Journal of Financial Econometrics, 2025, **23(1)**, nbae033
https://doi.org/10.1093/jjfinec/nbae033
Article

OXFORD

An Information-Theoretic Asset Pricing Model

Anisha Ghosh [1], Christian Julliard [2], and Alex P. Taylor [3]

[1]Desautels Faculty of Management, McGill University, Montreal, Canada
[2]Department of Finance and FMG, London School of Economics, and CEPR, London, UK
[3]Department of Finance, Manchester Business School, Manchester, UK

Address correspondence to Anisha Ghosh, Desautels Faculty of Management, McGill University, Montreal, Canada, or e-mail: anisha.ghosh@mcgill.ca

Abstract

We show that a non-parametric estimate of the pricing kernel, extracted using an information-theoretic approach, delivers smaller out-of-sample pricing errors and a better cross-sectional fit than leading multi-factor models. The information stochastic discount factor (I-SDF) identifies sources of risk not captured by standard factors, generating very large annual alphas (20–37%) and Sharpe ratio (1.1). The I-SDF extracted from a wide cross-section of equity portfolios is highly positively skewed and leptokurtic, and implies that about a third of the observed risk premia represent compensation for 2.5% tail events. The I-SDF offers a powerful benchmark relative to which competing theories and investment strategies can be evaluated.

Keywords: alpha, cross-sectional asset pricing, factor mimicking portfolios, factor models, pricing kernel, relative entropy

JEL classifications: G11, G12, C13, C53

Asset prices contain information about the stochastic discounting of possible future states, that is, about the pricing kernel, or stochastic discount factor (SDF). Based on this simple observation, and an information-theoretic approach, we propose a novel non-parametric method for the estimation of the pricing kernel and out-of-sample pricing of asset returns.

The proliferation of risk factors identified in the empirical asset pricing literature has brought forth concerns over data mining and spurious inference (see, e.g., Lewellen, Nagel, and Shanken 2010; Harvey and Liu 2015; McLean and Pontiff 2016; Bryzgalova 2016), and highlights the risk of over-parameterization of the pricing kernel. Therefore, a non-parametric approach to the recovery of the pricing kernel is a potentially valuable alternative to the *ad-hoc* construction of risk factors. Moreover, given its strong empirical

We benefited from helpful comments from Andrew Ang, George Constantinides, Magnus Dahlquist, Francis Diebold, Ralph Koijen, Dong Lou, Ian Martin, Toby Moskowitz, Christopher Polk, Tarun Ramadorai, Robert Stambaugh, Romeo Tedongap, Raman Uppal, Jessica Wachter, Irina Zviadadze, and seminar participants at Wharton, ICEF, Imperial College Business School, LSE, Rotman School of Management, Stockholm School of Economics, ESSFM 2015 Conference in Gerzensee, and the EEA 2015 Annual Meeting. Any errors or omissions are the responsibility of the authors. Christian Julliard thanks the Economic and Social Research Council (UK) [grant number: ES/K002309/1] for financial support.

Received: August 17, 2022. **Revised:** October 9, 2024. **Editorial decision:** October 22, 2024. **Accepted:** November 25, 2024

performance, it provides a benchmark model relative to which competing theories, as well as investment managers, can be evaluated.

The use of information-theoretic (or entropy-based) techniques has become increasingly popular since their introduction in the finance literature by Stutzer (1995, 1996) and Kitamura and Stutzer (2002). Recent research has found the information-theoretic approach to be a useful tool for addressing a number of central questions in economics and finance. Examples include Julliard and Ghosh (2012), who rely on this entropy-based inference approach to assess the empirical plausibility of the rare disasters hypothesis in explaining the equity premium puzzle; Almeida and Garcia (2012) and Backus, Chernov, and Zin (2014), who propose entropy-based performance measures to assess candidate asset pricing models (see also Bansal and Lehmann 1997; Almeida and Garcia 2016; Liu 2021); Borovicka, Hansen, and Scheinkman (2016), who propose using relative entropy minimization techniques to isolate a positive martingale component of the SDF process that contains information about long-term risk adjustments (see also Alvarez and Jermann 2005); Ghosh, Julliard, and Taylor (2017), who use a relative entropy minimization approach to derive entropy bounds for the SDF that are tighter and more flexible than the seminal Hansen and Jagannathan (1991, 1997) bounds, as well as recover the unobservable components of the SDF (e.g., habits, return on total wealth, etc.) for a broad class of consumption-based asset pricing models (see also Sandulescu, Trojani, and Vedolin 2021); Chen, Hansen, and Hansen (2020) and Ghosh and Roussellet (2020), who propose using the approach to recover investors' beliefs from observed asset prices; and Ghosh, Julliard, and Stutzer (2020), who use the approach to estimate the welfare costs of aggregate economic fluctuations.

In the above literature, the information-theoretic technique is typically used to construct an SDF *in sample*, which, by construction, correctly prices the cross-section of assets used in its estimation. Although this provides a very useful tool for the aforementioned applications, the fact that the test assets are priced perfectly means that standard tests of empirical asset pricing, where the success or failure of an asset pricing model is usually decided, are not meaningful. Indeed, it is an open question whether SDFs extracted using information-theoretic methods have, by the standards of conventional empirical asset pricing tests, any real ability to price assets *out-of-sample*. Given the increasing use of entropy techniques, we take this question to be of considerable importance. Therefore, in this article, we assess the ability of the entropy-based SDF to price broad cross-sections of assets out-of-sample, and compare its performance with those of the leading factor models popular in the empirical asset pricing literature. Our results suggest that the entropy-based SDF offers a powerful benchmark for asset pricing.

Ghosh, Julliard, and Taylor (2017) show how a pricing kernel can be estimated in the sample in a non-parametric fashion using only no arbitrage restrictions. Specifically, given the time series data of returns on a cross-section of assets, they utilize a model-free relative entropy minimization approach to estimate an SDF that prices the cross-section. The resulting SDF is a non-linear function of the asset returns and the Lagrange multipliers associated with the assets' cross-sectional pricing restrictions (i.e., the shadow value of relaxing the Euler equation restrictions).

The main methodological contribution of the present article is to extend the minimum-entropy SDF approach out-of-sample for the purposes of cross-sectional pricing. In particular, using the Lagrange multipliers estimated in a training sample, we construct the out-of-sample SDF in a rolling fashion, and use it as the single factor to price different cross-sections of test assets. This method ensures non-negativity, hence validity, of the SDF in- and out-of-sample. Our approach does not require taking a stance on either the number or the identity of the underlying risk factors or on the functional form of agents' risk preferences. Instead, the approach summarizes all the relevant pricing information (contained in, potentially multiple, priced risk factors) in the form of a single time series for the SDF. We refer to the out-of-sample SDF as the "Information SDF" (I-SDF). The question then becomes whether our non-parametric approach to the recovery of the pricing kernel provides meaningful information for the pricing of assets out-of-sample in empirically realistic scenarios. And if so, whether it can be considered a valuable alternative to the *ad-hoc* construction of risk factors, or even PCA-based methods, commonly used as proxies for the underlying sources of priced risk.

The relative entropy minimization approach delivers an SDF that is theoretically distinct from the mean–variance efficient tangency portfolio of the set of assets used in its construction. Specifically, the former approach recovers an I-SDF that has the minimum weighted sum of *all* of its moments. Recall that the tangency portfolio, on the other hand, is chosen to have the minimum variance. If the true underlying pricing kernel were log normal, then all of its higher order moments would be zero, such that minimizing relative entropy would be equivalent to minimizing the variance. If, on the other hand, the SDF is not log normal (e.g., if we believe that tail risk is an important source of priced risk), then the variance is not a sufficient statistic summarizing its distribution, and there is no *a priori* reason to rely on a solely variance minimizing criterion to recover the SDF.

One natural limitation of an entropy-driven approach is that, as any other large-T moment based estimator, it is more reliable with a relatively small cross-sectional dimension relative to the time series one. We address this limitation in two possible ways, by either: *a*) selecting a relatively small (15) set of anomaly portfolios, known to have good spanning properties for asset returns, as base assets to construct the I-SDF; and *b*) when considering a large set (135) of base assets to recover the SDF, we employ a L_1-penalization to deal with the high-dimensional moment conditions.

In the first case, we estimate the I-SDF from a small cross-section of 15 equity portfolios, that capture several well-known asset pricing anomalies—including portfolios sorted on the basis of size, book-to-market-equity, momentum, industry, and short- and long-term reversals—and analyze its ability to explain out of sample several broader cross-sections of test assets. Compared to leading multifactor models, such as the Fama and French (1993, 2015) three- and five-factor models (FF3 and FF5, respectively) or the Hou, Xue, and Zhang (2015) q-factor model (HXZ), the I-SDF typically delivers smaller pricing errors on all the different sets of test assets that we consider. Moreover, it explains a larger fraction of the cross-sectional variation of the returns. These results hold for a variety of measures commonly used in the literature to assess the cross-sectional fit in addition to the standard OLS R^2. We also show that the I-SDF (compared to the other factor models considered) more closely identifies—out-of-sample—the tangency portfolio, that is, the maximum Sharpe ratio portfolio. Furthermore, we find that the I-SDF extracts novel pricing information not captured by the FF3, HXZ4, or FF5 models: it leads to an "information anomaly," generating large and statistically significant intercepts (12.4–13.4% per annum) relative to these factor models, which only explain less than 13% of its time series variation.

When dealing with a large set of base assets to construct the SDF, we employ an extension of the penalized information-theoretic method proposed in Qiu and Otsu (2022). This enables the recovery of an I-SDF from a large cross-section, potentially greater than the length of the available time series. Specifically, we recover the I-SDF from a set of 135 equity portfolios that encompasses all the major anomaly variables identified in the literature—the five industry portfolios and the ten decile portfolios for univariate sorts on the basis of each of the following characteristics: size, book-to-market-equity, momentum, short-term reversals, long-term reversals, operating profitability, investment, accruals, new issues, beta, variance, residual variance, and earnings-to-price. Consistent with the results for the smaller cross-section of base assets, the I-SDF retains its favorable performance vis-á-vis the other factor models. Furthermore, even in this large N case, our entropy-based approach outperforms the PCA-based method of Kozak, Nagel, and Santosh (2020).

Overall, our results suggest that the I-SDF is well suited for studying broad cross-sections of assets, for example, portfolios, that have been explored in the literature with standard methods. It offers a single factor that prices such assets substantially better than standard benchmarks. However, being a large T moments-based method, without imposing a regularization as the one that we consider, the method might not be amenable to extracting a unique I-SDF that prices the universe of assets, such as, for example, the entire cross-section of individual stocks.

Furthermore, we offer an economic interpretation of the I-SDF. The I-SDF has a strong business cycle pattern: it is smaller during expansions and larger during recessions. However, the I-SDF also exhibits another interesting feature over and above its business cycle properties: it is particularly high when the recessions are concomitant with big downward movements in the stock market, such as during the 1973–1975, 2001, and 2007–2009 recessions (unlike the 1981–1982 and 1990–1991 recessions that were not accompanied by big stock market crash episodes). Our results suggest that while business cycle risk is an important source of systematic risk, it cannot fully explain the behavior of the I-SDF, and that the true underlying SDF may not be a function of business cycle variables alone, but also of additional variables related to the performance of the overall stock market. Also, the I-SDF has a strongly non-Gaussian distribution: it is highly positively skewed and leptokurtic. Furthermore, it implies that about one-half of the observed risk premia of stocks represent compensation for tail risk.

Our article contributes to the extensive cross-sectional asset pricing literature that seeks to identify priced risk factors to explain the cross-section of returns of different classes of financial assets. Harvey, Liu, and Zhu (2015) document 316 risk factors discovered by academics. Lewellen, Nagel, and Shanken (2010) offer a critical assessment of asset pricing tests and conclude that although many of the proposed factors seem to perform well in terms of producing high cross-sectional R^2 and small pricing errors, this result is largely driven by the strong factor structure of the size and book-to-market-equity sorted portfolio returns (which are often used as the sole test assets), which makes it quite likely for an arbitrarily chosen two or three factors, which have little correlation with the returns, to produce these results. Moreover, a large literature (e.g., Kan and Zhang 1999a, 1999b; Kleibergen 2009; Kleibergen and Zhan 2015), stresses that the apparent good performance of several factor models proposed in the literature might be the spurious outcome of a weak identification problem. We show that our I-SDF is robust to these concerns and that our approach provides a reliable benchmark against which competing models can be evaluated. Furthermore, Bryzgalova, Huang, and Julliard (2023) show that the true latent SDF is dense in the space of observable asset pricing factors, and that all low-dimensional factor models in the previous literature are misspecified with very high probability. Our method overcomes this issue by delivering an optimal non-linear combination of a large set of anomaly portfolios, that is, extracting a single SDF from a dense space of observable factors.

In spirit, our article is also close to, and builds upon, the long tradition of using asset prices to estimate the risk-neutral probability measure (see, e.g., Jackwerth and Rubinstein 1996; Ait-Sahalia and Lo 1998; Almeida and Freire 2022) and use this information to extract an implied pricing kernel (see, e.g., Ait-Sahalia and Lo 2000; Rosenberg and Engle 2002; Hansen 2014; Ross 2015). The main advantages of our approach relative to this literature are that *a*) we do not need to rely exclusively on options data that are only available over a much shorter sample period, and *b*) we can construct an out-of-sample pricing kernel and maximum Sharpe ratio portfolio. Moreover, as we show, our method is very general and can be applied to other asset classes including currencies, commodities, as well as international equities. In the usage of a L_1-norm penalty as regularization, our work is also connected to the recent developments on the estimation of the SDF in large cross-sections (e.g., Kozak, Nagel, and Santosh 2020; Korsaye, Quaini, and Trojani 2021).

The remainder of this article is organized as follows. Section 1 describes our method of extracting the pricing kernel from a vector of asset returns, as well as the different inference methods used in the empirical analysis. The data used in the empirical analysis are described in Section 2. The empirical performance of the I-SDF in explaining broad cross-sections of returns is presented in Section 3. Section 4 discusses the properties of the I-SDF. Section 5 concludes with suggestions for future research.

1 The Method

Our relative-entropy minimizing approach enables us to recover, for a given cross-section of assets, what we refer to as the I-SDF. Section 1.1 describes the information-theoretic method used to construct the I-SDF. Section 1.2 discusses the econometric tests used to assess the pricing performance of the I-SDF and compares its performance to some leading empirical asset pricing models commonly used in the literature.

1.1 Recovery of the I-SDF

The absence of arbitrage opportunities implies the existence of a strictly positive pricing kernel (also known as the SDF), M, such that the expectation of the product of the kernel and a vector of excess returns, $\mathbf{R}_t^e \in \mathbb{R}^N$, is zero under the physical probability measure, $\mathbb{P}$:

$$\mathbf{0} = \mathbb{E}^{\mathbb{P}}\left[M_t \mathbf{R}_t^e\right] = \int M_t \mathbf{R}_t^e d\mathbb{P},$$

where $\mathbf{0}$ denotes a conformable vector of zeros. Under weak regularity conditions, the above restrictions on the SDF can be rewritten as

$$\mathbf{0} = \int \frac{M_t}{\bar{M}} \mathbf{R}_t^e d\mathbb{P} = \int \mathbf{R}_t^e d\mathbb{Q} \equiv \mathbb{E}^{\mathbb{Q}}\left[\mathbf{R}_t^e\right], \tag{1}$$

where $\bar{x} := \mathbb{E}[x_t]$, and $\frac{M_t}{\bar{M}} = \frac{d\mathbb{Q}}{d\mathbb{P}}$ is the Radon–Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$. This change of measure is legitimate if the measure $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$. The above transformation also implies that using only the Radon–Nikodym derivative in the Euler equation for excess returns, one can recover the SDF only up to a scale. As below, we will use the Euler equation for the risk-free rate to recover the scale factor, hence the level of the SDF.

Given the above, an estimate of the risk-neutral probability measure $\mathbb{Q}$ can be obtained as the minimizer of its relative entropy with respect to the physical measure $\mathbb{P}$, that is, as[1]

$$\arg\min_{\mathbb{Q}} D(\mathbb{Q}||\mathbb{P}) \equiv \arg\min_{\mathbb{Q}} \int \frac{d\mathbb{Q}}{d\mathbb{P}} \ln\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right) d\mathbb{P} \quad \text{s.t.} \int \mathbf{R}_t^e d\mathbb{Q} = \mathbf{0}, \tag{2}$$

where $D(\mathbb{A}||\mathbb{B}) := \int \ln \frac{d\mathbb{A}}{d\mathbb{B}} d\mathbb{A} \equiv \int \frac{d\mathbb{A}}{d\mathbb{B}} \ln \frac{d\mathbb{A}}{d\mathbb{B}} d\mathbb{B}$ denotes the relative entropy of $\mathbb{A}$ with respect to $\mathbb{B}$, that is, the Kullback–Leibler Information Criterion (KLIC) divergence between $\mathbb{A}$

1 Minimizing the relative entropy to recover the risk neutral probability measure was first suggested by Stutzer (1995). Ghosh, Julliard, and Taylor (2017) extended the method to recover the unobserved component of the SDF for a broad class of consumption-based asset pricing models as well as to construct entropy bounds on the SDF and its components that are tighter and more flexible than the seminal Hansen–Jagannathan bounds.

and $\mathbb{B}$ (White 1982). Note that $D(\mathbb{A}||\mathbb{B})$ is always non-negative, and has a minimum at zero that is attained when $\mathbb{A}$ is identical to $\mathbb{B}$. This divergence measures the additional information content of $\mathbb{A}$ relative to $\mathbb{B}$ and, as pointed out by Robinson (1991), is very sensitive to any deviation of one probability measure from another. Therefore, the optimization in Equation (2) is a relative entropy minimization under the asset pricing restrictions coming from the Euler Equation (1).

