Smart Stochastic Discount Factors∗

Sofonias A. Korsaye† Alberto Quaini ‡ Fabio Trojani §

October 30, 2019

Abstract
We introduce model-free Smart Stochastic Discount Factors (S–SDFs) minimizing various notions of SDF dispersion, under general convex constraints on non zero pricing errors. S–SDFs can be naturally motivated by market frictions, asymptotic no-arbitrage conditions in an APT framework or a need for SDF regularization. More broadly, we show that they are always supported by a suitable viable economy with transaction costs. Minimum dispersion S–SDFs give rise to new nonparametric bounds for asset pricing models, under weaker assumptions on a model’s ability to price cross-sections of assets. They arise from a simple transformation of the optimal payoff in a dual penalized portfolio problem. We clarify the deep properties of S–SDFs induced by various economically motivated pricing error structures and develop a systematic tractable approach for their empirical analysis. For various APT settings, we demonstrate the improved out-of-sample pricing performance of minimum dispersion S–SDFs, which directly corresponds to highly profitable dual portfolio strategies.


First Version: December 2017

∗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, Olivier Scaillet, Paul Schneider, 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 and seminar participants at BI Norwegian Business School, Bocconi University, University of Zurich, University of Lugano and University of Geneva. All errors are ours.

†University of Geneva, Geneva Finance Research Institute & Swiss Finance Institute. Email: Sofonias.Korsaye@unige.ch
‡University of Geneva, Geneva Finance Research Institute. Email: Alberto.Quaini@unige.ch
§University of Geneva, Geneva Finance Research Institute & Swiss Finance Institute. Email: Fabio.Trojani@unige.ch
## Contents

1 Introduction 4

2 Theory of S–SDFs 7

2.1 Definition and existence of S–SDFs 8

2.2 Relevant examples of S–SDFs 11

2.3 Minimum dispersion S–SDFs 13

2.4 Modelling S–SDFs 18

2.5 Minimum dispersion S–SDFs and the APT 20

3 Estimation and inference for minimum dispersion S–SDFs 24

3.1 Consistency properties 24

3.2 Asymptotic distributions 25

3.2.1 Asymptotic distribution of estimators for S–SDF dispersion bounds 26

3.2.2 Asymptotic distribution of estimator for S–SDF dual portfolio weights 27

3.3 Testing S–SDF specifications using dispersion bounds 29

4 Empirical Analysis: S–SDFs in the APT Framework 31

4.1 Dataset 31

4.2 In–sample properties of APT–consistent minimum dispersion S–SDFs 32

4.3 Out-of-sample pricing properties of APT–consistent S–SDFs 35

4.3.1 S–SDF estimation setting and benchmark empirical asset pricing models 35

4.3.2 Out-of-sample pricing results for APT–consistent minimum variance S–SDFs 37

4.3.3 Out-of-sample pricing results for other APT–consistent minimum dispersion S–SDFs 37

4.4 Optimal trading strategies supporting APT–consistent S–SDFs 38

5 Conclusions 40

Appendices 42
1 Introduction

Hansen and Jagannathan [1991] provide a unified analysis and testing framework for asset pricing models under frictionless arbitrage-free markets, based on the minimum variance stochastic discount factor (SDF) that exactly prices a given set of securities. By construction, the minimum variance SDF provides both a lower bound on the variance of any admissible SDF and - by means of its dual portfolio characterization - an upper bound on the maximal attainable Sharpe ratio using linear portfolios of traded securities. In this way, it provides a useful reality check for any asset pricing model, in terms of both the required SDF variability and the attainable risk-return tradeoff. A large literature has built on these findings, by deriving the implications of more general notions of SDF dispersion for the admissible SDF variability and the attainable risk return tradeoff, giving rise to an extended family of model reality checks that can incorporate a concern for, e.g., higher-moment risk; see Backus et al. [2011], Juilliard and Ghosh [2012], Gosh et al. [2017] and Almeida and Garcia [2016], among others. In this paper, we extend this literature to general asset pricing settings where SDFs can incorporate widespread constraints on non-zero pricing errors. We call such SDFs Smart SDFs (S–SDFs).

Economically, S–SDFs naturally arise already in arbitrage-free economies with frictions that allow unbounded total wealth allocations to risky assets - such as proportional transaction costs, short-sale constraints or margin requirements - because of the sublinearity of the pricing functional in such settings; see, e.g, Luttmer [1996]. We first show that S–SDFs are more generally supported by viable markets with convex transaction costs that can directly constrain the total wealth allocation to risky assets, such as, e.g., quadratic transaction costs or constraints on total leverage. In such settings, we carefully embed total wealth penalizations in our S–SDF methodology, because they are essential to obtain well-behaved minimum dispersion S–SDFs also for markets where the standard SDF duality characterizations in Hansen and Jagannathan [1991], Almeida and Garcia [2016] and Luttmer [1996], among others, do not apply; see, e.g., Borwein [1993] for counter-examples. We further show that S–SDFs naturally arise also under asymptotic notions of no-arbitrage, which are typically assumed in Ross’s Arbitrage Pricing Theory (APT). In this context, we allow for flexible constraints on non zero pricing errors that can address also situations where the underlying factor model for returns may be misspecified; see, e.g., Uppal et al. [2018].

We define minimum dispersion S–SDFs as the solutions of general minimum SDF dispersion problems with flexible convex constraints on the pricing errors of a subset of assets. The generality of our setting allows
us to address with one coherent framework several specifications of frictions and total wealth constraints previously considered in the literature, various pricing error bounds implied by the APT, and more broadly any pricing error geometry induced by convex pricing error metrics. In parallel, our class of SDF dispersions is broad enough to embed a concern for second and higher-order risks, as measured, e.g., by notions of SDF variability induced by Cressie and Read [1984] power divergences; see also Kitamura et al. [2004] and Newey and Smith [2004], among others.

We formally link minimum dispersion S–SDFs to the solutions of dual portfolio selection problems with penalized portfolio weights, under a uniquely determined convex penalization function. This dual formulation explicitly measures the direct effect on optimal portfolio weights of, e.g., market frictions, of the type of portfolio weight regularization used to obtain a well-behaved S–SDF, of a possible violation of the Arbitrage Pricing Theory factor model assumptions, or a combination of these and other features. Equivalently, the one-to-one relation between the specification of convex pricing error constraints and the corresponding penalization on portfolio weights uniquely identifies the deep S–SDF properties that are induced, e.g., by a given market friction, a particular regularization choice or a specific form of misspecification in the APT.

The established duality between minimum dispersion S–SDFs and the solution of the corresponding penalized optimal portfolio problem allows us to develop a systematic approach also for the empirical analysis of minimum dispersion S–SDFs and their corresponding minimum dispersion bounds. In contrast to standard minimum dispersion SDFs, the portfolio optimization problem underlying minimum dispersion S–SDFs can imply a non smooth objective function, because its solution depends on a possibly nonsmooth penalization. In order to develop our empirical methodology for a broad class of convex pricing error specifications, we systematically make use of accurate approximations of portfolio weights penalties based on Moreau [1962] envelopes. This approach gives rise to smooth S–SDF estimation problems that are compatible with key S–SDF properties, such as pricing errors or portfolio weights sparsity, and are at the same time more tractable for applications.

In our empirical analysis, we systematically study minimum dispersion S–SDFs that are consistent with the pricing errors bounds induced by APT–like asymptotic no arbitrage conditions. We consider several such S–SDFs, using dispersion measures in the class of Cressie-Read power dispersions and several pricing error metrics that are compatible with various degrees of sparsity on pricing errors or portfolio weights.

We first show that pricing error sparsity is a key determinant of large S–SDF dispersions when pricing a
cross-section of asset returns without prior assumptions about exactly priced risk factors in the APT. In contrast, the trade-off between portfolio weight sparsity and S–SDF dispersion is less tight. This evidence indicates that S–SDFs with dense pricing errors may be better able to explain a given cross-section of asset returns, while targeting a given pricing error bound in the APT without taking a stand on which systematic risk factors are exactly priced.

Second, we explore the distinct properties of minimum dispersion S–SDFs induced by different notions of dispersion, including Hansen-Jagannathan distance, Kullback-Leibler, negative entropy and Hellinger dispersions. We find that these S–SDFs can be very different when no pricing error is admitted, due to the strict higher-moment effects in the empirical return distribution. However, differences tend to shrink substantially when introducing pricing errors that are consistent with the pricing errors bounds induced by APT-type no-arbitrage conditions. Hence, SDF regularizations induced by APT pricing error bounds make minimum dispersion S–SDFs less dependent on the dispersion criterion used.

We then study the out-of-sample pricing properties of minimum dispersion S–SDFs, relative to those of well-known linear SDF benchmarks in the literature that correspond to various APT–like factor model specifications. In doing so, we also systematically explore the implications of pricing error and portfolio weight sparsity for the out-of-sample pricing performance, based on GLS-adjusted $R^2$s in standard testing procedures for empirical asset pricing models. We find that Hansen-Jagannathan minimum variance S–SDFs with dense pricing errors and dense portfolio weights clearly outperform minimum variance S–SDFs imposing sparsity on pricing errors or portfolio weights. Moreover, the out-of-sample pricing performance of such S–SDFs dominates the one of minimum dispersion S–SDFs induced by notions of dispersion different from variance, such as negative entropy or Kullback-Leibler dispersion. We also find that the out-of-sample pricing performance of minimum variance S–SDFs is dramatically higher than the one of benchmark linear asset pricing models resulting from APT–like factor model specifications. For instance, we find that while the out-of-sample GLS-adjusted $R^2$ generated by benchmark three- and five-factor Fama-French models on an intermediate dimensional dataset with about 200 assets is only about 6%, the one generated by our data-driven minimum variance S–SDFs is about 42%.

The large out-of-sample pricing performance of Hansen-Jagannathan minimum variance S–SDFs with dense pricing errors and dense portfolio weights naturally corresponds to highly profitable optimal trading strategies. Indeed, in our out-of-sample period from July 1963 to June 2018, the unconditional annualized out-of-
sample Sharpe ratio of the portfolio supporting our minimum variance S–SDFs is about 1.4, while those of market capitalization-weighted and equally-weighted portfolio strategies are only 0.42 and 0.51, respectively. Importantly, the out-of-sample Sharpe ratios of data-driven minimum variance S–SDF portfolios are largely unrelated to the time-varying leverage in the underlying optimal S–SDF strategy. This feature indicates that the excellent performance of minimum variance S–SDF portfolios predominantly arises from their ability to successfully exploit the conditional mean-variance trade-offs, rather than from their dynamic timing features.

The remainder of the paper proceeds as follows. Section 2 presents our main theoretical findings. First, we formally define S–SDFs and prove their existence in arbitrage-free (viable) financial markets supported by corresponding sublinear (convex) transaction costs. Subsequently, we derive optimal model-free S–SDFs that minimize various established notions of statistical dispersion. To this end, we employ Fenchel duality to reduce the infinite dimensional optimization problem defining minimum dispersion S–SDFs to a corresponding finite dimensional penalized portfolio problem. We then fully characterize minimum dispersion S–SDFs as simple transformations of the optimal portfolio payoff resulting from the corresponding penalized dual portfolio problems. In between these theoretical results, we provide various relevant economic examples of S–SDFs, demonstrating how different friction, regularization and pricing error specifications in the literature are embedded into our approach, and we highlight their connections. Subsequently, we introduce our systematic modelling methodology based on Moreau envelopes to construct S–SDFs inducing tractable, i.e., smooth, dual portfolio problems, given any target pricing error specification with desirable properties such as sparsity. Finally, we introduce and explain in detail the notion of APT–consistent minimum dispersion S–SDFs. Section 3 addresses estimation and inference for S–SDFs and the corresponding minimum dispersion bounds. These findings provide the foundations for a general testing framework of asset pricing models based on S–SDFs, which is also developed. Section 4 presents our empirical analysis of S–SDFs induced by APT-type asymptotic no-arbitrage conditions. Finally, Section 5 concludes.

2 Theory of S–SDFs

Consider an economy consisting of a fixed number $N$ of basis securities with random payoffs $X := (X_n)_{n=1}^N$ at time 1 and associated quoted prices $P \in \mathbb{R}^N$ at time 0. These securities are partitioned into so-called sure and dubious securities, by means of two index sets, $S \subset \{1, \ldots, N\}$ and its complement $D := \{1, \ldots, N\} \setminus S$. 

7
having cardinalities $N_S$ and $N_D$, respectively.\footnote{Note that one out of these index sets might be empty, in which case either only sure or only dubious securities exist. The first case corresponds to the standard minimum dispersion SDF setting in, e.g., Hansen and Jagannathan [1991].} That is, $P_S := (P_n)_{n \in S}$ denotes the vector of prices and $X_S := (X_n)_{n \in S}$ the vector of payoffs of the sure assets. Similarly, $P_D$ and $X_D$ denote the vectors of dubious prices and payoffs, respectively. Sure securities are those for which quoted prices can be exactly matched by a linear pricing functional represented by a corresponding S–SDF. Dubious securities are those for which such a linear pricing rule implies instead non-zero pricing errors.

In the sequel, we first detail our approach for specifying non-zero pricing errors on dubious securities and the corresponding definition of an S–SDF. We then show that existence of an S–SDF is always supported by a viable market with trading frictions on the dubious assets, in which S–SDFs represent linear pricing functionals that are compatible with such trading frictions. Subsequently, we lay down various relevant examples of existing as well as new such S–SDFs. In line with Hansen and Jagannathan [1991], we then introduce minimum dispersion S–SDFs and derive their dual characterization in terms of the solutions to corresponding penalized optimal portfolio problems. Lastly, we introduce our systematic S–SDF modelling methodology based on Moreau envelopes and explain in detail the notion of an APT–consistent S–SDF.

### 2.1 Definition and existence of S–SDFs

The vector of pricing errors under a generic stochastic discount factor $M$ is given by $E[M X] - P$.\footnote{In more general multi-period economies, the vector of (conditional) pricing errors under stochastic discount factor $M$ is given by $E[M X | T] - P$, where $T$ is the information set available to investors at the time of the trade and $E[\cdot | T]$ is the corresponding conditional expectation operator. In this article, we do not model explicitly the information set $T$ and work without loss of generality with the vector of expected pricing errors $E[M X] - P$. As is standard in the literature, conditioning information can be incorporated in this setting by considering prices and payoffs instrumented by random variables in the information set $T$, which gives rise to unconditional pricing constraints for dynamically managed portfolios of asset payoffs.} Given the willingness or the necessity to tolerate some pricing errors for the dubious assets, a general approach to control their pricing error size and geometry can rely on following pricing error constraints:

$$E[M X_S] - P_S = 0 \quad \text{and} \quad h(E[M X_D] - P_D) \leq \tau,$$

where $\tau \geq 0$ and $h$ is a pricing error function in the set $\Gamma(\mathbb{R}^{N_D})$ of functions from $\mathbb{R}^{N_D}$ to $(-\infty, +\infty]$ that are proper, closed and convex.\footnote{Function $f : \mathbb{R}^N \to (-\infty, +\infty]$ has domain $\text{dom } f := \{x \in \mathbb{R}^N : f(x) < +\infty\}$; it is proper if $\text{dom } f$ is not empty and $f(x) > -\infty$ for every $x \in \mathbb{R}^N$; it is closed if the sublevel set $\{x \in \text{dom } f : f(x) \leq s\}$ is a closed set for each $s \in \mathbb{R}$.} Literally, linear pricing functional $E[M \cdot]$ in equation (1) matches exactly the quoted prices of sure payoffs and it implies pricing errors on dubious assets bounded by threshold $\tau$ under pricing error metric $h$. Here, set $\Gamma(\mathbb{R}^{N_D})$ of admissible pricing error metrics is chosen general enough to incorporate any convex pricing error constraint that is relevant for our work.
Granted the importance and generality of pricing constraints (1), we define S–SDFs as follows.

**Definition 1 (S–SDFs).** Given pricing error function \( h \in \Gamma(\mathbb{R}^{N_D}) \) and pricing error bound \( \tau \geq 0 \), an S–SDF is a non-negative random variable \( M \) satisfying pricing constraints (1).

Almost every existing asset pricing model implies SDFs that are consistent with pricing constraints of the form (1). Obvious ones are standard SDFs that price exactly all payoffs, which follow using the characteristic function \( h := \delta_\{0\} \) of singleton set \( \{0\} \), defined by:

\[
\delta_{\{0\}}(\mathbb{E}[MX] - P_D) = \begin{cases} 
0 & \mathbb{E}[MX] - P_D = 0 \\
\infty & \text{else}
\end{cases}
\]

Section 2.2 addresses in more detail many other relevant examples of S–SDFs. Before discussing them further, we first show that S–SDFs are supported in general by corresponding viable markets in the sense of Harrison and Kreps [1979], in which the trading of dubious assets is subject to convex transaction costs. This basically means that any S–SDF is supported by a suitable economy with frictions, in which at least one budget-constrained agent with well-behaved preferences is able to select an optimal consumption plan.

In order to formalize the above existence statement, let \((\Omega, \mathcal{F}, \mathbb{P})\) be a complete probability space and \(L^p\) with \( p \in [1, +\infty] \) be the space of \( p \)-integrable random variables on this probability space. Throughout this study, we assume that payoff \( X_n \in L^p \) for each basis security indexed by \( n = 1, \ldots, N \). Suppose further that portfolio positions \( \theta_S \in \mathbb{R}^{N_S} \) on the sure payoffs involve no transaction costs, while portfolio positions \( \theta_D \in \mathbb{R}^{N_D} \) on dubious payoffs involve transaction costs, consisting of a varying cost component measured by function \( h^* \in \Gamma(\mathbb{R}^{N_D}) \), the convex conjugate of \( h \), and a fix cost component \( \tau \geq 0 \).

We specify the set of traded payoffs as the set of portfolio payoffs involving finite transaction costs,

\[
Z := \{Z = X^{'\theta} : h^*(\theta_D) < +\infty\}
\]

where \( \theta := (\theta^*_S, \theta^*_D)' \), and accordingly define a pricing functional \( \pi \) on \( Z \) by:

\[
\pi(Z) := \inf_{\theta \in \mathbb{R}^N} \{P^{'\theta} + h^*(\theta_D) + \tau : Z = X^{'\theta}\}
\]

By definition, among all possible transaction prices \( P^{'\theta} + h^*(\theta_D) + \tau \) for a payoff \( Z \), \( \pi(Z) \) reproduces

---

4 We equip \( L^p := \{x : \Omega \to \mathbb{R} : x \) is measurable, \( \mathbb{E}[|x|^p]^{1/p} < +\infty \}, p \in [1, +\infty] \), with the partial order \( \leq \) defined by: \( x \leq y \) if and only if \( P(x \leq y) = 1 \). Also the duality pair between \( L^p \) and \( L^q \), for \( 1/p + 1/q = 1 \), is given by \( \mathbb{E}[xy] \), for any \( x \in L^p \), \( y \in L^q \). In what follows, we do not consider payoffs in \( L^p \) with \( p = 1 \) or \( p = +\infty \) in order to obtain compact proofs involving \( L^p \)–\( L^q \) duality and remain in the realm of the norm topology.

5 The convex conjugate \( h^* \) of a convex function \( h \) is defined for any \( \theta_D \in \mathbb{R}^{N_D} \) by \( h^*(\theta_D) := \sup_{\eta \in \mathbb{R}^{N_D}} \{\eta^{'\theta_D} - h(\eta)\} \). Further, as \( h \in \Gamma(\mathbb{R}^{N_D}) \), it follows that also \( h^* \in \Gamma(\mathbb{R}^{N_D}) \) and that the convex conjugate of \( h^* \) is \( h \) itself. Thus the two functions determine one another.

6 It is common to assume that the price of a traded payoff is not arbitrarily negative: \( \pi(Z) > -\infty \) for every payoff \( Z \in Z \).
the lowest price at which investors are willing to trade. Since the transaction cost function $h^*$ belongs to $\Gamma(\mathbb{R}^{N_D})$, the set of traded payoffs $Z$ and pricing functional $\pi$ are both convex. If in addition $\tau = 0$ and $h^*$ is sublinear, we obtain the arbitrage-free asset pricing framework with frictions in Jouini and Kallal [1995].\footnote{Function $h^*$ is sublinear if $h^*(\theta_D + \tilde{\theta}_D) \leq h^*(\theta_D) + h^*(\tilde{\theta}_D)$ and $h^*(a\theta_D) = ah^*(\theta_D)$ for any $a \geq 0$ and any $\theta_D \in \mathbb{R}^{N_D}$.} In this case, market $Z$ is a convex cone and pricing functional $\pi$ is sublinear as well. If, in addition, markets are frictionless, i.e., $h^*$ is the zero function, we obtain the classical arbitrage-free asset pricing framework in Harrison and Kreps [1979], in which $Z$ is a linear market and pricing functional $\pi$ is linear, i.e., the Law of One Price holds.

Given the economic setting just described, the next proposition states that existence of a S–SDF is equivalent to the viability of market $(Z, \pi)$, where viability is defined formally in Definition 5 of Appendix B.1 following Harrison and Kreps [1979].

**Proposition 1 (Existence of S–SDFs).** The set of traded payoffs $Z$ and pricing function $\pi$ satisfying $\pi(Z) > -\infty$ for every payoff $Z \in Z$ define a viable financial market if and only if there exists strictly positive $M \in L^q$, with $1/p + 1/q = 1$, satisfying pricing constraints (1),

$$E[MX_S] - P_S = 0 \quad \text{and} \quad h(E[MX_D] - P_D) \leq \tau ,$$

where functions $h$ and $h^*$ are convex conjugate to each other.

Proposition 1 provides a full characterization of the pricing properties of S–SDFs in terms of the traded payoffs and quoted prices of a supporting arbitrage-free economy with frictions. In this setting, viability is a necessary and sufficient condition for $(Z, \pi)$ to be a feasible model of an economic equilibrium. It requires existence of at least one budget-constrained agent with well-behaved preferences trading in market $(Z, \pi)$, who is able to select an optimal consumption plan. Viability is tightly connected to the common notion of absence of arbitrage. Chen et al. [2001] derive a version of the Fundamental Theorem of Asset Pricing for the general setting of viable financial markets with a convex set of traded payoffs and a convex pricing functional. When the set of traded payoffs is a convex cone and the pricing functional is sublinear, Chen et al. [2001] show that viability is equivalent to absence of arbitrage in that market. In contrast, in general convex financial markets viability implies absence of arbitrage and is thus a stronger condition.

Pricing error function $h$ in Definition 1 fixes the geometry of the pricing errors generated by an S–SDF and it thus crucially determines the deep S–SDFs properties. For instance, the chosen S–SDF pricing error geometry has direct implications for the S–SDF characterization of assets excess returns. Indeed, for any
sure asset payoff following standard expected excess return identity holds:

$$\mathbb{E}[X_n] - \frac{P_n}{\mathbb{E}[M]} = -\text{Cov}\left[\frac{M}{\mathbb{E}[M]}, X_n\right], \quad n \in S.$$  \hfill (5)

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

$$\mathbb{E}[X_n] - \frac{P_n}{\mathbb{E}[M]} = -\text{Cov}\left[\frac{M}{\mathbb{E}[M]}, X_n\right] \left(\mathbb{E}[MX_n] - \frac{P_n}{\mathbb{E}[M]}\right), \quad n \in D.$$  \hfill (6)

It follows that the expected excess returns of dubious assets are given by the sum of a risk premium component, reflecting exposure to systematic S–SDF risk, and an additional pricing error component. Hence, a given S–SDF geometry for pricing errors directly influences the way an S–SDF explains a cross-section of asset excess returns, in terms of their exposure to S–SDF risk and asset pricing errors, respectively.

