An Information-Theoretic Asset Pricing Model*

Anisha Ghosh† Christian Julliard‡ Alex P. Taylor§

July 5, 2016

Abstract

We show that a non-parametric estimate of the pricing kernel, extracted using an information-theoretic approach, delivers *out-of-sample* smaller pricing errors and better cross-sectional fit than leading factor models, and identifies the maximum Sharpe ratio portfolio. This *information SDF* identifies a novel source of risk not captured by Fama-French and momentum factors, revealing an 'information anomaly' that generates annualized alphas of about 9%–24%. A tradable information portfolio that mimics this kernel has high out-of-sample Sharpe ratio (about 1 or more), outperforming both the $1/N$ benchmark and Value and Momentum strategies combined. These results hold for wide cross-sections of test portfolios.

Keywords: Pricing Kernel, Relative Entropy, Factor Models, Factor Mimicking Portfolios, Alpha.
JEL Classification Codes: G11, G12, G13, C13, C53

*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, LSE, 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.

†Tepper School of Business, Carnegie Mellon University; anishagh@andrew.cmu.edu.

‡Department of Finance and FMG, London School of Economics, and CEPR; c.julliard@lse.ac.uk.

§Department of Finance, Manchester Business School; alex.taylor@mbs.ac.uk.

I Introduction

Asset prices contain information about the stochastic discounting of possible future states, i.e., 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 we evaluate its out-of-sample empirical performance in pricing assets and guiding portfolio investment decisions.

The proliferation of risk factors identified in the literature on empirical asset pricing 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 (2015)), and highlights the risk of collective over-paramterization 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, and provides a model-free test of the efficient market hypothesis. Moreover, given its strong empirical performance, it provides a benchmark model relative to which both competing theories, as well as investment managers, can be evaluated.

Building upon Ghosh, Julliard, and Taylor (2016), we show how the pricing kernel can be estimated in a non-parametric fashion using no arbitrage (Euler equation) restrictions. In particular, given the time series data of returns on a cross-section of assets, we rely on a model-free relative entropy minimization approach to estimate an SDF that prices the given cross-section. The solution to this problem 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). This approach delivers a non-parametric maximum likelihood estimate of the SDF and can, therefore, be interpreted as the *most likely* one-factor pricing model for the cross-section used for its construction.

We project the SDF *out-of-sample* for the purposes of cross-sectional pricing and optimal asset allocation. In particular, using the in-sample estimated Lagrange multipliers, we construct the out-of-sample SDF in a rolling fashion, and use it as the single factor to price the cross-section of test assets. 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 the pricing kernel. Instead, the approach allows us to conveniently summarize all the relevant information contained in, possibly 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).

We estimate the I-SDF for diverse sets of equity portfolios – including portfolios sorted on the basis of size, book-to-market-equity, momentum, industry, and long-term reversals – and analyse its ability to explain the cross-section of returns. Compared to leading multifactor models, such as the Fama–French 3-factor model (FF3) or the Carhart 4-factor model (which adds the momentum factor to the FF3), the I-SDF delivers smaller pricing errors on all the different sets of test assets despite being only a one-factor model. Moreover, it explains a larger fraction of the cross-sectional variation of the returns. These results hold for a variety of measures of cross-sectional fit as well as the standard OLS R^2. Most importantly, we show that the I-SDF (unlike the other factor models considered) seems to correctly identify – out-of-sample – the tangency portfolio, i.e. the maximum Sharpe ratio portfolio. Furthermore, we find that the I-SDF extracts novel pricing information not captured by the FF3 or Carhart 4-factor models: it leads to an 'information anomaly,' generating large and statistically significant intercepts (6.9%–17.2% per annum) relative to the FF3 and Carhart factor models, and these factors cannot explain more than one-fourth of its time series variation.

The I-SDF, being a nonlinear function of the asset returns used in its construction, is not a traded factor. Therefore, in order to exploit its ability to identify the tangency portfolio, we construct a tradable portfolio that mimics the estimated kernel, by projecting the I-SDF onto the set of test assets in sample, and using the projection coefficients (normalized to sum to one as portfolio weights) to construct, *out-of-sample*, what we refer to as the 'Information Portfolio' (I-P).

We show that the I-P consistently outperforms a number of standard benchmarks out-of-sample in terms of Sharpe ratios and certainty-equivalent (CEQ) returns. For example, when the 25 size and book-to-market-equity sorted portfolios are used as test assets, the I-P produces an annualized Sharpe Ratio of about 1.0. That is, the I-P delivers a Sharpe ratio that is more than three times that of the Market portfolio, more than twice what is achievable with the naïve $1/N$ diversification strategy[1] or with a "value" strategy, about 3.5 times what is delivered by a "momentum" strategy, and about one-third more than what can be achieved from combining value and momentum strategies.[2] And even after hedging with respect to the FF3 and momentum factors, the I-P produces large annualized hedged Sharpe ratios (up to 0.73) and an annualized α of about 8.6%–23.8%. Moreover, the data never reject the null hypothesis of the I-P delivering out-of-sample the maximum Sharpe ratio achievable using the cross-section of assets used to construct it. Furthermore, using the CEQ metric of DeMiguel, Garlappi, and Uppal (2009), we find that the I-P delivers annualized certainty-equivalent returns of about 14.0%–29.8%, while the other strategies considered

[1] DeMiguel, Garlappi, and Uppal (2009) show that the out-of-sample performance of commonly used mean–variance portfolio selection methods are typically worse than that of the $1/N$ rule in terms of Sharpe ratio and CEQ returns.

[2] Asness, Moskowitz, and Pedersen (2013) document consistent value and momentum return premia across diverse markets and asset classes. Moreover, they show that, thanks to the substantial Sharpe ratios of these strategies and their strong negative correlation, an extremely high Sharpe ratio can be achieved by combining the two.

deliver CEQ returns in the 2.4%–7.4% range, i.e. one order of magnitude smaller than the I-P. Interestingly, the I-P delivers such a strong investment performance with only yearly rebalancing (hence low trading costs), and a substantially smaller tail risk (as measured by skewness and kurtosis) than all the other strategies considered.

Our paper is close in spirit 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) and Ait-Sahalia and Lo (1998)) and use this information to extract an implied pricing kernel (see, e.g. Ait-Sahalia and Lo (2000), Hansen (2014), Rosenberg and Engle (2002), and Ross (2015)). The main advantages of our approach relative to this literature are that *a*) we do not need to use option data and *b*) we can construct an out-of-sample pricing kernel and maximum Sharpe ratio portfolio. Moreover, while the analysis in this paper focuses on equity portfolios, our method is very general and could be applied to other asset classes including bonds, derivatives, and currencies.

The use of an entropy metric is also closely related to Stutzer (1995, 1996) and Kitamura and Stutzer (2002), who first suggested using this information-theoretic alternative to the standard GMM approach to conduct inference for asset pricing models. Julliard and Ghosh (2012) relies on this entropy based inference approach to assess the empirical plausibility of the rare events hypothesis in explaining the equity premium puzzle. Moreover, our work is related to Ghosh, Julliard, and Taylor (2016), who use a relative entropy minimization to derive entropy bounds for the stochastic discount factor of consumption-based asset pricing models (see also Backus, Chernov, and Zin (2014)).

Our paper also 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) documents 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 only 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, Bryzgalova (2015) shows 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 information factors are robust to these concerns and that our approach provides a reliable benchmark against which empirical models can be evaluated.

Lastly, our paper contributes to the portfolio selection literature. While Markowitz (1952) derived the optimal portfolio rule in a static mean–variance setting, the implementation of that approach requires the estimation of the inputs, namely the expected returns and the variance–covariance matrix of the risky assets to be included in the portfolio. While extensive research effort has been dedicated to proposing approaches to reduce the estimation error in the inputs, DeMiguel, Garlappi, and Uppal (2009) show that the out-of-sample performance of the sample based mean–variance model, as well as its various extensions specifically designed to reduce the estimation error, is typically worse than that of the $1/N$ rule in terms of the Sharpe ratio and CEQ returns. We show that our approach robustly identifies the maximum Sharpe ratio portfolio out of sample and delivers very high CEQ returns.

The remainder of this paper is organized as follows. Section II 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 III. The empirical results are presented in Section IV. Section V concludes with suggestions for future research.

II The Method

Our relative-entropy minimizing approach enables us to recover, for a given cross-section of assets, what we refer to as the *Information SDF*. Section II.1 describes the information-theoretic method used to construct the SDF. Section II.2 discusses the econometric tests used to assess the pricing performance of the Information SDF and compare its performance to some leading empirical asset pricing models commonly used in the literature.

II.1 Recovery of the Information SDF

The absence of arbitrage opportunities implies the existence of a strictly positive pricing kernel (also known as the stochastic discount factor), 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

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

where $\bar{x} := \mathbb{E}[x_t]$, and $\frac{M_t}{M} = \frac{d\mathbb{Q}}{d\mathbb{P}}$ 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 $\mathbb{P}$.

Given the above, an estimate of the risk neutral probability measure can be obtained as the minimizer of its relative entropy with respect to the physical measure, i.e. as[3]

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

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 $\mathbb{B}$, i.e. the Kullback–Leibler Information Criterion (KLIC) divergence between $\mathbb{A}$ 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).

Ghosh, Julliard, and Taylor (2016) show that the above approach for the recovery of the pricing measure has desirable properties. First, the estimation in Equation (2) delivers a *non-parametric maximum likelihood* estimate of the risk neutral measure and the pricing kernel. Second, due to the presence of the logarithm in the objective functions in Equation (2), the use of relative entropy naturally enforces the non-negativity of the pricing kernel. 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.[4] 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 reweigh the observations to appropriately account for tail events that happened to occur in the data with a frequency

[3]Minimizing the relative entropy to recover the risk neutral probability measure was first suggested by Stutzer (1995). Ghosh, Julliard, and Taylor (2016) 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.

[4]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)).

lower than their true probability).[5]

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

$$\arg\min_{M_t} \; \mathbb{E}^{\mathbb{P}}\left[M_t \ln M_t\right] \quad \text{s.t. } \mathbb{E}^{\mathbb{P}}\left[M_t \mathbf{R}_t^e\right] = \mathbf{0}, \tag{3}$$

where, to simplify the exposition (but without loss of generality), we have used the innocuous normalization $\overline{M} = 1$.[6] 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,[7] obtaining

$$\arg\min_{\{M_t\}_{t=1}^T} \; \frac{1}{T}\sum_{t=1}^{T} M_t \ln M_t \quad \text{s.t. } \frac{1}{T}\sum_{t=1}^{T} M_t \mathbf{R}_t^e = \mathbf{0}. \tag{4}$$

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

$$\widehat{M_t} \equiv M_t\left(\widehat{\theta}_T, \mathbf{R}_t^e\right) = \frac{e^{\widehat{\theta}_T' \mathbf{R}_t^e}}{\sum_{t=1}^{T} e^{\widehat{\theta}_T' \mathbf{R}_t^e}}, \quad \forall t, \tag{5}$$

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

$$\widehat{\theta}_T := \arg\min_{\theta} \frac{1}{T}\sum_{t=1}^{T} e^{\theta' \mathbf{R}_t^e}, \tag{6}$$

and this last expression is the dual formulation of the entropy minimization problem in Equation (4). 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.[8]

We use the above method to recover the time series of the SDF in a rolling out-of-sample

[5] 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.

[6] 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.

