Smart Stochastic Discount Factors

SOFONIAS ALEMU KORSAYE, ALBERTO QUAINI and FABIO TROJANI*

First version: May 2019. This version: August 31, 2020

ABSTRACT

We develop a unifying theoretical framework for selecting model-free stochastic discount factors (SDFs) in arbitrage-free markets under general convex constraints on pricing errors, and show that such SDFs arise in a wide range of economies featuring, e.g., various forms of frictions, ambiguous asset payoffs, asymptotic no arbitrage conditions under Ross' Arbitrage Pricing Theory (APT), or a need for regularization in large asset markets. We introduce a new family of minimum variance SDFs incorporating APT pricing error bounds, which are designed to optimize the tradeoff between pricing accuracy and the SDF ability to comove with systematic asset return risks. Empirically, we find that a model-free adaptation of an SDF under the CAPM, which exactly prices market risk but otherwise constrains the amount of mispricing across assets with a model-free APT pricing error bound, generates an optimal tradeoff.

Keywords: SDF, Pricing Errors, Minimum Dispersion SDFs, SDF Bounds, Market Frictions, Portfolio Regularization, Arbitrage Pricing Theory, Factor Models, Fundamental Theorem of Asset Pricing

---

*Sofonias Alemu Korsaye (email: Sofonias.Korsaye@unige.ch) and Fabio Trojani (email: Fabio.Trojani@unige.ch) are with University of Geneva, Geneva Finance Research Institute & Swiss Finance Institute. Alberto Quaini (email: Alberto.Quaini@unige.ch) is with University of Geneva, Geneva Finance Research Institute. We thank Caio Almeida, Federico Bandi, Tony Berrada, Federico Carlini, Ines Chaieb, George Constantinides, Jerome Detemple, Patrick Gagliardini, Eric Ghysels, Lars Hansen, Oliver Linton, Stefan Nagel, Paolo Porchia, Olivier Scaillet, Paul Schneider, Andrea Vedolin, Michael Weber, Dacheng Xiu, conference participants at the 2019 SoFiE annual Meeting in Shanghai, the 2019 European Econometric Society Meeting in Manchester, the 2019 Workshop on Big Data and Economic Forecasting in Ispra, the 2019 Conference on Quantitative Finance and Financial Econometrics in Marseille, the 2019 Financial Econometrics Conference in Toulouse, the 2019 Vienna Congress on Mathematical Finance, the 2019 Swiss Finance Institute Research days, the 2019 ESSEC Workshop on Monte Carlo Methods and Approximate Dynamic Programming with Applications in Finance, the 2019 Paris December Finance Meeting, the 2020 Virtual World Congress of the Econometric Society in Milano, the 2020 Remote Seminar Series on Computational Economics and Finance, the 2020 Virtual Derivatives Workshop and seminar participants at BI Norwegian Business School, Bocconi University, Mc Gill University, University of Zurich, University of Lugano, University of Geneva, University of Lund, JRC in Ispra and Luiss University. All errors are ours.

This paper provides a novel unifying framework for selecting model-free stochastic discount factors in arbitrage-free markets under general convex constraints on pricing errors. In our approach, some assets may be priced exactly, e.g., because they correspond to well-understood tradable risk exposures delivering robust compensations for some form of systematic risk, while other assets may be priced with errors, e.g., because they may involve non negligible transaction costs in only partially liquid markets or because they may give rise to opaque or not perfectly understood risk exposures. We call SDFs allowing for such a joint treatment of exact and inexact pricing constraints Smart SDFs (S–SDFs) because of their affinity to the motivation underlying smart beta strategies in modern asset management.[1]

We first theoretically establish the economic importance of S–SDFs and show that they are founded in a large class of arbitrage-free economies with either frictions or ambiguity. In economies with frictions, S–SDFs represent positive linear pricing rules in markets with convex transaction costs. Hence, they give rise to pricing errors reflecting the given transaction cost structure. In economies with ambiguity, they represent positive linear pricing rules in a frictionless market, in which an unobserved marginal investor eliminates arbitrage opportunities perceived under her/his probability belief. Since the investor's belief and the reference belief describing asset payoffs may differ, asset valuations are ambiguous and pricing errors may arise. Using several seminal asset pricing approaches in the literature, we show how they can be explicitly founded and understood within our model-free S–SDF framework. Such approaches include, e.g., different versions of Ross [1976] Arbitrage Pricing Theory (APT), Cochrane and Saa-Requejo [2000] good-deal-bounds theories, Chen et al. [2020b] theory for a robust identification of investor beliefs, and recent regularizations of SDFs in large asset markets such as Kozak et al. [2020].

Second, in order to develop a coherent analysis and selection framework for S–SDFs, we introduce minimum dispersion S–SDFs as the solutions of general minimum dispersion problems with convex pricing constraints. Under the maintained assumption that the underlying economy with either frictions or ambiguity is arbitrage-free, we formally establish a strong duality between minimum dispersion S–SDF problems and corresponding penalized portfolio selection problems, in which the choice of the penalization stays in a one-to-one relation to the pricing constraint imposed on the S–SDF. This duality result characterizes the optimal S–SDF as a simple transformation of the penalized optimal portfolio payoff. Therefore, it also directly explains key S–SDF properties, such as sparsity, in terms of the choice of the underlying pricing error metric.

[1]Natural candidates of exactly priced assets for empirical asset pricing are tradeable portfolios replicating established systematic risk exposures in the literature, such market risk or smart beta exposures like, e.g., size, value and momentum, among many others.

Third, we exploit our general duality framework to build a family of minimum variance APT S–SDFs that is suited for empirically studying the relation between risk and return in cross-sections of asset returns. S–SDFs in this family satisfy a model-free APT pricing error bound and can account for a good range of S–SDF sparsity features. More importantly, they optimally synthesize the attainable tradeoffs between S–SDF spanning features for systematic risks and S–SDF pricing accuracy. Indeed, by definition minimum variance S–SDF bounds capture the downward sloping optimal relation between S–SDF variance and pricing error size. In parallel, the underlying S–SDF pricing error constraints force an inverse relation between S–SDF variance and the S–SDF co-movement with asset returns. Hence, any degree of S–SDF pricing accuracy implicitly fixes a corresponding degree of systematic S–SDF co-movement with asset returns, which in turn leads to a tradeoff between pricing accuracy and spanning features for systematic risks. We empirically characterize such tradeoffs for several cross-sections of characteristics sorted portfolio returns of varying dimension and relate them to key APT S–SDF building features, such as the underlying set of exactly priced assets and the implied degree of S–SDF or pricing error sparsity.[2]

Empirically, a well-defined tradeoff between S–SDF spanning features for systematic risks and S–SDF pricing accuracy requires existence of corresponding S–SDFs in arbitrage-free asset markets. We find that our family of APT S–SDFs produces a well-defined tradeoff also in markets with a growing number of assets, in which the standard SDF dualities in Hansen and Jagannathan [1991] and Almeida and Garcia [2016], among others, fail.[3] Indeed, we document that minimum variance empirical SDFs exactly pricing all assets can fail to exist in such settings, which is evidence of a violation of the standard no-arbitrage condition under the empirical distribution of asset payoffs. In contrast, we show that minimum variance APT S–SDFs still exist in such settings, which are founded in an underlying frictionless arbitrage-free economy with ambiguity.

We further follow a simple intuition to crucially exploit the joint treatment of exact and approximate pricing constraints in our APT S–SDF framework. By their nature, minimum variance S–SDFs always maximize the co-movement with any exactly priced asset return. Therefore, the maximal degree of S–SDF spanning for systematic return risks is always obtained by forcing exact pricing of the first principal component of asset returns alone. We empirically find that minimum variance APT S–SDFs exactly pricing only the first principal component of returns deliver the best out-of-sample tradeoff between S–SDF time series spanning

[2]Our closed-form minimum variance APT S–SDFs naturally synthesize these tradeoffs consistently with established cross-sectional GLS R^2 metrics of pricing accuracy in the literature. As emphasized in Lewellen et al. [2010], a cross-sectional GLS R^2 metric is preferable for interpreting the empirical asset pricing evidence derived from test assets with a possibly hidden strong factor structure, which is typically the case for tests assets given by characteristics-sorted portfolios.

[3]See also Borwein [1993] for examples of such theoretical duality failures.

features and pricing accuracy. Importantly, the structure of these S–SDFs is particularly transparent and simple, as it is sparse in two clearly interpretable portfolio returns: A traded systematic return spanned by the first principal component of asset returns and a second optimal portfolio of returns, which are orthogonal to the traded systematic return and feature a bounded mispricing under the APT. In comparison to S–SDFs restricted to depend on a sparse subset of individual asset returns or a sparse subset of principal components of returns, we find that such S–SDFs produce clearly suboptimal tradeoffs.

Finally, we show that a model-free adaptation of an empirical SDF under the CAPM, which exactly prices market excess returns but otherwise constrains the amount of mispricing across assets with a standard APT pricing error bound, produces a nearly optimal tradeoff between pricing accuracy and spanning features for systematic risks. In contrast, other intuitive choices of tradable systematic risk exposures, such as value, size or other Fama-French factors, imply a suboptimal tradeoff. This finding is a natural consequence of the typically large co-movement between market risk and the first principal component of sorted portfolio returns in most applications.

After a review of the literature, the remainder of the paper proceeds as follows. In Section 1, we define S–SDFs and prove their existence in corresponding arbitrage-free financial markets. Moreover, we study minimum dispersion S–SDFs and employ Fenchel duality to reduce the infinite dimensional optimization problem defining such S–SDFs to a corresponding finite dimensional penalized portfolio problem. We then characterize minimum dispersion S–SDFs as simple transformations of the optimal portfolio payoffs resulting from such penalized portfolio problems. In Section 2, we address in more detail various relevant examples of S–SDFs. Here, we put particular emphasis on the notion of minimum dispersion APT S–SDFs and clarify their main theoretical properties. Section 3 presents our empirical analysis of APT S–SDFs, while Section 4 concludes.

Related Literature

Our paper is related to various important strands of the literature. A first strand studies the no-arbitrage foundation of strictly positive SDFs with various versions of the fundamental theorem of asset pricing. Ross [1978], Harrison and Kreps [1979], and Jouini and Kallal [1995] prove early versions of the fundamental theorem for frictionless economies and for economies with sublinear transaction costs, respectively. In our arbitrage-free foundation of S–SDFs, we borrow from the market structure in Jouini and Kallal [1995] and build a class of suitable economies, in which a strictly positive S–SDF exists if and only if markets are arbitrage-free. This result provides a unique novel foundation for the existence of S–SDFs with general pricing error properties, which gives rise to two distinct economic interpretations based on corresponding arbitrage-free economies with either frictions or ambiguity.[4]

A second prominent direction in the literature studies model-free diagnostics for asset pricing models, based on non-parametric minimum variability bounds for admissible SDFs. Hansen and Jagannathan [1991] focus on minimum variance SDFs that impose zero pricing errors on all assets. Snow [2011], Stutzer [1995], Bansal and Lehmann [1997], Backus et al. [2014] and Almeida and Garcia [2016] extend this setting to SDF bounds for higher moments, Kullback-Leibler divergence, entropy, and an extended nesting family of Cressie and Read [1984] divergence measures allowing for a stronger sensitivity to negative skewness. Orlowski et al. [2016] develop a theory of arbitrage-free dispersion extending all these approaches, which gives rise to multivariate minimum dispersion SDF bounds in settings with multiple SDF components. Luttmer [1996] is the first to consider minimum variance S–SDFs with non zero pricing errors and the ensuing bounds on SDF variance within arbitrage-free economies with frictions such as short-selling constraints or margin requirements. Our framework extends this literature to markets with general pricing error structures, characterizes in closed-form the resulting minimum dispersion S–SDFs and S–SDF bounds, via the solution of a penalized portfolio problem, and establishes their existence in arbitrage-free markets.

Our general S–SDF framework is essential to provide a distinct foundation for other important directions in the literature. We show that the good-deal bound SDFs in Cochrane and Saa-Requejo [2000] and the robust SDFs in Kozak et al. [2020] are minimum variance S–SDFs corresponding to a particular choice of the pricing error metric, while the robust investor belief identified in Chen et al. [2020b] is interpretable as a S–SDF minimizing a robust notion of stochastic dispersion, such as Kullback-Leibler dispersion. Similarly, we embed the asset pricing predictions in Ross [1976] APT and its version with misspecification in Uppal et al. [2019] in our model-free S–SDF framework by means of a corresponding family of APT S–SDFs. We make use of these S–SDFs to obtain new useful interpretations of the APT pricing predictions and new model-free ways to implement them empirically. The formalizations of the APT in Hubermann [1982],

[4]The foundation of our S–SDF framework in arbitrage-free markets with ambiguity provides a useful link also to the large literature studying ambiguity and ambiguity aversion in asset pricing; see Epstein and Schneider [2010] and Guidolin and Rinaldi [2013] for a review. In most of this literature, ambiguity aversion gives rise to expected excess returns fully explained by the return covariance with a corresponding SDF, which consists of two distinct components pricing the risk and the ambiguity of asset returns, respectively. In our framework, ambiguity instead gives rise to S–SDFs that may imply non zero pricing errors on some assets, because the investor's belief and the belief used to describe asset payoffs may disagree on some of their zero probability assessments.

Chamberlain [1983], Chamberlain and Rothschild [1983], Ingersoll [1984], and Uppal et al. [2019] are based on an asymptotic no-arbitrage condition that gives rise to bounded pricing errors under a quadratic APT metric. Therefore, we construct a convenient family of minimum variance APT S–SDFs replicating the APT asset pricing predictions, under varying assumptions on the set of tradeable systematic risk factors and the S–SDF sparsity features. We show that such APT S–SDFs optimally capture the tradeoff between S–SDF pricing accuracy and the S–SDF ability to span systematic asset return risks.

Our approach is linked naturally also to recent directions applying various penalization techniques from the machine learning literature in empirical asset pricing. Freyberger et al. [2020] incorporate a nonparametric specification of nonlinear dependencies between expected returns and characteristics, finding that a small set of characteristics explains expected returns. Gu et al. [2020b] find evidence that machine learning methods allowing for nonlinear predictive relations, such as trees and neural networks, outperform in out-of-sample return prediction. Feng et al. [2020] propose a model-selection method for evaluating the contribution of new factors to an unknown SDF, above and beyond the contribution of a high-dimensional set of control factors, and find that only a few recently discovered factors in the literature have incremental explanatory power. Different from this literature, we propose a unifying framework that directly constructs minimum dispersion S–SDFs with target pricing error properties from a cross-section of asset returns. Importantly, allowing for pricing errors enables us to coherently address settings where assets' expected excess returns may not be fully explained by a covariation with an SDF.

A second part of the literature applying machine learning techniques in empirical asset pricing studies empirical SDFs with improved out-of-sample pricing properties. Kozak et al. [2020] and Qiu and Otsu [2018] introduce versions of such SDFs, using lasso, ridge, and elastic net penalizations. We show that the population versions of these SDFs are interpretable as minimum dispersion S–SDFs and we derive in closed-form their pricing error properties. Since such S–SDFs can fail to exist in markets of growing dimensions, we propose a simple data-driven methodology, which ensures existence of empirical minimum dispersion S–SDFs and optimizes the resulting tradeoff between S–SDF pricing accuracy and time series explanatory power. Finally, while in our empirical study we do not explicitly model time-varying conditional information, our framework provides a useful foundation also for more recent empirical SDFs that incorporate the information of a large set of conditioning variables, such as Gu et al. [2020a] and Chen et al. [2020a]. A common starting point of these approaches is a no-arbitrage assumption equivalent to the existence of a SDF exactly pricing all assets. Distinctions between these approaches arise instead in the way how they regularize the very high-dimensional unconditional pricing constraints that need to hold theoretically. We offer guidelines for understanding a penalization choice in terms of a constraint on the non-zero pricing errors of a corresponding S–SDF, which may exist in markets of growing dimension when an SDF may not.

1 Theory of Smart Stochastic Discount Factors

Consider an economy consisting of a fixed number N of securities with random payoffs $\boldsymbol{X} := (X_n)_{n=1}^N$ at time 1 and associated quoted prices $\boldsymbol{P} \in \mathbb{R}^N$ at time 0. The vector of pricing errors for these securities, under a generic linear pricing rule represented by a stochastic discount factor M, is given by $\mathbb{E}[M\boldsymbol{X}] - \boldsymbol{P}$.[5] In order to allow for the possibility of non zero pricing errors on some assets, we partition the market into what we label *sure* and *dubious* securities, by means of two index sets, $S \subset \{1, \ldots, N\}$ and its complement $D := \{1, \ldots, N\} \setminus S$, having cardinalities N_S and N_D, respectively. Note that one of these two index sets might be empty, in which case either only sure or only dubious securities exist. The choice of the partitioning is determined by the properties of the asset pricing problem at hand and its rationale is precisely to leave open the possibility that the dubious assets may be priced with errors.[6]

1.1 Definition of S–SDFs

Denote by $\boldsymbol{P}_S := (P_n)_{n\in S}$ ($\boldsymbol{P}_D := (P_n)_{n\in D}$) the vector of prices and by $\boldsymbol{X}_S := (X_n)_{n\in S}$ ($\boldsymbol{X}_D := (X_n)_{n\in D}$) the vector of payoffs of the sure (dubious) assets. We control the size and geometry of pricing errors by imposing the following pricing constraints:

$$\mathbb{E}[M\boldsymbol{X}_S] - \boldsymbol{P}_S = \boldsymbol{0} \qquad \text{and} \qquad h(\mathbb{E}[M\boldsymbol{X}_D] - \boldsymbol{P}_D) \leq \tau \ , \tag{1}$$

where $\tau \geq 0$ and $h : \mathbb{R}^{N_D} \to (-\infty, +\infty]$ is a convex pricing error function.[7]

Definition 1 (**Smart Stochastic Discount Factors**)**.** Given market partition $\{S, D\}$, pricing error function h and bound τ, a Smart Stochastic Discount Factor (S–SDF) is a nonnegative random variable M

[5] In multi-period economies, the (conditional) pricing errors under stochastic discount factor M are given by $\mathbb{E}[M\boldsymbol{X}|\mathcal{T}] - \boldsymbol{P}$, where $\mathcal{T}$ is the information set available at the time of the trade and $\mathbb{E}[\cdot|\mathcal{T}]$ the conditional expectation. As is standard in the literature, conditioning information can be incorporated by considering prices and payoffs instrumented by variables in information set $\mathcal{T}$, which gives rise to unconditional pricing constraints for managed portfolios of asset payoffs.

[6] We discuss below approaches for detecting endogenously stochastic discount factors implying ex-post zero pricing errors also for a non-empty subset of the dubious assets.

[7] An obvious implicit assumption that is always maintained in the sequel is that the set $\{\boldsymbol{\eta} \in \mathbb{R}^{N_D} : h(\boldsymbol{\eta}) \leq \tau\}$ of admissible pricing errors is not empty. Moreover, all convex functions are assumed for technical reasons as also closed and proper. A function h is closed if all its sub-level sets $\{\boldsymbol{\eta} : h(\boldsymbol{\eta}) \leq t\}$ are closed and it is proper if it is finite in at least one point. The set of closed convex and proper functions is broad enough to encompass all settings that are relevant for our work.

satisfying pricing constraints (1).

In Definition 1 positive linear pricing functional $\mathbb{E}[M\ \cdot]$ matches exactly the prices of sure payoffs and it implies pricing errors on dubious assets bounded by threshold τ under convex pricing error metric h. Clearly, any SDF that prices all assets exactly is a S–SDF, but the converse is not true. Intuitively, function h fixes the geometry of the pricing errors generated by an S–SDF and it thus crucially determines key S–SDFs properties, such as the structure of the resulting asset excess returns. Indeed, following standard expected excess return identity holds for any sure asset payoff under the given reference probability:

$$\mathbb{E}[X_n] - \frac{P_n}{\mathbb{E}[M]} = -Cov\left[\frac{M}{\mathbb{E}[M]}, X_n\right] \ , \quad n \in S \ . \tag{2}$$

In contrast, the expected excess returns of dubious assets satisfy:

$$\mathbb{E}[X_n] - \frac{P_n}{\mathbb{E}[M]} = -Cov\left[\frac{M}{\mathbb{E}[M]}, X_n\right] + \frac{\mathbb{E}[MX_n] - P_n}{\mathbb{E}[M]} \ , \quad n \in D \ . \tag{3}$$

It follows that the choice of the pricing error geometry directly influences the way how an S–SDF explains asset excess returns under a given reference probability, in terms of their exposure to S–SDF risk and asset pricing errors, respectively. With respect to sure assets, S–SDFs completely explain excess returns from their exposure to S–SDF risk, e.g., because they may be perceived as tradable exposures delivering robust compensations for some form of systematic risk. In contrast, on the dubious assets pricing errors may be tolerated, e.g., because they may be more difficult to trade if their trading involves non negligible transaction costs in possibly less liquid assets or because of the ambiguity of their payoffs.

Asset pricing constraints of the form (1) embed most of the existing asset pricing theories. An obvious one is the classical arbitrage-free framework of Harrison and Kreps [1979], in which all payoffs are exactly priced by a strictly positive SDF. Others include arbitrage-free economies with short-sale constraints or bid-ask spreads such as Luttmer [1996], in which unbounded pricing errors under a strictly positive SDF are only constrained to be negative. New examples of S–SDFs with bounded pricing errors, including S–SDFs consistent with Ross [1976]'s Arbitrage Pricing Theory, are introduced in Section 2. Before studying them in more detail, we characterize the no-arbitrage conditions ensuring existence of a strictly positive S–SDF with a given pricing error metric h. This provides a general economic foundation for an asset pricing approach based on S–SDFs.

