Which (Nonlinear) Factor Models?*

Caio Almeida[†] **Gustavo Freire**[‡]

January 31, 2023

Preliminary draft: Please do not circulate

Abstract

Traditional asset pricing tests boil down to evaluating the maximum Sharpe ratio obtained from the factors in a given model. This implicitly assumes the linear stochastic discount factor (SDF) that prices the factors as the asset pricing model. We generalize this approach by considering a comprehensive family of nonlinear SDFs pricing the model factors. The relevant metric for model comparison becomes the maximum Sharpe ratio of the mimicking portfolio constructed by projecting the nonlinear SDF onto the test assets. We show that nonlinearities matter empirically for both *absolute* and *relative* pricing performance of leading factor models.

JEL classification: C52, G11, G12

Keywords: Model Comparison, Factor Models, Anomalies, Stochastic Discount Factor, Nonlinearities

*We would like to thank Onno Kleen, Erik Kole, Alberto Quaini and seminar participants at the Erasmus School of Economics for useful comments and suggestions.

†E-mail: calmeida@princeton.edu, Department of Economics, Princeton University.

‡E-mail: freire@ese.eur.nl, Erasmus School of Economics, Erasmus University Rotterdam, and Tinbergen Institute.

1. Introduction

The ultimate goal of asset pricing is to find systematic factors that can explain the cross-section of asset expected returns. The seminal capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965) predicts that the expected excess return of any asset is equal to its exposure to the market factor times the market risk premium. This prediction has been eventually rejected empirically, giving rise not only to hundreds of cross-sectional anomalies as test assets that are challenging to price (Hou, Xue and Zhang, 2020), but also to hundreds of alternative factors beyond the market (Harvey, Liu and Zhang, 2016). Given the existing multitude of candidate factors, evaluating and comparing asset pricing models is of fundamental importance.

For traded factors like the market or the Fama and French (1993, 2015) factors, the usual approach for model evaluation is the Gibbons, Ross and Shanken (GRS, 1989) test. The null hypothesis is that the intercepts of time-series regressions of test assets excess returns on the factors, which are the pricing errors "alphas", are jointly equal to zero. The test statistic is a quadratic form in the alphas that is equivalent to the improvement in the squared Sharpe ratio obtained from investing in the test assets in addition to the factors. A nonzero alpha indicates that the factors do not span the tangency portfolio and the Sharpe ratio can be improved by taking positions in the test assets.

For nontraded factors like consumption growth (Breeden, 1979), a two-step approach is required as their risk premia must also be estimated. First, exposures to the factors or "betas" are obtained from time-series regressions, and then the risk premia of the factors are estimated from a cross-sectional regression (CSR) of average excess returns on the betas. The residuals of the second regression are the alphas as they capture deviations from the linear expected return-beta relation. As noted by Barillas and Shanken (2017), if the CRS is estimated with generalized least squares (GLS), a quadratic form in the alphas again reduces to analyzing the improvement in the squared Sharpe ratio from trading in the test assets, but now in addition to that of the mimicking portfolios of the factors. This is such that, if all factors are traded and included as test assets, the metrics of model mispricing under the GRS and GLS CRS approaches are equivalent.

When it comes to model comparison in terms of pricing performance, models that yield smaller alphas for a set of test assets have typically been preferred. More recently, Barrillas and Shanken (2017) argue that a model should be able to price not only the test assets, but also the traded factors in the competing models, i.e., the whole universe of assets under consideration. They show that this premise has a surprising implication under the alpha mispricing metric: model comparison boils down to comparing the maximum squared Sharpe ratio of the factors in each model, such that test assets are irrelevant.

In this paper, we generalize traditional asset pricing tests to incorporate nonlinearities. We start by showing that the GRS or GLS CSR approach for traded factors is equivalent to first computing the linear stochastic discount factor (SDF) that prices the factors (Hansen and Jagannathan, 1991) and then using it as a single factor in a GLS CSR.[1] This is because the mimicking portfolio of the linear SDF is precisely the portfolio of the factors with maximum squared Sharpe ratio, so that the quadratic form in the alphas (or the squared Sharpe ratio improvement) is the same in both cases.[2] We then propose to use a nonlinear SDF that prices the factors as the single factor in the GLS CSR. The relevant metric for pricing performance becomes the maximum squared Sharpe ratio of the mimicking portfolio of the nonlinear SDF.

Under no-arbitrage, there exists an infinity of admissible SDFs that price the factors beyond the linear one. Each of these admissible SDFs introduce nonlinearities that are irrelevant to price the factors, but that may be relevant to price the universe of test assets. In fact, there is a large literature showing that nonlinearities are important in asset pricing (Bansal and Viswanathan, 1993; Chapman, 1997; Harvey and Siddique, 2000; Dittmar, 2002; Vanden, 2006; Schneider, Wagner and Zechner, 2020). Therefore, nonlinearities may have significant implications for model evaluation and comparison. In particular, we derive a sufficient condition for a nonlinear SDF to improve upon the pricing performance of the linear one: the nonlinearity it adds must provide an insurance for systematic risk, i.e., it must covary positively with the economy-wide SDF.

Our framework also provides a reason why test assets are irrelevant for model comparison in the standard approach. If the SDF is a linear function of the factors, its mimicking portfolio (i.e., its projection onto the test assets returns) loads only on the factors themselves and has zero weights in the remaining test assets. Thus, a comparison of maximum squared Sharpe ratios between two models does not depend on the test assets. In contrast, if a nonlinear SDF pricing the factors of a given model is considered in the GLS CSR, its mimicking portfolio will load on the whole universe of test assets. That is, the relevance of test assets is restored as they are needed to mimic the nonlinearities.

[1] For a nontraded factor, we show that the standard GLS CSR approach is equivalent to using the linear SDF that prices the mimicking portfolio of the factor in a GLS CSR.

[2] This is true under the premise of Barillas and Shanken (2017) that traded factors are included in the set of test assets, which we follow throughout the paper.