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

[8] Note that since relative entropy is not symmetric, i.e., $D(\mathbb{Q}||\mathbb{P}) \neq D(\mathbb{P}||\mathbb{Q})$, 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.

fashion. In particular, for a given cross section of asset returns, we divide the time series of returns into rolling subsamples of length $\overline{T}$ and final date T_i, $i = 1, 2, 3, ...$, and constant $s := T_{i+1} - T_i$. In subsample i, we estimate the vector of Lagrange multipliers $\widehat{\theta}_{T_i}$ by solving the minimization in Equation (6). Using the estimates of the Lagrange multipliers, $\widehat{\theta}_{T_i}$, the out-of-sample *Information SDF* (I-SDF) $M\left(\widehat{\theta}_{T_i}, \mathbf{R}_t^e\right)$ is obtained for the subsequent s periods (i.e. for t such that $T_i + 1 \leq t \leq T_{i+1}$) using Equation (5). This process is repeated for each subsample to obtain the time series of the estimated kernel over the out-of-sample evaluation period.

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

The relative entropy minimizing pricing kernel, while being a function of asset returns, is not directly a traded asset or portfolio of assets. As a consequence, we create a mimicking portfolio, maximally correlated with the kernel, in a rolling out-of-sample fashion. We refer to this portfolio as the *Information Portfolio* (I-P). The I-P is constructed as follows. In subsample i, the estimates of the Lagrange multipliers, $\widehat{\theta}_{T_i}$, are used to construct the in-sample SDF $\widehat{M}_{i,t} \equiv M\left(\widehat{\theta}_{T_i}, \mathbf{R}_t^e\right)$, $t = T_i - \overline{T} + 1, T_i - \overline{T} + 2, ..., T_i$. Then $\widehat{M}_{i,t}$ is projected onto the space of excess 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

$$\omega_{T_i} := -\frac{\widehat{\mathbf{b}}_{T_i}}{\left|\widehat{\mathbf{b}}'_{T_i}\iota\right|}, \quad \left[\widehat{a}_{T_i}, \widehat{\mathbf{b}}'_{T_i}\right] := \underset{\left\{a_{T_i}, \mathbf{b}'_{T_i}\right\}}{\arg\min} \frac{1}{\overline{T}} \sum_{t=T_i-\overline{T}+1}^{T_i} \left(\widehat{M}_{i,t} - a_{T_i} - \mathbf{b}_{T_i}\mathbf{R}_t^e\right)^2, \tag{7}$$

where ι denotes a conformable column vector of ones. Using the portfolio weights vector, the *out-of-sample* I-P is obtained as $R_t^{IP} = \omega'_{T_i}\mathbf{R}_t^e$ for the subsequent s periods, i.e. for $t = T_i + 1, T_i + 2, ..., T_{i+1}$. This process is repeated for each subsample to obtain the time series of the information portfolio over the out-of-sample evaluation period. Note that in the scenario that the pricing kernel extracted using the relative entropy minimization approach prices assets perfectly in-sample, its projection, namely the I-P, identifies the mean–variance tangency portfolio of the test assets.

In the empirical analysis, we set $s = 12$ months (4 quarters) for monthly (quarterly) data. This corresponds to an annual rebalancing of the portfolio. The size of the rolling window, $\overline{T}$, is set to 30 years.

II.2 Asset Pricing Tests

For a given cross-section of test assets, we construct the out-of-sample I-SDF and I-P using the procedure described in Section II.1. We evaluate the empirical performance of the I-SDF and I-P at monthly and quarterly frequencies. We compare the performance of these factors to that of the one-factor CAPM, the three factor Fama–French model, and the Carhart four factor model.

We use the two-step method of Fama and MacBeth (1973) to assess the ability of each factor model to price the cross-section 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 a vector of regression slopes (that should be non-zero if the factors are priced), 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 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 $\overline{R}^2_{OLS}$). 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 and I-P since these are one-factor models, 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 $\overline{R}^2_{GLS}$) that is obtained from the cross-sectional regression of $\widehat{V}^{-1/2}\mu$ on $\widehat{V}^{-1/2}[\iota \ B]$, where $V := Var(\mathbf{R}^e)$. The $\overline{R}^2_{GLS}$ for a model, unlike $\overline{R}^2_{OLS}$, is completely determined by the model-implied factor's proximity to the minimum variance frontier and, in general, presents a more stringent hurdle for models.

Third, we present the cross-sectional T^2 statistic of Shanken (1985), given by $T^2 := \widehat{\alpha}' S_a^+ \widehat{\alpha}$, where S_a^+ is the pseudoinverse of the estimated $\Sigma_a := \left(1 + \gamma' \Sigma_F^{-1} \gamma\right) \frac{y \Sigma y}{T}$, $y := I - C\left(C'C\right)^{-1} C'$ 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 noncentrality parameter $\alpha'\Sigma_a^+\alpha = \alpha'(y\Sigma y)^+\alpha\frac{T}{\left(1+\gamma'\Sigma_F^{-1}\gamma\right)}$, 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, i.e., the vector of pricing errors $\alpha = 0$.

Fourth, we present the quadratic $q := \alpha'(y\Sigma y)^+\alpha$ which measures how far the factor is 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).

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 $\overline{R}^2_{OLS}$. 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 $\mu = hC\lambda + \varepsilon$, where $C \equiv [\iota, B]$, B denotes the vector of factor loadings in the historical sample, and $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. The constants h and σ_ε^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 plot 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 statistic in the historical data for $q = 0$.

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

III Data Description

We assess the empirical performance of the extracted pricing kernel (the I-SDF) and its tradable counterpart (the I-P) at monthly and quarterly frequencies. The out-of-sample evaluation covers the period 1963:07–2010:12. 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.

To illustrate the strength of our method, we analyse several cross-sections of equity portfolios.[10] In particular, we consider the 25 size and book-to-market-equity sorted portfolios, the 10 momentum-sorted portfolios, the 10 and 30 industry-sorted portfolios, and the 25 portfolios formed on long term reversal and size. We extract the I-SDF and the I-P from, and use them to price, each of these cross-sections, as well as several combinations of these cross-sections.

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 one-month Treasury Bill rate, also obtained from Kenneth French's data library. The quarterly returns on the equity portfolios as well as the quarterly risk free rate are obtained by compounding the monthly returns within each quarter. The excess returns on the portfolios are then computed by subtracting the risk free rate.

IV Empirical Evidence

In what follows, we evaluate the out-of-sample ability of the I-SDF and I-P to (a) explain the cross-section of returns and (b) deliver optimally diversified portfolios of the test assets. In particular, Section IV.1 presents the cross-sectional regression results for different sets of test assets, Section IV.2 presents the properties of the I-SDF and I-P, and Section IV.3 presents the performance of the I-P as an investment strategy.

IV.1 Cross-Sectional Pricing

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. Consider first Panel A, which presents the results at a monthly frequency. Row 1 shows that when the I-SDF

[10]We focus on portfolios, rather than individual asset returns, since our estimation method requires a large time series dimension relative to the cross-sectional one.

is used as the sole factor, its estimated price of risk has the correct sign and is strongly statistically significant with an absolute value of the t-statistic in excess of 7. 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 then 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 double 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 3.6%. Although this is statistically significant, part of it may be attributable to the differences in lending and borrowing rates (1%–2%). Moreover, rows 3 and 4 show that the CAPM and the FF3 model produce substantially higher annualized intercepts of 13.2%. The I-SDF produces an $\overline{R}^2_{OLS}$ of 67.0% and, more importantly, $\overline{R}^2_{GLS}$ is very similar to $\overline{R}^2_{OLS}$, at 56.6%. 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 0.077 and its 90% confidence interval includes 0, i.e., the I-SDF mimicking portfolio is statistically indistinguishable from the maximum Sharpe ratio portfolio of the test assets.

Row 2 shows that the I-P, when used as the single factor in the cross-sectional regression, produces results similar to those obtained with the I-SDF in row 1. Note that while a factor and its mimicking portfolio produce the same intercept, R^2, and pricing errors in a cross-sectional regression in-sample, the same does not hold out-of-sample. The small differences between rows 1 and 2 are because of the out-of-sample nature of the construction of the I-SDF and I-P.

In row 3, we present the results for the unconditional CAPM. The market risk premium has the wrong sign and is not statistically significantly different from zero. The intercept, on the other hand, is strongly significant with an annualized value of 13.2%. The OLS and GLS $\overline{R}^2$ are much smaller at 3.97% and 28.8%, respectively, compared to those obtained with the I-SDF and I-P. The T^2 statistic is almost double those obtained with the I-SDF and I-P, and has a p-value of zero: i.e. the model is strongly rejected. The q statistic is closely related to the $\overline{R}^2_{GLS}$ and the T^2 statistics and, therefore, not surprisingly, provides similar conclusions: the 90% confidence interval for the q statistic implies a large unexplained

Table 1: 25 Fama–French Portfolios

Row	*const.*	λ_{sdf}	λ_{IP}	λ_{Rm}	λ_{SMB}	λ_{HML}	$\overline{R}^2_{OLS}$ (%)	$\overline{R}^2_{GLS}$ (%)	T^2	q
					Panel A: Monthly					
(1)	0.003	−0.341					67.0	56.6	37.5	0.077
	(5.73)	(−7.06)					[39.5,100]	[52.4,100]	(0.207)	[0.00,0.09]
(2)	0.003		0.023				68.6	59.6	37.1	0.072
	(5.70)		(7.32)				[45.7,100]	[41.3,100]	(0.099)	[0.00,0.08]
(3)	0.011			−0.004			3.97	28.8	71.6	0.128
	(3.40)			(−1.41)			[−4.35,61.4]	[6.43,59.9]	(0.000)	[0.04,0.34]
(4)	0.011			−0.006	0.002	0.004	71.3	40.9	51.5	0.096
	(2.50)			(−1.53)	(3.86)	(6.87)	[21.1,90.9]	[20.5,90.8]	(0.002)	[0.03,0.16]
(5)	0.003	−0.383		0.002	0.002	0.004	83.8	59.0	29.5	0.063
	(0.865)	(−4.39)		(0.568)	(5.86)	(8.42)	[50.8,100]	[47.6,100]	(0.311)	[0.00,0.088]
(6)	0.004		0.025	0.0004	0.003	0.004	86.1	62.2	29.3	0.058
	(1.20)		(5.26)	(0.132)	(6.75)	(8.48)	[67.6,98.8]	[38.2,100]	(0.172)	[0.00,0.077]
					Panel B: Quarterly					
(1)	0.028	−5.46					26.8	30.8	41.3	0.332
	(11.33)	(−3.13)					[−1.22,100]	[12.3,70.5]	(0.451)	[0.00,0.60]
(2)	0.002		0.135				83.7	51.6	28.8	0.227
	(1.29)		(11.17)				[83.3,100]	[46.9,100]	(0.535)	[0.00,0.13]
(3)	0.024			−0.002			−3.92	8.50	80.9	0.431
	(2.79)			(−0.308)			[−4.35,25.9]	[−0.57,43.5]	(0.000)	[0.08,0.97]
(4)	0.028			−0.015	0.007	0.013	74.7	17.7	59.3	0.351
	(2.19)			(−1.15)	(4.95)	(7.20)	[30.3,93.1]	[−7.50,66.3]	(0.003)	[0.08,0.71]
(5)	0.002	−5.04		0.012	0.008	0.012	82.0	32.2	33.9	0.275
	(0.180)	(−3.19)		(0.884)	(6.11)	(6.83)	[44.8,100]	[13.3,100]	(0.391)	[0.00,0.21]
(6)	0.005		0.108	0.010	0.008	0.012	83.4	46.5	31.1	0.217
	(0.403)		(3.71)	(0.768)	(6.60)	(7.85)	[46.0,100]	[19.6,100]	(0.309)	[0.00,0.28]

Cross-sectional regressions of average excess returns of the 25 Fama–French portfolios on the estimated factor loadings for different asset pricing models. Panel A presents the monthly results and Panel B the quarterly results. In each panel, the first row presents the results when the factor is the information SDF and the second row, for the information portfolio. The information SDF and information portfolio are extracted from the 25 Fama–French portfolios using a relative entropy minimizing procedure, in a rolling out-of-sample fashion starting at 1963:07. Rows 3 and 4 present the results for the CAPM and the Fama–French 3-factor model, respectively. In row 5 the factors are the three FamaFrench factors plus the information SDF. In row 6 the factors are the three FamaFrench factors plus the information portfolio. For each model, the table presents the intercept and slopes, along with t-statistics in parentheses. It also presents the OLS adjusted R^2 and the GLS adjusted R^2, along with the 90% confidence intervals for the true underlying population adjusted 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 are from the mean–variance frontier.