### 2.2 Relevant examples of S–SDFs

Before introducing minimum dispersion S–SDFs, we discuss in more detail economically relevant examples of S–SDFs. From Definition 1, the deep S–SDF properties are characterized by the pricing error geometry induced by function \( h \) via pricing error constraints (1). Equivalently, since function \( h \) stays in a one-to-one relation with its convex conjugate \( h^* \), they are characterized with Proposition 1 in terms of the varying transaction cost component \( h^* \) in pricing functional (4). This dual interpretation of S–SDFs provides a tight link between the pricing error properties of a given S–SDF and the set of corresponding portfolios that may be accessible to investors to support it. In this way, it provides direct economic content to any S–SDF pricing error geometry induced by a pricing error metric \( h \in \Gamma(\mathbb{R}^{N_D}). \)

Using the above dual interpretation of S–SDFs, we discuss below various S–SDFs settings with concrete examples of different relevant pricing error geometries \( h \) associated to their corresponding interpretation in terms of varying transaction cost component \( h^* \). Examples 1–4 start from different concrete transaction cost interpretations and obtain the associated S–SDF pricing error geometry.\(^8\) Example 5 starts instead from a set of S–SDF pricing error constraints that are prescribed by Ross’s Arbitrage Pricing Theory (APT), Ross [2013], and derives the corresponding varying transaction cost component.

**Example 1 (Short-sale restrictions and bid-ask spreads).** Minimum variance SDFs incorporating the pricing implications of bid-ask spreads and short-sale constraints in arbitrage-free markets have been studied in Luttmer [1996]. Such market frictions can be modelled by considering zero transaction costs for portfolio positions \( \theta_D \) that belong to a specific convex cone \( K \subset \mathbb{R}^{N_D} \), such as the set \( \mathbb{R}_+^{N_D} \) of vectors with

---

\(^8\) Examples of studies involving (i) borrowing and short-sale constraints, bid-ask spreads and margin requirements are found in Luttmer [1996] and Naik and Uppal [1994], (ii) proportional transaction costs are found in Mei et al. [2016] and Constantinides [1986], and (iii) quadratic transaction costs are the focus in G"arleanu and Pedersen [2013] and DeMiguel et al. [2015]. The use of convex functions modelling transaction costs is advocated in, e.g., Lillo et al. [2009] and Constantinides [1979].
non-negative components, and arbitrarily large transaction costs for those portfolio positions that do not belong to $K$. That is, $h^* = \delta_K$, the characteristic function of set $K$. Luttmer [1996] shows that SDFs in such markets are characterized by pricing errors that are constrained to belong to the negative dual cone of $K$:  
\begin{equation}
    \mathbb{E}[M X_D] - P_D \in -K^* \iff \delta_{-K^*}(\mathbb{E}[M X_D] - P_D) \leq \tau .
\end{equation}
Hence, $h = \delta_{-K^*}$ defines the S–SDF pricing error geometry in these settings. For instance, if $K = K^* = \mathbb{R}^D_+$ one obtains $\mathbb{E}[M X_D] \leq P_D$, i.e., a sublinear pricing functional in an arbitrage-free market with short-selling constraints.

**Example 2 (Proportional, norm-based, transaction costs).** Transaction costs that are proportional to the portfolio positions on the dubious assets can be obtained by considering a transaction cost function $h^* = \lambda \| \cdot \|$, where $\| \cdot \|$ is some norm on $\mathbb{R}^D$ and $\lambda > 0$ a scaling parameter. The convex conjugate of a scaled norm is the characteristic function of the centered closed ball of its inversely scaled dual norm, $B(\frac{1}{\lambda} \| \cdot \|_*) = \lambda B(\| \cdot \|_*)$.\(^9\) Therefore, $h = \delta_{\lambda B(\| \cdot \|_*)}$ and the resulting S–SDF pricing error geometry implies pricing errors bounded by $\lambda$ according to the dual norm $\| \cdot \|_*$:
\begin{equation}
    \| \mathbb{E}[M X_D] - P_D \|_* \leq \lambda .
\end{equation}
Here, a higher scaling parameter $\lambda$ implies larger admissible pricing errors. Moreover, the sublinearity of $h^*$ implies that entry cost $\tau$ in Proposition 1 is in this case immaterial for determining pricing error bound (8). Hence, it can be put to zero without loss of generality.\(^11\)

**Example 3 (Leverage restrictions).** By switching the roles of $h$ and $h^*$ in Example 2, one easily obtains S–SDFs supported by markets characterized by restrictions on borrowings and leverage. If, for instance, investors can only form portfolios with absolute positions on the dubious assets bounded by a threshold $\lambda > 0$, i.e., $\| \theta_D \|_\infty \leq \lambda$, then the transaction cost function is the characteristic function of the centered closed ball $\lambda B_\infty$ of the $l_\infty$-norm with radius $\lambda$:
\begin{equation}
    h^*(\theta_D) = \delta_{\lambda B_\infty}(\theta_D) := \begin{cases} 0 & \| \theta_D \|_\infty \leq \lambda \\ +\infty & \text{else} \end{cases}.
\end{equation}
In this case, the pricing error function is the lasso penalty $h := \lambda \| \cdot \|_1$ and so there exists from Proposition 1 a S–SDF satisfying $\| \mathbb{E}[M X_D] - P_D \|_1 \leq \tau / \lambda$. Thus, the larger threshold $\lambda$ determining the admissible absolute portfolio positions on the dubious assets, the smaller the pricing errors on the dubious payoffs.

**Example 4 (Quadratic transaction costs).** A simple quadratic specification of the varying transaction cost component can be based, e.g., on a ridge penalization $h^* := \frac{\alpha}{2} \| \cdot \|^2_2$ with a scaling parameter $\alpha > 0$. Self-duality of the ridge penalization yields $h = \frac{1}{2\alpha \tau} \| \cdot \|^2_2$ and an S–SDF characterized by the pricing error bound $\| \mathbb{E}[M X_D] - P_D \|_2^2 \leq 2\alpha \tau$, i.e., once again a higher transaction cost scaling parameter $\alpha$ results in larger admissible pricing errors. More general quadratic transaction cost specifications can promote sparse portfolios based on, e.g., an elastic-net penalty $h^* = \lambda \| \cdot \|_1 + \frac{\alpha}{2} \| \cdot \|^2_2$ with $\alpha > 0$ and $\lambda > 0$, as in Nagel et al. [2018]. In this case, it follows from Corollary 2 in Appendix B.3 that the pricing error function is given by the squared Euclidean distance $h := d_{\lambda B_\infty}^2$ from the $l_\infty$-centered closed ball of radius $\lambda$. Hence, there exists from Proposition 1 a S–SDF such that the inequality in pricing errors constraints (1) reads:
\begin{equation}
    d_{\lambda B_\infty}^2(\mathbb{E}[M X_D] - P_D) \leq \tau .
\end{equation}
\(^9\) The dual cone of set $K \subseteq \mathbb{R}^D$ is defined by $K^* := \{ \theta_D \in \mathbb{R}^D : \theta_D \eta \geq 0 \text{ for all } \eta \in K \}$.
\(^10\) The dual norm $\| \cdot \|_*$ of norm $\| \cdot \|$ is defined as $\| \theta_D \|_* := \max_{\eta \in \mathbb{R}^D} \{ \theta_D \eta : \| \eta \| \leq 1 \}$ for any $\theta_D \in \mathbb{R}^D$. For instance, the dual norm of an $l_\alpha$-norm, that is $\| x \|_\alpha := (\sum_{i=1}^D x_i^\alpha)^{1/\alpha}$ when $p \in [1, +\infty)$ and $\| x \|_p := \max_{i \leq D} | x_i |$ when $p = +\infty$, is the $l_\alpha$-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. The centered closed ball of scaled dual norm $\| \cdot \|_*/\lambda$ is defined as $\lambda B(\| \cdot \|_*) := \{ x \in \mathbb{R}^D : \| x \| \leq \lambda \}$.
\(^11\) Explicitly, pricing error constraint (1) translates in this setting to the constraint:
\begin{equation}
    \tau \geq \delta_{\lambda B(\| \cdot \|_*)}(\mathbb{E}[M X_D] - P_D) := \begin{cases} 0 & \| \mathbb{E}[M X_D] - P_D \|_* \leq \lambda \\ +\infty & \text{else} \end{cases},
\end{equation}
which in turn is equivalent to pricing error constraint (8) on the dubious assets.
This inequality explicitly reproduces the hidden S–SDF pricing error geometry supported by an elastic-net penalization on the portfolio positions of dubious assets.

**Example 5 (Arbitrage Pricing Theory).** The asymptotic no-arbitrage condition in the APT yields a particular upper bound on pricing errors, both under a correctly-specified or under a misspecified factor model for returns; see, e.g., Uppal et al. [2018]. Consider, without loss of generality, the tradable factors as sure assets, i.e., they are exactly priced. In contrast, any other return not spanned by factor risk is dubious. Factor tradability implies that factor risk can be completely hedged away. Therefore, we can focus for convenience, but without loss of generality, on a vector of dubious gross returns $X_D$ having components

$$X_n = R_f + [R_n^e - \text{proj}_{\text{span} X_S^e}(R_n^e)] , \quad n \in D,$$

where $R_f$ denotes the gross risk-free rate, $R_n^e := R_n - R_f$ the original gross excess asset return and $\text{proj}_{\text{span} X_S^e}$ the orthogonal projection on the space of traded factor excess returns $X_S^e := R_F - 1R_f$. Denote by $\eta := E[M X_D] - 1$ the vector of pricing errors on dubious returns under S–SDF $M$ and by $\Sigma$ the covariance matrix of dubious returns. The bound on pricing errors in equation (1), which is implied by the APT for some $\tau > 0$, is obtained using a quadratic pricing error function:\textsuperscript{12}

$$h(\eta) = \eta^\top \Sigma^{-1} \eta.$$

The corresponding transaction costs associated with portfolio positions $\theta_D$ in the dubious assets is explicitly given by:

$$h^*(\theta_D) = \theta_D^\top \Sigma \theta_D . \quad (9)$$

It is important to note that the economic properties of transaction cost function (9) crucially depend on whether the factor model for returns is correctly specified, i.e., on whether matrix $\Sigma$ has some dominating eigenvalues in large asset markets. Under correct specification, no eigenvalue dominates. Therefore, the quadratic transaction costs of portfolios with weights proportional to any principal component of matrix $\Sigma$ are similar. When the factor model for returns is misspecified and there are dominating eigenvalues in matrix $\Sigma$, the transaction costs of portfolios with weights proportional to the dominating principal components prevail over those of portfolios with weights proportional to any other principal component, i.e., dominating principal component portfolios are penalized more. Equivalently, the pricing errors of portfolios corresponding to dominating principal components orthogonal to the traded factors are constrained more. Figure 1 illustrates the implications of variance covariance matrix $\Sigma$ for the APT pricing error function $h$ and corresponding transaction cost function $h^*$.

### 2.3 Minimum dispersion S–SDFs

Having highlighted the economic meaning of pricing error constraints (1) for various relevant asset pricing settings, in this section we derive optimal model-free S–SDFs that minimize various established notions of variability, under these general pricing error constraints. Borrowing from Rockafellar [1971], the variability of SDFs can be measured via a general family $\Phi$ of functions capturing established notions of stochastic dispersion in the literature. $\Phi$–dispersions are integral functionals of the form

$$\mathbb{E}[\phi(\cdot)] : L^d \to (-\infty, +\infty) , \quad (10)$$

\textsuperscript{12} Section 2.5 below provides details on the theoretical foundations of APT–induced pricing error bounds on S–SDFs.
where function \( \phi : \mathbb{R} \to (-\infty, +\infty] \) is such that \( \phi_+ \), the restriction of \( \phi \) to the nonnegative real line, is in \( \Gamma(\mathbb{R}) \) and is finite over the open interval \((0, +\infty)\).\(^{13}\) An important role in characterizing dual portfolio problems induced by minimum dispersion S–SDFs is played by \( \Phi \)--dispersions based on the convex conjugate of \( \phi_+ \),

\[
\mathbb{E}[\phi_+^* (\cdot)] : L^p \to (-\infty, +\infty] ,
\]

which measure the dispersion of portfolio payoffs. Many well-known measures of dispersion are of the form (10), such as the Hansen-Jagannathan distance Hansen and Jagannathan [1991], entropy-based dispersions and more generally dispersions in the Cressie-Read family introduced by Cressie and Read [1984]. Appendix A collects relevant examples of \( \Phi \)--dispersions in the Cressie-Read family, together with their \( \phi_+ \) functions and the corresponding convex conjugate \( \phi_+^* \).

We next formalize the notion of minimum dispersion S–SDFs as follows.

**Definition 2 (Minimum dispersion S–SDFs).** SDFs that are consistent with pricing constraints (1) and minimize a particular \( \Phi \)--dispersion are called minimum dispersion S–SDFs. That is, they are solutions to

\[
K(\tau) := \inf_{M \in \mathcal{M}} \{ \mathbb{E}[\phi(M)] : h(\mathbb{E}[M X_D] - P_D) \leq \tau \} ,
\]

where \( \mathcal{M} \subset L^q \) denotes the convex subset of nonnegative SDFs that exactly price the sure assets:

\[
\mathcal{M} := \{ M \in L^q : M \geq 0, \mathbb{E}[M X_S] - P_S = 0 \} .
\]

In Definition 2, the minimization problem defining minimum dispersion S–SDFs extends the setting in Hansen and Jagannathan [1991] and Luttmer [1996] to construct generalized minimum dispersion SDF bounds that can incorporate general pricing error properties. In Section 3.3 below, we exploit these generalized S–SDF bounds to develop diagnostic tools for asset pricing models that do not rely on the assumption of exact pricing.

If in Definition 2 an S–SDF \( M_0^S \) is feasible for \( K(+\infty) \), i.e., it is only required to price the sure assets exactly, then it attains by definition the lowest \( \Phi \)--dispersion among all S–SDFs that solve problem (12) for some \( \tau \geq 0 \). In particular, if:

\[
\tau^{max} := h(\mathbb{E}[M_0^S X_D] - P_D) < +\infty ,
\]

then this S–SDF is feasible for any \( K(\tau) \) such that \( \tau \geq \tau^{max} \) (i.e., \( E[\phi(M_0^S)] = K(\tau) \) for any \( \tau \geq \tau^{max} \)).

\(^{13}\)As we defined S–SDFs to be nonnegative random variables with certain pricing properties, we can focus on the restriction of \( \phi \) to the nonnegative real line, i.e., function \( \phi_+ : \mathbb{R} \to (-\infty, +\infty] \) is defined for every \( x \in \mathbb{R} \) by

\[
\phi_+(x) = \begin{cases} 
\phi(x) & x \geq 0 \\
+\infty & \text{else}
\end{cases} .
\]

Since \( \phi_+ \in \Gamma(\mathbb{R}) \), so is its convex conjugate \( \phi_+^* \).
Similarly, if in Definition 2 an S–SDF $M_0^{S,SDF}$ is able to exactly price the sure and the dubious assets, then it has to have the largest dispersion $K(0)$ among all S–SDFs solving problem (12) for some $\tau \geq 0.1$ between these two extremal cases, the S–SDF minimum dispersion bound $K : \mathbb{R}_+ \to (-\infty, +\infty]$ is convex and non increasing in $\tau$, by [Luenberger, 1997, Prop. 1 and 2, p. 216-217].

Working directly with optimization problems of the form (12) is not computationally feasible, as they are infinite-dimensional problems defined in terms of unobservable random variables. Therefore, we shall provide a dual characterization of S–SDFs, in terms of the solutions of finite-dimensional dual portfolio problems with penalized portfolio weights. To this end, it is convenient to first reinterpret constrained S–SDF problems (12) as more general penalized S–SDF problems, for any pricing error penalty $\psi \in \Gamma(\mathbb{R}^{N_D})$:

$$\Pi := \inf_{M \in M} \{ \mathbb{E}[\phi(M)] + \psi(\mathbb{E}[MX_D] - P_D) \} . \quad (15)$$

Note that penalized S–SDF problem $\Pi$ can always be constructed directly from constrained S–SDF problem $K(\tau)$ in equation (12), by setting $\psi := \delta_C$ with $C := \{ \eta \in \mathbb{R}^{N_D} : h(\eta) \leq \tau \}$. Written in this way, the constrained and the penalized S–SDF problems are equivalent. Nevertheless, it is useful to study dual portfolio problem characterizations for the broader family of penalized S–SDF problems (15) with generic penalization $\psi \in \Gamma(\mathbb{R}^{N_D})$, because they include relevant settings in which constrained S–SDF problem (12) obtains a tractable dual portfolio problem characterization only indirectly, via the Lagrange Multiplier Theorem; see Proposition 9 in Appendix B.2.

Before formulating our main S–SDF duality results, we recall for easy reference our basic assumptions regarding the class of $\Phi$–dispersions and penalization function $\psi$ used in our framework.

**Assumption 1.** The restriction $\phi_+$ of function $\phi : \mathbb{R} \to (-\infty, +\infty]$ to the non negative real line lies in $\Gamma(\mathbb{R})$ and is finite over open interval $(0, +\infty)$. Further, $\psi \in \Gamma(\mathbb{R}^{N_D})$.

Moreover, following constraint qualification is required to focus on the relevant S–SDF settings for which a duality result in terms of a corresponding dual optimal portfolio can be established.

**Assumption 2.** There exists $\tilde{M} > 0$ with $\mathbb{E}[\phi(\tilde{M})] < +\infty$, exactly pricing the sure payoffs and such that either (i) $\mathbb{E}[MX_D] - P_D \in \text{ri(dom } \psi)$, or (ii) $\psi(\mathbb{E}[MX_D] - P_D) < +\infty$ is $\psi$ is piecewise linear.$^{14}$

When $\psi = \delta_C$, the characteristic function of a non-empty closed and convex set $C$, or in cases where $\text{dom } \psi = \mathbb{R}^{N_D}$, Assumption 2 is directly satisfied in a viable market in which transaction costs are measured by $\psi^*$, the convex conjugate of $\psi$, as shown in Proposition 1.$^{15}$ These two settings cover all relevant cases

---

$^{14}$ The relative interior of $\text{dom } \psi$ is by definition the interior of the affine hull of $\text{dom } \psi$.

$^{15}$ Note that the integrability condition on $\phi(\tilde{M})$ in Assumption 2 can be ensured by requiring that basis payoffs and S–SDFs
for our analysis.

Under the above assumptions, we can now formally state the desired duality between S–SDF penalized problem (15) and a corresponding penalized optimal portfolio problem.

**Proposition 2 (Dual portfolio problem).** Consider the portfolio problem

\[ \Delta := \min_{\theta \in \mathbb{R}^N} \left\{ \mathbb{E}[\phi_+^*(-X'\theta)] + P'\theta + \psi^*(\theta_D) \right\}. \quad (16) \]

*Under Assumptions 1 and 2, \( \Pi = -\Delta \).*

In view of Proposition 2, the optimization problem underlying minimum dispersion S–SDFs is uniquely linked to a finite-dimensional portfolio optimization problem, in which the dispersion of portfolio payoffs is measured by the dual \( \Phi \)–dispersion (11) and in which portfolio weights corresponding to dubious securities are penalized by \( \psi^* \), the convex conjugate of the pricing error penalty \( \psi \). From Proposition 1, \( \psi^* \) can be interpreted as a measure of the transaction costs associated with portfolio positions on the dubious assets.

Proposition 2 does not take a stand on the relation between the solutions to the primal penalized S–SDF problem \( \Pi \) and the dual penalized portfolio problem \( \Delta \). Uniqueness of the solution to the penalized S–SDF problem \( \Pi \) follows by imposing additional structure on \( \Phi \)–dispersions, while uniqueness of the portfolio weights solving problem \( \Delta \) follows under an additional non-degeneracy assumption on the vector of traded payoffs \( X \).

**Proposition 3 (Uniqueness of solutions).** *Given the assumptions of Proposition 2, it follows:*

(i) If \( \phi_+ \) is strictly convex, then any optimal solution to problem \( \Pi \) is unique.\(^{16}\)

(ii) If \( \phi_+^* \) is strictly convex and vector \( X \) consists of linearly independent payoffs, then any optimal solution to problem \( \Delta \) is unique.

\( \phi_+ \) and \( \phi_+^* \) are both strictly convex for all Cressie-Read \( \Phi \)–dispersions in Examples 1 of Appendix A, except for \( \phi_+^* \) in case \((3a)\), which includes the case of Hansen-Jagannathan distance. In this case, \( \phi_+^* \) is strictly convex only on the non-negative real line.

The next proposition characterizes the link between the unique S–SDF solving problem \( \Pi \) in Proposition 3 and the solution of the corresponding dual optimal portfolio problem \( \Delta \). Further, it shows how one can compute the dubious pricing errors under S–SDF solving problem \( \Pi \) in terms of the gradient of \( \psi^* \), when it exists.

---

\(^{16}\) A function \( f \) is strictly convex if it is so only on its domain \( \text{dom} \ f \). If \( f \in \Gamma(\mathbb{R}) \) is strictly convex, then the convex conjugate \( f^* \) is smooth on the (non-empty) interior of its domain, see [Borwein and Lewis, 1991, Thm. 4.6].
Proposition 4 (S–SDF Link). Given the assumptions of Proposition 2, let \( \phi_+ \) be strictly convex and \( \theta_0 \) be a solution of dual portfolio problem \( \Delta \) such that \(-X'_\theta \theta_0 < d_\phi := \lim_{x \to +\infty} \phi(x)/x \) almost surely.\(^{17}\)

(i) Then, the unique S–SDF solving problem \( \Pi \) is given by:\(^{18}\)
\[
M_0 = (\phi_+^\prime)^{-1}(-X'_\theta \theta_0). \tag{17}
\]

(ii) If in addition \( \psi^* \) is differentiable at \( \theta_{0D} \), then:\(^{19}\)
\[
\mathbb{E}[M_0X_D] - P_D = \nabla \psi^*(\theta_{0D}). \tag{18}
\]

In Proposition 4, the form of the unique S–SDF solving problem \( \Pi \), as a function of the optimal portfolio payoff solving dual problem \( \Delta \), is completely determined by the shape of \( \phi_+^\prime \), through its derivative. Corollary 1 of Appendix A reports the explicit expressions of this derivative for the class of Cressie-Read \( \Phi \)–dispersions in Examples 1 of Appendix A. Moreover, the geometry and size of the pricing errors under minimum dispersion S–SDFs are determined by the transaction cost function \( \psi^* \), through its subgradient \( \partial \psi^*(\theta_{0D}) \). When dual penalization function \( \psi^* \) is differentiable, the vector of pricing errors of a minimum dispersion S–SDF is always exactly reproduced via equation (18) by the gradient of \( \psi^* \) at the dual solution. Therefore, in this case it is always possible to explicitly compute the dependence of pricing errors of minimum dispersion S–SDFs on the optimal portfolio weights in the corresponding penalized portfolio problem. Given the one-to-one connection between any proper closed and convex function and their convex conjugates, the deep properties of minimum dispersion S–SDFs are intrinsically determined by the choice of the \( \Phi \)–dispersion to be minimized and the chosen geometry for bounding pricing errors.

It is important to note that Assumption 2 is a key requirement for Propositions 2–4. A violation of this assumption can imply that a candidate S–SDF solution associated in Proposition 4(i) with dual portfolio problem \( \Delta \) is not a solution of the primal S–SDF problem \( \Pi \). Figure 2 illustrates such a possible duality failure empirically, based on the datasets used in the empirical study of Section 4. We solve the empirical version of dual problem \( \Delta \) for various pricing error bounds \( \tau \), using sample variance as a measure of S–SDF dispersion and an \( l_2 \)–penalization function for dual portfolio weights, which is APT–consistent in the sense discussed in Section 2.5 below. We then estimate a candidate minimum variance S–SDF \( M_0 \) given by Proposition 4(i). Finally, we estimate the resulting pricing error vector \( \mathbb{E}[M_0X_D] - P_D \) in problem (15) for this candidate S–SDF and we verify whether indeed the condition \( h(\mathbb{E}[M_0X_D] - P_D) \leq \tau \) is satisfied.