A natural question is then *which* nonlinear SDFs to consider from the no-arbitrage admissible set. This set is in general very large and may contain SDFs that are not economically meaningful. The standard approach implicitly uses the linear SDF, which minimizes variance and, by duality, bounds the Sharpe ratio obtained from the factors, being directly related to quadratic utility. We propose to work with SDFs minimizing Cressie and Read (1984) discrepancy functions that generalize the variance metric. These SDFs take into account higher moments and bound generalized notions of Sharpe ratio, being related to the broad class of hyperbolic absolute risk aversion (HARA) utility functions (Almeida and Garcia, 2017; Almeida and Freire, 2022). Importantly, the minimum discrepancy SDFs and their nonlinearities are indexed by a single parameter γ, allowing us to track and economically interpret how pricing performance depends on it.

Empirically, we investigate whether incorporating nonlinearities matter for the absolute and relative pricing performance of leading factor models. More specifically, we test (1) how the nonlinear models compare to the linear model for a given set of factors and (2) how the best nonlinear model (within the Cressie-Read family) of a set of factors compares to that of a competing set of factors. We consider 10 factor models encompassing 19 unique factors: the market factor (CAPM); the 2-factor intermediary asset pricing model of He, Kelly and Manela (HKM, 2017); the betting-against-beta extension of the CAPM of Frazzini and Pedersen (BAB, 2014); the factor model of Daniel, Hirshleifer and Sun (DHS, 2018), which adds 2 behavioral factors to the market; the Fama and French (1993) 3-factor model (FF3); the investment q-factor model of Hou, Xue and Zhang (2015) (q4); the Fama and French (2015) 5-factor model (FF5); the hedged FF5 of Daniel et al. (FF5*, 2020); FF5 plus momentum (Carhart, 1997) (FF6); and the Barrilas and Shanken (2018) 6-factor model (BS). Our baseline set of test assets is given by the 19 unique factors and 44 anomalies from Kozak, Nagel and Santosh (2020).

We find that nonlinearities substantially improve the absolute pricing performance of nearly all factor models considered. In some cases, such as for the CAPM and the BAB, the maximum squared Sharpe ratio can even double. These results are statistically significant and also hold out-of-sample. This is striking as the minimum discrepancy SDFs are not optimized to maximize pricing performance across the universe of test assets. In fact, just as in the linear case, only information on the factors is used, such that improvements come solely from economically meaningful nonlinearities embedded in the SDF. The minimum discrepancy SDFs that yield those improvements are the ones that minimize skewness and put less weight in extreme factor returns. In contrast, nonlinear SDFs that maximize skewness and overweight extreme returns lead to worst performance compared to the linear SDF.

Nonlinear SDFs also have strong implications for factor model comparison. First, nonlinearities improve substantially the performance of the CAPM, imposing a stronger hurdle to beat it. The CAPM outperforms both the HKM and FF3 models when nonlinearities are allowed, while the opposite happens under the linear specification. This is consistent with papers showing that nonlinear versions of the CAPM perform well in cross-sectional asset pricing (Harvey and Siddique, 2000; Dittmar, 2002). Second, since the relative benefits of nonlinearities vary across models, accounting for it often leads to different rankings between factor models. For instance, even though BAB is outperformed by FF5 under the traditional linear approach, its best nonlinear SDF yields a maximum Sharpe ratio that is 46% higher than that of the best nonlinear model of FF5. Overall, the best performing factor model is the DHS, followed by BS and FF5*. This is true both under the linear and nonlinear specifications. In this sense, we can say this ranking is robust within the family of nonlinearities we consider.

The remainder of the paper is organized as follows. After a brief review of the related literature, Section 2 summarizes standard asset pricing tests and their implications for model comparison. Section 3 presents our generalization to allow for nonlinearities and discusses the specific nonlinear SDFs we consider. Section 4 contains the results for our empirical analysis of the importance of nonlinearities for the pricing performance of leading factor models. Section 5 concludes the paper.

1.1. Related literature

Our paper is mainly related to three strands of the literature. The first strand studies asset pricing tests. Gibbons, Ross and Shanken (1989) provide a test for the efficiency of a model with traded factors. Shanken (1985) develops a test based on a quadratic form of GLS CSR pricing errors, which Kan and Robotti (2008) show is analogous to a modified Hansen and Jagannathan (1997) distance. Lewellen, Nagel and Shanken (2010) discuss how to improve empirical tests. Barillas and Shanken (2017) show that standard tests imply that the preferred model is the one with higher maximum squared Sharpe ratio. Barillas and Shanken (2018) derive a Bayesian asset pricing test, while Barillas et al. (2020) provide asymptotic tests for model comparison based on maximum

Sharpe ratios. Detzel, Novy-Marx and Velikov (2022) take into account transaction costs when evaluating asset pricing models. While all these papers focus on linear models, we generalize the usual regression-based approaches to allow for nonlinearities.

The second strand of the literature introduces nonlinearities in asset pricing. Bansal and Viswanathan (1993) and Chapman (1997) use neural networks and orthonormal polynomials, respectively, to estimate a nonlinear SDF as a function of a few state variables. Harvey and Siddique (2000) consider a conditional version of the three-moment CAPM of Kraus and Litzenberger (1976) where coskewness is priced, while Dittmar (2002) proposes a cubic SDF taking into account preferences for cokurtosis with the market. Vanden (2006) provides conditions under which the economy SDF depends on quadratic terms of index option returns. Schneider, Wagner and Zechner (2020) document that an SDF that is a quadratic function of market returns can explain low risk anomalies. We show that incorporating nonlinear functions of a given set of factors in asset pricing tests substantially improves the pricing performance of leading factor models and can lead to different conclusions regarding model comparison relative to the linear case.