1.2 Economic Foundation of S–SDFs

To characterize existence of strictly positive S–SDFs, we make use of a suitable arbitrage-free market structure. Asset payoffs are defined on probability space $(\Omega,\ \mathcal{F},\ \mathbb{P})$ and have finite $p-$th moment for some $p \in (1,\infty)$.[8] Portfolio positions $\boldsymbol{\theta}_S \in \mathbb{R}^{N_S}$ on the sure payoffs are modelled as involving no additional cost beside the uniquely given portfolio price $\boldsymbol{\theta}_S'\boldsymbol{P}_S$, while dubious portfolio positions $\boldsymbol{\theta}_D \in \mathbb{R}^{N_D}$ involve an additional cost measured by a sublinear cost function $\sigma : \mathbb{R}^{N_D} \to [0,+\infty]$.[9] In this way, we obtain a market of traded payoffs simply given by the set of all portfolio payoffs that are implementable with finite costs:

$$\mathcal{Z} := \{Z = \boldsymbol{X}'\boldsymbol{\theta}\ :\ \sigma(\boldsymbol{\theta}_D) < +\infty\}\ , \tag{4}$$

where $\boldsymbol{\theta} := (\boldsymbol{\theta}_S', \boldsymbol{\theta}_D')'$. Accordingly, we price traded payoffs with their smallest portfolio replication cost, which gives rise to a pricing functional π on $\mathcal{Z}$ given by:

$$\pi(Z) := \inf_{\boldsymbol{\theta}\in\mathbb{R}^N} \{\boldsymbol{P}'\boldsymbol{\theta} + \sigma(\boldsymbol{\theta}_D)\ :\ Z = \boldsymbol{X}'\boldsymbol{\theta}\}\ . \tag{5}$$

The next proposition makes explicit the class of functions σ allowing an arbitrage-free foundation to S–SDFs. The notion of arbitrage used for this result is that of a free lunch; see, e.g., Harrison and Kreps [1979].[10]

Proposition 1 (Existence of S–SDFs). *Consider a market $(\mathcal{Z},\pi)$ with cost function $\sigma = \sigma_h$, where*

$$\sigma_h(\boldsymbol{\theta}_D) := \sup_{\boldsymbol{\eta}\in\mathbb{R}^{N_D}} \{\boldsymbol{\eta}'\boldsymbol{\theta}_D\ \ :\ \ h(\boldsymbol{\eta}) \le \tau\}\ . \tag{6}$$

Such a market has no free lunches if and only if there exists a strictly positive S–SDF M, with finite $q-$th moment for $q = p/(p-1)$, satisfying pricing constraints (1).

According to Proposition 1, there always exists an arbitrage-free economy that founds a given S–SDF, and the converse holds true as well. Moreover, the geometry of the pricing errors of a given S–SDF, i.e., the set of pricing errors satisfying constraints (1), stays in a one-to-one relation with function σ_h founding such

[8] Denoting by L^p the space of $p-$integrable random variables, we adopt the partial order $\le$ defined by: $x \le y$ if and only if $\mathbb{P}(x \le y) = 1$. Also, the duality pair between L^p and L^q, for $1/p + 1/q = 1$, is given by $\mathbb{E}[xy]$, for any $x \in L^p$, $y \in L^q$ with $p \notin \{1,\infty\}$, in order to obtain compact proofs based on $L^p - L^q$ duality.

[9] Function σ is sublinear if $\sigma(\boldsymbol{\theta}_D + \tilde{\boldsymbol{\theta}}_D) \le \sigma(\boldsymbol{\theta}_D) + \sigma(\tilde{\boldsymbol{\theta}}_D)$ and $\sigma(a\boldsymbol{\theta}_D) = a\sigma(\boldsymbol{\theta}_D)$ for any $a \ge 0$ and any $\boldsymbol{\theta}_D, \tilde{\boldsymbol{\theta}}_D \in \mathbb{R}^{N_D}$. Intuitively, sublinear costs reflect the fact that a single execution of a portfolio position bears a cost no greater than that of multiple executions of the same position in separate orders.

[10] For a formal definition of a free lunch, see Assumption A.5 in Clark [1993]. A free lunch arises from the possibility to get arbitrarily close to an arbitrage opportunity, i.e., a traded positive payoff with zero or negative price. If the set of traded payoffs with zero or strictly negative price is a polyhedral cone, the notion of free lunch and arbitrage opportunity coincide; see Theorem 6 in Clark [1993]. Such a situation arises, e.g., when function σ has as a domain $\mathbb{R}^{N_D}$ or another polyhedral cone.

S–SDF.[11] It follows that, given a strictly positive S–SDF satisfying pricing error constraints (1), we can always uniquely identify the arbitrage-free economy with sublinear cost function σ_h founding it. Similarly, given sublinear cost function σ_h, we can always uniquely identify the convex set of pricing errors admitted by corresponding S–SDFs. Moreover, as we show in Section 2.1 below, function (6) is fully available in closed-form for a broad variety of practically relevant S–SDF settings.

The arbitrage-free characterization of S–SDFs in Proposition 1 relies on a definition of arbitrage opportunities that explicitly depends on the probability assessments under reference probability belief $\mathbb{P}$. Depending on the asset pricing context under scrutiny, such probability may or may not imply identical zero probability assessments as under the probability belief of the marginal investor in the underlying founding economy. If such zero probability assessments are identical, arbitrage opportunities perceived under each of these beliefs coincide and the arbitrage-free economy in Proposition 1 is directly interpretable as an economy with transaction costs, in which the marginal investor has eliminated any possible free lunch. In this context, the pricing errors associated with the strictly positive S–SDF M of Proposition 1 are a direct consequence of the sublinearity of pricing functional (4), which is explicitly rationalized by the marginal investor as the consequence of transaction costs of the form $\sigma = \sigma_h$. In particular, in settings where σ_h equals zero for long portfolios of dubious assets, price vector $\boldsymbol{P}_D$ is interpretable as the actual vector of transaction prices for the dubious assets in the founding economy with transaction costs σ_h,. This situation arises, e.g., in economies with short sale-constraints or bid-ask spreads, as in Luttmer [1996].[12]

When some zero probability assessments under the reference and the marginal investor's belief differ, Proposition 1 gives rise to an economically different foundation of S–SDFs, which relies on a frictionless economy with ambiguity. In such settings, while arbitrage opportunities perceived under the marginal investor's belief are naturally eliminated, some arbitrage opportunities perceived under the reference belief may not. Denoting by $\tilde{\mathbb{P}}$ the marginal investor's belief, the absence of free lunches under belief $\tilde{\mathbb{P}}$ is equivalent to the existence of a strictly positive linear pricing functional exactly pricing all traded assets, which is represented by a strictly positive SDF $\tilde{M}$ such that: $\boldsymbol{P} = \tilde{\mathbb{E}}[\tilde{M}\boldsymbol{X}]$; see again, Harrison and Kreps [1979] and Clark [1993], among others. However, this pricing functional does not admit an analogous representation with a strictly

[11] σ_h is the support function of set $C := \{\boldsymbol{\eta} \in \mathbb{R}^{N_D} : h(\boldsymbol{\eta}) \leq \tau\}$, which is non-empty, closed and convex by the properties of h. By [Hiriart-Urruty and Lemaréchal, 2012, Thm. 2.2.2], support functions and closed convex sets stay in a one-to-one relation. Section F of the Online Appendix proves a version of Proposition 1 under the generally stronger no-arbitrage condition of market viability, in which a one-to-one relation is obtained between pricing error metric h and a cost function $\sigma = h^*$, the convex conjugate of h. Here, any closed and convex, not necessarily sublinear, cost function h^* corresponds in a one-to-one way to and S–SDF with pricing error metric h.

[12] When σ_h is not zero when evaluated at the unit vectors, $\boldsymbol{P}_D$ is interpretable as a vector of quoted prices for the dubious assets, which does not incorporate part of the transaction costs for trading these assets in the founding economy with frictions.

positive SDF under reference belief $\mathbb{P}$ when this belief has not identical zero probability assessments as $\tilde{\mathbb{P}}$.

In the above frictionless settings with ambiguity, Proposition 1 allows us to tackle economies where an SDF exactly pricing all assets does not exist, but an S–SDF may exist. This is achieved by means of clearly interpretable existence conditions, which arise from following equivalent expression for pricing rule (5) in Proposition 1:

$$\pi(Z) = \inf_{\boldsymbol{\theta}\in\mathbb{R}^N} \sup_{\boldsymbol{\eta}\in\mathbb{R}^{N_D}} \{\boldsymbol{P}_S'\boldsymbol{\theta}_S + (\boldsymbol{P}_D + \boldsymbol{\eta})'\boldsymbol{\theta}_D \ : \ h(\boldsymbol{\eta}) \leq \tau \text{ and } Z = \boldsymbol{X}'\boldsymbol{\theta}\} \ . \tag{7}$$

In essence, $\pi(Z)$ is interpretable as the min-max replication cost of a payoff Z, after considering all ambiguous valuations of dubious payoffs in a frictionless market, which result from admissible price adjustments $\boldsymbol{\eta}$ relative to the marginal investor's valuation $\boldsymbol{P}_D = \tilde{\mathbb{E}}[\tilde{M}\boldsymbol{X}_D]$.[13] Proposition 1 characterizes existence of a strictly positive S–SDF satisfying pricing constraints (1), by means of a no-arbitrage condition based on pricing functional (7). This no-arbitrage condition is verifiable for any specification of the reference belief and is weaker than the requirement of no-arbitrage under exact pricing of all assets, because payoff prices under pricing rule (7) are by definition an upper bound for the actual prices in the underlying frictionless economy. In this sense, the pricing predictions derived while assuming existence of an S–SDF in ambiguous frictionless asset pricing settings are by construction less fragile than those derived when assuming existence of an SDF that exactly prices all assets in unambiguous frictionless settings.

In summary, we have shown that the asset pricing predictions of S–SDFs are founded either by unambiguous arbitrage-free markets with frictions or by frictionless arbitrage-free markets with ambiguity. Which of these foundations is the appropriate one, depends on the concrete asset pricing problem under scrutiny.

1.3 Minimum Dispersion S–SDFs

Given the economic foundation just discussed, an analysis and selection framework for S–SDFs is required. For this reason, we study optimal S–SDFs satisfying pricing error constraints (1) and minimizing various established notions of stochastic dispersion. We measure the S–SDF dispersion via $\Phi-$dispersions of the form $M \mapsto \mathbb{E}[\phi(M)]$, where $\phi : \mathbb{R} \to [0, +\infty]$ is a strictly convex function.[14] Many well-known measures of SDF dispersion are $\Phi-$dispersions, such as the variance, entropy-based dispersions and more generally

[13] Analogously, π is interpretable as a robust pricing functional, which treats the vector of dubious prices $\boldsymbol{P}_D$ as potentially contaminated by pricing errors $\boldsymbol{\eta}$ such that $h(\boldsymbol{\eta}) \leq \tau$. Accordingly, market $\mathcal{Z}$ in definition (4) is interpretable as the set of payoffs that are robustly replicable using portfolios of the existing assets, at a cost that is (uniformly) bounded across all asset valuations compatible with such pricing errors.

[14] For technical reasons, we always require in the sequel that ϕ is finite on the interval $(0, +\infty)$.

dispersions in the Cressie and Read [1984] family.[15]

Definition 2 (Minimum dispersion S–SDFs). S–SDFs that satisfy pricing constraints (1) and minimize a particular $\Phi-$dispersion are called minimum dispersion S–SDFs. They are defined by the solution of following minimum S–SDF dispersion problem:

$$\Pi(\tau) := \inf_{M \in \mathcal{M}} \{\mathbb{E}[\phi(M)] \;:\; h(\mathbb{E}[M\boldsymbol{X}_D] - \boldsymbol{P}_D) \leq \tau\} \;, \tag{8}$$

where $\mathcal{M}$ denotes the set of nonnegative SDFs that exactly price the sure assets:

$$\mathcal{M} := \{M \text{ with finite } q-th \text{ moment } \;:\; M \geq 0 \text{ and } \mathbb{E}[M\boldsymbol{X}_S] - \boldsymbol{P}_S = \boldsymbol{0}\} \;.$$

Minimum dispersion S–SDF problems in Definition 2 give rise to an extended family of minimum SDF dispersion bounds, relative to the bounds in Hansen and Jagannathan [1991], Luttmer [1996] and Almeida and Garcia [2016], which can account for general pricing error features parametrized by pricing error function h and by bound τ. Figure 1 illustrates the properties of such generalized bounds. Let M_0^S be the unique S–SDF solving problem $\Pi(\infty) := \lim_{\tau\to+\infty} \Pi(\tau)$, i.e., M_0^S is only required to price the sure assets exactly, and define:

$$\tau^{max} := h(\mathbb{E}[M_0^S\boldsymbol{X}_D] - \boldsymbol{P}_D) \;. \tag{9}$$

By construction, τ^{max} is the smallest pricing error threshold on the dubious assets, for which the pricing error constraint on the dubious assets is satisfied by a SDF only required to price exactly the sure assets. The resulting generalized dispersion bound $\Pi(\tau)$ is strictly convex and strictly decreasing for any $\tau \leq \tau^{max}$, while it is constant at $\Pi(\tau) = E[\phi(M_0^S)]$ if $\tau \geq \tau^{max}$.[16]

Working directly with problem (8) is inconvenient, as it is an infinite-dimensional optimization problem. Therefore, we provide the dual characterization of minimum dispersion S–SDFs, in terms of the solutions of penalized portfolio problems of the form:

$$\Delta(\tau) := \min_{\boldsymbol{\theta}\in\mathbb{R}^N} \left\{\mathbb{E}[\phi_+^*(-\boldsymbol{X}'\boldsymbol{\theta})] + \boldsymbol{P}'\boldsymbol{\theta} + \sigma_h(\boldsymbol{\theta}_D)\right\} \;, \tag{10}$$

with the penalization function σ_h in equation (6) of Proposition 1, the restriction ϕ_+ of dispersion function ϕ to $[0,\infty)$ and the convex conjugate of ϕ_+, defined for any $y \in \mathbb{R}$ by $\phi_+^*(y) := \sup_{x\in\mathbb{R}}\{yx - \phi_+(x)\}$.[17] In

[15] Online Appendix A collects relevant explicit examples of $\Phi-$dispersions in the Cressie and Read [1984] family.

[16] This follows by [Luenberger, 1997, Prop. 1 and 2, p. 216-217], the strict convexity of $\Phi-$dispersions and the convexity of pricing error function h.

[17] Restriction ϕ_+ is defined by $\phi_+(x) := \phi(x)$ if $x \geq 0$ and $\phi_+(x) := +\infty$ if $x < 0$.

these portfolio problems, dual $\Phi-$dispersion $\mathbb{E}[\phi_+^*(\cdot)]$ measures the dispersion of portfolio payoffs, while σ_h is a penalization term of the form (6) arising only for the portfolio weights of the dubious assets.

Recalling the two interpretations of market $(\mathcal{Z}, \pi)$ in Proposition 1, in terms of a founding economy with either frictions or ambiguity, we obtain two corresponding interpretations for portfolio problem (10). With respect to the economy with frictions, problem (10) is directly interpretable as a portfolio problem in which investors incorporate transaction costs $\sigma_h(\boldsymbol{\theta}_D)$ for implementing portfolios of dubious payoffs. With respect to the founding frictionless economy with ambiguity, following equivalent form of problem (10) helps in providing useful intuition:

$$\Delta(\tau) := \min_{\boldsymbol{\theta}\in\mathbb{R}^N} \sup_{\boldsymbol{\eta}\in\mathbb{R}^{N_D}} \left\{ \mathbb{E}[\phi_+^*(-\boldsymbol{X}'\boldsymbol{\theta})] + \boldsymbol{P}_S'\boldsymbol{\theta}_S + (\boldsymbol{P}_D + \boldsymbol{\eta})'\boldsymbol{\theta}_D : h(\boldsymbol{\eta}) \le \tau \right\} . \tag{11}$$

This problem is a robust min-max problem, in which investors minimize a worst case criterion that incorporates the most costly valuations of portfolios $\boldsymbol{\theta}_D$ of dubious assets, out of the set of admissible ambiguous valuations $(\boldsymbol{P}_D + \boldsymbol{\eta})'\boldsymbol{\theta}_D$ such that $h(\boldsymbol{\eta}) \le \tau$.

The duality between minimum dispersion S–SDFs and optimal portfolios solving penalized problem (10) is established next, under the no-arbitrage assumption that was adopted in Proposition 1.[18]

Proposition 2 (Dual portfolio problem). *Given $\tau > 0$, suppose that the economy $(\mathcal{Z}, \pi)$ in Proposition 1 admits no free lunches for some $\tilde{\tau} < \tau$. It then follows: $\Pi(\tau) = -\Delta(\tau)$. Moreover, denoting by $\boldsymbol{\theta}_0$ the solution of problem (10), the unique solution of minimum S–SDF dispersion problem (8) is:*[19]

$$M_0 = (\phi_+^*)'(-\boldsymbol{X}'\boldsymbol{\theta}_0) . \tag{12}$$

Proposition 2 states that penalized portfolio problems (10) can be used to compute the minimum dispersion bounds (8). Moreover, the minimum dispersion S–SDF M_0 can be easily recovered from link (12), via a simple transformation of the portfolio dispersion function ϕ_+^*.[20] Therefore, Proposition 2 provides a powerful device to compute minimum S–SDF dispersion bounds and minimum dispersion S–SDFs in a variety of practically relevant asset pricing settings.

The duality result in Proposition 2 relies on the same no-arbitrage condition of Proposition 1, which ensures

[18]Proposition 2 holds for all closed-form pricing error metrics considered in this work.

[19]With a slight abuse of notation, we denote by $(\phi_+^*)'(y)$ in equation (12) the first derivative of function ϕ_+^* in an interior point y.

[20]Online Appendix A collects closed-form expressions of ϕ_+, ϕ_+^* and the minimum dispersion S–SDF (12) for the family of Cressie and Read [1984] dispersions.

existence of a strictly positive S—SDF in the first place. Therefore, when duality does not hold, market $(\mathcal{Z}, \pi)$ in Proposition 1 is not arbitrage-free. In such situations, a solution to dual portfolio problem $\Delta(\tau)$ in equation (10) may still exist, but random variable M_0 in equation (12) may not be an admissible minimum dispersion S–SDF. Emergence of such a duality failure can be assessed empirically, by verifying whether M_0 satisfies the pricing error constraints in the primal minimum S–SDF dispersion problem (8). We borrow from this insight in our empirical analysis of Section 3, in order to distinguish empirically asset pricing settings with solid S–SDF foundation from those with weak S–SDF foundation.

2 S–SDF Applications

Minimum dispersion S–SDFs and their dual portfolio problems are directly related to various important asset pricing settings, which correspond to concrete specifications of pricing error metric h and/or penalty function σ_h; see Table 1 for a partial list of closed-form such (h, σ_h) pairs. Borrowing from the arbitrage-free characterizations of Section 1.2, such settings can be founded economically in two ways. First, via an economy with frictions where transaction costs are explicitly reflected in the evaluation of arbitrage opportunities. Second, via a frictionless economy with ambiguity where arbitrage opportunities are evaluated by means of a robust min-max pricing rule. With this second characterization, we obtain novel arbitrage-free S–SDF foundations and corresponding minimum dispersion S–SDFs for a variety of important asset pricing backgrounds that do no explicitly incorporate market frictions. These include recent settings proposing new methods for computing SDFs in large asset markets, based on several regularization techniques from the machine learning literature, and key asset pricing relations derived under APT-type no-arbitrage assumptions.

2.1 S–SDFs with bounded pricing errors

Minimum variance S–SDFs incorporating conic portfolio constraints, such as short selling constrains and constraints arising in markets with bid-ask spreads, are studied in Luttmer [1996]. As detailed in Online Appendix B, they are obtained with Proposition 2 under a penalization σ_h implying zero (infinite) transaction costs for each portfolio satisfying (violating) the constraint. A specific property of these S–SDFs is that pricing errors, while delimited by a convex cone, are unbounded. In the sequel, we address instead key asset pricing settings with bounded pricing errors.

A useful class of transaction costs on portfolio positions can be modelled using norms: $\sigma_h = \tau \|\cdot\|$, where $\|\cdot\|$

is some norm in $\mathbb{R}^{N_D}$ and $\tau > 0$ is a scaling parameter. From Proposition 2, S–SDFs in these markets give rise to pricing errors bounded with respect to the dual norm $\|\cdot\|_*$:[21]

$$\|\mathbb{E}[M\boldsymbol{X}_D] - \boldsymbol{P}_D\|_* \leq \tau \ . \tag{13}$$

For instance, a specification $\sigma_h = \tau \|\cdot\|_1$ of proportional transaction costs with an l_1-norm implies pricing errors uniformly bounded in absolute value by τ.[22] Therefore, S–SDFs implying uniformly bounded pricing errors are founded in a corresponding arbitrage-free economy with proportional transaction costs on portfolio positions. Obvious modifications of these transaction cost settings can incorporate, e.g., asset specific transaction costs and/or portfolio reallocation costs relative to a reference point, such as those in DeMiguel et al. [2020], among others.

For asset pricing settings with no explicit transaction costs, S–SDFs implying norm bounded pricing errors are most naturally founded in frictionless arbitrage-free economies with ambiguity. In such settings, bound (13) constrains the discrepancies between the marginal investors' arbitrage-free valuations of dubious payoffs under her probability belief and the S–SDF valuations under the reference belief. Consistent with robust pricing functional (7), arbitrage opportunities are then assessed with respect to an inflated price system, which penalizes the ambiguity of the payoff valuations under the reference belief. For any traded payoff Z, the robust pricing functional (7) corresponding to norm-based pricing error constraint (13) is given in closed-form by:

$$\pi(Z) := \inf_{\boldsymbol{\theta} \in \mathbb{R}^N} \{\boldsymbol{P}'\boldsymbol{\theta} + \tau \|\boldsymbol{\theta}_D\| : Z = \boldsymbol{X}'\boldsymbol{\theta}\} \ . \tag{14}$$

Note that specification $\sigma_h = \tau \|\cdot\|_1$ with an l_1-norm reproduces the widely used lasso penalty in the machine learning literature; see Tibshirani [1996]. This penalty is known to produce sparse optimal solutions, which in our setting implies sparse optimal portfolio weights in the solution of the dual portfolio problem of Proposition 2. Given the link (12) between optimal portfolios and minimum dispersion S–SDFs, this choice translates to S–SDFs that depend on a strict subset of dubious asset returns. According to Proposition 2, these penalization techniques give rise to robust S–SDFs with pricing errors constrained by closed-form bounds of the form (13). Therefore, they also admit an explicit no-arbitrage foundation within a frictionless

[21] The dual norm $\|\cdot\|_*$ is defined for any $\boldsymbol{\theta}_D \in \mathbb{R}^{N_D}$ by $\|\boldsymbol{\theta}_D\|_* := \max_{\boldsymbol{\eta} \in \mathbb{R}^{N_D}} \{\boldsymbol{\theta}_D'\boldsymbol{\eta} : \|\boldsymbol{\eta}\| \leq 1\}$, i.e., it is equal to the support function of the unit $\|\cdot\|-$ball.