\(^{17}\)See Appendix A for the values of \( d_\phi \) for most of the Cressie-Read dispersions.

\(^{18}\)For any real number \( x \), we denote by \( (\phi_+^\prime)^{-1}(x) \) the first derivative of function \( \phi_+^\prime \) in \( x \). Moreover, we denote by \( \nabla \psi^*(\theta_{0D}) \) the subgradient of function \( \psi^* \) in \( \theta_{0D} \). Whenever \( \psi^* \) is differentiable, \( \nabla \psi^*(\theta_{0D}) = \{ \nabla \psi^*(\theta_{0D}) \} \), where \( \nabla \psi^*(\theta_{0D}) \) is the gradient of function \( \psi^* \) in \( \theta_{0D} \).

\(^{19}\)Note that in general the pricing errors of the dubious securities under S–SDF \( M_0 \) satisfy \( \mathbb{E}[M_0X_D] - P_D \in \partial \psi^*(\theta_{0D}) \).
consistently with the pricing error restriction in constrained primal S–SDF problem (12). We find that for a range of values for parameter \( \tau \) strictly smaller than the point of discontinuity in the plot of Figure 2, the candidate SDF \( M_0 \) implied by the dual portfolio problem yields a serious violation of the inequality \( h(\mathbb{E}[M_0 X_D] - P_D) \leq \tau \), i.e., \( M_0 \) is not a solution of the empirical primal S–SDF problem in those cases. Therefore, no empirical minimum variance SDF exactly pricing all returns according to the finite-sample pricing restrictions exists in such cases. In contrast, an empirical minimum variance S–SDF solution exists for sufficiently large pricing error bounds \( \tau \).\(^{20}\)

### 2.4 Modelling S–SDFs

Pricing error penalty \( \psi \), or equivalently transaction cost penalty \( \psi^* \), crucially determines the S–SDF pricing error geometry. A useful property for solving penalized dual portfolio problem (16) in applications, is tractability of the transaction cost penalty \( \psi^* \). For instance, when \( \psi^* \) is smooth, pricing errors under minimum dispersion S–SDFs are easily characterized by the gradient of \( \psi^* \), as shown in equation (18). Such smoothness is also useful to obtain some of the asymptotic results developed in Section 3 for the empirical analysis of S–SDFs, under our unified approach for a broad class of penalizations.

We develop a unifying methodology to model the geometry of S–SDF pricing errors and the hidden portfolio weight penalizations, by means of a convenient class of pricing error penalty functions \( \psi \) and conjugate transaction cost functions \( \psi^* \), which ensure tractability and smoothness of dual portfolio problem (16). Given a target pricing error function \( f \in \Gamma(\mathbb{R}^{N_D}) \), we propose to accurately approximate it, based on a ridge-corrected penalization function, defined by:

\[
\psi = f + \frac{\alpha}{2} \| \cdot \|_2^2 ,
\]

for some \( \alpha \geq 0 \). By definition, in identity (19) one always has \( \psi \in \Gamma(\mathbb{R}^{N_D}) \). Importantly, this family of penalizations contains many specifications considered in the literature. Moreover, it always leads to smooth transaction cost functions \( \psi^* \) when \( \alpha > 0 \), as the next proposition shows.\(^{21}\)

**Proposition 5.** Let \( f \in \Gamma(\mathbb{R}^{N_D}) \) and penalization function \( \psi \) in penalized S–SDF problem (15) be given by (19). Then it follows for any \( \theta_D \in \mathbb{R}^{N_D} \):

\[
\psi^*(\theta_D) = \begin{cases} 
(f_0 f^*)(\theta_D) & \alpha > 0 \\
\alpha(\theta_D) & \alpha = 0
\end{cases}
\]

\(^{20}\) In the empirical analysis of Section 4, we systematically ensure that our estimated S–SDFs are solutions of the empirical primal S–SDF problem, using a sufficiently large pricing error threshold \( \tau \).

\(^{21}\) The proofs of the results in this section are collected in Appendix B.3.
where \((e, f^*) \in \Gamma(\mathbb{R}^N_D)\), the Moreau approximation of \(f^*\), is defined as

\[
(e, f^*)(\theta_D) := \inf_{z \in \mathbb{R}^N_D} \left\{ f^*(z) + \frac{1}{2\alpha} \| z - \theta_D \|_2^2 \right\}.
\]

Moreover, for any \(\alpha > 0\) function \(\psi^*\) is differentiable, with \(\alpha^{-1}\)-Lipschitz continuous gradient given by

\[
\nabla \psi^*(\theta_D) = \frac{1}{\alpha} (\theta_D - \text{prox}_\alpha f^*(\theta_D)) ,
\]

where \(\text{prox}_\alpha f^* : \mathbb{R}^N_D \to \mathbb{R}^N_D\) is the proximal operator of \(f^*\), defined by

\[
\text{prox}_\alpha f^*(\theta_D) = \arg \min_{z \in \mathbb{R}^N_D} \left\{ f^*(z) + \frac{1}{2\alpha} \| z - \theta_D \|_2^2 \right\}.
\]

In Proposition 5, transaction cost penalty \(\psi^*\) is continuously differentiable whenever \(\alpha > 0\), with a gradient that is \(\alpha^{-1}\)-Lipschitz continuous, independently of the choice of function \(f\). Moreover, both \(\psi^*\) and its gradient are available in simplified- or even closed-form in many relevant applications, which are reported in Corollary 2 of Appendix B.3. Below, we discuss in more detail some of these important cases.

**Example 6 (Loose short-sale restrictions).** The analysis in Luttmer [1996] can be extended to include, e.g., loose short-sale constraints, as detailed by Statement (ii)a of Corollary 2 in Appendix B.3. Here, the target pricing error penalty \(f = \delta_K\) is the characteristic function of a convex cone \(K \subset \mathbb{R}^N_D\), giving rise to the pricing error penalty \(\psi = \delta_K + \frac{\alpha}{2} \| \cdot \|_2^2\). It then follows, for any \(\theta_D \in \mathbb{R}^N_D\):

\[
\psi^*(\theta_D) = \frac{1}{2\alpha} d_{-K^*}^2(\theta_D),
\]

where \(d_{-K^*}\) is the Euclidean distance from the negative dual convex cone \(-K^*\). For instance, consider the case \(K = -\mathbb{R}^N_D\) and \(-K^* = \mathbb{R}^N_D\). In this setting, pricing errors must satisfy \(\mathbb{E}[M X_D] \leq P_D\) under target penalization \(f\), but they are also subject to the quadratic penalization \(\frac{\alpha}{2} \| \mathbb{E}[M X_D] - P_D \|_2^2\), which shrinks them more and more towards zero as \(\alpha\) grows. Concerning the underlying portfolio weights, this setting induces loose short-sale restrictions when \(\alpha > 0\): transaction costs are zero for non-negative positions while negative positions involve quadratic price impact. Importantly, as \(\alpha \downarrow 0\) this framework approximates with increasing precision the portfolio weight penalization \(f^* = \delta_{\mathbb{R}^N_D}\) corresponding to sharp short-selling constraints.

**Example 7 (Sparsity inducing pricing error geometries).** Consider again the case of borrowing or leverage constraints in Example 3 of Section 2.2, in which portfolios are constrained by maximum absolute positions on dubious assets bounded by \(\lambda > 0\). Here, the relevant transaction cost penalty \(f^* = \delta_{\lambda B_{\infty}}\) and the resulting pricing error penalty is the lasso penalty \(f = \lambda \| \cdot \|_1\) in, e.g., Tilshirani [1996]. It is well-known that this penalization induces an increasing degree of sparsity on its argument as parameter \(\lambda\) becomes larger, where as \(\lambda \to +\infty\), i.e., portfolio weights are unrestricted, all pricing errors are zero and all assets are exactly priced. Therefore, this framework can be naturally used to determine endogenously dubious assets with no pricing errors under a given minimum dispersion S-SDF. A popular extension of the lasso penalty is the elastic-net penalty, Zou and Hastie [2005], which is a ridge-corrected lasso penalty \(\psi = \lambda \| \cdot \|_1 + \frac{\alpha}{2} \| \cdot \|_2^2\) falling in our general penalty framework (19). While Nagel et al. [2018] utilize elastic-net as a portfolio penalty to induce sparse portfolio weights, when used as a pricing error penalty it can help to model sparse pricing errors. The corresponding dual transaction cost function induces loose borrowing or leverage constraints and is obtained explicitly with Statement (i) of Corollary 2 in Appendix B.3. It is given for any \(\theta_D \in \mathbb{R}^N_D\) by:

\[
\psi^*(\theta_D) = \frac{1}{2\alpha} d_{\lambda B_{\infty}}^2(\theta_D),
\]

\(^{22} d_{-K^*}(\theta_D) := \theta_D - \text{proj}_{-K^*}(\theta_D) \|_2\), where \text{proj}_{-K^*}\) denotes the Euclidean projection on \(-K^*\).
where \( d_{\lambda B_\infty} \) denotes the Euclidean distance from the closed centred ball \( \lambda B_\infty \). Under this transaction cost function, quadratic costs \( \frac{[\theta_n - \lambda]^2}{2\alpha} \) are paid on any long or short portfolio position \( \theta_n \), \( n \in D \), larger than \( \lambda \) in absolute value, while no transaction costs are paid for portfolio positions such that \( |\theta_n| \leq \lambda \). Such quadratic transaction costs can model price impact costs arising when an investor places an absolute trade larger than \( \lambda \); see, e.g., Lillo et al. [2003] and Engle and Ferstenberg [2006]. The first two panels of Figure 3 illustrate the properties of the lasso and elastic-net penalties, together with the properties of their convex conjugates, by plotting the two-dimensional unit balls induced by these penalizations. As expected, both elastic-net and lasso penalties feature non-differentiability along the variables axes, which is the necessary condition for inducing strict sparsity. Consistently with the result in Proposition 5, the convex conjugate of elastic-net produces instead a unit ball with a differentiable surface.

**Example 8** (Approximate sparsity in minimum dispersion S–SDF dual portfolio weights). Sparsity in minimum dispersion S–SDF dual portfolio weights can be obtained with hidden proportional transaction costs modelled by a sparsity inducing penalization function, such as the lasso penalty \( f^* = \lambda \|\cdot\|_1 \) where \( \lambda > 0 \) is a scaling parameter. Since \( f = \delta_{\lambda B_\infty} \), the resulting pricing errors on the dubious assets have maximum absolute value bounded by \( \lambda \). If we introduce the ridge-corrected penalization function \( \psi = \delta_{\lambda B_\infty} + \frac{\alpha}{2} \|\cdot\|_2^2 \) with \( \alpha > 0 \), we obtain a mixed proportional and quadratic transaction cost function \( \psi^* = \gamma_{\alpha, \tau} \), which is explicitly given by the multivariate Huber function in Statement (ii)b of Corollary 2 of Appendix B.3. Indeed, the Huber function penalizes in a smooth way large absolute portfolio weights above a threshold \( \alpha \tau \) with an \( \ell_1 \)-penalty, while it penalizes small absolute weights below the same threshold quadratically. The third panel of Figure 3 illustrates the properties of the Huber penalty by plotting its induced two-dimensional unit ball. Importantly, the Huber penalty approximates the lasso penalty in a smooth way, and its shape can be rendered arbitrarily close to that of the lasso, by means of shrinking \( \alpha \)-corrections that induce an approximate sparsity in S–SDF dual portfolio weights.

### 2.5 Minimum dispersion S–SDFs and the APT

In this section, we characterize in detail the theoretical relations between minimum dispersion S–SDFs and the asset pricing error bounds induced by the APT; see again Example 5 above. To provide proper background to our approach, we start from a standard linear factor model for a \( N \times 1 \) vector \( \mathbf{R}^e := \mathbf{R} - R_f \mathbf{1} \) of asset excess returns:

\[
\mathbf{R}^e = \mathbf{\beta} \mathbf{F}^e + \zeta, \tag{24}
\]

where \( \mathbf{F}^e := \mathbf{F} - R_f \mathbf{1} \) is a \( N_S \times 1 \) vector of traded factor excess returns, \( \mathbf{\beta} \) a \( N \times N_S \) matrix of factor loadings and \( R_f \) is the risk-free return. The \( N \times 1 \) vector of residuals \( \zeta \) is assumed orthogonal to traded factor excess returns, but potentially cross-sectionally correlated and not zero mean. Its (full rank) variance covariance matrix is denoted with \( \Sigma \).

As shown, e.g., in Ingersoll Jr [1984], and more broadly in Uppal and Zaffaroni [2016], the absence of asymptotic arbitrage opportunities implies tight constraints on asset expected excess returns and pricing errors in factor model (24).\(^{23}\) Precisely, [Uppal and Zaffaroni, 2016, Thm 3.3] implies the existence of a sequence of portfolios \( \{\mathbf{\theta}_k\}_{k \in \mathbb{N}} \), corresponding to a sequence of growing economies \( \{\mathbf{R}_k\}_{k \in \mathbb{N}} \) with \( k \to \infty \), is said to generate an asymptotic arbitrage opportunity if along some subsequence \( \{\mathbf{\theta}'_{k'}\}_{k' \in \mathbb{N}} \):

\[
\text{Var}(\mathbf{\theta}'_{k'}, \mathbf{R}_{k'}) \to 0 \quad \text{and} \quad \mathbb{E}[\mathbf{\theta}'_{k'}(\mathbf{R}_{k'} - R_f \mathbf{1})] \geq \tau \to 0 . \tag{25}
\]
$N_S \times 1$ vector $\lambda = E[F^e]$ of traded factor risk premia such that for any $N$ there exists $\tau > 0$ satisfying following asset pricing bound in the APT factor model:\footnote{The original APT pricing error bound is $\alpha^T \Sigma^{-1} \alpha \leq \tau^2$ for some $\tau > 0$. [Uppal and Zaffaroni, 2016, Thm 3.3] states the existence of $\tau_\eta \geq 0$ such that $\eta^T \Sigma^{-1} \eta \leq \tau_\eta^2$, where $\Sigma$ is the covariance matrix of the non-pervasive component $\epsilon$ of $\zeta$. However, note that $\Sigma^{-1} - \Sigma^{-1}$ is positive semi-definite under the assumptions of [Uppal and Zaffaroni, 2016, Thm 3.3]. Hence bound (26) directly follows.}

$$
\Sigma^{-1/2} \eta := \sqrt{\eta^T \Sigma^{-1} \eta} \leq \tau ; \quad \alpha := E[\zeta] = \beta^H \lambda^H + \eta ,
$$

(26)

where $\beta^H \lambda^H$ is an expected return component that may be induced by an unknown number of $K$ hidden systematic risk factors, $\beta^H$ a $N \times K$ matrix of associated factor loadings, $\lambda^H$ a $K \times 1$ vector of corresponding factor risk premia and $\eta$ an asset specific expected return component. Note that by construction in this statement the pricing error on traded factor excess returns is zero, i.e., we can naturally treat traded factors as sure assets in our S–SDF setting. On the other hand, $\alpha$ in equation (26) has the interpretation of an expected excess return component not generated by an exposure to observed traded factor risk, but rather by exposures to hidden systematic factor risks and some form of asset mispricing. These properties motivate following notion of an APT–consistent S–SDF.

**Definition 3 (APT–consistent S–SDF).** In the above economies, an S–SDF $\tilde{M}$ is APT–consistent if:

1. It exactly prices the risk-free return $R_f$ and the vector of traded factor returns, i.e, $E[\tilde{M} R_f] = 1$ and $E[\tilde{M} F] = 1$.

2. It implies pricing errors $\eta := E[\frac{\tilde{M}}{E[M]}(R^e - \beta F^e)]$ that satisfy the APT bound (26) for some $\tau > 0$.

In summary, APT–consistent S–SDFs induce traded factor risk premia that exactly reproduce the physical expected excess returns of traded factors. In parallel, they induce a particular expected component in the excess returns orthogonal to traded factor risk, which reflects exposures to hidden factor risks and asset mispricing consistently with the APT bound (26). More precisely, borrowing from identity (6) we easily obtain following asset expected excess return decomposition for APT–consistent S–SDFs:

$$
E[R^e] = -Cov \left( \frac{\tilde{M}}{E[M]}, \beta F^e \right) - Cov \left( \frac{\tilde{M}}{E[M]}, R^e - \beta F^e \right) + \eta .
$$

(27)

In the above equation, the first component on the RHS is the risk premium component for exposure to traded factor risk and it equals $\beta E[F^e]$. The second component is a risk premium component due to a potential exposure to hidden systematic risk factors and it is equal to $E[(1 - \frac{\tilde{M}}{E[M]})(R^e - \beta F^e)]$. The third component $\eta = E[\frac{\tilde{M}}{E[M]}(R^e - \beta F^e)]$ captures the part of expected excess returns due to (bounded) asset mispricing.
where for some $\tau > 0$: \(^{25}\)

$$
\Sigma^{-1/2} \mathbb{E}[\tilde{\mathcal{M}}(\mathbf{R}^e - \beta\mathbf{F}^e)] \leq \mathbb{E}[\tilde{\mathcal{M}}] \tau .
$$

(28)

Given that an APT–consistent S–SDF exactly prices observable traded factor returns, it is convenient to decompose asset excess returns into the two orthogonal components that are spanned and unspanned by traded factor risk, respectively. To this end, we introduce the projection of excess returns on the space orthogonal to the traded factor excess return space:

$$
\mathbf{R}^e := \mathbf{R}^e - \text{proj}_{\text{span}F^e}(\mathbf{R}^e) .
$$

(29)

Following Definition 3, we can finally define by $\mathbf{R}_S := (R_f, \mathbf{F}')'$ the vector of sure asset returns and by $\mathbf{R}_D := R_f \mathbf{1} + \mathbf{R}^e$ the vector of dubious asset returns for an APT–consistent S–SDF. With this notation, the notion of an APT–consistent minimum dispersion S–SDF naturally follows.

**Definition 4 (Minimum dispersion APT–consistent S–SDF).** Let sure and dubious asset return vectors be given by $\mathbf{R}_S := (R_f, \mathbf{F}')'$ and $\mathbf{R}_D := R_f \mathbf{1} + \mathbf{R}^e$. An APT–consistent minimum $\Phi$–dispersion S–SDF is defined for some $\tau > 0$ by the solution of a constrained optimization problem:

$$
K(\tau) := \inf_{M \in \mathcal{M}} \left\{ \mathbb{E}[\phi(M)] : h(\mathbb{E}[M \mathbf{R}_D] - \mathbf{1}) \leq \tau \right\} ,
$$

(30)

in which pricing error metric $h : \mathbb{R}^N \to (-\infty, +\infty]$ is such that for any pricing error vector $\eta_D \in \mathbb{R}^N$:

$$
h(\eta_D) \geq h_2(\eta_D) := \Sigma^{-1/2} \eta_D \Sigma^{-1/2} .
$$

(31)

In other words, a minimum dispersion S–SDF is APT–consistent when its pricing error metric is an upper bound for the pricing error metric $h_2$ that defines the APT bound (26).

Beside pricing error metric $h_2$ in equation (31), natural examples of pricing error metrics inducing APT–consistent minimum dispersion SDFs that are used in Section 4 are, e.g.,

$$
h_1(\eta_D) := (1 - \lambda) \Sigma^{-1/2} \eta_D + \lambda \Sigma^{-1/2} \eta_D
$$

(32)

and

$$
h_\infty(\eta_D) := (1 - \lambda) \sqrt{N_D} \Sigma^{-1/2} \eta_D + \lambda \Sigma^{-1/2} \eta_D ,
$$

(33)

where $\lambda \in [0, 1]$ and $N_D$ is the number of dubious assets.\(^{26}\)

\(^{25}\)In all cases where only excess returns have to be priced, the natural normalization $\mathbb{E}[\tilde{\mathcal{M}}] = 1$ applies. More generally, when a risk free asset is priced exactly, bound (26) holds for $\tau > 0$ if and only if bound (28) holds for $\tilde{\tau} := \mathbb{E}[\tilde{\mathcal{M}}] \tau > 0$, i.e., both bounds are equivalent. Therefore, for simplicity we do not explicitly distinguish between the threshold values $\tau$ and $\tilde{\tau}$ in these two bounds in our developments on APT–consistent S–SDFs below.

\(^{26}\)APT–consistency of these pricing error metrics directly follows from the standard norm inequalities $\| \cdot \|_2 \leq \| \cdot \|_1$ and $\| \cdot \|_2 \leq \sqrt{N_D} \| \cdot \|_\infty$. 

---

---
Figure 4 illustrates the relation between these APT–consistent pricing error metrics for a setting with two-dimensional pricing errors such that $\Sigma = I$. Here, the unit ball under function $h_1$ is actually the unit ball under a convex combination of $l_1$–norm and $l_2$–norm, which is included in the unit ball under $h_2$, i.e., the unit ball under the $l_2$–norm. Similarly, the unit ball under function $h_\infty$, which is a convex combination of a scaled $l_\infty$–norm and the $l_2$–norm, is also included in the unit ball under the $l_2$–norm. Therefore, these unit balls are all consistent with a common upper bound with respect to the $l_2$–pricing error function. However, recall that they may imply potentially very different sparsity properties for admissible APT–consistent pricing errors.

Using our S–SDF theory of Section 2 and the fact that for $i = 1, 2, \infty$ pricing error function $h_i$ is a norm, the dual portfolio problem for constrained, APT–consistent, S–SDF problem (30)–(31) is easily written as:

$$\Delta = \min_{\theta \in \mathbb{R}^N} \{E[(\phi_+)^*(-R^T\theta)] + 1^T\theta + \tau h_1(\theta_D)\} ,$$

with corresponding dual norms given by:\textsuperscript{27}

(i) When $\lambda = 1$:

$$h_1(\theta_D) = h_\infty(\theta_D) = h_2(\theta_D) = \Sigma^{1/2} \theta_D .$$

(ii) When $\lambda = 0$:

$$h_1(\theta_D) = \Sigma^{1/2} \theta_D ; \quad h_\infty(\theta_D) = \Sigma^{1/2} \theta_D / \sqrt{N_D} .$$

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

$$h_1(\theta_D) = \min_{z \in \mathbb{R}^N} \max \left\{ \frac{\Sigma^{1/2} \theta_D \infty}{1 - \lambda}, \frac{\Sigma^{1/2}(\theta_D - z) Z}{\lambda} \right\} ,$$

$$h_\infty(\theta_D) = \min_{z \in \mathbb{R}^N} \max \left\{ \frac{\Sigma^{1/2} \theta_D 1}{\sqrt{N_D(1 - \lambda)}}, \frac{\Sigma^{1/2}(\theta_D - z) Z}{\lambda} \right\} .$$

Using this family of pricing error functions, in Section 4 we study in detail the empirical properties of APT–consistent S–SDFs implied by various choices of Cressie-Read $\Phi$–dispersions in Examples 1 of Appendix A.

\textsuperscript{27}Note that in solving dual problem (34) there is no need to compute the inverse of matrix $\Sigma$, which is a useful feature when working with large asset markets.
3 Estimation and inference for minimum dispersion S–SDFs

In this section, we study estimation and inference for S–SDF dispersion bounds and dual portfolio weights. Consider to this end a time series \( \{(X_{t+1}, P_t)\}_{t \in \mathbb{N}} \) of payoffs and quoted prices of the \( N \) basis assets that are defined on the filtered probability space \((\Omega, \{\mathcal{F}_t\}_{t \in \mathbb{N}}, \mathbb{P})\). The \( N \)-dimensional vectors of payoffs \( X_{t+1} \) and reference prices \( P_t \) are observed at time \( t + 1 \) and time \( t \), respectively. We develop estimation and inference for our setting under standard stationarity and ergodicity assumptions. Similarly to Hansen et al. [1995], we consider \((X, P)\) to be the starting point \((X_1, P_0)\) of our time series.