The third strand makes use of Cressie-Read discrepancies for different purposes in finance. A number of papers have considered SDFs minimizing the Cressie-Read family (Almeida and Garcia, 2012, 2017) or particular members of the family such as variance (Hansen and Jagannathan, 1991), entropy (Stutzer, 1995; Bansal and Lehmann, 1997; Alvarez and Jermann, 2005; Backus, Chernov and Zin, 2014; Ghosh, Julliard and Taylor, 2017) and generalized entropy (Snow, 1991; Liu, 2021) for diagnosing asset pricing models. Stutzer (1996) and Almeida and Freire (2022) analyze the option pricing implications of the minimum entropy SDF and the whole family of minimum discrepancy SDFs, respectively. Ghosh, Julliard and Taylor (2019) show that an out-of-sample extension of the minimum entropy SDF estimated from test assets outperforms popular factor models, while Sandulescu, Trojani and Vedolin (2021) examine ratios of minimum entropy SDFs from international markets. We analyze the implications of using minimum discrepancy SDFs pricing a given set of factors as the asset pricing model instead of the minimum variance SDF.

2. Traditional asset pricing tests

Consider N assets with excess returns R and K traded factors with returns f that are also in excess of the risk-free rate or return spreads of long-short portfolios. Asset pricing is concerned about whether the linear expected return-beta relation holds:

$$\mathbb{E}(R) = \beta\lambda, \tag{1}$$

where β is $N \times K$ and contains the covariances between the assets and the factors, and λ is the K-vector of expected excess returns of the factors, i.e., their risk premia. The relation above states that any expected return beyond the risk-free rate should come as a compensation for exposure to systematic factor risk. Deviations from this relation in the form of pricing errors $\alpha \equiv \mathbb{E}(R) - \beta\lambda$ indicate that the factor model f does not contain the true systematic factors.

The traditional approach to test a model of traded factors is the Gibbons, Ross and Shanken (1989) test. This approach consists of a multivariate linear regression with time-series observations on R_t and f_t:

$$R_t = \alpha + \beta f_t + \epsilon_t, \qquad t = 1, ..., T, \tag{2}$$

where all variables are N-vectors, with the exception of the $N \times K$ β matrix and the K-vector of factors. The error term ϵ_t has zero mean and an invertible covariance matrix Σ. The null hypothesis is that the entries of α are jointly equal to zero, that is, that relation (1) holds. The GRS test is based on a quadratic form in the alphas that they show is equivalent to the improvement in the maximum squared Sharpe ratio $Sh^2(.)$ attainable from investing in the assets in addition to the factors:[3]

$$\alpha' \Sigma^{-1} \alpha = Sh^2(f, R) - Sh^2(f). \tag{3}$$

In other words, a nonzero alpha indicates that the factors do not span the tangency portfolio, or, equivalently, do not attain the maximum squared Sharpe ratio in the economy.

For models where the factors f are nontraded, the same relation (1) should hold, but now the means of the factors are uninformative and different from their risk premia. In

[3] For any set of returns $\tilde{R}$, the maximum squared Sharpe ratio is given by $Sh^2(\tilde{R}) = \mathbb{E}(\tilde{R})' Var(\tilde{R})^{-1} \mathbb{E}(\tilde{R})$.

this case, the GRS test is not applicable, and a two-step approach is needed instead as λ must also be estimated. First, the betas with respect to the factors are obtained from time-series regressions for each asset $i = 1, ..., N$:

$$R_{i,t} = c_i + \beta_i f_t + u_{i,t}, \qquad t = 1, ..., T. \tag{4}$$

Then, a cross-sectional regression (CSR) of expected excess returns on betas obtains the risk premia as the slope coefficients and the pricing errors as the residuals:

$$\mathbb{E}(R) = \beta\lambda + \alpha. \tag{5}$$

The regression does not contain an intercept, such that the residuals capture deviations from the linear expected return-beta relation. The null hypothesis of $\alpha = 0$ can also be tested with a quadratic form (Shanken, 1985). Using results from Lewellen, Nagel and Shanken (2010), Barillas and Shanken (2017) show that, if the CRS is estimated with generalized least squares (GLS), a quadratic form in the alphas again reduces to the improvement in the squared Sharpe ratio from trading in the assets, but now in addition to that of the mimicking portfolios of the factors. That is:

$$\alpha' V^{-1} \alpha = Sh^2(R) - Sh^2(f_p), \tag{6}$$

where V is the covariance matrix of R and f_p are the returns of the mimicking portfolios of the original factors f.[4]

In sum, standard asset pricing tests boil down to evaluating the maximum squared Sharpe ratio obtained from the factors. In fact, both approaches above are equivalent if the factors are traded and included in the set of asset returns R (Barillas and Shanken, 2017). To see that, note that in this case R includes both the tests assets and the factors and the mimicking portfolios of traded factors are the factors themselves. This implies that the expression in (6) equals that in (3).

[4]The mimicking portfolio for a factor f is given by the projection of the factor on the returns and a constant. More specifically, $f_{p,t} = AR_t$, where A is obtained from the time-series regression $f_t = a + AR_t + \eta_t$.

2.1. *Model comparison*

Given that all models are misspecified, it is of interest to be able to compare models. Considering the alpha mispricing metric implied by traditional asset pricing tests, Barillas and Shanken (2017) provide a surprising result. While most of the empirical literature has compared the performance of factor models in pricing different sets of test assets, they argue that a model should also be able to price the traded factors in competing models, i.e., the whole universe of assets under consideration. As it turns out, this implies that comparing two models is equivalent to comparing the maximum squared Sharpe ratio of the factors in each model, such that test assets are irrelevant.

The argument is simple. Let f_1 and f_2 be two competing models of traded factors and R the returns of a set of basis test assets, such that the whole universe of test assets is given by $R_{all} = [R, f_1, f_2]$. According to the alpha mispricing metric, factor model f_1 is preferred if its improvement in Sharpe ratio when investing in the test assets is smaller than that for f_2, that is, if:

$$Sh^2(R_{all}) - Sh^2(f_1) < Sh^2(R_{all}) - Sh^2(f_2). \tag{7}$$

The common term above drops out and we have that the better model is the one which factors yield the higher maximum squared Sharpe ratio: $Sh^2(f_1) > Sh^2(f_2)$. Throughout the paper, we follow the premise that traded factors are included in the set of test assets.