There are many different functions that could be used to measure the divergence between two probability measures ($\mathbb{Q}$ and $\mathbb{P}$ in this case). Equation (2) relies on one particular choice of this function. In fact, if the function is chosen to be quadratic, minimizing the divergence between $\mathbb{P}$ and $\mathbb{Q}$ is akin to the continuous updating GMM estimator. An important advantage of using the function in Equation (2) is that the objective function can be re-written as a weighted sum of *all* the moments of the distribution of $\log(\mathbb{Q})$, whereas the quadratic choice of the function is akin to minimizing the variance. If the underlying pricing kernel is not lognormal, its variance would not be a sufficient statistic for its distribution and there is no obvious reason to minimize the variance alone.

Ghosh, Julliard, and Taylor (2017) show that the above approach to the recovery of the pricing measure has several desirable properties. First, the estimation in Equation (2) delivers a non-parametric maximum likelihood estimate of the risk-neutral measure. A formal proof of this property of the estimator is as follows. Consider the following procedure for constructing the series $\{q_t\}_{t=1}^T$ (up to a positive scale). Given an integer $N \gg 0$, distribute randomly to the various points in time $t = 1,\ldots,T$, the value $1/N$ in N independent draws. That is, draw a series of probability weights $\{\tilde{q}\}_{t=1}^T$, given by $\tilde{q}_t \equiv \frac{n_t}{N}$, where n_t measures the number of times that the value $1/N$ has been assigned to time t. Subsequently, check whether the drawn series $\{\tilde{q}\}_{t=1}^T$ satisfies the asset pricing restrictions $\sum_{t=1}^T R_t^e \tilde{q}_t = 0$. If it does, use this series as the estimator of $\{q_t\}_{t=1}^T$, and if it does not, draw another series. In other words, an estimate for q_t would correspond to the most likely outcome of the above procedure. Noticing that the distribution of the $\tilde{q}_t$ is, by construction, a multinomial distribution with support given by the historical sample, the likelihood of any particular sequence $\{\tilde{q}_t\}_{t=1}^T$ is given by:

$$L\left(\{\tilde{q}_t\}_{t=1}^T\right) = \frac{N!}{n_1!n_2!\ldots n_T!} \times T^{-N} = \frac{N!}{N\tilde{q}_1!N\tilde{q}_2!\ldots N\tilde{q}_T!} \times T^{-N}.$$

Hence, as $N \to \infty$ (and, therefore, the approach becomes more accurate), the log likelihood is given by[2]

$$\lim_{N\to\infty} \ln L\left(\{\tilde{q}_t\}_{t=1}^T\right) = -\sum_{t=1}^T \tilde{q}_t \ln \tilde{q}_t.$$

Therefore, taking into account the asset pricing constraints, the MLE of q_t solves

$$\{\hat{q}_t\}_{t=1}^T \equiv \arg\max - \sum_{t=1}^T \tilde{q}_t \ln \tilde{q}_t, \quad \text{s.t.} \quad \{\tilde{q}_t\}_{t=1}^T \in \Delta^T, \ \sum_{t=1}^T \mathbf{R}_t^e \tilde{q}_t = 0.$$

Note that the solution to the above likelihood maximization problem is also the solution of the relative entropy minimization problem in Equation (2) (see, e.g., Csiszar 1975). Therefore, the KLIC minimization is equivalent to maximizing the likelihood of finding the risk-neutral measure q_t.

2 Recall that from Stirling's formula, we have $\lim_{N\tilde{q}_t\to\infty} \frac{N\tilde{q}_t!}{\sqrt{2\pi N\tilde{q}_t}\left(\frac{N\tilde{q}_t}{e}\right)^{N\tilde{q}_t}} = 1.$

The second desirable property of this approach to the recovery of the risk-neutral measure is that, due to the presence of the logarithm in the objective function in Equation (2), the use of relative entropy naturally enforces the non-negativity of the pricing kernel—a property that, as shown below, extends to the out-of-sample pricing kernel. Note that the absence of arbitrate opportunities implies the existence of a strictly positive SDF that prices assets (see, e.g., Harrison and Kreps 1979). In other words, negative realizations of the SDF imply the presence of arbitrage opportunities that should not be present in well-functioning financial markets. SDFs that take on negative values, particularly for a non-trivial proportion of the time, are, therefore, indicative of misspecification and should be interpreted with this caveat even if they price assets well. It is for this reason that, in theoretical asset pricing models, the SDF is modeled as being strictly positive. Thus, we consider an SDF that takes on negative values as not being valid. Third, the approach satisfies Occam's razor, or the law of parsimony, since it adds the minimum amount of information needed for the pricing kernel to price assets. Fourth, it is straightforward to add conditioning information: given a vector of conditioning variables $\mathbf{Z}_{t-1}$, one simply has to multiply (element by element) the argument of the integral constraint in Equation (2) by the conditioning variables in $\mathbf{Z}_{t-1}$. Fifth, there is no *ex-ante* restriction on the number of assets that can be used in constructing M. Nevertheless, as for other moment-based estimators (e.g., GMM), the cross-sectional dimension (the number of assets) should be ideally kept small relative to the length of the time series used.[3] Sixth, as implied by Brown and Smith (1990), the use of entropy is desirable if one believes that tail events are an important component of the risk measure (minimum entropy estimators endogenously re-weight the observations to appropriately account for tail events that happened to occur in the data with a frequency lower than their true probability).[4]

In this article, we focus on the out-of-sample asset pricing performance of an SDF constructed using the above relative entropy minimization approach. In particular, note that since $\frac{M_t}{\bar{M}} = \frac{d\mathbb{Q}}{d\mathbb{P}}$, the optimization in Equation (2) can be rewritten as

$$\underset{M_t}{\text{argmin}}\ \mathbb{E}^{\mathbb{P}}[M_t \ln M_t] \quad \text{s.t.} \quad \mathbb{E}^{\mathbb{P}}\left[M_t \mathbf{R}_t^e\right] = \mathbf{0},$$

where, to simplify the exposition, we have used the normalization $\bar{M} = 1$ without loss of generality.[5] Given a sample of size T and a history of excess returns $\{\mathbf{R}_t^e\}_{t=1}^T$, the above expression can be made operational by replacing the expectation with a sample analogue, as is customary for moment-based estimators,[6] obtaining

$$\underset{\{M_t\}_{t=1}^T}{\arg\min} \frac{1}{T}\sum_{t=1}^{T} M_t \ln M_t \quad \text{s.t.} \quad \frac{1}{T}\sum_{t=1}^{T} M_t \mathbf{R}_t^e = \mathbf{0}. \tag{3}$$

The above formulation is convenient because a solution is easily obtainable via Fenchel's duality (see, e.g., Csiszar 1975):

[3] The approach does not require a decomposition of M into short- and long-run components (cf. Alvarez and Jermann 2005), and it does not rely on the existence of a continuum of options price data (cf. Ross 2015).

[4] Based on this insight, Julliard and Ghosh (2012) used a relative entropy estimation approach to analyse the empirical plausibility of the rare events hypothesis to explain a host of asset pricing puzzles.

[5] This normalization is innocuous since the estimate of M_t is identified up to a strictly positive scale constant. This positive scale constant can be recovered from the Euler equation for the risk free rate. Specifically, the recovered I-SDF $\hat{M}_t = kM_t$, where M_t denotes the true underlying SDF and k is a positive constant. Since the true SDF must price the risk free asset, we have, $E[M_t] = E[\hat{M}_t]/k = E[1/R_{f,t}]$. Hence, the constant k can be recovered as $k = E[\hat{M}_t]/E[1/R_{f,t}]$.

[6] This amounts to assuming ergodicity for both the pricing kernel and asset returns.

$$\hat{M}_t \equiv M_t\left(\hat{\theta}_T, \mathbf{R}_t^e\right) = \frac{Te^{\hat{\theta}'_T \mathbf{R}_t^e}}{\sum_{t=1}^{T} e^{\hat{\theta}'_T \mathbf{R}_t^e}}, \quad \forall t, \tag{4}$$

where $\hat{\theta} \in \mathbb{R}^N$ is the vector of Lagrange multipliers that solve the unconstrained convex problem

$$\hat{\theta}_T := \underset{\theta}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^{T} e^{\theta' \mathbf{R}_t^e}, \tag{5}$$

and this last expression is the dual formulation of the entropy minimization problem in Equation (3). The above duality result implies that the number of free parameters available in estimating $\{M_t\}_{t=1}^{T}$ is equal to the dimension of (the Lagrange multiplier) θ: that is, it is simply equal to the number of assets considered in the Euler equation.

Note that since relative entropy is not symmetric, that is, $D(\mathbb{Q}||\mathbb{P}) \neq D(\mathbb{P}||\mathbb{Q})$, we can reverse the roles of the probability measures $\mathbb{P}$ and $\mathbb{Q}$ in Equation (2) to obtain an alternative definition of relative entropy and, therefore, a second approach to estimating the pricing kernel. This approach is described in Appendix A.1. The empirical results obtained using this approach are similar to those obtained using Equation (2) and are, hence, omitted for brevity. Furthermore, this alternative approach does not guarantee the non-negativity of the SDF out-of-sample.[7]

We use the above method to recover the time series of the SDF in a rolling out-of-sample fashion. In particular, for a given cross-section of asset returns, we divide the time series of returns into rolling subsamples of length $\bar{T}$ and final date T_i, $i = 1, 2, 3, \ldots$, and constant $s := T_{i+1} - T_i$. In subsample i, we estimate the vector of Lagrange multipliers $\hat{\theta}_{T_i}$ by solving the minimization in Equation (5). Using the estimates of the Lagrange multipliers, $\hat{\theta}_{T_i}$, the out-of-sample I-SDF $M(\hat{\theta}_{T_i}, \mathbf{R}_t^e)$ is obtained for the subsequent s periods (i.e., for t such that $T_i + 1 \leq t \leq T_{i+1}$) using Equation (4). This process is repeated for each subsample to obtain the time series of the estimated kernel over the out-of-sample evaluation period. Note that this procedure ensures, thanks to the exponential form in Equation (4), a non-negative SDF not only in-sample but also out-of-sample.

This procedure is analogous in spirit to the canonical approach of forming portfolios (e.g., the SMB and HML portfolios) based on past asset return characteristics (e.g., by sorting on size and book-to-market-equity in the past calendar year). The key difference is that $M(\hat{\theta}_{T_i}, \mathbf{R}_t^e)$ is a non-linear function of the portfolio $\hat{\theta}'_{T_i} \mathbf{R}_t^e$ and the weights θ are chosen to deliver an MLE of the SDF in each (past) subsample.

In our baseline empirical analysis, we set $s = 12$ months, corresponding to annual rebalancing. The size of the rolling window, $\bar{T}$, is set to 30 years, allowing the out-of-sample tests to begin in 1963:07.

1.2 Asset Pricing Tests

For a given cross-section of test assets, we construct the out-of-sample I-SDF using the procedure described in Section 1.1. We compare the performance of the I-SDF to that of the one-factor CAPM, the three and five factor Fama–French models (FF3 and FF5, respectively), the Hou, Xue, and Zhang q-factor model (HXZ), the principal components based

[7] More generally, both of these entropy based criteria corresponds to two particular cases of the so-called Cressie–Read family of discrepancies (see, e.g., Almeida, Ardison, and Garcia 2020).

SDF of Kozak, Nagel, and Santosh (2020) (KNS), and the minimum variance SDF of Hansen and Jagannathan (1991).

We use the standard two-step regression approach (with Shanken correction of the standard errors) [8] to assess the ability of each model to price broad cross-sections of test assets. In the first step, the factor loadings for the test assets are estimated from a time series regression of the excess returns on the factors:

$$\mathbf{R}_t^e = a + BF_t + \varepsilon_t.$$

In the second step, the factor risk premia are obtained from a cross-sectional regression of the average excess asset returns, $\mu \in \mathbb{R}^N$, on the factor loadings estimated from the first stage:

$$\mu = z\iota + B\gamma + \alpha = C\lambda + \alpha, \quad C := [\iota \; B], \lambda' := [z \; \gamma'],$$

where ι denotes a conformable vector of ones, γ denotes the vector of factor risk premia, z is a scalar constant (that should be zero if the zero-beta rate matches the risk-free rate), and $\alpha \in \mathbb{R}^N$ is the vector of pricing errors (that should be zero if the factors price the test assets accurately).

Following the suggestions of Lewellen, Nagel, and Shanken (2010), we present several alternative measures of performance for the above cross-sectional regressions. First, we present the standard OLS cross-sectional adjusted R^2 (hereafter denoted by $\bar{R}^2_{OLS}$). This measure suffers from the shortcoming that if the returns have a strong factor structure (such as, e.g., the size and book-to-market equity sorted portfolio returns), then an arbitrarily chosen set of two or three factors, that have little correlation with the returns, are quite likely to produce large values of this statistic. This is obviously less of an issue for our I-SDF since it is a one-factor model, but it is likely to affect the performance of the other factor models that we consider for comparison.

Second, we present the GLS-adjusted R^2 (hereafter denoted by $\bar{R}^2_{GLS}$) that is obtained from the cross-sectional regression of $\hat{V}^{-1/2}\mu$ on $\hat{V}^{-1/2}[\iota \; B]$, where $V := Var(\mathbf{R}^e)$. The $\bar{R}^2_{GLS}$ for a model, unlike $\bar{R}^2_{OLS}$, is completely determined by the model-implied factors' proximity to the minimum variance frontier and, in general, presents a more stringent hurdle for models (Lewellen, Nagel, and Shanken 2010).

Third, we present the cross-sectional T^2 statistic of Shanken (1985), given by $T^2 := \hat{\alpha}' S_a^+ \hat{\alpha}$, where S_a^+ is the pseudoinverse of the estimated $\Sigma_a := \left(1 + \gamma' \Sigma_F^{-1} \gamma\right) \frac{y\Sigma y}{T}$, $y := I - C(C'C)^{-1}C'$ and $\Sigma := Var(\varepsilon_t)$. The T^2 statistic has an asymptotic χ^2 distribution with $N-K-1$ degrees of freedom, where K denotes the number of factors, and the non-centrality parameter $\alpha' \Sigma_a^+ \alpha = \alpha'(y\Sigma y)^+ \alpha \frac{T}{(1+\gamma'\Sigma_F^{-1}\gamma)}$, where Σ_F denotes the covariance matrix of the factors. We compute the p-value of this statistic under the null hypothesis that the model explains the vector of expected returns perfectly, that is, the vector of pricing errors $\alpha = 0$.

Fourth, we present the quadratic $q := \alpha'(y\Sigma y)^+ \alpha$, which measures how far a candidate model's factors are from the mean–variance frontier.[9] In particular, it is equal to the difference between the squared Sharpe ratio of the tangency portfolio of the test assets and the maximum squared Sharpe ratio attainable from the model-implied factors (or their mimicking portfolios in the case of non-traded factors).

8 We follow Cochrane (2005, Chapter 12) and apply the Shanken (1992) correction to the standard errors therein.

9 See Uppal and Zaffaroni (2015) for an alternative economic interpretation of this statistic.

Lastly, we present the simulated 90% confidence intervals for the statistics. The simulated confidence intervals are obtained using the approach suggested by Stock (1991) (see also Lewellen, Nagel, and Shanken 2010 for a detailed discussion). Consider first the construction of the confidence intervals for the $\bar{R}^2_{OLS}$. The simulations have two steps. First, we fix a true (population) cross-sectional R^2 that we want the model to have and alter the $(N \times 1)$ vector of expected returns, μ, to be $\mu = hC\lambda + e$, where $C \equiv [\iota, B]$, B denotes the vector of factor loadings in the historical sample, and $e \sim N(0, \sigma_e^2)$. The constants h and σ_e^2 are chosen to produce the right cross-sectional R^2 and maintain the historical cross-sectional dispersion of the average returns. Second, we jointly simulate an artificial time series of the factor and the returns of the same length as the historical data by sampling, with replacement, from the historical time series. We then use the two-pass regression method to estimate the sample cross-sectional R^2 of the simulated sample. We repeat the second step 1,000 times to construct a sampling distribution of the R^2 statistic conditional on the given population R^2. This procedure is repeated for all values of the population R^2 between 0 and 1. The 90% confidence interval for the true R^2 represents all values of the population R^2 for which the estimated R^2 in the historical sample falls within the 5th and 95th percentiles of the sample distribution.

A confidence interval for q is found using a method similar to that used to obtain the confidence interval for the true (population) cross-sectional R^2. Specifically, a given population R^2 implies a specific value of q. We obtain the sample distribution of the T^2 statistic as a function of q. The confidence interval for the true q represents all values of the q for which the estimated T^2 in the historical sample falls within the 5th and 95th percentiles of the sample distribution.

For the T^2 statistic, we present its finite-sample p-value, obtained from the above simulations, as the probability that the T^2 statistics in the simulated samples exceed the value of the statistics in the historical data for $q = 0$.

