factor selection. The above is achieved by the following loss function:

$$\hat{\beta}_i = \arg\min_{\beta_\rangle} ||r_i \ - \sum_{j=1}^{P} \beta_{i,j} f_i||^2 \ + \lambda_i \sum_{j=1}^{P} w_j |\beta_{i,j}| \tag{4}$$

where w_j is a vector of pre-specified weights, which is defined by using OLS estimates as $w_j = \frac{1}{|\hat{\beta}_j^{OLS}|}$.

To select the factors with the highest explanatory power on equity returns, we use our set of assets as targets for the adaptive Lasso following Chen et al. (2024). We first calculate the OLS estimates to compute the "adaptive weights" for each factor based on the $\hat{\beta}^{OLS}$ estimate of each factor. Then, we standardize our factors before implementing the adaptive Lasso methodology, which is scale-sensitive. After transforming the factors with these adaptive weights, we perform a grid search for the optimal value of the penalty parameter λ. With the optimal λ, the Adaptive Lasso regression is applied on each individual asset, resulting in a set of selected factors per asset. Then, we evaluate how frequently each factor is selected across all asset regressions. Consistently selected factors are considered strong candidate factors. In the final step, we perform OLS regressions on the equally weighted portfolio returns using all possible combinations of these candidate factors. Thus, we determine the final model based on explanatory power and statistical significance.

- **ALASSO**: including the market factor (MKT), the size factor (SMB), the

market-equity (ME), the *Devil in HML's Details* (HML^m), and the return-on-assets (ROA) factor.

5. Equities and Factor Data

5.1. Equities data

We obtain our test assets following a similar philosophy as Giglio and Xiu (2021), including 202 equity portfolios from Kenneth French's data library and a more complete set of 360 anomaly portfolios sorted by characteristics from Andrew Chen's Open Source Asset Pricing data library [4] The Kenneth French set of portfolios consists of 25 portfolios sorted by size and book-to-market ratio, 17 industry portfolios, 25 portfolios sorted by operating protability and investment, 25 portfolios sorted by size and variance, 35 portfolios sorted by size and net issuance, 25 portfolios sorted by size and accruals, 25 portfolios sorted by size and beta, and 25 portfolios sorted by size and momentum. As for Andrew Chen's set, we choose quintile portfolios sorted based on 108 characteristics. We remove portfolios with missing values during our sample period or are highly correlated (i.e., more than 99%) with each other. Our final sample comprises 562 test portfolios of monthly frequency from January 1990 to December 2023.

[4]The former have been obtained from Kenneth French's (`http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`) and the latter from Andrew Chen's Open Source Asset Pricing website (`https://www.openassetpricing.com/data/`).

5.2. SD-Sparse Potential Factors

We consider a set of 24 factors, consisting of all non-overlapping factors included in each of the leading factors and the adaptive LASSO model, along with some extra factors or anomalies found to be significant in the relevant literature (see, Pastor and Stambaugh, 2003; Stambaugh et al., 2012; Moskowitz et al., 2012; Frazzini and Pedersen, 2014. The latter include asset growth, betting-against-beta, quality-minus-junk, composite equity issues, gross profitability, illiquidity, net operating assets, net stock issues, return on assets, and time-series momentum. In this way, we are trying to answer our research question of whether we can use the SD-Sparse methodology to create a more informative set of factors from all leading factor models and informative factors found in the equities literature. [5]

6. Empirical Application

6.1. In-sample analysis

6.1.1. Time-series Analysis: *Model Comparison*

We start our in-sample analysis with a preliminary time-series analysis of the SD-Sparse against the benchmark leading factor models by performing a time-series regression of the equities' returns on each factor model to assess their goodness of

[5]We have also employed a more extensive set of 153 factors of Jensen et al. (2023) (https:www.jkpfactors.com), clustered into 13 broad themes, as inputs for the SD-Sparse and Lasso frameworks to identify the most important factors describing equity portfolio returns. Our results are at least qualitatively similar. However, we would like to focus our study on whether we can build a more informative factor model from a small set of factors established in the literature as significant in explaining equity returns.

fit. Table 1 presents the intercept (i.e., alpha), the R^2, and the adjusted R^2 of those regressions.

Table 1: **Time Series Analysis of Factor models**. The table reports the average intercept (i.e., alpha), the R^2, and the adjusted R^2 (i.e., Adk. R^2) of a time-series regression of the universe of equities on SD-Sparse and the benchmark factor models for the period January 1990 to December 2023.

Factor Models	**Alpha (%)**	R^2 **(%)**	**Adj.** R^2 **(%)**
Carhart	0.10	87.7	87.6
FF6	0.11	88.8	88.6
FF6L	0.13	88.8	88.6
BS	0.07	88.4	88.2
HOU	0.20	86.3	86.1
DHS	0.33	76.2	76.1
SD-Sparse	0.04	90.0	89.7
ALASSO	0.02	88.3	88.1

The SD-Sparse factor model reports the highest explanatory power by reporting a large R^2, an adjusted R^2, and the second-smallest intercept. The adaptive LASSO framework closely follows, the Barillas and Shanken (2018) model being the third best. While the above findings show the superiority of SD-Sparse in a preliminary time-series setting, we also need to

6.1.2. *Scaled Intercepts & Lucky Factors*

Following the philosophy of Harvey and Liu (2021), we compute the scaled intercept statistics of a baseline model- in our case, each benchmark- and an augmented model- in our case, the SD-Sparse. The scaled intercept statistic is the ratio of the scaled absolute difference between the augmented and the baseline models' intercepts over

the average scaled absolute intercept in the cross-section. The standard error of the regression intercept under the baseline model scales the ratio's numerator and denominator. For instance, the scaled intercept statistic measuring the percentage difference in the scaled mean absolute intercept is given by

$$SI^m = \frac{1/N \sum_{i=1}^{N} (|\alpha_i^+| - |\alpha_i|)/s_i}{1/N \sum_{i=1}^{N} \alpha_i/s_i} \tag{5}$$

where (α_i^N) and $(\alpha_i^N)^+$ denotes the cross-section of baseline and augmented model intercepts, respectively. While (s_i^N) denotes the cross-section of baseline intercept's standard errors. A scaled intercept indicates that the augmented model is better than the baseline model. The unique feature of scaled intercept statistics is that they are economically and statistically sound. It is closely related to the *information ratio* taking a long position in the test asset and a short position in the zero-cost portfolio, in our case, each factor. Also, it controls for heterogeneity in asset return volatility, assigning less weight to more volatile assets in the model.