[22] Given an l_p-norm such that $\|\boldsymbol{x}\|_p := (\sum_{i=1}^{N_D} |x_i|^p)^{1/p}$ when $p \in [1, +\infty)$ and $\|\boldsymbol{x}\|_p := \max_{i=1}^{N_D} |x_i|$ when $p = +\infty$, its dual norm is the l_q-norm with $1/p + 1/q = 1$. In particular, the $l_\infty-$norm is the dual norm of the l_1-norm and vice-versa, while the l_2-norm is self-dual.

economy with ambiguity. This foundation is very different from the econometric motivation for using these techniques in empirical asset pricing, which focuses on estimators with desirable finite sample properties when estimating an SDF that exactly prices a large number of assets. Instead, in our approach the minimum dispersion S–SDF solving the primal problem of Proposition 2 is directly the object of interest. In particular, no further assumptions, regarding, e.g., some spanning properties of returns, is needed to found it.[23]

A key insight of our theory is that sparsity cannot be obtained both for pricing errors and optimal portfolio weights of minimum dispersion S–SDFs. This sparsity trade-off arises also for penalization devices implying a more general relation between shrinkage and sparsity. One such example is the Elastic Net penalty proposed in Zou and Hastie [2005] and used in Kozak et al. [2020], among others, to shrink the cross-section of returns. In our framework, a more flexible relation between shrinkage and sparsity of optimal portfolio weights is easily obtained with a norm-based penalization that combines the sparsity properties of the lasso with the shrinkage properties of l_2-penalizations: $\sigma_h := \alpha \left\|\cdot\right\|_1 + \tau \left\|\cdot\right\|_2$. The resulting closed-form pricing error metric $h = \text{dist}_{\alpha B_\infty}$ in Table 1 is the Euclidean distance from the $l_\infty-$norm ball of radius α. Here, maximal pricing errors larger than α are penalized quadratically, which gives rise to a quadratic constraint above a threshold implying non sparse pricing errors. The limit case $\alpha = 0$ corresponds to a standard ridge pricing error metric $h = \left\|\cdot\right\|_2$ with penalization $\sigma_h = \tau \left\|\cdot\right\|_2$, which gives rise to pricing errors and optimal portfolio weights that are neither sparse.[24]

2.2 Minimum dispersion S–SDFs as pricing error minimizers

As formally shown in Proposition 4 of Appendix B, minimum dispersion S–SDFs can be equivalently defined by means of a minimum pricing error problem, subject to a bound $\nu := \nu(\tau) > 0$ on the S–SDF dispersion:

$$\tilde{\Pi}(\nu) := \inf_{M \in \mathcal{M}} \left\{ h(\mathbb{E}[M\boldsymbol{X}_D - \boldsymbol{P}_D]) \,:\, \mathbb{E}[\phi(M)] \leq \nu \right\} \ . \tag{15}$$

This definition establishes a link to additional important directions in the literature, such as the good-deal bounds theory in Cochrane and Saa-Requejo [2000], the robust investor's belief identification approach in

[23] Qiu and Otsu [2018] define with a link of the form (12) a particular S–SDF from the solution of a portfolio problem of the form (10), in which the penalty is given by lasso and the dispersion of portfolio payoffs (function ϕ_+^*) is measured by the convex conjugate of the Kullback-Leibler SDF dispersion (function ϕ_+); see Case 1 of Examples 1 in Online Appendix A for the closed-form expressions of these functions. The authors use this S–SDF to estimate a target SDF that exactly prices all assets, under the assumption that the logarithm of the latter is well approximated by linear combinations of excess returns as the cross-section of assets grows.

[24] Many other pricing error metrics and portfolio weight penalties can be implemented in our framework, covering all sublinear penalties corresponding to any convex and closed pricing error metric. Moreover, Section F of the Online Appendix proves a version of Proposition 1, under the no-arbitrage condition of market viability, in which any closed and convex penalization h^* corresponds in a one-to-one way to a convex and closed pricing error metric h. Hence, convex penalizations are also covered by our theory.

Chen et al. [2020b] and the SDFs for large asset markets proposed in Kozak et al. [2020].[25] Given a vector $\boldsymbol{F}^e$ of N_D characteristics-based factor excess returns, the latter ones are parametric S–SDFs of the form $M(\boldsymbol{\theta}_0) := \theta_{0S} - \boldsymbol{\theta}_{0D}'(\boldsymbol{F}^e - \mathbb{E}[\boldsymbol{F}^e])$, with $\boldsymbol{\theta}_0$ defined for some $\nu > 0$ as the solution of following optimization problem:[26]

$$\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^N} \left\{ \frac{1}{2} \left\| \mathbb{E}[M(\boldsymbol{\theta})\boldsymbol{F}^e] \right\|_2^2 \ : \ \frac{1}{2}\mathbb{E}[M^2(\boldsymbol{\theta})] \le \nu \text{ and } \mathbb{E}[M(\boldsymbol{\theta})] = 1 \right\} .$$

This problem is specified using a single sure asset, with unit payoff and price, and dubious asset payoffs $\boldsymbol{X}_D := \boldsymbol{F}^e$ with price vector $\boldsymbol{P}_D := \boldsymbol{0}$. Hence, Proposition 2 and the equivalence of problems (8) and (15) show that $M(\boldsymbol{\theta}_0)$ is the minimum variance S–SDF induced by the ridge pricing error metric $h = \frac{1}{2}\left\|\cdot\right\|_2^2$. Note that the geometry of pricing errors induced by this metric is identical to the one under the norm-based metric $h = \frac{1}{\sqrt{2}}\left\|\cdot\right\|_2$, i.e., $M(\boldsymbol{\theta}_0)$ is in the class of S–SDF with norm bounded pricing errors discussed in Section 2.1. Therefore, it is founded in a frictionless arbitrage-free economy under ambiguity, based on the closed-form robust pricing functional defined for any traded payoff Z by:

$$\pi(Z) := \inf_{\boldsymbol{\theta}\in\mathbb{R}^N} \left\{ \boldsymbol{P}'\boldsymbol{\theta} + \sqrt{2\tau}\left\|\boldsymbol{\theta}_D\right\|_2 : Z = \boldsymbol{X}'\boldsymbol{\theta} \right\} . \tag{16}$$

As noted at the end of Section 2.1, this S–SDF gives rise to pricing errors and optimal portfolio weights that are neither sparse. More generally, a sparse minimum dispersion S–SDF is obtained using an Elastic Net norm based portfolio penalization $\sigma_h = \alpha\left\|\cdot\right\|_1 + \tau\left\|\cdot\right\|_2$, which gives rise to non sparse pricing errors constrained by the closed-form pricing error metric $h = \mathrm{dist}_{\alpha B_\infty}$; see again Table 1. As elaborated in Section 2.1, such S–SDFs as well are founded in frictionless arbitrage-free economies with ambiguity.[27]

2.3 APT S–SDFs

An essential constraint underlying the APT is a finite maximal Sharpe ratio in markets with no (asymptotic) arbitrage opportunities.[28] Such a constraint gives rise to approximate linear valuation rules when returns

[25]See Online Appendix C for a detailed derivation of the relation between our S–SDF methodology, good deal bounds theories and robust belief identification approaches.

[26]Such S–SDFs are motivated by a Bayesian prior on expected returns that shrinks the Sharpe ratio of low eigenvalue principal component portfolios relative to the Sharpe ratio of high eigenvalue principal component portfolios.

[27]The corresponding closed-form robust pricing functional is defined for any traded payoff Z by:

$$\pi(Z) := \inf_{\boldsymbol{\theta}\in\mathbb{R}^N} \left\{ \boldsymbol{P}'\boldsymbol{\theta} + \alpha\left\|\boldsymbol{\theta}_D\right\|_1 + \tau\left\|\boldsymbol{\theta}_D\right\|_2 : Z = \boldsymbol{X}'\boldsymbol{\theta} \right\} .$$

Note that since penalization σ_h in Proposition 1 is sublinear, definitions of sparse SDFs based on not sublinear portfolio weight penalizations are not embedded in Proposition 1. One such example is the Elastic Net penalization $\sigma = \alpha_1\left\|\cdot\right\|_1 + \frac{\alpha_2}{2}\left\|\cdot\right\|_2^2$ in Kozak et al. [2020]. However, Section F of the Online Appendix provides the arbitrage-free foundation also for all closed and convex penalizations under the generally stronger no-arbitrage condition of market viability.

[28]See, e.g., [Chamberlain and Rothschild, 1983, Corollary 1] for a formal statement.

satisfy a suitable factor structure. Consider a general factor model representation for a $N_D \times 1$ vector of excess returns $\boldsymbol{R}^e_{N_D}$:

$$\boldsymbol{R}^e_{N_D} = \boldsymbol{\Lambda}_{N_D} \boldsymbol{F}^e + \boldsymbol{\zeta}_{N_D} \ , \tag{17}$$

where $\boldsymbol{F}^e$ is a $(N_S - 1) \times 1$ vector of observed traded excess factor returns and $\boldsymbol{\Lambda}_{N_D}$ a $N_D \times (N_S - 1)$ matrix of factor loadings. The $N_D \times 1$ vector of residuals $\boldsymbol{\zeta}_{N_D}$ is orthogonal to traded factor risk, but potentially cross-sectionally correlated, with variance covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}_{N_D}} = \boldsymbol{B}_{N_D} \boldsymbol{B}'_{N_D} + \boldsymbol{Q}_{N_D}$, where $\{\boldsymbol{B}_{N_D}\}$ is a sequence of $N_D \times N_K$ matrices and $\{\boldsymbol{Q}_{N_D}\}$ a sequence of symmetric positive definite $N_D \times N_D$ matrices with uniformly bounded eigenvalues.[29] In this setting, it follows from [Chamberlain and Rothschild, 1983, Corollary 2] that the absence of asymptotic arbitrage opportunities, under a sequence of factor models (17) with growing dimension N_D, yields existence of a constant $\tau \geq 0$ and a $N_K \times 1$ vector $\boldsymbol{\lambda}$ of risk premia for unobservable factor risks such that:[30]

$$\|\boldsymbol{\eta}_{N_D}\|_{2,\boldsymbol{\Sigma}^{-1/2}_{\boldsymbol{\zeta}_{N_D}}} := \sqrt{\boldsymbol{\eta}'_{N_D} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\zeta}_{N_D}} \boldsymbol{\eta}_{N_D}} \leq \tau \ , \tag{18}$$

where

$$\boldsymbol{\eta}_{N_D} := \mathbb{E}[\boldsymbol{\zeta}_{N_D}] - \boldsymbol{B}_{N_D} \boldsymbol{\lambda} \ . \tag{19}$$

In equation (19), $\boldsymbol{\eta}_{N_D}$ has the interpretation of an expected excess return component not explained by priced exposures to unobservable factor risk in model (17). Moreover, since factors $\boldsymbol{F}^e$ are themselves traded excess returns, $\mathbb{E}[\boldsymbol{F}_e]$ is the vector of traded factor risk premia and $\boldsymbol{F}_e$ has by construction a price of zero.

The next corollary to Proposition 1 provides the economic foundation for APT S–SDFs, which are S–SDFs reproducing the observable APT asset pricing predictions (18)-(19).[31]

Corollary 1 (APT S–SDF). *Given vector $\boldsymbol{F}^e$ of traded factor excess returns, let sure and dubious asset payoffs be given by $\boldsymbol{X}_S := (1, \boldsymbol{F}^{e\prime})'$ and $\boldsymbol{X}_D := \boldsymbol{R}^e$, with corresponding sure and dubious price vectors $\boldsymbol{P}_S := (1, \boldsymbol{0}')'$ and $\boldsymbol{P}_D := \boldsymbol{0}$. Consider a market structure $(\mathcal{Z}, \pi)$ as in equations (4) and (5), with penalization σ_h defined by:*

$$\sigma_h(\boldsymbol{\theta}_D) = \tau \, \|\boldsymbol{\theta}_D\|_{2,\boldsymbol{\Sigma}^{1/2}_{\boldsymbol{\zeta}}} \ . \tag{20}$$

[29] Sufficient conditions for existence and uniqueness of such a factor representation are provided in [Chamberlain and Rothschild, 1983, Theorem 4].

[30] Constant τ in inequality (18) is explicitly given as $\delta^2 \lambda_{k+1}/\lambda_0$, where $\lambda_{k+1} < \infty$ $(\lambda_0 > 0)$ is a uniform upper (lower) bound on the $k+1$ largest (the smallest) eigenvalue of $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}_N}$ and δ^2 is the squared maximal Sharpe ratio in the underlying arbitrage-free economy.

[31] Given that bound (18) is independent of N_D, we consider in the sequel an excess return vector $\boldsymbol{R}^e$ of fix dimension and drop the N_D subscripts.

Such market has no free lunches if and only if there exists a strictly positive APT S–SDF, M_{APT}, with normalized expectation $\mathbb{E}[M_{APT}] = 1$, such that:

$$\mathbb{E}[M_{APT}\boldsymbol{F}^e] = \mathbf{0} \text{ and } \|\mathbb{E}[M_{APT}\boldsymbol{R}^e]\|_{2,\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}^{-1/2}} \leq \tau \ . \tag{21}$$

Corollary 1 provides a novel arbitrage-free foundation to APT asset pricing relations, within a frictionless economy with ambiguity, in which pricing errors on excess returns orthogonal to traded factor risk are bounded under standardized pricing error metric $h = \|\cdot\|_{2,\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}^{-1/2}}$. This foundation is in itself independent of APT assumptions regarding, e.g., asymptotic no-arbitrage conditions or a linear factor representation for excess returns. Moreover, using the APT S–SDF in Corollary 1 a distinct characterization of asset excess returns easily follows. Indeed, traded factor risk obeys standard characterization (2), in which a factor risk premium is completely explained by the covariance of traded factor excess returns with the APT S–SDF. In contrast, excess returns orthogonal to traded factor risk satisfy the more general characterization (3):

$$\mathbb{E}[\boldsymbol{R}^e - \boldsymbol{\Lambda}\boldsymbol{F}^e] = -Cov(M_{APT}, \boldsymbol{R}^e - \boldsymbol{\Lambda}\boldsymbol{F}^e) + \mathbb{E}[M_{APT}\boldsymbol{R}^e] \ . \tag{22}$$

These excess returns are the sum of a systematic risk premium component, generated by their covariance with S–SDF risk, and a pure mispricing component $\mathbb{E}[M_{APT}\boldsymbol{R}^e]$, which satisfies by definition the APT pricing error bound (18). In this sense, decomposition (22) is a model-free reproduction of the APT asset pricing predictions based on APT S–SDFs founded by Corollary 1.

Importantly, Corollary 1 provides the weakest no-arbitrage condition for the existence of at least one APT S–SDF satisfying pricing constraints (21). However, several other APT–SDFs satisfying such constraints may exist. For instance, any strictly positive random variable M is another such APT S–SDF if under a norm $\|\cdot\|$ stronger than the l_2–norm ($\|\cdot\| \geq \|\cdot\|_2$) it satisfies the tighter pricing constraints:[32]

$$\mathbb{E}[M\boldsymbol{F}^e] = \mathbf{0} \text{ and } \|\mathbb{E}[M\boldsymbol{R}^e]\|_{\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}^{-1/2}} \leq \tau \ . \tag{23}$$

Consistent with Proposition 1, the arbitrage-free foundation of APT S–SDFs satisfying tighter pricing constraint (23) directly succeeds with a closed-form penalization $\sigma_h = \tau \|\cdot\|_{*,\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}^{1/2}}$. However, since this penalization is bounded from above by penalization (20), such foundation has to rely on a more stringent no-arbitrage condition than the one underlying Corollary 1.

[32]Closed-form examples of such norms are convex combinations of l_1– and l_2–norms or of (scaled) l_∞– and l_2–norms, which are applied in our empirical analysis of Section 3 to investigate APT S–SDFs with varying sparsity properties.

A key advantage of APT S–SDFs is that they are naturally suited for an analysis and selection framework based on minimum dispersion S–SDFs that reproduce the APT pricing predictions. Borrowing from Proposition 2, this framework is made explicit by the next proposition, which provides the premise for our empirical analysis of APT S–SDFs in Section 3.

Proposition 3. *Given the specification of sure and dubious assets in Corollary 1, consider in Definition 2 a minimum dispersion APT S–SDF problem with pricing error metric $h := \|\cdot\|_{\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}^{-1/2}}$ induced by a norm $\|\cdot\| \geq \|\cdot\|_2$. For any $\tau > 0$, the corresponding closed-form penalization from Proposition 1 is $\sigma_h = \tau \|\cdot\|_{*,\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}^{1/2}}$ and dual portfolio problem (10) is explicitly given by:*

$$\Delta_{\|\cdot\|}(\tau) := \min_{\boldsymbol{\theta}\in\mathbb{R}^N} \left\{ \mathbb{E}[\phi_+^*(-\boldsymbol{X}'\boldsymbol{\theta})] + \boldsymbol{P}'\boldsymbol{\theta} + \tau \|\boldsymbol{\theta}_D\|_{*,\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}^{1/2}} \right\} . \tag{24}$$

Moreover, if a strictly positive S–SDF exists satisfying the pricing error constraints (23) with a strict inequality, then $\Pi(\tau) = -\Delta_{\|\cdot\|}(\tau)$ and the unique solution of minimum S–SDF dispersion problem $\Pi(\tau)$ is $M_0 = (\phi_+^*)'(-\boldsymbol{X}'\boldsymbol{\theta}_0)$, where $\boldsymbol{\theta}_0$ is the solution of problem $\Delta_{\|\cdot\|}(\tau)$.

Minimum dispersion APT S–SDFs are computable as closed-form transformations of an optimal portfolio payoff from closed-form penalized portfolio problem $\Delta_{\|\cdot\|}(\tau)$.[33] Moreover, consistent with the discussion at the end of Section 1.3, non existence of a minimum dispersion APT S–SDF indicates a violation of the no-arbitrage condition in the founding economy of Proposition 3.

Minimum dispersion APT S–SDFs are natural benchmarks for evaluating empirical asset pricing models motivated by the APT and for constructing model-free S–SDFs with desirable out-of-sample properties. In this context, different APT consistent norms $\|\cdot\| \geq \|\cdot\|_2$ give rise to varying S–SDF properties with respect to, e.g., pricing errors and S–SDFs sparsity.

3 Empirics of APT S–SDFs

Minimum dispersion APT S–SDFs give rise to a new convenient approach for constructing model-free S–SDFs with desirable time series and cross-sectional properties. Our approach exploits two key properties of these S–SDFs. First, they optimally capture the relation between cross-sectional explanatory power and minimum S–SDF dispersion. Second, the fraction of expected returns explained by a covariation with S–SDF risk is proportional to the S–SDF volatility. Together, these two features generate an optimal tradeoff between

[33] In addition, they can be consistently estimated in real data applications, using the methodology detailed in Section 3.1.

cross-sectional and time series S–SDF explanatory power. We exploit this tradeoff to identify model-free single-factor empirical asset pricing settings with desirable properties along these two dimensions. These settings differ with respect to key building blocks of minimum dispersion APT S–SDFs, such as (i) the degree of sparsity in pricing errors or S–SDFs and (ii) the choice of the traded sure payoffs replicating relevant systematic risk exposures.

3.1 Empirical setting

In order to study APT S–SDFs with varying sparsity properties, we consider for $\lambda \in [0,1]$ the following APT pricing error norms, in addition to the benchmark APT norm (18):[34]

$$h_{1,\boldsymbol{\Sigma}_{\zeta}^{-1/2}}(\boldsymbol{\eta}_D) := (1-\lambda)\left\|\boldsymbol{\eta}_D\right\|_{1,\boldsymbol{\Sigma}_{\zeta}^{-1/2}} + \lambda\left\|\boldsymbol{\eta}_D\right\|_{2,\boldsymbol{\Sigma}_{\zeta}^{-1/2}} \ , \tag{25}$$

and

$$h_{\infty,\boldsymbol{\Sigma}_{\zeta}^{-1/2}}(\boldsymbol{\eta}_D) := (1-\lambda)\sqrt{N_D}\left\|\boldsymbol{\eta}_D\right\|_{\infty,\boldsymbol{\Sigma}_{\zeta}^{-1/2}} + \lambda\left\|\boldsymbol{\eta}_D\right\|_{2,\boldsymbol{\Sigma}_{\zeta}^{-1/2}} \ . \tag{26}$$

While norm h_1 induces varying degrees of sparsity on the resulting pricing errors, h_∞ implies sparsity on the portfolio weights supporting minimum dispersion S–SDFs. These features follow from the presence of the l_1-norm in pricing error norm h_1 and in the dual penalization $\sigma_{h_\infty} = \tau h_{\infty*,\boldsymbol{\Sigma}_{\zeta}^{1/2}}$ from Proposition 3, respectively.[35]

We make use of standard databases of different dimensions, in order to benchmark our results with related empirical evidence in the literature. We use three datasets of sorted portfolio returns that differ in the dimension of the set of assets considered, in order to better isolate the empirical merits of APT S–SDFs for return spaces of different dimensions. Our datasets consist of monthly gross returns of sorted portfolios from Kenneth French's data library.

1. **Low dimensional dataset:** The returns consist of the risk-free return, 25 portfolio returns sorted on size and book to market, 10 portfolio returns sorted on momentum, and 25 portfolio returns sorted on size and long term reversal.

2. **Intermediate dimensional dataset:** The returns consist of the risk-free return, 100 portfolio returns sorted on size and book to market, 25 portfolio returns sorted on momentum, 25 portfolio returns

[34] Note that all these norms are stronger than the benchmark APT norm (18), because $\|\cdot\|_2 \le \|\cdot\|_1$ and $\|\cdot\|_2 \le \sqrt{N_D}\,\|\cdot\|_\infty$.

[35] See again Table 1 and Lemma 1 in Appendix B, which reports the closed-form expressions for penalization σ_{h_i} $(i = 1, \infty)$ from Proposition 3.

sorted on size and long term reversal, 25 portfolio returns sorted on size and short term reversal, and 49 portfolio returns sorted on industry.

3. **High dimensional dataset:** For this dataset, the returns in the intermediate dimensional dataset are augmented with characteristics-based factors from the WRDS financial ratios (WFR) dataset, which consists of 69 ratios for 10 industries based on Fama-French industry classification. The construction of the factor returns follows Kozak et al. [2020].

We also consider the returns of various Fama-French factors for constructing simple benchmark linear SDFs and for incorporating well-known systematic risk exposures as sure assets in our APT S–SDFs.