Assumption 3. Stochastic process \( \{(X_{t+1}, P_t)\}_{t \in \mathbb{N}} \) is stationary and ergodic. Further, \( \mathbb{E}[P_t] = \mathbf{P} \in \mathbb{R}^N \).

Let function \( Q: \mathbb{R}^N \to (-\infty, +\infty] \) be defined for any portfolio weight \( \theta \) by:

\[
Q(\theta) := \mathbb{E} \left[ \phi^*_+(-X'\theta) \right] = P'\theta + \psi^*(\theta_D) ,
\]

which is the objective function corresponding to the population dual portfolio problem (16). Denoting with \( \Theta \) the effective domain of function \( Q \), we obtain:

\[
\Delta = \min \{ Q(\theta) : \theta \in \Theta \} ,
\]

which, from our previous duality findings, corresponds to the minimum \( \Phi \)-dispersion attainable by an S–SDF.

Given a sample of size \( T > 0 \), the empirical analogue to population function (35) is a function \( Q_T: \mathbb{R}^N \to (-\infty, +\infty] \), given for any portfolio weight \( \theta \) by:

\[
Q_T(\theta) := \frac{1}{T} \sum_{t=1}^{T} \phi^*_+(-X'_{t+1}\theta) + P'_t\theta + \psi^*(\theta_D) .
\]

Accordingly, we define

\[
\Delta_T := \min \{ Q_T(\theta) : \theta \in \mathbb{R}^N \} ,
\]

as the natural estimator for population value \( \Delta \).

3.1 Consistency properties

We study the asymptotic properties of estimating sequence \( \Delta_T \) for \( \Delta \), under following standard assumption on the set of minimizers of function \( Q \).
**Assumption 4.** The set of minimizers of $Q$ is a non-empty compact subset of the interior of $\Theta$.

Assumption 4 naturally rules out pathological cases in which $\Delta = +\infty$. Consistent estimation of the optimal portfolio weight $\theta_0 := \arg\min\{Q(\theta) : \theta \in \Theta\}$ corresponding to dual portfolio problem (16) is based on estimator $\theta_T \in \arg\min \{Q_T(\theta) : \theta \in \mathbb{R}^N\}$. We prove consistency of this estimator for $\theta_0$ under following standard identification assumption.

**Assumption 5.** $\theta_0$ is the unique minimizer of $Q$.

In line with Proposition 3(ii), a sufficient condition for both $Q$ and $Q_T$ to have unique minimizers is, e.g., that vector $X_{t+1}$ consists of linearly independent payoffs and that $\phi^*_+ \neq 0$ is strictly convex on its domain. Consistency of estimators for S-SDF dispersion bounds and dual portfolio weights is stated in the next result.

**Proposition 6 (Consistent estimation of $\Delta$ and $\theta_0$).** Suppose Assumptions 1 and 3–4 are satisfied. (i) Then $\Delta_T$ converges to $\Delta$ almost surely as $T \to \infty$. (ii) Let additionally Assumption 5 be satisfied, then $\theta_T$ converges to $\theta_0$ in probability as $T \to \infty$.

### 3.2 Asymptotic distributions

We next study the asymptotic distribution of S–SDF dispersion bound estimator $\Delta_T$ and the corresponding dual portfolio weight estimator $\theta_T$. For the sequel, it is convenient to make use of differentiable penalization functions $\psi^*$, as our methodology in Section 2.4 for modelling the S-SDF pricing error geometry allows us to smoothly approximate also non-smooth dual penalization functions, starting from primal penalization functions of the form (19) with $\alpha > 0$.

**Assumption 6.** Dual penalty $\psi^*$ is of type (20) with $\alpha > 0$ and dual dispersion $\phi^*_+$ is continuously differentiable in a neighborhood of $\theta_0$.

Differentiability condition on $\phi^*_+$ in Assumption 6 is verified, e.g., for primal dispersion functions in the Cressie-Read family of Examples 1 in Appendix A.

---

28Note that the set of minimizers of function $Q$ is a singleton and therefore trivially compact under the assumptions of Proposition 2(ii). As discussed above, this covers the class of Cressie-Read $\Phi$-dispersions in Examples 1 of Appendix A, except for case (3a), which includes Hansen-Jagannathan dispersion. Proposition 10 of Appendix B.4 shows which conditions are needed to obtain compactness of the set of minimizers of function $Q$ for this last case.

29It is not necessary for our consistency results that the set of minimizers of $Q_T$ is a singleton. On the other hand, we need this set to be not empty for some sufficiently large $T$, which follows from Brown et al. [1973] because of the lower semi-continuity of $\phi^*_+$ and $\psi^*$ and the properness of function $Q$ under Assumption 4. Note that a proper and convex function is lower semi-continuous if and only if it is closed.

30The proofs of the results in this section are collected in Appendix B.4.

31For instance, a not everywhere differentiable lasso dual penalization induced by a primal $l_\infty$ pricing error geometry is well approximated by a continuously differentiable Huber dual penalization (A-11). Similarly, a primal elastic-net pricing error penalization directly induces a continuously differentiable dual penalization function.
3.2.1 Asymptotic distribution of estimators for S–SDF dispersion bounds

To obtain the asymptotic distribution of S–SDF dispersion bounds estimators, we first assume interchangeability between differentiation and integration in gradient $\nabla E[\phi^*_+(-X'_{t+1}\theta) + P_t'\theta]_{\theta=\theta_0}$; see, e.g., [Newey and McFadden, 1994, Lem. 3.6]. Any assumption allowing such an interchangeability can be used instead of Assumption 7 below.

**Assumption 7.** Given $\theta_0$, the unique minimizer of $Q$, and the continuous differentiability of $\phi^*_+$ in an open neighborhood of $\theta_0$, denoted by $\mathcal{N}(\theta_0)$, the following condition holds:

$$E \left[ \sup_{\theta \in \mathcal{N}(\theta_0)} P_t - (\phi^*_+)'(-X'_{t+1}\theta_0)X_{t+1} \right] < \infty.$$  

The high-level Central Limit Theorem assumption for the asymptotic distribution of the S–SDF dispersion bound estimator reads as follows.

**Assumption 8.** The sequence of random variables

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \{ \phi^*_+(-X'_{t+1}\theta_0) + P_t'\theta_0 \} - E[\phi^*_+(-X'_{t+1}\theta_0) + P_t'\theta_0]$$

converges in distribution to a normally distributed random variable with mean zero and strictly positive variance

$$v(\theta_0) := \lim_{T \to \infty} \text{Var} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \{ \phi^*_+(-X'_{t+1}\theta_0) + P_t'\theta_0 \} - E[\phi^*_+(-X'_{t+1}\theta_0) + P_t'\theta_0] \right]. \tag{39}$$

With the above conditions, the asymptotic Gaussian distribution of dispersion bound estimator $\Delta_T$ follows with standard methods, using an additional boundedness in probability assumption.

**Proposition 7 (Asymptotic distribution of dispersion bound estimator $\Delta_T$).** Let Assumptions 1, and 3–8 be satisfied and the sequence of random vectors

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \{ P_t - (\phi^*_+)'(-X'_{t+1}\theta_0)X_{t+1} \} - \text{E}[P_t - (\phi^*_+)'(-X'_{t+1}\theta_0)X_{t+1}]$$

be bounded in probability. Then, $\sqrt{T}(\Delta_T - \Delta)$ converges in distribution as $T \to \infty$ to a normally distributed random variable with mean zero and variance $v(\theta_0)$ given in equation (39).32

Despite the presence of regularization, the asymptotic distribution of estimator $\Delta_T$ is Gaussian, because of the convexity and differentiability under Assumption 6 of the objective function defining the bound in the penalized dual optimal portfolio problem. The boundedness in probability of sequence (40) follows, e.g., under the assumptions of a suitable Central Limit Theorem, such as the one in Assumption 10 below.

---

32 The same result can be obtained for general dual penalty function $\psi$, by assuming that it is continuously differentiable in a neighborhood of $\theta_0$ instead of Assumption 6.
Boundedness in probability of sequence (40) is typically assumed to obtain the asymptotic distribution of estimators for dispersion bounds of SDFs that are subject only to exact pricing constraints. Similarly, Central Limit Theorems of the type of the one stated in Assumption 8 are also assumed in such settings. Therefore, our asymptotic results in Proposition 7 follow essentially by extendig the standard assumptions in unpencilized settings by Assumptions 6. Such assumption is consistent with our general approach in Section 2.4 for modelling S–SDF pricing error geometries with primal pricing error metrics of the form (19), which always induce smooth dual penalization functions when \( \alpha > 0 \) and can in practice accurately approximate also non smooth penalizations using a sufficiently small such parameter.

### 3.2.2 Asymptotic distribution of estimator for S–SDF dual portfolio weights

To derive the asymptotic distribution of estimator \( \theta_T \) for S–SDF dual optimal portfolio weights vector \( \theta_0 \), recall that the objective function of the corresponding dual penalized portfolio problem is not in general twice differentiable, because penalization function \( \psi^* \) is typically not twice differentiable everywhere. However, we can still work under a sufficient degree of generality using the standard high-level assumption of stochastic equicontinuity.

**Assumption 9.** Following stochastic equicontinuity condition holds for any \( u \in \mathbb{R}^N \):

\[
Q_T \left( \theta_0 + \frac{u}{\sqrt{T}} \right) - Q_T(\theta_0) - \nabla Q_T(\theta_0) \cdot \frac{u}{\sqrt{T}} - \left( Q \left( \theta_0 + \frac{u}{\sqrt{T}} \right) - Q(\theta_0) \right) = o_p(1/T).
\]

Stochastic equicontinuity Assumption 9 allows us to derive the asymptotic distribution of portfolio weight estimator \( \theta_T \) without requiring twice differentiability of the sample objective function \( Q_T \).\(^{33}\) Clearly, in settings where \( Q_T \) and \( Q \) are twice differentiable Assumption 9 holds. High-level conditions implying stochastic equicontinuity are provided, e.g., in Andrews [1994].

In order to obtain the asymptotic distribution of estimator \( \theta_T \), we finally need a standard high-level Central Limit Theorem assumption for the gradient of the objective function in the unpencilized dual portfolio problem, which is complemented by a nondegeneracy assumption for the corresponding Hessian matrix of the expectaton.

**Assumption 10.** The sequence of random vectors (40) converges in distribution as \( T \to \infty \) to a zero mean vector.

\(^{33}\) Stochastic equicontinuity Assumption 9 can also be more commonly stated as follows. Let \( R_T^2(u) := Q_T \left( \theta_0 + \frac{u}{\sqrt{T}} \right) - Q_T(\theta_0) - \nabla Q_T(\theta_0) \cdot \frac{u}{\sqrt{T}} - \left( Q \left( \theta_0 + \frac{u}{\sqrt{T}} \right) - Q(\theta_0) \right) \). Stochastic equicontinuity then requires that for any sequence \( \delta_n \to 0 \) following condition holds: \( \sup_{\|u\| \leq \delta_n} \frac{\sqrt{T} R_T^2(u)}{\|u/\sqrt{T}\|} = o_p(1) \).
normally distributed random vector with positive definite variance covariance matrix:

\[
V(\theta_0) := \lim_{T \to \infty} \text{Var} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left\{ P_t - (\phi^*_t)'(-X'_{t+1}\theta_0)X_{t+1} - \mathbb{E}[P_t - (\phi^*_t)'(-X'_{t+1}\theta_0)X_{t+1}] \right\} \right].
\]  

(41)

Moreover, matrix \( G(\theta_0) := \nabla^2 \mathbb{E}[\phi^*_t(-X'_{t+1}\theta_0) + P_t]\theta = \theta_0 \) exists.

In Assumption 10, we require twice differentiability of convex function \( \theta \mapsto \mathbb{E}[\phi^*_t(X'_{t+1}\theta) - P_t\theta] \) in \( \theta_0 \), which does not necessarily require twice differentiability of \( \phi^*_t \) everywhere, due to the smoothing properties of the expectation operator. Note that the convexity of \( \phi^*_t \) directly yields positive semi-definiteness of matrix \( G(\theta_0) \). Positive definiteness of this matrix further follows under Assumption 5.

Under the above conditions, the asymptotic distribution of dual optimal portfolio weight estimator \( \theta_T \) is characterized as follows.

**Proposition 8 (Asymptotic distribution of optimal portfolio weight estimator \( \theta_T \)).** Let Assumptions 1, 3–7 and 9–10 be satisfied. Then, random sequence \( \sqrt{T}(\theta_T - \theta_0) \) converges to random variable

\[
u_0 = \arg \min_u Z(u),
\]

in distribution as \( T \to \infty \), where

\[
Z(u) := u'Y + \frac{1}{2} u'G(\theta_0)u + R(u),
\]

(42)

\( Y \) is a zero mean normally distributed random vector with variance covariance matrix \( V(\theta_0) \) and \( R(\cdot) \) is the limit function of the sequence \( TR_T(\cdot) \), with:

\[
R_T(u) := \psi^* \left( \theta_0D + \frac{u_D}{\sqrt{T}} \right) \left( \psi^*(\theta_0D) - \nabla \psi^*(\theta_0D) \frac{u_D}{\sqrt{T}} \right).
\]

(43)

The convergence of the process \( TR_T \) is implied by Assumption 6. From Proposition 8, \( \sqrt{T}(\theta_T - \theta_0) \) is not in general asymptotically normally distributed, due to the contribution of function \( R(u) \) in equation (42).

It is easy to see that the assumption of twice continuous differentiability of penalization function \( \psi^* \) directly implies a corresponding limit function \( R \) given by \( R(u) = \frac{1}{2} u_D' \nabla^2 \psi^*(\theta_0D)u_D \) for any \( u \in \mathbb{R}^N \). Therefore, under this additional assumption \( \sqrt{T}(\theta_T - \theta_0) \) is asymptotically normally distributed with mean zero and covariance matrix \( (J^{-1}(\theta_0))V(\theta_0)(J^{-1}(\theta_0))' \), where:

\[
J(\theta_0) = G(\theta_0) + \left[ \begin{array}{cc} 0_{N_D \times N_D} & 0_{N_D \times N_D} \\ 0_{N_D \times N_D} & \nabla^2 \psi^*(\theta_0D) \end{array} \right].
\]

Such a setting arises, e.g., in equation (19) for the pure Ridge regularization where \( f = 0 \) and \( \alpha > 0 \). Moreover, for most penalty specifications in equation (19) the set of points in \( \mathbb{R}^{N_D} \) where \( \psi^* \) is not twice

\[34\] The same result can be obtained for general dual penalty function \( \psi^* \) by assuming that \( TR_T \) converges pointwise to limit function \( R \) as \( T \to \infty \).
differentiable is easy to identify. For instance, when \( f = \delta_C \) and \( C \) is a closed convex cone, this set is given by the border of \(-C^*\). Similarly, when \( f = \| \cdot \| \) this set is given by the border of the unit ball of the dual norm \( \| \cdot \|_\ast \).

### 3.3 Testing S–SDF specifications using dispersion bounds

Under the null of the existence of possible pricing errors on dubious assets, we can develop natural corresponding diagnostics for asset pricing models, which may incorporate non zero pricing error bounds with respect to relevant pricing error geometries. To this end, suppose there is available the time series of S–SDF realizations \( \{ M_{t+1}^\ast \}_{t=1}^T \) induced by a given S–SDF specification \( M^\ast \) in a corresponding asset pricing model. Given a pricing error function \( h \in \Gamma(\mathbb{R}^{N_D}) \) and a corresponding pricing error threshold \( \tau \geq 0 \), we are interested in diagnostics for asset pricing models incorporating following null hypothesis:

\[
\mathcal{H}_0 : \mathbb{E}[M^\ast X] - P^S = 0 \quad \text{and} \quad h(\mathbb{E}[M^\ast X] - P^D) \leq \tau . \tag{44}
\]

The null hypothesis of correct SDF specification under exact pricing in Hansen and Jagannathan [1991] is obtained as a special case when, e.g., \( \tau = 0 \) and \( h \) is a norm. Similarly, the null hypothesis in Hansen et al. [1995] of correct SDF specification in presence of short-selling constraints and bid-ask spreads is obtained by setting \( h = \delta_{-\mathbb{R}_+^{N_D}} \). However, our testing framework provides model diagnostics for much broader pricing error geometries corresponding to any function \( h \in \Gamma(\mathbb{R}^{N_D}) \) or, equivalently, much broader transaction cost specifications induced by any function \( h^\ast \in \Gamma(\mathbb{R}^{N_D}) \). A natural diagnostics for asset pricing models in our setting simply tests whether an S–SDF \( M^\ast \) satisfies the minimum S–SDF dispersion bounds introduced in Section 2.3, given that constraint (44) holds and while parametrizing the bound as a function of the bond price \( \mathbb{E}[M^\ast] \) in a bond market without frictions. We obtain standard asymptotically normal bounds using our modelling approach with Moreau envelopes from equation (19).

The relevant primal minimum dispersion S–SDF problem (15) incorporating constraint (44) and the exact pricing constraint in bond markets is given by:

\[
\Pi(\tau) := \inf_{M \in \mathcal{M}} \{ \mathbb{E}[\phi(M)] + \frac{\alpha}{2} \| \mathbb{E}[M X^D] - P^D \|_2^2 : \mathbb{E}[M] = \mathbb{E}[M^\ast], \ h(\mathbb{E}[M X^D] - P^D) \leq \tau \} .
\]

From Proposition 2, we obtain \( \Pi(\tau) = -\Delta(\tau) \), with the corresponding penalized portfolio problem given by:

\[
\Delta(\tau) := \min_{\gamma \in \mathbb{R}, \theta \in \mathbb{R}^N} \{ \mathbb{E}[\hat{\varphi}_\gamma(\gamma - X^T \theta)] + \gamma \mathbb{E}[M^\ast] + P^D \theta + \psi^\ast(\theta^D) \} ,
\]

29
where, from Corollary 2 (ii) in Appendix B.3, penalization function \( \psi^* \) is given by:

\[
\psi^*(\theta_D) = \frac{1}{2\alpha} \left( \|\theta_D\|_2^2 - d_{C(\alpha\tau)}^2(\theta_D) \right),
\]

with sublevel set \( C(\alpha\tau) := \{ \eta \in \mathbb{R}^D : h(\eta) \leq \alpha\tau \} \). By construction, under null hypothesis (44), following null hypothesis also holds:

\[
\mathcal{H}_0' : \xi(\tau) := \Delta(\tau) + \frac{1}{\tau} \sum_{t=1}^{T} \phi(M^*_{t+1}) + \frac{\alpha}{2} \| \mathbb{E}[M^* X_D] - P^D \|_2^2 \geq 0,
\]

(45)

where the sample version of population quantity \( \xi(\tau) \) is the statistic:

\[
\xi_T(\tau) := \Delta_T(\tau) + \frac{1}{\tau} \sum_{t=1}^{T} \phi(M^*_{t+1}) + \frac{\alpha}{2} \sum_{t=1}^{T} (M^*_{t+1} X^D_{t+1} - P^D_t)^2.
\]

(46)

By suitably adjusting the arguments in the proof of Proposition 7, we obtain that \( \sqrt{T}(\xi_T(\tau) - \xi(\tau)) \) converges in distribution to a normally distributed zero-mean random variable with variance:

\[
v(\theta_0) = \lim_{T \to \infty} Var \sqrt{T}(\Delta_T(\tau) - \Delta(\tau)) + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left( \phi(M^*_{t+1}) - \mathbb{E}[\phi(M^*_{t+1})] \right) + \frac{\alpha}{2} \sqrt{T}(\Gamma_T - \Gamma),
\]

(47)

where

\[
\Gamma_T - \Gamma := \frac{1}{\tau} \sum_{t=1}^{T} \mathbb{E}[M^*_{t+1} X^D_{t+1} - P^D_t]^2 \left\{ M^*_{t+1} X^D_{t+1} - P^D_t - \mathbb{E}[M^*_{t+1} X^D_{t+1} - P^D_t] \right\}.
\]

(48)

In equation (47), the first term isolates the variance contribution of the generalized minimum S–SDF dispersion bound estimators \( \Pi_T(\tau) := -\Delta_T(\tau) \), the second term the contribution of the estimated dispersion of S–SDF \( M^* \) and the third term the contribution of the estimated pricing error penalization for S–SDF model \( M^* \), which derives from the Ridge regularization in our modelling approach (19) with Moreau envelopes.

We can now formulate a model diagnostics testing procedure based on the composite null hypothesis (45).

The most conservative critical value for statistic (46) under composite null hypothesis (45) is the one implied by the simple null hypothesis \( \xi(\tau) = 0 \). Given this simple null, we can use the above asymptotic distribution to construct conservative asymptotic critical values for null hypothesis (45).

Figure 5 reports the estimated dispersion bound \( \Pi_T(\tau) \), using various values of threshold \( \tau \) and target bond price \( \mathbb{E}[M^*] \), using the low-dimensional dataset described in our empirical analysis of Section 4. The bounds decrease rapidly already when introducing small pricing error thresholds \( \tau > 0 \). The bound sensitivity with respect to the target bond price level \( \mathbb{E}[M^*] \) also significantly decreases as small pricing errors are allowed. Overall, this evidence highlights that asset pricing models not satisfying minimum SDF dispersion bounds
implied by exact pricing may well satisfy generalized S–SDF minimum dispersion bounds incorporating pricing error properties that are economically plausible.

4 Empirical Analysis: S–SDFs in the APT Framework

In order to exhibit the usefulness of S–SDFs in analysing asset pricing models, this section explores the empirical properties of minimum dispersion S–SDFs under an Arbitrage Pricing Theory setting, in which economic pricing error bounds are motivated by the absence of asymptotic arbitrage opportunities; see again Section 2.5. We estimate various minimum dispersion S–SDFs consistent with APT pricing error bounds and study (i) their pricing properties, (ii) the implications of different relevant dispersion metrics and geometries for pricing errors and dual portfolio weights, and (iii) the tightness of the resulting minimum dispersion tradeoffs with respect to the degree of pricing error and portfolio weight sparsity, from both an in-sample and an out-of-sample standpoint. Finally, we study the performances of portfolio strategies supporting APT–consistent S–SDFs.

4.1 Dataset

Our empirical evidence relies on standard databases of different dimensions, in order to easily benchmark our results with related empirical evidence in the literature. We use three datasets that differ in the dimension of the set of assets considered, in order to better isolate the empirical merits of APT–consistent 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 considered are 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 considered are the risk-free return, 100 portfolio returns sorted on size and book to market, 25 portfolio returns sorted on momentum, 25 portfolio returns 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. **Augmented intermediate 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 and French industry classification. The construction of the factor returns follows Nagel et al. [2018].

We also consider five Fama-French factor returns for constructing simple benchmark linear SDFs. Data availability allows accessing a sample of asset returns from January 1931 (January 1970 for the characteristic-based factors in the augmented intermediate dimensional dataset) until 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 augmented intermediate dataset has 260 asset returns with 561 time series observations.

4.2 In–sample properties of APT–consistent minimum dispersion S–SDFs

We first explore the in-sample properties of APT–consistent S–SDFs induced by their deep characteristics: the S–SDF dispersion criterion (function $\phi$) and the S–SDF pricing error geometry (function $h$). We find that allowing pricing errors on a subset of assets can help (i) restoring the SDF dispersion to more realistic levels and (ii) avoiding to reach the zero lower bound for an S–SDF when variance is used to measure dispersion. These two S–SDF features are more easily attained by pricing error geometries with dense pricing errors than by sparsity inducing geometries, showing that pricing error sparsity can imply a large price in terms of minimal S–SDF dispersion. We also find that the importance of introducing a concern for higher order risks by means of S–SDF dispersion criteria other than variance diminishes as pricing errors are allowed.