We also compute and present the multiple test p-value of the approach of Harvey and Liu (2021) for selecting incremental factors. The scope of their method is to identify factors from a set of candidates that can be added to a set of pre-determined factors, established as explanatory, while controlling for luck. The so-called *Lucky Factors* approach controls the family-wise error rate to select *new* factors via bootstrapping. The null hypothesis is that the candidate factor has zero cross-sectional explanatory power. The method is performed in a panel regression setup to address the impact of extreme observations and model misspecifications, which can appear in cross-sectional regressions. Another advantage of the *Lucky Factors* method is

its suitability for testing on a large cross-section of assets relative to the time-series dimension, which also holds in our case. Such a feature, along with the test assessing the incremental information of each candidate factor, makes the method more robust and powerful than previous methods for factor model comparison, such as the (GRS) test of Gibbons et al. (1989).[6]

Table 2 reports the scaled intercepts as the percentage difference in the scaled mean and median absolute intercept under equal weighting (columns two and four of the table). We set the SD-sparse as our augmented model and each benchmark model as a baseline. We also report the corresponding multiple testing p-values as given by the *Lucky Factors* test. At the same time, a small p-value means that at least one factor from the SD-Sparse significantly decreases the baseline model intercept and improves the explanatory power of the baseline model. We only test factors that do not overlap between the baseline and the SD-Sparse model. In other words, if an SD-Sparse factor is also included in the baseline, then it is not assessed as a candidate. [7]

Table 2 findings highlight the superiority of the SD-Sparse against most of the leading and ALASSO factor models concerning the scaled intercept findings. Both mean and median scaled intercept statistics' negative values reveal that the regression

[6]As highlighted in Harvey and Liu (2021) it is challenging to accurately estimate the residual covariance matrix in large cross-sections, leading to large GRS statistics and more frequent rejections of factor models. Indeed, we faced a similar issue when we applied the GRS test separately for the 202 portfolios of Fama and French and the 360 portfolios obtained from Andrew Chen's data library.

[7]The test provides the scaled intercept and the multiple test p-value for each factor model separately under its stepwise structure. We decided to compute and present a scaled intercept statistic including all candidate factors and the multiple test p-value of the candidate factor with the smallest due to the lack of space. The detailed results for each SD-sparse factor are available upon request.

Table 2: **Scaled Intercepts and *Lucky Factors*.** The table reports the scaled intercepts measuring the difference in the equally weighted scaled mean/median absolute regression intercept between the SD-Sparse and the benchmark factor models (i.e., leading and ALASSO factor models). The mutiple test p-value corresponds to the multiple hypothesis testing of the explanatory power of a candidate factor model under the *Lucky Factors* procedure of Harvey and Liu (2021)

Baseline	**Mean**		**Median**	
	SI	Mult.p-value	SI	Mult.p-value
Carhart	-0.457	0.001	-0.506	0.001
FF6	-0.363	0.000	-0.379	0.000
FF6_Liq	-0.402	0.000	-0.393	0.000
BS	-0.351	0.000	-0.378	0.000
Hou_Q	-0.521	0.235	-0.534	0.224
DHS_Q	-0.661	0.000	-0.726	0.000
ALASSO	-0.340	0.000	-0.371	0.001

scaled intercept provided by the SD-Sparse is always smaller than the scaled intercept generated by each benchmark model. The multiple test p-values under both mean and median-based scaled intercept statistics are usually lower than the 1% statistical significance level. Such a result shows that at least one factor from the SD-Sparse model improves the explanatory power of the baseline model while controlling for luck except for the case of Hou et al. (2020) model used as the baseline.

6.1.3. Cross-sectional Analysis: *Pricing Performance*

We explore the pricing performance of our proposed SD-Sparse factor model for equity returns against the leading factor models and the factor model provided by the Lasso specification. To do so, we use the approach of Kan et al. (2013), considering the impact of model misspecification on the variability of cross-sectional estimates. Deriving the asymptotic distribution of the cross-sectional R^2 estimator, and so the distribution of the difference between R^2 for competing models, is crucial for robust

factor model comparison. For that purpose, we compute the cross-sectional R^2 in an excess returns specification, including a constant for a zero-beta rate different from the risk-free rate. We also compute the risk-premia estimates and prices of risk for each factor model along with their corresponding misspecification robust t-ratios. In this subsection, we focus our analysis only on the 360 portfolios obtained from Andrew Chen. Performing the approach of Kan et al. (2013) for model misspecification cross-sectionally is infeasible for the whole universe of portfolios since the number of test assets (i.e., 562) is greater than the number of monthly periods used in the estimation (i.e., 408). [8]

We run cross-sectional regressions for each portfolio on a constant, and the factor returns to obtain the above estimates for the SD-Sparse and benchmark factor models. Following Lewellen et al. (2010), we focus our analysis on GLS estimators rather than OLS estimators since the latter do not appear to be very informative about the true performance of a factor model and often provide inflated cross-sectional R^2s. Also, GLS R^2 provides a more practical economic interpretation regarding the relative mean-variance efficiency of the model's factor portfolios, as it is determined purely by the factor's location in the mean-variance space. Hence, GLS R^2 is a more meaningful statistic than the OLS R^2 when it comes to the question of the ability of a factor model to better explain the risk-return opportunities available in the market.

We start our in-sample analysis by presenting pairwise tests of equality of cross-sectional R^2s for the SD-Sparse and the competing factor models. Following Kan et al. (2013), we apply the normal test (i.e., $H_0 : 0 < R_A^2 = R_B^2 < 1$) for non-

[8]The corresponding results of factor models cross-sectional pricing performance on the Kenneth French's 202 equity portfolios are available upon request.

nested models, assuming that the factor models provide some explanatory power on the cross-section of equity portfolio returns. Table 3 reports the pairwise differences between the sample R^2s for all the factor models considered and their corresponding two-tailed p - values in parentheses. The negative sign of the difference denotes that the column model has a higher R^2 than the row model.

Table 3: **Equality tests of cross-sectional $\mathbf{R^2}$s**. The table reports the pairwise differences of the GLS cross-sectional R^2s of SD-Sparse and the seven benchmark factor models. All models are estimated based on a universe of 360 equity portfolio returns from January 1990 to December 2023. We define the difference as the row minus the column model and present its corresponding p - value for non-nested models, testing the null hypothesis of equality in R^2s, in parenthesis. Statistical significance at 10%, 5% and 1% is denoted as *, ** , and ***, repsctively.