Our sample starts in January 1931 (January 1970 for the characteristic-based factors in the high dimensional dataset) and ends in June 2018. Portfolios with missing time series observations are removed. After such removal, the low dimensional dataset consists of 57 assets and 1054 time series observations, the intermediate dimensional dataset consists of 188 assets and 1054 time series observations, while the high dimensional dataset has 260 asset returns with 561 time series observations.

Consistent estimation procedures of S–SDF dispersion bounds and dual portfolio weights in stationary dynamic economies can be naturally developed within our approach. To this end, let $\{(\boldsymbol{X}_{t+1}, \boldsymbol{P}_t)\}_{t\in\mathbb{N}}$ be a time series of payoffs and quoted prices of the N basis assets, which are defined on the filtered probability space $(\Omega, \{\mathcal{F}_t\}_{t\in\mathbb{N}}, \mathbb{P})$. The N–dimensional vectors of payoffs $\boldsymbol{X}_{t+1}$ and prices $\boldsymbol{P}_t$ are observed at time $t+1$ and time t, respectively. By Proposition 2, the minimum S–SDF dispersion bound is given by $\Delta(\tau)$ and the optimal portfolio weight vector $\boldsymbol{\theta}_0$ is the minimizer of the corresponding dual portfolio problem.

Given a sample of size $T > 0$, the estimator for population value $\Delta(\tau)$ is thus:

$$\Delta_T(\tau) := \min \left\{Q_T(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathbb{R}^N\right\} , \tag{27}$$

with the empirical objective function

$$Q_T(\boldsymbol{\theta}) := \frac{1}{T}\sum_{t=1}^{T}[\phi_+^*(-\boldsymbol{X}_{t+1}'\boldsymbol{\theta}) + \boldsymbol{P}_t'\boldsymbol{\theta}] + \sigma_h(\boldsymbol{\theta}_D) \; .$$

Accordingly, $\boldsymbol{\theta}_T \in \arg\min\{Q_T(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathbb{R}^N\}$ is the corresponding estimator of $\boldsymbol{\theta}_0$. We apply these estimators to compute minimum dispersion APT S–SDF from Proposition 3 for pricing error norms of the form (25)-(26). Under the standard assumptions detailed in Section D of the Online Appendix, $\Delta_T(\tau)$ and $\boldsymbol{\theta}_T$ are consistent estimators of $\Delta(\tau)$ and $\boldsymbol{\theta}_0$, respectively.[36]

3.2 In–sample analysis

In the in-sample analysis, we characterize in detail the properties of the attainable tradeoffs between time series and cross-sectional explanatory power generated by APT S–SDFs under varying key assumptions on the latter.

3.2.1 Existence of minimum dispersion APT S–SDFs

A well-defined tradeoff between spanning features for systematic risks and pricing accuracy requires in the first place existence of corresponding S–SDFs in arbitrage-free asset markets. Recall from Proposition 2 and 3 that existence of a strictly positive S–SDF is a sufficient condition for strong duality to hold. Consequently, a duality failure is an indication of a weak economic foundation of the corresponding S–SDF setting.

Empirically, we find that a duality failure can arise quite often, especially with growing cross-sectional dimensions. In Figure 2, we document this fact for the high dimensional dataset, where we solve for various pricing error thresholds τ the empirical dual problem (27) corresponding to population problem (24), using variance as a measure of SDF dispersion and the APT pricing error metric (18).[37] We then compute with Proposition 3 an empirical candidate S–SDF solution $\hat{M}_{0t+1} = \max\{-\boldsymbol{X}'_{t+1}\hat{\boldsymbol{\theta}}_0, 0\}$, where $\hat{\boldsymbol{\theta}}_0$ is the solution of the empirical portfolio problem (27) corresponding to population problem (24) and $\boldsymbol{X}_{t+1}$ denotes the vector of sure and dubious excess returns at time $t+1$. Further, we compute the empirical expected pricing error vector $\hat{\mathbb{E}}[\hat{M}_{0t+1}\boldsymbol{X}_{t+1}]$, where expectation $\hat{\mathbb{E}}[\cdot]$ is taken under the empirical distribution of excess returns, and verify whether the pricing error constraint $h\left(\hat{\mathbb{E}}[\hat{M}_{0t+1}\boldsymbol{X}_{Dt+1}]\right) \leq \tau$ in the primal empirical S–SDF problem from Proposition 2 is satisfied. We find that for a range of τ thresholds smaller than discontinuity point $\hat{\tau}^{min}$ in Figure 2 an empirical duality failure emerges, i.e., no strictly positive empirical APT S–SDF exists. In contrast, such S–SDFs exist for pricing error thresholds not smaller than $\hat{\tau}^{min}$.

By recognizing the possibility of non-existence of a strictly positive empirical APT S–SDF, we can discipline economically with Proposition 2 and 3 the choice of penalization parameter τ in our minimum dispersion

[36] Such assumptions are satisfied, e.g., for general minimum dispersion S–SDF settings based on (i) Cressie and Read [1984] measures of dispersion in Section A of the Online Appendix and (ii) lower semi-continuous penalization functions σ_h, where a convex and proper function is lower semi-continuous if and only if it is closed. An asymptotic inference framework for minimum dispersion S–SDFs and S–SDF bounds can also be developed, for which a study of the asymptotic distribution of estimators $\Delta_T(\tau)$ and $\boldsymbol{\theta}_T$ is needed. Such a distribution depends on the smoothness properties of penalization function σ_h. Korsaye et al. [2019] make use of Moreau [1962] envelopes theory to develop a unified asymptotic inference approach for minimum dispersion S–SDFs and S–SDF bounds under general penalizations σ_h.

[37] Analogous evidence for the intermediate dimensional dataset is collected in Figure 1 of Section H in the Online Appendix.

APT S–SDFs. Therefore, in our empirical analysis admissible penalization parameters are always constrained to be not smaller than the lower bound $\hat{\tau}^{min}$. Moreover, recall that as pricing error threshold τ increases, the pricing error constraint on minimum dispersion S–SDFs becomes progressively more relaxed, to the point that a minimum dispersion S–SDF only required to exactly price the sure assets can satisfy the constraint for a sufficiently large pricing error threshold $\hat{\tau}^{max}$; see again Figure 1 and equation (9). Therefore, in our analysis we can ensure a well-defined empirical tradeoff between S–SDF spanning features for systematic risks and S–SDF pricing accuracy simply by constraining the admissible penalization parameters in the interval $[\hat{\tau}^{min}, \hat{\tau}^{max}]$.

3.2.2 Tradeoff between cross-sectional and time series S–SDF explanatory power

We next empirically explore the tradeoffs between pricing accuracy and time series explanatory power of minimum variance APT S–SDFs.[38] Figure 3 quantifies such tradeoffs in the low dimensional dataset, based on different pricing error norms (25) and (26) such that $\lambda = 0, 0.5, 1$. Here, we take the market excess return as the single sure asset payoff.[39] In this case, we obtain APT S–SDFs that are interpretable as perturbations of an empirical SDF under the CAPM, which allow for bounded pricing errors on payoffs orthogonal to market risk. Note that when $\lambda = 1$ norm $h_1 = \|\cdot\|_1$ ($h_\infty = \sqrt{N_D}\,\|\cdot\|_\infty$) implies extremely sparse pricing errors (S–SDFs), while for $\lambda = 0$ no sparsity in pricing errors and S–SDFs arises since $h_1 = h_\infty = \|\cdot\|_2$. Parameter value $\lambda = 0.5$ implies instead a less extreme degree of sparsity in pricing errors (norm h_1) or S–SDFs (norm h_∞). Hence, these choices cover a good spectrum of economically relevant perturbations of an empirical SDF under the CAPM.

Panel (A) of Figure 3 summarizes the tradeoff between minimum S–SDF volatility and maximal pricing error size, in an extended version of the minimum S–SDF volatility bounds in Hansen and Jagannathan [1991] and Luttmer [1996]. Since different pricing error norms give rise to a different maximal pricing error threshold $\hat{\tau}^{max}$ for the S–SDF only required to price exactly market excess returns, we collect minimum S–SDF volatilities parametrized by relative pricing error thresholds $\sigma := \tau/\hat{\tau}^{max} \in [\hat{\tau}^{min}/\hat{\tau}^{max}, 1]$.

In the right plot of Panel (B) in Figure 3, the minimum S–SDF volatilities induced by pricing error constraints under the APT l_2−norm (l_∞−norm) correspond to the highest (lowest) S–SDF cross-sectional explanatory

[38] We concentrate without loss of generality on minimum variance APT S–SDFs, because as we explain below they produce by construction the best tradeoff between S–SDF pricing accuracy and time series explanatory power. Section G of the Online Appendix reports the evidence for minimum dispersion APT S–SDFs based on notions of dispersion different from variance.

[39] Section 3.2.4 below addresses in detail the implication of the choice of the sure assets for the resulting S–SDF tradeoff between pricing accuracy and explanatory power in the time series.

power, when the explanatory power is measured by the standard cross-sectional GLS R^2 metric:[40]

$$R^2_{GLS} := 1 - \frac{\left\|\mathbb{E}[M_0(\boldsymbol{X} - \bar{\boldsymbol{X}}\mathbf{1})]\right\|^2_{2,\boldsymbol{\Sigma}^{-1/2}}}{\left\|\mathbb{E}[\boldsymbol{X} - \bar{\boldsymbol{X}}\mathbf{1}]\right\|^2_{2,\boldsymbol{\Sigma}^{-1/2}}} , \tag{28}$$

where $\mathbf{1}$ is a $N \times 1$ vector of ones, $\bar{\boldsymbol{X}} := (\mathbf{1}'\boldsymbol{X})/N$ the average of the components of vector $\boldsymbol{X}$ and $\boldsymbol{\Sigma}$ the $N \times N$ covariance matrix of excess return vector $\boldsymbol{X}$. This correspondence offers a second useful interpretation of penalization parameter τ in Proposition 3, which can be understood as a parametrization of the relative loss $1 - R^2_{GLS}$ in cross-sectional explanatory power that an APT S–SDF suffers with respect to the minimum dispersion SDF exactly pricing all assets. Because of the scale independence of R^2_{GLS}, this interpretation simplifies the comparison of minimum dispersion APT S–SDFs induced by different pricing error norms.

Given a set of sure systematic factor excess returns, the minimum S–SDF volatilities induced by APT pricing error constraints are also directly related to the S–SDF ability to co-move with systematic excess return risks. Indeed, since dubious expected excess returns are given and satisfy decomposition (3), lower S–SDF pricing errors are linked to both larger absolute S–SDF covariances with dubious excess returns and larger S–SDF volatilities. This constraint directly impacts the time series S–SDF co-movement with any excess return X, as measured by a standard time series OLS R^2:

$$R^2_{OLS} := \frac{(Cov(M_0, X))^2}{Var(M_0)Var(X)} = \frac{1}{Var(M_0)} \cdot \frac{(\mathbb{E}[X] - \mathbb{E}[M_0 X])^2}{Var(X)} . \tag{29}$$

Given a cross-section of asset (squared) Sharpe ratios, identity (29) constrains the S–SDF average time series explanatory power in the left plot of Panel (B) from Figure 3, as a function of the varying relative pricing error threshold σ. As intuitively expected, a higher cross-sectional explanatory power in the bottom right plot corresponds to a lower time series explanatory power in the bottom left plot. However, what matters for an economic S–SDF comparison is the generated relative tradeoff between time series and cross-sectional explanatory powers. Panel (B) of Figure 3 shows that minimum variance S–SDFs under the APT l_2-pricing error norm always imply the most favourable tradeoff in-sample. For instance, while under this norm a cross-sectional GLS R^2 of 50% can be obtained together with an average time series R^2 of about 28%, under an APT l_1- and $l_\infty-$norm the resulting average time series R^2 is only about 20% and 14%, respectively. This evidence naturally follows from the fact that the l_2-pricing error bound is the only one producing for any given target value of cross-sectional GLS R^2 metric (28) the corresponding minimum variance S–SDF. Since

[40] By definition, R^2_{GLS} equals one minus the ratio of the cross-sectional variance of pricing errors and the cross-sectional variance of expected returns, after a standardization by the return volatility matrix $\boldsymbol{\Sigma}^{1/2}$.

minimizing the S–SDF variance is equivalent to maximizing the time-series R^2 metric (29), the minimum variance S–SDF with l_2-pricing error bound has to be optimal with respect to the standard GLS metric (29) of cross-sectional explanatory power.

3.2.3 Metrics of pricing accuracy and robustness of S–SDF tradeoff

Two further important aspects for understanding the S–SDF tradeoff between pricing accuracy and time series explanatory power cover: (i) the role of the metric of pricing accuracy and (ii) the robustness of the tradeoff in presence of deviations from the in-sample assumptions.[41] Regarding the former aspect, recall that in general APT S–SDFs satisfy APT pricing error constraint (23), under a norm $||\cdot||$ that may be different from the l_2-norm. Therefore, the natural equivalent of GLS R^2 metric (28) for measuring the S–SDF cross-sectional explanatory power is:

$$R^2_{GLS,\|\cdot\|} := 1 - \frac{\left\|\mathbb{E}[M_0(\boldsymbol{X} - \bar{\boldsymbol{X}}\mathbf{1})]\right\|^2_{\boldsymbol{\Sigma}^{-1/2}}}{\left\|\mathbb{E}[\boldsymbol{X} - \bar{\boldsymbol{X}}\mathbf{1}]\right\|^2_{\boldsymbol{\Sigma}^{-1/2}}} \ . \tag{30}$$

A natural question is how the tradeoff in Section 3.2.2 between cross-sectional and time series explanatory power is impacted by the choice of the metric of pricing accuracy. For each minimum variance APT S–SDF in Panel (B) of Figure 3, Figure 4 plots this tradeoff using cross-sectional GLS R^2 metric (30) with norms $\|\cdot\|_1$, $\|\cdot\|_2$, and $\sqrt{N_D}\,\|\cdot\|_\infty$. The three left plots show that as expected minimum variance APT S–SDFs produce an optimal tradeoff when metric (30) is defined by the same norm as the one in APT pricing error bound (23). At the same time, it appears that the tradeoff induced by APT S–SDFs with l_2-pricing error bound is less sensitive to the choice of the GLS R^2 metric than the tradeoffs induced by l_1- and (scaled) $l_\infty-$pricing error bounds.

Concerning the second aspect above, the three right plots of Figure 4 study the effects of moderate data perturbations, creating a discrepancy of about 10 years of monthly observations between S–SDF estimation and validation samples, for the S–SDF tradeoff between time series and cross-sectional explanatory powers. We find that APT S–SDFs satisfying an l_2-pricing error bound tend to produce a clearly more consistent tradeoff on perturbed data, essentially outperforming the other S–SDFs with respect to all GLS R^2 metrics.

[41] Intuitively, the second aspect relates to the fact that empirical S–SDFs with better resistance to varying in-sample data features are likely to perform better out-of-sample.

3.2.4 The role of sparsity in APT S–SDFs

As emphasized in Section 2.1, sparsity cannot be obtained for both pricing errors and dual optimal portfolios of minimum dispersion S–SDFs. Therefore, in an APT context, one can either impose sparsity on scaled pricing errors $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}^{-1/2}\mathbb{E}[M_0\boldsymbol{X}_D]$ or on scaled portfolio weights $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}^{1/2}\boldsymbol{\theta}_{0D}$. Figure 5 documents these sparsity properties for the three minimum variance S–SDFs with 50% target GLS R^2 in Panel (B) of Figure 3.

In Panel (A) of Figure 5, the l_1-pricing error norm gives rise to sparse pricing errors. In contrast, the (scaled) $l_\infty-$norm produces extreme shrinkage and no pricing error sparsity. The l_2-norm induces shrinkage and no sparsity as well, but with maximal pricing errors larger than under the (scaled) $l_\infty-$norm. The properties of S–SDF dual portfolio weights follow from the corresponding penalizations $\sigma_h = \tau \left\|\cdot\right\|_{*,\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}}$ in Proposition 3. Since the dual norm of the l_1-norm is the $l_\infty-$norm, the l_1-pricing error bound produces extreme shrinkage in dual portfolio weights and no sparsity. Analogously, the (scaled) $l_\infty-$pricing error norm gives rise with a l_1-penalization to extreme sparsity in dual portfolio weights. Finally, the self-duality of the l_2-norm induces shrinkage but no sparsity in dual portfolio weights.

While linked to S–SDFs with different sparsity properties, the S–SDF time series in Panel (B) of Figure 5 are superficially quite similar in a number of dimensions.[42] However, they give rise to economically important differences in time series explanatory power for asset returns. As a consequence, the choice of a pricing error norm and the corresponding S–SDF sparsity features is in the end strongly linked to a particular tradeoff choice between S–SDF cross-sectional and time series explanatory power. Our minimum dispersion APT S–SDF theory in Proposition 3 makes this tradeoff empirically measurable and indicates that S–SDF sparsity is quite costly with respect to the implied tradeoff between cross-sectional and time series explanatory powers under metrics (28) and (29).

3.2.5 Sure assets and systematic S–SDF risks

Intuitively, sure assets in our framework correspond to tradable systematic excess returns with a clearly understood risk compensation. Importantly, the exact pricing condition on sure excess returns forces via equation (29) a maximal co-movement with minimum variance S–SDFs, i.e., these S–SDFs maximize the co-movement with the systematic risks spanned by the sure excess returns. Therefore, the choice of the sure asset payoffs directly influences the S–SDF tradeoff between cross-sectional and time series explanatory

[42]All S–SDF time series exhibit quite similar volatilities between about 1.75 and 1.85, as well as a quite substantial co-movement, with time series correlations above 0.8 between S–SDFs induced by the l_2-pricing error norm and the other S–SDFs.

power.

Figure 6 explores these tradeoffs for various minimum variance S–SDFs with the l_2-APT pricing error bound (18), under different choices of the sure assets: (i) no sure asset, (ii) one single sure asset given by the market excess return or by the first principal component of excess returns, and (iii) three sure assets given by the three Fama-French factor excess returns. We find that all S–SDFs imply a virtually identical cross-sectional GLS R^2 curve in-sample.[43] However, these S–SDFs in part strongly differ in the way how they span systematic excess return risks. The S–SDFs including as a single sure asset the market excess return or the first principal component of excess returns produce a quite similar and highest average time series R^2 curve. In constast, S–SDFs including no sure asset uniformly produce a very low time series explanatory power that is essentially independent of the chosen degree of penalization. Therefore, including sure tradable risk exposures that reflect well-understood systematic risks allows to best optimize the S–SDF trade-off between time series and cross-sectional explanatory power under metrics (28) and (29).

Assuming the market excess return as a single sure payoff gives rise to APT S–SDFs that are interpretable as perturbations of an empirical SDF under the CAPM, with large S–SDF average time series R^2s for returns that reflect the systematic nature of market risk. Interestingly, extending the set of sure excess returns to include the three Fama-French factor returns produces a perturbation of an empirical SDF under the three-factor Fama-French model. In our sample, we find that this extension does not help to improve on the simpler SDF perturbation of the CAPM.[44]

Overall, we conclude that minimum variance APT S–SDFs bounding pricing errors on dubious returns with the APT pricing error bound (21) and exactly pricing market excess returns or the first principal component of sorted portfolio excess returns consistently offer the best in-sample tradeoff between pricing accuracy and time series explanatory power.

3.2.6 APT S–SDFs, Non APT S–SDFs and Relation to Kozak et al. [2020]

When pricing accuracy is measured by the standard, not self-standardized, cross-sectional R^2 metric:

$$R^2_{OLS} := 1 - \frac{\left\|\mathbb{E}[M_0(\boldsymbol{X} - \bar{\boldsymbol{X}}\mathbf{1})]\right\|_2^2}{\left\|\mathbb{E}[\boldsymbol{X} - \bar{\boldsymbol{X}}\mathbf{1}]\right\|_2^2} , \tag{31}$$

[43] With the only distinction that in the intermediate dimensional dataset the minimum variance S–SDF including the Fama-French factors as sure assets does not exist for a sufficiently low APT pricing error threshold.

[44] This evidence may depend on the fact that the Fama-French size and book-to-market factors are not strong systematic factors for our panels of returns, given that the latter strictly extend the set of double-sorted size and book-to-market portfolio returns. It may also reflect the fact that Fama-French size and book-to-market factor returns are less easily tradable than market returns, which naturally motivates their treatment as dubious rather than sure returns.

the corresponding optimal minimum variance S–SDFs simply need to bound pricing errors with a constrained form of l_2-pricing error bound (21), in which $\mathbf{\Sigma} := \boldsymbol{I}_{N\times N}$, i.e., such optimal S–SDFs are by construction non APT–SDFs.

The two top panels of Figure 7 illustrate for the high dimensional dataset the optimal tradeoffs implied by minimum variance APT and non APT S–SDFs exactly pricing the first principal component of excess returns, under metrics (28) and (31) of pricing accuracy, respectively.[45] Importantly, both such S–SDFs are sparse in two corresponding portfolios of excess returns: the first principal component of excess returns itself and a second optimal portfolio of returns orthogonal to traded systematic risk, which bounds the overall S–SDF mispricing with an APT bound (28) and a non APT bound (28) such that $\mathbf{\Sigma}_{\boldsymbol{\zeta}} := \boldsymbol{I}_{N_D\times N_D}$, respectively.

Kozak et al. [2020] study the sparsity implications of minimum variance S–SDFs in large asset markets. As elaborated in Section 2, their S–SDFs can be obtained with an Elastic Net norm penalization $\sigma_h = \alpha\left\|\cdot\right\|_1 + \tau\left\|\cdot\right\|_2$ and a pricing error metric $h = \text{dist}_{\alpha B_\infty}$; see again Table 1. Therefore, for $\alpha = 0$ these S–SDFs are not sparse and optimize the cross-sectional pricing accuracy under metric (31). In contrast, for $\tau \neq 0$ they are sparse and give rise to a suboptimal tradeoff between the S–SDF time series explanatory power and pricing accuracy under metric (31). However, Kozak et al. [2020] show that this sparsity-induced suboptimality is less serious when it is imposed on the set of all principal components of returns, rather than on the original returns.

The bottom right panel of Figure 7 illustrates the suboptimality of the tradeoff between S–SDF time series explanatory power and pricing accuracy under metric (31), when sparsity is induced on the endogenously selected principal components in Kozak et al. [2020] S–SDFs. It additionally shows that the tradeoff of a non APT S–SDF exactly pricing the first principal component of excess returns alone is essentially identical to the one of the optimal non sparse Kozak et al. [2020] S–SDF, which depends on all principal components of excess returns. Finally, the bottom left panel of Figure 7 shows that principal component APT S–SDFs not forcing exact pricing of the first principal component are on average essentially uncorrelated with excess returns, i.e., they are unable to generate a tradeoff between S–SDF time series explanatory power and pricing accuracy.