3. Incorporating nonlinearities

To introduce nonlinearities into asset pricing tests, we propose a simple generalization of the traditional methods described in Section 2. Our approach can be seen as a three-step procedure. First, for a given traded factor model f, we identify an SDF m that prices the factors, i.e., that satisfies the Euler equation:

$$\mathbb{E}(mf) = 0. \tag{8}$$

Then, we run the two-step GLS CSR using the SDF m as a single nontraded factor to obtain the pricing errors α. From Equation (6), the following holds:

$$\alpha' V^{-1} \alpha = Sh^2(R_{all}) - Sh^2(m_p), \tag{9}$$

where m_p is the mimicking portfolio of the SDF and R_{all} contains the basis test assets and the factors f (and any other competing factors).

We now show that the standard asset pricing tests in Section 2 are the particular case of our approach that uses in the first step the unique linear SDF that prices the factors (Hansen and Jagannathan, 1991):

$$m^* = 1 - b'[f - \mu_f], \quad b = \Sigma_f^{-1}\mu_f, \tag{10}$$

where $\mu_f = \mathbb{E}(f)$ and $\Sigma_f = Var(f)$. The SDF above is a linear function of the portfolio of the factors $b'f$ with maximum squared Sharpe ratio. Therefore, if we run the GLS CSR on m^* and analyze the quadratic form in the pricing errors alphas, we obtain:

$$\alpha' V^{-1}\alpha = Sh^2(R_{all}) - Sh^2(m_p^*) = Sh^2(R_{all}) - Sh^2(-b'f) = Sh^2(R_{all}) - Sh^2(f). \tag{11}$$

The second equality stems from the fact that the mimicking portfolio m_p^* of m^*, i.e., its projection on R_{all}, recovers precisely the portfolio $-b'f$ since m^* is linear in the factors.[5] The third equality holds because $b'f$ is already the portfolio of the factors that yields the maximum squared Sharpe ratio $Sh^2(f)$. In other words, the GRS or GLS CSR approaches on the factors f are equivalent to using the linear SDF pricing f in the GLS CSR.

Our framework also provides additional insights into the test asset irrelevance result of Barillas and Shanken (2017). Test assets are irrelevant in the usual approach because they are not needed to mimic the SDF, which is already a linear function of factors f. In contrast, if a nonlinear SDF pricing the factors of a given model is considered, its mimicking portfolio will load on the entire universe of test assets. That is, test assets become relevant as they are needed to mimic the nonlinearities.

We thus propose to use a nonlinear SDF that prices the factors as the single factor in the GLS CSR. Under no-arbitrage, there exists an infinity of admissible SDFs satisfying the Euler equation (8) beyond the linear one.[6] Each of these alternative SDFs introduce nonlinearities that are irrelevant to price the factors, but that may be relevant to price the whole universe of test assets. In the next subsection, we make this statement more precise and discuss the implications of nonlinearities for the pricing performance of a traded factor model f.

[5] That is, in the regression $1 - b'[f_t - \mu_f] = a + AR_t + \eta_t$, A is equal to $-b$ for the factors f_t and zero for the remaining test assets in R_t, while $a = 1 + b'\mu_f$.

[6] This is true under incomplete markets, which is the realistic case.

While we focus on the case of traded factors f, our framework also analogously generalizes traditional asset pricing tests for nontraded factors. In this case, our approach would first identify an SDF m that prices the mimicking portfolios of the nontraded factors, and then use this SDF in the GLS CSR. As detailed in Appendix A, the traditional GLS CSR approach applied directly to the nontraded factors is a particular case of our framework that uses the linear SDF pricing the mimicking portfolios of the factors.

3.1. Implications of nonlinearities for pricing performance

We start with a decomposition that any admissible SDF m satisfies (Cochrane, 2001):

$$m = m^* + e, \quad E(e) = 0, \; E(ef) = 0. \tag{12}$$

The decomposition shows that the nonlinear term e simply adds noise for pricing f. However, since any model is potentially misspecified, it is reasonable to assume that the factors f do not fully capture the systematic risk in the extended economy with all test assets. In this case, the nonlinearity *can* improve the pricing performance relative to the linear SDF.

More precisely, using the decomposition (12), we can rewrite the metric of model mispricing in (9) as:

$$\alpha' V^{-1} \alpha = Sh^2(R_{all}) - Sh^2(m_p) = Sh^2(R_{all}) - Sh^2(m_p^* + e_p), \tag{13}$$

where e_p is the mimicking portfolio of the nonlinearity e. The maximum squared Sharpe ratio of m_p can be further simplified into:

$$Sh^2(m_p^* + e_p) = \frac{\mathbb{E}(m_p^* + e_p)^2}{Var(m_p^* + e_p)} = \frac{\mu^2_{m_p^*}}{\sigma^2_{m_p^*} + \sigma^2_{e_p}} + \frac{\mu^2_{e_p}}{\sigma^2_{m_p^*} + \sigma^2_{e_p}} + \frac{2\,\mu_{m_p^*}\,\mu_{e_p}}{\sigma^2_{m_p^*} + \sigma^2_{e_p}}, \tag{14}$$

where μ_x and σ^2_x denote expected value and variance, respectively, of the variable x in the subscript. The expression above tells us that, everything else constant, a more volatile nonlinearity hurts the pricing performance of the nonlinear SDF. However, for reasonable distortions of m^*, the variance of e (which is a difference of SDFs that has mean zero) should be small relative to the variance of m^*. Under this assumption, we can write $\sigma^2_{m_p^*} + \sigma^2_{e_p} \approx \sigma^2_{m_p^*}$ and obtain the following approximation:

$$Sh^2(m_p^* + e_p) \approx Sh^2(f) + \frac{\mu_{e_p}^2}{\sigma_{m_p^*}^2} + \frac{2\,\mu_{m_p^*}\,\mu_{e_p}}{\sigma_{m_p^*}^2}. \tag{15}$$