	FF6	FF6_Liq	BS	Hou_Q	DHS	Lasso5	SD-Sparse
Carhart	-0.0039	-0.0045	-0.0030	-0.0044	-0.0006	-0.0040	-0.0087
	(0.02)	(0.02)	(0.03)	(0.01)	(0.39)	(0.02)	(0.00)
FF6		-0.0005	0.0009	-0.0004	0.0033	-0.0000	-0.0047
		(0.52)	(0.57)	(0.82)	(0.06)	(0.97)	(0.04)
FF6_Liq			0.0014	0.0000	0.0038	0.0004	-0.0042
			(0.42)	(0.96)	(0.06)	(0.827)	(0.06)
BS				-0.0013	0.0024	-0.0010	-0.0056
				(0.39)	(0.07)	(0.60)	(0.00)
Hou_Q					0.0037	0.0034	-0.0043
					(0.02)	(0.88)	(0.00)
DHS						-0.0034	-0.0080
						(0.07)	(0.00)
ALASSO							-0.0047
							(0.06)

The results show the superiority of our SD-Sparse factor model in generating significantly higher cross-sectional R^2s than most benchmark factor models at the 1% level and the FF6, FF6_Liq, and ALASSO factor models at the 5% level. In other words, our stochastic dominance factor model dominates most of the leading and the adaptive LASSO factor models in explaining equity portfolio returns. The

DHS factor model comes second by dominating most of the remaining benchmark factor models at the 5% level, except for the Carhart (1997) factor model. The above results justify the crucial property of the SD-Sparse framework in considering the entire distribution of equity portfolio returns to select the dominant factors rather than the existing factor models constructed based only on the mean-variance efficiency statement. We also observe that this property of SD-Sparse is even more advantageous than nonlinear machine learning methods in building a factor model that better explains equity portfolio returns.

Next, we investigate the exact pricing performance of the factors comprising the SD-Sparse, leading and ALASSO models. Table 4 reports the asset pricing findings based on the zero-beta rate and the risk premia estimators for each factor along with their corresponding model misspecification robust t-ratios from Kan et al. (2013), which are more informative on the true explanability of each factor, especially when the tested factor has low correlation with asset returns.[9] We also report the GLS and OLS R^2 for each factor model.

At first glance, we observe that the market factor (MKT) risk premium is not statistically significant across all factor models since it reports low misspecification robust t-ratios. Interestingly, only for SD-Sparse and the Hou et al. (2020) factor models do at least half of the factors generate significant premia. For the remaining models, only a few factors offer significant risk premia controlling for miss-pricing. For example, the return on equity (ROE), the expected growth factor (EG), the

[9]The Fama and MacBeth (1973) t-ratios under correctly specified models, the Shanken (2015) and the Jagannathan and Wang (1998) t-ratios under correctly specified models that account for the EIV problem are also available upon request.

Table 4: **Zero-beta rate and risk premia estimates.** The table presents the zero-beta rates and the risk premia estimates of the SD-Sparse and the seven benchmark factor models in % terms. The models are estimated using cross-sectional regression of monthly equity portfolio returns on the factor returns from January 1990 to December 2023. We also report the model misspecification robust t-ratios from Kan et al. (2013) and the GLS and OLS R^2 for each factor model. Statistical significance at 10%, 5% and 1% is denoted as *, ** , and ***, repsctively

Factors	Carhart		FF6		FF6Liq		BS		HXZ		DHS		ALASSO		SD-Sparse	
	Premium	t-ratio	Premium	t-ratio	Premium	t-ratio	Premium	t-ratio	Premium	t-ratio	Premium	t-ratio	Premium	t-ratio	Premium	t-ratio
Premium	t-ratio															
Constant	0.0045***	3.34	0.0044***	3.31	0.0044***	3.23	0.0044***	3.29	0.0044***	3.25	0.0044***	3.31	0.0045***	3.33	0.0043***	3.17
Mkt-RF (From FF5)	0.0026	1.02	0.0027	1.03	0.0027	1.04	0.0027	1.04	0.0027	1.06	0.0027	1.03	0.0026	1.02	0.0028	1.08
SMB	0.0017	1.00	0.0017	1.01	0.0017	1.01	0.0017	1.01					0.0017	1.025	0.0017	1.01
HML	-0.0012	-0.67	-0.0013	-0.72	-0.0013	-0.69										
Mom	0.0027	0.94	0.0054***	3.54	0.0054***	3.55										
RMW			0.0022	1.48	0.0022	1.49										
CMA			0.0027	0.94	0.0027	0.94										
Agg. Liq					-0.0095	-1.33										
IA							0.0028*	1.71	0.0028*	1.76						
ROE							0.0059***	3.43	0.0058***	3.41					0.0059***	3.43
HML^m							0.0007	0.33					0.0007	0.33	0.0007	0.34
TSMOM							0.0024	0.42							0.0023	0.39
R_ME									0.0002	0.10			0.0001	0.06		
EG									0.0065***	4.11					0.0065***	4.10
PEAD											0.0015	0.84				
FIN											0.004	1.58				
QMJ															0.0048***	3.18
Asset growth															0.01***	3.90
Gross profitability															0.0021*	1.49
Net_operating_assets															0.0067***	3.07
Return_on_assets													0.0061***	3.5		
GLS R^2	0.03		0.42		0.47		0.33		0.46		0.09		0.43		0.91	
OLS R^2	20.23		35.15		35.15		28.56		44.28		4.36		25.43		64.79	

Quality Minus Junk, the asset growth, and the net operating assets factors generate positive and statistically significant premia at 1% level for the SD-Sparse model. The gross profitability generates positive and statistically significant premia at 10% level. The return on equity (ROE) and the expected growth factor (EG) are also statistically significant at 1% for the Hou et al. (2020) model, along with the investment factor (IA), which is positive and significant at the 10% level. The investment (IA) and the return on equity (ROE) factors are also significant for the Barillas and Shanken (2018) factor model. Regarding the remaining factors, only the momentum factor for the FF6 and FF6, including liquidity factor, models and the return on assets for the ALASSO model are significant at 1% level. Overall, our proposed SD-Sparse factor model selects four extra significant factors, the *Quality Minus Junk*,

the asset growth, the gross profitability and the net operating assets, with high misspecification t-ratios, which increase the model's explanatory power. This is also revealed by the highest generated cross-sectional GLS and OLS R^2 estimators from the SD-Sparse model compared to the benchmark. Surprisingly, commonly used factor models, such as the Carhart and FF6 models for empirical asset pricing exercises, present much lower GLS R^2s, with that of Hou et al. (2020) being the second highest. Similarly, ALASSO yields a quite low R^2, which can be justified by the "black-box" properties of machine learning models lacking an economic framework. [10]

Kan et al. (2013) stress that the usual risk premium for a given factor does not allow for robust inference on the added value of that factor on a multifactor model unless the factors are uncorrelated. It is the factor's price of covariance risk which permits accurate inference. We provide the relevant findings in Table 5. In particular, we report risk prices and their corresponding misspecification robust t-ratios for each factor model.

Our findings support that stochastic sparse spanning can better select and spotlight existing factors with high explanatory power for the cross-section of equity portfolios' expected returns, resulting in the highest cross-sectional R^2 across all models examined. Specifically, the asset growth and *Quality Minus Junk*, chosen from SD-Sparse as extra factors, carry significant risk prices of 6.39 and 12.45 with large corresponding t-ratios of 2.59 and 1.9, respectively. The same factors also report significant risk premia. Similarly, return on equity (ROE) and expected growth (EG) factors generate high and statistically significant risk prices of 8.96 and 15.61,

[10]The substantial difference between GLS and OLS R^2s is because obtaining a high population GLS R^2 in practice is much harder than finding a high OLS R^2 (see, Lewellen et al., 2010).