2 Data Description

In our baseline results, we assess the out-of-sample pricing performance of the extracted pricing kernel (the I-SDF) at the monthly frequency, over the period 1963:07–2017:06. The start date 1963:07 is chosen to coincide with that in Fama and French (1993), Lewellen, Nagel, and Shanken (2010), as well as DeMiguel, Garlappi, and Uppal (2009). This facilitates a useful comparison of our results with the existing literature. Given our 30-year training period, this requires data from 1933:07.

We first estimate the I-SDF from a small cross-section of fifteen equity portfolios, which includes industry portfolios and the top and bottom deciles of portfolios formed by univariate sorts of stocks based on size, book-to-market equity, momentum, and short- and long-term reversals. We also recover an I-SDF from a larger cross-section of 135 equity portfolios, which includes the five industry portfolios and the ten decile portfolios for univariate sorts on the basis of each of the following characteristics: size, book-to-market equity, momentum, short-term reversals, long-term reversals, operating profitability, investment, accruals, new issues, beta, variance, residual variance, and earnings-to-price. For the latter I-SDF, the out-of-sample evaluation period covers 1993:07–2017:06 because data on many of the portfolios are only available from the mid-sixties and our baseline procedure uses a 30-year training period.

Monthly returns data on the above portfolios are obtained from Kenneth French's data library. An estimate of the monthly risk-free rate is subtracted from the portfolio returns to produce the excess returns. Our proxy for the risk-free rate is the 1-month Treasury Bill rate, also obtained from Kenneth French's data library.

We also evaluate the performance of the I-SDF for other asset classes. In particular, we consider (i) the six currency portfolios from Lustig, Roussanov, and Verdelhan (2011), formed by sorting the currencies of developed and emerging economies on the basis of their forward discounts and rebalancing every month. Monthly returns on these currency portfolios are available from 1983:11 to 2017:06, (ii) portfolios of commodity futures from Asness, Moskowitz, and Pedersen (2013), the monthly returns on which are available over the period 1972:02–2017:06, and (iii) portfolios of global individual stocks from Asness, Moskowitz, and Pedersen (2013), available over the 1970:07–2017:06 period.

3 Cross-Sectional Pricing

In this section, we evaluate the out-of-sample ability of the I-SDF to explain the cross-section of returns for a variety of test assets representing the central anomalies in empirical asset pricing. Section 3.1 presents the results when the set of assets used to recover the I-SDF coincides with the set of test assets that the SDF is subsequently asked to price. Section 3.2, on the other hand, recovers a unique I-SDF from a small cross-section of fifteen portfolios and uses it to price several larger cross-sections of test assets. Finally, Section 3.3 uses an extension of the methodology to recover a unique I-SDF from a high dimensional cross-section, that includes portfolios formed based on all the major anomaly variables identified in the literature.

3.1 A Unique I-SDF for Each Cross-Section

Panel A of Table 1 presents the cross-sectional pricing results when the test assets consist of the 25 size and book-to-market-equity sorted portfolios of Fama and French (hereafter referred to as FF25). Row 1 shows that when the I-SDF is used as the sole factor, its estimated price of risk has the correct sign and is strongly statistically significant. Specifically, the (Shanken corrected) t-statistic has an absolute value in excess of 4. Harvey, Liu, and Zhu (2015) argue that a t-statistic of around 2.0 is too low a hurdle to establish the statistical significance of a given factor in the presence of extensive data mining. Using a new framework that allows for multiple tests, they show that a t-statistic greater than 3.0 would be required for a factor to be deemed as being statistically significant. Row 1 shows that the I-SDF has a t-statistic more than the value needed to establish statistical significance even after taking into account the possibility of data mining.

Since the regression uses the monthly excess returns as the dependent variable, the intercept can be interpreted as the estimated monthly zero beta rate over and above the risk-free rate. The estimated annualized zero beta rate is 3.1% (annualized) but is not statistically different from zero.[10] The I-SDF produces an $\bar{R}^2_{OLS}$ of 84.5% and, more importantly, $\bar{R}^2_{GLS}$ is very similar to $\bar{R}^2_{OLS}$, at 72.6%. Note that the GLS R^2 is high if and only if the factor is close to the mean–variance frontier and, in general, provides a more stringent hurdle for asset pricing models. The T^2 statistic shows that the model is not rejected at conventional significance levels. Lastly, the q statistic, which equals the difference between the squared Sharpe ratio of the tangency portfolio of the test assets and the squared Sharpe ratio of the factor-mimicking portfolio, is 0.045 and its 90% confidence interval includes 0, that is, the I-SDF mimicking portfolio is statistically indistinguishable from the maximum Sharpe ratio portfolio of the test assets.

In Row 2, we present the results for the unconditional CAPM. The market risk premium has the wrong sign and is not statistically different from zero. The intercept, on the other hand, is strongly significant with an annualized value of 13.9%—more than four times higher than the 3.1% value obtained with the I-SDF. The OLS and GLS $\bar{R}^2$ are much smaller at 3.95% and 30.3%, respectively, compared to those obtained with the I-SDF. The T^2 statistic is triple that obtained with the I-SDF, and has a p-value of zero: that is, the model is strongly rejected. The q statistic is closely related to the $\bar{R}^2_{GLS}$ and the T^2 statistics

[10] Note that part of it may be attributable to the differences between lending and borrowing rates (1–2%).

Table 1. I-SDF versus factor models, monthly 1963:07–2017:06

Row	$const.(\%)$	λ_{sdf}	λ_{Rm}	λ_{SMB}	λ_{HML}	λ_{RMW}	λ_{CMA}	λ_{IA}	λ_{ROE}	$\bar{R}^2_{OLS}$	$\bar{R}^2_{GLS}$	T^2	q
Panel A: 25 FF Portfolios													
$I-SDF_{(0\%<0)}$	0.26 (0.98)	−0.259 (−4.2)								84.5 [78.1,100]	72.6 [75.8,100]	24.4 (0.384)	0.045 [0.00,0.034]
CAPM	1.16 (3.09)		−0.004 (−0.97)							3.95 [−4.4,61.4]	30.3 [6.43,59.9]	74.5 (0.000)	0.116 [0.04,0.34]
FF3	1.26 (4.71)		−0.007 (−2.22)	0.002 (1.32)	0.004 (3.27)					66.3 [21.1,90.9]	40.2 [20.5,90.8]	55.3 (0.000)	0.089 [0.03,0.16]
FF5	1.02 (3.34)		−0.005 (−1.45)	0.002 (1.70)	0.003 (2.90)	0.005 (2.73)	−0.001 (−0.31)			74.2 [17.9,94.9]	37.1 [−7.0,86.8]	49.3 (0.000)	0.085 [0.01,0.21]
HXZ	0.98 (3.08)		−0.005 (−1.25)	0.003 (2.46)				0.004 (3.07)	0.004 (1.6)	67.2 [5.2,91.6]	36.9 [1.5,87.2]	53.4 (0.000)	0.096 [0.02,0.18]
$KNS-SDF$ $(0.46\%<0)$	−0.012 (−0.10)	−0.059 (−3.60)								59.8 [27.0,100]	34.4 [17.2,100]	68.0 (0.000)	0.112 [0.02,0.28]
$HJ-SDF$ $(3.7\%<0)$	0.32 (1.24)	−0.193 (−3.94)								78.5 [63.5,100]	70.5 [64.2,100]	27.9 (0.218)	0.048 [0.00,0.06]
Panel B: 10 Momentum-Sorted Portfolios													
$I-SDF$ $(0\%<0)$	0.36 (1.72)	−0.152 (−3.81)								94.1 [64.0,100]	81.6 [45.6,100]	7.66 (0.467)	0.013 [0.00,0.053]
CAPM	1.51 (4.33)		−0.010 (−2.45)							16.8 [−12.5,78.6]	6.41 [−7.9,42.4]	39.3 (0.000)	0.063 [0.03,0.43]
FF3	2.57 (3.09)		−0.019 (−2.11)	0.005 (0.84)	−0.008 (−1.16)					82.9 [−78.2,100]	31.7 [−25.5,92.2]	15.8 (0.015)	0.035 [0.00,0.34]
FF5	1.41 (0.99)		−0.013 (−0.87)	0.035 (1.75)	0.010 (0.61)	0.009 (0.99)	0.014 (0.88)			87.8 [−122.8,100]	54.0 [−26.2,100]	2.38 (0.667)	0.016 [0.00,0.24]
HXZ	1.73 (2.10)		−0.011 (−1.41)	0.007 (0.77)				−0.002 (−0.27)	0.004 (1.4)	83.3 [−72.8,100]	42.7 [−26.5,100]	11.5 (0.042)	0.023 [0.00,0.29]
$KNS-SDF$ $(0\%<0)$	−0.11 (−0.37)	−0.045 (−3.55)								72.6 [−11.4,100]	33.5 [−3.9,89.0]	27.8 (0.001)	0.046 [0.01,0.29]
$HJ-SDF$ $(1.5\%<0)$	0.28 (1.26)	−0.115 (−3.88)								96.1 [89.9,100]	86.6 [74.3,100]	5.61 (0.691)	0.009 [0.00,0.021]

(continued)

Table 1. (continued)

Row	$const.(\%)$	λ_{sdf}	λ_{Rm}	λ_{SMB}	λ_{HML}	λ_{RMW}	λ_{CMA}	λ_{IA}	λ_{ROE}	$\bar{R}^2_{OLS}$	$\bar{R}^2_{GLS}$	T^2	q
Panel C: Smallest and Largest Deciles of ME, B/M, Mom, STR, LTR, and 5 Industry Portfolios													
$I-SDF$ (0% < 0)	0.32 (1.39)	−0.368 (−5.4)								91.4 [90.3,100]	89.9 [95.8,100]	6.90 (0.907)	0.013 [0.00,0.00]
CAPM	0.99 (3.54)		−0.004 (−1.15)							−2.11 [−7.69,44.0]	2.59 [−1.39,52.6]	72.1 (0.000)	0.112 [0.03,0.51]
FF3	1.40 (4.87)		−0.008 (−2.49)	0.002 (1.52)	0.001 (1.05)					3.57 [−27.3,50.4]	−2.55 [−12.4,47.6]	61.5 (0.000)	0.100 [0.03,0.41]
FF5	0.17 (0.46)		0.004 (0.98)	0.003 (2.22)	−0.001 (−0.63)	0.001 (0.67)	0.005 (3.19)			11.0 [−55.6,53.3]	−11.3 [−14.4,56.3]	45.6 (0.000)	0.089 [0.02,0.40]
HXZ	0.32 (1.20)		0.002 (0.66)	0.005 (3.08)				0.002 (1.91)	0.005 (3.4)	52.4 [−40.0,86.0]	16.3 [−18.4,69.5]	38.9 (0.000)	0.072 [0.02,0.37]
$KNS-SDF$ (0% < 0)	0.43 (1.57)	−0.003 (−0.39)								−7.09 [−7.70,34.3]	10.47 [−1.0,50.7]	68.0 (0.000)	0.105 [0.03,0.46]
$HJ-SDF$ (3.6% < 0)	0.30 (1.33)	−0.228 (−5.71)								91.2 [90.3,100]	87.8 [85.6,100]	8.80 (0.788)	0.015 [0.00,0.009]

Cross-sectional regressions of average excess returns of different sets of test assets on the estimated factor loadings for different asset pricing models. Panel A presents the results when the test assets consist of the twenty-five size and BM sorted portfolios of FF. Panel B presents results when the test assets consist of the ten Momentum-sorted portfolios. Panel C presents results when the test assets consist of five industry-sorted portfolios and the top and bottom deciles of the ten size-sorted, ten BM-sorted, ten momentum-sorted, ten short-term reversal sorted, and ten long-term reversal sorted. In each panel, the first row presents the results when the factor is the information SDF, extracted from the corresponding set of test assets using a relative entropy minimizing procedure, in a rolling out-of-sample fashion starting in 1963:07. Rows 2–5 present the results for the CAPM, the Fama–French three-factor (FF3) and five-factor models (FF5), and the Hou, Xue, and Zhang four-factor model (HXZ), respectively. Row 6 reports the results when the factor is the SDF extracted with the KNS PCA approach. Finally, Row 7 presents the results for the HJ-SDF. For each model, the table presents the intercept and slopes, along with Shanken t-statistics in parentheses. It also presents the OLS and GLS adjusted R^2, along with their 90% confidence intervals for the true population R^2 (in square brackets). The confidence intervals are constructed via simulations using the approach suggested by Stock (1991) and used by Lewellen, Nagel, and Shanken (2010). The last two columns present, respectively, Shanken's (1985) cross-sectional T^2 statistic along with its asymptotic p-value in parentheses, and the q statistic that measures how far the factor-mimicking portfolios for the various models are from the mean–variance frontier.

and, therefore, not surprisingly, provides similar conclusions: the 90% confidence interval for the q statistic implies a large unexplained Sharpe ratio between 0.20 and 0.58, that is, the model fails to identify the maximum Sharpe ratio portfolio.

Row 3 presents the results for the FF3 model. The results show that the market risk premium is marginally statistically significant, albeit with the wrong sign. Only the risk premium associated with the factor proxying for risks related to book-to-market equity is significantly positive. However, the intercept is statistically and economically large, with an annualized value of 15.1%, similar to that obtained with the market risk factor alone in Row 2. The $\bar{R}^2_{OLS}$ is high at 66.3% (although substantially smaller than the 84.5% value obtained with the I-SDF), consistent with existing empirical evidence that the three FF factors explain a large fraction of the time series and cross-sectional variation in the returns of the twenty-five FF portfolios. However, moving to a GLS cross-sectional regression, $\bar{R}^2$ drops sharply to 40.2%, consistent with the observation that a GLS regression offers a more stringent hurdle for models than does the OLS. This is in stark contrast to the I-SDF, which delivers very similar $\bar{R}^2$ using both the OLS and GLS procedures. The T^2 statistic is larger than that obtained with the I-SDF (55.3 versus 24.4) and has a p-value of zero, implying a statistical rejection of the model. The q statistic is also double that obtained with the I-SDF (0.089 versus 0.045). Moreover, the 90% confidence interval of the q statistic does not include 0, that is, the maximum Sharpe ratio obtainable from the three FF factors is statistically smaller than the Sharpe ratio of the tangency portfolio of the test assets. Finally, the 90% confidence intervals for the true population $\bar{R}^2_{OLS}$ and $\bar{R}^2_{GLS}$ are tighter for the I-SDF than those obtained with the three FF factors. Specifically, the 95% confidence interval for the $\bar{R}^2_{OLS}$ includes the range [21.1–90.9] for the three FF factors, whereas it includes a narrower range of [78.1–100] for the I-SDF; the corresponding ranges for the $\bar{R}^2_{GLS}$ are [20.5–90.8] and [75.8–100] for the three FF factors and the I-SDF, respectively. Note that the confidence intervals for both the $\bar{R}^2_{OLS}$ and $\bar{R}^2_{GLS}$ include the highest possible value of 100 for the I-SDF but not for the three FF factors.

Row 4 shows that the FF5 model does not improve much upon the FF3 model. Specifically, the estimated intercept remains unchanged relative to the FF3 model at 12.2% annualized, the $\bar{R}^2_{OLS}$ increases marginally from 66.3% for the FF3 to 74.2% for the FF5 model, the $\bar{R}^2_{GLS}$ in fact, decreases marginally from 40.2% for the FF3 to 37.1% for the FF5 model, and the T^2 test rejects both models with a p-value of 0.00. Finally, the performance of the HXZ model, presented in Row 5, is also very similar to the FF3 and FF5 models. Interestingly, among the market risk factor for the CAPM, and the three, four, and five risk factors for the FF3, HXZ, and FF5 models, respectively, only the HML factor in FF3 and the IA factor in HXZ have t-statistics in excess of 3 once the first-stage estimation error in the factor betas are taken into account via the Shanken correction.

Row 5 presents results obtained using the Kozak, Nagel, and Santosh (2020, KNS henceforth) shrinkage-based method of constructing an SDF using the principal components of the base assets. The tuning parameters are set using a three-fold cross-validation that uses the full sample due to the relatively short time series available (this gives the method a potential advantage since it might induce a look-ahead bias). The KNS-SDF has overall a good performance, with large measures of fit and small and insignificant intercepts, and small share of negative realizations (0.46% of the cases). Nevertheless, at the point estimates, it fairs quite worse than the I-SDF in terms of cross-sectional fit, and the latter comes with a very large sampling uncertainty.

The last row of Panel A presents results for the Hansen–Jagannathan variance-minimizing SDF (hereafter referred to as HJ-SDF). This represents an alternative nonparametric approach to the recovery of the underlying SDF, albeit does not present a valid SDF in that it can take on negative values. Row 6 shows that the HJ-SDF performs slightly worse than the I-SDF – $\bar{R}^2_{OLS}$ of 84.5% versus 78.5%, respectively, and intercept of 3.8%

versus 3.1%. Importantly, unlike the I-SDF which is always positive by construction, the HJ-SDF takes on negative values about 3.7% of the time.

We next show that the strong performance of our model holds not only for the FF25 portfolios but also for portfolios formed by sorting stocks on the basis of other characteristics, such as prior returns, industry, etc. Panel B of Table 1 presents the cross-sectional regression results when the set of test assets consists of the 10 momentum sorted portfolios. The results are very similar to those obtained with the FF25 portfolios in Panel A.