Sharpe ratio between 0.2 and 0.58, i.e. the model fails to identify the maximum Sharpe ratio portfolio.

Row 4 presents the results for the FF 3-factor model. The results show that the market risk premium is not statistically significant but the risk premia associated with the factors proxying for risks related to size and book-to-market-equity are both significantly positive. However, the intercept is statistically and economically large, with an annualized value of 13.2%, the same as that obtained with the market risk factor alone in row 3. The $\overline{R}^2_{OLS}$ is high at 71.3%, consistent with existing empirical evidence that the 3 FF factors explain a large fraction of the time series and cross-sectional variation in the returns of the 25 FF portfolios. However, moving to a GLS cross-sectional regression, $\overline{R}^2$ drops sharply to 40.9%, 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 and I-P, which deliver very similar $\overline{R}^2$ using both the OLS and GLS procedures. The T^2 statistic is larger than those obtained with the I-SDF (51.5 vs 37.5) and I-P (51.5 vs 37.1), and has a p-value of zero, implying a statistical rejection of the model. The q statistic is also larger than those obtained with the I-SDF (0.096 vs 0.077) and I-P (0.096 vs 0.072). Moreover, the 90% confidence interval of the q statistic does not include 0, i.e. the maximum Sharpe ratio obtainable from the 3 FF factors is statistically different from the Sharpe ratio of the tangency portfolio of the test assets.

Row 5 presents the results when the I-SDF is used in conjunction with the 3 FF factors in the cross-sectional regression. Note that the risk premium for the I-SDF remains strongly statistically significant even in the presence of the 3 FF factors and its magnitude is very similar to that obtained when the I-SDF is used as the sole factor in row 1. Although $\overline{R}^2_{OLS}$ is higher, at 83.8% compared to 67.0% in row 1, the $\overline{R}^2_{GLS}$ for the two rows are very similar (59.0% vs 56.6%). Similar results are obtained in row 6 when the I-P is used in conjunction with the 3 FF factors.

Similar results are obtained at a quarterly frequency in Panel B. Both the I-SDF and I-P deliver a strongly significant λ, the T^2 statistic implies that these pricing models are not rejected, and the q statistic implies that the these factors seem to identify the maximum Sharpe ratio portfolio. Nevertheless, some of the $\overline{R}^2_{OLS}$ and $\overline{R}^2_{GLS}$ are somewhat reduced, but this reduction is not informative since the confidence intervals for this statistics include values as high as 100%. The CAPM, on the other hand, produces a negative $\overline{R}^2_{OLS}$, an $\overline{R}^2_{GLS}$ of 8.8%, and a T^2 statistic with a p-value of 0%. For the FF 3-factor, although $\overline{R}^2_{OLS}$ is high at 74.7% (but smaller than that for the I-P), the GLS $\overline{R}^2$ drops sharply, to only 17.7% (whereas that for the I-P is 51.6%). Moreover, the T^2 test rejects the FF 3-factor specification while the q statistics suggests that this factor model fails to identify the capital market line (while both the I-SDF and I-P succeed in this task). Lastly, combining the information factors with the FF 3-factor leaves both the point estimates and the statistical significance of those information factors unaffected.

Note that, as noted in Lewellen, Nagel, and Shanken (2010), it is relatively easy to find factors that produce large $\overline{R}^2_{OLS}$ for the 25 FF portfolios because of their strong factor structure. What is more impressive is that a single factor,[11] namely the I-SDF or the I-P,

[11] Of course, any multi-factor model can be rewritten as a single factor model (see e.g. Back (2010)), nevertheless this requires the knowledge of 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

does even better than the FF3 factors. Moreover, similar conclusions are obtained if, rather than relying on $\overline{R}^2_{OLS}$ alone, more stringent hurdles are imposed on the model via the $\overline{R}^2_{GLS}$, T^2, and q statistics, and their confidence bands.

We next show that the superior performance of our model holds not only for the size and book-to-market-equity sorted portfolios, but also for portfolios formed by sorting stocks on the basis of other characteristics, such as prior returns, industry, etc. Tables 2–4 present the cross-sectional regression results when the set of test assets consists of (a) the 10 momentum sorted portfolios, (b) the 25 portfolios formed on the basis of size and long-term reversal, and (c) the 10 industry portfolios and the smallest and largest deciles of portfolios formed on the basis of size, B/M, and momentum. The results, in each case, are very similar to those obtained with the 25 FF portfolios in Table 1.

Overall, Tables 2–4 show that: the I-SDF and I-P tend to produce smaller pricing errors and larger cross-sectional R^2s than the Fama–French 3-factor and the Carhart 4-factor models, despite being only a one-factor model (i.e. despite having substantially fewer degrees of freedom for fitting the data than the other models); the risk premia associated with the I-SDF and I-P are statistically significant, even after controlling for the FF and Carhart factors; the T^2 statistics of the I-SDF and I-P imply that these factors are never rejected at standard confidence levels (while the other factor models considered are almost always rejected); the q statistics imply that the I-SDF and I-P successfully identify the capital market line, i.e. they are statistically undistinguishable from the maximum Sharpe ratio portfolio (while the other factor models considered fail in this respect); in 29 cases out of 32 (or 37 out of 40 if Table 1 is included) the t-statistics of the information factors are larger than 3, hence clearing the higher hurdle for statistical significance recommended by Harvey and Liu (2015). Moreover, as an additional robustness check of the results in Tables 1 to 4, we have also run cross-sectional estimates using the Pen-FM (Penalized Fama–MacBeth) estimator of Bryzgalova (2015), that by design has the ability to detect spurious factors and shrinking (in a 'lasso' fashion) their λ's to zero. Using this approach, we found virtually identical results, for the information factors, to those discussed above.[12]

In Tables 1–4, the cross-section of assets used to extract the I-SDF and I-P coincide with the test assets that the model is then challenged to price. As described in Section II, if the I-SDF prices perfectly the cross-section in-sample, then the I-P is identical, in sample, to the mean–variance tangency portfolio of the test assets. However, this is not the case when the set of assets used to estimate the kernel differs from the set of test assets. The relative performance of the I-SDF and I-P in such a scenario is shown in Table 5. The cross-section

freedom that a model has for fitting the data.

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

Table 2: 10 Momentum Portfolios

Row	*const.*	λ_{sdf}	λ_{IP}	λ_{Rm}	λ_{SMB}	λ_{HML}	λ_{MOM}	$\overline{R}^2_{OLS}$	$\overline{R}^2_{GLS}$	T^2	q
Panel A: Monthly											
(1)	0.004 (9.99)	−0.27 (−9.16)						90.2 [66.3,100]	68.1 [16.0,100]	12.37 (0.325)	0.024 [0.00,0.05]
(2)	0.002 (7.80)		0.034 (13.66)					95.4 [55.0,100]	83.6 [74.5,100]	6.37 (0.665)	0.012 [0.00,0.018]
(3)	0.014 (2.04)			−0.009 (−1.45)				10.9 [−12.5,78.6]	−2.0 [−7.85,42.4]	40.15 (0.000)	0.074 [0.03,0.43]
(4)	0.022 (1.27)			−0.013 (−0.75)	−0.011 (−0.61)	−0.032 (−1.18)	0.007 (6.02)	78.9 [−78.2,100]	2.59 [−25.5,92.2]	8.81 (0.386)	0.044 [0.00,0.34]
(5)	0.005 (0.59)	−0.284 (−3.05)		0.002 (0.28)	−0.008 (−1.03)	−0.023 (−1.90)	0.006 (10.48)	95.9 [−62.0,100]	69.6 [−38.3,100]	3.25 (0.640)	0.011 [0.00,0.318]
(6)	0.003 (0.42)		0.039 (4.36)	0.004 (0.60)	−0.005 (−0.80)	−0.015 (−1.60)	0.006 (14.64)	97.7 [14.5,100]	82.1 [−30.9,100]	2.61 (0.605)	0.006 [0.00,0.30]
Panel B: Quarterly											
(1)	0.008 (6.39)	−1.07 (−7.93)						87.3 [16.8,100]	75.2 [63.2,100]	8.12 (0.529)	0.056 [0.00,0.19]
(2)	0.006 (5.60)		0.107 (9.90)					95.4 [73.0,100]	78.6 [70.4,100]	7.36 (0.589)	0.047 [0.00,0.07]
(3)	0.038 (2.40)			−0.024 (−1.59)				14.4 [−12.5,80.9]	−6.49 [−6.8,31.2]	39.55 (0.000)	0.226 [0.05,1.15]
(4)	0.061 (1.08)			−0.050 (−0.83)	0.038 (1.03)	0.021 (0.63)	0.022 (5.46)	75.4 [−72.8,94.6]	1.56 [−23.6,77.3]	9.55 (0.188)	0.131 [0.00,0.96]
(5)	0.028 (1.88)	−1.13 (−5.10)		−0.18 (−1.13)	0.035 (3.72)	0.011 (1.31)	0.023 (22.58)	98.4 [41.5,100]	85.8 [12.7,100]	1.21 (0.849)	0.015 [0.00,0.194]
(6)	0.036 (2.43)		0.084 (3.69)	−0.024 (−1.54)	0.032 (3.28)	0.0005 (0.05)	0.021 (20.80)	98.4 [84.3,100]	86.7 [39.7,100]	1.35 (0.826)	0.014 [0.00,0.14]

Cross-sectional regressions of average excess returns of the 10 momentum-sorted portfolios on the estimated factor loadings for different asset pricing models. Panel A presents the monthly results and Panel B the quarterly results. The first two rows in each panel present the results when the factor is the information SDF (row 1) and the information portfolio (row 2). The information SDF and information portfolio are extracted from the 10 momentum-sorted portfolios using a relative entropy minimizing procedure, in a rolling out-of-sample fashion starting at 1963:07. Rows 3 and 4 present the results for the CAPM and the Carhart 4-factor model, respectively. In row 5 the factors are the four Carhart factors plus the information SDF. In row 6 the factors are the four Carhart factors plus the information portfolio. For each model, the table presents the intercept and slopes, along with t-statistics in parentheses. It also presents the OLS adjusted R^2 and the GLS adjusted R^2, along with the 90% confidence intervals for the true underlying population adjusted 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 are from the mean–variance frontier.