Table 5: **Zero-beta rate and prices of risk estimates** The table presents the zero-beta rates (in % terms) and the covariance prices of risk estimates of the SD-Sparse and the eight benchmark factor models. The models are estimated using cross-sectional regression of monthly equity portfolio returns on the factor returns from January 1990 to December 2023. We also report the model misspecification robust t-ratios from Kan et al. (2013) and the GLS and OLS R^2 for each factor model. Statistical significance at 10%, 5% and 1% is denoted as *, ** , and ***, repsectively

Factors	Carhart		FF6		FF6Liq		BS		HXZ		DHS		ALASSO		SD-Sparse	
	Price Risk	t-ratio	Price Risk	t-ratio	Price Risk	t-ratio	Price Risk	t-ratio	Price Risk	t-ratio	Price Risk	t-ratio	Price Risk	t-ratio	Price Risk	t-ratio
Constant	0.005***	3.34	0.004***	3.31	0.004***	3.26	0.004***	3.29	0.004***	3.25	0.004***	3.31	0.005***	3.33	0.004***	3.17
Mkt-RF (From FF5)	1.64	1.04	5.026***	2.73	5.71***	2.81	4.533**	2.46	6.771***	3.36	3.726**	2.01	3.129***	2.04	10.116***	3.74
SMB	1.099	0.65	7.006***	3.01	6.846***	2.94	7.329***	3.08					34.063***	3.51	9.118***	3.06
HML	-0.231	-0.12	-9.635***	-2.61	-9.121**	-2.46										
Mom	1.656	1.06	16.336***	4.75	16.164***	4.71										
RMW			14.783***	2.49	14.133***	2.38										
CMA			0.575	0.34	0.784	0.47										
Agg. Liq					-2.283	-1.25										
IA							5.791**	1.96	8.297***	2.17						
ROE							13.931***	5.87	3.753	1.63					8.969*	1.81
HML^m							2.188	0.61					4.295***	2.05	5.974**	2.01
TSMOM							1.195	0.23							0.645	0.13
R_ME									4.614**	2.12			-27.737***	-2.94		
EG									17.592***	3.31					15.612***	2.58
PEAD											5.796	1.27				
FIN											4.203**	2.30				
QMJ															12.458*	1.91
asset growth															6.392***	2.59
gross profitability															-3.989	-0.79
net_operating_assets															3.294	1.09
return_on_assets													15.911***	4.95		
GLS R^2	0.03		0.42		0.47		0.33		0.46		0.09		0.43		0.91	
OLS R^2	20.23		35.15		35.15		28.56		44.28		4.36		25.43		64.79	

at 10% and 1% level, respectively. The Hou et al. (2020) also validates the significant explanatory power of those factors regarding risk premia and risk prices. This is probably the reason for the Hou et al. (2020) model generating the second-highest cross-sectional R^2. Investment (IA) and momentum factors are the remaining factors that demonstrate significant risk premia and risk price. There are also numerous factors which present statistically significant prices of risk, but not risk premium, such as market equity, size (SMB), high minus low (HML, HML^m), robust-minus-weak factor (RMW), and long-horizon behavioural factor (FIN).

6.1.4. Non-Parametric Stochastic dominance tests

To complete our in-sample analysis, we compare the explanatory power of the proposed SD-Sparse model with all the other factor models we consider by comparing

the performance of the optimal portfolios formed on the residuals plus alpha of equity portfolios for each factor model. The optimal portfolios are constructed using the stochastic bounding method of Arvanitis et al. (2021). The idea is that the true factor model should explain equity returns. On the other hand, if the asset pricing model cannot explain documented return patterns, then the model itself ignores the existence of a real systematic risk, meaning that systematic risk factors are missing from the model. Then the residuals and the alpha coefficient contain any missing information. The higher the explanatory power of a particular factor model, the lower the information left in the abnormal returns (i.e., residuals plus alpha), and the poorer the performance of the optimal equity portfolio. For each factor model, we have:

$$R_i = \hat{\alpha}_i + \sum_{m=1}^{M} \hat{\beta}_{i,m} f_m + \hat{\epsilon}_i, \tag{6}$$

$$R_i^{abn} = R_i - \hat{\beta}_{i,m} f_m \tag{7}$$

where R_i are the excess returns of the equity portfolio i, f_m is the return of the m-th factor, and $\hat{\beta}_{i,m}$ is the m-th factor loading for the equity portfolio i and $\hat{\epsilon}$ is a zero-mean residual for the equity portfolio i.

The Stochastic Bounding equities portfolio dominates any portfolio that can be constructed from the set of the 693 equity portfolios with respect to the second stochastic dominance criterion. If such a combination exists, then it spans any other combination of equities, and it is not spanned by any of them.

Definition 2.2.1 (Stochastic Bounding): *Portfolio set* Λ *stochastically bounds*

portfolio set K *if it contains a portfolio which stochastically dominates all portfolios* $\boldsymbol{\kappa} \in \mathrm{K}$ *by SSD:* $\exists \boldsymbol{\lambda} \in \Lambda : (\boldsymbol{\lambda} \succeq_F \boldsymbol{\kappa} \, \forall \boldsymbol{\kappa} \in \mathrm{K})$.

To get the optimal equity portfolio on the abnormal returns from the set Λ of the 693 equity portfolios we consider, we set $\Lambda = \mathrm{K}$. So, now we search for a portfolio $\lambda \in \Lambda$ that stochastically dominates every other portfolio in Λ. We use the Arvanitis et al. (2021) method to get the portfolio $\lambda \in \Lambda$ that stochastically bounds any portfolio of the equities set Λ. The computational strategy for the bounding test is given in the Appendix.

We then solve the bounding model on the abnormal returns generated by each factor model and get the optimal composition of the equity portfolios. We use the whole sample size, and after obtaining the optimal weights of each equity for each factor model, we construct the optimal bounding portfolios and calculate the performance measures. Stochastic Bounding amounts to the one-fund separation that the optimal solutions for all permissible utility functions are equivalent. Such universal optimal solutions are 'greatest elements'. If the greatest elements exist, then they naturally constitute the efficient set. Multiple non-equivalent greatest elements cannot exist due to the antisymmetry property of the dominance relation. If the factor model contains the "true factors," then all combinations of the equities, after controlling for the systematic risk, should exhibit poor performance. Expected hedged return will show cross-sectional variation, depending on the goodness of the factor model.

We present the performance metrics of each optimal equity portfolio abnormal returns and for each factor model in Table 6. In particular, we calculate the mean

return, the annualized Sharpe ratio and the upside-potential ratio of the optimal portfolio of equities' abnormal returns as given the examined models.