In summary, our minimum variance APT and non APT S–SDFs exactly pricing the first principal component of excess returns provide an optimal tradeoff between pricing accuracy and time series explanatory power

[45] Analogous evidence for the low dimensional dataset is reported in Figure 4 of Section H in the Online Appendix.

under metrics (28) and (31) of cross-sectional pricing accuracy. These S–SDFs are by construction sparse in the returns of two distinct portfolios, given by the first principal component of returns and a second optimal portfolio that bounds the mispricing across assets under l_2−APT pricing error bound (21) and under l_2−non APT pricing error bound (21) with $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}} := \boldsymbol{I}_{N_D \times N_D}$, respectively.

3.3 Out-of-sample analysis

We next characterize the out-of-sample tradeoff between pricing accuracy and time series explanatory power generated by minimum variance APT S–SDFs.[46] To this end, similarly to Ghosh et al. [2016], we sequentially estimate minimum variance APT S–SDFs on rolling windows of 30 years of monthly returns. We then evaluate the out-of-sample S–SDF tradeoff between pricing accuracy and time series explanatory power according to metrics (28) and (29). Given our earlier results, we focus on minimum variance APT S–SDFs implying varying sparsity properties based on pricing errors constrained with norms $\|\cdot\|_1$, $\|\cdot\|_2$, and $\sqrt{N_D}\,\|\cdot\|_\infty$, and consider various choices for the set of sure assets, such as the market excess return alone, in which case our S–SDFs reflect economically relevant perturbations of an empirical SDF under the CAPM.

3.3.1 Basic out-of-sample framework and evidence

For every in-sample window of 30 years of monthly observations with last observation in month y, the in-sample S–SDF estimation requires the specification of a pricing error threshold, denoted by $\hat{\tau}_y$.[47] Similar to the in-sample analysis, our theoretical findings offer a natural way to determine the range of relevant empirical pricing error thresholds. We obtain the maximal admissible threshold $\hat{\tau}_y^{max}$ as the empirical version of equation (9). For instance, under a set of sure payoffs including the market excess return alone, this corresponds to the empirical pricing error threshold implied by a minimum variance SDF that is only required to price exactly the market excess return in the given estimation window. The minimal admissible threshold $\hat{\tau}_y^{min}$ is obtained similarly to Section 3.2.1 as the smallest threshold value for which in the given estimation window no empirical duality failure arises. This yields for any estimation window a range $[\hat{\tau}_y^{min}, \hat{\tau}_y^{max}]$ of admissible thresholds $\hat{\tau}_y$ and a range $[\hat{\sigma}_y^{min}, 1]$ of admissible relative thresholds $\hat{\sigma}_y$, where $\hat{\sigma}_y^{min} := \hat{\tau}_y^{min}/\hat{\tau}_y^{max}$.

[46] We adopt a simple approach that does not require an explicit modelling of conditional information structures. While the modelling of conditioning information is clearly an important dimension, it would require a more extended treatment formally dealing with a potentially large number of information variables and a growing cross-sectional asset dimension. While recent important research has incorporated high-dimensional conditioning information in the construction of model-free SDFs founded by a standard no-arbitrage condition in frictionless markets (see again Gu et al. [2020a] and Chen et al. [2020a]), to our knowledge no such attempt has been proposed yet for S–SDFs embedding convex non zero pricing errors on some assets.

[47] Given our data sample, this gives us an out-of-sample period of monthly observations from July 1963 to June 2018. The first in-sample window of 30 years consists of monthly observations from July 1933 to June 1963.

Given a sequence of admissible relative thresholds $\{\hat{\sigma}_y\}$, we apply the estimation methodology in Section 3.1, to obtain a sequence $\{\hat{\boldsymbol{\theta}}_y\}$ of estimated S–SDF dual portfolio weights, which is updated at a semi-annual frequency. For each of the six months following month y, we denote by $\boldsymbol{X}_{y+m}$ the vector of excess returns in month $y+m$ and estimate with Proposition 2 and 3 the corresponding sequence of out-of-sample monthly minimum variance S–SDFs, i.e., $\hat{M}_{y+m} := \max\{-\hat{\boldsymbol{\theta}}_y' \boldsymbol{X}_{y+m}, 0\}$. In this way, we obtain a monthly time series $\{\hat{M}_{y+m}\}$ of data driven APT S–SDFs, for which we can evaluate the out-of-sample time series and cross-sectional explanatory power for the sequence of out-of-sample return vectors $\{\boldsymbol{X}_{y+m}\}$.

Regarding the choice of the sequence of relative thresholds $\{\hat{\sigma}_y\}$, we first adopt in this section a simple approach based on constant relative thresholds $\hat{\sigma}_y = \overline{\sigma}$. This approach produces a useful description of the attainable out-of-sample tradeoffs between time series and cross-sectional S–SDF explanatory power, which may arise from data-driven threshold selections holding the relative threshold roughly constant across estimation windows, but it does not produce a criterion for a non forward-looking choice of an optimal threshold. The first two rows of Figure 8 report the resulting curves of out-of-sample average time series R^2 (28) and cross-sectional GLS R^2 (29), in dependence of the value of constant relative threshold $\overline{\sigma} \in [\max_y\{\hat{\sigma}_y^{min}\}, 1]$, for minimum variance S–SDFs incorporating the market excess return as the single sure payoff. All out-of-sample GLR R^2 curves are inversely U–shaped. In the low (intermediate) dimensional data set, the maximal GLR R^2 is as high as about 70% (45%) for the APT S–SDF with l_2-constrained pricing errors, when the relative pricing error threshold $\overline{\sigma}$ is about 40% (65%). These are large increases in pricing accuracy, compared, e.g., to the mildly negative out-of-sample GLR R^2s of an empirical CAPM S–SDF exactly pricing only the market return in-sample ($\overline{\sigma} = 1$).

Importantly, also in the out-of-sample evidence we observe a tradeoff between time series and cross-sectional S–SDF explanatory power, which is summarized in the two bottom panels of in Figure 8, for relative pricing error thresholds between the threshold value at which the maximum of the GLS R^2 curve is attained and the maximum admissible value $\overline{\sigma} = 1$. Therefore, also out-of-sample the S–SDF pricing accuracy has to be balanced against the S–SDF time series explanatory power. Similar to the in-sample findings of Section 3.2.3, we find that APT S–SDFs with pricing error metric different from the l_2-norm imply a less favourable out-of-sample tradeoff between cross-sectional and time series explanatory power. This evidence shows that S–SDF sparsity is quite costly also from the perspective of the S–SDF out-of-sample tradeoff between cross-sectional and time series explanatory power.

3.3.2 Optimal data-driven minimum variance APT S–SDFs

To implement optimal minimum variance APT S–SDFs in a non forward-looking way, we propose a transparent data-driven approach, which incorporates varying views on the target in-sample S–SDF tradeoff between times series and cross-sectional explanatory powers. We split each window of 30 years of monthly data into a training window of 25 years and a separate window of 5 years. For the prevailing range of admissible pricing error thresholds $\hat{\sigma}_y \in [\hat{\sigma}_y^{min}, 1]$ in the training window, we estimate various minimum variance APT S–SDFs with bounded APT pricing errors using only the training data. Based on the whole window of 30 years of data, we then compute the resulting curves of S–SDF metrics $R^2_{GLS}(\hat{\sigma}_y)$ and $R^2(\hat{\sigma}_y)$ defined in equations (28) and (29). Finally, we select an optimal pricing error threshold $\hat{\sigma}_y^\star := \hat{\sigma}_y^\star(p)$, by maximizing cross-sectional metric $R^2_{GLS}(\hat{\sigma}_y)$ under a constraint on the average time series metric $\overline{R^2}(\hat{\sigma}_y)$ across assets. This constraint stipulates that $\overline{R^2}(\hat{\sigma}_y)$ has to be not lower than a fraction $p \in [0, 1]$ of the average time series metric $\overline{R^2}(1)$ under an empirical SDF only required to price exactly the sure assets:

$$\hat{\sigma}_y^\star = \arg \max_{\hat{\sigma}_y \in [\hat{\sigma}_y^{min}, 1]} \{R^2_{GLS}(\hat{\sigma}_y) : \overline{R^2}(\hat{\sigma}_y) \geq p\overline{R^2}(1)\} \ . \tag{32}$$

We compute optimal thresholds $\hat{\sigma}_y^\star$ for the choices $p = 0, 0.25, 0.5, 0.75, 1$, where the first choice corresponds to an optimization (32) with no constraint on the S–SDF time series explanatory power.

From the sequence of optimal thresholds $\{\hat{\sigma}_y^\star\}$, we obtain the sequence $\{\hat{\boldsymbol{\theta}}_y^\star\}$ of optimal S–SDF dual portfolio weights, which is updated at a semi-annual frequency. We then estimate the sequence $\{\hat{M}_{y+m}\}$ of out-of-sample monthly minimum variance S–SDFs $\hat{M}_{y+m}^\star := \max\{-\boldsymbol{X}_{y+m}' \hat{\boldsymbol{\theta}}_y^\star, 0\}$. Finally, we evaluate the out-of-sample time series and cross-sectional explanatory power for returns $\{\boldsymbol{X}_{y+m}\}$ using metrics (28) and (29) computed from out-of-sample return and S–SDF data.

Figure 9 collects the out-of-sample evidence on data-driven minimum variance S–SDFs for various intuitive choices of the set of sure assets. Given our earlier results, we focus in the sequel on l_2-bounded APT pricing errors. In the top two panels of Figure 9, we obtain an out-of-sample tradeoff between S–SDF time series and cross-sectional explanatory power that is consistent with the in-sample selection of the sequence of optimal thresholds $\{\hat{\sigma}_y^\star\}$. For instance, for the S–SDFs including the market excess return as the single sure payoff, the out-of-sample GLS R^2 metric (28) in the low (intermediate) dimensional data set decreases monotonically with the tightness of the constraint on the in-sample time series explanatory power, yielding GLS R^2s between a maximum of 68% (38%) and a minimum of -2% (-2%) for parameters

$p = 0, 0.25, 0.5, 0.75, 1$. Correspondingly, the out-of-sample average time series R^2s are increasing with the tightness of the in-sample constraint from a minimum of 2.8% (1%) to a maximum of 68% (62%). An analogous pattern arises for other choices of sure assets, given by the first principal component of returns and the three Fama-French factors, respectively.

The parameter choice $p = 1$ corresponds to minimum variance S–SDFs only required to price exactly the sure excess returns in-sample. Given the choice of the market excess return as the single sure return, the resulting S–SDF produces a mildly negative cross-sectional GLS R^2, together with a large average time series R^2 of 62% (68%) in the intermediate (low) dimensional dataset. In comparison, a minimum variance S–SDF only required to price exactly the three Fama-French factors produces an even more negative GLS R^2 and an average time series R^2 of only 18% (22%) in the intermediate (low) dimensional dataset. This explains why the minimum variance S–SDFs incorporating the three Fama-French factors as sure assets produce a clearly less favourable out-of-sample tradeoff between time series and cross-sectional explanatory power than those incorporating the market as the single sure asset.

A more favourable tradeoff is achieved by minimum variance S–SDFs incorporating the first principal components of returns as a single sure payoff in-sample. Interestingly, while these S–SDFs tend to produce larger average time series R^2s, due to the more consistent out-of-sample time series explanatory power of the first principal component of returns relative to market returns, the resulting tradeoff between time series and cross-sectional explanatory power is quite comparable to the one induced by the S–SDF pricing exactly market returns alone. Notably, this S–SDF is a direct and simple model-free adaptation of an empirical SDF under the CAPM, which exactly prices market risk but otherwise constrains the amount of mispricing across assets with a standard APT pricing error bound.

Following the intuition underlying the APT, model-free APT S–SDFs able to partly span systematic return shocks need to correlate with potential common risk factors in asset returns. The bottom two panels of Figure 9 illustrate this intuition more precisely, by reporting the fraction of out-of-sample variation in model-free APT S–SDFs that is explained by variations of the first principal component of sorted portfolios excess returns. Consistent with our previous findings, APT S–SDFs maximizing the in-sample pricing accuracy produce the highest out-of-sample pricing accuracy, but they are also almost uncorrelated with the first principal component of excess returns out-of-sample. Conversely, APT S–SDFs only required to exactly price systematic risk exposures in-sample produce the lowest out-of-sample pricing accuracy and the largest co-movement with the first principal component of excess returns. Such co-movement is naturally highest, and rather similar, for S–SDFs exactly pricing either market returns or the first principal component of sorted portfolios excess returns in-sample. Further, it appears that the tradeoff between pricing accuracy and the S–SDF ability to span systematic risks is rather steep. For instance, in the low (intermediate) dimensional dataset, a cross-sectional GLS R^2 of about 20% and 33% (12% and 18%) is attained under a time series R^2 with the first principal component of excess returs amounting to about 60% and 38% (43% and 18%), respectively. This evidence is compatible with the existence of a non negligible expected excess return component, not reflecting a compensation for systematic asset return risks, that is proportionally more apparent in higher dimensional datasets.

Finally, we also find that the APT S–SDFs in Figure 9 clearly outperform linear SDF specifications globally minimizing the in-sample pricing error size according to a standard GMM criterion. More specifically, given the linear SDF specification:

$$M(\boldsymbol{\theta}) := 1 + \boldsymbol{\theta}'(\boldsymbol{F}^e - \mathbb{E}[\boldsymbol{F}^e]) \ , \tag{33}$$

where $\boldsymbol{F}^e$ is a $k \times 1$ random vector of factor excess returns, we estimate the optimal SDF $M_0 := M(\boldsymbol{\theta}_0)$ implied by following solution of a GMM estimation problem:[48]

$$\boldsymbol{\theta}_0 := \arg\min ||\mathbb{E}[M(\boldsymbol{\theta})\boldsymbol{X}]||_{2,\boldsymbol{\Sigma}^{-1/2}} \ . \tag{34}$$

As for the APT S–SDFs of Figure 9, we estimate such SDFs on rolling windows of 30 years of monthly data, while updating the resulting GMM estimates at a semi-annual frequency. We then evaluate their out-of-sample cross-sectional and time series explanatory power for different choices of the set of factors. In all cases, we obtain pairs of cross-sectional and time series explanatory power that are dominated by corresponding pairs induced by some S–SDF in Figure 9. For instance, a CAPM SDF specification with a single market factor implies a cross-sectional GLS R^2 of -0.2% (0.5%) and an average time series R^2 of 51.8% (52%) in the low (intermediate) dimensional dataset. Similarly a three-factor Fama-French SDF specification implies a cross-sectional GLS R^2 of 2.8% (0.4%) and an average time series R^2 of 26.9% (26.3%) in the low (intermediate) dimensional dataset.

[48]Note that in contrast to the S–SDFs in Figure 9, all these SDFs exclusively depend on risks that are spanned by the given set of factors in vector $\boldsymbol{F}_e$.

4 Conclusions

This paper provides a unifying framework for selecting model-free stochastic discount factors that satisfy general convex pricing constraints in arbitrage-free markets (S–SDFs). We theoretically establish the economic importance of S–SDFs and show that they are rooted in a large class of arbitrage-free economies with either transaction costs or ambiguity. This setting nests several seminal asset pricing approaches in the literature that can be studied with our S–SDF methodology. Our S–SDF selection framework is based on minimum dispersion S–SDFs defined by the solution of a general minimum dispersion problem with convex pricing constraints. We establish a strong duality between minimum dispersion S–SDF problems and corresponding penalized portfolio selection problems, in which the choice of the penalization stays in a one-to-one relation to the imposed pricing constraint. This result identifies minimum dispersion S–SDFs as a known transformation of an optimal portfolio payoff, crystallizing the joint implications of pricing error constraints and portfolio penalization choices for minimum dispersion S–SDFs.

Exploiting our S–SDF methodology, we propose a family of minimum variance APT S–SDFs for empirically studying the relation between risk and return in cross-sections of asset returns. We show that this family optimally synthesizes the attainable tradeoffs between S–SDF pricing accuracy vs. S–SDF comovent with systematic asset returns risks, and we characterize these tradeoffs empirically for several cross-sections of characteristics sorted portfolio returns. We find that while APT S–SDFs produce a well-defined tradeoff also in markets with a growing number of assets, empirical SDFs exactly pricing all assets may fail to exist in such settings. We further find that the best tradoff is attained by minimum variance APT S–SDFs that exactly price only the first principal component of excess returns. These S–SDFs are sparse in two interpretable portfolios: the first principal component of excess returns and a second optimal portfolio of returns orthogonal to traded systematic risks, which bounds the overall S–SDF mispricing under a standard APT bound. Given the large co-movement between market returns and the first principal component of sorted portfolio returns, we conclude that a model-free adaptation of an empirical SDF under the CAPM, which exactly prices market risk but otherwise constrains pricing errors with a standard APT pricing error bound, also produces an optimal tradeoff.

Appendix A - Figures and Tables

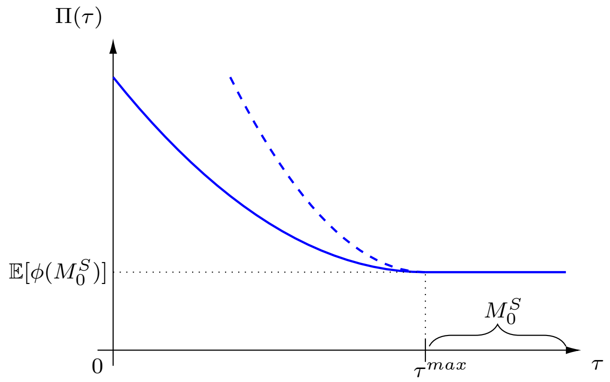


Figure 1: **Properties of minimum S–SDF dispersion curve.** The solid and dashed curves illustrate two possible shapes of the minimum dispersion bound $\Pi(\tau)$, which may change due to varying $\Phi-$dispersion functions, pricing error functions h, and market partitions $\{S, D\}$.

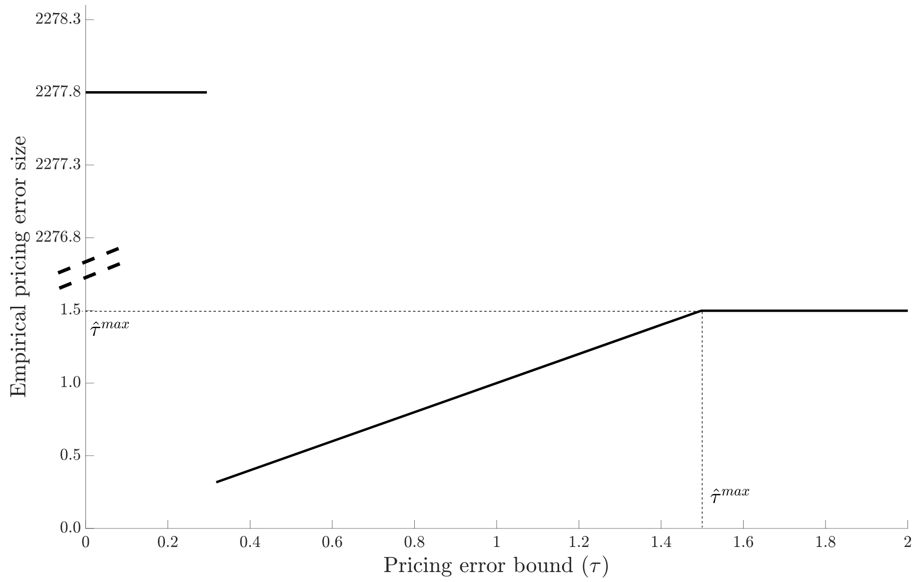
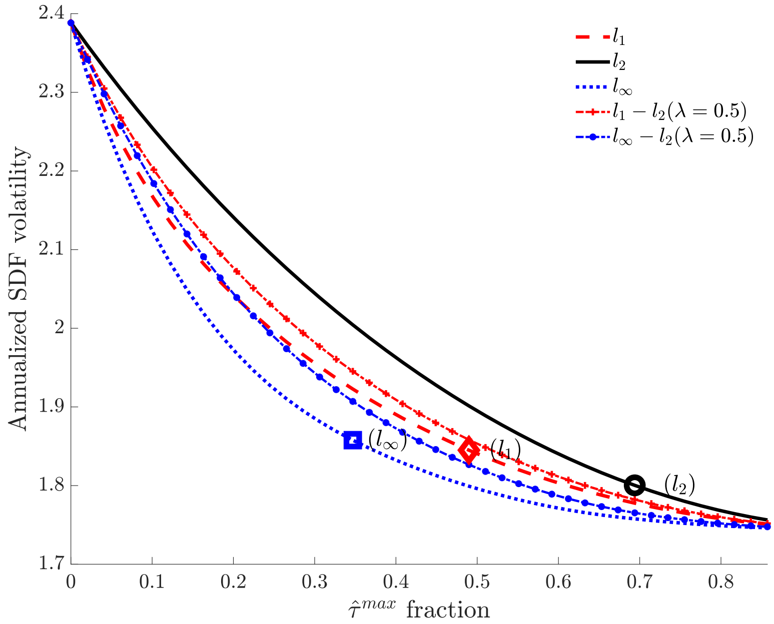
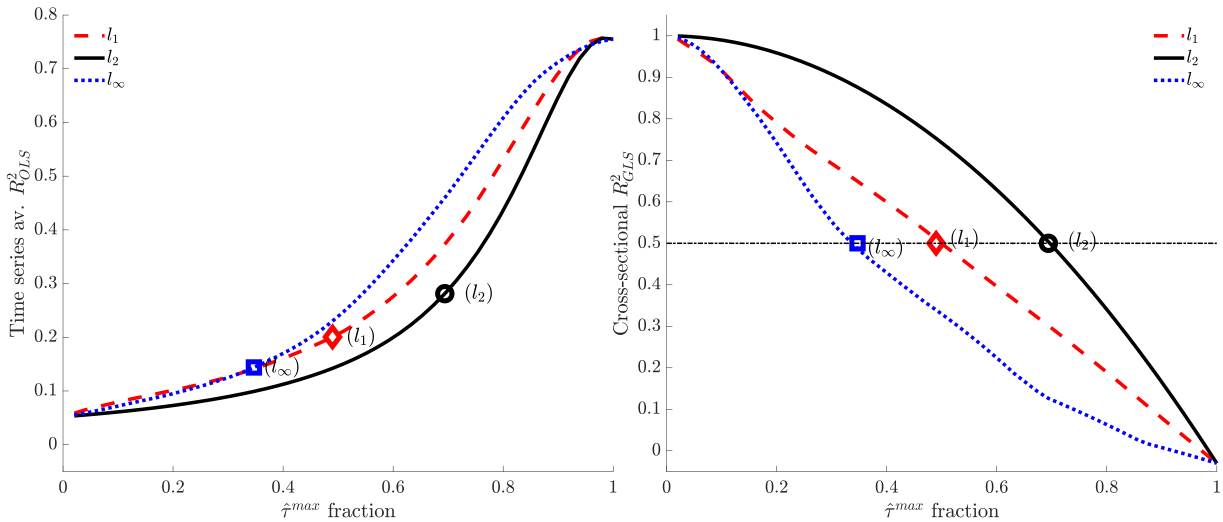


Figure 2: **Empirical duality failure.** We compute the minimum variance APT S–SDF M_0 with l_2-pricing error function for varying pricing error thresholds $\tau \geq 0$. On the y-axis we report the estimate of the pricing error function value $h(\mathbb{E}[M_0\boldsymbol{X} - \boldsymbol{P}])$ for each τ. The point of discontinuity in the plot identifies the smallest pricing error bound τ, for which a solution of the empirical primal S–SDF problem exists. The largest pricing error threshold $\hat{\tau}^{max}$ in the plot is computed as the sample version of the maximal threshold τ^{max} in equation (9). All calculations are based on the high dimensional dataset from June 1990 to June 2018, using the risk-free asset as the only sure asset and all sorted portfolios as the dubious assets.