Table 3: 25 Portfolios Formed on Long-Term Reversal and Size

Row	*const.*	λ_{sdf}	λ_{IP}	λ_{Rm}	λ_{SMB}	λ_{HML}	$\overline{R}^2_{OLS}$ (%)	$\overline{R}^2_{GLS}$ (%)	T^2	q
Panel A: Monthly										
(1)	0.006	−0.22					13.5	61.5	25.07	0.047
	(7.06)	(−2.18)					[−4.35,100]	[47.1,90.2]	(0.603)	[0.00,0.03]
(2)	0.002		0.024				72.3	68.0	18.35	0.037
	(2.13)		(7.98)				[48.9,100]	[66.4,100]	(0.830)	[0.00,0.01]
(3)	0.005			0.002			−1.6	10.1	58.14	0.103
	(1.38)			(0.78)			[−4.35,37.4]	[0.15,33.1]	(0.001)	[0.03,0.18]
(4)	0.002			0.002	0.001	0.007	74.3	26.1	40.37	0.077
	(0.68)			(0.93)	(1.62)	(4.96)	[34.9,100]	[1.66,100]	(0.019)	[0.01,0.10]
(5)	−0.001	−0.370		0.005	0.004	0.003	86.0	71.1	13.93	0.029
	(−0.291)	(−5.24)		(2.40)	(4.77)	(2.35)	[71.2,100]	[92.3,100]	(0.862)	[0.00,0.005]
(6)	−0.002		0.022	0.006	0.003	0.004	84.5	66.4	16.69	0.033
	(−0.86)		(5.50)	(2.77)	(3.99)	(3.39)	[59.2,100]	[78.5,100]	(0.723)	[0.00,0.02]
Panel B: Quarterly										
(1)	0.023	−0.33					−3.94	27.2	54.50	0.291
	(11.80)	(−0.300)					[−4.35,36.3]	[21.1,45.3]	(0.353)	[0.00,0.44]
(2)	0.008		0.075				56.8	56.5	23.96	0.167
	(3.13)		(5.71)				[23.8,100]	[48.7,100]	(0.732)	[0.00,0.07]
(3)	0.008			0.013			8.7	1.33	68.46	0.372
	(1.04)			(1.81)			[−4.35,70.8]	[−2.26,31.0]	(0.002)	[0.09,0.68]
(4)	0.006			0.009	0.005	0.020	77.8	11.96	48.86	0.301
	(0.651)			(1.03)	(2.59)	(4.76)	[18.9,100]	[−7.61,84.3]	(0.018)	[0.04,0.49]
(5)	0.002	−2.75		0.014	0.009	0.012	84.2	32.8	32.58	0.219
	(0.133)	(−3.35)		(1.87)	(4.32)	(2.93)	[58.0,100]	[−4.38,100]	(0.292)	[0.00,0.257]
(6)	0.002		0.070	0.012	0.011	0.007	86.7	53.0	22.61	0.153
	(0.329)		(5.28)	(1.80)	(5.13)	(1.47)	[56.8,100]	[63.1,100]	(0.531)	[0.00,0.082]

Cross-sectional regressions of average excess returns of the 25 long term reversal and size sorted portfolios on the estimated factor loadings for different asset pricing models. Panel A presents the monthly results and Panel B the quarterly results. The first two rows in each panel present the results when the factor is the information SDF (row 1) and the information portfolio (row 2). The information SDF and the information portfolio are extracted from the 25 long term reversal and size sorted portfolios using a relative entropy minimizing procedure, in a rolling out-of-sample fashion starting at 1963:07. Row 3 presents the results for the CAPM, and row 4, for the Fama–French 3-factor model. In row 5 the factors are the three FamaFrench factors plus the information SDF. In row 6 the factors are the three FamaFrench factors plus the information portfolio. For each model, the table presents the intercept and slopes, along with t-statistics in parentheses. It also presents the OLS adjusted R^2 and the GLS adjusted R^2, along with the 90% confidence intervals for the true underlying population adjusted R^2 in square brackets below. 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 are from the mean–variance frontier.

Table 4: Small, Large, Growth, Value, Winners, Losers, 10 Industry

Row	*const.*	λ_{sdf}	λ_{IP}	λ_{Rm}	λ_{SMB}	λ_{HML}	λ_{MOM}	$\overline{R}^2_{OLS}$	$\overline{R}^2_{GLS}$	T^2	q
Panel A: Monthly											
(1)	0.001 (2.49)	−1.12 (−16.7)						94.9 [99.9,100]	88.0 [86.2,100]	5.39 (0.984)	0.013 [0.00,0.013]
(2)	0.003 (7.45)		0.027 (8.62)					83.0 [74.3,100]	84.2 [82.0,100]	9.89 (0.872)	0.019 [0.00,0.009]
(3)	0.007 (2.02)			−0.002 (−0.70)				−3.5 [−7.14,57.1]	−1.17 [−3.75,45.0]	62.76 (0.000)	0.112 [.038,.413]
(4)	0.003 (1.25)			0.002 (1.09)	0.002 (2.42)	0.002 (1.92)	0.009 (8.78)	84.8 [11.4,100]	40.3 [−19.2,91.7]	27.23 (0.018)	0.052 [0.005,0.262]
(5)	−0.002 (−1.12)	−1.49 (−5.67)		0.006 (4.30)	0.003 (5.92)	0.0004 (0.872)	0.008 (10.65)	95.1 [94.0,100]	88.6 [81.6,100]	3.03 (0.955)	0.009 [0.00,0.014]
(6)	−0.003 (−1.58)		0.035 (8.10)	0.007 (4.45)	0.003 (5.44)	0.0003 (0.51)	0.008 (10.65)	94.7 [56.5,100]	84.8 [81.7,100]	5.73 (0.891)	0.012 [0.00,0.013]
Panel B: Quarterly											
(1)	0.014 (12.44)	−3.88 (−6.92)						75.8 [56.1,100]	60.9 [28.6,100]	17.15 (0.555)	0.145 [0.00,0.554]
(2)	0.009 (5.81)		0.100 (6.61)					74.0 [31.4,100]	79.9 [68.1,100]	10.85 (0.828)	0.077 [0.00,0.181]
(3)	0.021 (2.10)			−0.005 (−0.52)				−5.1 [−7.14,61.4]	0.73 [0.83,65.9]	67.17 (0.000)	0.358 [0.093,1.345]
(4)	0.012 (1.47)			0.004 (0.45)	0.006 (2.50)	0.005 (2.01)	0.027 (7.97)	82.9 [25.0,100]	37.3 [−19.9,84.9]	27.86 (0.026)	0.178 [0.015,1.323]
(5)	0.009 (1.10)	−2.37 (−2.54)		0.006 (0.82)	0.007 (2.94)	0.005 (2.12)	0.026 (7.94)	84.4 [41.5,100]	50.1 [−20.5,100]	18.7 (0.118)	0.129 [0.00,0.520]
(6)	−0.002 (−0.26)		0.107 (4.62)	0.017 (2.16)	0.008 (3.89)	0.002 (1.10)	0.024 (8.43)	89.6 [49.0,100]	75.8 [37.7,100]	8.27 (0.701)	0.062 [0.00,0.258]

Cross-sectional regressions of average excess returns of the 10 industry portfolios and the top and bottom deciles of portfolios sorted on the basis of size, book-to-market-equity, and momentum on the estimated factor loadings for different asset pricing models. Panel A presents the monthly results and Panel B the quarterly results. In each panel, row 1 presents the results when the factor is the information SDF, and row 2, when the factor is the information portfolio. The information SDF and the information portfolio are extracted from the 10 industry portfolios and the top and bottom deciles of portfolios sorted on the basis of size, book-to-market-equity, and momentum, using a relative entropy minimizing procedure, in a rolling out-of-sample fashion starting at 1963:07. Rows 3 and 4 present the results for the CAPM and the Carhart 4-factor model, respectively. In row 5 the factors are the three FamaFrench factors plus the information SDF. In row 6 the factors are the three FamaFrench factors plus the information portfolio. For each model, the table presents the intercept and slopes, along with the t-statistics in parentheses. It also presents the OLS adjusted R^2 and the GLS adjusted R^2, along with the 90% confidence intervals for the true underlying population adjusted 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.

Table 5: I-SDF and I-P Extracted from Small, Big, Growth, Value, Winner, Loser Portfolios plus 10 Industry Portfolios

Row	Assets	*const.*	λ_{sdf}	λ_{IP}	$\overline{R}^2_{OLS}$ (%)	$\overline{R}^2_{GLS}$ (%)	T^2	q
			Panel A: Monthly					
(1)	25 ME & Mom	.0014	−1.13		83.6	19.2	65.5	.161
		(2.84)	(−11.10)		[66.6,100]	[−1.22,50.6]	(0.028)	[.021,.209]
(2)	25 ME & Mom	.0037		.031	82.2	24.2	76.1	.151
		(9.79)		(10.58)	[68.7,100]	[5.64,59.2]	(0.000)	[.042,.221]
(3)	25 FF + 30 Ind	.0013	−1.03		35.2	31.2	100.4	.235
		(1.45)	(−5.51)		[23.6,100]	[16.8,100]	(0.238)	[.000,.282]
(4)	25 FF + 30 Ind	.0020		.038	45.2	33.1	107.6	.226
		(3.34)		(6.75)	[31.7,100]	[29.4,100]	(0.058)	[.000,.267]
(5)	25 FF + 30 Ind + 10 Mom	.0016	−.94		53.4	28.5	129.5	.291
		(3.06)	(−8.62)		[45.1,100]	[22.8,100]	(0.147)	[.000,.341]
(6)	25 FF + 30 Ind + 10 Mom	.0030		.028	58.6	28.8	147.1	.288
		(9.00)		(9.57)	[53.3,100]	[12.7,77.1]	(0.002)	[.032,.300]
			Panel B: Quarterly					
(1)	25 ME & Mom	.018	−5.29		85.2	28.5	37.7	.326
		(19.70)	(−11.78)		[78.1,100]	[8.54,100]	(0.371)	[.000,.598]
(2)	25 ME & Mom	.018		.108	75.0	18.4	51.7	.373
		(15.51)		(8.54)	[52.0,100]	[3.11,45.3]	(0.069)	[.000,.318]
(3)	25 FF + 30 Ind	.014	−4.11		24.2	27.6	106.3	.776
		(10.08)	(−4.27)		[3.21,100]	[7.36,92.4]	(0.508)	[.000,.680]
(4)	25 FF + 30 Ind	.013		.107	35.8	25.8	110.4	.795
		(9.10)		(5.58)	[9.32,100]	[10.0,86.9]	(0.526)	[.000,.613]
(5)	25 FF + 30 Ind + 10 Mom	.014	−4.51		45.9	24.8	138.0	1.07
		(14.88)	(−7.44)		[37.0,100]	[6.40,100]	(0.512)	[.000,.875]
(6)	25 FF + 30 Ind + 10 Mom	.014		.099	53.0	25.6	153.0	1.06
		(14.97)		(8.56)	[48.2,100]	[3.31,90.5]	(0.449)	[.000,.431]

Cross-sectional regressions of average excess returns listed in column 2 on the estimated factor loadings for the information SDF (odd rows) and portfolio (even rows). The information SDF and information portfolio are extracted from only a subset of the portfolios (the Small, Big, Growth, Value, Winners and Losers portfolios plus the 10 industry portfolios) in a rolling out-of-sample fashion starting at 1963:07. Panel A presents the monthly results and Panel B the quarterly results. For each model, the table presents the intercept and slopes, along with t-statistics in parentheses. It also presents the OLS adjusted R^2 and the GLS adjusted R^2, along with the 90% confidence intervals for the true underlying population adjusted 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.

used to estimate the I-SDF and I-P is the same as that in Table 4, i.e. the 10 industry portfolios and the smallest and largest deciles of portfolios formed on the basis of size, B/M, and momentum. The set of test assets consist of a larger set formed by finer sortings of stocks into portfolios on the basis of the same characteristics, namely, industry, size, B/M, and momentum.