Table 6: **Performance Measures of the Optimal Portfolio of Abnormal Returns**. The table reports the in-sample performance measures, namely, the mean, standard deviation (St.Dev.), the annualized Sharpe ratio and UP ratio, of the optimal bounding portfolios formed based on the abnormal returns (i.e., residuals plus alpha) given for each equity portfolio by each factor model.

Performance Measures				
Models	Mean %	St.Dev %	An.Sharpe	U-P ratio
Carhart	0.72	1.81	0.28	0.98
FF6	1.02	2.47	0.32	1.17
FF6Liq	1.04	2.44	0.34	1.21
BS	1.01	2.57	0.31	1.17
DHS	1.19	3.32	0.29	1.14
HXZ	1.08	2.47	0.35	1.27
Lasso	0.97	2.38	0.32	1.20
SD-Sparse	0.53	1.30	0.24	0.93

We observe that the optimal equities portfolios constructed based on the abnormal returns of the SD-Sparse model exhibit the lowest average return and standard deviation, as well as the lower Sharpe and U-P ratios. Those findings suggest that the SD-Sparse factor model's abnormal returns contain less information than the benchmark factor models. However, we still need to validate those findings in an out-of-sample set-up.

6.2. Out-of-sample analysis

We proceed with the out-of-sample analysis of the SD-Sparse factor against the benchmark factor models to better explain equity portfolio returns. First, we per-

form a prediction experiment on one-month ahead test assets' returns.Second, we perform a non-parametric test, similar to the one in-sample, using the abnormal returns (i.e., alpha and residuals) extracted from each factor model, to form optimal equity portfolios with the Stochastic Bounding methodology and evaluate their performance.

6.2.1. Optimal prediction

First, we evaluate the out-of-sample predictive performance of SD-Sparse and the benchmark factor models on the one-month ahead equity portfolios' returns. In particular, we run time series linear regressions of the test portfolios on each factor model in a rolling structure. We use the monthly data from January 1990 to December 2009 as in-sample and from January 2010 to December 2023 as out-of-sample (i.e., the final fourteen years of our sample). The first time series regression is performed using the initial in-sample period. Then, we roll the in-sample window forward by one month at each step, maintaining a fixed window size, until we reach the end of our sample period. We perform our prediction experiment in both a panel structure of all test portfolios and the equally weighted portfolio of the test portfolios separately. We calculate the corresponding mean absolute (MAE) and root mean squared (RMSE) errors to evaluate the accuracy of each factor model's equity portfolio predictions. Table 7 presents the relevant results. Our findings show that the SD-Sparse model outperforms all linear benchmark and Lasso factor models by generating the smallest MAE and MSE metrics compared to the benchmarks. The results are more sound for the prediction case of the equally weighted portfolio of test assets. The SD-Spanning model demonstrates superior predictive out-of-sample

performance, achieving the lowest errors compared to the models.

Table 7: **Prediction Accuracy**. The table reports the mean absolute error (MAE) and root mean squared error (RMSE) metrics of generated forecasts of the equally weighted portfolio (EW) and the test assets excess returns panel for each factor model. We run rolling window regressions using the corresponding factors of each model as predictors.

	EW		**Panel**	
Factor model	MAE	RMSE	MAE	RMSE
Carhart	0.0580	0.0781	0.0185	0.0284
FF6	0.0578	0.0780	0.0185	0.0283
FF6Liq	0.0578	0.0779	0.0185	0.0283
BS	0.0575	0.0780	0.0184	0.0282
DHS	0.0574	0.0770	0.0229	0.0335
HOU	0.0570	0.0770	0.0186	0.0287
SD-Sparse	0.0561	0.0769	0.0183	0.0279
ALASSO	0.0579	0.0784	0.0184	0.0282

Next, we perform the Diebold and Mariano (1995) test for superior predictive ability to assess the statistical significance of the predictions of the SD-Sparse factor model against the predictions of the benchmark factor models in a pairwise set-up. The Diebold and Mariano (1995) test examines the null hypothesis that two forecasts possess equal predictive ability by assessing the difference between their loss functions. The alternative hypothesis suggests that this difference is non-zero. A negative and statistically significant t-statistic leads to rejecting the null hypothesis, indicating that the SD-Sparse factor model outperforms the benchmark by achieving a lower loss. Table 8 presents the relevant t-statistics and their corresponding p-values in parenthesis under the MAE loss function.

The findings validate the superiority of the SD-Sparse factor model in predicting

Table 8: **Diebold and Mariano (1995) test for superior predictive accuracy**. The table presents t-statistics and their corresponding p-values in parentheses from the Diebold and Mariano (1995) test, comparing the SD-Sparse factor model predictive accuracy against each benchmark on a pairwise basis for the January 2010 to December 2023 out-of-sample period. The null hypothesis asserts that the SD-Sparse and the benchmark factor model possess equal predictive accuracy under the MAE criterion. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	DM Test						
EW Predictions	Carhart	FF6	FF6Liq	BS	DHS	HXZ	ALASSO
SD-Sparse	-3.71^{***}	-3.42^{***}	-3.40^{***}	-3.60^{***}	-1.08	-2.08^{**}	-4.41^{***}
	(0.000)	(0.001)	(0.001)	(0.000)	(0.282)	(0.039)	(0.000)
Panel Predictions							
SD-Sparse	-4.72^{***}	-4.98^{***}	-4.68^{***}	-4.04^{***}	-14.83^{***}	-3.06^{***}	-3.49^{***}
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.001)

asset returns out-of-sample under linear regression specification. The Diebold and Mariano (1995) t-statistics and their corresponding p-values show that the SD-Sparse factor model predictions are significantly more accurate than almost all leading factor models at 1% level of statistical significance. The SD-Sparse report equivalent performance only in the case of the DHS model predictions on the equally-weighted portfolio of test portfolios.

6.2.2. Non-parametric test

We conduct out-of-sample backtesting experiments on a rolling window basis to compare the optimal bounding portfolios of equity portfolios formed on the residuals plus alpha for each factor model. Those experiments allow us to validate the robustness of in-sample results. Although we are aware that the performance of the optimal

portfolio formed on the residuals of the SD-Sparse model exhibited the worst performance in-sample, it is unknown a priori whether this holds in an out-of-sample analysis. This is because by construction, we form these portfolios at time t, based on the information prevailing at time t, while we reap the portfolio returns over the next month $[t, t+1]$. The out-of-sample test is a real-time exercise that avoids a potential look-ahead bias and mimics how a real-time investor acts in practice.