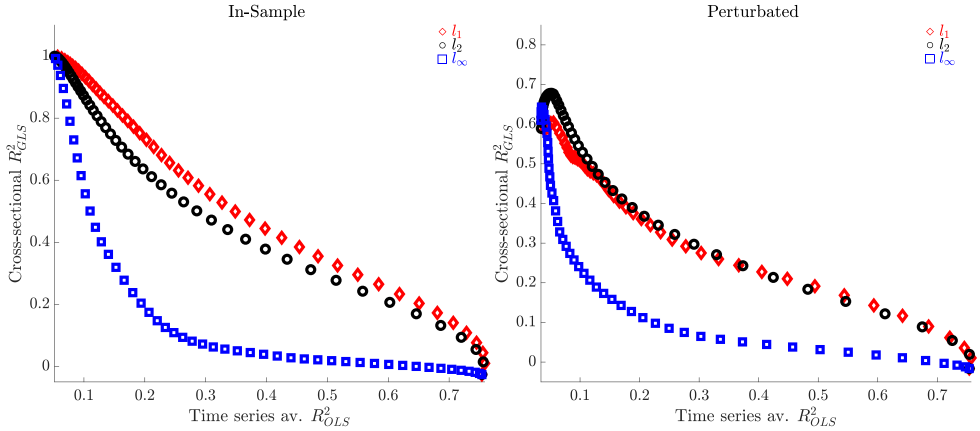


(1) **Panel (A)**

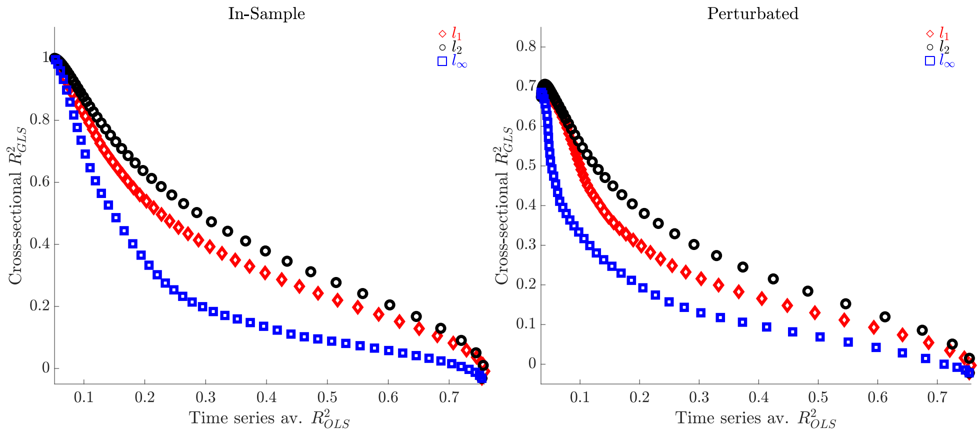


(2) **Panel (B)**

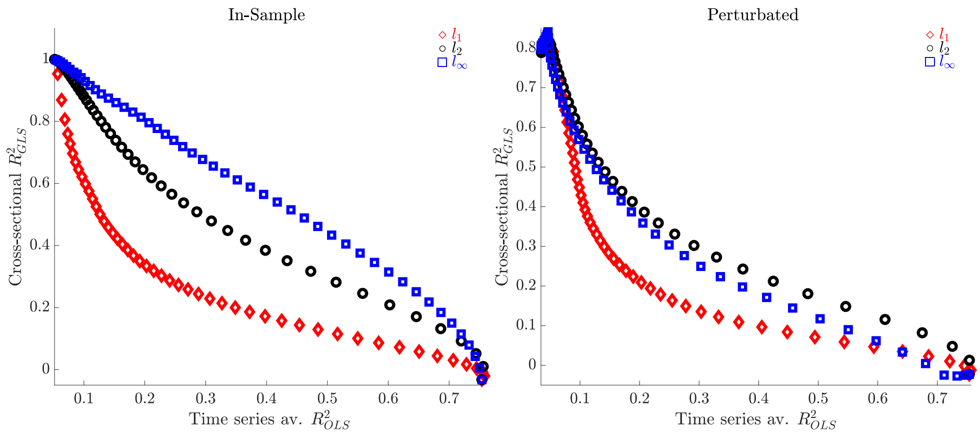
Figure 3: **Minimum APT S–SDF volatility curve and tradeoff between time-series and cross-sectional S–SDF explanatory power. Panel (A)** reports the annualized volatility of minimum variance APT S–SDFs as a function of the relative pricing error threshold $\sigma \in (\tau^{min}/\tau^{max}, 1]$, with threshold τ^{max} defined in equation (9). We consider the pricing error metrics $h = \|\cdot\|_1$ (dashed line), $h = \|\cdot\|_2$ (solid line) and $h = \sqrt{N_D}\,\|\cdot\|_\infty$ (dotted line). The two additional curves report the minimum S–SDF volatility curve for pricing error metrics $h = \lambda\,\|\cdot\|_1 + (1-\lambda)\,\|\cdot\|_2$ (dashed-crossed line) and $h = \sqrt{N_D}\lambda\,\|\cdot\|_\infty + (1-\lambda)\,\|\cdot\|_2$ (dashed dotted line), where $\lambda = 0.5$. **Panel (B)** reports the average time series R^2 metric (29) (left panel) and the cross-sectional GLS R^2 metric (28) (right panel) for the first three minimum variance S–SDFs from Panel (A). Finally, the diamonds, the circles, and the squares in Panel (B) identify the corresponding S–SDFs attaining a cross-sectional GLS R^2 of 50%. All calculations are based on the low dimensional dataset from July 1963 to June 2018, using the risk-free asset and the market as the only sure assets, and all sorted portfolios as the dubious assets.



(1) **Panel (A)**

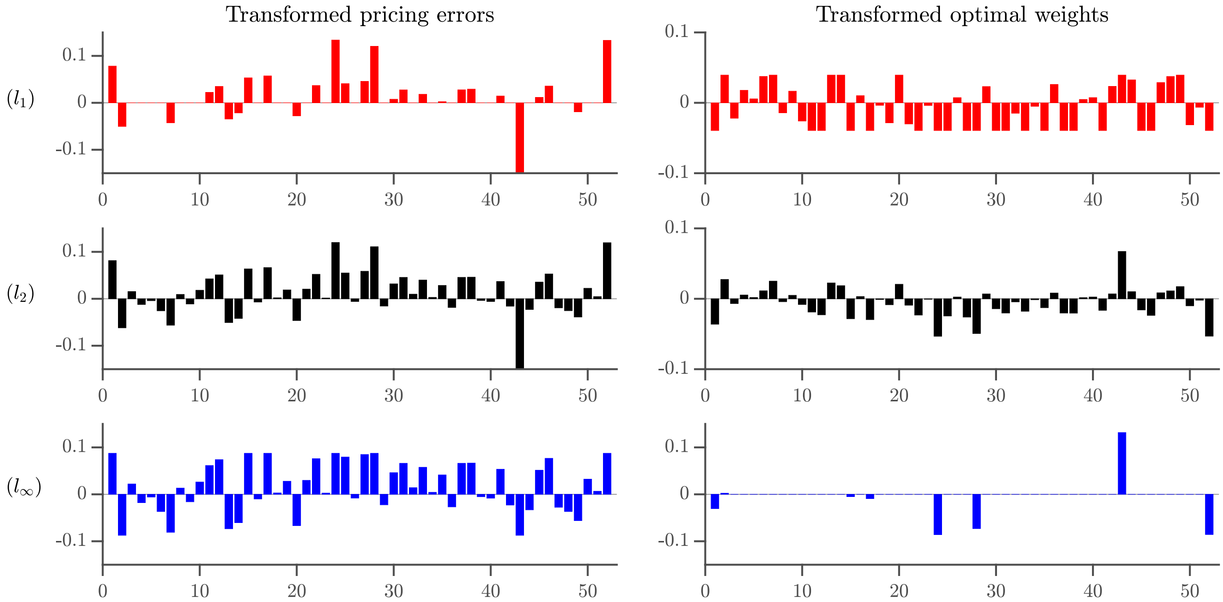


(2) **Panel (B)**

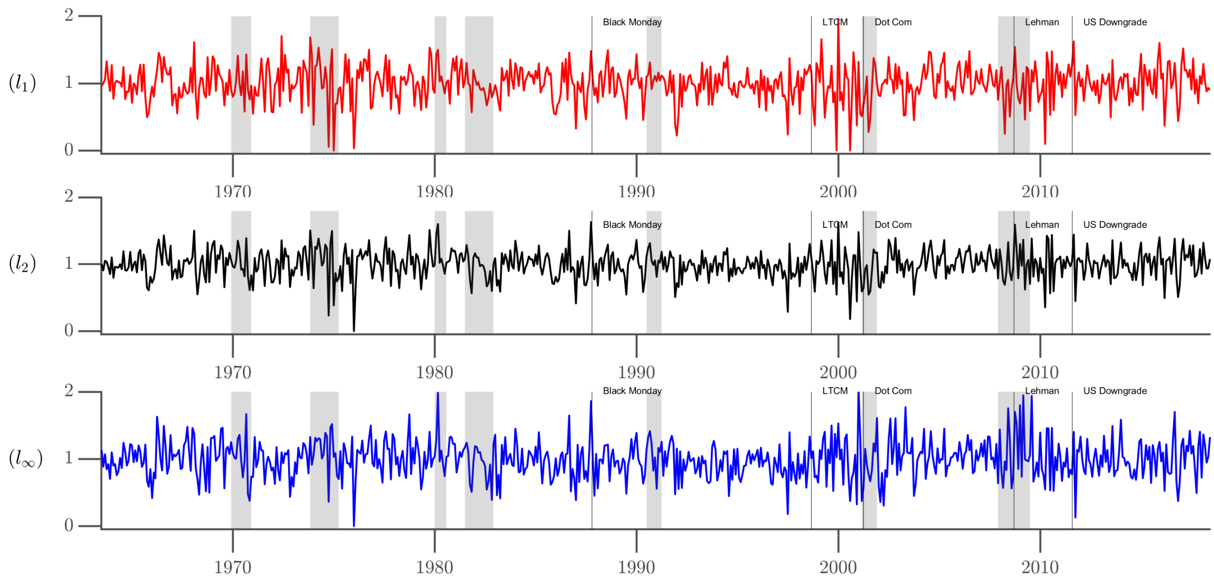


(3) **Panel (C)**

Figure 4: **Tradeoff between time series and cross-sectional APT S–SDF explanatory power.** The figure reports the tradeoff between average time-series R^2 metric (29) and cross-sectional GLS R^2 metric (30), using the l_1-norm in **Panel (A)**, the l_2-norm in **Panel (B)**, and the (scaled) $l_\infty-$norm in **Panel (C)**, for various minimum variance APT S–SDFs based on a l_1-, l_2-, and (scaled) $l_\infty-$pricing error metric, and while taking as sure assets the risk-free asset and the market. Left panels report in-sample results where the estimation and evaluation periods overlap on the 40 year window 1965-2005. Right panels report results for a 40 year estimation window 1965-2005 and a 40 year evaluation window 1975-2015. All computations are based on the low dimensional dataset.

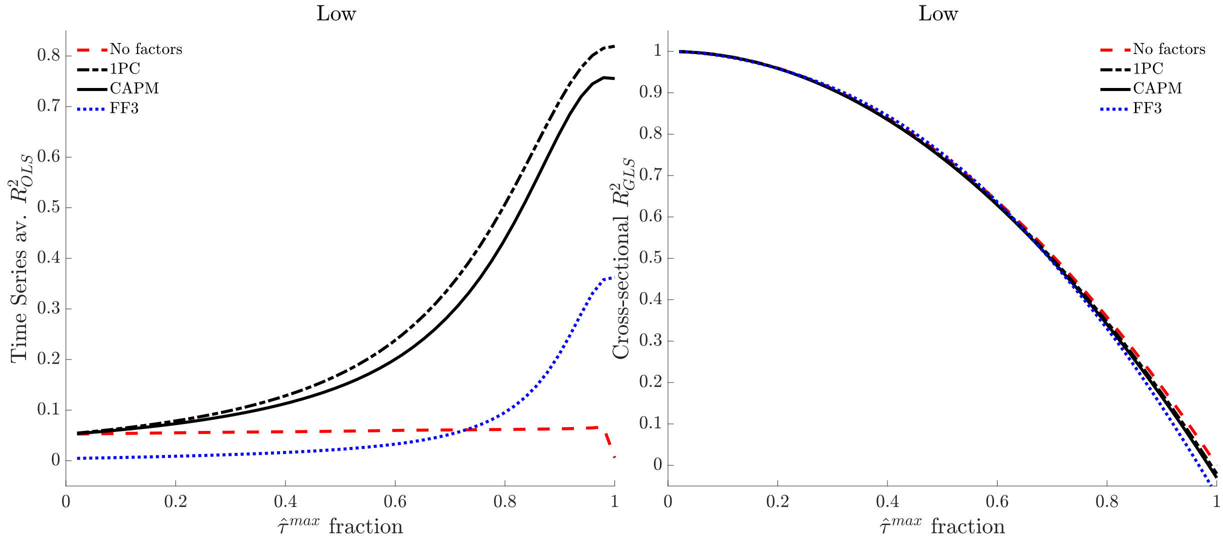


(1) **Panel (A)**

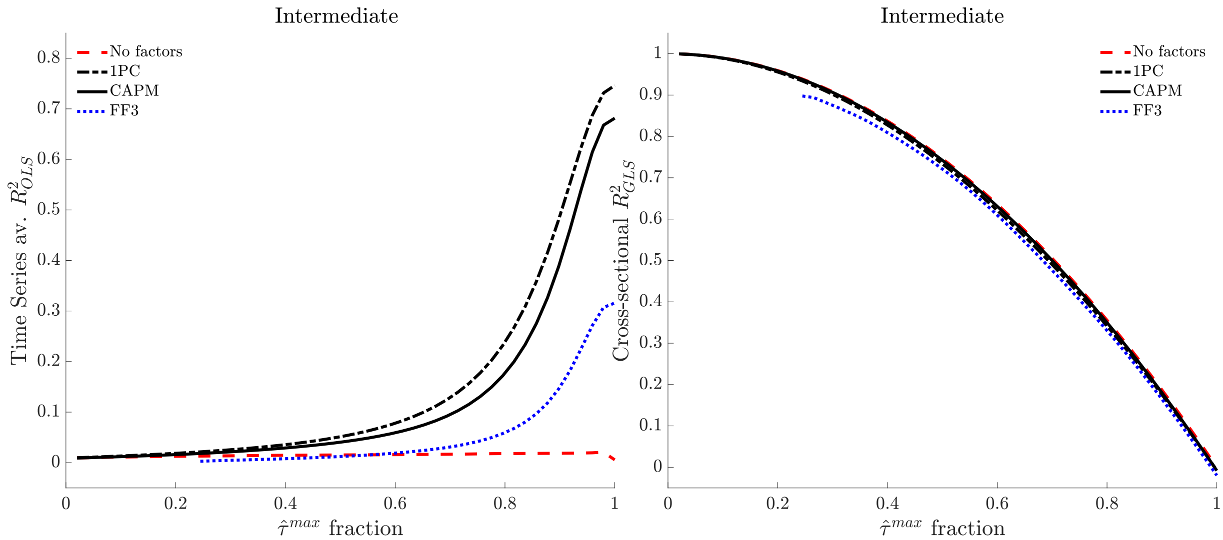


(2) **Panel (B)**

Figure 5: **Properties of APT S–SDFs.** **Panel (A)** reports, from top to bottom, $\mathbf{\Sigma}_\zeta^{-1/2}-$transformed pricing errors (left) and $\mathbf{\Sigma}_\zeta^{1/2}-$transformed optimal portfolio weights (right) of the three minimun variance APT S–SDFs highlighted with a diamond, a circle and a square in Panel (B) of Figure 3, which correspond to a l_1- (first row), l_2- (second row), and scaled $l_\infty-$ (third row) APT pricing error metric, respectively. **Panel (B)** reports the time-series of these three minimum variance APT S–SDFs. All calculations are based on the low dimensional dataset from July 1963 to June 2018, using the risk-free asset and the market as the only sure assets, and all sorted portfolios as the dubious assets. Grey shaded areas in Panel (B) highlight NBER recession periods.

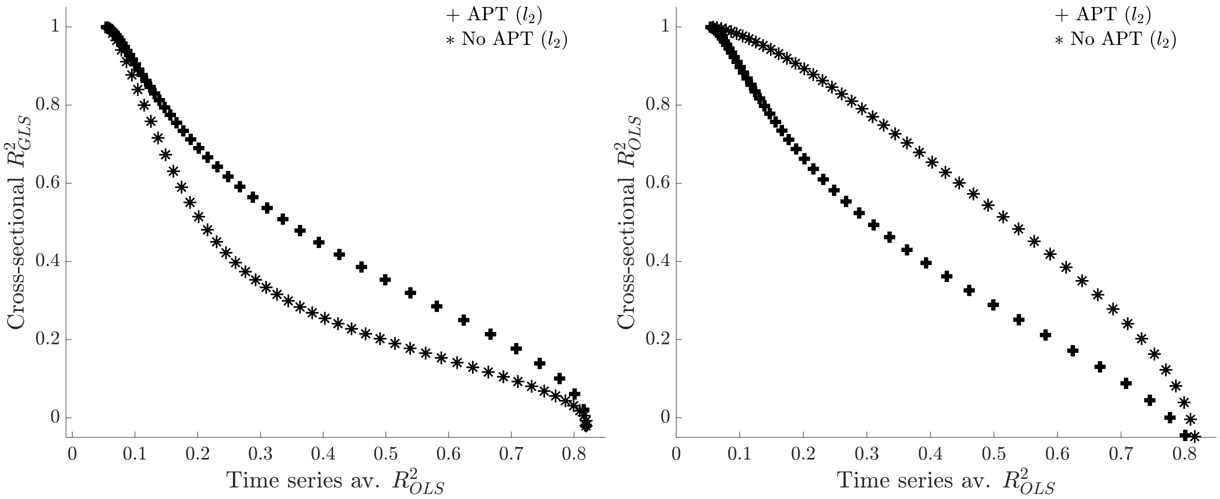


(1) **Panel (A)**

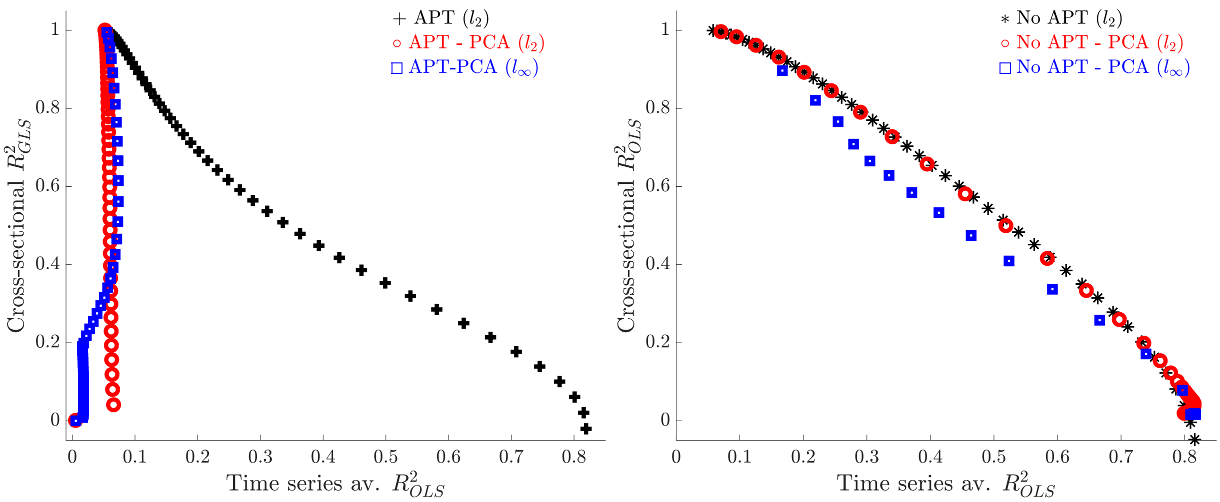


(2) **Panel (B)**

Figure 6: **Role of sure assets in minimum dispersion APT S–SDFs.** The figure reports the average time-series R^2 metric (29) (left panel) and cross-sectional GLS R^2 metric (28) for various minimum variance APT S–SDFs, corresponding to different choices for the set of sure assets, in the low dimensional (**Panel (A)**) and intermediate dimensional (**Panel (B)**) datasets. We consider following choices of sure assets: (1) No factors, corresponding to APT SDFs pricing exactly only the risk-free rate; (2) CAPM, corresponding to APT S–SDFs pricing exactly the market return and the risk free rate; (3) FF3, corresponding to APT S–SDFs pricing exactly the three Fama-French factor returns and the risk-free rate; (4) 1PC, corresponding to APT S–SDFs pricing exactly the first principal component of excess returns and the risk-free rate. All calculations are for APT S–SDFs based on the standard l_2-pricing error metric. Both the low and intermediate dimensional datasets run from July 1963 to June 2018,