Consider first Panel A of Table 5, which presents the results at a monthly frequency. In rows 1–2, the test assets consist of 25 portfolios formed on the basis of size and momentum. The I-SDF delivers a substantially smaller intercept of 0.14% (row 1) compared to 0.37% (row 2) obtained with the I-P. Although the former intercept is statistically significant, its magnitude can, in principle, be fully explained by differences between lending and borrowing rates. The intercept obtained with the I-P, on the other hand, is economically large and is too big to be explained by differences between lending and borrowing rates. Similar results are obtained in rows 3–4, when the set of test assets consists of the 25 size and B/M sorted portfolios and the 30 industry-sorted portfolios: the I-SDF delivers an intercept of 0.13% that is not statistically different from zero, whereas the I-P produces a larger and highly statistically significant intercept of 0.2%. Lastly, rows 5–6, where the test assets consist of the combination of the 25 size and B/M sorted portfolios, the 30 industry-sorted portfolios, and the 10 momentum-sorted portfolios, produce, once again, similar results: the I-SDF and the I-P produce estimated intercepts of 0.16% and 0.3%, respectively. The results are less stark at the quarterly frequency: the estimated intercepts are similar for the I-SDF and the I-P, although the former produces smaller pricing errors, as indicated by the values of the T^2 and q statistics, than the latter for all three sets of test assets.

The above results suggest that the I-SDF accurately identifies the underlying sources of priced risk. When the cross-section of assets used to extract the I-SDF and, therefore, the I-P is the same as the set of assets used in the cross-sectional tests, the I-SDF and I-P deliver similar cross-sectional fits. However, when the two sets of assets differ, the I-SDF delivers a better fit than the I-P, at least at a monthly frequency, and smaller pricing errors. This is because the I-SDF provides an estimate of the underlying kernel or sources of systematic risk while its projection, the I-P, isolates the component of the kernel most relevant for pricing that particular set of assets. This difference in performance suggests that the non-linearity of the I-SDF in asset returns (see Equation 5) is actually informative, and that part of this information is lost when working with the linear I-P.

IV.2 Properties of the Information SDF

We now show that the I-SDF and the I-P contain novel pricing information not captured by standard multifactor asset pricing models, such as the FF 3-factor and the Carhart 4-factor models. Table 6 presents the time series regressions of the I-SDF and I-P, constructed from each set of test assets in Tables 1–4 (and indicated in the second column), on the FF3 factors. Whenever the assets used to construct the information SDF and portfolio include momentum-sorted portfolios, we also include the momentum factor as a regressor in addition to the FF3 factors. If the factors fully explain the variation in the I-SDF and I-P, the intercepts from the time series regressions should be indistinguishable from zero and the R^2 of the regressions should be high.

Panel A presents the results at a monthly frequency. In rows 1–2, the 25 size and book-to-market-equity sorted portfolios are used to extract the kernel and its mimicking portfolio. Row 1 shows that the 3 FF factors explain only 17.6% of the variation in the I-SDF. Moreover,

Table 6: Explaining I-SDF and I-P with FF3 and Momentum Factors

Row	Assets	α_{sdf}	α_{IP} (%)	β_{Rm}	β_{SMB}	β_{HML}	β_{MOM}	$\overline{R}^2_{OLS}$ (%)	Info-Ratio
				Panel A: Monthly					
(1)	FF25	1.17 (35.5)		−4.65 (−6.01)	−1.36 (−1.26)	−12.25 (−10.55)		17.6	
(2)	FF25		1.31 (4.89)	0.47 (7.44)	0.21 (2.41)	1.30 (13.82)		26.9	0.211
(3)	10 Momentum	1.19 (31.8)		−5.00 (−5.70)	1.94 (1.61)	−3.88 (−2.93)	−12.06 (−14.23)	27.1	
(4)	10 Momentum		1.37 (3.06)	0.94 (9.03)	−0.30 (−2.09)	0.26 (1.67)	1.67 (16.52)	36.1	0.135
(5)	25 Long-Term Reversal & Size	1.16 (28.3)		−3.89 (−4.03)	−1.49 (−1.11)	−9.23 (−6.36)		7.34	
(6)	25 Long-Term Reversal & Size		0.69 (2.79)	0.40 (6.92)	0.35 (4.39)	0.92 (10.56)		20.7	0.121
(7)	S, B, G, V, W, L, 10 Industry	1.33 (18.2)		−8.73 (−5.09)	−6.52 (−2.77)	−13.2 (−5.10)	−15.1 (−9.09)	16.1	
(8)	S, B, G, V, W, L, 10 Industry		0.93 (3.28)	0.59 (9.00)	0.35 (3.90)	1.28 (12.87)	1.25 (19.70)	46.6	0.145
				Panel B: Quarterly					
(1)	FF25	2.69 (4.57)		6.61 (0.86)	−0.22 (−0.02)	−6.93 (−0.70)		−0.50	
(2)	FF25		5.49 (4.01)	0.59 (3.29)	0.43 (1.72)	1.12 (4.85)		15.4	0.310
(3)	10 Momentum	1.69 (12.36)		−8.11 (−4.80)	4.82 (2.02)	−3.26 (−1.46)	−12.37 (−7.57)	27.5	
(4)	10 Momentum		2.63 (1.77)	1.20 (6.52)	−0.21 (−0.82)	0.43 (1.76)	1.92 (10.76)	42.6	0.147
(5)	25 Long-Term Reversal & Size	2.39 (4.61)		−0.365 (−0.054)	9.67 (1.01)	−14.3 (−1.62)		0.83	
(6)	25 Long-Term Reversal & Size		2.39 (2.57)	0.50 (4.16)	0.01 (0.06)	0.93 (5.94)		18.0	0.199
(7)	S, B, G, V, W, L, 10 Industry	3.22 (8.66)		−6.47 (−1.40)	−17.63 (−2.71)	−19.71 (−3.24)	−29.81 (−6.70)	19.7	
(8)	S, B, G, V, W, L, 10 Industry		3.21 (2.77)	0.67 (4.70)	0.29 (1.42)	1.16 (6.15)	1.23 (8.89)	33.6	0.230

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 information SDF (odd rows) and portfolio (even rows) on the Fama–French and Carhart factors. Each row presents the results when the information SDF and portfolio are constructed using the cross-section of assets listed in column 2. Since the α_{IP} is presented in percentage terms, α_{sdf} is presented as the intercept multiplied by 100 for the sake of comparability. Note that, since the I-P weights in equation (7) 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. Panels A and B present the results at monthly and quarterly frequencies.

the estimated intercept is strongly statistically significant, with an annualized value of 14.0%. Note that since the I-SDF is not a tradeable factor, the intercept is not interpretable as an alpha. Row 2 shows that the FF factors can explain a larger fraction of the variation in the I-P than does the I-SDF (26.9% versus 17.6%). However, even in this case, about three-quarters 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 15.7% per annum. These results, together with the observation that the I-SDF and I-P perform substantially better at pricing the cross-section of the 25 size and B/M sorted portfolios (Table 1), suggest that the FF factors do not fully capture the sources of priced risk even for the size and book-to-market portfolios.

Similar results are obtained for the 10 momentum sorted portfolios (rows 3–4), the 25 portfolios formed on the basis of size and long term reversal (rows 5–6), and the smallest and largest deciles of portfolios formed on the basis of size, B/M, and momentum and the 10 industry-sorted portfolios (rows 7–8). The $\overline{R}^2_{OLS}$ from the I-SDF regressions vary from 7.3% (for the size and long-term reversal sorted portfolios) to 27.1% (for the 10 momentum-sorted portfolios), showing that a substantial proportion of the variability in the I-SDFs cannot be explained by the movements in the FF3 and momentum factors. The corresponding $\overline{R}^2_{OLS}$ from the I-P regressions are higher, varying from 20.7%–46.6%, but still a substantial fraction of the variability is left unexplained by the standard multifactor models. The estimated annualized intercepts are all statistically significant and economically large, varying from 8.3%–16.4% for the I-SDF and from 13.9%–16.0% for the I-P.

The last column for each I-P regression presents the so-called Information Ratio, defined as the estimated alpha divided by the standard deviation of the residual from each regression. The Information Ratio, therefore, measures the Sharpe ratio of a hedged strategy that has an alpha equal to the estimated alpha and that has no systematic risk with respect to the FF3 or momentum factors (i.e., its beta with respect to each of these factors is zero). The results reveal that the Information Ratios are economically large, varying from 0.42–0.73 per annum. That is, a portfolio strategy that is long the I-P but perfectly hedged with respect to the market, size, book-to-market and momentum risk factors would deliver an annual return of 13.9%–16.0% and an annual Sharpe ratio of 0.42–0.73. Moreover, note that such a portfolio strategy would require rebalancing only once per year. This is remarkable if compared to the annualized excess return (less than 5%) and Sharpe ratio (about 0.31) on the U.S. stock market during the same period. This suggests that, as discussed extensively in the next section, the information SDF and portfolio are not only useful for pricing assets, but also as an asset allocation approach. As a robustness check (not reported), we also added as regressors the profitability and investment factors of Fama and French (2015), obtaining very similar results, in terms of intercepts and measures of fit, to the one reported in the table.[13]

The results obtained at the quarterly frequency in Panel B are largely similar. In fact, the FF3 or 4 factors explain an even smaller fraction of the variability of the I-SDF at a quarterly frequency compared to that at a monthly frequency. For two out of the four sets of test assets, $\overline{R}^2_{OLS}$ is less than 1%, and the estimated intercepts are statistically significant in all four cases. For the I-P regressions, $\overline{R}^2_{OLS}$ is lower at the quarterly frequency in three out of the four sets of test assets. The estimated α's are all statistically significant (with the exception of the momentum-sorted portfolios) and economically large, varying from 9.6% to 22.0% (annualized) and the Information Ratios are also economically large, varying from 0.29 to 0.62 (annualized).

IV.3 An Asset-Allocation Perspective

The previous results show that both the I-SDF and I-P offer a good one-factor benchmark model for pricing broad cross-sections of equity portfolios. Moreover, our cross-sectional asset pricing tests suggest that (as one should expect from a good pricing model) I-SDF and I-P identify correctly the capital market line, i.e. the maximum Sharpe ratio portfolio. As a consequence, since the information portfolio is easily tradable, we next investigate the implications of the I-P for strategic asset allocation.