Since the second term of (15) is always positive, it is possible to arrive at a sufficient condition for the nonlinear SDF to improve upon the performance $Sh^2(f)$ of the linear SDF by studying the signs of $\mu_{m_p^*}$ and μ_{e_p}. For that, it is helpful to understand what determines their signs. Both m_p^* and e_p are portfolios of traded assets in the extended economy. Hence, their expected returns depend on how they covary with the economy-wide SDF, i.e., the benchmark linear SDF m_{all}^* that prices the whole universe of assets and is associated with $Sh^2(R_{all})$.[7] More specifically, starting from the Euler equations $\mathbb{E}(m_{all}^*\, m_p^*) = 0$ and $\mathbb{E}(m_{all}^*\, e_p) = 0$, it is easy to show that:

$$\mu_{m_p^*} = -cov(m_p^*, m_{all}^*) = -cov(m^*, m_{all}^*), \tag{16}$$

$$\mu_{e_p} = -cov(e_p, m_{all}^*) = -cov(e, m_{all}^*). \tag{17}$$

That is, an asset gets a negative expected excess return if it provides insurance for marginal utility by covarying with the economy SDF. Arguably, we should expect that any model consisting of traded factors f produces an SDF m^* that covaries positively with the economy-wide SDF due to the comovement of asset returns. In fact, this is the case empirically for all the factor models we consider in Section 4. This implies that $\mu_{m_p^*} < 0$. Therefore, under approximation (15), a sufficient condition for nonlinearities to improve upon the linear case is that $\mu_{e_p} < 0$. In other words, if the nonlinearity e is an insurance for systematic risk, the nonlinear SDF m outperforms the linear one m^*. More than that, the better an insurance e is (the more it covaries with m_{all}^*), the stronger is the outperformance of m relative to m^*.

The results above show how nonlinearities can improve the absolute pricing performance of a given factor model f relative to the traditional linear case. For model comparison between a nonlinear SDF pricing f_1 and a nonlinear SDF pricing f_2, the first model would be preferred if $Sh^2(m_{p,1}) > Sh^2(m_{p,2})$. Using again (15), this relation can be approximated by:

[7] Note that, *for the whole economy*, it suffices to work with the linear SDF m_{all}^* as any nonlinear SDF m_{all} simply adds noise for pricing the entire universe of assets R_{all}.

$$Sh^2(f_1) + \frac{\mu_{e_{1,p}}^2}{\sigma_{m_{1,p}^*}^2} + \frac{2\,\mu_{m_{1,p}^*}\,\mu_{e_{1,p}}}{\sigma_{m_{1,p}^*}^2} > Sh^2(f_2) + \frac{\mu_{e_{2,p}}^2}{\sigma_{m_{2,p}^*}^2} + \frac{2\,\mu_{m_{2,p}^*}\,\mu_{e_{2,p}}}{\sigma_{m_{2,p}^*}^2}. \tag{18}$$

While it is not possible to derive explicit conditions, nonlinearities *can* lead to different conclusions about the preferred model compared to the linear case. That is, we may have $Sh^2(f_1) < Sh^2(f_2)$ but $Sh^2(m_{p,1}) > Sh^2(m_{p,2})$ depending on the relative performance of the nonlinearities in each model (i.e., on the covariance of the nonlinear term of each model with the economy-wide SDF).

Given that nonlinearities can have important implications for both absolute and relative asset pricing performance, the natural question is *which* nonlinear SDFs to consider from the no-arbitrage admissible set, which is in general very large. In the next subsection, we propose to work with an economically meaningful and tractable family of nonlinear SDFs that naturally generalize the linear case.

3.2. Minimum discrepancy SDFs

As we have shown, traditional asset pricing tests use the linear SDF pricing the factors f as the asset pricing model. The linear SDF is the projection of any admissible SDF on the space of factors excess returns, such that m^* has the minimum variance among all candidate SDFs (this can also be seen from decomposition 12). By duality, the minimum variance SDF also imposes an upper bound on the Sharpe ratio that can be attained from the factors and can be seen as the marginal utility of an investor solving an optimal portfolio problem by maximizing expected quadratic utility (Cerny, 2003).

We propose to use nonlinear SDFs that naturally generalize the minimum variance one. More specifically, we consider SDFs minimizing the Cressie and Read (1984) family of discrepancies:

$$\begin{aligned} &\underset{m}{min}\ \mathbb{E}\left[\tfrac{m_\gamma^{\gamma+1}-1}{\gamma(\gamma+1)}\right] \\ &\text{s.t. } \mathbb{E}(m_\gamma\, f) = 0,\ \mathbb{E}(m_\gamma) = 1,\ m_\gamma \geq 0, \end{aligned} \tag{19}$$

where the parameter $\gamma \in \mathbb{R}$ indexes the particular Cressie-Read loss function and the corresponding minimum discrepancy SDF. The minimum variance SDF is a particular case when $\gamma = 1$, with the difference that we impose a nonnegativity constraint in the SDF.[8] This constraint is important to guarantee that the nonlinear SDFs we identify are

[8]This constraint is also considered in Hansen and Jagannathan (1991) as an alternative specification.

consistent with no-arbitrage in the extended economy, which is not necessarily satisfied by m^* as it can reach negative values. Whenever the constraint is not binding, $m_{\gamma=1}$ and m^* are equivalent.

In their seminal work, Hansen and Jagannathan (1991) propose to diagnose asset pricing models via moment restrictions to pricing kernels. They derive the minimum variance SDF placing a lower bound on the variance of candidate pricing kernels. This approach is generalized by Almeida and Garcia (2017) by deriving the Cressie-Read family of minimum discrepancy SDFs, which embeds as particular cases virtually all SDF moment restrictions proposed in the literature.[9] These papers investigate whether the pricing kernel of a candidate model is able to generate enough dispersion as measured by a particular Cressie-Read loss function. In contrast, we analyze how minimum discrepancy SDFs identified from a set of factors f compare to the linear SDF in pricing the universe of test assets R_{all} under the traditional alpha mispricing metric.