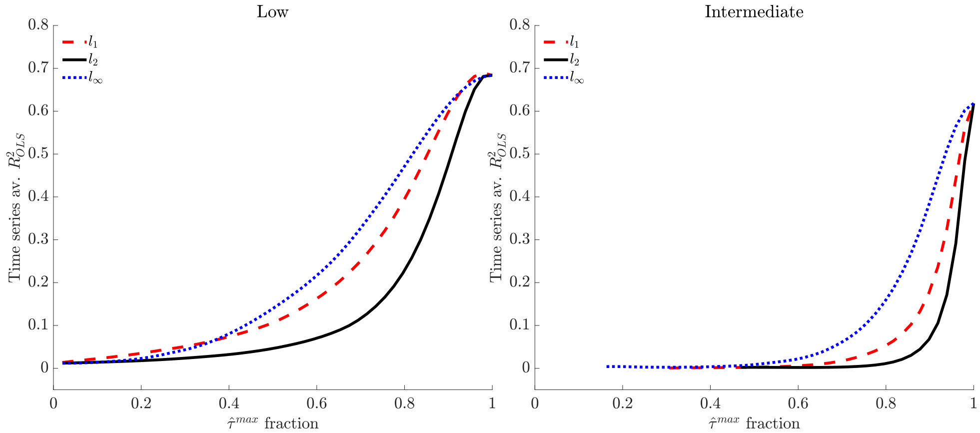


(1) **Panel (A)**

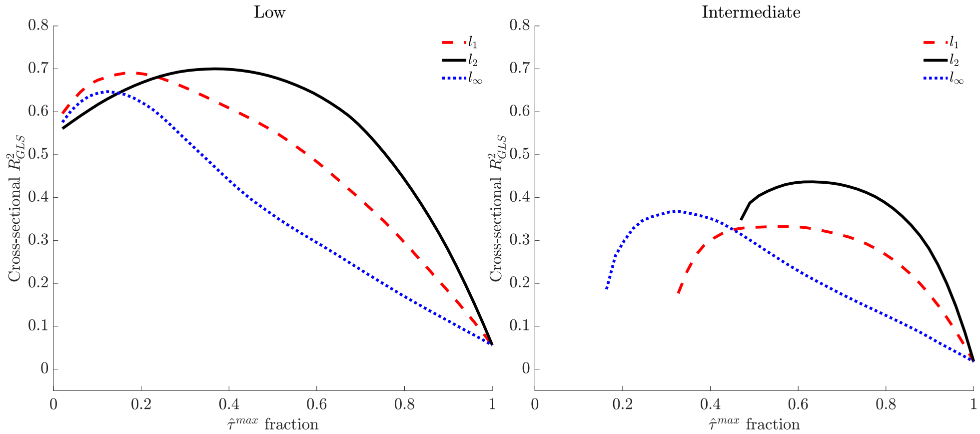


(2) **Panel (B)**

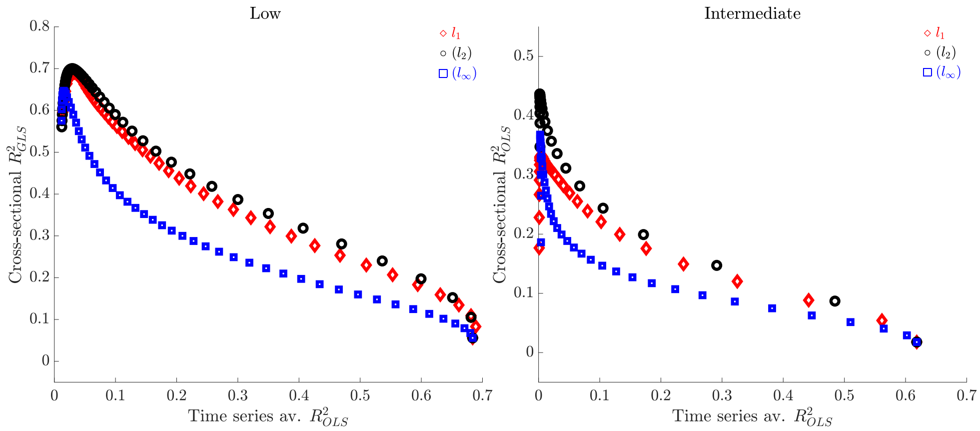
Figure 7: **Tradeoff between time-series and cross-sectional S–SDF explanatory power of APT S–SDFs, non APT S–SDFs, and principal components based S–SDFs.** For various minimum variance S–SDFs, the figure reports the tradeoff between average time series R^2 metric (29) and cross-sectional GLS R^2 metric (28) (denoted by R^2_{GLS}, left panels) or cross-sectional R^2 metric (31) (denoted by R^2_{OLS}, right panels). Following S–SDFs are compared in Panel (A). First, APT S–SDFs based on the l_2-pricing error metric. Second, non APT S–SDFs based on the l_2-pricing error metric, but with a diagonal weighting matrix $\mathbf{\Sigma}_{\boldsymbol{\eta}} = \boldsymbol{I}_{N_D \times N_D}$. Both these S–SDFs incorporate as sure assets the risk-free rate and the first principal component of excess returns. In each figure of Panel (B), results for two additional principal components based S–SDFs are reported, which incorporate only the risk-free rate as a sure asset and impose the corresponding pricing error bound on the principal components of sorted portfolio excess returns, instead of the original excess returns. The first and second of these S-SDFs impose in the left (right) figure an APT l_2- and $l_\infty-$ pricing error bound (a non APT l_2- and $l_\infty-$ pricing error bound based on a diagonal weighting matrix $\mathbf{\Sigma}_{\boldsymbol{\eta}} = \boldsymbol{I}_{N_D \times N_D}$), respectively. All results are based on the high dimensional dataset running from 1963-2018.



(1) **Panel (A)**

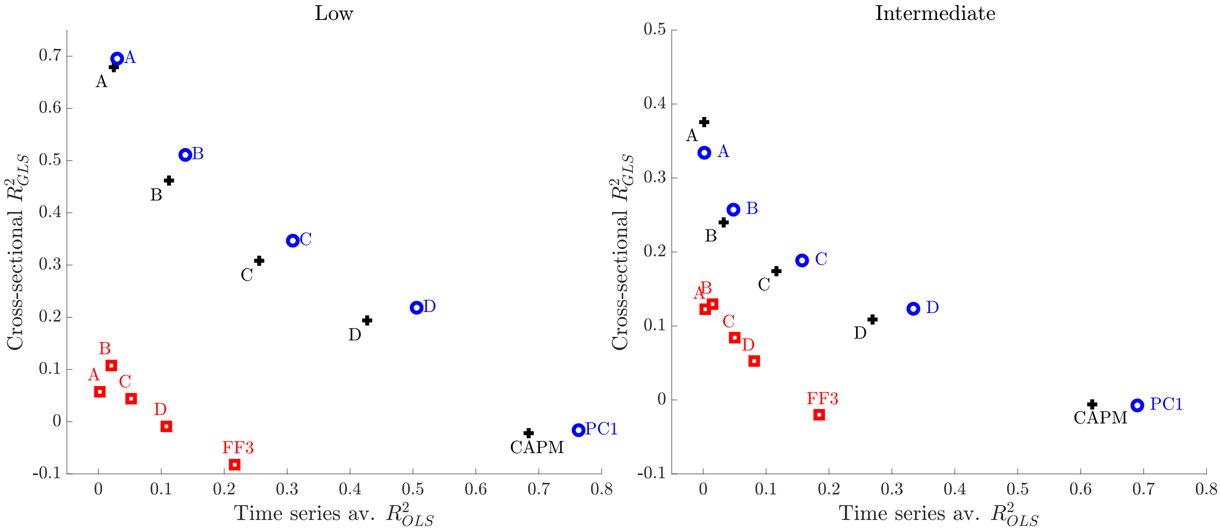


(2) **Panel (B)**

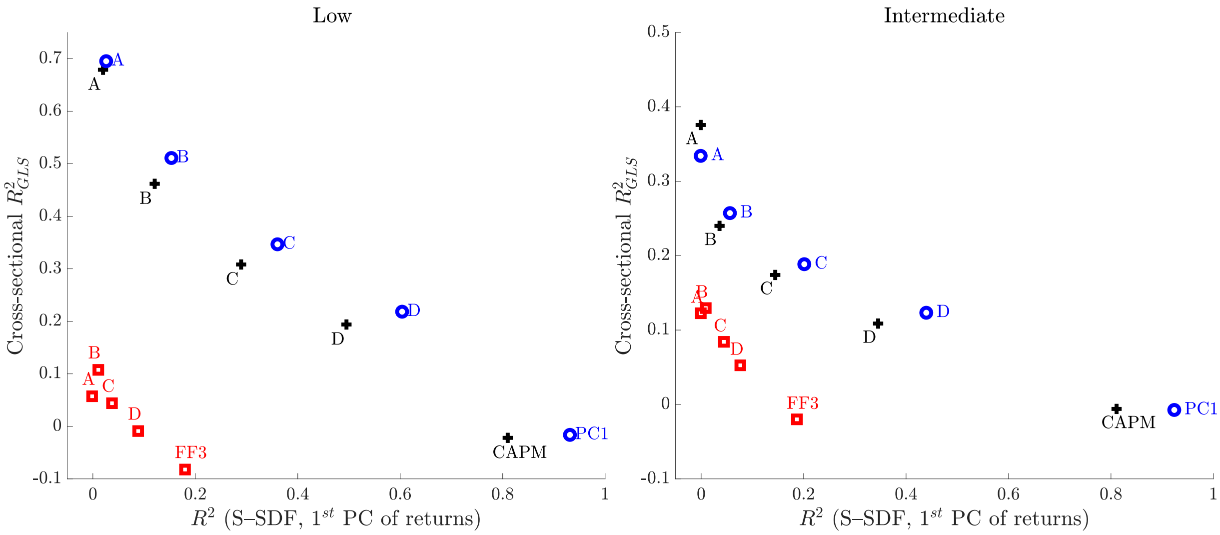


(3) **Panel (C)**

Figure 8: **Out-of-sample tradeoff between time-series and cross-sectional APT S–SDF explanatory power.** The figure reports the out-of-sample average time-series R^2 metric (29) (top panels) and the out-of-sample cross-sectional GLS R^2 metric (28) (middle panels) for the sequence of minimum variance APT S–SDFs from Section 3.3.1. These S–SDFs are estimated on rolling windows of 30 years under a constant relative threshold $\overline{\sigma} \in [\max_y\{\hat{\sigma}_y^{min}\}, 1]$ and based on a l_1-, l_2-, and (scaled) $l_\infty-$pricing error metric, respectively. In each estimation window, S–SDFs incorporate the risk-free asset and the market as the only sure assets, while all sorted portfolios are treated as dubious assets. Bottom panels directly report the tradeoff between out-of-sample average time series R^2 metric (29) and out-of-sample cross-sectional GLS R^2 metric (28). Results are based on the low dimensional (left panels) and the intermediate dimensional (right panels) datasets.



(1) **Panel (A)**



(2) **Panel (B)**

Figure 9: **Out-of-sample tradeoff between time-series and cross-sectional explanatory power of optimal data-driven APT S–SDFs.** The figure reports in **Panel (A)** (**Panel (B)**) the tradeoff between out-of-sample average time-series R^2 metric (29) (out-of-sample time-series R^2 metric (29) with respect to the first principal component of excess returns) and out-of-sample cross-sectional GLS R^2 metric (28), for the sequences of optimal data-driven minimum variance APT S–SDFs with l_2-pricing error metric from Section 3.3.2. These S–SDFs are estimated on rolling windows of 30 years by maximizing the in-sample cross-sectional metric $R^2_{GLS}(\hat{\sigma}_y)$ in constrained optimization problem (32), for parameter choices $p = 0$ (label A), $p = 0.25$ (label B), $p = 0.5$ (label C), $p = 0.75$ (label D), and $p = 1$. We consider APT S–SDFs based on three different choices for the set of sure asset returns, in addition to the risk-free rate: (1) the market excess return, (2) the three Fama-French factor excess returns, and (3) the first principal component of sorted portfolio excess returns. Label CAPM corresponds to case (1) for $p = 1$, label FF3 to case (2) for $p = 1$ and label PC1 to case (3) for $p = 1$. In each case, all sorted portfolios are treated as dubious assets. Results are based on the low dimensional (left panels) and the intermediate dimensional (right panels) datasets.

Table 1: **Pricing error constraints and their supporting cost functions**

Pricing error and supporting cost function	h	σ_h
Conic	δ_K	δ_{K^o}
Norm	$\lVert\cdot\rVert$	$\tau \lVert\cdot\rVert_*$
Self-standardized norm	$\lVert\cdot\rVert_{\boldsymbol{W}^{-1}}$	$\tau \lVert\cdot\rVert_{*,\boldsymbol{W}}$
Convex combination of norms	$(1-\lambda)\lVert\cdot\rVert^{(1)} + \lambda \lVert\cdot\rVert^{(2)}$	$\tau \left(\min_{\boldsymbol{\eta}\in\mathbb{R}^{N_D}} \left\{\max\left\{\frac{\lVert\cdot\rVert_*^{(1)}}{1-\lambda}, \frac{\lVert\cdot-\boldsymbol{\eta}\rVert_*^{(2)}}{\lambda}\right\}\right\}\right)$
Euclidean distance from norm ball	$\mathrm{dist}_{\alpha B(\lVert\cdot\rVert)}$	$\alpha \lVert\cdot\rVert_* + \tau \lVert\cdot\rVert_2$
Short-sale constraint	$\delta_{\mathbb{R}_+^{N_D}}$	$\delta_{\mathbb{R}_+^{N_D}}$
l_2-norm	$\lVert\cdot\rVert_2$	$\tau \lVert\cdot\rVert_2$
l_1-norm	$\lVert\cdot\rVert_1$	$\tau \lVert\cdot\rVert_\infty$
Self-standardized $l_\infty-$norm	$\lVert\cdot\rVert_{\infty,\boldsymbol{W}^{-1}}$	$\tau \lVert\cdot\rVert_{1,\boldsymbol{W}}$
Convex combination of l_1- and l_2-norms	$(1-\lambda)\lVert\cdot\rVert_1 + \lambda \lVert\cdot\rVert_2$	$\tau \left(\min_{\boldsymbol{\eta}\in\mathbb{R}^{N_D}} \left\{\max\left\{\frac{\lVert\cdot\rVert_\infty}{1-\lambda}, \frac{\lVert\cdot-\boldsymbol{\eta}\rVert_2}{\lambda}\right\}\right\}\right)$
Distance from $l_\infty-$ball	$\mathrm{dist}_{\alpha B_\infty}$	$\alpha \lVert\cdot\rVert_1 + \tau \lVert\cdot\rVert_2$

The table collects various specifications of pricing error functions h and associated supporting cost functions σ_h, along with closed-form examples. From top to bottom, we consider: (i) conic pricing constraints obtained via characteristic functions of convex cone K and its polar cone K^o; for instance, short-sale constraints feature $K = K^o = \mathbb{R}_+^{N_D}$ and bid-ask spreads feature $K = K^o = \mathbb{R}_+^{N_D} \times \mathbb{R}_-^{N_D}$. (ii) norm constraints obtained via $h = \lVert\cdot\rVert$ with $\sigma_h = \tau \lVert\cdot\rVert_*$, where $\lVert\cdot\rVert_*$ is the dual norm of $\lVert\cdot\rVert$; for instance, the l_2-norm is self-dual, while the l_1- and $l_\infty-$norms are dual to each other. (iii) self-standardized norms defined via $\lVert\cdot\rVert_{\boldsymbol{W}^{-1}} := \lVert \boldsymbol{W}^{-1}\cdot\rVert$. (iv) Euclidean distance from a ball of radius α under norm $\lVert\cdot\rVert$ (i.e., $\alpha B(\lVert\cdot\rVert) := \{\boldsymbol{\eta} \in \mathbb{R}^{N_D} : \lVert\boldsymbol{\eta}\rVert \leq \alpha\}$) defined via $\mathrm{dist}_{\alpha B(\lVert\cdot\rVert)} := \inf_{\boldsymbol{\eta}\in\mathbb{R}^{N_D}} \{\lVert\cdot - \boldsymbol{\eta}\rVert_2 : \lVert\boldsymbol{\eta}\rVert \leq \alpha\}$. In the table, $\tau \geq 0$, $\lambda \in [0,1]$ and $\boldsymbol{W}$ denotes a generic symmetric positive definite weighting matrix.

Appendix B - Proofs

Proof of Proposition 1. By [Rockafellar, 1970, Thm. 13.2] σ_h is proper, closed and sublinear.[49] This implies that the pricing function π is sublinear and that the set of payoffs, $\mathcal{Z}$ is a convex cone. Hence, by [Chen, 2001, Thm. 1, 5 and Cor. 1], absence of free lunches is equivalent to the existence of an almost surely strictly positive SDF, M, such that $\mathbb{E}[MZ] \leq \pi(Z)$ for any $Z \in \mathcal{Z}$.[50]

[49] Properness follows from the implicit assumption that the set of admissible pricing errors induced by function h and threshold τ is not empty, i.e., that there exists $\boldsymbol{\eta} \in \mathbb{R}^{N_D}$ such that $h(\boldsymbol{\eta}) \leq \tau$.

[50] Technically, in order to use [Chen, 2001, Thm. 5], we require the regulatory assumption that there is some payoff in the payoff space that is strictly positive almost surely. Such assumption is easily satisfied when, for e.g., there is a risk-free asset with positive payoff.

Now, by definition of cost functional π and that of the payoff space $\mathcal{Z}$, it equivalently follows for any $\boldsymbol{\theta} \in \mathbb{R}^N$ that $\boldsymbol{\theta}'\boldsymbol{P} + \sigma_h(\boldsymbol{\theta}_D) \geq \mathbb{E}[M\boldsymbol{\theta}'\boldsymbol{X}]$, where σ_h is given in equation (6), i.e., for any $\boldsymbol{\theta} \in \mathbb{R}^N$:

$$\boldsymbol{\theta}'_S\boldsymbol{P}_S + \boldsymbol{\theta}'_D\boldsymbol{P}_D + \sigma_h(\boldsymbol{\theta}_D) \geq \boldsymbol{\theta}'_S\mathbb{E}[M\boldsymbol{X}_S] + \boldsymbol{\theta}'_D\mathbb{E}[M\boldsymbol{X}_D] \ . \tag{B-1}$$

Inequality (B-1) is equivalent to the following two inequalities: $\boldsymbol{\theta}'_S\boldsymbol{P}_S \geq \boldsymbol{\theta}'_S\mathbb{E}[M\boldsymbol{X}_S]$ for any $\boldsymbol{\theta}_S \in \mathbb{R}^{N_S}$ and $\boldsymbol{\theta}'_D(\mathbb{E}[M\boldsymbol{X}_D] - \boldsymbol{P}_D) \leq \sigma_h(\boldsymbol{\theta}_D)$ for any $\boldsymbol{\theta}_D \in \mathbb{R}^{N_D}$. The first inequality holds if and only if $\mathbb{E}[M\boldsymbol{X}_S] - \boldsymbol{P}_S = \mathbf{0}$ and the second inequality, by [Bauschke and Combettes, 2011, Prop. 13.10 (i)], if and only if

$$(\sigma_h)^*(\mathbb{E}[M\boldsymbol{X}_D] - \boldsymbol{P}_D) \leq 0 \ . \tag{B-2}$$

By [Rockafellar, 1970, Thm. 13.2], the convex conjugate $(\sigma_h)^*$ is given by the characteristic function δ_C, where $C := \{\boldsymbol{y} : \ h(\boldsymbol{y}) \leq \tau\}$, i.e.,

$$(\sigma_h)^*(\mathbb{E}[M\boldsymbol{X}_D] - \boldsymbol{P}_D) = \begin{cases} 0 & h(\mathbb{E}[M\boldsymbol{X}_D] - \boldsymbol{P}_D) \leq \tau \\ +\infty & \text{else} \end{cases} .$$

Thus, inequality (B-2) is equivalent to condition $h(\mathbb{E}[M\boldsymbol{X}_D] - \boldsymbol{P}_D) \leq \tau$. This concludes the proof. □

Proof of Proposition 2. We first prove that $\Pi(\tau) = -\Delta(\tau)$. To do so, we rewrite problem $\Pi(\tau)$ according to the notation given in [Borwein and Lewis, 1992, Sec. 4] and employ Fenchel's duality to obtain its dual problem. Define for any stochastic discount factor M the linear function $A : L^q \to \mathbb{R}^N$ by $A(M) := \mathbb{E}[M\boldsymbol{X}]$ and let $R := \{\boldsymbol{P}_S\} \times \{\boldsymbol{P}_D + C\}$, where $C := \{\boldsymbol{\eta}_D \in \mathbb{R}^{N_D} \ : \ h(\boldsymbol{\eta}_D) \leq \tau\}$. With the notation $I_f := \mathbb{E}[f(\cdot)]$ for any function $f : \mathbb{R} \to [-\infty, +\infty]$, it follows that $I_{\phi_+}(M) = I_\phi(M) + \delta_{L^q_+}(M)$.[51] Thus we can write:

$$\Pi(\tau) = \inf_{M \in L^q} \{I_{\phi_+}(M) \ : \ A(M) \in R\} \ .$$

As payoffs are in L^p with $1/p + 1/q = 1$, A is continuous, while the properties of ϕ_+ and h imply that I_{ϕ_+} is a closed convex function and that R is a convex set. Thus, in order to apply [Borwein and Lewis, 1992, Cor. 4.6] and obtain the dual problem of Π we need to check that:[52]

$$A[\mathrm{qri}(\mathrm{dom}\, I_{\phi_+})] \cap \mathrm{ri}(R) \neq \emptyset \ . \tag{B-3}$$

As shown in [Borwein and Lewis, 1991, Cor. 2.6], our requirement $(0, +\infty) \subset \mathrm{dom}\, \phi$ implies that $A(\mathrm{dom}\, I_\phi \cap L^q_{++}) \subset A[\mathrm{qri}(\mathrm{dom}\, I_{\phi_+})]$. Hence, showing that

$$A(\mathrm{dom}\, I_\phi \cap L^q_{++}) \cap \mathrm{ri}(R) \neq \emptyset$$

is enough to prove (B-3). Further, notice that $\mathrm{ri}(R) = \{\boldsymbol{P}_S\} \times (\{\boldsymbol{P}_D\} + \mathrm{ri}(C))$ and, by [Rockafellar, 1970, Thm. 7.6], that $\mathrm{ri}(C) = \mathrm{ri}(\mathrm{dom}\, h) \cap C_0$, where $C_0 := \{\boldsymbol{\eta} \in \mathbb{R}^{N_D} : h(\boldsymbol{\eta}) < \tau\}$. Thus the result follows from absence of arbitrage in a supporting economy with parameter $\tilde{\tau} < \tau$. Indeed, by Proposition 1, there exists $\tilde{M} \in \mathrm{dom}\, I_\phi \cap L^q_{++}$, hence $A(\tilde{M}) \in A(\mathrm{dom}\, I_\phi \cap L^q_{++})$.[53] Moreover, $\boldsymbol{P}_S = \mathbb{E}[\tilde{M}\boldsymbol{X}_S]$ and $\mathbb{E}[\tilde{M}\boldsymbol{X}_D] \in \boldsymbol{P}_D + C_0$. Therefore, under the assumption $\mathrm{ri}(\mathrm{dom}\, h) = \mathrm{dom}(h)$, condition (B-3) is satisfied.[54]

Thus, by [Borwein and Lewis, 1992, Cor. 4.6] we obtain

$$\Pi(\tau) = \max_{\boldsymbol{\theta} \in \mathbb{R}^N} \{-I^*_{\phi_+}({}^tA(\boldsymbol{\theta})) - \delta^*_R(-\boldsymbol{\theta})\} \ , \tag{B-4}$$

where $I^*_{\phi_+} : L^p \to [0, +\infty]$ is the conjugate function of I_{ϕ_+} and ${}^tA : \mathbb{R}^N \to L^p$ is the adjoint map of A,

[51] Remember that ϕ_+ is the restriction of ϕ to $\mathbb{R}_+$, i.e., $\phi_+(x) = \phi(x)$ if $x \geq 0$ and $\phi_+(x) = +\infty$ if $x < 0$.