Even stronger results are obtained in Panel C, where the set of test assets is much broader, consisting of the five industry-sorted portfolios and the top and bottom deciles of the ten size-sorted, ten BM-sorted, ten momentum-sorted, ten short-term reversal sorted, and ten long-term reversal sorted portfolios. Specifically, the I-SDF produces high $\bar{R}^2_{OLS}$ and $\bar{R}^2_{GLS}$ of 91.4% and 89.9%, respectively, and the T^2 test fails to reject the model at conventional significance levels. For the CAPM, on the other hand, the $\bar{R}^2_{OLS}$ and $\bar{R}^2_{GLS}$ are close to zero at −2.1% and 2.6%, respectively. For the FF3 and FF5 models, these statistics remain close to zero at 3.6% and 11.0%, respectively, for the $\bar{R}^2_{OLS}$ and −2.6% and −11.3%, respectively, for the $\bar{R}^2_{GLS}$. The HXZ model performs better than the FF3 and FF5 models with $\bar{R}^2_{OLS}$ and $\bar{R}^2_{GLS}$ of 52.4% and 16.3%, respectively. However, its performance is significantly worse than that of the I-SDF. Moreover, for each of these multi-factor models, the T^2 test strongly rejects the hypothesis that the model accurately prices the cross-section producing zero pricing errors. The KNS-SDF performs relatively well (but worse than the I-SDF) in Panel B (yielding a good fit and a small intercept), but does significantly worse in Panel C (where it achieves a negative $\bar{R}^2_{OLS}$ and has a large, albeit statistically insignificant, intercept). The HJ-SDF has comparable performance to the I-SDF in Panels B and C, but, once again, takes on negative values of 1.5% and 3.6%, respectively, of the time, thereby invalidating its use as a benchmark SDF.

Overall, Table 1 shows that: the I-SDF tends to produce smaller pricing errors and larger cross-sectional R^2s than the Fama–French three- and five-factor and the HXZ four-factor models, despite being only a one-factor model; the risk premium associated with the I-SDF is statistically significant; the T^2 statistic of the I-SDF implies that this factor is never rejected at standard confidence levels (while the other factor models considered are almost always rejected); the q statistic implies that the I-SDF successfully identifies the capital market line, that is, the mimicking portfolio is statistically undistinguishable from the maximum Sharpe ratio portfolio (while the other factor models are considered to fail in this respect); in all three cases, the (Shanken corrected) t-statistics of the information factor are larger than 3, hence clearing the higher hurdle for statistical significance recommended by Harvey and Liu (2015) and the 95% confidence intervals for the $\bar{R}^2_{OLS}$ and $\bar{R}^2_{GLS}$ are tighter for the I-SDF than those obtained with the other multi-factor models. Moreover, as an additional robustness check of the results in Table 1, we have also obtained cross-sectional estimates using the Pen-FM (Penalized Fama–MacBeth) estimator of Bryzgalova (2016), that by design has the ability to detect spurious factors and shrink (in a "lasso" fashion) their λ's to zero. Using this approach, we found virtually identical results for the information factor to those discussed above.[11]

Among the alternative methods for constructing a single factor SDF that we consider, only the HJ method has a similar performance to that of the I-SDF, but it often yields negative realization of the SDF.

Note that Table 1 focused on U.S. equities with long available histories of data. While U.S. equities are undoubtedly the most widely studied asset class among both academics and practitioners, other asset classes such as currencies and commodities have gained prominence in recent times. Moreover, international financial markets are playing an

[11] We are thankful to Svetlana Bryzgalova for providing us with the necessary computer code to implement this test.

Table 2. Performance of I-SDF on other asset classes

Row	Assets	$const.(\%)$	λ_{sdf}	$\bar{R}^2_{OLS}(\%)$	$\bar{R}^2_{GLS}(\%)$	T^2	q
Monthly							
(1)	Currencies	−0.14 (−1.26)	−0.243 (−4.55)	92.6	84.2	2.66 (0.61)	0.011
(2)	Commodities	0.28 (1.14)	−0.173 (−3.56)	89.1	96.9	0.47 (0.79)	0.001
(3)	Global Equities	−0.25 (−1.28)	−0.100 (−4.16)	98.4	98.9	0.128 (0.938)	0.000

Cross-sectional regressions of average excess returns are listed in column 2 on the estimated factor loadings for the information SDF, at the monthly frequency. For each set of returns, the table presents the intercept and slopes, along with Shanken t-statistics in parentheses. It also presents the OLS-adjusted R^2 and the GLS-adjusted R^2. The last two columns present, respectively, Shanken's (1985) cross-sectional T^2 statistic along with its asymptotic p-value in parentheses, and the q statistic.

ever-increasing role in the global landscape. In Table 2, we provide empirical evidence on the ability of the I-SDF to price currencies, commodities, and international equities out of sample. Because of the relatively short sample over which the returns data on these assets are available, we use a rolling window of 10 (rather than 30) years for the estimation of the I-SDF.

Row 1 presents the results for currency markets. The test assets are the six currency portfolios from Lustig, Roussanov, and Verdelhan (2011), formed by sorting the currencies of (at most) thirty-five developed and emerging economies on the basis of their forward discounts and rebalancing every month. Row 1 shows that the I-SDF produces an annualized intercept of −1.7%, which is statistically indistinguishable from zero and economically small. The risk premium associated with the I-SDF, on the other hand, is strongly statistically significant with a t-statistic exceeding 4 in absolute terms. Moreover, the risk premium has a similar magnitude and sign to those obtained for the U.S. equity market in Table 1. The OLS $\bar{R}^2$ looks impressive at 92.6% and, more importantly, is quite close to the GLS $\bar{R}^2$ of 84.2%. Consistent with the above statistics, the T^2 statistic has a p-value of 61%, suggesting that the sum of squared pricing errors is not statistically different from zero.

Row 2 presents the results for commodities. The test assets consist of four portfolios: the top and bottom portfolios formed by univariate sorting of twenty-seven different commodity futures into three portfolios on the basis of their book-to-market equity or "value" and momentum or past performance (details of the construction of these portfolios can be found in Asness, Moskowitz, and Pedersen 2013). The results are quite similar to those obtained with currency portfolios in Row 1—the estimated intercept is statistically insignificant and economically small with an annualized value of −3.4%; the risk premium associated with the I-SDF is statistically significant, with the absolute value of the t-statistic exceeding 3; the OLS and GLS $\bar{R}^2$ are both large at 89.1% and 96.9%, respectively; and the T^2 test fails to reject the model.

Row 3 presents the results for global equities. The test assets consist of four portfolios: the top and bottom portfolios formed by univariate sorting of individual stocks globally across four equity markets—the United States, the UK, Europe, and Japan—into three portfolios on the basis of their book-to-market equity and momentum (details of the construction of these portfolios can be found in Asness, Moskowitz, and Pedersen 2013). Row 3 shows that, in this case, the estimated intercept is statistically insignificant and economically small with an annualized value of 3.0%. The risk premium associated with the I-SDF, on the other hand, is strongly statistically significant with a magnitude similar to those obtained in the U.S. equity market in Table 1. Moreover, the OLS and GLS $\bar{R}^2$ are both large at 98.4% and 98.9%, respectively, and the T^2 test fails to reject the model.

The above results suggest that the I-SDF accurately identifies the underlying sources of priced risk, for broad cross-sections of assets. The latter finding is an important robustness check, since the methodology used relies on a large time series dimension (T) relative to the cross-sectional one (N). Hence, the stability of the results when the I-SDF is estimated from other asset classes with short histories is reassuring about the performance of the approach with shorter time series of returns data.

3.2 Unique I-SDF Recovered from a Small Cross-Section

In Section 3.1, the set of assets used to recover the out-of-sample I-SDF coincided with the set of test assets that the recovered I-SDF was subsequently challenged to price. In other words, a different I-SDF was recovered for each set of test assets. In this section, we assess whether a unique I-SDF, extracted from a small cross-section of assets, can price a variety of larger cross-sections of test assets.

Specifically, we consider the I-SDF obtained from the small, but broad, cross-section in Table 1, Panel C. The cross-section consists of the five industry portfolios and the top and bottom deciles of portfolios formed by sorting the universe of U.S. stocks based on several observable characteristics established in earlier literature as earning a risk premium—namely, size, the ratio of the book value of equity to its market value, momentum, short-term reversals, and long-term reversals. We use the I-SDF recovered from these fifteen portfolios to price larger cross-sections of test assets. We select this limited cross-section of anomaly portfolios (and industry ones) for two reasons. First, it captures well-documented risk factors that tend to be salient for more than just the unique cross-sections that they were designed to price. Second, the relative entropy approach's empirical performance, as any large T moment-based estimator, is more reliable with a relatively small cross-sectional dimension (N) relative to the time dimension (ideally, we would have $N/\sqrt{T}$ approaching zero).

The results are presented in Table 3. Consider first Panel A, where the test assets consist of the FF25 portfolios. Column 2 shows that the I-SDF produces a substantially smaller (and statistically insignificant) intercept than the other factor models—its intercept is very close to zero and at least two orders of magnitude smaller than those implied by CAPM, FF3, FF5, HXZ4, and KNS. The $\bar{R}^2_{OLS}$ and $\bar{R}^2_{GLS}$ for the I-SDF are comparable with those obtained with the FF3, FF5, and HXZ4 models. However, it must be borne in mind that these multi-factor models were specifically proposed to explain the returns on these sizes and book-to-market equity sorted portfolios. Another consideration documented by Lewellen, Nagel, and Shanken (2010) is that the additional degrees of freedom enjoyed by multifactor models over a single-factor model can artificially inflate the R^2 values, and we consider this effect in Table 4 below. The HJ-SDF produces results similar to those obtained with the I-SDF, albeit at the cost of producing negative realizations of the SDF in about 3.6% of the time periods.

The superior performance of the I-SDF vis-à-vis other multi-factor models is perhaps better demonstrated in Panel B, where the test assets consist of the ten size sorted, ten book-to-market-equity sorted, ten momentum sorted, ten short-term reversals sorted, ten long-term reversals sorted, and five industry-sorted portfolios. The I-SDF produces $\bar{R}^2_{OLS}$ of 74.7%, $\bar{R}^2_{GLS}$ of 43.1%, and a T^2 statistic that cannot be rejected at the 5% level of significance. The $\bar{R}^2_{OLS}$ for the FF3 and FF5 models are smaller at 22.9% and 31.9%, respectively, and the $\bar{R}^2_{GLS}$ close to zero at 9.0% and 8.1%, respectively. Moreover, for both models, the T^2 test rejects the null hypothesis of zero pricing errors at all significance levels. The HXZ model performs better than the FF3 and FF5 models for this set of test assets. Specifically, it produces an $\bar{R}^2_{OLS}$ of 56.5%, which is still substantially smaller than the

Table 3. Unique I-SDF Recovered From 15 portfolios, Monthly 1963:07–2017:06

Row	$const.(\%)$	λ_{sdf}	λ_{Rm}	λ_{SMB}	λ_{HML}	λ_{RMW}	λ_{CML}	λ_{IA}	λ_{ROE}	$\bar{R}^2_{OLS}(\%)$	$\bar{R}^2_{GLS}(\%)$	T^2	q
Panel A: Twenty-five FF Portfolios													
$I-SDF$ (0% < 0)	−0.00 (−0.013)	−0.741 (−3.20)								68.6	36.1	37.3 (0.030)	0.107
CAPM	1.16 (3.09)		−0.004 (−0.97)							3.95	30.3	74.5 (0.000)	0.116
FF3	1.26 (4.71)		−0.007 (−2.22)	0.002 (1.32)	0.004 (3.27)					66.3	40.2	55.3 (0.000)	0.089
FF5	1.02 (3.34)		−0.005 (−1.45)	0.002 (1.70)	0.003 (2.90)	0.005 (2.73)	−0.001 (−0.31)			74.2	37.1	49.3 (0.000)	0.085
HXZ	0.98 (3.08)		−0.005 (−1.25)	0.003 (2.46)				0.004 (3.07)	0.004 (1.59)	67.2	36.9	53.4 (0.000)	0.096
$KNS-SDF$ (0% < 0)	0.85 (2.50)	0.003 (0.29)								−3.56	22.3	82.5 (0.000)	0.127
$HJ-SDF$ (3.6% < 0)	−0.06 (−0.18)	−0.446 (−3.51)								67.5	34.3	46.6 (0.003)	0.110
Panel B: Fifty-five Test Assets													
$I-SDF$ (0% < 0)	0.35 (1.66)	−0.313 (−4.69)								74.7	43.1	69.5 (0.063)	0.124
CAPM	0.69 (2.64)		−0.001 (−0.26)							−1.51	5.88	132.5 (0.000)	0.205
FF3	1.27 (4.39)		−0.007 (−2.09)	0.002 (1.79)	0.001 (1.15)					22.9	8.96	118.4 (0.000)	0.190
FF5	0.48 (1.60)		0.001 (0.16)	0.003 (2.23)	0.001 (0.44)	0.001 (0.64)	0.004 (3.01)			31.9	8.10	108.0 (0.000)	0.185
HXZ	0.19 (0.75)		0.003 (1.03)	0.004 (2.77)				0.002 (2.25)	0.004 (2.71)	56.5	16.1	91.8 (0.000)	0.167
$KNS-SDF$ (0% < 0)	0.37 (1.46)	−0.006 (−0.75)								1.26	10.32	127.3 (0.000)	0.197
$HJ-SDF$ (3.6% < 0)	0.32 (1.53)	−0.198 (−4.94)								76.4	46.6	68.1 (0.080)	0.116

(continued)

Table 3. (continued)

Row	$const.(\%)$	λ_{sdf}	λ_{Rm}	λ_{SMB}	λ_{HML}	λ_{RMW}	λ_{CML}	λ_{IA}	λ_{ROE}	$\bar{R}^2_{OLS}(\%)$	$\bar{R}^2_{GLS}(\%)$	T^2	q
Panel C: Twenty-five FF + 30 Industry + 10 Momentum													
$I-SDF$ (0%<0)	0.38 (1.70)	−0.313 (−4.34)								50.3	32.3	143.1 (0.000)	0.255
CAPM	0.86 (3.76)		−0.002 (−0.72)							1.50	19.6	194.4 (0.000)	0.301
FF3	1.25 (5.55)		−0.007 (−2.38)	0.002 (1.53)	0.002 (1.44)					24.9	23.4	173.2 (0.000)	0.277
FF5	0.97 (3.44)		−0.004 (−1.32)	0.002 (1.91)	0.001 (1.02)	0.002 (1.22)	0.002 (0.84)			32.2	21.8	170.4 (0.000)	0.274
HXZ	0.79 (3.36)		−0.002 (−0.83)	0.003 (2.31)				0.002 (1.63)	0.004 (2.51)	45.5	23.6	168.5 (0.000)	0.293
$KNS-SDF$ (0%<0)	0.64 (2.77)	0.00 (−0.03)								−1.6	18.5	199.4 (0.000)	0.308
$HJ-SDF$ (3.6%<0)	0.34 (1.57)	−0.196 (−4.46)								49.0	28.9	157.5 (0.000)	0.268

Cross-sectional regressions of average excess returns of different sets of test assets on the estimated factor loadings, for different asset pricing models. Panel A presents the results when the test assets consist of the twenty-five size and BM sorted portfolios of FF. Panel B presents results when the test assets consist of ten size-sorted, ten BM-sorted, ten momentum-sorted, ten short-term reversal sorted, ten long-term reversal sorted, and five industry-sorted portfolios. Panel C presents results when the test assets consist of the twenty-five FF, thirty Industry, and ten Momentum-sorted portfolios. In each panel, the first row presents the results when the factor is the information SDF. This SDF is constructed from fifteen portfolios—consisting of five industry portfolios and the smallest and largest deciles of size, BM, momentum, short-term reversal, and long-term reversal sorted portfolios—using a relative entropy minimizing procedure, in a rolling out-of-sample fashion starting in 1963:07. Rows 2–5 present the results for the CAPM, the FF3, the FF5, and the HXZ models, respectively. Row 6 reports the results when the factor is the SDF extracted with the KNS PCA approach. Finally, Row 7 presents the results for the HJ-SDF. For each model, the table presents the intercept and slopes, along with Shanken t-statistics in parentheses. It also presents the OLS-adjusted R^2 and the GLS adjusted R^2. The last two columns present, respectively, Shanken's (1985) cross-sectional T^2 statistic along with its asymptotic p-value in parentheses, and the q statistic that measures how far the factor-mimicking portfolios are from the mean–variance frontier of the test assets.