While problem (19) is of infinite dimension, Almeida and Garcia (2017) show that it can be solved via a much simpler dual problem of dimension equal to the number of pricing restrictions. Under no-arbitrage, it is equivalent to solve, for $\gamma < 0$:[10]

$$\theta_\gamma = \underset{\theta \in \{1-\gamma\theta' f>0\}}{arg\ max}\ \mathbb{E}\left[-\frac{1}{\gamma+1}\left(1-\gamma\,\theta' f\right)^{\left(\frac{\gamma+1}{\gamma}\right)}\right], \tag{20}$$

where the minimum discrepancy SDF can be recovered from the first-order condition of (20) with respect to θ, evaluated at θ_γ:

$$m_\gamma = \left(1-\gamma\,\theta_\gamma' f\right)^{\frac{1}{\gamma}}. \tag{21}$$

Mathematically, θ_γ is the vector of Lagrange multipliers associated with the Euler equations for the factors in (19). Economically, Almeida and Freire (2022) show that (20) can be interpreted as an optimal portfolio problem for an investor maximizing a HARA utility function with concavity parameter γ, where θ_γ is proportional to the optimal

[9] Snow (1991) places restrictions on the p-th moment of the pricing kernel, with $p > 1$, which corresponds to $\gamma > 0$. Stutzer (1995) derives an information bound minimizing the Kullback–Leibler divergence ($\gamma \to 0$), while Bansal and Lehmann (1997) obtain a growth-optimal bound that minimizes entropy ($\gamma \to -1$). More recently, Liu (2021) complements Snow (1991) by restricting the p-th moment of the SDF, with $p \in (-\infty, 1)$, corresponding to $\gamma < 0$.

[10] For $\gamma > 0$, the problem is unconstrained with an indicator function in the objective function: $\mathbb{E}\left[-\frac{1}{\gamma+1}\left(1-\gamma\theta' f\right)^{\left(\frac{\gamma+1}{\gamma}\right)} I_{\Theta_\gamma(f)}(\theta)\right]$, where $\Theta_\gamma(f) = \{\theta \in \mathbb{R}^K : (1-\gamma\theta' f) > 0\}$ and $I_A(x) = 1$ if $x \in A$, and 0 otherwise. For $\gamma = 0$, the problem is unconstrained and the objective function is exponential: $\mathbb{E}\left[-e^{-\theta' f}\right]$.

allocation of wealth in the factors f. The SDF m_γ is the marginal utility of the investor. For $\gamma = 1$, the investor maximizes quadratic utility, which is equivalent to maximizing the Sharpe ratio obtained from the factors. For different γ's, the maximization of HARA utility maps to maximizing generalized notions of Sharpe ratio attainable from the factors that take into account nonlinearities (Cerny, 2003; Almeida and Freire, 2022).

The parameter γ controls the relative importance of higher moments for the minimum discrepancy SDF and, consequently, the particular shape of its distortion of the linear SDF. This can be seen by Taylor expanding the expected value of the Cressie-Read loss function $\phi_\gamma(m) = \frac{m^{\gamma+1}-1}{\gamma(\gamma+1)}$ around the SDF mean of 1:

$$\mathbb{E}\left[\phi_\gamma\left(m\right)\right] = \frac{1}{2}\mathbb{E}\left(m-1\right)^2 + \frac{(\gamma-1)}{3!}\mathbb{E}\left(m-1\right)^3 + \frac{(\gamma-1)(\gamma-2)}{4!}\mathbb{E}\left(m-1\right)^4 + \ldots \quad (22)$$

For $\gamma = 1$, the minimum discrepancy problem (19) minimizes variance as all higher moments are given zero weights. All other discrepancies give the same weight to the variance, but each one weighs nonlinearities of the SDF differently. The main distinction comes from whether γ is below or above one. For $\gamma < 1$, skewness is maximized as it is given a negative weight. In contrast, skewness is minimized for $\gamma > 1$. The more extreme the γ, the higher the relative importance of skewness. On the other hand, kurtosis is minimized for essentially any γ other than one.[11]

To illustrate how γ affects the shape of the SDF, the left panel of Figure 1 plots three SDFs obtained from the market factor as a function of the factor returns.[12] Each market return state corresponds to one time-series observation in our data. The minimum variance SDF m^* is a linear function of the market. Since it never attains negative values, m^* is equal to the SDF minimizing the Cressie-Read loss function with $\gamma = 1$. For $\gamma = -2$, the SDF maximizes skewness and is thus a convex function of market returns, giving more weight to both large negative and positive returns. In particular, "bad" states of nature with the lowest returns get highly overweighted. In contrast, for $\gamma = 4$, skewness is minimized and the SDF is a concave function of the market, reducing the weight to extreme returns compared to the linear case. For the largest returns, which represent "good" states of nature with low marginal utility, the positive constraint is binding and

[11] More precisely, the weight to kurtosis is negative for γ between 1 and 2. Even so, it is very close to zero in the case.

[12] As can be seen in Equation (21), the SDF is a function of the returns of the optimal portfolio of the factors: $\theta_\gamma' f$. Since the CAPM is a one-factor model, θ_γ is only a scaling parameter, such that we can plot the SDF directly as a function of f. In the case of a multi-factor model, the same patterns discussed below will hold, but with the SDF as a function of $\theta_\gamma' f$.

an indicator function sets SDF values to zero.

The shape of the SDF is directly related to the shape of the nonlinearity it adds to m^*. The right panel of Figure 1 depicts, as a function of market returns, the nonlinear term $e_\gamma = m_\gamma - m^*$ for both $\gamma = -2$ and $\gamma = 4$. Nonlinear SDFs minimizing discrepancies with $\gamma < 1$ increase skewness of m^* by adding more weight to extreme returns, while those for $\gamma > 1$ decrease skewness by reducing compensation for both bad and good states of nature. The smaller (greater) the γ below (above) one, more (less) importance is given to extreme returns. In other words, a single parameter γ indexes the particular nonlinearities that minimum discrepancy SDFs add to the linear SDF.

3.3. Restricting the admissible set

The no-arbitrage admissible set of nonlinear SDFs pricing a given set of factors is very large. This is such that one needs to impose additional structure to study how nonlinearities affect pricing performance. We propose to restrict the admissible set to SDFs minimizing Cressie-Read discrepancies. These SDFs satisfy a number of important properties. More specifically, as shown in the previous subsection, they embed the minimum variance SDF as a particular case; impose no-arbitrage by construction; are associated with economically meaningful preferences; and are indexed by a single parameter γ that controls the nonlinear term added to the benchmark linear SDF. The latter property allows us to look at pricing performance as a function of γ and interpret it in light of the nonlinearities this parameter represents.