[52] See [Borwein and Lewis, 1992, Def. 2.3] for the definition of relative interior, ri, and quasi relative interior, qri.

[53] Again, the implicit assumption behind primal problem 8 is that the $L^p - L^q$ dual spaces are chosen so that the integrability condition $\mathbb{E}[\phi(M)] < +\infty$ is satisfied.

[54] Such assumption is satisfied for all pricing error functions h in Table 1. For instance, any function h whose domain is $\mathbb{R}^{N_D}$ or whose domain is a finitely-generated cone satisfies this assumption.

given by ${}^tA(\boldsymbol{\theta}) = \boldsymbol{X}'\boldsymbol{\theta}$.[55] As ϕ_+ is convex and closed, we can apply [Rockafellar, 1968, Thm. 2] to obtain $I^*_{\phi_+} = I_{\phi^*_+}$. Moreover, for every $\boldsymbol{\theta} \in \mathbb{R}^N$,

$$\begin{aligned}
\delta^*_R(-\boldsymbol{\theta}) &= \delta^*_{\{\boldsymbol{P}_S\}}(-\boldsymbol{\theta}_S) + \delta^*_{\{\boldsymbol{P}_D + C\}}(-\boldsymbol{\theta}_D) \\
&= \sup_{\boldsymbol{\eta}_S \in \mathbb{R}^{N_S}} \{-\boldsymbol{\theta}'_S \boldsymbol{\eta}_S \ : \ \boldsymbol{\eta}_S = \boldsymbol{P}_S\} + \sup_{\boldsymbol{\eta}_D \in \mathbb{R}^{N_D}} \{-\boldsymbol{\theta}'_D(\boldsymbol{P}_D + \boldsymbol{\eta}_D) \ : \ h(\boldsymbol{\eta}_D) \leq \tau\} \\
&= -\boldsymbol{P}'\boldsymbol{\theta} + \sigma_h(-\boldsymbol{\theta}_D) \ .
\end{aligned}$$

Thus, after a change of variable, (B-4) reads:

$$\Pi(\tau) = -\min_{\boldsymbol{\theta} \in \mathbb{R}^N} \left\{ \mathbb{E}[\phi^*_+(-\boldsymbol{X}'\boldsymbol{\theta})] + \boldsymbol{P}'\boldsymbol{\theta} + \sigma_h(\boldsymbol{\theta}_D) \right\} \ ,$$

thereby proving strong duality between $\Pi(\tau)$ and $-\Delta(\tau)$.

We now prove the relation between the optimal solutions of $\Pi(\tau)$ and $\Delta(\tau)$ given in (12). With change of variable in equation (B-4), the dual problem can be written as

$$\Delta(\tau) = \min_{\boldsymbol{\theta} \in \mathbb{R}^N} \{ I_{\phi^*_+}({}^tA(-\boldsymbol{\theta})) + \delta^*_R(\boldsymbol{\theta}) \} \ . \tag{B-5}$$

Since ϕ_+ is strictly convex on its domain, assuming that $-\boldsymbol{X}'\boldsymbol{\theta}_0$ is almost surely an element of the interior of the domain of ϕ^*_+, by [Borwein and Lewis, 1991, Thm. 4.6] ϕ^*_+ is differentiable a.s. in $-\boldsymbol{X}'\boldsymbol{\theta}_0 =^t A(-\boldsymbol{\theta}_0)$.[56] Let us denote such derivative by $M_0(\omega) := (\phi^*_+)'(-\boldsymbol{X}'(\omega)\boldsymbol{\theta}_0)$. Below, we will use the fact that M_0 is the unique element of $\partial I_{\phi^*_+}({}^tA(-\boldsymbol{\theta}_0)) \subset L^q$. To show this, we first claim that for any $\bar{M} \in \partial I_{\phi^*_+}({}^tA(-\boldsymbol{\theta}_0))$ we must have $\bar{M}(\omega) \in \partial \phi^*_+({}^tA(-\boldsymbol{\theta}_0)(\omega)) \subset \mathbb{R}$ a.s.. If this is true, the fact that ϕ^*_+ is differentiable in $-X(\omega)'\boldsymbol{\theta}_0$ a.s., that is, the only element of its subdifferential is given by $M_0(\omega)$, implies that $\bar{M} = M_0$ almost surely, i.e., M_0 is the unique element of $\partial I_{\phi^*_+}({}^tA(-\boldsymbol{\theta}_0))$. To show uniqueness, we start from following identity, which holds by [Rockafellar, 1970, Thm. 23.5]:

$$I_{\phi^*_+}({}^tA(-\boldsymbol{\theta}_0)) + I_{\phi_+}(\bar{M}) = \langle {}^tA(-\boldsymbol{\theta}_0), \bar{M} \rangle \ .$$

Explicitly, this gives $\int \left[\phi^*_+({}^tA(-\boldsymbol{\theta}_0)(\omega)) + \phi_+(\bar{M}(\omega)) -^t A(-\boldsymbol{\theta}_0)(\omega)\bar{M}(\omega) \right] d\mathbb{P}(\omega) = 0$. Here, since by Fenchel's inequality the integrand is nonnegative, it is zero a.s.. Applying again [Rockafellar, 1970, Thm. 23.5] to this integrand, we obtain $\bar{M}(\omega) \in \partial \phi^*_+({}^tA(-\boldsymbol{\theta}_0)(\omega))$ a.s., hence $M_0 = \bar{M} \in L^q$, as claimed.

To show the primal feasibility of M_0, i.e., that M_0 satisfies the pricing condition (1) and that $M_0 \geq 0$, notice first that since $\boldsymbol{\theta}_0$ is a solution to problem (B-5), if we denote with Q the dual objective function, the following first order condition holds:

$$0 \in \partial Q(\boldsymbol{\theta}_0) = \partial I_{\phi^*_+}({}^tA(-\boldsymbol{\theta}_0)) + \partial \delta^*_R(\boldsymbol{\theta}_0) \ .$$

Moreover, by [Borwein, 1981, Thm. 4.1], $\partial_{\boldsymbol{\theta}} I_{\phi^*_+}({}^tA(-\boldsymbol{\theta}_0)) = -{}^{tt}A(\partial I_{\phi^*_+}({}^tA(-\boldsymbol{\theta}_0))$, which is given by the singleton $\{-A(M_0)\}$ since ${}^{tt}A|_{L^q} = A$. Hence, there exists $\boldsymbol{\nu} \in \partial \delta^*_R(\boldsymbol{\theta}_0) \subseteq \mathbb{R}^N$, such that

$$-A(M_0) + \boldsymbol{\nu} = 0 \ . \tag{B-6}$$

Equation (B-6), together with the fact that $\boldsymbol{\nu} \in \partial \delta^*_R(\boldsymbol{\theta}_0) \subset \mathbb{R}^N$ and $M_0 \in \partial I_{\phi^*_+}({}^tA(\boldsymbol{\theta}_0))$, implies the primal feasibility of M_0. Indeed, by [Rockafellar, 1970, Thm. 23.5], $\boldsymbol{\nu} \in \partial \delta^*_R(\boldsymbol{\theta}_0) \subset \mathbb{R}^N$ implies $\delta_R(\boldsymbol{\nu}) = \boldsymbol{\nu}'\boldsymbol{\theta}_0 - \delta^*_R(\boldsymbol{\theta}_0)$. Since δ^*_R is proper and by definition of a subgradient, we must thus have $\delta^*_R(\boldsymbol{\theta}_0) < +\infty$. Hence, $\delta_R(\boldsymbol{\nu})$ is finite, which from equation (B-6) means that also $\delta_R(A(M_0))$ is finite. Using the explicit definition of δ_R, this means $\mathbb{E}[M\boldsymbol{X}_S] - \boldsymbol{P}_S = \boldsymbol{0}_S$ and $h(\mathbb{E}[M\boldsymbol{X}_D] - \boldsymbol{P}_D) \leq \tau$. Similarly, using again [Rockafellar, 1970,

[55] The adjoint map of A, ${}^tA$, is characterized by the identity $\mathbb{E}[{}^tA(\boldsymbol{\theta})M] = \boldsymbol{\theta}'\mathbb{E}[M\boldsymbol{X}]$, for each $M \in L^q$ and each portfolio weights $\boldsymbol{\theta} \in \mathbb{R}^N$.

[56] The condition that $-\boldsymbol{X}'\boldsymbol{\theta}_0$ is in the interior of the domain of ϕ^*_+ a.s. can be ensured by requiring that $-\boldsymbol{X}'\boldsymbol{\theta}_0 < d := \lim_{y\to\infty} \phi(y)/y$ a.s., see [Borwein and Lewis, 1991, Lem. 4.2]. This condition holds for all the Cressie-Read dispersions listed in Online Appendix A, see Borwein and Lewis [1991] for more details.

Thm. 23.5], the fact that $M_0 \in \partial I_{\phi_+^*}({}^t A(-\boldsymbol{\theta}_0))$, the definition of subgradient and properness of $I_{\phi_+^*}$ imply $M_0 \in \operatorname{dom} I_{\phi_+}$, so that $M_0 \in L_+^q$ is indeed primal feasible.

We finally show that M_0 is a primal solution, i.e. $I_{\phi_+}(M_0) + \delta_R(A(M_0)) = \Pi(\tau)$, To this end, consider the following equalities:

$$I_{\phi_+^*}({}^t A(-\boldsymbol{\theta}_0)) + I_{\phi_+}(M_0) = \langle {}^t A(-\boldsymbol{\theta}_0), M_0\rangle = \langle -\boldsymbol{\theta}_0, A(M_0)\rangle = -\boldsymbol{\nu}'\boldsymbol{\theta}_0 \ ,$$

where the first equality follows again from [Rockafellar, 1970, Thm. 23.5], the second one from the definition of the adjoint map, and the third one from optimality condition (B-6). Thus, we can explicitly write the primal objective function computed in M_0 as:

$$I_{\phi_+}(M_0) + \delta_R(A(M_0)) = -I_{\phi_+^*}({}^t A(-\boldsymbol{\theta}_0)) - \delta_R^*(\boldsymbol{\theta}_0) + [-\boldsymbol{\nu}'\boldsymbol{\theta}_0 + \delta_R^*(\boldsymbol{\theta}_0) + \delta_R(A(M_0))] \ . \tag{B-7}$$

Using equation (B-6), $-\boldsymbol{\nu}'\boldsymbol{\theta}_0 + \delta_R^*(\boldsymbol{\theta}_0) + \delta_R(A(M_0)) = -A(M_0)'\boldsymbol{\theta}_0 + \delta_R^*(\boldsymbol{\theta}_0) + \delta_R(A(M_0))$, which is zero by [Rockafellar, 1970, Thm. 23.5]. This shows that the RHS of equation (B-7) is equal to $-\Delta(\tau)$, hence that M_0 is a primal solution. Uniqueness of this solution can be shown following [Borwein and Lewis, 1991, Prop. 2.11]. This concludes the proof. $\square$

Proposition 4 (Equivalence with dispersion constrained minimum pricing error problems). *Suppose that $\Pi(\tau)$ with $\tau \geq 0$ and*

$$\mathcal{P}(\nu) := \inf_{M\in\mathcal{M}} \{h(\mathbb{E}[M\boldsymbol{X}_D - \boldsymbol{P}_D]) \ : \ \mathbb{E}[\phi(M)] \leq \nu\} \tag{B-8}$$

with $\nu \geq 0$ are finite and attained. (i) If M_0 solves $\Pi(\tau)$, then there exists $\nu_0 \geq 0$ such that M_0 solves $\mathcal{P}(\nu_0)$. (ii) If M_0 is the unique solution of $\mathcal{P}(\nu)$, then there exists $\tau_0 \geq 0$ such that M_0 solves $\Pi(\tau_0)$.

Proof. (i) Let $\nu_0 := \mathbb{E}[\phi(M_0)]$. By strict convexity of ϕ and [Borwein and Lewis, 1991, Prop. 2.11], M_0 is the unique solution of $\Pi(\tau)$. Therefore, M_0 is the unique element of $\{M \in \mathcal{M} \ : \ \mathbb{E}[\phi(M)] \leq \nu_0, \ h(\mathbb{E}[M\boldsymbol{X}_D - \boldsymbol{P}_D]) \leq \tau\}$. Thus, M_0 solves $\mathcal{P}(\nu_0)$. (ii) Let $\tau_0 := h(\mathbb{E}[M_0\boldsymbol{X}_D - \boldsymbol{P}_D])$. If M_0 is the unique solution of $\mathcal{P}(\nu)$, it is the unique element of $\{M \in \mathcal{M} \ : \ \mathbb{E}[\phi(M)] \leq \nu, \ h(\mathbb{E}[M\boldsymbol{X}_D - \boldsymbol{P}_D]) \leq \tau_0\}$. Thus, M_0 solves $\Pi(\tau_0)$. $\square$

Lemma 1 (Closed-form penalizations for pricing error norms in equations (25) and (26))**.** *Consider norms h_1 and h_∞ in equations (25) and (26), respectively. The corresponding penalizations $\sigma_{h_i}(\boldsymbol{\theta}_D) = \tau h_{i*,\boldsymbol{\Sigma}_\zeta^{1/2}}(\boldsymbol{\theta}_D)$ from Proposition 3 follow in closed-form as:*

(i) When $\lambda = 1$:

$$h_{1*}(\boldsymbol{\theta}_D) = h_{\infty*}(\boldsymbol{\theta}_D) = \|\boldsymbol{\theta}_D\|_2 \ .$$

(ii) When $\lambda = 0$:

$$h_{1*}(\boldsymbol{\theta}_D) = \|\boldsymbol{\theta}_D\|_\infty \ ; \ h_{\infty*}(\boldsymbol{\theta}_D) = \|\boldsymbol{\theta}_D\|_1 / \sqrt{N_D} \ .$$

(iii) When $\lambda \in (0,1)$:

$$h_{1*}(\boldsymbol{\theta}_D) = \min_{\boldsymbol{z}\in\mathbb{R}^{N_D}} \max\left\{\frac{\|\boldsymbol{\theta}_D\|_\infty}{1-\lambda}, \frac{\|\boldsymbol{\theta}_D - \boldsymbol{z}\|_2}{\lambda}\right\} \ ,$$

$$h_{\infty*}(\boldsymbol{\theta}_D) = \min_{\boldsymbol{z}\in\mathbb{R}^{N_D}} \max\left\{\frac{\|\boldsymbol{\theta}_D\|_1}{\sqrt{N_D}(1-\lambda)}, \frac{\|\boldsymbol{\theta}_D - \boldsymbol{z}\|_2}{\lambda}\right\} \ .$$

References

Caio Almeida and René Garcia. Economic implications of nonlinear pricing kernels. *Management Science*, 63(10):3361–3380, 2016.

David Backus, Mikhail Chernov, and Stanley Zin. Sources of entropy in representative agent models. *Journal of Finance*, 64(1):51–99, 2014.

Ravi Bansal and Bruce N. Lehmann. Growth-optimal portfolio restrictions on asset pricing models. *Macroeconomic Dynamics*, 108(1):333–354, 1997.

Heinz H Bauschke and Patrick L Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.

Jonathan M Borwein. A lagrange multiplier theorem and a sandwich theorem for convex relations. *Mathematica Scandinavica*, 48:189–204, 1981.

Jonathan M Borwein. On the failure of maximum entropy reconstruction for fredholm equations and other infinite systems. *Mathematical programming*, 61(1-3):251–261, 1993.

Jonathan M Borwein and Adrian S Lewis. Duality relationships for entropy-like minimization problems. *SIAM Journal on Control and Optimization*, 29(2):325–338, 1991.

Jonathan M Borwein and Adrian S Lewis. Partially finite convex programming, part i: Quasi relative interiors and duality theory. *Mathematical Programming*, 57(1-3):15–48, 1992.

Gary Chamberlain. Funds, factors and diversification in arbitrage pricing models. *Econometrica*, 51(5): 1305–1324, 1983.

Gary Chamberlain and Michael Rothschild. Arbitrage, factor structure and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304, 1983.

Luyang Chen, Markus Pelger, and Jason Zhu. Deep learning in asset pricing. Technical report, Working Paper, 2020a.

Xiaohong Chen, Lars P Hansen, and Peter G Hansen. Robust identification of investor beliefs. Technical report, 2020b.

Zhiwu Chen. Viable costs and equilibrium prices in frictional securities markets. *Annals of Economics and Finance*, 2(2):297–323, 2001.

Stephen A Clark. The valuation problem in arbitrage price theory. *Journal of Mathematical Economics*, 22 (5):463–478, 1993.

John H Cochrane and Jesus Saa-Requejo. Beyond arbitrage: Good-deal asset price bounds in incomplete markets. *Journal of Political Economy*, 108(1):79–119, 2000.

Noel Cressie and Timothy RC Read. Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 46(3):440–464, 1984.

Victor DeMiguel, Alberto Martín-Utrera, Francisco J Nogales, and Raman Uppal. A transaction-cost perspective on the multitude of firm characteristics. *Review of Financial Studies*, 33(5):2180–2222, 2020.

Larry Epstein and Martin Schneider. Ambiguity and asset markets. *Annual Review of Financial Economics*, 2:315–346, 2010.

Guanhao Feng, Stefano Giglio, and Dacheng Xiu. Taming the factor zoo: A test of new factors. *Journal of Finance*, 75(3):1327–1370, 2020.

Joachim Freyberger, Andreas Neuhierl, and Michael Weber. Dissecting characteristics nonparametrically. *The Review of Financial Studies*, 33(5):2326–2377, 2020.

Anisha Ghosh, Christian Julliard, and Alex P Taylor. An information-theoretic asset pricing model. Technical report, Working Paper, London School of Economics, 2016.

Shihao Gu, Brian Kelly, and Dacheng Xiu. Autoencoder asset pricing models. *Journal of Econometrics, forthcoming*, 2020a.

Shihao Gu, Bryan T Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5):2223–2273, 2020b.

Massimo Guidolin and Francesca Rinaldi. Ambiguity in asset pricing and portfolio choice: a review of the literature. *Theory and Decision*, 74:183–217, 2013.

Lars Peter Hansen and Ravi Jagannathan. Implications of security market data for models of dynamic economies. *Journal of Political Economy*, 99(2):225–262, 1991.

J Michael Harrison and David M Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic theory*, 20(3):381–408, 1979.

Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of convex analysis*. Springer Science & Business Media, 2012.

Gur Hubermann. A simple approach to the arbitrage pricing theory. *Journal of Economic Theory*, 28(1): 183–191, 1982.

Jonathan E Ingersoll. Some results in the theory of arbitrage pricing. *Journal of Finance*, 39(4):1021–1039, 1984.

Elyes Jouini and Hédi Kallal. Martingales and arbitrage in securities markets with transaction costs. *Journal of Economic Theory*, 66(1):178–197, 1995.

Sofonias A Korsaye, Alberto Quaini, and Fabio Trojani. Econometric evaluation of asset pricing models with smart sdfs. Technical report, University of Geneva and Swiss Finance Institute, 2019.

Serhiy Kozak, Stefan Nagel, and Shrihari Santosh. Shrinking the cross-section. *Journal of Financial Economics*, 135(2):271–292, 2020.

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

David G Luenberger. *Optimization by vector space methods.* John Wiley & Sons, 1997.

Erzo GJ Luttmer. Asset pricing in economies with frictions. *Econometrica*, 64(6):1439–1467, 1996.

Jean-Jaques Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *Comptes Rendus de l'Académie des Sciences*, pages 2897–2899, 1962.

Piotrek Orlowski, Andras Sali, and Fabio Trojani. Arbitrage free dispersion. Technical report, University of Lugano and Swiss Finance Institute, 2016.

Chen Qiu and Taisuke Otsu. Information theoretic approach to high dimensional multiplicative models: Stochastic discount factor and treatment effect. Technical report, London School of Economics, 2018.

Ralph T Rockafellar. Integrals which are convex functionals. *Pacific Journal of Mathematics*, 24(3):525–539, 1968.

Ralph T Rockafellar. *Convex analysis.* Princeton University Press, 1970.

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

Stephen Ross. Mutual fund separation in financial theory – the separating distributions. *Journal of Economic Theory*, 17(2):254–286, 1978.

Karl Snow. Diagnosing asset pricing models using the distribution of asset returns. *Journal of Finance*, 46 (3):955–983, 2011.

Michael Stutzer. A bayesian approach to diagnosis of asset pricing models. *Journal of Econometrics*, 68(2): 367–397, 1995.

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

Raman Uppal, Paolo Zaffaroni, and Irina Zviadadze. Correcting misspecified stochastic discount factors. Technical report, EDHEC Business School, HEC Paris and Imperial College, 2019.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67(2):301–320, 2005.