Table 4. Performance of model-implied SDFs

Row	Assets	*const.*(%)	λ_{sdf}	$\bar{R}^2_{OLS}(\%)$	$\bar{R}^2_{GLS}(\%)$	T^2	q	$Euler\ Err_{[RMSE]}$
Panel A: I-SDF (0% < 0), 1963:07–2017:06								
(1)	FF 25	−0.00 (−0.042)	−0.741 (−7.31)	68.6	36.1	37.3 (0.030)	0.107	0.001 [0.002]
(2)	*55 Decile Portfolios* (*ME,BE/ME, Mom, STR, LTR, 5Ind.*)	0.35 (1.65)	−0.313 (−4.69)	74.7	43.1	69.5 [0.063]	0.124	0.001 [0.002]
(3)	25 FF + 30 Ind + 10 Mom	0.38 (1.70)	−0.313 (−4.33)	50.3	32.3	143.1 (0.000)	0.255	0.002 [0.002]
Panel B: FF3-SDF (0.15% < 0), 1963:07–2017:06								
(1)	FF 25	0.22 (0.53)	−0.038 (−1.38)	0.67	24.5	78.0 (0.000)	0.124	0.002 [0.002]
(2)	*55 Decile Portfolios* (*ME, BE/ME, Mom, STR, LTR, 5Ind.*)	0.99 (3.90)	0.030 (1.28)	3.92	5.79	130.3 [0.000]	0.205	0.001 [0.002]
(3)	25 FF + 30 Ind + 10 Mom	0.86 (4.60)	0.016 (0.86)	0.73	18.2	197.7 (0.000)	0.307	0.002 [0.003]
Panel C: FF5-SDF (11.5% < 0), 1993:07–2017:06								
(1)	FF 25	0.81 (2.62)	0.020 (0.017)	−4.34	38.7	58.1 (0.000)	0.202	0.025 [0.030]
(2)	*55 Decile Portfolios* (*ME, BE/ME, Mom, STR, LTR, 5Ind.*)	0.81 (3.05)	1.11 (1.10)	22.6	22.7	64.7 (0.131)	0.233	0.018 [0.024]
(3)	25 FF + 30 Ind + 10 Mom	0.80 (2.88)	0.672 (0.732)	7.01	26.9	147.8 (0.000)	0.520	0.028 [0.033]
Panel D: HXZ-SDF (5.56% < 0), 1993:07–2017:06								
(1)	FF 25	0.82 (2.90)	−0.044 (−0.66)	10.7	36.9	59.9 (0.000)	0.209	0.010 [0.012]
(2)	*55 Decile Portfolios* (*ME, BE/ME, Mom, STR, LTR, 5Ind.*)	0.74 (3.00)	−0.030 (−0.469)	3.73	24.0	65.7 (0.113)	0.229	0.010 [0.011]
(3)	25 FF + 30 Ind + 10 Mom	0.75 (2.82)	−0.025 (−0.42)	2.21	23.9	156.7 (0.000)	0.545	0.008 [0.011]

Cross-sectional regressions of average excess returns are listed in column 2 on the estimated factor loadings for different model-implied SDFs. For each model, the model-implied SDF is recovered in a rolling out-of-sample fashion using only past returns data on a cross-section of fifteen equity portfolios: five industry portfolios and the smallest and largest deciles of size, BM, momentum, short-term reversal, and long-term reversal sorted portfolios. Panels A–D report results for the I-SDF, FF3, FF5, and HXZ models, respectively. For each set of test assets, the table presents the intercept and slopes, along with Shanken t-statistics in parentheses, the OLS and GLS adjusted R^2 Shanken's (1985) cross-sectional T^2 statistic along with its asymptotic p-value in parentheses, and the q statistic that measures how far the factor-mimicking portfolios are from the mean–variance frontier. The last column reports the average absolute Euler Equation error for each cross-section of test assets, along with the cross-sectional root mean squared error in parentheses.

74.7% implied by the I-SDF. It also implies a substantially smaller $\bar{R}^2_{GLS}$ than that produced by the I-SDF (16.1% versus 43.1%). This is also reflected in the higher value of the closely related q-statistic for the HXZ model compared to the I-SDF. Also, the T^2 test rejects the HXZ model at all significance levels. The KNS-SDF does not perform well in this cross-section of test assets, yielding a statistically insignificant λ coefficient and low measures of fit. Once again, only the HJ-SDF has similar performance to that of the I-SDF, yet once again at the cost of negative realizations.

Finally, Panel C presents the pricing performance of the various models when the set of test assets includes the FF25 portfolios, the thirty industry-sorted portfolios, and the ten momentum-sorted portfolios. As with the test assets in Panel B, the I-SDF exhibits the best performance in terms of the smallest estimated intercept, the highest $\bar{R}^2_{OLS}$ and $\bar{R}^2_{GLS}$, the lowest T^2 statistic, and the lowest q-statistic amongst the factor models.

Overall, Table 3 reinforces the conclusion from Table 1, namely that the I-SDF tends to produce smaller pricing errors and larger cross-sectional R^2s than the Fama–French three- and five-factor and the HXZ four-factor models, despite being only a one-factor model. As shown in Lewellen, Nagel, and Shanken (2010), it is relatively easy to find two or three factors that produce large $\bar{R}^2_{OLS}$ for test assets such as the twenty-five FF portfolios because of their strong factor structure. It is noteworthy that a single factor, namely the I-SDF, does even better than the FF3, FF5, and HXZ factors. Of course, any multi-factor model can be rewritten as a single-factor model (see e.g., Back 2010), nevertheless this requires knowledge of the projection coefficients that are available only *ex post* to the econometrician. Hence, *ex ante*, the number of factors is a relevant metric for assessing the degrees of freedom that a model has for fitting the data.

To assess the extent of this bias in favor of the multifactor models (FF3, FF5, and HXZ), we also test one-factor versions of these models, where the single factor is created from a linear combination of the multiple factors in the rolling training sample. These tests restrict the degrees of freedom of the multifactor models to that of a single factor, allowing better comparison with the I-SDF than when the coefficients are left unrestricted in the cross-sectional regressions. Specifically, we present the empirical performance of the multi-factor models when a *multi-factor-model-implied SDF* is constructed as a linear function of the risk factors, with the coefficients estimated in a rolling out-of-sample fashion using only past returns data on the same cross-section of portfolios used to recover the I-SDF. For instance, we define the FF3 model-implied SDF as:

$$M_t^{FF3} = \gamma_0 + \sum_{j=1}^{3} \gamma_j f_{j,t}, \tag{6}$$

where the coefficients γ_j, $j = 0, 1, 2, 3$, are estimated in a rolling out-of-sample fashion using only past returns data on the cross-section of fifteen equity portfolios, so as to satisfy the Euler equation restrictions for these portfolios:

$$0 = E\left[(R_{i,t} - R_{f,t})\left(\gamma_0 + \sum_{j=1}^{3} \gamma_j f_{j,t}\right)\right], \quad i = 1, 2, \ldots, 15. \tag{7}$$

The resulting M_t^{FF3} is then used as the single risk factor in the cross-sectional regressions for different sets of test assets to assess its empirical performance. Similarly, the FF5 model-implied SDF is defined as $M_t^{FF5} = \gamma_0 + \sum_{j=1}^{5} \gamma_j f_{j,t}$, and so on for the HXZ model-implied SDF.

The results are presented in Table 4. For the ease of comparison, Panel A reiterates the pricing performance of the I-SDF for the three different cross-sections (Table 3, Row 1 of each panel). Consider first Panel B, which presents the results for the FF3 model-implied SDF, M_t^{FF3}. The table shows that the performance of the one-factor M_t^{FF3} is substantially worse than when the FF3 factors are treated as three separate risk factors with unrestricted coefficients in the cross-sectional regressions (Table 3, Panels A–C, Row 3). Specifically, regardless of the cross-section of test assets used, the magnitudes of the $\bar{R}^2_{OLS}$ are substantially smaller than those obtained in Table 3 where the coefficients of the three factors are left unrestricted in the cross-sectional regressions—when the test assets are the FF25 portfolios, the $\bar{R}^2_{OLS}$ reduces from 66.3% in Table 3 to 0.67% in Table 4; and for the FF25, thirty Industry, and ten momentum-sorted portfolios, it reduces from 24.9% to 0.73%.

We next turn to the FF5 model-implied SDF. Since data on the profitability and investment factors are only available from 1963 onwards and we use a 30-year rolling sample to recover the model-implied SDFs, our out-of-sample evaluation period for this model extends over 1993:07–2017:06. The results are presented in Panel C of Table 4 and are similar to those obtained in Panel B for the FF3 model-implied SDF. Similar conclusions are also obtained in Panel D for the HXZ model-implied SDF over 1993:07–2017:06. Note that in Table 4, we do not report results for the HJ- and KNS-SDF since these would be identical to those reported in Table 3.

Finally, note that, we have, thus far, focused on tests based on the expected return–beta representation of models. However, one can also perform SDF-based tests, that is, tests based on the Euler restrictions $E[M\mathbf{R}^{\mathbf{e}}] = 0$. This is because our information-theoretic methodology delivers a direct estimate of the pricing kernel M. Therefore, for each cross-section of test assets in Table 4, we can compute the average absolute Euler Equation error, $EER \equiv \frac{1}{N}\sum_{i=1}^{N} |\frac{1}{T}\sum_{t=1}^{T} M_t R^e_{i,t}|$, along with the cross-sectional root mean squared error, $RMSE \equiv \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} M_t R^e_{i,t}\right)^2}$. These statistics can be similarly computed using the model-implied SDFs for the FF3, FF5, and HXZ multi-factor models. The results, presented in the last column of Table 4, show, once again, that the I-SDF compares favorably with respect to the multi-factor models, particularly the FF5 and HXZ models.

Note that the performance of the FF3, FF5, and HXZ multi-factor models might be surprising relative to the results in the literature. The reason for this discrepancy is that, unlike most previous studies, we restrict them in the out-of-sample to be a single SDF, hence allowing them only one degree of freedom—like for the I-SDF—to price asset returns.

Overall, the results suggest that the one-factor I-SDF has an impressive pricing performance compared to popular multi-factor models widely used in the literature.

3.3 Unique I-SDF Recovered from a Large Cross-Section

The unique I-SDF in Section 3.2 was recovered from a small cross-section of fifteen equity portfolios. This includes portfolios representing many, but not all, of the anomalies identified in the literature. The restricted cross-section is required because the methodology, being an information-theoretic alternative to the widely used generalized method of moments approach, relies on the number of moment restrictions being small relative to the length of the available time series. In this section, we apply an extension of the approach, developed by Qiu and Otsu (2022), to recover the I-SDF from a large cross-section of assets that includes all the major anomaly variables identified in the literature.

Qiu and Otsu (2022) show that, in a high-dimensional scenario, the SDF can be recovered as:

$$\hat{M}_t \equiv M_t\left(\hat{\theta}_T, \mathbf{R}_t^e\right) = \frac{Te^{\hat{\theta}'_T \mathbf{R}_t^e}}{\sum_{t=1}^{T} e^{\hat{\theta}'_T \mathbf{R}_t^e}}, \quad \forall t, \tag{8}$$

where

$$\hat{\theta}_T := \underset{\theta}{\text{argmin}} \frac{1}{T}\sum_{t=1}^{T} e^{\theta' \mathbf{R}_t^e} + \gamma_N ||\theta||_1, \tag{9}$$

where $||\theta||_1$ is the L1 norm of the vector of Lagrange multipliers θ and γ_N is a penalty level chosen by the researcher. A comparison of Equations (8) and (9) for the high-dimensional case with Equations (4) and (5) for a low-dimensional scenario shows that the only difference between them is the introduction of the penalty term in the former case. This penalty term serves to shrink towards zero the weights of redundant assets in the resultant I-SDF. This approach can be applied even when the size of the cross-section is larger than the length of the available time series, thereby allowing one to form a unique I-SDF from a much richer sample of assets.

We apply this extension of the methodology to a cross-section of $N = 135$ portfolios. These include the five industry portfolios and the ten decile portfolios for univariate sorts on the basis of each of the following characteristics: size, book-to-market equity, momentum, short-term reversals, long-term reversals, operating profitability, investment, accruals, new issues, beta, variance, residual variance, and earnings-to-price. As with our earlier analysis, we use a training sample of 30 years. Since data on many of these portfolios are only available from the mid-sixties, we present out-of-sample pricing performance over the 25 year period 1993:07–2017:06.

For selecting the penalty parameter γ_N, we consider two different approaches. First, we follow the approach of Qiu and Otsu (2022). Specifically, we create a grid for γ_N from 0.1 to 2 with 0.05 increments, estimate the I-SDF using our method, and implement the cross-sectional regression for each penalty level. In the absence of penalization, the I-SDF performs poorly with the R^2_{OLS} and R^2_{GLS} close to zero. As the penalty level increases, the performance of the I-SDF improves, with the performance peaking when the penalty level is around $0.5 - 0.6$. Further increases in the penalty level lead to worsening of the model performance since the number of portfolios selected becomes too small for very large penalty levels. Based on this, we set the penalty level at 0.6. Note that the above approach to selecting the penalty parameter has a look-ahead bias. Therefore, in our second approach, the penalty parameter is selected in a rolling fashion and then used to construct the I-SDF out-of-sample. Specifically, in the 30-year training sample, the first 25 years are used to estimate the Lagrange multipliers. The penalty parameter is then selected using the subsequent five years of data. Using the estimates of the Lagrange multipliers and the selected value of the penalty parameter based on the 30-year window, the I-SDF is then constructed out-of-sample for the following 12 months. This process is then repeated using a rolling 30-year training sample to obtain the time series of the I-SDF over the full out-of-sample evaluation period. This removes any look-ahead bias from the procedure.

Our results are presented in Table 5, Panel A. Row 1 shows that, when the penalty parameter is selected using the full out-of-sample period, the I-SDF has a price of risk that has the same negative sign as in all the previous tables, albeit not statistically significant in this case, and an $\bar{R}^2_{OLS}$ of 22.0%. Row 2 presents the results when the penalty parameter is chosen in a rolling manner thereby avoiding any look-ahead bias. A comparison of Rows 1 and 2 shows that the results without any look-ahead bias are actually quite similar to the ones obtained with a penalty parameter selected over the full sample. Row 3 shows that,

Table 5. Unique I-SDF recovered from large cross-section

Row	$const.(\%)$	λ_{sdf}	λ_{Rm}	λ_{SMB}	λ_{HML}	λ_{RMW}	λ_{CML}	λ_{IA}	λ_{ROE}	$\bar{\mathrm{R}}^2_{OLS}(\%)$	$\bar{\mathrm{R}}^2_{GLS}(\%)$	T^2	q
Panel A: 135 Portfolios, Rolling Window = 30 years													
$I-SDF$ (full sample penalty) (0% < 0)	0.51 (1.38)	−0.020 (−1.29)								22.0	10.6	303.5 (0.000)	1.09
$I-SDF$ (OOS penalty) (0% < 0)	0.77 (3.42)	−0.065 (−0.71)								14.9	13.8	301.0 (0.000)	1.06
CAPM	0.85 (3.11)		−0.001 (−0.367)							3.85	10.0	316.9 (0.000)	1.10
FF3	0.91 (3.87)		−0.002 (−0.630)	0.001 (0.529)	0.001 (0.570)					9.52	9.02	313.8 (0.000)	1.10
$FF3-SDF$ (0.34% < 0)	0.72 (2.41)	0.001 (0.047)								−0.007	10.9	316.4 (0.000)	1.10
$FF3-SDF$ (Mkt, S, B, G, V:(0.34% < 0))	0.88 (3.21)	0.031 (0.885)								4.24	11.4	311.3 (0.000)	1.09
FF5	0.27 (1.12)		0.004 (1.12)	0.003 (1.23)	−0.000 (−0.040)	0.003 (1.60)	0.002 (1.65)			32.2	10.4	274.5 (0.000)	1.06
$FF5-SDF$ (10.1% < 0)	0.71 (2.69)	−0.042 (−0.160)								−0.08	10.3	316.2 (0.000)	1.10
$FF5-SDF$ (Mkt, S, B, G, V, R, W, C, A:0.69% < 0)	0.69 (2.19)	−0.014 (−0.269)								1.19	9.96	317.0 (0.000)	1.10
HXZ	0.45 (2.13)		0.002 (0.66)	0.002 (1.18)				0.001 (0.59)	0.003 (1.36)	23.9	8.90	301.6 (0.000)	1.09
$HXZ-SDF$ (2.08% < 0)	0.70 (2.56)	−0.029 (−0.56)								8.66	10.3	315.1 (0.000)	1.10
$KNS-SDF$ (0% < 0)	0.84 (3.1)	0.002 (0.34)								3.09	10.22	316.4 (0.000)	1.10
$HJ-SDF$ (23.6% < 0)	0.78 (44.8)	−0.338 (−4.98)								15.1	14.3	297.5 (0.000)	1.05

(continued)

Table 5. (continued)

Row	$const.(\%)$	λ_{sdf}	λ_{Rm}	λ_{SMB}	λ_{HML}	λ_{RMW}	λ_{CML}	λ_{IA}	λ_{ROE}	$\bar{R}^2_{OLS}(\%)$	$\bar{R}^2_{GLS}(\%)$	T^2	q
Panel B: 135 Portfolios, Rolling Window = 20 years													
$I-SDF$ (full sample penalty) (0% < 0)	0.53 (1.91)	−0.335 (−2.21)								60.9	22.3	257.1 (0.000)	0.720
$I-SDF$ (OOS penalty) (0% < 0)	0.66 (2.76)	−0.273 (−1.94)								49.8	16.3	283.9 (0.000)	0.775
CAPM	1.10 (4.11)		−0.004 (−1.20)							22.2	12.0	329.2 (0.000)	0.814
FF3	1.00 (4.42)		−0.003 (−1.04)	−0.001 (−0.474)	0.002 (1.22)					24.2	11.0	328.0 (0.000)	0.810
$FF3-SDF$ (0.25% < 0)	0.59 (2.18)	−0.011 (−0.505)								0.13	12.2	331.4 (0.000)	0.813
$FF3-SDF$ (Mkt, S, B, G, V:1.23% < 0)	0.88 (4.40)	0.044 (1.18)								4.51	12.9	327.8 (0.000)	0.811
FF5	0.46 (2.15)		0.002 (0.551)	0.001 (0.419)	0.001 (0.540)	0.004 (2.62)	0.002 (1.72)			43.4	13.4	297.2 (0.000)	0.776
$FF5-SDF$ (1.23% < 0)	0.56 (1.97)	−0.123 (−1.27)								19.3	12.4	322.9 (0.000)	0.811
$FF5-SDF$ (Mkt, S, B, G, V, R, W, C, A:0.74% < 0)	0.54 (1.80)	−0.065 (−1.18)								14.5	12.0	327.2 (0.000)	0.814
HXZ	0.38 (1.90)		0.002 (0.838)	0.001 (0.810)				0.002 (1.32)	0.004 (2.52)	44.7	11.1	309.2 (0.000)	0.803
$HXZ-SDF$ (1.23% < 0)	0.53 (1.80)	−0.074 (−1.55)								32.5	12.4	321.5 (0.000)	0.811
$KNS-SDF$ (0% < 0)	0.846 (3.1)	0.003 (0.35)								3.46	10.19	316.4 (0.000)	1.10
$HJ-SDF$ (31.6% < 0)	0.75 (3.21)	−2.38 (−2.03)								57.5	30.0	200.1 (0.000)	0.648