We resort to backtesting experiments on a rolling horizon basis. The rolling windows cover the period from January 1990 to December 2023. At each month, we use the abnormal returns obtained from the previous 20 years (i.e., 240 monthly residuals plus the factor alpha) from each factor model for each equity portfolio to calibrate the procedure. Then we solve the resulting optimization problem for the Stochastic Bounding and get the optimal portfolio weights. We use these weights to compute the out-of-sample realized return of the portfolio over the period [t, t + 1]. We repeat this process by advancing t by one month, and dropping the first observation from the previous window, until we reach the end of the sample. Thus, we compute the ex-post realized returns over the entire period for the portfolios for each factor model. The out-of-sample performance metrics of the optimal bounding portfolio of equities' abnormal returns are presented in Table 9.

The above findings are even more evident than the relevant one in-sample. We observe now that the optimal bounding portfolio constructed on the abnormal returns given by the SD-Sparse factor mode, generate by far the lowest mean return, annualized Sharpe and U-P ratios over the out-of-sample period. The Lasso is the second-best model in terms of the lowest portfolio performance.

Table 9: **Performance Measures of the Optimal Portfolio of Abnormal Returns**. The table reports the out-of-sample performance measures, namely, the mean, standard deviation (St.Dev.), the annualized Sharpe ratio and UP ratio, of the optimal bounding portfolios formed based on the abnormal returns (i.e., residuals plus alpha) given for each equity portfolio by each factor model.

Performance Measures				
Models	Mean %	St.Dev %	An.Sharpe %	U-P ratio %
Carhart	0.20	1.48	8.52	66.43
FF6	0.20	1.96	6.48	62.94
FF6Liq	0.23	2.01	8.06	65.69
BS	0.21	1.92	7.15	68.17
HXZ	0.32	2.73	10.82	71.58
DHS	0.37	2.73	19.82	71.58
Lasso	0.13	1.71	3.63	61.98
SD-Sparse	0.07	0.115	-0.27	52.72

More importantly, the average return and standard deviation of returns of the optimal SD-Sparse portfolio are almost equal to the moments of the riskless rate for the same period (0,71% and 1,15% respectively), while the Sharpe ratio is very closed to zero. These observations indicate that the SD-Sparse factor model approximates the "true factor model", since after controlling for the systematic risk, the behavior of this diversified portfolio approximates the riskless asset. Such evidence justifies that the SD-Sparse factor model's abnormal returns contain minimum information compared to other factor models, suggested that the SD-Sparse better explains equity portfolio returns.

7. Conclusion

This study advances the asset pricing literature by proposing a novel factor model firmly grounded in the Stochastic Dominance (SD) framework. Utilizing the Sparse Spanning method of Arvanitis et al. (2025), we systematically identify a dominant combination of factors from a comprehensive set initially comprising 24 prominent factors featured in existing factor models. Extending our analysis further, we incorporate an expanded universe consisting of 177 factors, including the 153 factors from Jensen et al. (2023). Empirical findings consistently indicate that our sparse second-order stochastic dominance (SSD)-based selection methodology robustly identifies an optimal ten-factor model, which continues to deliver superior explanatory and predictive performance.

Specifically, our proposed SD-Sparse model substantially outperforms widely acknowledged benchmarks such as Carhart (1997), Fama and French (2015), Hou et al. (2018), and Daniel et al. (2020) across rigorous parametric and non-parametric asset pricing tests conducted both in- and out-of-sample. Additionally, the SD-Sparse model demonstrates clear superiority over advanced machine learning methodologies, such as the adaptive LASSO specification used by Chen et al. (2024), emphasizing both statistical robustness and economic interpretability.

By relaxing traditional restrictive assumptions of mean-variance frameworks (Huberman and Kandel, 1987), our Sparse Spanning approach explicitly optimizes generalized utility functions reflecting investor preferences for higher-order risks such as skewness and kurtosis. The resulting SD-Sparse model achieves the highest ex-

planatory power (cross-sectional R^2), significantly improves prediction accuracy—as evidenced by statistically significant Diebold and Mariano (1995) tests—and produces abnormal equity returns whose optimized portfolios closely mimic risk-free performance, underscoring its proximity to the theoretically ideal factor set.

This research demonstrates that stochastic dominance methods provide a robust, economically intuitive alternative to purely statistical machine learning factor-selection techniques, directly addressing issues of interpretability and economic relevance. By integrating theoretical rigor with robust empirical validation, our study bridges advanced statistical methodologies and conventional asset pricing theories, setting a strong precedent for further applications of stochastic dominance in portfolio choice and empirical asset pricing research.

References

Arvanitis, S., T. Post, and N. Topaloglou (2019). Stochastic spanning. *Journal of Business & Economic Statistics 37*(4), 573–585.

Arvanitis, S., T. Post, and N. Topaloglou (2021). Stochastic bounds for reference sets in portfolio analysis. *Management Science 67*(12), 7737–7754.

Arvanitis, S., O. Scaillet, and N. Topaloglou (2024). Spanning analysis of stock market anomalies under prospect stochastic dominance. *Management Science 70*(7), 6002–6025.

Arvanitis, S., O. Scaillet, and N. Topaloglou (2025). Sparse spanning portfolios and under-diversification with second-order stochastic dominance. *Available at SSRN*.

Asness, Clifford Frazzini, A. and L. H. Pedersen (2019). Quality minus junk. *Review of Accounting Studies 24*(1), 34–112.

Asness, C. and A. Frazzini (2013). The devil in hml's details. *Journal of Portfolio Management 39*(4), 49–68.

Barillas, F. and J. Shanken (2018). Comparing asset pricing models. *The Journal of Finance 73*(2), 715–754.

Berk, J. B. and J. H. van Binsbergen (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics 119*(1), 1–23.

Bryzgalova, S., J. Huang, and C. Julliard (2023). Bayesian solutions for the factor zoo: We just ran two quadrillion models. *Journal of Finance 78*(1), 487–557.

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

Chalamandaris, G., K. Pukthuanthong, and N. Topaloglou (2024). Examining anomalies from an investment perspective. Working paper.

Chen, A. Y. and T. Zimmermann (2022). Open source cross-sectional asset pricing. *Critical Finance Review 11*(2), 207–264.

Chen, Y., Z. Li, Y. Tang, and G. Zhou (2024). Anomalies as new hedge fund factors. Working paper.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance 66*(4), 1047–1108.

Constantinides, G. M., M. Czerwonko, J. Carsten Jackwerth, and S. Perrakis (2011). Are options on index futures profitable for risk-averse investors? empirical evidence. *The Journal of Finance 66*(4), 1407–1437.

Daniel, K., D. Hirshleifer, and L. Sun (2020). Short- and Long-Horizon Behavioral Factors. *The Review of Financial Studies 33*(4), 1673–1736.

De Roon, F. A., T. E. Nijman, and B. J. Werker (2003). Currency hedging for international stock portfolios: The usefulness of mean–variance analysis. *Journal of Banking & Finance 27*(2), 327–349.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636.

Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance 75*(3), 1327–1370.

Frazzini, A. and L. H. Pedersen (2014). Betting against beta. *Journal of Financial Economics 111*(1), 1–25.