Figure 6 shows the tradeoff between minimum S–SDF dispersion and APT-type pricing error bounds, for various levels of parameter $\tau$, under the family of APT–consistent S–SDFs induced by portfolio weight penalizations $h_i^*$ in Section 2.5 for $i = 1, 2, \infty$. Results in this section are based on the intermediate dimensional dataset and variance as a measure of S–SDF dispersion, where we take the risk-free rate as the only sure asset. By construction, the minimum S–SDF variance decreases in the APT–induced pricing error bound $\tau$ for all penalization choices in Section 2.5. However, this tradeoff is different for S–SDFs induced by an $l_1$–pricing error metric, i.e., APT–consistent S–SDFs with sparse pricing errors require a proportionally higher variance to satisfy the same APT pricing error bound. By construction, there is a tradeoff also for minimum variance S–SDFs induced by a (scaled) $l_\infty$–pricing metric with dense pricing errors. However, this second tradeoff appears as clearly less tight than for the $l_1$–pricing error metric.

Remark 1. Note that the tight tradeoff in Figure 6 between minimum S–SDF dispersion and APT-type
pricing error bounds is not explained by estimation uncertainty. To show this concretely, we can make use of our regularization approach in Section 2.4 based on Moreau envelopes, in order to derive the asymptotically normal confidence intervals for minimum \( S\text{-SDF} \) dispersion estimates with Proposition 7. Indeed, recall that given a penalty function of the form in equation (19), \( \sqrt{T}(\Delta_T - \Delta) \) converges in distribution to a zero mean normally distributed random variable, having an asymptotic variance \( v_0(\theta_0) \) in equation (39) that can be estimated consistently, e.g., with a Newey-West estimator. While Proposition 7 does not directly apply to the non smooth penalization function \( f^*_i := h^*_i \) in dual portfolio problem (34), we can make use of approach (19) and Proposition 5 to obtain accurate smooth approximations of dual portfolio problem (34), which allows us to apply the asymptotic distribution in Proposition 7. Figure 7 reports estimated APT–consistent minimum \( S\text{-SDF} \) variances based on both a non smooth penalization \( f^*_i := h^*_i \) (dashed line) and the corresponding smooth approximation \( e_0 f^*_i \) (solid line) provided in Proposition 5 by Moreau envelopes. The right panel shows that for a sufficiently low parameter value \( \alpha = 0.002 \), both minimum \( S\text{-SDF} \) dispersion curves essentially overlap.\(^{35}\) While the width of the confidence intervals provided by the smooth approximation decreases with the pricing error bound \( \tau \), both the upper and lower 90% confidence bands decrease rapidly with the pricing error bound \( \tau \), confirming the tight tradeoff between \( S\text{-SDF} \) dispersion and APT-type pricing error bounds.

Using the APT pricing error threshold highlighted by the vertical line in Figure 6, we next illustrate in more detail, in the three rows of Figures 8 and 9, the deep properties of APT–consistent \( S\text{-SDF} \). In order to show explicitly the different geometries induced by the \( l_1 \), \( l_2 \) and the (scaled) \( l_\infty \) APT–consistent pricing error metrics, we report in all panels of Figure 8 pricing errors in their standardized APT metric, i.e., we report the empirical version of vector \( \Sigma^{-1/2}(\mathbb{E}[M_0 R_D] - 1) \). Similarly, optimal portfolio weights are reported in their penalized APT metric, i.e., we report the empirical counterpart of \( \Sigma^{1/2} \theta_0D \).

In Figure 8, sparse pricing errors under the \( l_1 \)–pricing error metric are linked to dense portfolio weights and to the largest \( S\text{-SDF} \) variance. On the other hand, dense pricing errors under the (scaled) \( l_\infty \)–pricing error metric are linked to sparse portfolio weights and the second largest \( S\text{-SDF} \) variance. Finally, the \( l_2 \)–pricing error metric induces an \( S\text{-SDF} \) with dense portfolio weights and pricing errors. From Figure 6, this \( S\text{-SDF} \) has a moderately higher variance and lower pricing errors, in the APT pricing error metric, than the minimum variance SDF pricing exactly only the sure assets. It also implies a qualitatively identical pattern in APT pricing errors and dual portfolio weights, due to both the self-duality of the \( l_2 \)–pricing error metric and the linearity of minimum variance \( S\text{-SDF} \) in their optimal dual portfolio returns.

Figure 9 reports the time series of \( S\text{-SDF} \) estimated for the various APT–consistent pricing error metrics. \( S\text{-SDF} \) with dense portfolio weights and pricing errors imply the lowest volatility and rarely attain the zero lower bound. Their lowest volatility is the result of the portfolio weights in Figure 8, which are on average smaller in absolute value than the portfolio weights of \( S\text{-SDF} \) with sparse pricing errors or sparse portfolio weights. The larger portfolio weights implied by the two latter \( S\text{-SDF} \) follow from the more restrictive

\(^{35}\)Under a higher parameter value \( \alpha = 0.02 \), the two curves are still relatively near, especially for low pricing error bounds \( \tau \).
pricing error geometries under the $l_1$– and the (scaled) $l_\infty$–metrics, which increase the minimal S–SDF volatility. Such higher volatility also induces minimum variance S–SDFs attaining the zero lower bound more frequently.

We next address the properties of APT–consistent minimum dispersion S–SDFs linked to notions of dispersion beyond variance. To this end we further consider Kullback Leibler (KL), Negative Entropy (NE) and Hellinger (HE) dispersions, where the pricing error metrics is given by the APT-consistent $l_2$-norm. Note that while minimizing variance entails S–SDFs that are piecewise linear in the dual optimal portfolio payoffs, minimizing the KL dispersion entails S–SDFs exponential in the payoffs and minimizing NE or HE dispersion provides a hyperbolic relation between minimum dispersion S–SDFs and optimal payoffs; see Corollary 1 of Appendix A. Therefore, while minimum variance S–SDFs can attain the zero lower bound, minimum KL, NE or HE dispersion S–SDFs do not.

In Figure 10 we report the time series of minimum variance S–SDFs for different pricing error thresholds, expressed as fractions of $\hat{\tau}^{\text{max}}$, the estimate of the maximum pricing error threshold $\tau^{\text{max}}$ in equation (14). The minimum variance SDF pricing all assets exactly is highly volatile and often degenerate, i.e., it reaches the zero lower bound. In contrast, APT–consistent minimum variance S–SDFs feature by construction much less volatility and hence reach less often, if at all, the zero lower bound, when higher pricing error thresholds are allowed.

APT–consistent minimum KL divergence S–SDFs in Figure 11 are strictly positive in all cases and more positively skewed, especially for low pricing error thresholds. However, as the admissible pricing error threshold increases to, e.g., 60% of the maximal pricing error threshold $\hat{\tau}^{\text{max}}$, minimum variance and minimum KL divergence S–SDFs become more and more similar and virtually perfectly positively related. This feature holds also with respect to minimum KL and HL dispersion S–SDFs not reported in the scatter plots of Figure 11. To understand this last evidence better, Figure 12 reports the correlations between the optimal portfolio payoffs underlying minimum variance S–SDFs and the optimal portfolio payoffs underlying minimum KL, NE and HE dispersion S–SDFs, as a function of the corresponding pricing error threshold. Here, all S–SDF correlations are increasing in the pricing error threshold and the smallest correlation between minimum variance S–SDFs and the other S–SDFs is no less than 97.5% when the corresponding pricing error threshold is at least as large as $0.5 \cdot \hat{\tau}^{\text{max}}$. Overall, this evidence demonstrates that the differences in minimum dispersion S–SDFs induced by different notions of S–SDF dispersion tend to shrink in presence of
less constrained pricing errors that correspond to stronger portfolio weight penalizations.

4.3 Out-of-sample pricing properties of APT–consistent S–SDFs

We explore the out-of-sample pricing properties of purely data-driven APT–consistent S–SDFs. To this end, we sequentially estimate various minimum dispersion S–SDFs, based on several APT–consistent pricing error metrics from Section 2.5. We adopt a simple fully data-driven approach and estimate S–SDFs on rolling windows of 30 years of monthly returns, treating all returns as dubious, with the exception of the risk-free return. In this way, our approach is independent of the specification of the relevant traded risk factors in a potentially misspecified APT factor model for returns. We then evaluate the S–SDF out-of-sample pricing performance, by quantifying goodness of fit with GLS adjusted cross-sectional $R^2$—metrics, evaluated on out-of-sample returns and out-of-sample S–SDF realizations.

The main finding is that allowing for pricing errors in-sample helps to significantly improve the S–SDF out-of-sample pricing performance. Such performance clearly dominates the one of benchmark empirical asset pricing models, including three- and five-factor Fama-French models. The S–SDF out-of-sample pricing performance also clearly dominates the one of linear SDFs implied by informal regularizations based on the principal components of asset returns. Within our family of APT–consistent S–SDFs, we find that minimum variance S–SDFs based on the standard APT pricing error metric in equation (26) achieve the highest and most robust out-of-sample pricing performance. Other APT–consistent pricing error metrics or different notions of S–SDF dispersion than variance produce inferior out-of-sample pricing results.

4.3.1 S–SDF estimation setting and benchmark empirical asset pricing models

For every in-sample window of 30 years of monthly observations having last observation in June of year $y - 1$, the in-sample S–SDF estimation requires the specification of an APT pricing error threshold, denoted by $\tau_y$.\footnote{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.} We use our S–SDF duality framework to naturally determine the range of relevant empirical pricing error thresholds. We estimate the maximal admissible threshold $\hat{\tau}_y^{\max}$ with the finite-sample version of equation (14), which corresponds to the estimated pricing error threshold of a standard minimum variance SDF that exactly prices only the risk-free asset in the estimation window of 30 years previous to year $y$. Similarly, the minimal admissible threshold $\hat{\tau}_y^{\min}$ is estimated as the smallest threshold value for which in the given estimation window no empirical duality failure arises; see again the discussion in Section 2.3 following
Proposition 4.

Given estimated range $[\tilde{z}_y^{\min}, \tilde{z}_y^{\max}]$ of admissible thresholds for a fixed estimation window of 30 years of monthly observations, we estimate an optimal threshold $\hat{\tau}_y$ that maximizes the estimated GLS-adjusted cross-sectional $R^2$–metric within the given estimation window. Using the methodology introduced in Section 3, this gives rise to a sequence $\{\hat{\tau}_y, \hat{\theta}_y\}$ of estimated pricing error thresholds and S–SDF dual portfolio weights, which is updated with an annual frequency. For any given month $m$ in year $y$, we denote by $\mathbf{R}_y^{(m)}$ the vector of returns for month $m$ and estimate with Proposition 4 the corresponding sequence of out-of-sample monthly minimum dispersion S–SDFs, i.e. $\{\hat{\mathbf{M}}_y^{(m)} := (\hat{\phi}_y)^{\dagger}(\hat{\theta}_y \hat{\mathbf{R}}_y^{(m)})\}$. In this way, we obtain a monthly time series $\{\hat{\mathbf{M}}_y^{(m)}\}$ of fully data driven APT–consistent S–SDFs with no forward looking bias, for which we can evaluate the out-of-sample fit, based on standard testing methodologies for empirical asset pricing models applied to out-of-sample monthly S–SDF and asset return realizations.

We compare the out-of-sample fit of our data-driven APT–consistent S–SDFs with the following benchmarks. First, a three-factor Fama-French model with factors given by the excess returns of market, size and book-to-market portfolios. Second, a 5-factor Fama-French model that extends the three-factor model by an operational profitability and an investment factor. Third, a linear factor model with factors consisting of the first three principal components of returns estimated from out-of-sample return realizations. While this empirical asset pricing model implies a forward-looking bias due to the estimation of the covariance matrix of returns from out-of-sample observations, it provides a natural reference for assessing the out-of-sample pricing performance that is achievable by linear factor models with time-invariant factor compositions.

Our last benchmark empirical asset pricing model exploits in a different way forward-looking information from the out-of-sample period and is based on a sequence of corresponding S–SDFs. We obtain these S–SDFs using the threshold parametrization $\hat{\tau}_y(w) := w\tilde{z}_y^{\max}$, where parameter $w$ is estimated using out-of-sample data and is such that $\hat{\tau}_y(w) \geq \max_y \{\tilde{z}_y^{\min}\}$, in order to ensure existence of an empirical minimum dispersion S–SDF satisfying pricing error threshold $\hat{\tau}_y(w)$ for every estimation window.$^{37}$ We obtain the ex-post optimal sequence $\hat{\tau}_y := \hat{\tau}_y(\hat{w})$ of these SDFs using an estimated weight $\hat{w}$ that maximizes the GLS-adjusted cross-sectional $R^2$–metric with respect to the out-of-sample S–SDF and asset return realizations. While as mentioned this last empirical asset pricing model implies a forward-looking bias generated by the estimation of parameter $w$, it is a natural benchmark for assessing the out-of-sample pricing performance of data driven S–SDFs in which time-varying threshold $\hat{\tau}_y$ varies roughly proportionally to the maximal

$^{37}$See again the discussion in Section 2.3 following Proposition 4.
threshold $\hat{\tau}_{y}^{max}$.

4.3.2 Out-of-sample pricing results for APT–consistent minimum variance S–SDFs

For the intermediate dimensional dataset, Figure 13 reports various curves of out-of-sample GLS adjusted $R^2$s implied by data-driven minimum variance S–SDFs for weights $w$ in the admissible range, together with the optimal weight $\hat{w}$ and GLS adjusted $R^2$ of the optimal forward looking minimum variance S–SDF. The left and right panels report results for the APT–consistent pricing error metrics in equations (32) and (33), respectively. We find that the maximal forward-looking GLS adjusted $R^2$ in the left (right) panel ranges between a value of about 44.1% for the $l_2$ APT–consistent pricing error metric, to a value of about 32.3% (39.5%) for the $l_1$ (the scaled $l_\infty$) APT–consistent pricing error metric.\textsuperscript{38}

The above evidence indicates that data driven S–SDFs based on a time-varying pricing error bound roughly proportional to maximal threshold $\hat{\tau}_{y}^{max}$ may hardly produce out-of-sample GLS adjusted $R^2$ larger than about 45%, which is a quite high benchmark for the out-of-sample pricing performance in the intermediate dimensional dataset. For comparison, we find that the out-of-sample GLS adjusted $R^2$ of the three-factor Fama-French model is only about 6.3%, while that of the five-factor Fama-French model is only about 5.9%. In parallel, the GLS adjusted $R^2$ of a linear three-factor model with factors given by the first three principal components of returns is only about 5.7%. In contrast, we find that the out-of-sample GLS adjusted $R^2$ of APT–consistent S–SDFs without forward looking bias based on a $l_2$–pricing error metric is as large as 41.8%, which is only slightly lower than the maximal GLS adjusted $R^2$ of forward-looking APT–consistent S–SDFs. Table 1 produces additional details on the out-of-sample fit provided by the various S–SDF approaches considered above.

In summary, we conclude that our data-driven APT–consistent minimum variance S–SDFs with no forward looking bias yield a substantial out-of-sample pricing performance relative to all benchmark empirical asset pricing settings considered.

4.3.3 Out-of-sample pricing results for other APT–consistent minimum dispersion S–SDFs

Our previous analysis has focused on APT–consistent minimum variance S–SDFs for the intermediate dimensional dataset. A natural question is how the corresponding out-of-sample results depend on the notion of S–SDF dispersion used and on the dimension of the set of traded assets under scrutiny. To answer these

\textsuperscript{38}The minimal forward-looking GLS adjusted $R^2$ in all panels of Figure 13 is about 10% and corresponds to a sequence of standard minimum variance SDFs exactly pricing in each estimation window only the risk free rate.
questions, Figures 15 and 16 report the evidence obtained for minimum Kullback Leibler divergence S–SDFs, using again the set of APT–consistent pricing error metrics from Section 2.5. In each figure, the left (right) panels produce the results for the low dimensional (intermediate dimensional) dataset. Bottom panels report the evidence for minimum KL divergence S–SDFs, while top panels report for comparison the evidence based on minimum variance SDFs.

Overall, we find that minimum KL divergence S–SDFs do not improve on the out-of-sample fit of minimum variance S–SDFs. For instance, in the intermediate dimensional dataset, the maximal GLS adjusted $R^2$ under a forward-looking APT–consistent minimum KL divergence S–SDF is even lower than the GLS adjusted $R^2$ under the minimum variance APT–consistent S–SDF with no forward looking bias introduced above.

Second, we find that the region of pricing error thresholds with large GLS adjusted $R^2$ under the minimum variance APT–consistent S–SDF is quite broad. In contrast, the same region under the minimum KL divergence S–SDF is much thinner, reflecting a higher sensitivity of the out-of-sample fit of these S–SDFs with respect to the choice of the relevant APT pricing error bound. Such a large sensitivity can be a challenge for constructing fully data-driven minimum KL divergence S–SDFs with good out-of-sample fit, based on a corresponding estimated sequence of thresholds $\{\hat{\tau}^*_y\}$ with no forward looking bias.

Third, in the low-dimensional data set the optimal weight $\hat{w}$ for minimum KL divergence SDFs is basically equal to 1, indicating that the optimal S–SDF sequence actually consists of standard minimum KL divergence SDFs that exactly price in each estimation window only the risk-free rate. Related results for minimum variance S–SDFs show that the out-of-sample pricing performance is essentially independent of the pricing error threshold used. Here, the fit of standard minimum variance SDFs exactly pricing the sure assets alone or the sure and the dubious assets together is basically identical to the one of any other minimum variance S–SDF. Therefore, in the low dimensional data set the degree of portfolio weight penalization does not seem to matter much for the resulting out-of-sample fit, especially when using minimum variance SDFs based on APT-consistent pricing errors under the $l_2$ metric.

### 4.4 Optimal trading strategies supporting APT–consistent S–SDFs

The time series of our data-driven monthly minimum variance S–SDFs $\{\hat{M}^{(m)}_y\}$ is spanned by a corresponding time series of dual optimal portfolio monthly payoffs $\{\hat{\theta}^*_y R^{(m)}_y\}$, in which portfolio weights $\{\hat{\theta}^*_y\}$ are estimated on a training window of previous 30 years and rebalanced yearly. Hence, the price $\hat{\theta}^*_y 1$ of these payoffs varies
with a yearly frequency and implies, e.g., a total leverage that varies at the same frequency.

To obtain an excess return with a normalized total investment, we compute following excess return, which corresponds to a one dollar investment in the optimal portfolio financed by borrowing at the risk-free rate $R_y^{f(m)}$:

$$R_y^{e(m)} := \hat{\theta}_y' \left( R_y^{(m)} - R_y^{f(m)} \mathbf{1} \right) \left/ \hat{\theta}_y' \mathbf{1} \right..$$

(49)

Note that since excess return (49) conditionally spans S–SDF $\hat{M}_y^{(m)}$, it induces the same conditional pricing fit and is thus a natural candidate to define a single-factor linear asset pricing model based on a traded excess return.\(^{39}\)

To obtain an excess return with a uniformly bounded leverage on each asset, we can further scale $R_y^{e(m)}$ proportionally to its (time-varying) maximal allocation to individual assets, e.g., by twice the maximal allocation to individual assets, which ensures a bound of 0.5 on the maximal absolute asset allocation:

$$\tilde{R}_y^{e(m)} := \frac{\hat{\theta}_y' \left( R_y^{(m)} - R_y^{f(m)} \mathbf{1} \right) \left/ 2 \cdot \max_{i \in \{1, \ldots, N\}} |\hat{\theta}_{yi}| \right.}{\hat{\theta}_y' \mathbf{1}}.$$  

(50)

Similar to excess return (49), excess return (50) induces by construction the same conditional pricing fit as S–SDF $\hat{M}_y^{(m)}$.

Table 2 documents the out-of-sample fit of empirical asset pricing models based on single factor excess returns (49) and (50) implied by an $l_2$ APT-consistent pricing error metric, in comparison to benchmark single factor models based on the Fama-French excess returns and the excess return of an equally-weighted portfolio, respectively. Consistently with our previous findings, we find that the model based on excess return factors (49) and (50) clearly outperforms all other single factor benchmarks.

Table 3 collects summary statistics of the various excess returns underlying the single-factor models of Table 2. It shows that excess return (49) produces a monthly Sharpe ratio larger than about three times the market Sharpe ratio, the Sharpe ratio of the equally weighted portfolio or the Sharpe ratios of size and book-to-market excess returns. In contrast to the market and the equally weighted portfolio returns, excess return (49) is positively skewed. However, it also implies a higher kurtosis. The time-varying rescaling underlying the definition of excess return $\tilde{R}_y^{e(m)}$ in Table 3 is reflected in the lower expected return and volatility of

\(^{39}\)The unconditional pricing fit of $\hat{M}_y^{(m)}$ and $R_y^{e(m)}$ can in principle be different, due to the unconditional variations in portfolio weights $\hat{\theta}_y$. However, the low-frequency (yearly) character of these variations implies small differences in out-of-sample pricing, as we document below.
$\tilde{R}_y^{(m)}$ relative to $R_y^{(m)}$. While the Sharpe ratio of these two excess returns is very similar, $\tilde{R}_y^{(m)}$ dominates both the market and the equally weighted portfolio returns in a mean variance sense.

Figure 17 reports the cumulative returns of a one dollar investment in various excess returns, namely the market excess return, excess return (49) leveraged to have the same volatility as the market portfolio, excess return (50) leveraged to have the same volatility as the market portfolio, the excess return of the SMB portfolio, the excess return of the HML portfolio and the excess return of an equally-weighted portfolio of all sorted portfolios in the intermediate dimensional dataset. As is apparent from Figure 17, the optimal portfolios implied by our data-driven S–SDFs $\{\tilde{M}_y^{(m)}\}$ largely outperform all other benchmarks in terms of cumulative performance. They also produce on average lower conditional return volatilities and are less subject to downside risks during recessions or periods of financial distress.

Overall, we conclude that the high profitability of our trading strategies induced by minimum variance S–SDFs predominantly arises from their ability to optimize conditional mean-variance trade-offs and less from their dynamic leverage of timing features.

5 Conclusions

We propose a general approach to study model-free smart stochastic discount factors (S–SDFs), i.e., stochastic discount factors incorporating widespread constraints on the non-zero pricing errors of a subset of assets. To this end, we introduce minimum dispersion S–SDFs as the solutions of a general minimum SDF dispersion problem with convex constraints on some subset of nonzero pricing errors. This approach allows us to incorporate, e.g., several specifications of frictions and total wealth constraints in the literature, various useful approaches to regularize SDFs in arbitrage-free asset markets, as well as various pricing error bounds implied by the Arbitrage Pricing Theory, both under a correctly specified or a misspecified factor model for returns.

We first show that S–SDFs are generally always supported by a corresponding viable market with convex transaction costs, such as, e.g., quadratic transaction costs or constraints on total leverage. We then formally link minimum dispersion S–SDFs to the solution of a dual portfolio selection problem with penalized portfolio weights, under a uniquely determined convex penalization function. This dual penalization explicitly measures the direct effect on optimal portfolio weights of a given pricing error metric for non zero pricing errors. Equivalently, it uniquely identifies the deep S–SDF properties that are induced by a given market
friction, a particular regularization choice or a specific APT pricing error bound.

We develop the necessary methodology for the empirical analysis of S–SDFs, under a broad choice of convex pricing error constraints. In contrast to standard minimum dispersion SDFs, the dual optimization problem of minimum dispersion S–SDFs often yields a non smooth objective function. For a broad class of pricing error constraints and convex portfolio penalization, we systematically make use of Moreau envelopes to obtain tractable estimation problems that are compatible with various geometries of sparse pricing errors or sparse optimal portfolio weights.