Cross-sectional regressions of average excess returns of a set of 135 test assets on the estimated factor loadings for different asset pricing models. The test assets consist of 135 equity portfolios, that include the five industry portfolios and the ten decile portfolios for univariate sorts on the basis of each of the following characteristics: size, book-to-market equity, momentum, short-term reversals, long-term reversals, operating profitability, investment, accruals, new issues, beta, variance, residual variance, and earnings-to-price. Panels A and B report results for training samples of length 30 and 20 years, respectively. The first row in each panel presents the results when the factor is the information SDF, that is constructed from the above 135 portfolios using a relative entropy minimizing procedure, in a rolling out-of-sample fashion starting in 1993:07. Rows 2, 3, 6, and 9 present the results for the CAPM, the FF3, the FF5, and the HXZ models, respectively. Rows 4 and 5 report results for the FF3 model-implied SDFs, Rows 7 and 8 for the FF5 model-implied SDFs, and Row 10 for the HXZ model-implied SDF. Row 11 reports the results when the factor is the SDF extracted with the KNS PCA approach. Finally, Row 12 presents the results for the HJ-SDF. For each model, the table presents the intercept and slopes, along with Shanken t-statistics in parentheses, the OLS and GLS adjusted R^2 Shanken's (1985) cross-sectional T^2 statistic along with its asymptotic p-value in parentheses, and the q statistic that measures how far the factor-mimicking portfolios are from the mean–variance frontier of the test assets.

perhaps not surprisingly, the CAPM fares quite poorly at explaining this large cross-section, producing an annualized intercept of 10.2%, a wrong sign for the price of market risk, and an $\bar{R}^2_{OLS}$ of only 3.9%. Row 4 shows that the FF3 model improves upon the CAPM but still fares worse than the I-SDF, both in terms of the annualized intercept and $\bar{R}^2_{OLS}$. This is particularly striking given that the I-SDF is a one-factor model compared to the three-factor FF3 model. To this effect, Row 5 shows the performance of the *one-factor-FF3-model-implied SDF*, denoted M^{FF3}, constructed as a linear function of the three risk factors, with the coefficients estimated in a rolling out-of-sample fashion using past returns data on the same cross-section of 135 portfolios used to recover the I-SDF. The table shows that the one-factor M^{FF3} performs substantially worse than the three-factor FF3, producing a price of risk that is not statistically significant and an $\bar{R}^2_{OLS}$ of -0.7% (compared to 9.5% in Row 3 for the FF3 model). Row 6 shows that the one-factor SDF implied by a disaggregated five-factor variant of the FF3 model—with the factors being the return on the market, and returns on portfolios of small market capitalization, big market capitalization, growth, and value stocks—fares no better than the one-factor-FF3-model-implied SDF in Row 5.

Row 7 shows that the unrestricted FF5 model performs better than the I-SDF—it produces a smaller intercept and larger $\bar{R}^2_{OLS}$. However, Row 8 shows that this is largely driven by the extra flexibility of a five-factor model relative to that of a one-factor model in matching the data. Specifically, the one-factor M^{FF5}, constructed as a linear combination of the five FF factors in a similar manner as the M^{FF3}, performs substantially worse than the I-SDF—its estimated price of risk is not statistically significant and $\bar{R}^2_{OLS}$ is negative. Row 9 shows that the one-factor SDF implied by a disaggregated nine-factor variant of the FF5 model—with the factors being the return on the market, and returns on portfolios of small market capitalization, big market capitalization, growth, value, robust profitability, weak profitability, conservative investment, and robust investment stocks—fares no better than the one-factor-FF5-model-implied SDF in Row 8. Similar conclusions were obtained for the unrestricted four-factor HXZ model in Row 10 and its restricted one-factor specification in Row 11. Surprisingly, the KNS-SDF in Row 12, despite being designed for handling large cross-sections of base and test assets, does not perform well in this setting.

Finally, Row 13 presents results for the HJ-SDF. The HJ-SDF performs slightly worse than the I-SDF in Row 1 – $\bar{R}^2_{OLS}$ of 15.1% versus 22.0%, respectively, and (annualized) intercept of 9.4% versus 6.1%. More importantly, as discussed earlier, the HJ-SDF is not a valid SDF in that it produces negative realizations of the SDF. Not surprisingly, this issue becomes more severe when the size of the cross-section used to extract it increases. In this case, where 135 portfolios are used, the HJ-SDF is negative 23.6% of the time. The I-SDF, on the other hand, always produces a valid SDF.

Panel B presents results for a shorter training window of 20 years. The pricing performance of all the models improve in terms of their $\bar{R}^2_{OLS}$ relative to Panel A, where the length of the training window corresponded to the baseline value of 30 years. However, the conclusions regarding the relative performance of the models remain unaltered with respect to those obtained in Panel A. Specifically, the one-factor I-SDF performs substantially better than the one-factor SDFs implied by all the multi-factor models. The HJ-SDF, which has a performance comparable to that of the I-SDF, leads to negative realizations 31.6% of the time in this case.

4 Properties of the I-SDF

Having shown that the I-SDF is successful at pricing broad cross-sections of financial assets out-of-sample, we next turn to an investigation of the properties of the I-SDF.

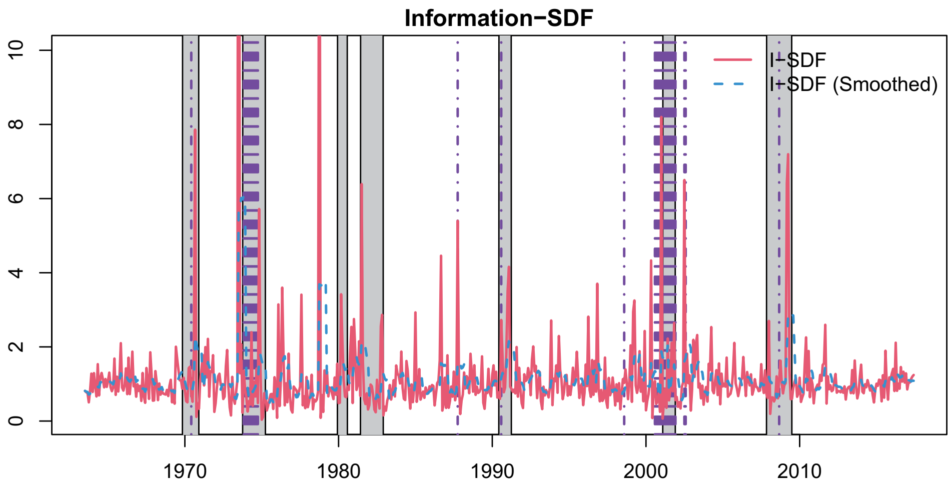

Figure 1. Time series of the I-SDF (red line) and the 1-year moving average of the I-SDF (blue-dashed line). The I-SDF is constructed from fifteen portfolios—consisting of five industry portfolios and the smallest and largest deciles of size, BM, momentum, short-term reversal, and long-term reversal sorted portfolios—using a relative entropy minimizing procedure, in a rolling out-of-sample fashion starting in 1963:07. NBER recession periods are marked by the gray vertical bands. The vertical dot-dashed lines indicate market crashes identified using the Mishkin and White (2002) approach.

4.1 The I-SDF

Figure 1 plots the time series of the unique I-SDF (red solid line) from Section 3.2. To isolate visually the low frequency behavior of the I-SDF, the figure also reports the 1-year moving average of the I-SDF (blue dashed line). The I-SDF shows a clear business cycle pattern: it is low during expansions and higher during NBER recession periods (depicted by the grey vertical bands).

However, Figure 1 also reveals another interesting characteristic of the I-SDF over and above the co-movement related to the business cycle. The I-SDF is especially high when the recession episodes are concomitant with periods of big stock market downturns, like the 2001, and 2007–2009 recessions, unlike the 1981–1982 and 1990–1991 recessionary periods that were not accompanied by big stock market crash episodes. The correlation between the smoothed I-SDF and a recession dummy is 12.4%, while that of the I-SDF and a market crash dummy is 9.2%.

Note also that using an alternative relative entropy minimization approach for the construction of the I-SDF (in Appendix A1), we obtain a pricing kernel with very similar business cycle properties. This stresses that business cycle risk seems to be priced in the cross-section of asset returns. Hence, minimum entropy constructions of the SDF, and more generally Cressie–Read discrepancy-based methods (see, e.g., Almeida and Garcia 2012),[12] of which relative entropy is a particular case, are likely to recover this feature of the data.

Our results suggest that business cycle risk is an important source of priced risk. Note that this result obtains even though we do not use any macroeconomic variables, but only

[12] The Cressie–Read discrepancies have one free parameter that controls the sensitivity of the SDF to the higher moments of the base assets.

Table 6. Characterizing the I-SDF

Panel A: Summary statistics

Row	Median	Volatility	Skewness	Kurtosis	Max	Min
(1)	0.85	0.80	4.52	35.5	8.91	0.032

Panel B: Quantifying Tail Risk

Row	Assets	$M_{0.95}$	$M_{0.90}$	$R^e_{0.95}$	$R^e_{0.90}$	$M_{0.95}$ and $R^e_{0.95}$	$M_{0.90}$ and $R^e_{0.90}$
(1)	Market	33.0 [5.09]	45.7 [10.0]	50.4 [4.94]	57.6 [9.88]	42.7 [1.23]	61.5 [3.40]
(2)	Cross-Section	39.7 [5.09]	48.4 [10.0]	42.7 [4.94]	50.0 [9.88]	37.4 [0.93]	60.7 [3.09]

Panel A reports the summary statistics from the distribution of the I-SDF. The I-SDF is constructed using the cross-section of ffiteen portfolios, consisting of five industry portfolios and the smallest and largest deciles of size, BM, momentum, short-term reversal, and long-term reversal sorted portfolios, in a rolling out-of-sample fashion over 1963:07–2017:06. Panel B presents the contribution of tail risk to the overall risk premium of the aggregate stock market portfolio (Row 1) and to the risk premium of the cross-section (Row 2). The contribution of tail risk is computed as the ratio of (i) the covariance between the I-SDF and the excess returns on the market (Row 1) or each asset in the cross-section (Row 2), computed using only those observations that lie in the tail of the I-SDF (Columns 3 and 4), the excess returns (Columns 5 and 6) or both (Columns 7 and 8) and (ii) the overall covariance between the I-SDF and the excess returns on the market (Row 1) or each asset in the cross-section (Row 2), computed using all the available observations. The two numbers in each cell denote the ratio for the excess market return in Row 1 (the mean of the ratio across the assets in the cross-section for Row 2), along with the fraction of the total number of observations belonging to the corresponding tail in square brackets below.

stock market variables in the recovery of the I-SDF. However, our results also indicate that, while business cycle risk is undoubtedly important and spanned by the asset returns used to construct the I-SDF, it cannot fully explain its behavior. We find evidence that the true underlying SDF may not solely be a function of the business cycle, but also of additional variables related to the performance of the stock market (that is only imperfectly correlated with the business cycle).

4.2 I-SDF and Tail Risk

The I-SDF has a strong non-Gaussian distribution. Panel A of Table 6 presents the summary statistics from the distribution of the I-SDF. The I-SDF is strongly positively skewed with a coefficient of skewness equal to 4.5. The coefficient of excess kurtosis is also very high at 35.5. These results suggest that tail risk is an important source of priced risk.

With the estimated I-SDF at hand, it is also possible to quantify the share of the risk premium of an asset attributable to tail risk compensation. Recall that, given an SDF M_t, the lack of arbitrage opportunities implies that the risk premium on an asset i is

$$\mathbb{E}\left[R^e_{i,t}\right] = -\frac{\int \left(M_t - \bar{M}\right) R^e_{i,t} \mathbf{1}_{\{(M_t, R^e_{i,t}) \in \mathbb{A}\}} d\mathbb{P} + \int \left(M_t - \bar{M}\right) R^e_{i,t} \mathbf{1}_{\{(M_t, R^e_{i,t}) \notin \mathbb{A}\}} d\mathbb{P}}{\mathbb{E}\left[1/R^f_t\right]}$$

for any arbitrary set $\mathbb{A}$, where $\mathbf{1}_{\{\}}$ denotes the indicator function that takes value 1 when the condition in brackets is satisfied, and the numerator of the above expression is simply the covariance between the SDF and the excess return on asset i. As a consequence, using our estimate of the I-SDF, choosing a set $\mathbb{A}$ in the tail of the empirical distribution of $\hat{M}_t$ and $R^e_{i,t}$, and replacing the integral with a sample analogue, we can compute the share of the asset's observed risk premium generated by tail risk. Note that the above decomposition only holds for an SDF that prices asset *i* perfectly, that is, has zero pricing errors in the Euler equation for asset *i*. Since the out-of-sample I-SDF has non-zero pricing errors (as shown in Tables 1–5), for the purpose of this decomposition, we use the in-sample SDF constructed from the small but broad cross-section of fifteen portfolios in Table 1, Panel C that, by construction, prices the assets perfectly.

Panel B of Table 6 reports the share of the risk premium attributable to tail risk. Consider first Row 1 that presents the fraction of the market risk premium attributable to tail risk. Columns 3 and 4 show that the 5% (10%) most extreme (positive) realizations of the SDF drive 33.0% (45.7%) of the total market risk premium. Columns 5 and 6 show that the 5% (10%) most extreme (negative) realizations of market returns generate 50.4% (57.6%) of the observed market risk premia. Finally, the last two columns report the share of observed risk premia generated by the joint tail of asset returns and the I-SDF. The joint tail of the I-SDF and the market portfolio accounts on average for 42.7–61.5% of the market risk premium. Specifically, 1.2% of the possible joint states generate almost half (42.7%) of the market equity premium and 3.4% of the possible joint states generate 61.5% of the premium.

Similar results, presented in Row 2 of Panel B, obtain for the proportion of the observed risk premia on the cross-section of equity portfolios attributable to tail risk. Across the cross-section of 55 test assets—that consists of the five industry portfolios and the decile portfolios formed from univariate sorts on size, book-to-market equity, momentum, short-term reversals, and long-term reversals—the 5% (10%) most extreme realizations of the I-SDF drive 39.7% (48.4%) of the assets' risk premia on average, the 5% (10%) most extreme realizations of the asset returns generate 42.7% (50.0%) of the premia on average, and the joint tail of the I-SDF and the asset return accounts on average for 37.4–60.7% of the premia.

To facilitate interpretation, we also offer rough guidance on what these numbers would look like in a Gaussian world. The observations in the 2.5% right tail of the distribution of the I-SDF account for almost 30% (27.2% to be exact) of the market risk premium. If, however, the underlying SDF were Gaussian, these observations would all lie in the 0.46% right tail of its distribution, that is, such events would be more than five times less frequent. We obtain this latter number from a Gaussian SDF with the same mean and variance as the recovered I-SDF. Similarly, the observations in the 2.5% left tail of the distribution of the market return account for over 30% (32.3% to be precise) of the market risk premium. If, however, the market return were Gaussian, these observations would all lie in the 1% left tail of its distribution, that is, such events would be more than two and a half times less frequent. Once again, we obtain this by looking at a Gaussian distribution for the market return with the same mean and variance estimated from the historical sample.

Overall, our results suggest that a very large part of the observed risk premia are due to a compensation for tail risk. A large literature on asset pricing highlights the crucial role of tail events in explaining several observed features of stock market data. Examples include the rare disasters paradigm in a representative agent setting that models negative skewness in aggregate consumption growth (see, e.g., Barro 2006) as well as models with incomplete markets that feature negative skewness in the cross-sectional distribution of idiosyncratic labor income growth (see, e.g., Mankiw 1982; Constantinides and Ghosh 2017). By their very definition, such disaster events are rare making it difficult to rigorously calibrate such models. Our methodology offers a data-driven approach to calibrating models featuring tail risks as an important source of systematic risk.

4.3 I-SDF versus Tangency Portfolio

As highlighted in the introduction, using relative entropy minimization as a criterion to recover the underlying pricing kernel or its components has been recently advocated in the literature. Equations (2) and Appendix (A1) provide the methodology to recover the kernel, for two alternative definitions of relative entropy. Section 1.1 enumerates a number of desirable characteristics of using this approach to recover the SDF. Theoretically, this procedure is distinct from estimating the mean–variance efficient tangency portfolio from a given cross-section of assets.