Freyberger, J., A. Neuhierl, and M. Weber (2020, 04). Dissecting characteristics nonparametrically. *The Review of Financial Studies 33*(5), 2326–2377.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica 57*(5), 1121–1152.

Giglio, S. and D. Xiu (2021). Asset pricing with omitted factors. *Journal of Political Economy 129*(7), 1947–1990.

Harvey, C. R. and Y. Liu (2021). Lucky factors. *Journal of Financial Economics 141*(2), 413–435.

Hodder, J. E., J. C. Jackwerth, and O. Kolokolova (2015). Improved portfolio choice using second-order stochastic dominance. *Review of Finance 19*(4), 1623–1647.

Hou, K., C. Xue, and L. Zhang (2018, 12). Replicating anomalies. *The Review of Financial Studies 33*(5), 2019–2133.

Hou, K., C. Xue, and L. Zhang (2020). Replicating Anomalies. *The Review of Financial Studies 33*(5), 2019–2133.

Huberman, G. and S. Kandel (1987). Mean-variance spanning. *The Journal of Finance 42*(4), 873–888.

Jackwerth, J. C., G. M. Constantinides, and S. Perrakis (2009). Mispricing of s&p 500 index options. *Review of Financial Studies 22*(3), 1247–1277.

Jagannathan, R. and Z. Wang (1998). An asymptotic theory for estimating beta-pricing models using cross-sectional regression. *The Journal of Finance 53*(4), 1285–1309.

Jensen, T. I., B. Kelly, and L. H. Pedersen (2023). Is there a replication crisis in finance? *The Journal of Finance 78*(5), 2465–2518.

Kan, R., C. Robotti, and J. Shanken (2013). Pricing model performance and the two-pass cross-sectional regression methodology. *The Journal of Finance 68*(6), 2617–2649.

Kim, S., R. A. Korajczyk, and A. Neuhierl (2020, 09). Arbitrage portfolios. *The Review of Financial Studies 34*(6), 2813–2856.

Levy, H. (2016). *Stochastic dominance: Investment Decision Making Under Uncertainty*. 3rd Edition, Springer.

Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics 96*(2), 175–194.

Moskowitz, T. J., Y. H. Ooi, and L. H. Pedersen (2012). Time series momentum. *Journal of Financial Economics 104*(2), 228–250. Special Issue on Investor Sentiment.

Pastor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*(3), 642–685.

Perrakis, S. (2019). *Stochastic Dominance Option Pricing: An Alternative Approach to Option Market Research*. Number ISBN: 9783030115906. Springer.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*(3), 341–360.

Russel and Seo (1989). *Representative Sets for Stochastic Dominance Rules, in Studies in the Economics of Uncertainty*. Springer-Verlag.

Shanken, J. (2015, 05). On the Estimation of Beta-Pricing Models. *The Review of Financial Studies 5*(1), 1–33.

Stambaugh, R. F., J. Yu, and Y. Yuan (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics 104*(2), 288–302. Special Issue on Investor Sentiment.

Stambaugh, R. F. and Y. Yuan (2017). Mispricing Factors. *The Review of Financial Studies 30*(4), 1270–1315.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological) 58*(1), 267–288.

Whang, Y.-J. (2019). *Econometric Analysis of Stochastic Dominance: Concepts, Methods, Tools, and Applications*. Cambridge University Press.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101*(476), 1418–1429.

A. Appendix: Computational strategies

Here we give the computational strategies for the spanning and the bounding tests.

A.1. Computational Strategy for Spanning

The test statistic η can be also represented in terms of expected utility as:

$$\eta(\mathbb{P}) := \sup_{\boldsymbol{\lambda}\in\Lambda; u\in\mathcal{U}} \inf_{\kappa\in \mathrm{K}} \mathbb{E}_{\mathbb{P}} \left[u\left(X^{\mathrm{T}}\boldsymbol{\lambda}\right) - u\left(X^{\mathrm{T}}\boldsymbol{\kappa}\right)\right]; \tag{8}$$

$$\mathcal{U} := \left\{ u \in \mathcal{C}^0 : u(y) = \int_{\underline{x}}^{\overline{x}} v(x) r(y;x) dx \; v \in \mathcal{V} \right\}; \tag{9}$$

$$\mathcal{V} := \left\{ v : \mathcal{X} \to \mathbb{R}_+ : \int_{\mathcal{X}} v\left(x\right) = 1 \right\} \tag{10}$$

$$r(y;x) := (y-x)1(y \leq x),\ (x,y) \in \mathcal{X}^2. \tag{11}$$

$\mathcal{U}$ is comprised of normalized, increasing and concave utility functions that are constructed as convex mixtures of elementary Russell and Seo (1989) ramp functions $r(y;x),\ x \in \mathcal{X}$. This implies that K spans Λ, iff for any $\boldsymbol{\lambda} \in \Lambda$ there exists some $\boldsymbol{\kappa} \in \mathrm{K}$, weakly preferred to the former, by every utility in $\mathcal{U}$. Equivalently, spanning occurs iff no risk averter in $\mathcal{U}$ loses expected utility from the excision Λ-K. This representation can be used for the numerical implementation of the associated testing procedure.

The test statistic can be expressed as:

$$\eta_T := \sqrt{T} \sup_{u\in\mathcal{U}} \left(\sup_{\boldsymbol{\lambda}\in\Lambda} \mathbb{E}_{\mathbb{P}_T} \left[u\left(X^{\mathrm{T}}\boldsymbol{\lambda}\right)\right] - \sup_{\boldsymbol{\kappa}\in \mathrm{K}} \mathbb{E}_{\mathbb{P}_T} \left[u\left(X^{\mathrm{T}}\boldsymbol{\kappa}\right)\right] \right). \tag{12}$$

The computational complexity of evaluating η_T stems from the functional complexity of the set $\mathcal{U}$. However, due to the properties of the admissible utilities, Arvanitis et al., (2019) approximate every element of $\mathcal{U}$ with arbitrary prescribed accuracy using a finite set of increasing and concave piecewise-linear functions in the following way:

Let N_1, N_2 denote integers greater than or equal to 2. First $\mathcal{X}$ is partitioned into N_1 equally spaced values as $\underline{x} = z_1 < \cdots < z_{N_1} = \overline{x}$, where $z_n := \underline{x} + \frac{n-1}{N_1-1}(\overline{x} - \underline{x})$, $n = 1, \cdots, N_1$. Second, $[0, 1]$ is partitioned as $0 < \frac{1}{N_2-1} < \cdots < \frac{N_2-2}{N_2-1} < 1$. Using those partitions, consider:

$$\underline{\eta_T} := \sqrt{T} \sup_{u \in \underline{\mathcal{U}}} \left(\sup_{\boldsymbol{\lambda} \in \Lambda} \mathbb{E}_{\mathbb{P}_T} \left[u\left(X^{\mathrm{T}} \boldsymbol{\lambda}\right) \right] - \sup_{\boldsymbol{\kappa} \in \mathrm{K}} \mathbb{E}_{\mathbb{P}_T} \left[u\left(X^{\mathrm{T}} \boldsymbol{\kappa}\right) \right] \right); \tag{13}$$

$$\underline{\mathcal{U}} := \left\{ u \in \mathcal{C}^0 : u(y) = \sum_{n=1}^{N_1} v_n r(y; z_n) \; v \in V \right\}; \tag{14}$$

$$V := \left\{ v \in \left\{ 0, \frac{1}{N_2 - 1}, \cdots, \frac{N_2 - 2}{N_2 - 1}, 1 \right\}^{N_1} : \sum_{n=1}^{N_1} v_n = 1 \right\}. \tag{15}$$