Assuming that the investors' utility functions depend only on the mean and variance of a portfolio's return, Markowitz (1952) derived the optimal rule for allocating wealth across a set of risky assets. However, the practical implementation of that approach requires estimating the expected returns and the variance–covariance matrix of the assets. For instance, with $N = 25$ risky assets, estimating these moments via their sample analogues requires the estimation of $N + \frac{N(N+1)}{2} = 350$ parameters. Not surprisingly, these optimal portfolios often have extreme weights on constituent assets that fluctuate substantially over time, and perform poorly out-of-sample. Given the widespread use of the mean–variance approach to asset allocation among both academics and practitioners, substantial research effort has been devoted to trying to reduce the estimation error and improving the performance of the model. DeMiguel, Garlappi, and Uppal (2009) evaluate the out-of-sample performance of the sample based mean–variance approach, as well as a broad set of its extensions designed to reduce the effect of estimation error, using several different sets of test assets. Using several performance evaluation measures, they conclude that optimally diversified portfolios constructed using these approaches typically underperform a naïve diversification strategy

[13] Intercepts and $\overline{R}^2_{OLS}$ are virtually identical when the dependent variable is the I-SDF, while for the I-P the αs are somehow reduced and the $\overline{R}^2_{OLS}$ is minimally increased.

consisting of an equally-weighted ($1/N$) portfolio of the test assets.

We evaluate the out-of-sample performance of the I-P using the same performance measures as in DeMiguel, Garlappi, and Uppal (2009), namely (i) the Sharpe ratio and (ii) the certainty-equivalent (CEQ) return for the expected utility of a mean–variance investor. The Sharpe ratio is defined as $\widehat{SR}_{I\text{-}P} = \widehat{\mu}_{I\text{-}P}/\widehat{\sigma}_{I\text{-}P}$, where $\widehat{\mu}_{I\text{-}P}$ and $\widehat{\sigma}_{I\text{-}P}$ are the sample mean and sample standard deviation, respectively, of the out-of-sample excess returns on the I-P. The CEQ return is defined as the risk free rate that would make an investor with mean–variance preferences and coefficient of risk aversion $\gamma = 1$ indifferent between the risky I-P and the risk free rate: $\widehat{CEQ}_{I\text{-}P} = \widehat{\mu}_{I\text{-}P} - \frac{\gamma}{2}\widehat{\sigma}^2_{I\text{-}P}$.

For each cross-section of assets used to construct the information portfolio, we compute the Sharpe Ratio, the CEQ return, and the first four moments of the I-P. The results are presented in Table 7. Panels A and B use monthly and quarterly frequencies. As a benchmark to facilitate comparison, we also compute the corresponding statistics for the $1/N$ portfolio of the test assets. In addition to the equally-weighted portfolio, we also compare the performance of the I-P to other standard benchmarks, including: the market portfolio (row 2), the value and size portfolios (HML and SMB in rows 3 and 4 respectively) of Fama and French (1993) that are meant to exploit the value and size premia; the momentum portfolio (row 6) of Carhart (1997); and the combined value and momentum portfolio that is meant to exploit the negative correlation between value and momentum strategies (see Asness, Moskowitz, and Pedersen (2013)).

Consider first row 1 of Panel A, where the I-P is constructed from the 25 size and book-to-market sorted portfolios. Its 2.1% monthly (23.3% annual) return is about three times that of the $1/N$ portfolio (presented in parenthesis below), about 5 times that of the market and HML (rows 2 and 3) portfolios, 7 times that of the SMB portfolio (row 4), about three times that of the momentum portfolio (row 6), and about three and a half times that of the value and mometum strategies combined (row 9). These very high returns are obtained with a volatility that is only about two-thirds larger than that of the market and momentum portfolios.

Moreover, the I-P monthly Sharpe ratio is 0.288 (about 1.0 annualized), while the Sharpe ratio of the corresponding $1/N$ benchmark (presented in parentheses below) is only 0.128 monthly (or 0.44 annualized), i.e. less than one-half that of the I-P. The I-P's Sharpe ratio not only outperforms the $1/N$ benchmark, but also the market portfolio, by a factor of more than three, the HML portfolio (row 3) by a factor of more than two, the SMB portfolio by a factor of 3.5, the momentum factor by a factor of 1.75, and the combined Value and Momentum portfolio (that has an annualized SR of about 0.8) by a factor of 1.25. Note also that this last comparison might seem unfair to the information portfolio since, in row

Table 7: Summary Statistics of Information Portfolio & Returns

Row	Assets	Mean	Volatility	Sharpe Ratio	Skewness	Kurtosis	CEQ
	Panel A: Monthly						
(1)	R^{IP}	0.021	0.073	0.288	0.384	5.541	0.018
	(FF25)	(0.007)	(0.051)	(0.128)	(−0.575)	(5.589)	(0.006)
(2)	Market - Risk Free	0.004	0.045	0.091	−0.567	5.028	0.003
(3)	HML	0.004	0.029	0.139	−0.034	5.440	0.004
(4)	SMB	0.003	0.032	0.083	0.527	8.452	0.002
(5)	R^{IP}	0.030	0.127	0.235	−0.352	8.022	0.022
	(10 Momentum)	(0.004)	(0.048)	(0.085)	(−0.326)	(4.793)	(0.003)
(6)	Momentum Portfolio	0.007	0.044	0.164	−1.419	13.65	0.006
(7)	R^{IP}	0.013	0.064	0.206	−0.212	5.111	0.011
	(25 Long-Term Reversal & Size)	(0.007)	(0.051)	(0.137)	(−0.444)	(5.865)	(0.006)
(8)	R^{IP}	0.027	0.088	0.306	−0.679	6.180	0.023
	(S, B, G, V, W, L, 10 Industry)	(0.005)	(0.046)	(0.106)	(−0.490)	(4.953)	(0.004)
(9)	HML & Momentum	0.006	0.024	0.231	−0.961	10.59	0.006
	Panel B: Quarterly						
(1)	R^{IP}	0.080	0.194	0.413	0.410	3.955	0.061
	(FF25)	(0.021)	(0.103)	(0.207)	(−0.183)	(3.576)	(0.016)
(2)	Market - Risk Free	0.013	0.087	0.150	−0.435	3.635	0.009
(3)	HML	0.012	0.060	0.204	0.109	4.754	0.010
(4)	SMB	0.009	0.059	0.146	0.299	2.602	0.002
(5)	R^{IP}	0.085	0.239	0.354	−0.090	5.295	0.056
	(10 Momentum)	(0.013)	(0.093)	(0.143)	(−0.231)	(3.805)	(0.009)
(6)	Momentum Factor	0.020	0.081	0.254	−1.411	10.13	0.017
(7)	R^{IP}	0.042	0.134	0.313	−0.168	3.833	0.033
	(25 Long-Term Reversal & Size)	(0.023)	(0.104)	(0.220)	(−0.057)	(3.865)	(0.018)
(8)	R^{IP}	0.083	0.173	0.480	0.181	3.463	0.068
	(S, B, G, V, W, L, 10 Industry)	(0.016)	(0.090)	(0.175)	(−0.315)	(3.794)	(0.012)
(9)	HML & Momentum	0.019	0.042	0.443	−0.070	5.350	0.018

Mean, volatility, Sharpe ratio, skewness, kurtosis, and CEQ statistic for the portfolios listed in column 2: the information portfolio, R^{IP}, constructed from various cross-sections of assets (listed in parentheses) with the corresponding statistics for an equally-weighted portfolio of the underlying assets presented in parentheses below; the market minus the risk free rate portfolio; the value portfolio (HML); the size portfolio (SMB); the momentum portfolio; and the value and momentum portfolio. Panels A and B present the results at monthly and quarterly frequencies.

1, it is constructed without using the momentum sorted portfolios, and hence the high SR achievable by exploiting jointly the value and momentum anomalies. Indeed, when we allow the I-P to exploit these features of the data (in row 8), its SR becomes one-third higher than what is achievable by combining the value and moment strategies (in row 9).

Note that the very high returns and Sharpe ratio of the I-P in row 1 do not seem to be a compensation for negative skewness and tail risk: the I-P's skewness is positive (about 0.384), while that of the market, HML and momentum portfolios (both individually and combined) are all negative (and very large for momentum based strategies), and its kurtosis is similar to that of the market and HML portfolios, and much smaller than those of the momentum, HML plus momentum, and SMB strategies.

Similar conclusions are obtained using the CEQ return as the measure of performance. A mean–variance investor with $\gamma = 1$ would need an annualized risk free rate of 22.0% (or about 1.8% monthly) in order to not invest in the I-P, whereas a risk free rate of only 7.2% (or about 0.6% monthly) is required for such an agent to not invest in the $1/N$ portfolio. Similarly, annual (monthly) risk free rates of only 3.6% (0.3%), 4.8% (0.4%), 7.2% (0.6%), and 7.1% (0.6%), respectively, are required in order to be indifferent between the risk free rate and the market, the HML, the momentum, and the HML plus momentum portfolios.

To show that the performance of the I-P in row 1 is not driven by just a subset of the data, panel A of Figure 1 plots the path of $1 invested in the I-P over the entire out-of-sample evaluation period. Note that because we use excess returns in the construction of the I-P, this corresponds to a long–short strategy that is short $1 in the risk free rate and uses the proceeds to invest in the optimal portfolio of risky assets. For comparison, and since the plotted I-P is constructed using the FF25 portfolios (hence it might exploit the size and value premia), we also plot the path of $1 invested in the HML and SMB portfolios, as well as the excess return on the market and the equally weighted portfolios. Note also that the graph is in log scale, so that the slopes of the various lines are directly comparable across the various strategies at each point in time. As is evident from the figure, the I-P outperforms, by a wide margin, each of the benchmarks. Moreover, the I-P outperformance is robust across sub-periods: the average slope of the I-P line is higher in virtually all the 10-year sub-periods. For robustness, Panel B of Figure 1 presents the same cumulated returns as Panel A but with the benchmark portfolios leveraged in order to have the same volatility as the Information Portfolio. The figure shows that only for a very brief period at the end of the 60s did the leveraged SMB and $1/N$ outperform the I-P, and that only in the late 70s did the leveraged HML have a similar performance as the I-P. In all other periods, I-P clearly outperforms the various benchmarks. Moreover, the I-P tends to have less severe contractions in returns than the other portfolios during, and following, market-wide crashes (vertical dot-dashed lines in the figure).[14]

The I-P in row 1 of Table 7 and Figure 1 is an optimally weighted portfolio of the 25 size and book-to-market sorted portfolios. Therefore, the question arises as to whether our approach relies on extreme weights on the constituent portfolios that also fluctuate wildly over time. Figure 2a plots the time series of weights on each of the 25 portfolios in the I-P. The figure makes clear that the vast majority of the weights lie in the $[-2, 2]$ interval and, therefore, are not extreme. Moreover, these weights evolve smoothly, implying that the I-P has low turnover and, therefore, low trading costs (note also that the rebalancing is done once a year, in June).

In order to provide more intuition regarding the composition of the I-P in row 1 of

[14]We follow Mishkin and White (2002) and identify a stock market crash as a period in which either the Dow Jones Industrial, the S&P500, or the NASDAQ index drops by at least 20 percent in a time window of either one day, five days, one month, three months, or one year.