To see this, note that a cumulant expansion of the minimized objective function in Equation (A1) gives:

$$D\left(\mathbb{P}||\hat{\mathbb{Q}}\right) = \frac{\kappa_2^{\hat{M}}}{2!} + \frac{\kappa_3^{\hat{M}}}{3!} + \frac{\kappa_4^{\hat{M}}}{4!} + \ldots, \tag{10}$$

where $\hat{\mathbb{Q}}$ solves Equation (A1), and $\hat{M} = \frac{d\hat{\mathbb{Q}}}{d\mathbb{P}}$ with $E(\hat{M}) = 1$. In the above equation, $\kappa_3^{\hat{M}}$ denotes the j-th moment of $\ln(\hat{M})$, that is, $\kappa_2^{\hat{M}}$ denotes its variance, $\kappa_3^{\hat{M}}$ its skewness, $\kappa_4^{\hat{M}}$ its kurtosis, and so on.

Therefore, the relative entropy minimization approach recovers a pricing kernel $\hat{M}$ that has the minimum weighted sum of *all* its moments. Recall that, the tangency portfolio, on the other hand, is chosen to have the minimum variance. If the true underlying pricing kernel were log normal, then all of its higher order moments would be zero, such that minimizing relative entropy would be equivalent to minimizing the variance (see the above equation). If, on the other hand, the SDF is not log normal (for example, if we believe that tail risk is an important source of priced risk), then the variance is not a sufficient statistic summarizing its distribution, and there is no *a priori* reason to recover the SDF to minimize solely the variance.

Note that, the I-SDF, by construction, prices perfectly in-sample the cross-section of assets used to recover it. Therefore, a linear projection of the I-SDF on to the set of asset returns used to construct it identifies the tangency portfolio for that set of assets (we refer to this portfolio as the I-P). Appendix A.2 provides a derivation of this result. In this section, we evaluate whether the non-linear dependency of the I-SDF on the underlying assets carries valuable pricing information. We do so by comparing the I-SDF's pricing ability to the pricing performance of the corresponding I-P.

The I-P is constructed, in a rolling out-of-sample fashion, from time series regressions of the I-SDF on to the set of asset returns used in its construction. Specifically, using the notation in Section 1.1, in subsample i, the estimates of the Lagrange multipliers, $\hat{\theta}_{T_i}$, are used to construct the in-sample SDF $\hat{M}_{i,t} \equiv M\left(\hat{\theta}_{T_i}, \mathbf{R}_t^e\right)$, $t = T_i - \bar{T} + 1,\ T_i - \bar{T} + 2,\ \ldots,\ T_i$. Then $\hat{M}_{i,t}$ is projected onto the space of returns to obtain the vector of portfolio weights $\omega_{T_i} \in \mathbb{R}^N$ (normalized to sum to unity). That is, the mimicking portfolio weights ω_{T_i} are given by

$$\omega_{T_i} := -\frac{\hat{\mathbf{b}}_{T_i}}{|\hat{\mathbf{b}}'_{T_i}\iota|},\ \left[\hat{a}_{T_i}, \hat{\mathbf{b}}'_{T_i}\right] := \underset{\{a_{T_i}, \mathbf{b}'_{T_i}\}}{\operatorname{argmin}} \frac{1}{\bar{T}} \sum_{t=T_i-\bar{T}+1}^{T_i} \left(\hat{M}_{i,t} - a_{T_i} - \mathbf{b}_{T_i}\mathbf{R}_t\right)^2, \tag{11}$$

where ι denotes a conformable column vector of ones. Using the portfolio weights vector, the *out-of-sample* I-P is obtained as $R_t^{IP} = \omega'_{T_i}\mathbf{R}_t$ for the subsequent s periods, that is, for $t = T_i + 1,\ T_i + 2,\ \ldots,\ T_{i+1}$. This process is repeated for each subsample to obtain the time series of the I-P over the out-of-sample evaluation period.

To compare the performance of the I-SDF vis-à-vis the I-P, we consider the unique I-SDF recovered in Section 3.2 from the small, but broad, cross-section of assets, encompassing several well-documented anomalies. The I-P is obtained as the linear projection of the

Table 7. I-SDF versus its linear projection

Row	Assets	$const.(\%)$	λ_{sdf}	λ_{proj}	$\bar{R}^2_{OLS}(\%)$	$\bar{R}^2_{GLS}(\%)$	T^2	q
Panel A: Rolling Sample = 30 years								
(1)	FF25	−0.00 (−0.013)	−0.741 (−3.20)		68.6	36.1	37.3 (0.030)	0.107
		0.14 (0.52)		0.069 (3.63)	67.8	35.3	50.2 (0.001)	0.109
(2)	*55 Decile Portfolios* *ME,BE/ME,Mom,STR,LTR,5Ind.*	0.35 (1.66)	−0.313 (−4.69)		74.7	43.1	69.5 (0.063)	0.124
		0.38 (1.91)		0.034 (5.3)	80.2	46.9	68.1 (0.080)	0.116
(3)	25 FF + 30 Ind + 10 Mom	0.38 (1.70)	−0.313 (−4.34)		50.3	32.3	143.1 (0.000)	0.255
		0.41 (1.93)		0.033 (4.46)	51.8	29.1	158.1 (0.000)	0.267
Panel B: Rolling Sample = 10 years								
(1)	FF25	0.20 (0.51)	−1.10 (−2.42)		50.3	36.7	36.4 (0.037)	0.106
		0.74 (3.82)		−0.517 (−0.315)	−3.91	22.4	81.7 (0.000)	0.128
(2)	*55 Decile Portfolios* *ME,BE/ME,Mom,STR,LTR,5Ind.*	0.38 (1.70)	−0.55 (−3.96)		62.8	34.1	76.0 (0.021)	0.143
		0.64 (3.40)		−2.05 (−1.22)	2.65	6.09	106.5 (0.000)	0.204
(3)	25 FF + 30 Ind + 10 Mom	0.41 (1.79)	−0.528 (−3.94)		42.4	28.1	145.5 (0.000)	0.271
		0.66 (3.59)		−0.87 (−0.73)	−0.46	18.3	191.0 (0.000)	0.308

Cross-sectional regressions of the average excess returns listed in column 2 on the estimated factor loadings for the I-SDF and its linear projection (the I-P), at the monthly frequency over 1963:07–2017:06. The I-SDF is constructed using a cross-section of fifteen portfolios, consisting of five industry portfolios and the smallest and largest deciles of size, BM, momentum, short-term reversal, and long-term reversal sorted portfolios, in a rolling out-of-sample fashion over 1963:07–2017:06. Panels A and B present the results for rolling samples of length 30 years and 10 years, respectively. For each set of returns, the table presents the intercept and slopes, along with Shanken t-statistics in parentheses. It also presents the OLS-adjusted R^2 and the GLS-adjusted R^2. The last two columns present, respectively, Shanken's (1985) cross-sectional T^2 statistic along with its asymptotic p-value in parentheses, and the q statistic.

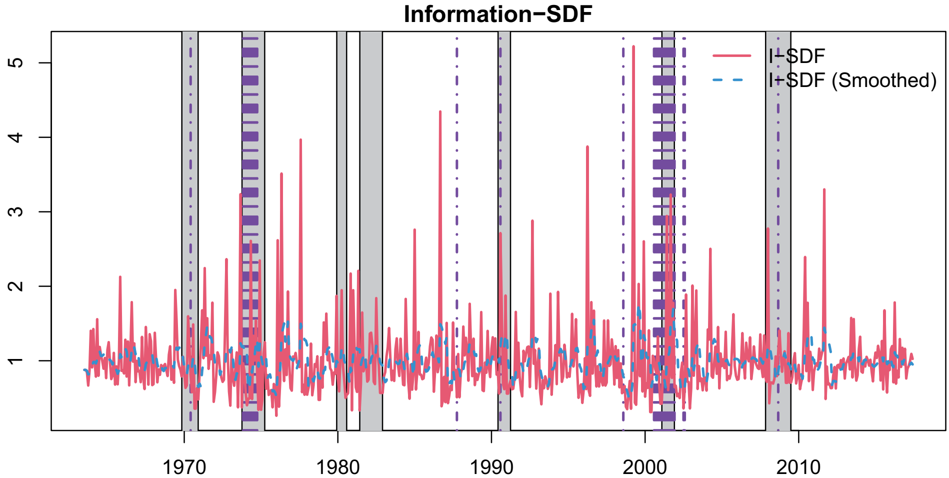

Figure 2. Time series of the I-SDF (red line) and the 1-year moving average of the I-SDF (blue-dashed line). The I-SDF is constructed from fifteen portfolios—consisting of five industry portfolios and the smallest and largest deciles of size, BM, momentum, short-term reversal, and long-term reversal sorted portfolios—using the alternative relative entropy minimizing procedure, in a rolling out-of-sample fashion starting in 1963:07. NBER recession periods are marked by the gray vertical bands. The vertical dot-dashed lines indicate market crashes identified using the Mishkin and White (2002) approach.

I-SDF as described above. The results are presented in Table 7. Panels A and B report the results using a rolling sample period of 30 and 10 years, respectively. Panel A shows that the I-P has a comparable performance to that of the I-SDF for all different cross-sections of test assets when the length of the training sample is long, that is, the rolling out-of-sample period used to construct it is long. Specifically, the I-P produces similar intercepts and $\bar{R}^2_{OLS}$ and $\bar{R}^2_{GLS}$ compared to those generated by the I-SDF. However, Panel B shows that the pricing performance of the I-P is volatile and unstable, relying critically on long training periods. Specifically, its estimated price of risk changes sign when moving from a 30-year training period in Panel A to a 10-year period in Panel B—the estimate in Panel A is always positive with an average value of 0.045 across the three cross-sections, compared to being always negative with an average of -1.15 in Panel B. Also, its $\bar{R}^2_{OLS}$ falls drastically when moving from Panel A to B: 67.8% to −3.9% when the 25 FF portfolios are the test assets; 80.5% to 2.7% when the test assets are the ten decile portfolios formed from univariate sorts on size, B/M, momentum, short-term reversals, and long-term reversals, and the five industry portfolios; and 51.8% to −.46% when the test assets consist of the 25 FF, 30 Industry, and 10 momentum portfolios. The I-SDF, on the other hand, has a consistently negative price of risk in both panels, which is strongly statistically significant in all six cases. Its $\bar{R}^2_{OLS}$ also does fall when moving from Panel A to B—from 68.6% to 50.3% for the 25 FF portfolios, 74.7% to 62.8% for the 55 decile portfolios, and 50.3% to 42.4% for the twenty-five FF, thirty Industry, and ten momentum portfolios—albeit much less so compared to the I-P.

We next proceed to show that the I-SDF contains novel information, over and above that contained in its linear projection (namely, the I-P), that is not captured by standard multi-factor asset pricing models. Table 8 presents time series regressions of the I-SDF on the FF3, FF5, and HXZ factors. If the factors fully explain the variation in the I-SDF, the intercepts from the time series regressions should be indistinguishable from zero and the R^2 of the regressions should be high. Note that, in the case of similar time series regressions

Table 8. Explaining I-SDF and I-P with FF factors

Row	Assets	$\alpha_{sdf}(\%)$	$\alpha_{\mathrm{IP}}(\%)$	β_{Rm}	β_{SMB}	β_{HML}	β_{RMW}	β_{CMA}	β_{IA}	β_{ROE}	$\bar{\mathrm{R}}^2_{OLS}(\%)$
Monthly, 1963:07–2017:06											
(1)	I-SDF (FF3)	1.03 (33.0)		−3.07 (−4.13)	−1.50 (−1.43)	−4.87 (−4.29)					4.42
(2)	I-P (FF3)		3.06 (7.39)	0.40 (4.08)	0.42 (3.03)	0.93 (6.16)					7.61
(3)	I-SDF (FF5)	1.07 (34.5)		−4.78 (−6.21)	−2.62 (−2.39)	1.20 (0.80)	−4.74 (−3.15)	−13.6 (−6.18)			10.0
(4)	I-P (FF5)		2.28 (5.66)	0.70 (7.08)	0.67 (4.71)	−0.09 (−0.46)	1.01 (5.19)	2.30 (8.05)			17.4
(5)	I-SDF (HXZ4)	1.12 (33.1)		−4.61 (−5.91)	−4.10 (−3.75)				−13.2 (−7.30)	−6.50 (−4.98)	12.6
(6)	I-P (HXZ4)		1.65 (3.82)	0.65 (6.53)	0.96 (6.84)				2.18 (9.46)	1.25 (7.49)	21.7

The table presents the intercept and slope coefficients, along with the *t*-statistics in parentheses, as well as the OLS-adjusted R^2, from time series regressions of the I-SDF (odd rows) and I-P (even rows) on factors identified by popular multi-factor models. The I-SDF and I-P are constructed from a cross-section of fifteen portfolios, consisting of five industry portfolios and the smallest and largest deciles of size, BM, momentum, short-term reversal, and long-term reversal sorted portfolios, in a rolling out-of-sample fashion over 1963:07–2017:06. Note that, since the I-P weights in Equation (11) are proportional to minus the projection coefficients, and are normalized to sum to 1, one would expect the betas of I-P and I-SDF to have opposite signs and their magnitudes are not directly comparable.

of the I-P on a set of candidate risk factors, the intercepts have the interpretation of a standard α.

Row 1 shows that the three FF factors explain only 4.4% of the variation in the I-SDF. Moreover, the estimated intercept is strongly statistically significant, with an annualized value of 12.4%. Note that since the I-SDF is not a tradeable factor, the intercept is not interpretable as a tradeable alpha. Row 2 shows that the 3 FF factors can explain a larger fraction of the variation in the I-P than in the I-SDF (7.6% versus 4.4%). However, even in this case, the bulk of the variation is left unexplained by the FF factors. Moreover, the estimated intercept, which in this case has the interpretation of a standard α, is statistically and economically large, at 36.7% per annum.

Similar results are obtained for the five FF factors. The $\bar{R}^2_{OLS}$ from the I-SDF regression is only 10.0% (Row 3), showing that a substantial proportion of the variability in the I-SDF cannot be explained by the movements in the FF5 factors. Row 4 shows that the $\bar{R}^2_{OLS}$ from the I-P regression is higher at 17.4%, but still, a substantial fraction of the variability is left unexplained. In both cases, the estimated annualized intercepts are statistically significant and economically large, varying from 12.8% for the I-SDF to 27.4% for the I-P. The results obtained for the four HXZ factors, reported in Rows 5 and 6, are, once again, very similar.

Overall, the results in Tables 7 and 8 suggest that the non-linearity of the I-SDF contains valuable information about the underlying sources of priced risk—information that is partially lost by the linear projection of the I-SDF on to the set of asset returns used in its construction.

5 Conclusion

Given a cross-section of asset returns, we show how an information-theoretic approach can be used to estimate, non-parametrically, an out-of-sample pricing kernel. We show that this “information SDF” prices broad cross-sections of asset returns better than commonly employed multi-factor models (e.g., FF3, HXZ, and FF5 models) and that, unlike these factor models, it seems to more closely pin down the tangency portfolio out-of-sample, as a correct SDF should. Moreover, the I-SDF extracts novel pricing information not captured by the Fama–French and HXZ factors (which explain only a small share of the I-SDF’s time variation). The I-SDF offers a useful benchmark against which competing theories and investment strategies can be evaluated.

Appendix A

A.1. An Alternative Minimum Entropy Pricing Kernel

The definition of relative entropy, or KLIC, implies that this discrepancy metric is not symmetric, that is, generally $D(\mathbb{A}||\mathbb{B}) \neq D(\mathbb{B}||\mathbb{A})$ unless $\mathbb{A}$ and $\mathbb{B}$ are identical (in which case their divergence would be zero). This implies that for measuring the information divergence between $\mathbb{Q}$ and $\mathbb{P}$, we can also interchange the roles of $\mathbb{Q}$ and $\mathbb{P}$ in Equation (2) to recover $\mathbb{Q}$ as

$$\underset{\mathbb{Q}}{\operatorname{argmin}}\, D(\mathbb{P}||\mathbb{Q}) \equiv \underset{\mathbb{Q}}{\operatorname{argmin}} \int \ln \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{P} \quad \text{s.t.} \quad \int \mathbf{R}_t^e d\mathbb{Q} = \mathbf{0}. \tag{A1}$$

Since $\frac{M_t}{M} = \frac{d\mathbb{Q}}{d\mathbb{P}}$, the optimization in Equation (A1) can be rewritten as

$$\underset{M_t}{\operatorname{argmin}}\, \mathbb{E}^{\mathbb{P}}[\ln M_t] \quad \text{s.t.} \quad \mathbb{E}^{\mathbb{P}}\left[M_t \mathbf{R}_t^e\right] = 0.$$

where, to simplify the exposition, we have used the innocuous normalization $\bar{M} = 1$. Replacing the expectation with a sample analog yields

$$\underset{\{M_t\}_{t=1}^T}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1}^{T} \ln M_t \quad \text{s.t.} \quad \frac{1}{T}\sum_{t=1}^{T} M_t \mathbf{R}_t^e = 0. \tag{A2}$$

Thanks to Fenchel's duality theorem (see, e.g., Csiszar 1975) this entropy minimization is solved by

$$\hat{M}_t \equiv M_t\left(\hat{\theta}_T, \mathbf{R}_t^e\right) = \frac{1}{T(1+\hat{\theta}_T'\mathbf{R}_t^e)}, \quad \forall t \tag{A3}$$

where $\hat{\theta}_T \in \mathbb{R}^N$ is the solution to

$$\underset{\theta}{\operatorname{argmin}} \; -\frac{1}{T}\sum_{t=1}^{T} \log(1+\theta'\mathbf{R}_t^e),$$

and this last expression is the dual formulation of the entropy minimization problem in Equation (A2). Note also that this dual problem is analogous to estimating the so-called growth-optimal portfolio (i.e., the portfolio with the maximum log return).