By construction, every $u \in \underline{\mathcal{U}}$ consists of at most N_2 linear line segments with endpoints at N_1 possible outcome levels. Furthermore $\underline{\mathcal{U}} \subset \mathcal{U}$, it is finite as it has $N_3 := \frac{1}{(N_1-1)!} \prod_{i=1}^{N_1-1} (N_2 + i - 1)$ elements and $\underline{\eta_T}$ approximates η_T from below as the partitioning scheme is refined ($N_1, N_2 \to \infty$). Then for every $u \in \underline{\mathcal{U}}$, the two embedded maximization problems in (13) can be solved using linear programming.

Consider

$$c_{0,n} := \sum_{m=n}^{N_1} \left(c_{1,m+1} - c_{1,m}\right) z_m; \tag{16}$$

$$c_{1,n} := \sum_{m=n}^{N_1} w_m; \tag{17}$$

$$\mathcal{N} := \{n = 1, \cdots, N_1 : v_n > 0\} \bigcup \{N_1\}. \tag{18}$$

Then for any given $u \in \underline{\mathcal{U}}$, $\sup_{\boldsymbol{\lambda}\in\Lambda} \mathbb{E}_{\mathbb{P}_T}\left[u\left(X^{\mathrm{T}}\boldsymbol{\lambda}\right)\right]$ is the optimal value of the objective function of the following linear program:

$$\max T^{-1} \sum_{t=1}^{T} y_t \tag{19}$$

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

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

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

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

The linear programming problem always has a feasible solution and has $\mathcal{O}(T + M)$ variables and constraints, making it manageable for typical data dimensions.

The empirical application is based on the entire available history of quarterly investment returns to a standard set of benchmark assets ($N = 12$, $T = 360$), and uses $N_1 = 10$ and $N_2 = 5$. This gives $N_3 = \frac{1}{9!}\prod_{i=1}^{9}(4 + i) = 715$ distinct utility functions and $2N_3 = 1,430$ small linear programming problems. This is perfectly

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

A.2. Computational Strategy for Bounding

The bounding methodology considers all portfolios in Λ rather than a single given portfolio and the joint empirical support generally consists of infinitely many points, introducing the need for discretization.

Let $\hat{\mathcal{X}}_\Lambda := \left[\hat{A}_\Lambda, \hat{B}_\Lambda\right]$, $\hat{A}_\Lambda := \min_{\Lambda,\lambda} \boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda}$ and $\hat{B}_\Lambda := \max_{\Lambda,t} \boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda}$. The computational strategy partitions $\hat{\mathcal{X}}_\Lambda$ using J equally spaced grid points $\hat{x}_j := \hat{A}_\Lambda + (j-1)\left(\hat{B}_\Lambda - \hat{A}_\Lambda\right)(J-1)^{-1}$, $j = 1, \cdots, J$. For every grid point, let $\hat{L}^*_{\Lambda,j} := \min_\Lambda L(\boldsymbol{\lambda}, \hat{x}_j, \hat{\mathbb{P}})$,$j = 1, \cdots, J$.

The approximation $\xi(\Lambda, \Lambda, \sqrt{T}\hat{\mathbb{P}}) \approx \sqrt{T}\chi_J(\Lambda, \Lambda, \hat{\mathbb{P}})$ is used, where

$$\begin{aligned}\chi_J(\Lambda, \Lambda, \hat{\mathbb{P}}) &:= \min_{\Lambda} \max_{\Lambda,j} D(\boldsymbol{\lambda}, \boldsymbol{\lambda}, \hat{x}_j, \hat{\mathbb{P}}) \\ &= \min_{\Lambda} \max_{j} \left(L(\boldsymbol{\lambda}, \hat{x}_j, \hat{\mathbb{P}}) - \hat{L}^*_{\Lambda,j}\right) \\ &= \min_{\Lambda,\sigma} \left(\sigma : L(\boldsymbol{\lambda}, \hat{x}_j, \hat{\mathbb{P}}) - \sigma \leq \hat{L}^*_{\Lambda,j},\ j = 1, \cdots, J\right).\end{aligned} \tag{20}$$

The number of grid points J can be chosen to achieve the desired balance between accuracy and computer time. In the experience of the authors, $J = 100$ yields very high accuracy for typical applications.

The approximate $\chi_J(\Lambda, \Lambda, \hat{\mathbb{P}})$ can be computed by solving a series of linear pro-

grams using linear relaxations in spirit of Rockafellar et al. (2000). Each value $\hat{L}^*_{\Lambda,j}$, $j = 1, \cdots, J$, can be computed as the optimal value of the objective function of the following program:

$$\begin{aligned} &\min T^{-1} \sum_{t=1}^{T} \eta_{j,t} \\ -\eta_{j,t} - \boldsymbol{x}_t^{\mathrm{T}} \boldsymbol{\lambda} &\leq -\hat{x}_j, \ t = 1, \cdots, T; \\ \eta_{j,t} &\geq 0, \ t = 1, \cdots, T; \\ \boldsymbol{\lambda} &\in \Lambda. \end{aligned} \tag{21}$$

Given the solutions to the J problems, $\chi_J(\Lambda, \Lambda, \hat{\mathbb{P}})$ can be computed by solving the linear program:

$$\begin{aligned} &\min \sigma \\ T^{-1} \sum_{t=1}^{T} \theta_{j,t} - \sigma &\leq \hat{L}^*_{\Lambda,j}, \ j = 1, \cdots, J; \\ -\theta_{j,t} - \boldsymbol{x}_t^{\mathrm{T}} \boldsymbol{\lambda} &\leq -\hat{x}_j, \ j = 1, \cdots, J; \ t = 1, \cdots, T; \\ \theta_{j,t} &\geq 0, \ j = 1, \cdots, J; \ t = 1, \cdots, T; \\ \boldsymbol{\lambda} &\in \Lambda; \\ \sigma &\text{ free.} \end{aligned} \tag{22}$$

The optimal solution $\lambda \in \Lambda$ identifies the factors portfolio that stochastically spans any other portfolio in Λ, and is not spanned by any portfolio of factors.