In order to systematically construct more robust model-free stochastic discount factors motivated by plausible no-arbitrage assumptions, we make use of Cressie-Read power dispersions to obtain APT–consistent S–SDFs in asset markets where APT-type no arbitrage conditions can be invoked. We empirically quantify the tradeoff between pricing error sparsity and pricing error size in getting APT–consistent S–SDFs with a plausible dispersion. We then clarify the deep properties of pricing error and dual portfolio weight geometries induced by various APT–consistent penalization choices. Finally, we propose a systematic data-driven approach to construct APT–consistent minimum variance S–SDFs that yield substantial improvements in out-of-sample pricing performance, relative to a variety of natural empirical asset pricing benchmarks. We directly show how the large out-of-sample pricing performance of APT–consistent minimum variance S–SDFs naturally corresponds to highly profitable optimal trading strategies, which are able to very successfully optimize the conditional mean-variance trade-off without implying hardly practicable timing features.

Our S–SDF methodology is applicable to study several important topics that could not be addressed in this paper, such as the estimation of conditional S–SDFs incorporating information from time-varying asset characteristics and market frictions, the validation of macro-based asset pricing models in presence of general transaction costs, or the measurement of segmentation between markets with nonlinear (convex) pricing rules. We plan to address these topics in future research.
Appendix A - Examples of tractable $\Phi-$dispersions

This Appendix collects relevant explicit examples of $\Phi-$dispersions, based on corresponding functions $\phi$, from the the Cressie-Read family; see also Kitamura et al. [2004] and Newey and Smith [2004], among others. For the sake of exposition, we present the various relevant subcases in this family and provide for each one the corresponding parameter $d_\phi := \lim_{x \to +\infty} \phi(x)/x$ for Proposition 4 in the main text.

**Examples 1.** Consider in equation (10) functions $\phi$ implying a corresponding restriction $\phi_+$ to the nonnegative real line from the following family:

1. Kullback-Leibler dispersion:

$$\phi_+(x) = \begin{cases} \ln x - x + 1 & x > 0 \\ 0 & x = 0, \quad \text{with } \phi_+^*(y) = \exp(y) - 1, \quad \text{and } d_\phi = +\infty. \\ -\infty & x < 0 \end{cases}$$

2. Negative entropy:

$$\phi_+(x) = \begin{cases} -\ln x + x - 1 & x > 0 \\ +\infty & x \leq 0, \quad \text{with } \phi_+^*(y) = \begin{cases} \ln(1 - y), & y < 1 \\ -\infty & y \geq 1, \quad \text{and } d_\phi = 1. \end{cases} \end{cases}$$

3. Power dispersion:

(a) for $\gamma > 1$ and $\beta = \gamma/(\gamma - 1),^{40}$

$$\phi_+(x) = \begin{cases} x^{\gamma - 1}/\gamma & x \geq 0 \\ +\infty & x < 0, \quad \text{with } \phi_+^*(y) = \begin{cases} y^{1 + \beta}, & y < -\beta \\ +\infty & y \geq -\beta, \quad \text{and } d_\phi = +\infty, \end{cases} \end{cases}$$

(b) for $0 < \gamma < 1$ and $\beta = \gamma/(\gamma - 1)$,

$$\phi_+(x) = \begin{cases} (x^{1 - \gamma})/(1 - \gamma) & x \geq 0 \\ +\infty & x < 0, \quad \text{with } \phi_+^*(y) = \begin{cases} (1 + y/\beta)^{\beta}, & y < -\beta \\ +\infty & y \geq -\beta, \quad \text{and } d_\phi = -\beta, \end{cases} \end{cases}$$

(c) for $\gamma < 0$ and $\beta = \gamma/(\gamma - 1)$,

$$\phi_+(x) = \begin{cases} (1 - x^{\gamma})/\gamma & x > 0 \\ +\infty & x \leq 0, \quad \text{with } \phi_+^*(y) = \begin{cases} -((1 - y)^{1 + \beta})/\beta, & y < 1 \\ +\infty & y \geq 0, \quad \text{and } d_\phi = 1. \end{cases} \end{cases}$$

Note that function $\phi_+$ is strictly convex on its domain in all these cases. In contrast, function $\phi_+^*$ is strictly convex on its domain in all these cases, but case (3a).

The next corollary collects the analytical expressions for minimum dispersion S–SDFs corresponding to the particular choices of $\Phi-$dispersions from Examples 1.

**Corollary 1.** Under the conditions of Proposition 4, consider the $\Phi-$dispersions induced by functions $\phi_+$ in Examples 1. It then follows:

1. $M_0 = \exp(-X'\theta_0)$.
2. $M_0 = 1/(1 + X'\theta_0)$, when $-X'\theta_0 < 1$.
3a. $M_0 = (-X'\theta_0)^{+(-\beta - 1)}.$

$^{40}$ The notation $y^+$ denotes $\max\{0, y\}$.  

42
\[ M_0 = (1 - X'\theta_0/\beta)^\beta - 1, \text{ when } -X'\theta_0 < -\beta. \]

(3c) \[ M_0 = (1 + X'\theta_0)^{\beta - 1}, \text{ when } -X'\theta_0 < 1. \]

**Proof of Corollary 1.** Given the conditions of Proposition 4, for every \( \Phi \)-dispersion considered under Examples 1, \( \phi_\nu \) is strictly convex on its domain. Moreover, for \( \Phi \)-dispersions (1) and (3a) \( \nu = +\infty \), i.e., there is no restriction on the values of the optimal portfolio payoff. With regard to the other cases, \( \nu < +\infty \). Therefore, the optimal portfolio payoff is restricted to be essentially strictly smaller than the corresponding value of \( \nu \).

**Appendix B - Proofs and auxiliary results**

**Appendix B.1 - Proofs of Section 2.1**

Before proving Proposition 1, we report the viability condition introduced by Harrison and Kreps [1979]. First, suppose that agents can be represented by preferences \( \succeq \) on the space of net trades \( \mathbb{R} \times L^p \), where \( (r, x) \in \mathbb{R} \times L^p \) represent \( r \) units of consumption at time 0 and \( x \) units of consumption at time 1, respectively. Let \( A \) be the set of complete and transitive preferences satisfying the following conditions:

(i) For any \( (r, x) \in \mathbb{R} \times L^p \), the set \( \{(r', x') \in \mathbb{R} \times L^p : (r', x') \succeq (r, x)\} \) is convex.

(ii) For any \( (r, x) \in \mathbb{R} \times L^p \), the sets \( \{(r', x') \in \mathbb{R} \times L^p : (r', x') \succeq (r, x)\} \) and \( \{(r', x') \in \mathbb{R} \times L^p : (r, x) \succeq (r', x')\} \) are \( \sigma \)-closed, where \( \sigma \) is the product topology of the Euclidean topology on \( \mathbb{R} \) and the norm topology on \( L^p \).

(iii) For any \( (r, x) \in \mathbb{R} \times L^p \), \( r' > 0 \) and strictly positive \( x' \in L^p \), \( (r + r', x) \succeq (r, x) \) and \( (r, x + x') \succeq (r, x) \).

Now, the viability condition can be made precise via the following definition.

**Definition 5 (Viable financial market).** A financial market defined by a set of traded payoffs \( \mathcal{Z} \) and a pricing function \( \pi \) is viable if there exists some agent with preferences \( \succeq \in A \) and \( (r^*, x^*) \in \mathbb{R} \times \mathcal{Z} \) such that \( r^* + \pi(z^*) \leq 0 \) and \( (r^*, z^*) \succeq (r, z) \) for every \( (r, z) \in \mathbb{R} \times \mathcal{Z} \) with \( r + \pi(z) \leq 0 \).

**Proof of Proposition 1.** Convexity of \( h^* \) implies convexity of pricing functional \( \pi \) and the set of traded payoffs \( \mathcal{Z} \). By [Chen et al., 2001, Theorem 1], there exists a strictly positive, continuous linear functional \( \lambda \) on \( L^p \), such that \( \lambda|\mathcal{Z} \leq \pi \), where \( \lambda|\mathcal{Z} \) indicates the restriction of \( \lambda \) to \( \mathcal{Z} \).

Since \( \lambda \) is a linear continuous functional on \( L^p \), it belongs to the dual space of \( L^p \). Then, Riesz Representation Theorem implies that there exists \( M \in L^q \), with \( 1/p + 1/q = 1 \), such that \( \lambda(Z) = E[MZ] \), see e.g., [Royden and Fitzpatrick, 1988, Ch. 8]. The positivity of \( \lambda \) implies that \( M > 0 \) almost surely. Moreover, by construction, such \( M \) also satisfies the inequality \( E[MZ] \leq \pi(Z) \) for any \( Z \in \mathcal{Z} \). From the definition of pricing functional \( \pi \), it then also follows for any \( \theta \in \mathbb{R}^N \): \[ \theta_S P_S + \theta_D P_D + h^*(\theta_D) + \tau \geq \theta_S E[MX_S] + \theta_D E[MX_D]. \]

This inequality implies, for any \( \theta_S \in \mathbb{R}^N \): \[ \theta_S P_S + \tau \geq \theta_S E[MX_S]. \]

Thus, \( E[MX_S] - P_S = 0_S \). Moreover, inequality (A-1) also implies, for any \( \theta_D \in \mathbb{R}^N \):

\[ \theta_D (E[MX_D] - P_D) - \tau \leq h^*(\theta_D). \]

From [Bauschke et al., 2011, Prop. 13.10 (i)], we hence finally obtain \( h(E[MX_D] - P_D) \leq \tau \). This concludes the proof.

\[ \square \]
Appendix B.2 - Proofs of Section 2.3

Proposition 9 (Lagrange Multiplier Theorem for S-SDFs). Suppose \( h \in \Gamma(\mathbb{R}^n) \), \( \phi \in \Gamma(\mathbb{R}) \) is such that \( \phi_+ \) is strictly convex on \((0, +\infty) \subset \text{dom} \phi_+ \), and let:

\[
\Pi_\lambda := \inf_{M \in \mathcal{M}} \left\{ \mathbb{E}[\phi(M)] + \lambda h(\mathbb{E}[MX_D] - P_D) \right\}.
\]

Further suppose that for some fixed \( \tau \geq 0 \), \( K(\tau) > -\infty \) and there exists \( \tilde{M} \in \mathcal{M} \) such that \( \mathbb{E}[\phi(\tilde{M})] \) is finite and \( h(\mathbb{E}[M X_D] - P_D) < \tau \). Then there exists \( \lambda_0 \geq 0 \) such that \( K(\tau) = \Pi_{\lambda_0} - \lambda_0 \tau \) and the unique solution of \( \Pi_{\lambda_0} \), if it exists, is the unique solution of \( K(\tau) \).

Proof of Proposition 9. Under the assumptions of the proposition, the Lagrange Multiplier Theorem implies existence of \( \lambda_0 \geq 0 \) such that \( K(\tau) = \Pi_{\lambda_0} - \lambda_0 \tau \) and such that any solution \( M_0 \) of \( K(\tau) \) also solves \( \Pi_{\lambda_0} \); see, e.g., [Luenberger, 1997, Thm. 1 pag. 217]. Since by Proposition 3 the set of solutions of \( \Pi_{\lambda_0} \), if it is not empty, is a singleton, we can conclude that the unique solution of \( \Pi_{\lambda_0} \), if it exists, is the unique solution of \( K(\tau) \).

Proof of Proposition 2. First we rewrite the penalized S-SDF problem (15) according to the notation given in [Borwein and Lewis, 1992, Sec. 4]. Subsequently, we check that the conditions of [Borwein and Lewis, 1992, Cor. 4.3] are satisfied and we obtain the dual portfolio problem.

Define for any stochastic discount factor \( M \) the linear function \( A : L^p \to \mathbb{R}^N \) by \( A(M) := \mathbb{E}[MX] \). Further, for every vector \( y \in \mathbb{R}^N \), let \( g : \mathbb{R}^N \to (-\infty, +\infty] \) be defined by \( g(y) := \delta_{\{P_S\}}(y_S) + \psi(y_D - P_D) \). Finally, let \( I_\phi := \mathbb{E}[\phi(\cdot)] \) and notice that

\[
I_\phi(M) + \delta_{\lambda_+}(M) = I_{\phi_+}(M) = \begin{cases} I_\phi(M) & M \geq 0 \\ \infty & \text{else} \end{cases}.
\]

Then:

\[
\Pi = \inf_{M \in L^p} \left\{ I_{\phi_+}(M) + g(A(M)) \right\}.
\]

As payoffs are in \( L^p \) with \( 1/p + 1/q = 1 \), \( A \) is continuous, while Assumptions 1 and 2 imply that \( I_{\phi_+} \) and \( g \) are both proper and convex functions. Thus, in order to apply [Borwein and Lewis, 1992, Cor. 4.3] and obtain the dual problem of \( \Pi \) we need to check that

\[
A[qri(\text{dom} I_{\phi_+})] \cap \text{ri(dom } g) \neq \emptyset,
\]

or simply \( A[qri(\text{dom } I_{\phi_+})] \cap \text{dom } g \neq \emptyset \) in case \( \psi \) is piecewise linear.\(^{42}\)

As proved in [Borwein and Lewis, 1991, Cor.2.6], our requirement \((0, +\infty) \subset \text{dom } \phi \) in Assumption 1 implies that

\[
A(\text{dom } I_\phi \cap L^q_{++}) \subset A[qri(\text{dom } I_{\phi_+})].
\]

Moreover, \( \text{ri(dom } g) = \{P_S\} \times \text{ri(dom } \psi) \) because the relative interior distributes over cartesian products and \( P_S \in \text{ri} \{P_S\} \).\(^{43}\) Hence, showing that \( A(\text{dom } I_\phi \cap L^q_{++}) \cap \text{ri(dom } g) \neq \emptyset \) is enough to prove (A-3). Again, in case \( \psi \) is piecewise linear, it is sufficient to show \( A(\text{dom } I_\phi \cap L^q_{++}) \cap \text{dom } g \neq \emptyset \). This result follows from Assumption 2. Indeed, we have \( \tilde{M} \in \text{dom } I_\phi \cap L^q_{++} \), hence \( A(\tilde{M}) \in A(\text{dom } I_\phi \cap L^q_{++}) \). Moreover, \( P_S = \mathbb{E}[\tilde{M} X_S] \) and \( \mathbb{E}[\tilde{M} X_D] - P_D \in \text{ri(dom } \psi) \), or \( \mathbb{E}[\tilde{M} X_D] - P_D \in \text{dom } \psi \) in case \( \psi \) is piecewise linear, which means that \( A(\tilde{M}) \in \text{dom } g \), \( A(\tilde{M}) \in \text{dom } g \) in case \( \psi \) is piecewise linear.

\(^{41}\) For the class of Cressie-Read \( \Phi \)-dispersions in Examples 1 of Appendix A, \( \phi \in \Gamma(\mathbb{R}) \) is such that \( \phi_+ \) is strictly convex in \((0, +\infty)\).

\(^{42}\) See [Borwein and Lewis, 1992, Def. 2.3] for the definition of quasi relative interior, qri.

\(^{43}\) The fact that \( P_S \in \text{ri}(P_S) \) trivially follows from the characterization of relative interior given in [Borwein and Lewis, 1992, Prop. 2.1], as the cone generated by \( \{P_S\} - P_S = \{0\} \) is a subspace.
Therefore, condition (A-3) is satisfied and by [Rockafellar and Lewis, 1992, Cor. 4.3] we have
\[
\Pi = \max_{\theta \in \mathbb{R}^N} \{-I_{\phi_+}^*(t\mathcal{A}(\theta)) - g^*(-\theta)\} \tag{A-4}
\]
where \(I_{\phi_+}^* : L^p \to (-\infty, +\infty]\) is the conjugate function of \(I_{\phi_+}\) and \(t\mathcal{A} : \mathbb{R}^N \to L^p\) is the adjoint map of \(A\), given by \(t\mathcal{A}(\theta) = X'\theta\).\(^{44}\)

We now show that the right hand side of equation (A-4) is equivalent to the dual portfolio problem defined in equation (16). As \(\phi_+\) is in \(\Gamma(\mathbb{R})\), we can apply [Rockafellar, 1968, Thm. 2] to obtain \(I_{\phi_+}^* = I_{\phi_+^*}\). Moreover, for every \(\theta \in \mathbb{R}^N\),
\[
g^*(-\theta) = \sup_{y \in \mathbb{R}^N} \{-\theta'y - g(y)\} = \sup_{y \in \mathbb{R}^N} \{\langle \theta', y \rangle - \delta_{\{\mathcal{P}\}}(y)\} + \sup_{y_D \in \mathbb{R}^N, D} \{-\theta' y_D - \psi(y_D - P_D)\}
= -P'_S \theta_S + \sup_{k \in \mathbb{R}^N, D} \{-\theta' (k + P_D) - \psi(k)\}
= -P'\theta + \psi^*(-\theta_D) .
\]

Thus, we conclude
\[
\Pi = \max_{\theta \in \mathbb{R}^N} \{-I_{\phi_+}^*(X'\theta) + P'\theta - \psi^*(-\theta_D)\} ,
\]
thereby proving strong duality between \(\Pi\) and \(-\Delta\), with dual attainment. \(\square\)

**Proof of Proposition 3.** The statements of Proposition 3 follow from the proofs of [Rockafellar and Lewis, 1991, Prop. 2.11 and Thm. 4.6, 4.7]. \(\square\)

**Proof of Proposition 4.** Following the proof of Proposition 2,
\[
\Pi = \inf_{M \in L^p} \{I_{\phi_+}(M) + g(A(M))\} = \max_{\theta \in \mathbb{R}^N} \{-I_{\phi_+}^*(t\mathcal{A}(\theta)) - g^*(-\theta)\} = -\Delta .
\]

The dual problem can be furthermore rewritten as
\[
\Delta = \min_{\theta \in \mathbb{R}^N} \{I_{\phi_+}^*(t\mathcal{A}(\theta)) + g^*(\theta)\} ,
\]
since \(I_{\phi_+}^* = I_{\phi_+}\) by [Rockafellar, 1968, Thm. 2].

We first show that \(\phi_+^*\) is differentiable in \(t\mathcal{A}(\theta_0)(\omega)\) almost surely. By [Rockafellar and Lewis, 1991, Lem. 4.2], we have that \(\text{int}(\text{dom } \phi_+^*) = (-\infty, d_+).\) Moreover, since \(\phi_+^*\) is strictly convex on its domain, by [Rockafellar and Lewis, 1991, Thm. 4.6] \(\phi_+^*\) is differentiable on \(\text{int}(\text{dom } \phi_+^*)\), hence, on \((-\infty, d_+).\) This, together with the assumption that \(-X'\theta_0 < d_+\) a.s., implies that \(\phi_+^*\) is differentiable in \(t\mathcal{A}(\theta_0)(\omega)\) almost surely. Let us denote such derivative by \(M_0(\omega) := (\phi_+^*)'(X'(\omega)\theta_0)\).

Below, we will use the fact that \(M_0\) is the unique element of \(\partial I_{\phi_+}^*(t\mathcal{A}(\theta_0)) \subset L^q\). To show this, we first claim that for any \(\tilde{M} \in \partial I_{\phi_+}^*(t\mathcal{A}(\theta_0))\) we must have \(\tilde{M}(\omega) \in \partial \phi_+^*(t\mathcal{A}(\theta_0)(\omega)) \subset \mathbb{R}\) almost surely. If this is true, the fact that \(\phi_+^*\) is differentiable in \(-X(\omega)\theta_0\) a.s., that is, the only element of its subdifferential is given by \(M_0(\omega)\), implies that \(M = M_0\) almost surely, i.e., \(M_0\) is the unique element of \(\partial I_{\phi_+}^*(t\mathcal{A}(\theta_0))\). To show this uniqueness, we start from following identity, which holds by [Rockafellar, 1970, Thm. 23.5]:
\[
I_{\phi_+}^*(t\mathcal{A}(\theta_0)) + I_{\phi_+}(\tilde{M}) = \langle t\mathcal{A}(\theta_0), \tilde{M} \rangle .
\]

Explicitly, this gives \(\int \{\phi_+^*(t\mathcal{A}(\theta_0)(\omega)) + \phi_+^*(\tilde{M}(\omega))\} \frac{d\mathbb{P}(\omega)}{d\mathbb{P}(\omega)} = 0\). Here, the integrand is non-negative by Fenchel’s inequality. Therefore, it is zero almost surely. Applying again [Rockafellar, 1970, \(^{44}\) The adjoint map of \(A\), \(t\mathcal{A}\), is characterized by the identity \(E[t\mathcal{A}(\theta)M] = \theta' E[M X]\), for each \(M \in L^q\) and each portfolio weights \(\theta \in \mathbb{R}^N\).]
Thm. 23.5) to this integrand, we obtain $\tilde{M}(\omega) \in \partial \phi_+^*(t^\prime A(-\theta_0)(\omega))$ almost surely and hence $M_0 = \tilde{M} \in L^\nu$, as claimed.

To show the primal feasibility of $M_0$, notice first that since $\theta_0$ is a solution to problem (A-5), if we denote with $Q$ the dual objective function, the following first order condition holds:

$$0 \in \partial Q(\theta_0) = \partial I_{\phi_+}^*(t^\prime A(-\theta_0)) + \partial g^*(\theta_0) .$$

Moreover, by [Borwein, 1981, Thm. 4.1], $\partial I_{\phi_+}^*(t^\prime A(-\theta_0)) = -\mu A(\partial I_{\phi_+}^*(t^\prime A(-\theta_0)))$, which is given by the singleton $\{A(M_0)\}$ since $\mu A|_{L^\nu} = A$. Hence, there exists $\nu \in \partial g^*(\theta_0) \subseteq \mathbb{R}^N$, such that

$$-A(M_0) + \nu = 0 .$$

Equation (A-6), together with the fact that $\nu \in \partial g^*(\theta_0) \subseteq \mathbb{R}^N$ and $M_0 \in \partial I_{\phi_+}^*(t^\prime A(\theta_0))$, implies the primal feasibility of $M_0$. Indeed, by [Rockafellar, 1970, Thm. 23.5], $\nu \in \partial g^*(\theta_0) \subseteq \mathbb{R}^N$ implies $g(\nu) = \nu^\prime \theta_0 - g^*(\theta_0)$. Since $g^*$ is proper and by definition of the subgradient, we must thus have $g^*(\theta_0) < +\infty$. Hence, $g(\nu)$ is finite, which from equation (A-6) means that also $g(A(M_0))$ is finite. Using the explicit definition of $g$, this means $\mathbb{E}[M X_S] - P_S = 0_S$ and $\mathbb{E}[M X_D] - P_D \in \text{dom } \psi$. Similarly, using again [Rockafellar, 1970, Thm. 23.5], the fact that $M_0 \in \partial I_{\phi_+}^*(t^\prime A(-\theta_0))$, the definition of the subgradient and properness of $I_{\phi_+}$ imply $M_0 \in \text{dom } I_{\phi_+}$, so that $M_0 \in L^\nu_+$ is indeed primal feasible.

We finally show that $M_0$ is a primal solution. To this end, consider the following equalities:

$$I_{\phi_+}^*(t^\prime A(-\theta_0)) + I_{\phi_+}(M_0) = \langle t^\prime A(-\theta_0), M_0 \rangle = \langle -\theta_0, A(M_0) \rangle = -\nu^\prime \theta_0 ,$$

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 (A-6). Thus, we can explicitly write the primal objective function computed in $M_0$ as:

$$I_{\phi_+}(M_0) + g(A(M_0)) = -I_{\phi_+}^*(t^\prime A(-\theta_0)) - g^*(\theta_0) + [-\nu^\prime \theta_0 + g^*(\theta_0) + g(A(M_0))] .$$

Using equation (A-6), $-\nu^\prime \theta_0 + g^*(\theta_0) + g(A(M_0)) = -A(M_0)^\prime \theta_0 + g^*(\theta_0) + g(A(M_0))$, which is zero by [Rockafellar, 1970, Thm. 23.5]. This shows that $M_0$ is a primal solution. Uniqueness of this solution follows from Proposition 3. Moreover, since $g^*(\theta) = P^\prime \theta + \psi^*(\theta_D)$, from equation (A-6) we also obtain $\mathbb{E}[M_0 X_S] - P_S = 0$ and $\mathbb{E}[M_0 X_D] - P_D \in \partial \psi^*(\theta_{0D})$. This concludes the proof of claim (i).