We next proceed to show that the entropy minimization problem in Equation (A1) also delivers a maximum likelihood estimate of the risk-neutral measure. Let the vector $\mathbf{z}_t$ be a sufficient statistic for the state of the economy at time t. Given $\mathbf{z}_t$, the equilibrium quantities, such as asset returns $\mathbf{R}^e$ and the SDF M, are just a mapping from $\mathbf{z}$ on to the real line, that is,

$$M(\mathbf{z}) : \mathbf{z} \to \mathbb{R}_+, \quad \mathbf{R}^e(\mathbf{z}) : \mathbf{z} \to \mathbb{R}^N, \quad M_t \equiv M(\mathbf{z}_t), \quad \mathbf{R}_t^e \equiv \mathbf{R}^e(\mathbf{z}_t),$$

where $\mathbf{z}_t$ is the time t realization of $\mathbf{z}$.

Equipped with the above definition, we can rewrite the Euler Equation (1) as

$$0 = \mathbb{E}\left[\mathbf{R}_t^e M_t\right] \equiv \int \mathbf{R}_t^e M_t dP = \int \mathbf{R}^e(\mathbf{z}) M(\mathbf{z}) p(\mathbf{z}) d\mathbf{z}, \tag{A4}$$

where $p(\mathbf{z})$ is the pdf associated with the physical measure P. Moving to the risk-neutral measure we have

$$0 = \mathbb{E}\left[\mathbf{R}_t^e M_t\right] = \mathbb{E}^Q\left[\mathbf{R}_t^e\right] = \int \mathbf{R}^e(\mathbf{z}) q(\mathbf{z}) d\mathbf{z}, \tag{A5}$$

where $q(\mathbf{z})$ is the pdf associated with the risk-neutral measure Q and $M = dQ/dP$. Note that

$$D(P||Q) = \int \ln \frac{dP}{dQ} dP = \int p(\mathbf{z}) \ln p(\mathbf{z}) d\mathbf{z} - \int p(\mathbf{z}) \ln q(\mathbf{z}) d\mathbf{z}.$$

Since the first term on the right-hand side of the above expression does not involve q, $D(P||Q)$ is minimized, with respect to Q, by choosing the distribution that maximizes the second term, that is,

$$Q^* \equiv \underset{Q}{\operatorname{argmin}}\, D(P||Q) \equiv \underset{q}{\operatorname{argmax}}\, \mathbb{E}\left[\ln q(\mathbf{z})\right] \text{ s.t. } \mathbb{E}^Q\left[\mathbf{R}_t^e\right] = 0.$$

That is, the minimum entropy estimator in Equation (A1) maximizes the expected—risk-neutral—log likelihood. Approximating the continuous distribution $q(\mathbf{z})$ with a multinomial distribution $\{q_t\}_{t=1}^T$ that assigns probability weight q_t to the time t realizations of $\mathbf{z}$, a non-parametric maximum likelihood estimator of Q can be obtained as

$$\{q_t^*\}_{t=1}^T = \operatorname{argmax} \quad \frac{1}{T}\sum_{t=1}^T \ln q_t \tag{A6}$$

$$\text{s.t. } q_t \in \Delta^T \equiv \{(q_1, q_2, \ldots, q_T) : q_t \geqslant 0, \sum_{t=1}^T q_t = 1\} \text{ and holds,}$$

provided that

$$\frac{1}{T}\sum_{t=1}^T \ln q_t \underset{T\to\infty}{\overset{p.}{\to}} \mathbb{E}\left[\ln q(\mathbf{z})\right].$$

Since the SDF estimate obtained with the EL approach in Equation (A3) can, in principle, lead to negative realizations out-of-sample, we focused on the SDF estimated with Equation (4) in the main text which guarantees the positivity of the SDF. Here we present

Table A.1. Unique I-SDF Recovered From 15 portfolios, 1963:07–2017:06

Row	$const.(\%)$	λ_{sdf}	$\bar{R}^2_{OLS}(\%)$	$\bar{R}^2_{GLS}(\%)$	T^2	q
Panel A: 25 FF Portfolios						
I-SDF	−0.10 (−0.43)	−0.526 (−3.55)	32.6	25.8	40.1 (0.015)	0.124
Panel B: 55 Test Assets						
I-SDF	0.21 (3.47)	−0.272 (−7.01)	47.1	26.0	82.2 (0.006)	0.161
Panel C: 25 FF + 30 Industry + 10 Momentum						
I-SDF	0.04 (0.407)	−0.410 (−6.61)	40.0	25.7	114.1 (0.000)	0.283

Cross-sectional regressions of average excess returns of different sets of test assets on the estimated factor loading, for the I-SDF recovered using the alternative relative entropy minimization approach. Panel A presents the results when the test assets consist of the twenty-five size and BM sorted portfolios of FF. Panel B presents results when the test assets consist of ten size-sorted, ten BM-sorted, ten momentum-sorted, ten short-term reversal sorted, ten long-term reversal sorted, and five industry-sorted portfolios. Panel C presents results when the test assets consist of the twenty-five FF, thirty Industry, and ten Momentum-sorted portfolios. The I-SDF is constructed from fifteen portfolios—consisting of five industry portfolios and the smallest and largest deciles of size, BM, momentum, short-term reversal, and long-term reversal sorted portfolios—using a relative entropy minimizing procedure, in a rolling out-of-sample fashion starting in 1963:07. The table presents the intercept and slopes, along with Shanken t-statistics in parentheses. It also presents the OLS-adjusted R^2 and the GLS-adjusted R^2. The last two columns present, respectively, Shanken's (1985) cross-sectional T^2 statistic along with its asymptotic p-value in parentheses, and the q statistic that measures how far the factor-mimicking portfolios are from the mean–variance frontier of the test assets.

the pricing performance of the alternative SDF in Equation (A3) for the cross-sectional out-of-sample exercise in which a unique I-SDF is recovered from a small cross-section of fifteen anomaly portfolios and is then used to price, out-of-sample, several other larger cross-sections (analogous to Table 3). Overall, the performance of the EL method is only marginally worse than that of the ET in Table 3 and, despite this not being guaranteed, it never yields a negative SDF out-of-sample in the applications we consider. Figure 2 plots the time series of the EL-based I-SDF. Overall, its time series behavior is similar to the one obtained with the ET approach, and their correlation is 38.5%.

A.2. I-SDF versus Tangency Portfolio

The sample-based mean–variance approach identifies the tangency (maximum Sharpe Ratio) portfolio weights as:

$$\omega = \frac{[Var(\mathbf{R})]^{-1}\mathbb{E}(\mathbf{R}^e)}{\iota'[Var(\mathbf{R})]^{-1}\mathbb{E}(\mathbf{R}^e)}, \tag{A7}$$

The projection of the I-SDF, M, onto the set of assets used for its construction is obtained from the regression:

$$M_t = a + b'\mathbf{R}_t + \varepsilon_t. \tag{A8}$$

Therefore, the (normalized) weights for the Information Portfolio (I-P) are given by:

$$b = \frac{[Var(\mathbf{R}_t)]^{-1}Cov(\mathbf{R}_t, M_t)}{\iota'[Var(\mathbf{R}_t)]^{-1}Cov(\mathbf{R}_t, M_t)}, \tag{A9}$$

Now, note that since M_t, by construction, is a valid pricing kernel in that it prices the cross-section of test assets perfectly in sample, we have

$$\mathbb{E}\left(\mathbf{R}_t^e\right) = -\frac{Cov(\mathbf{R}_t, M_t)}{\mathbb{E}(M_t)}. \tag{A10}$$

Substituting $Cov(\mathbf{R}_t, M_t)$ from Equation (A10) in the expression for b in Equation (A9) (the weights on the I-P), we have

$$b = \frac{[Var(\mathbf{R})]^{-1}\mathbb{E}(\mathbf{R}^e)}{\iota'[Var(\mathbf{R})]^{-1}\mathbb{E}(\mathbf{R}^e)} \equiv \omega. \tag{A11}$$

Hence, in sample, the portfolio weights for the I-P in Equation (A11) are identical to that of the mean–variance efficient tangency portfolio ones in Equation (A7).

References

Ait-Sahalia, Y., and A. W. Lo. 1998. Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices. *The Journal of Finance* 53: 499–547.

Ait-Sahalia, Y., and A. W. Lo. 2000. Nonparametric Risk Management and Implied Risk Aversion. *Journal of Econometrics* 94: 9–51.

Almeida, C., K. Ardison, and R. Garcia. 2020. Nonparametric Assessment of Hedge Fund Performance. *Journal of Econometrics* 214: 349–378.

Almeida, C., and G. Freire. 2022. Pricing of Index Options in Incomplete Markets. *Journal of Financial Economics* 144: 174–205.
Almeida, C., and R. Garcia. 2012. Assessing Misspecified Asset Pricing Models with Empirical Likelihood Estimators. *Journal of Econometrics* 170: 519–537.
Almeida, C., and R. Garcia. 2016. Economic Implications of Nonlinear Pricing Kernels. *Management Science* 63: 3361–3380.
Alvarez, F., and U. J. Jermann. 2005. Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth. *Econometrica* 73: 1977–2016.
Asness, C. S., T. J. Moskowitz, and L. H. Pedersen. 2013. Value and Momentum Everywhere. *The Journal of Finance* 68: 929–985.
Back, K. 2010. *Asset Pricing and Portfolio Choice Theory Financial Management Association Survey and Synthesis*. Oxford: Oxford University Press.
Backus, D., M. Chernov, and S. E. Zin. 2014. Sources of Entropy in Representative Agent Models. *The Journal of Finance* 69: 51–99.
Bansal, R., and B. N. Lehmann. 1997. Growth-Optimal Portfolio Restrictions on Asset Pricing Models. *Macroeconomic Dynamics* 1: 333–354.
Barro, R. J. 2006. Rare Disasters and Asset Markets in the Twentieth Century. *Quarterly Journal of Economics* 121: 823–866.
Borovicka, J. L. Hansen, and J. Scheinkman. 2016. Misspecified Recovery. *The Journal of Finance* 71: 2493–2544.
Brown, D. E., and R. L. Smith. 1990. A Correspondence Principle for Relative Entropy Minimization. *Naval Research Logistics* 37: 191–202.
Bryzgalova, S. 2016. *Spurious Factors in Linear Asset Pricing Models*. Working paper.
Bryzgalova, S., J. Huang, and C. Julliard. 2023. Bayesian Solutions for the Factor Zoo: We Just Ran Two Quadrillion Models. *The Journal of Finance* 78: 487–557.
Chen, X., L. P. Hansen, and P. G. Hansen. 2020. Robust Identification of Investor Beliefs. *Proceedings of the National Academy of Sciences of the United States of America* 117: 33130–33140.
Cochrane, J. H. 2005. *Financial Markets and the Real Economy*. Norwell, MA:Now Publishers Inc.
Constantinides, G., and A. Ghosh. 2017. Asset Pricing with Countercyclical Household Consumption Risk. *The Journal of Finance* 72: 415–460.
Csiszar, I. 1975. I-Divergence Geometry of Probability Distributions and Minimization Problems. *Annals of Probability* 3: 146–158.
DeMiguel, V., L. Garlappi, and R. Uppal. 2009. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy. *Review of Financial Studies* 22: 1915–1953.
Fama, E. F., and K. R. French. 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33: 3–56.
Fama, E. F., and K. R. French. 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116: 1–22.
Ghosh, A, C. Julliard, and M. Stutzer. 2020. *The Market Cost of Business Cycle Fluctuations*. Manuscript. Norwell, MA: McGill University.
Ghosh, A., C. Julliard, and A. P. Taylor. 2017. What is the Consumption-CAPM Missing? An Information-Theoretic Framework for the Analysis of Asset Pricing Models. *Review of Financial Studies* 30: 442–504.
Ghosh, A., and G. Roussellet. 2020. *Identifying Beliefs from Asset Prices*. Manuscript. Norwell, MA: McGill University.
Hansen, L. P. 2014. Nobel Lecture: Uncertainty Outside and Inside Economic Models. *Journal of Political Economy* 122: 945–987.
Hansen, L. P., and R. Jagannathan. 1991. Implications of Security Market Data for Models of Dynamic Economies. *Journal of Political Economy* 99: 225–262.
Hansen, L. P., and R. Jagannathan. 1997. Asssessing Specification Errors in Stochastic Discount Factor Models. *The Journal of Finance* 52: 557–590.
Harrison, M. J, and D. M. Kreps. 1979. Martingales and Arbitrage in Multiperiod Securities Markets. *Journal of Economic Theory* 20: 381–408.
Harvey, C. R., and Y. Liu. 2015. *Lucky Factors*. Durham, NC: Duke University manuscript.
Harvey, C. R., Y. Liu, and H. Zhu. 2015. The Cross-Section of Expected Returns. *Review of Financial Studies* 29: 5–68.

Hou, K., C. Xue, and L. Zhang. 2015. Digesting Anomalies: An Investment Approach. *Review of Financial Studies* 28: 650–705.

Jackwerth, J. C., and M. Rubinstein. 1996. Recovering Probability Distributions from Option Prices. *The Journal of Finance* 51: 1611–1631.

Julliard, C., and A. Ghosh. 2012. Can Rare Events Explain the Equity Premium Puzzle. *Review of Financial Studies* 25: 3037–3076.

Kan, R., and C. Zhang. 1999a. GMM Tests of Stochastic Discount Factor Models with Useless Factors. *Journal of Financial Economics* 54: 103–127.

Kan, R., and C. Zhang. 1999b. Two-Pass Tests of Asset Pricing Models with Useless Factors. *The Journal of Finance* 54: 203–235.

Kitamura, Y., and M. Stutzer. 2002. Connections between Entropic and Linear Projections in Asset Pricing Estimation. *Journal of Econometrics* 107: 159–174.

Kleibergen, F. 2009. Tests of Risk Premia in Linear Factor Models. *Journal of Econometrics* 149: 149–173.

Kleibergen, F., and Z. Zhan. 2015. Unexplained Factors and Their Effects on Second Pass R-Squared's. *Journal of Econometrics* 189: 101–116.

Korsaye, S. A., A. Quaini, and F. Trojani. 2021. "Smart Stochastic Discount Factors." Swiss Finance Institute Research Paper Series 21–51, Zürich, Switzerland, Swiss Finance Institute.

Kozak, S., S. Nagel, and S. Santosh. 2020. Shrinking the Cross-Section. *Journal of Financial Economics* 135: 271–292.

Lewellen, J., S. Nagel, and J. Shanken. 2010. A Skeptical Appraisal of Asset Pricing Tests. *Journal of Financial Economics* 96: 175–194.

Liu, Y. 2021. Index Option Returns and Generalized Entropy Bounds. *Journal of Financial Economics* 139: 1015–1036.

Lustig, H., N. Roussanov, and A. Verdelhan. 2011. Common Risk Factors in Currency Markets. *Review of Financial Studies* 24: 3731–3777.

Mankiw, N. G. 1982. Hall's Consumption Hypothesis and Durable Goods. *Journal of Monetary Economics* 10: 417–425.

McLean, R. D., and J. Pontiff. 2016. Does Academic Research Destroy Stock Return Predictability. *The Journal of Finance* 71: 5–32.

Mishkin, F. S., and E. N. White. 2002. "US Stock Market Crashes and Their Aftermath: Implications for Monetary Policy." NBER Working Papers 8992, Cambridge, MA, National Bureau of Economic Research, Inc.

Qiu, C., and T. Otsu. 2022. Information Theoretic Approach to High Dimensional Multiplicative Models: Stochastic Discount Factor and Treatment Effect. *Quantitative Economics* 13: 63–94.

Robinson, P. M. 1991. Consistent Nonparametric Entropy-Based Testing. *The Review of Economic Studies* 58: 437–453.

Rosenberg, J. V., and R. F. Engle. 2002. Empirical Pricing Kernels. *Journal of Financial Economics* 64: 341–372.

Ross, S. A. 2015. The Recovery Theorem. *The Journal of Finance* 70: 615–648.

Sandulescu, M., F. Trojani, and A. Vedolin. 2021. Model-Free International Stochastic Discount Factors. *The Journal of Finance* 76: 935–976.

Shanken, J. 1985. Multivariate Tests of the Zero-Beta CAPM. *Journal of Financial Economics* 14: 327–348.

Shanken, J. 1992. On the Estimation of Beta-Pricing Models. *Review of Financial Studies* 5: 1–33.

Stock, J. 1991. Confidence Intervals for the Largest Autoregressive Root in US Macroeconomic Time Series. *Journal of Monetary Economics* 28: 435–459.

Stutzer, M. 1995. A Bayesian Approach to Diagnosis of Asset Pricing Models. *Journal of Econometrics* 68: 367–397.

Stutzer, M. 1996. A Simple Nonparametric Approach to Derivative Security Valuation. *Journal of Finance* LI: 1633–1652.

Uppal, R., and P. Zaffaroni. 2015. *Portfolio Choice with Model Misspecification: A Foundation for Alpha and Beta Portfolios*. Working paper.

White, H. 1982. Maximum Likelihood of Misspecified Models. *Econometrica* 50: 1–25.

Journal of Financial Econometrics, 2025, 23, 1–39
https://doi.org/10.1093/jjfinec/nbae033
Article