In principle, the parameter γ coves the whole real line. However, Almeida and Freire (2022) show that there is no solution to the minimum discrepancy problem when $\gamma \to -\infty$ or $\gamma \to \infty$. This is because for large negative and positive γ's distortions become too extreme to still be able to satisfy the pricing restrictions for the basis assets (in our case, the factors f). In fact, they show that solution exists within an interval $[\underline{\gamma}, \overline{\gamma}]$ and provide an algorithm to find this set. This interval depends on the basis assets under consideration. In our empirical analysis, to be able to compare different models on the same basis, we consider a fixed interval between $\underline{\gamma} = -3$ and $\overline{\gamma} = 30$ that guarantees solution for all the factor models we analyze.[13] Our results do not depend on this particular choice as this interval is broad enough to capture the main effects of nonlinearities. In practice,

[13] The absolute value of $\overline{\gamma}$ is much higher than that of $\underline{\gamma}$ because the set of solutions for negative γ's is smaller as they enforce distortions that are more extreme than positive γ's.

for each factor model f, we estimate minimum discrepancy SDFs pricing f indexed by $\gamma \in [-3, 30]$, with a grid with spacing of 1.

4. Empirical analysis

In this section, we describe the factor models and test assets we consider in our analysis and discuss the empirical results. More specifically, we first analyze how the nonlinear models compare to the linear model for a given factor model f, i.e., we study the implications of nonlinearities for absolute pricing performance. Then, we investigate the implications for model comparison by examining how the best nonlinear model of factors f_1 compares to that of f_2 and contrasting that with the relative performance under the linear case.

4.1. Factor models and test assets

We consider 10 factor models in total, ranging from more classical models to recent specifications proposed by the literature. The first model is the seminal CAPM, consisting of the value-weighted market excess return (MKT). The second model, by He, Kelly and Manela (HKM, 2017), adds a financial intermediary capital risk factor (FIRFT) to the market factor. Their motivation is that intermediaries are marginal investors in many markets, such that their financial soundness should be important for asset prices. The third is the Frazzini and Pedersen (2014) model, which adds to the MKT a portfolio long on low-market-beta stocks and short on high-beta stocks (BAB). The economic intuition behind their factor is that constrained investors who cannot use leverage bid up high beta assets, causing those assets to offer lower returns. The fourth factor model, by Daniel, Hirshleifer and Sun (DHS, 2018), augments the market factor with two factors that capture long- and short-horizon mispricing (FIN and PEAD).[14] These factors are based on behavioral theories of investor overconfidence and limited attention.

Fifth, we consider the three-factor model of Fama and French (FF3, 1993), which includes the small-minus-big (SMB) and high-minus-low (HML) factors capturing the size effect and the value effect, respectively. The sixth model is the investment q-factor model (q4) of Hou, Xue and Zhang (2015). Motivated by the neoclassical q-theory of

[14] FIN is a financing factor exploiting underreactions to issuance/repurchase activity. PEAD is based on the post-earnings announcement drift, which reflects delayed price response to information.

investment, they include beyond the market their own size factor (ME), an investment factor (IA) and a profitability factor (ROE). The seventh model is the five-factor model of Fama and French (FF5, 2015), that adds to the FF3 two factors capturing profitability (RMW) and investment (CMA) patterns in stock returns. The eighth is the hedged-FF5 (FF5*) of Daniel et al. (2020) that statistically removes unpriced risk from each of the original FF5 factors. Ninth, we add the momentum factor (UMD) to FF5 to obtain a six-factor model (FF6). The momentum factor is motivated by Carhart (1997). Finally, the last model is the Barrilas and Shanken (BS, 2018) factor model that statistically selects six factors in their analysis. The model includes the market, the q4 investiment and profitability factors, the small-minus-big of FF3, the high-minus-low updated monthly (HMLm) from Asness and Frazzini (2013) and the momentum factor.

Since there is some overlap across the 10 factors models, in the end we have 19 unique factors. Our sample ranges monthly from July 1972 to October 2018, encompassing 556 months. This is the largest sample range for which data on all factors is readily available.[15] Table 1 provides summary statistics for the monthly returns of each factor. All factors have positive average returns. The FIRFT is the factor with the highest premium, but is also the most volatile one. Average returns are all statistically significant at the 5% level, with the exception of the size factors. In fact, SMB, ME and SMB* yield the lowest Sharpe ratios. The highest Sharpe ratios come from the PEAD and BAB factors. Moreover, while the hedged FF5* factors command smaller premiums than their original counterparts, they reduce the volatility substantially, ultimately increasing the t-statistic and the Sharpe ratio.

Figure 2 reports the factor correlations. Alternative versions of the same factor (e.g., SMB and ME, CMA and IA, RMW and ROE, HML and HMLm) are naturally highly correlated. Similarly, each of the FF5 factors has a strong positive correlation with its hedged FF5* counterpart. FIRFT and FIN have the highest positive and highest negative correlations with the market factor, respectively. FIN also correlates substantially with HML, IA, RMW and CMA, while PEAD has very low correlations with other factors. The UMD factor also mostly displays low correlations, with the exception of a strong negative correlation with HMLm. Finally, BAB correlates mildly with FIN and MKT*.

As test assets, we follow the common practice in the recent empirical asset pricing literature of considering anomaly portfolios. We use 44 anomalies from Kozak, Nagel

[15] In Appendix B we describe our data sources.

and Santosh (2020) that are available for the same sample period that the factors. The complete list of anomalies is provided in Appendix B. In unreported tests, we also run our empirical analysis for traditional test assets such as the 25 size/book-to-market portfolios of Fama and French (1993) and industry portfolios. We find that results are qualitatively similar. The entire universe of test assets R_{all} in our analysis consists of the 44 anomalies plus the 19 unique factors, totaling 63 assets. That is, we assess the ability of each model to price not only the basis test assets, but also the factors in the competing models.