If $\psi^*$ is differentiable in $\theta_0$, then $\partial \psi^*(\theta_{0D}) = \{\nabla \psi^*(\theta_{0D})\}$. Hence, $\mathbb{E}[M_0 X_D] - P_D = \nabla \psi^*(\theta_{0D})$, proving claim (ii). This concludes the proof.

**Appendix B.3 - Proofs of Section 2.4**

**Proof of Proposition 5.** For $\alpha > 0$ and since $f^* \in \Gamma(\mathbb{R}^{N_D})$, [Bauschke et al., 2011, Prop. 13.21 (i)] yields:

$$(e_{\alpha} f^*)^* = f + \frac{\alpha}{2} \Vert \cdot \Vert_2^2 .$$

Moreover, from [Bauschke et al., 2011, Prop. 12.15], $e_{\alpha} f^* \in \Gamma(\mathbb{R}^{N_D})$. Therefore,

$$(f + \frac{\alpha}{2} \Vert \cdot \Vert_2^2)^* = ((e_{\alpha} f^*)^*)^* = e_{\alpha} f^* .$$

(A-7)

The result for $\alpha = 0$ follows from the fact that $e_{\alpha} f^*$ converges to $f^*$ as $\alpha \downarrow 0$. This proves result (20). Result (21) now follows from [Bauschke et al., 2011, Prop. 12.29]. This concludes the proof.

**Corollary 2.** Let function $f \in \Gamma(\mathbb{R}^{N_D})$ and consider $\psi$ given by (19) with $\alpha > 0$. It then follows:
(i) If \( f = \| \cdot \| \) is a norm:

\[
\psi = \| \cdot \| + \frac{\alpha}{2} \| \cdot \|_2^2 \quad \text{and} \quad \psi^* = \frac{1}{2\alpha} d_{B(\| \cdot \|)}^2 ,
\]

where \( d_{B(\| \cdot \|)} \) is the Euclidean distance from the closed unit ball \( B(\| \cdot \|) \) under dual norm \( \| \cdot \|_* \).

(ii) If \( f = \delta_C \) is the characteristic function of a non empty closed convex set \( C \):

\[
\psi = \delta_C + \frac{\alpha}{2} \| \cdot \|_2^2 \quad \text{and} \quad \psi^* = \frac{1}{2\alpha} d_{\alpha C}^2 ,
\]

where \( d_{\alpha C} \) is the Euclidean distance from convex set \( \alpha C := \{ x \in \mathbb{R}^D : x = \alpha z \text{ for some } z \in C \} \). More specifically:

(a) If \( C \) is a closed convex cone \( K \):

\[
\psi = \delta_K + \frac{\alpha}{2} \| \cdot \|_2^2 \quad \text{and} \quad \psi^* = \frac{1}{2\alpha} d_{-K^*}^2 ,
\]

where \( d_{-K^*} \) is the Euclidean distance from the negative dual cone \( -K^* = \{ \theta_D \in \mathbb{R}^D : \theta_D^T x \leq 0 \text{ for all } x \in K \} \).

(b) If \( C \) is the closed ball \( \tau B_\infty \) of radius \( \tau \) under the \( l_\infty \)-norm:

\[
\psi = \delta_C + \frac{\alpha}{2} \| \cdot \|_2^2 \quad \text{and} \quad \psi^* = \Upsilon_{\alpha, \tau} ,
\]

with the multivariate Huber function \( \Upsilon_{\alpha, \tau} \), defined for any \( \theta_D \in \mathbb{R}^D \) by

\[
\Upsilon_{\alpha, \tau}(y) := \sum_{i=1}^N \upsilon_{\alpha, \tau}(y_i) .
\]

Here, \( \upsilon_{\alpha, \tau} : \mathbb{R} \to \mathbb{R} \) is the univariate Huber function, defined for any real number \( z \) by:

\[
\upsilon_{\alpha, \tau}(z) = \left\{ \begin{array}{ll}
\tau \left( |z| - \frac{\tau^2}{2\alpha} \right) & |z| \geq \tau \alpha \\
\frac{\tau^2}{2\alpha} & |z| \leq \tau \alpha .
\end{array} \right.
\]

(iii) Let \( f \in \Gamma(\mathbb{R}^D) \) and consider \( \psi \) given by (19) with \( \alpha = 0 \). It then follows:

(i) If function \( f = \| \cdot \| \) is a norm:

\[
\psi = \| \cdot \| \quad \text{and} \quad \psi^* = \delta_{\mu(\| \cdot \|)} .
\]

(ii) If function \( f = \delta_K \) is the characteristic function of a non empty convex cone \( K \):

\[
\psi = \delta_K \quad \text{and} \quad \psi^* = \delta_{-K^*} .
\]

**Proof of Corollary 2.** Consider function \( f \in \Gamma(\mathbb{R}^D) \) and function \( \psi \) given by (19) with \( \alpha > 0 \).

(i) If \( f = \| \cdot \| \), its convex conjugate is given by \( f^* = \delta_{\mu(\| \cdot \|)} \). Moreover, since \( f \in \Gamma(\mathbb{R}^D) \), Proposition 5 implies \( \psi^* = e_\alpha f^* \). Explicit Moreau envelope calculations then give:

\[
\psi^*(y) = \inf_z \left\{ \mu_B(\| \cdot \|)(z) + \frac{1}{2\alpha} \| y - z \|_2^2 \right\} = \inf_{z \in B(\| \cdot \|)} \left\{ \frac{1}{2\alpha} \| y - z \|_2^2 \right\} = \frac{1}{2\alpha} d_{B(\| \cdot \|)}^2 (y) .
\]

(ii) Given \( f = \delta_C \in \Gamma(\mathbb{R}^D) \), Proposition 5 yields \( \psi^* = e_\alpha f^* \). Since \( C \) is a generic closed convex set, \( f^* \) is not known explicitly. Therefore, we first apply Moreau decomposition (see, e.g., [Bauschke et al., 2011, Thm. 14.3]), to write \( \psi^* = e_\alpha f^* = \frac{1}{\alpha} \| y - z \|_2^2 - e_1(y/\alpha) \). Explicit Moreau envelope calculations then

\[^{45} d_{B(\| \cdot \|)}(\theta_D) \) := \( \theta_D - \text{proj}_{B(\| \cdot \|)}(\theta_D) \) \( \| \cdot \|_2 \), where \( \text{proj}_{B(\| \cdot \|)} \) denotes the projection on the centred unit ball of norm \( \| \cdot \|_*$.
yield:

\[
\psi^*(y) = \frac{1}{2\alpha} \|y\|^2_2 - \inf_z \left\{ f(z) + \frac{\alpha}{2} \|y/\alpha - z\|^2_2 \right\} = \frac{1}{2\alpha} \|y\|^2_2 - \frac{\alpha}{2} \inf_{z \in C} \left\{ \frac{\alpha}{2} \|y/\alpha - z\|^2_2 \right\} = \frac{1}{2\alpha} \|y\|^2_2 - \frac{\alpha}{2} d^2_C(y/\alpha) = \frac{1}{2\alpha} \left( \|y\|^2_2 - d^2_{\alpha C}(y) \right).
\]

(a) If \( C \) is a closed convex cone, \( \alpha C = C \) and

\[
\psi^*(y) = \frac{1}{2\alpha} \left( \|y\|^2_2 - d^2_C(y) \right).
\]

An equivalent expression arises by directly calculating the convex conjugate:

\[
(\delta_C)^*(y) = \sup_{x} \{ y'x - \delta_C(x) \} = \sup_{x \in C} \{ y'x \} = \delta_{-C^*}(y), \tag{A-14}
\]

with the polar cone \( -C^* = \{ y : y'x \geq 0 \text{ for all } x \in C \} \). Direct Moreau envelope calculations then yield:

\[
\psi^*(y) = c_\alpha \delta_{-C^*}(y) = \frac{1}{2\alpha} d^2_{\alpha C^*}(y). \tag{A-15}
\]

(b) If \( C \) is the closed ball \( \tau B_\infty \), we obtain:

\[
\psi^*(y) = \frac{1}{2\alpha} \left( \|y\|^2_2 - d^2_{\alpha \tau B_\infty}(y) \right) = \frac{1}{2\alpha} \left( \|y\|^2_2 - \|y - \tilde{y}\|^2_2 \right),
\]

where \( \tilde{y} \) is the Euclidean projection of \( y \) on the ball \( \alpha \tau B_\infty \). Explicit calculations of this projection finally give:

\[
\psi^*(y) = \frac{1}{2\alpha} \sum_{i=1}^{N} v_i, \quad \text{where} \quad v_i = \begin{cases} y_i^2, & |y_i| \leq \alpha \tau \\ \alpha \tau y_i - \alpha^2 \tau^2, & y_i \geq \alpha \tau \\ 2\alpha \tau y_i - \alpha^2 \tau^2, & y_i \leq -\alpha \tau \end{cases}.
\]

(c) Taking the limit \( \alpha \downarrow 0 \) naturally gives the convex conjugate \( \psi^* \) of pricing error functions \( \psi \) given by norms or characteristic functions.

This concludes the proof. \( \square \)

**Appendix B.4 - Proofs of Section 3**

Before proving Proposition 6, we provide conditions under which the solution set of problem \( \Delta \) is compact for settings where \( \Phi \)--dispersion is defined as in Examples 1 of Appendix A, case (3a).

**Proposition 10 (Compactness of set of minimizers for power \( \Phi \)--dispersions with \( \gamma > 1 \)).** Suppose \( Q \) is proper, payoffs \( X_{t+1} \) are linearly independent, function \( \psi^* \in \Gamma(\mathbb{R}^N) \) is bounded from below and Assumptions 2 and 3 hold. Consider problem \( \Delta \) defined in terms of the \( \Phi \)--dispersions in Examples 1 of Appendix A, case (3a). Then, the set of solutions to \( \Delta \) is compact.

**Proof of Proposition 10.** Given Assumptions 1, 3 and the properness of function \( Q, Q \in \Gamma(\mathbb{R}^N) \). Hence, \( Q \) has a closed and convex set of minimizers [Rockafellar, 1970, Section 7]. In order to show that the set of minimizers is compact, we need to show that it is bounded, which is the case if and only if \( Q \) is coercive; see [Bauschke et al., 2011, Prop. 11.12]. Therefore, we now show that \( Q \) is coercive.
Since $Q$ is defined in terms of $\Phi$--dispersions in Examples 1 of Appendix A, case (3a), we need to show that for every nonzero portfolio weight $\theta$:

$$\lim_{y \to +\infty} y^\beta E[(X'_{t+1}\theta)^{+\beta}/\beta] - yE[P'_t\theta] + \psi^\ast(y\theta_D) = +\infty,$$

where $\beta = \gamma/(\gamma - 1) > 1$.

Recall first that by assumption $\psi^\ast(y\theta_D)$ is bounded from below for all $y \in \mathbb{R}$ and dubious portfolio weights $\theta_D$. Let us consider directions $\theta$ such that $X'_{t+1}\theta > 0$ with positive probability. Then $E[(X'_{t+1}\theta)^{+\beta}/\beta]$ is strictly positive and therefore

$$y^\beta E[(X'_{t+1}\theta)^{+\beta}/\beta] - yE[P'_t\theta] \to +\infty,$$

as $y > 0$ increases, since $\beta > 1$. Now consider directions $\theta$ for which $X'_{t+1}\theta \leq 0$ almost surely. Here $E[(X'_{t+1}\theta)^{+\beta}/\beta] = 0$ and, by the linear independence of $X_{t+1}$, $X'_{t+1}\theta < 0$ with positive probability. Then Assumption 2, which implies absence of arbitrage by Proposition 1, implies that $E[P'_t\theta]$ is strictly negative. Therefore, for these directions $-yE[P'_t\theta]$ grows unboundedly as $y > 0$ increases.

In summary, we have shown that function $Q$ is coercive. Hence, the set of solutions to problem $\Delta$ is closed and bounded. This concludes the proof.

**Proof of Proposition 6.** Given claim (i), convergence in probability of $\theta_T$ to $\theta_0$ in statement (ii) follows using, e.g., [Newey and McFadden, 1994, Thm 2.7]. We now prove claim (i).

Under Assumptions 1, 3 and 4, stochastic process $\{\phi^+_t(-X'_{t+1}\theta) + P'_t\theta\}_{t \in \mathbb{N}}$ is ergodic, stationary and integrable for all $\theta \in \Theta$. We can thus use the Ergodic theorem (see, e.g., [Davidson, 1994, Thm. 13.12]) to conclude that $Q_T$ converges to $Q$ almost surely on $\Theta$. Then, by [Rockafellar, 1970, Thm. 10.8], we obtain almost sure convergence of $Q_T$ to $Q$, uniformly on any compact subset of the interior of $\Theta$.

Under Assumption 4, $\Theta$ contains in its interior the compact set of minimizers of $Q$, which we denote by $\mathcal{Q}$. Hence, the exists a compact subset $\Xi$ of the interior of $\Theta$, which contains $\mathcal{Q}$ in its interior. Therefore, the uniform convergence of sequence $Q_T$ to $Q$ on $\Xi$ implies that for a sufficiently large $T$ the minimum of $Q_T$ on $\Xi$ is taken almost surely in set $Q$. For such a large $T$, convexity of $Q_T$ implies that also the minimum of $Q_T$ on $\mathbb{R}^N$ is taken almost surely in set $Q$, i.e., $\Delta_T := \min\{Q_T(\theta) : \theta \in \mathbb{R}^N\} = \min\{Q_T(\theta) : \theta \in \Xi\}$ almost surely. Therefore, given the almost sure uniform convergence of sequence $Q_T$ to $Q$ on $\Xi$, we finally obtain the almost sure convergence of $\Delta_T$ to $\min\{Q(\theta) : \theta \in \Xi\} = \min\{Q(\theta) : \theta \in \Theta\} =: \Delta$. This concludes the proof.

**Proof of Proposition 7.** We have:

$$\sqrt{T}(\Delta_T - \Delta) = -\sqrt{T}(Q_T(\theta_T) - Q(\theta_0))$$

$$= -\sqrt{T}(Q_T(\theta_T) - Q_T(\theta_0)) - \sqrt{T}(Q_T(\theta_0) - Q(\theta_0)) .$$

(A-16)

We analyze the two components of the RHS of equation (A-16). The second component, $\sqrt{T}(Q_T(\theta_0) - Q(\theta_0))$, reads:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \{\phi^+_t(-X'_{t+1}\theta_0) + P'_t\theta_0 - E[\phi^+_t(-X'_{t+1}\theta_0) + P'_t\theta_0] \} .$$

(A-17)

Assumption 8 implies that the random sequence (A-17) converges in distribution to a zero mean normally distributed random variable with variance $v(\theta_0)$.

Now, recall that $Q_T$ is a convex, differentiable function minimized in $\theta_T$. Therefore, the first component on
the RHS of equation (A-16) satisfies:

$$0 \leq -\sqrt{T} (Q_T(\theta_T) - Q_T(\theta_0)) \leq -\sqrt{T} \nabla Q_T(\theta_0)(\theta_T - \theta_0),$$  (A-18)

where

$$\sqrt{T} \nabla Q_T(\theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left\{ \left( \phi^*_+ \right)'(-X'_{t+1} \theta_0)X_{t+1} + P_t + [0'_S, \nabla \psi^*(\theta_{0D})]' \right\}. $$

From the first-order condition $\nabla Q(\theta_0) = 0$ for an extremum in the interior of $\Theta$ and Assumption 7, we have $[0'_S, \nabla \psi^*(\theta_{0D})]' = -E \left[ \left( \phi^*_+ \right)'(-X'_{t+1} \theta_0)X_{t+1} + P_t \right]$. Moreover, given the boundedness in probability of the sequence

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left\{ \left( \phi^*_+ \right)'(-X'_{t+1} \theta_0)X_{t+1} + P_t - E[(-\phi^*_+)'(-X'_{t+1} \theta_0)X_{t+1} + P_t] \right\}$$  (A-19)

in equation (40), we obtain that $\sqrt{T} \nabla Q_T(\theta_0)$ is bounded in probability as well. This, together with the consistency of $\theta_T$ in Proposition 6, shows that the last term of on the RHS of inequalities (A-18) converges in probability to zero, implying $-\sqrt{T}(Q_T(\theta_T) - Q_T(\theta_0)) = o_p(1)$. Therefore, $\sqrt{T}(\Delta T - \Delta)$ is asymptotically equivalent to the random sequence (A-17), which converges in distribution to a zero mean normally distributed random variable with variance $v(\theta_0)$. This concludes the proof.

**Proof of Proposition 8.** We first define for $u, z \in \mathbb{R}^N$ the auxiliary functions: $\hat{Q}_T(u) = Q_T\left(\theta_0 + \frac{u}{\sqrt{T}}\right)$,

$$\hat{Q}(u) = Q\left( \theta_0 + \frac{u}{\sqrt{T}} \right)\left( gr(z) := \frac{1}{T} \sum_{t=1}^{T} \left( \phi^*_+(-X'_{t+1}z) + P'_t z \right) \right) $$

and $g(z) := E[\phi^*_+(-X'_{t+1}z) + P'_t z]$. We then define the estimation problem in terms of local parameter $u \in \mathbb{R}^N$, using the process:

$$Z_T(u) = T \left\{ \hat{Q}_T(u) - \hat{Q}_T(0) \right\}. $$  (A-20)

Note that $u_T := \sqrt{T}(\theta_T - \theta_0)$ minimizes $Z_T(u)$, because $Q_T(\theta)$ is minimized in $\theta_T$. Using the stochastic equicontinuity Assumption 9, we obtain:

$$\hat{Q}_T(u) - \hat{Q}_T(0) = \frac{1}{\sqrt{T}} \nabla \hat{Q}_T(0) u + \left( \hat{Q}(u) - \hat{Q}(0) \right) + o_p(1/T). $$  (A-21)

The second component of the RHS of equation (A-21) reads:

$$\hat{Q}(u) - \hat{Q}(0) = g\left( \theta_0 + \frac{u}{\sqrt{T}} \right) - g(\theta_0) + \psi^*(\theta_{0D}) \left( \theta_0 + \frac{u}{\sqrt{T}} \right) - \psi^*(\theta_{0D}). $$  (A-22)

Assumption 10 now allows us to compute a second order Taylor expansion of auxiliary function $g$ in $\theta_0$. Together with the expression of the remainder sequence $R_T$, this yields:

$$\hat{Q}(u) - \hat{Q}(0) = \frac{1}{\sqrt{T}} \left( \nabla g(\theta_0) + [0'_S, \nabla \psi^*(\theta_{0D})]' \right) u + \frac{1}{2T} u' G(\theta_0) u + R_T(u) + o_p(1/T). $$  (A-23)

Given that $\theta_0$ minimizes $Q(\theta)$ in the interior of $\Theta$, the first component on the RHS of equation (A-23) vanishes. Substituting identity (A-23) in equation (A-21) and multiplying by $\sqrt{T}$, we obtain:

$$Z_T(u) = \sqrt{T} \left( \nabla g(\theta_0) + [0'_S, \nabla \psi^*(\theta_{0D})]' \right) u + \frac{1}{2T} u' G(\theta_0) u + TR_T(u) + o_p(1).$$

Assumptions 7 and 10, together with the first-order condition $\nabla Q(\theta_0) = 0$, imply the convergence in distribution of term $\sqrt{T} \left( \nabla g(\theta_0) + [0'_S, \nabla \psi^*(\theta_{0D})]' \right)$ to a zero mean normally distributed random vector $Y$ with variance covariance matrix $V(\theta_0)$. Finally, Lipschitz continuity of $\nabla \psi^*$ implies that the term $TR_T(u)$ converges pointwise to a limit $R(u)$ described by convex function $R$. 

50
In summary, we obtain that all finite-dimensional distributions of process $Z_T$ converge to the corresponding finite-dimensional distributions of a process $Z$ defined by $Z(u) := u'Y + \frac{1}{2}u'G(\theta_0)u + R(u)$, as $T \to \infty$. Assumption 5 and the convexity of function $Q$ also imply that matrix $G(\theta_0)$ is positive definite. Hence, random function $u \mapsto Z(u)$ is strictly convex in $u$ and has a unique minimizer $u_0 := \arg \min_{u \in \mathbb{R}^N} Z(u)$ almost surely. This implies that $\sqrt{T}(\theta_T - \theta_0) \to \arg \min_{u \in \mathbb{R}^N} Z(u)$ in distribution as $T \to \infty$; see, e.g., Geyer [1996]. This concludes the proof.
Appendix C - Minimum variance S–SDFs without sign constraint

Lemma 1. Given a vector of asset returns $\mathbf{R}$ and convex conjugate penalization function $\psi^*$, the normalized minimum variance S–SDF such that $\mathbb{E}[\mathbf{M}] = 1$ and no non negativity constraint holds, is given by:

$$M_0 = 1 - \theta_0^*(\mathbf{R} - \mathbb{E}[\mathbf{R}]) \quad (A-24)$$

Here, vector $\theta_0 \in \mathbb{R}^N$ is the solution of the dual optimization problem:

$$\Delta(\tau) = \min_{\theta \in \mathbb{R}^N} \left\{ \frac{1}{2} \left( (\mu - V\theta)'V^{-1}(\mu - V\theta) - 1 - \mu'V^{-1}\mu \right) + \psi^*(\theta_D) \right\}$$

where

$$\mu := \mathbb{E}[\mathbf{R} - 1], \quad V := \mathbb{E}[(\mathbf{R} - \mathbb{E}[\mathbf{R}])(\mathbf{R} - \mathbb{E}[\mathbf{R}])'] \quad (A-25)$$

Moreover, whenever penalization function $\psi^*$ is differentiable, $\theta_0$ is such that:

$$\theta_0 = V^{-1}(\mu - \nabla \psi^*(\theta_D)) \quad (A-28)$$

Proof. From Proposition 2, the dual optimization problem resulting for the primal minimum variance S–SDF problem with dispersion function $\phi(x) := \frac{1}{2}x^2$ and convex conjugate dual penalization function $\psi^*$ reads:

$$\Delta(\tau) = \min_{\gamma \in \mathbb{R}, \theta \in \mathbb{R}^N} \left\{ \mathbb{E}[\phi^*_+(\gamma - \theta'\mathbf{R})] + \gamma + \theta'\mathbf{1} + \psi^*(\theta_D) \right\}$$

where $\phi^*_+(y) := \frac{1}{2}y^2$. From Proposition 4, the resulting minimum variance S–SDF is given by:

$$M_0 = -\gamma_0 - \theta_0^*\mathbf{R} \quad (A-30)$$

with the solution $(\gamma_0, \theta_0)^*$ to problem $\Delta(\tau)$. The first order condition $\gamma_0 = -1 - \theta_0^*\mathbb{E}[\mathbf{R}]$ then yields the concentrated dual problem:

$$\Delta(\tau) = \min_{\theta \in \mathbb{R}^N} \left\{ \mathbb{E}[\phi^+_+(1 - \theta'\mathbf{R} - \mathbb{E}[\mathbf{R}])] - 1 - \theta'\mathbb{E}[\mathbf{R}] - 1 + \psi^*(\theta_D) \right\}$$

$$= \min_{\theta \in \mathbb{R}^N} \left\{ \frac{1}{2} \theta'\mathbb{E}[(\mathbf{R} - \mathbb{E}[\mathbf{R}])(\mathbf{R} - \mathbb{E}[\mathbf{R}])'] 1 - 1 + \psi^*(\theta_D) \right\}$$

$$\quad (A-31)$$

Using equation (A-30), the resulting minimum variance S–SDF is hence given by:

$$M_0 = 1 - \theta_0^*(\mathbf{R} - \mathbb{E}[\mathbf{R}]) \quad (A-33)$$

Further, after introducing notation $(A-26)-(A-27)$, the concentrated dual problem reads:

$$\Delta(\tau) = \min_{\theta \in \mathbb{R}^N} \left\{ \frac{1}{2} (\theta'V\theta - 1) - \theta'\mu + \psi^*(\theta_D) \right\}$$

$$= \min_{\theta \in \mathbb{R}^N} \left\{ \frac{1}{2} (\theta'V\theta - 1) - (\theta'\mu + \frac{1}{2}\mu'V^{-1}\mu - 1) + \psi^*(\theta_D) \right\}$$

$$= \min_{\theta \in \mathbb{R}^N} \left\{ \frac{1}{2} (\mu - V\theta)'V^{-1}(\mu - V\theta) - 1 + \frac{1}{2}\mu'V^{-1}\mu + \psi^*(\theta_D) \right\}$$

Hence, whenever function $\psi^*$ is differentiable, vector $\theta_0$ satisfies the first order condition:

$$0 = V\theta_0 - \mu + \nabla \psi^*(\theta_D) \quad (A-34)$$

i.e., equation (A-28) holds. This concludes the proof. \qed

52
Remark 2. The minimum variance SDF in Lemma 1 is equivalently obtained by letting $\theta_0$ be the solution of following dual problem:

$$
\hat{\Delta}(\tau) = \min_{\theta \in \mathbb{R}^N} \left\{ \frac{1}{2} (\mu - V \theta)' V^{-1} (\mu - V \theta) + \psi^*(\theta_D) \right\},
$$

(A-35)

which is a penalized minimum Hansen-Jagannathan distance problem. In absence of penalization ($\psi^* = 0$), the solution of this problem is identical to the solution of the ordinary minimum Least Squares problem:

$$
\theta_0 = \arg \min_{\theta \in \mathbb{R}^N} (\mu - V \theta)'(\mu - V \theta).
$$

(A-36)

In this case, $\mu = V \theta_0$ and the choice of a weighting matrix in the above minimum Least Squares and minimum Hansen-Jagannathan distance problems is irrelevant. In contrast, whenever pricing errors arise and $\mu \neq V \theta_0$, the choice of the weighting matrix matters and the two criteria give rise to different solutions in presence of penalization.