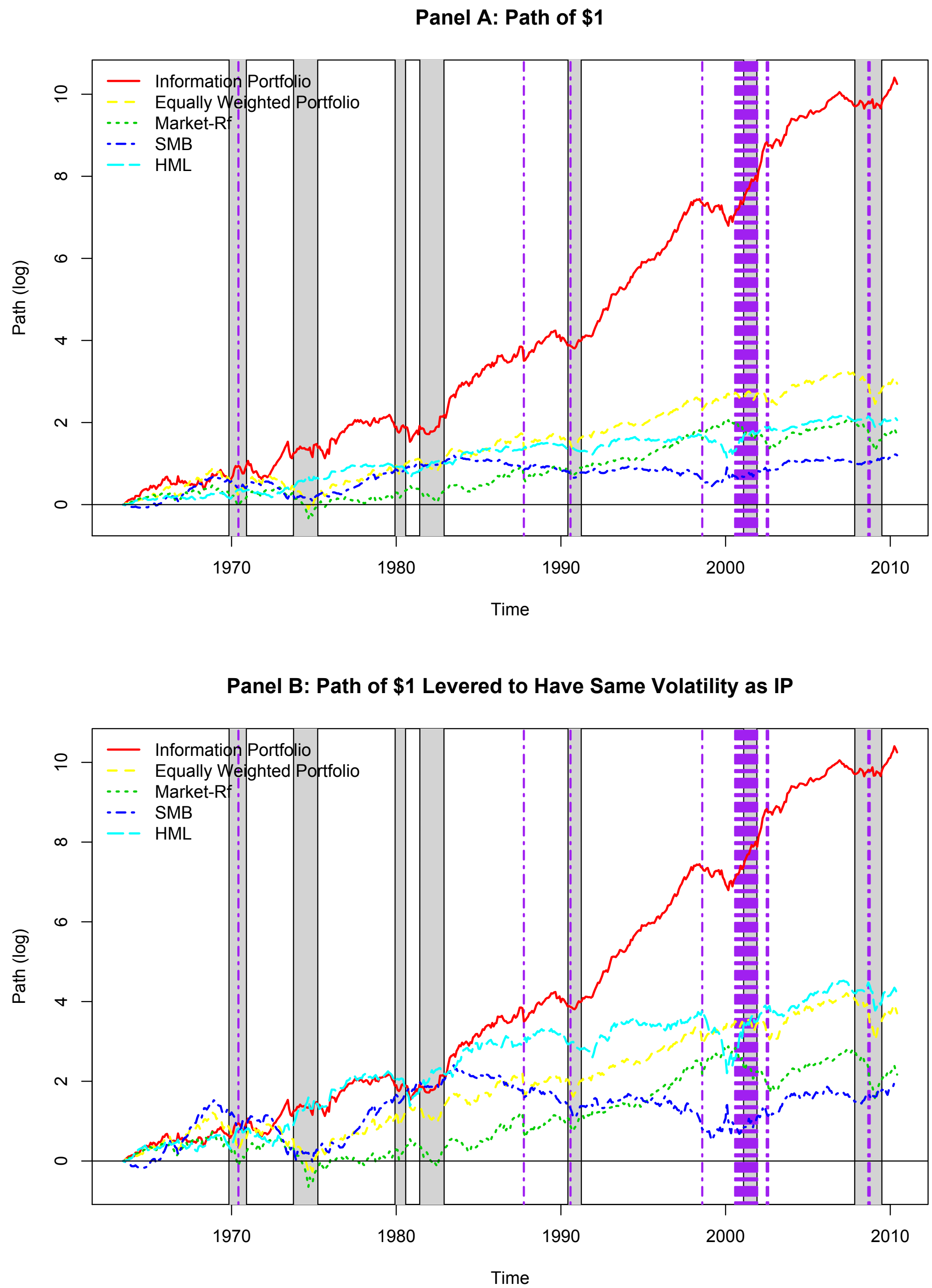
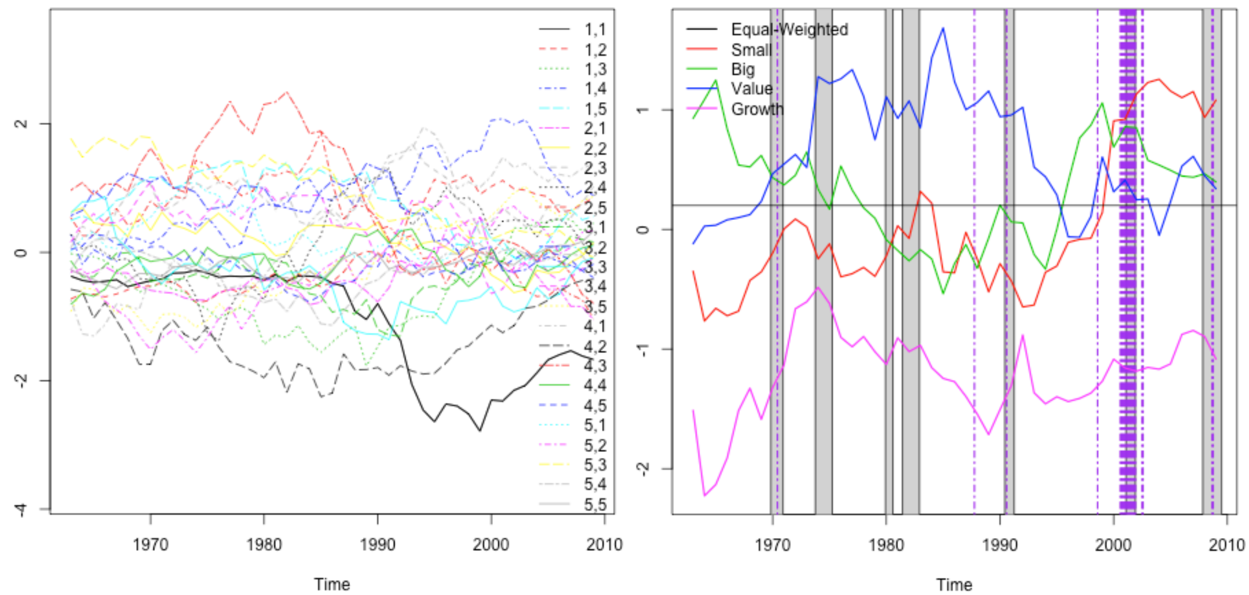


Figure 1: Panel A: cumulated log returns of a zero wealth $1 invested in: information portfolio (red solid line); market portfolio in excess of the risk free rate (green dotted line); SMB portfolio (dark blue dash-dot line); HML portfolio (pale blue long-dash line); $1/N$ portfolio (yellow dashed line). Panel B: same series as Panel A but with portfolios leveraged to the same volatility as the information portfolio. The information portfolio is non-parametrically extracted at a monthly frequency from the 25 Fama–French portfolios using a relative entropy minimization procedure in a rolling out-of-sample fashion starting at 1963:07. Shaded areas indicate NBER recession dates while the vertical dot-dashed lines indicate market crashes identified using the Mishkin and White (2002) approach.



(a) portfolio weights

(b) Small, Big, Value and Growth weights

Figure 2: Portfolio weights of the information portfolio extracted monthly from the 25 Fama–French portfolios. Panel (a): time series of weights assigned to each of the 25 size and book-to-market-equity sorted portfolios. Panel (b): time series of weights assigned to the 'Small', 'Big', 'Growth', and 'Value' portfolios.

Table 7, Figure 2b plots the aggregate weights on portfolios of small, big, growth, and value stocks in the I-P. For instance, writing $(1, 5)$ for the portfolio with stocks in the smallest size quintile and the largest book-to-market-equity quintile, the line labeled 'Small' in the figure plots the sum of the weights on portfolios $(1, 1)$, $(1, 2)$, $(1, 3)$, $(1, 4)$, and $(1, 5)$ at each date. The 'Big,' 'Growth,' and 'Value' curves are similarly defined. The Growth and Value curves reveal that the I-P typically takes a long position in value stocks and a short position in growth stocks, much like the HML factor of Fama–French. However, unlike the latter, the weights on the long and short ends are not constant in the I-P. Although the weights almost always lie between -2 and $+2$, they do vary over time. The Small and Big curves offer a less clean interpretation as a long–short strategy and resemble less the SMB factor. Overall, the weights on the small, large, growth, and value stocks in the I-P are quite different from those implied by the SMB and HML factors. Moreover, our results suggest that this alternative weighting scheme leads to substantially better performance, both in terms of out-of-sample pricing as well as constructing optimally diversified portfolios.

But the I-P outperforms the various benchmark portfolios not only when it is constructed using the FF25 portfolios, but also when different cross-sections are used. In particular, row 5 of Table 7 presents the results when the I-P is constructed from the 10 momentum sorted portfolios. In this case, the returns on the I-P are even higher: about 3% per month. Moreover, once again, the I-P has a Sharpe ratio almost triple that of the $1/N$ portfolio