For instance, given an APT consistent norm $||\Sigma^{-1/2} \cdot ||$, the APT pricing constraint

$$
||\Sigma^{-1/2} \mathbb{E}[MR_D - 1]|| \leq \tau,
$$

implies a penalization function given by $\psi^*(\theta_D) = \tau ||\Sigma^{1/2} \theta_D||_*$, which yields an explicit dual problem of the form:

$$
\hat{\Delta}(\tau) = \min_{\theta \in \mathbb{R}^N} \left\{ \frac{1}{2} (\mu - V \theta)' V^{-1} (\mu - V \theta) + \tau ||\Sigma^{1/2} \theta_D||_* \right\},
$$

(A-37)

Whenever no sure assets exist, i.e., $\theta = \theta_D$ and $V = \Sigma$, this dual problem equivalently reads:

$$
\hat{\Delta}(\tau) = \min_{\theta \in \mathbb{R}^N} \left\{ \frac{1}{2} (V^{-1/2} \mu - \theta)'(V^{-1/2} \mu - \theta) + \tau ||\theta||_* \right\},
$$

(A-38)

which is a penalizes ordinary Least Squares problem with rescaled expected excess returns $V^{-1/2} \mu$ and an unscaled norm penalization $\tau ||\theta||_*$ on portfolio weights.
Appendix D - Figures

Figure 1: Two-dimensional unit balls for pricing errors (orange) and portfolio weights (blue), under the APT pricing error function \( h = \frac{1}{2} \Sigma^{-1/2} \Sigma^{1/2} \), thus \( h^* = \frac{1}{2} \Sigma^{1/2} \Sigma^{1/2} \), for the cases: \( \Sigma = I \) (left panel); \( \Sigma_{11} = 1, \Sigma_{12} = 0, \Sigma_{22} = 1.5 \) (middle panel); \( \Sigma_{11} = 2, \Sigma_{12} = 0.5, \Sigma_{22} = 1 \) (right panel). The pricing error (portfolio weight) unit ball is given by the set \( \{ \mathbf{x} \in \mathbb{R}^2 : h(\mathbf{x}) \leq 1 \} \) (\( \{ \mathbf{x} \in \mathbb{R}^2 : h^*(\mathbf{x}) \leq 1 \} \)).

Figure 2: Empirical duality failure. We solve for the minimum variance APT-consistent S-SDF with \( l_2 \)-pricing error function for varying pricing error bounds \( \tau \geq 0 \). On the y-axis we report the estimate of the pricing error function value \( h(\mathbb{E}[\mathbf{P} - M_0 \mathbf{X}]) \) for each \( \tau \). The point of discontinuity in the plot identifies the smallest pricing error bound \( \tau \), for which the sample version of Assumption 2 holds and a solution of the empirical primal S-SDF problem exists. The largest pricing error threshold \( \hat{\tau}^{\text{max}} \) in the plot is computed as the sample version of the maximal threshold \( \tau^{\text{max}} \) in equation (14). All calculations are based on the augmented intermediate dimensional dataset from January 1970 to June 2018, using the risk-free asset as the only sure asset while sorted portfolios and financial ratios constitute the dubious assets.
Figure 3: (Top Panel) Two-dimensional unit balls of lasso (left), elastic net (middle) and Huber (right) pricing error penalization function $\psi$. These unit balls are given by the sets $\{y \in \mathbb{R}^2 : \psi(y) \leq 1\}$ for $\psi = \|\cdot\|_1$ (lasso), $\psi = (\lambda \|\cdot\|_1 + \frac{\alpha}{2} \|\cdot\|_2^2)$ (elastic net) and $\psi = (\Phi_{\alpha,\tau})$ (Huber). (Bottom Panel) Unit balls of the convex conjugate of the functions given in the top panel.

Figure 4: Various APT–consistent pricing error bounds. Two-dimensional unit balls induced by various APT–consistent pricing error functions on pricing errors transformed by $\Sigma^{-1/2}$. The pricing error functions are given by the elastic-net norm $h = (1 - \lambda)\|\cdot\|_1 + \lambda \|\cdot\|_2$ on the left panel and by $h = (1 - \lambda)\sqrt{N_D} \|\cdot\|_\infty + \lambda \|\cdot\|_2$ on the right panel, for various $\lambda \in [0,1]$. 

55
Figure 5: Generalized Hansen-Jagannathan minimum dispersion bounds. The top left panel reports the estimate of the generalized Hansen-Jagannathan minimum dispersion bound $\Pi(\tau)$ defined in Section 3.3, as a function of the target bond price $\mathbb{E}[M^*]$ and the pricing error threshold $\tau$, under an APT-consistent $l_2$ pricing error function. Level curves of the minimum S–SDF dispersion surface are reported in the top right panel. The bottom left panel reports the bounds as a function of the target bond price for different pricing error thresholds $\tau$. Bounds as a function of pricing error threshold $\tau$ for different target bond prices are reported in the bottom right panel. The low dimensional dataset from July 1963 to June 2018 is used, with the three Fama French factors and the risk-free bond as sure assets, while sorted portfolios constitute the dubious assets.
Figure 6: **Tradeoff between minimum S–SDF dispersion bound and APT–consistent pricing errors.** The two panels report the empirical minimum variance bounds for various APT–consistent S–SDFs, as a function of pricing error threshold $\tau$. S–SDFs are determined for various $\lambda \in [0, 1]$ by pricing error function $h = (1 - \lambda) \| \cdot \|_1 + \lambda \| \cdot \|_2$ in the left panel, and by pricing error function $h = (1 - \lambda) \sqrt{N_D} \| \cdot \|_\infty + \lambda \| \cdot \|_2$ in the right panel. All calculations are based on the intermediate dimensional dataset from July 1963 to June 2018, with the risk-free asset being the only sure asset and sorted portfolios constituting the dubious assets. The dashed-dotted vertical line indicates the pricing error threshold $\tau = 0.7$ used in Figures 8 and 9.

Figure 7: **S–SDF minimum dispersion confidence intervals.** The figure reports the empirical minimum variance bound as a function of the pricing error threshold $\tau$. In both panels, the blue dashed line represents the empirical minimum variance bound of S–SDFs pricing exactly the risk-free asset and implying a bound $\tau$ on the APT–consistent $l_\infty$–pricing error function for the sorted portfolios. Quantity $\hat{\tau}_{\max}$ is computed as the empirical counterpart of maximal threshold $\hat{\tau}_{\max}$ in equation (14). The black solid line represents the corrected empirical minimum variance bound under a penalization function given in (19), with $f$ given by the characteristic function of the $l_\infty$ ball of radius $\tau/N_D$. 90% confidence bounds are reported with dotted curves. For the left (right) panel, we use a smoothing parameter value $\alpha = 2 \times 10^{-2}$ ($2 \times 10^{-3}$). Results are based on the intermediate dimensional dataset from July 1963 to June 2018.
Figure 8: Properties of S–SDFs (I). The figures report, from top to bottom, $\Sigma^{-1/2}$–transformed pricing errors and $\Sigma^{1/2}$–transformed optimal portfolio weights for APT–consistent $l_2$ (first row), $l_1$ (second row) and scaled $l_\infty$ (third row) pricing error functions. These results are based on an identical pricing error threshold $\tau = 0.7$ from Figure 6. All calculations are based on the intermediate dimensional dataset from July 1963 to June 2018, with the risk-free asset being the only sure asset and sorted portfolios constituting the dubious assets.
Figure 9: Properties of S–SDFs (II). Time series of minimum variance S–SDFs with, from top to bottom, APT–consistent $l_2$ (first row), $l_1$ (second row) and scaled $l_\infty$ (third row) pricing error functions. These results are based on an identical pricing error threshold $\tau = 0.7$ from Figure 6. All calculations are based on the intermediate dimensional dataset from July 1963 to June 2018, with the risk-free asset being the only sure asset and sorted portfolios constituting the dubious assets. Grey shaded areas represent NBER recessions.
Figure 10: **Time series of minimum variance APT–consistent S–SDFs.** Time series of minimum variance APT–consistent S–SDFs with $l_2$ pricing error function for pricing error threshold $\tau = 0$ (top-left panel), $\tau = \hat{\tau}_{\text{max}} \times 0.3$ (top-right panel), $\tau = \hat{\tau}_{\text{max}} \times 0.6$ (bottom-left panel) and $\tau = \hat{\tau}_{\text{max}} \times 0.9$ (bottom-right panel), where $\hat{\tau}_{\text{max}}$ is the empirical version of population quantity $\tau_{\text{max}}$ defined by equation (14). The intermediate dimensional dataset from July 1963 to June 2018 is used, with the risk-free asset being the only sure asset and sorted portfolios constituting the dubious assets.
Figure 11: Scatter plots of minimum variance $S$–$S$DFs and minimum Kullback Leibler divergence $S$–$S$DFs. Scatter plots of the time series of minimum Kullback Leibler divergence and minimum variance APT–consistent $S$–$S$DFs, based on the $l_2$ pricing error function computed with pricing error threshold $\tau = 0$ (top-left panel), $\tau = \hat{\tau}_{max} \times 0.3$ (top-right panel), $\tau = \hat{\tau}_{max} \times 0.6$ (bottom-left panel) and $\tau = \hat{\tau}_{max} \times 0.9$ (bottom-right panel), where $\hat{\tau}_{max}$ is the empirical version of population quantity $\tau_{max}$ defined by equation (14). The intermediate dimensional dataset from July 1963 to June 2018 is used, with the risk-free asset being the only sure asset and sorted portfolios constituting the dubious assets.
Figure 12: Correlation between optimal portfolio payoffs underlying different minimum dispersion S–SDFs. The figure reports for various pairs of optimal portfolio payoffs their time series correlation, in dependence of different pricing error thresholds that correspond to different fractions of threshold $\hat{\tau}^{\text{max}}$, the empirical version of population quantity $\tau^{\text{max}}$ in equation (14). The payoffs are induced by S–SDFs minimizing various notions of dispersions collected in Appendix A, i.e., Variance (Hansen-Jagannathan, HJ), Kullback-Leibler divergence (KL), Negative entropy (NE) and Hellinger divergence (HE). The intermediate dimensional dataset from July 1963 to June 2018 is used, with the risk-free asset being the only sure asset and sorted portfolios constituting the dubious assets.
Figure 13: **Out-Of-Sample (OOS) GLS adjusted $R^2$ of various S–SDFs.** The two panels report the OOS GLS adjusted $R^2$ of the cross-sectional regressions of average excess returns on estimated factor loadings for different empirical asset pricing models. The red, black and blue curves report the OOS GLS adjusted $R^2$ for the forward-looking APT–consistent minimum variance S–SDFs described in Section 4.3, based on pricing error functions given for $\lambda \in [0,1]$ by $h = (1 - \lambda) ||\cdot||_1 + \lambda ||\cdot||_2$ (left panel) and $h = (1 - \lambda) \sqrt{N_D} ||\cdot||_\infty + \lambda ||\cdot||_2$ (right panel). We compute these S–SDFs for various fractions, reported on the x-axis, of the empirical maximal pricing error thresholds $\{\tau^m_{\max}\}$. The dashed horizontal line reports the OOS GLS adjusted $R^2$ achieved by the Fama-French three-factor model. The dashed-dotted horizontal line reports the OOS GLS adjusted $R^2$ attained by the data-driven minimum variance S–SDF (OPT(A)) with $l_2$ APT–consistent pricing error function and no forward-looking bias; see again Section 4.3 for details. The intermediate dataset is used with an OOS period starting in July 1963 and ending in June 2018. For each of these S–SDFs, the risk-free asset is the only sure asset and sorted portfolios constitute the dubious assets. See Section 4.3 for details.
Figure 14: **Optimal APT–consistent OOS S–SDFs time series.** We report, from top to bottom, the time series of following out-of-sample S–SDFs from Figure 13: (A), OPT(A), (B) and (C). The out-of-sample valuation period starts in July 1963 and ends in June 2018. Shaded areas correspond to NBER recessions. See Section 4.3 for details.
Figure 15: **OOS comparison of minimum variance (H-J) and minimum Kullback Leibler (KL) dispersion S–SDFs.** The figure reports the OOS GLS adjusted $R^2$'s of the cross-sectional regressions of average excess returns in the low dimensional dataset (left panels) and the intermediate dimensional dataset (right panels) on estimated factor loadings of minimum H-J dispersion (top panels) and minimum KL dispersion (bottom panels) S–SDFs for various APT–consistent pricing error functions. The OOS period starts in July 1963 and ends in June 2018. The solid black curve depicts the OOS GLS adjusted $R^2$'s obtained with an APT–consistent $l_2$ pricing error function, while the dashed red lines correspond to a pricing error function given for various $\lambda \in (0, 1)$ by $h = (1 - \lambda) \| \cdot \|_1 + \lambda \| \cdot \|_2$. In every panel, the risk-free asset is the only sure asset, while sorted portfolios constitute the dubious assets. See Section 4.3 for details.
Figure 16: OOS comparison of minimum variance (H-J) and minimum Kullback Leibler (KL) dispersion S–SDFs. The figure reports the OOS GLS adjusted $R^2$s of the cross-sectional regressions of average excess returns in the low dimensional dataset (left panels) and the intermediate dimensional dataset (right panels) on estimated factor loadings of minimum H-J dispersion (top panels) and minimum KL dispersion (bottom panels) S–SDFs for various APT–consistent pricing error functions. The OOS period starts in July 1963 and ends in June 2018. The solid black curve depicts the OOS GLS adjusted $R^2$s obtained with an APT–consistent $l_2$ pricing error function, while the dashed red lines correspond to a pricing error function given for various $\lambda \in (0, 1)$ by $h = (1 - \lambda) \sqrt{N_D} ||\cdot||_\infty + \lambda ||\cdot||_2$. In every panel, the risk-free asset is the only sure asset, while sorted portfolios constitute the dubious assets. See Section 4.3 for details.
Figure 17: **OOS portfolios cumulative returns.** The figure reports the cumulative returns of a one dollar investment in different portfolios, financed by borrowing at the risk-free rate: the market portfolio (Mkt, yellow line), Portfolio (49) leveraged to have the same volatility as the market portfolio ($R^{(m)}_y$, blue line), Portfolio (50) leveraged to have the same volatility as the market portfolio ($\bar{R}^{(m)}_y$, red line), portfolio SMB (purple line), portfolio HML (green line) and the equally-weighted portfolio of all sorted portfolios in the intermediate dataset (EW, light blue line). The out-of-sample valuation period starts in July 1963 and ends in June 2018. Shaded areas correspond to NBER recessions.
Table 1: Two-step out-of-sample asset pricing tests

|        | Const(%) | \(\lambda_1\) (%) | \(\lambda_2\) (%) | \(\lambda_3\) (%) | \(\lambda_4\) (%) | \(\lambda_5\) (%) | \(R^2_{OLS}\)(%) | \(R^2_{GLS}\)(%)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>OPT(A)</td>
<td>0.86</td>
<td>-48.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22.07</td>
<td>41.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(36.65)</td>
<td>(-7.33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A)</td>
<td>0.80</td>
<td>-16.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24.45</td>
<td>44.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(46.95)</td>
<td>(-7.82)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(B)</td>
<td>0.65</td>
<td>-16.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18.58</td>
<td>32.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(33.51)</td>
<td>(-6.59)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(C)</td>
<td>0.89</td>
<td>-41.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12.25</td>
<td>39.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(26.17)</td>
<td>(-5.19)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FF3</td>
<td>1.01</td>
<td>-0.40</td>
<td>0.23</td>
<td>0.16</td>
<td></td>
<td></td>
<td>27.51</td>
<td>6.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.91)</td>
<td>(-3.26)</td>
<td>(6.27)</td>
<td>(3.12)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FF5</td>
<td>0.71</td>
<td>-0.13</td>
<td>0.29</td>
<td>0.09</td>
<td>-0.01</td>
<td>0.34</td>
<td>29.23</td>
<td>5.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.53)</td>
<td>(-0.86)</td>
<td>(7.42)</td>
<td>(3.60)</td>
<td>(-0.16)</td>
<td>(3.60)</td>
<td></td>
</tr>
<tr>
<td>PCA3</td>
<td>0.99</td>
<td>-3.22</td>
<td>2.02</td>
<td>0.06</td>
<td></td>
<td></td>
<td>22.70</td>
<td>5.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8.26)</td>
<td>(-2.01)</td>
<td>(5.67)</td>
<td>(0.32)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table summarizes the results of out-of-sample two-step cross-sectional regressions of average excess returns on the estimated factor loadings of different empirical asset pricing models. The columns report from the left to the right the estimated intercept, factor risk premia, adjusted \(R^2\) and GLS adjusted \(R^2\) in percentage terms, with t-statistics in parenthesis. The first row reports results for the data-driven S–SDF OPT(A) with no forward-looking bias described in Section 4.3. Rows two to four report results for the S–SDFs (A), (B) and (C) with forward-looking bias from Figure 13; see again Section 4.3 for details. The fifth and sixth rows report results, respectively, for the Fama-French three-factor and five-factor models. Row seven reports results for a linear SDF based on the first three PCA factor returns computed using data from the out-of-sample period. The intermediate dataset is used with an out-of-sample period starting in July 1963 and ending in June 2018.
Table 2: Two-step out-of-sample tests of various single factor linear asset pricing models

<table>
<thead>
<tr>
<th></th>
<th>Const(%)</th>
<th>λ(%)</th>
<th>$R^2_{OLS}$ (%)</th>
<th>$R^2_{GLS}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_y^{(m)}$</td>
<td>0.78</td>
<td>30.99</td>
<td>34.40</td>
<td>40.61</td>
</tr>
<tr>
<td></td>
<td>(52.47)</td>
<td>(9.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{R}_y^{(m)}$</td>
<td>0.78</td>
<td>1.31</td>
<td>28.69</td>
<td>38.95</td>
</tr>
<tr>
<td></td>
<td>(49.45)</td>
<td>(8.71)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mkt</td>
<td>0.54</td>
<td>0.17</td>
<td>0.99</td>
<td>-0.19</td>
</tr>
<tr>
<td></td>
<td>(5.00)</td>
<td>(1.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMB</td>
<td>0.54</td>
<td>0.24</td>
<td>16.78</td>
<td>1.71</td>
</tr>
<tr>
<td></td>
<td>(15.56)</td>
<td>(6.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HML</td>
<td>0.77</td>
<td>0.17</td>
<td>4.00</td>
<td>3.07</td>
</tr>
<tr>
<td></td>
<td>(34.99)</td>
<td>(2.96)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EW</td>
<td>0.39</td>
<td>0.34</td>
<td>6.31</td>
<td>-0.37</td>
</tr>
<tr>
<td></td>
<td>(4.27)</td>
<td>(3.69)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table summarizes the results of out-of-sample two-step cross-sectional regressions of average excess returns on the estimated factor loadings of different single factor empirical asset pricing models. The columns report from the left to the right the estimated intercept, factor risk premium, adjusted $R^2$ and GLS adjusted $R^2$ in percentage terms, with $t$-statistics in parenthesis. The first two rows report results for the data-driven excess returns $R_y^{(m)}$ and $\tilde{R}_y^{(m)}$ described in Section 4.4. Rows three to six report results for single-factor models based on the three Fama-French excess return factors and the equally weighted portfolio excess return, respectively. The intermediate dataset is used with out-of-sample period starting in July 1963 and ending in June 2018.
Table 3: Summary statistics of various excess returns

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Sd</th>
<th>Sharpe ratio</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_y^{(m)}$</td>
<td>0.301</td>
<td>0.740</td>
<td>0.407</td>
<td>0.541</td>
<td>11.894</td>
</tr>
<tr>
<td>$\tilde{R}_y^{(m)}$</td>
<td>0.015</td>
<td>0.036</td>
<td>0.414</td>
<td>0.967</td>
<td>17.816</td>
</tr>
<tr>
<td>Mkt</td>
<td>0.005</td>
<td>0.044</td>
<td>0.121</td>
<td>-0.539</td>
<td>5.029</td>
</tr>
<tr>
<td>SMB</td>
<td>0.002</td>
<td>0.031</td>
<td>0.073</td>
<td>0.493</td>
<td>8.426</td>
</tr>
<tr>
<td>HML</td>
<td>0.003</td>
<td>0.028</td>
<td>0.118</td>
<td>0.087</td>
<td>5.103</td>
</tr>
<tr>
<td>EW</td>
<td>0.007</td>
<td>0.050</td>
<td>0.147</td>
<td>-0.525</td>
<td>5.670</td>
</tr>
</tbody>
</table>

The table reports out-of-sample summary statistics of various monthly excess return strategies. The columns report from the left to the right the sample mean, standard deviation, Sharpe ratio, skewness and kurtosis of various monthly portfolio excess returns. The first two rows report results for the data-driven excess returns $R_y^{(m)}$ and $\tilde{R}_y^{(m)}$ described in Section 4.4. Rows three to six report results for the excess returns on the three Fama-French factors and on the equally weighted portfolio, respectively. The intermediate dataset is used with out-of-sample period starting in July 1963 and ending in June 2018.

References