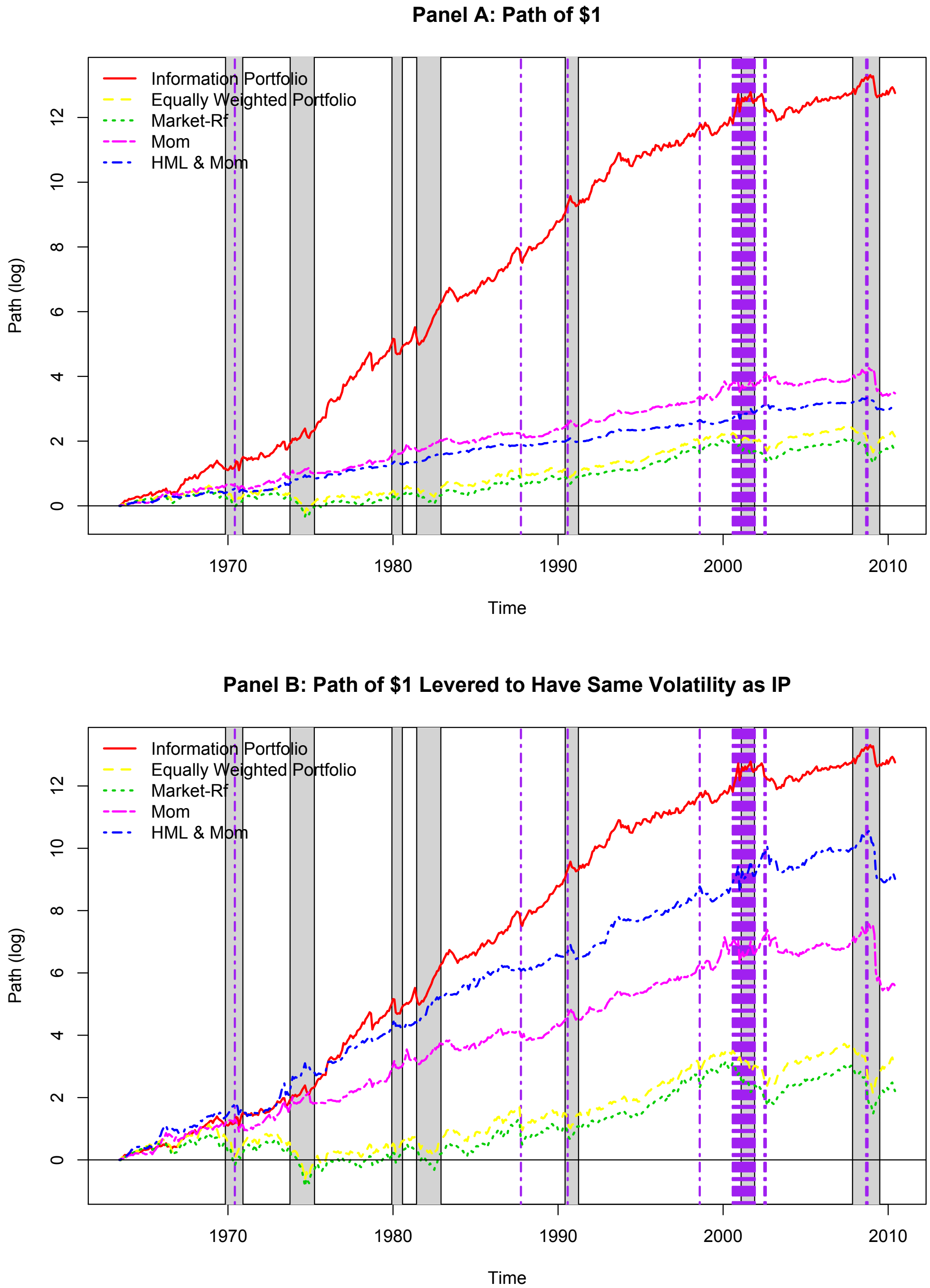


Figure 3: Panel A: cumulated log returns of a zero wealth $1 invested in: information portfolio (red solid line); market portfolio in excess of the risk free rate (green dotted line); $1/N$ portfolio (yellow dashed line); momentum portfolio (purple long-dash-dot line); value and momentum portfolio (dark blue dash-dot line). Panel B: same series as Panel A but with portfolios leveraged to the same volatility as the information portfolio. The information portfolio is extracted monthly from the Small, Big, Value, Growth, Winners, Losers and 10 Industry portfolios, using a rolling out-of-sample fashion starting at 1963:07. Shaded areas indicate NBER recession dates. Vertical dot-dashed lines indicate market crashes identified using the Mishkin and White (2002) approach.

and a CEQ return more than 7 times higher, and similarly outperforms the market, HML, momentum, and HML plus momentum strategies, with neither a large negative skewness risk nor extremely thick tails in the returns distribution. Similar results are obtained when the I-P is constructed from the 25 long-term reversal and size sorted portfolios (row 7).

Moreover, the I-P portfolio shows an even stronger performance (in terms of SR and CEQ) when it is constructed using the the Small, Big, Growth, Value, Winners, and Losers portfolios as well as the 10 industry sorted portfolios (row 8). To show once again that this result is not driven by a particular sub-sample, and in order to offer a time series comparison of this portfolio with the momentum, and the joint value and momentum strategies, Figure 3 plots the path of $1 invested in the I-P (from row 8 of Table 7) over the entire out-of-sample evaluation period. Comparing unleveraged strategies (in Panel A), it is clear that the I-P outperforms the momentum, and value plus momentum, strategies in each 10-year sub-period. Comparing leveraged strategies, the momentum based strategies have a performance comparable to the I-P one in the first ten years or so of the data, but are strongly outperformed by the I-P from the mid 70s onward. Moreover, the I-P tends to have less severe contractions in its returns than the other portfolios during, and following, market-wide crashes (vertical dot-dashed lines in the figure), consistently with the smaller negative skewness and tail risk for this portfolio found in Table 7.

Furthermore, Panel B of Table 7 shows that results similar to those discussed above are obtained when the information portfolio is estimated using quarterly data. This is an important robustness check, since the method proposed in this papers relies on a large time series dimension (T) relative to the cross-sectional one (N). Hence, the stability of the results when the information factor is estimated quarterly is reassuring about the performance of the approach with smaller time series of returns data.

Overall, our results show that the I-P typically outperforms the naïve $1/N$ portfolio as well as other standard benchmarks out-of-sample, in terms of the Sharpe ratio and CEQ return. Moreover, these results seem quite robust with respect to the set of risky assets used for its construction, the data frequency, and the subsample considered. This is consistent with the findings in Section IV.1 that the I-SDF correctly identifies the tangency portfolio and that the I-P is statistically indistinguishable from the ex post maximum Sharpe ratio portfolio of the test assets out-of-sample. Therefore, the I-P offers an attractive procedure for optimal asset allocation across risky assets. Moreover, note that the above results have been obtained using *i*) a very simple approximation of the I-SDF with the I-P and *ii*) without searching for either an optimal rolling window or an optimal rebalancing frequency. As a consequence, the strong performance of the I-P for investment purposes outlined in this section should probably be interpreted as a lower bound on the potential performance of a tradable version of the I-SDF.

V Conclusion and Extensions

Given a set of test assets, we show how an information-theoretic approach can be used to estimate non-parametrically the pricing kernel that prices the given cross-section. We show that this 'information SDF' prices out-of-sample asset returns as well as, or better than, commonly employed multi-factor models (FF3 and Carhart 4-factor models) and that, unlike these factor models, it seems to correctly 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 momentum factors (which explain only a small share of its time variation). These results hold independently of the set of test assets used.

Furthermore, a (low turnover) tradable portfolio that mimics this kernel, which we have referred to as the 'information portfolio', has several interesting out-of-sample properties. First, it delivers smaller pricing errors than the canonical multi-factor models, despite being only a one-factor model. Second, it has a very high Sharpe ratio (about 1 in annualized terms), consistently outperforming the $1/N$ benchmark out-of-sample as well as the value and momentum strategies (whether combined or separate). Third, it leads to an 'information anomaly', generating high alphas of around 8.6%–23.8% per annum relative to the FF3 and momentum factors. Lastly, these results hold for a wide cross-section of assets consisting of size, book-to-market-equity, momentum, industry, and long term reversal sorted portfolios.

The analysis in this paper focuses on the construction of the pricing kernel and the mimicking information portfolio for a *given set of assets*. While this is undoubtedly an important step, the broader economic question is whether there exists a pricing kernel that can successfully price *all* the assets. While the absence of arbitrage opportunities implies the existence of an SDF, the SDF is unique only under the additional condition of market completeness. Our information-theoretic method can help shed light on how the pricing kernels constructed from different asset classes differ from one another, thereby offering guidance regarding the reasons (if any) for market incompleteness and segmentation.

Lastly, the present paper focuses on common stocks. However, our method is very general and could be applied to other asset classes, including bonds, derivatives, currencies, mutual funds and even alternative investment vehicles.

References

Ait-Sahalia, Y., and A. W. Lo (1998): "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, 53(2), 499–547.

——— (2000): "Nonparametric risk management and implied risk aversion," *Journal of Econometrics*, 94(1-2), 9–51.

Alvarez, F., and U. J. Jermann (2005): "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth," *Econometrica*, 73(6), 1977–2016.

Asness, C. S., T. J. Moskowitz, and L. H. Pedersen (2013): "Value and Momentum Everywhere," *The Journal of Finance*, 68(3), 929–985.

Back, K. (2010): *Asset Pricing and Portfolio Choice Theory*, Financial Management Association Survey and Synthesis. Oxford University Press.

Backus, D., M. Chernov, and S. E. Zin (2014): "Sources of Entropy in Representative Agent Models," *The Journal of Finance*, 69(1), 51–99.

Brown, D. E., and R. L. Smith (1990): "A Correspondence Principle for Relative Entropy Minimization," *Naval Research Logistics*, 37(2).

Bryzgalova, S. (2015): "Spurious Factors in Linear Asset Pricing Models," LSE manuscript.

Carhart, M. M. (1997): "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52(1), 57–82.

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(5), 1915–1953.

Fama, E. F., and K. R. French (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *The Journal of Financial Economics*, 33, 3–56.

——— (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116(1), 1–22.

Fama, E. F., and J. MacBeth (1973): "Risk, Return and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607–636.

Ghosh, A., C. Julliard, and A. P. Taylor (2016): "What is the Consumption-CAPM Missing? An Information-Theoretic Framework for the Analysis of Asset Pricing Models," forthcoming in *The Review of Financial Studies*.

Hansen, L. P. (2014): "Nobel Lecture: Uncertainty Outside and Inside Economic Models," *Journal of Political Economy*, 122(5), 945 – 987.

Harvey, C. R., and Y. Liu (2015): "Lucky Factors," Duke University manuscript.

Harvey, C. R., Y. Liu, and H. Zhu (2015): ". . . and the Cross-Section of Expected Returns," *Review of Financial Studies*.

Jackwerth, J. C., and M. Rubinstein (1996): "Recovering Probability Distributions from Option Prices," *Journal of Finance*, 51(5), 1611–32.

Julliard, C., and A. Ghosh (2012): "Can Rare Events Explain the Equity Premium Puzzle?," *Review of Financial Studies*, 25(10), 3037–3076.

Kitamura, Y., and M. Stutzer (2002): "Connections between entropic and linear projections in asset pricing estimation," *Journal of Econometrics*, 107(1-2), 159–174.

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

Markowitz, H. M. (1952): "Portfolio Selection," *Journal of Finance*, 7, 77–91.

McLean, R. D., and J. Pontiff (2016): "Does Academic Research Destroy Stock Return Predictability?," *The Journal of Finance*, 71(1), 5–32.

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

Robinson, P. M. (1991): "Consistent Nonparametric Entropy-based Testing," *Review of Economic Studies*, 58(3), 437–453.

Rosenberg, J. V., and R. F. Engle (2002): "Empirical pricing kernels," *Journal of Financial Economics*, 64(3), 341–372.

Ross, S. A. (2015): "The Recovery Theorem," *The Journal of Finance*, 70(2), 615–648.

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

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(2), 367 – 397.

——— (1996): "A simple Nonparametric Approach to Derivative Security Valuation," *Journal of Finance*, LI(5), 1633–1652.

Uppal, R., and P. Zaffaroni (2015): "Portfolio Choice with Model Misspecification: A Foundation for Alpha and Beta Portfolios," EDHEC Manuscript.

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

A Appendix

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

$$\arg\min_{\mathbb{Q}} D(\mathbb{P}||\mathbb{Q}) \equiv \arg\min_{\mathbb{Q}} \int \ln \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{P} \quad \text{s.t.} \quad \int \mathbf{R}_t^e \, d\mathbb{Q} = \mathbf{0}. \tag{8}$$

Since $\frac{M_t}{\overline{M}} = \frac{d\mathbb{Q}}{d\mathbb{P}}$, the optimization in Equation (8) can be rewritten as

$$\arg\min_{M_t} \ \mathbb{E}^{\mathbb{P}}[\ln M_t] \quad \text{s.t.} \ \mathbb{E}^{\mathbb{P}}[M_t \mathbf{R}_t^e] = \mathbf{0}.$$

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

$$\arg\min_{\{M_t\}_{t=1}^T} \ \frac{1}{T}\sum_{t=1}^T \ln M_t \quad \text{s.t.} \ \frac{1}{T}\sum_{t=1}^T M_t \mathbf{R}_t^e = \mathbf{0}. \tag{9}$$

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

$$\widehat{M}_t \equiv M_t\left(\widehat{\theta}_T, \mathbf{R}_t^e\right) = \frac{1}{T(1+\widehat{\theta}_T' \mathbf{R}_t^e)}, \quad \forall t \tag{10}$$

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

$$\arg\min_{\theta} \ -\frac{1}{T}\sum_{t=1}^T \log(1+\theta' \mathbf{R}_t^e),$$

and this last expression is the dual formulation of the entropy minimization problem in Equation (9). 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).

Since the correlation of the SDF estimates obtained with either Equations (5) or (10) is extremely high (more than 95%), and the pricing performances of the two are almost indistinguishable, to simplify the exposition we present only the results based on the former.