4.2. In-sample results

In Section 3, we show how to incorporate nonlinearities into asset pricing tests. The relevant metric of pricing performance and model comparison becomes the maximum squared Sharpe ratio (SR^2) of the mimicking portfolio of the nonlinear SDF pricing the factors of a given model. While we discuss population results in the theory, it is straightforward to implement our approach by using sample analogues. Therefore, for each factor model f, we compute $Sh^2(m_p^*)$, which is simply the SR^2 attainable from the factors, that is, the usual metric $Sh^2(f)$ from traditional tests. Then, for each γ in the grid we consider, we obtain m_γ from (21) and compute the SR^2 of its mimicking portfolio. This analysis uses the whole sample from July 1972 to October 2018.

Figure 3 plots, for each factor model, the SR^2 across γ. The horizontal line depicts the performance of the linear SDF. A clear pattern can be observed across all factor models. For $\gamma < 1$, the SR^2 is always below the linear benchmark and rapidly decreases as γ decreases. In contrast, with the exception of FF3, pricing performance substantially increases for $\gamma > 1$ relative to the linear SDF. In some cases, such as for the CAPM and the BAB, the SR^2 can even double. This is striking as the minimum discrepancy SDFs do not use any information about the test assets in their construction, nor are they optimized to maximize performance across the entire universe of test assets. Instead, such improvement in absolute pricing performance comes solely from economically meaningful nonlinearities in the factors embedded in $m_\gamma(f)$.

To help understand the different patterns in SR^2 across γ, we report in Appendix C how the volatility and mean of the mimicking portfolio of m_γ varies with γ for each factor model. First, the volatility of the mimicking portfolio of all the nonlinear SDFs is above that of the linear SDF, which is natural as the latter is the minimum variance SDF. However, the volatility is much higher for γ's below one, while it is close to the linear case for γ's above one. This is because, as discussed in Section 3.2, for $\gamma < 1$ skewness is maximized, such that the SDF is convex and reaches much higher values for extreme factor returns. This makes the SDF considerably more volatile than the ones for $\gamma > 1$ that minimize skewness and give less weight to extreme returns.

The result above could in principle explain the low SR^2 associated with $\gamma < 1$. However, for $\gamma > 1$, the mean of the SDF mimicking portfolio must be compensating its additional volatility to yield a higher SR^2 than the linear SDF. In fact, we find that this is the case for all factor models except for FF3: for γ's above one, the mean of the mimicking portfolio is more negative than that of the linear SDF. This means that the nonlinearity e_γ covaries positively with the economy-wide SDF, i.e., it offers an additional insurance against systematic risk. In contrast, for $\gamma < 1$, the nonlinear term e_γ has a positive mean, reducing the mimicking portfolio insurance capacity relative to the linear case and contributing even more for a low SR^2.

To assess whether the improvements in SR^2 coming from nonlinearities in Figure 3 are statistically significant, we rely on the asymptotic test for differences in SR^2 from Barillas et al. (2020). More specifically, we consider the version of their test for nontraded factors, which is the case for us since we use an SDF as the asset pricing model. This test takes into account the additional uncertainty involved in computing the mimicking portfolio of a nontraded factor for the calculation of its SR^2. It does not, however, account for the uncertainty of estimating the SDF m_γ itself, which depends on obtaining the parameter θ_γ in (20). Even so, in unreported tests, we verify that a bootstrap test that reestimates the SDF at each bootstrap sample yields very similar results. This suggests that such uncertainty is rather small and we can work with the simpler asymptotic test.

Figure 4 plots, for each factor model and each γ, the t-statistics for the difference between the SR^2 of the mimicking portfolio of m_γ and that of the linear SDF m^*. As can be seen, for 7 out of the 9 factor models (excluding FF3), there are statistically significant improvements in SR^2 associated with γ's above one. The exceptions are the HKM and the FF5 factor models. This indicates that the substantial increase in pricing performance by allowing for nonlinearities cannot be explained by estimation uncertainty.

While so far we have analyzed the impact of nonlinearities for absolute pricing performance, we now focus on its implications for comparison across different factor models. Figure 5 depicts, for each factor model, the SR^2 associated with the best nonlinear SDF (within the Cressie-Read family) and the one coming from the linear SDF. Interestingly, nonlinearities substantially improve the performance of the CAPM, imposing a stronger hurdle to beat it. For instance, while the traditional linear approach implies that both the HKM and FF3 models outperform the CAPM, the opposite is true allowing for nonlinearities. This is aligned with the literature showing that nonlinear versions of the CAPM perform well in cross-sectional asset pricing (Harvey and Siddique, 2000; Dittmar, 2002).

Another model that benefits substantially from incorporating nonlinearities is the BAB. Even though it is outperformed by FF5 under the linear metric, its best nonlinear SDF yields a SR^2 that is 46% higher than that of the best nonlinear model of FF5. A similar change in ranking happens between q4 and FF6, where the latter becomes the preferred model under nonlinearities. Overall, the best performing factor model is the DHS, followed by BS and FF5*. Since these three factor models benefit similarly from nonlinearities, the ranking between them is the same compared to the linear case.

It is worth noting that for nested factor models like the CAPM and FF3, it is not possible under the traditional linear approach for the nested model to have a higher SR^2 in-sample than the nesting model. This is simply because with the nesting model one has access to more investment opportunities and can find a risk-return trade-off at least as good as the one attainable with the nested factors. The same is not true when nonlinearities are allowed for. In fact, we see that the best nonlinear model of the CAPM outperforms that of the FF3. The reason is that, in contrast to the linear case, the mimicking portfolio of a nonlinear SDF loads on the entire universe of test assets, such that investment opportunities are not nested anymore.

In sum, we show that allowing for nonlinearities has deep implications for asset pricing. For nearly all factor models, absolute pricing performance improves substantially by considering nonlinear SDFs. This means that economically meaningful nonlinear versions of these factor models come closer to spanning the mean-variance frontier. Furthermore, model comparison is also affected, as it is often the case that the ranking between two factor models changes if we incorporate nonlinearities. In the next subsection, we provide further evidence from out-of-sample tests.

4.3. Out-of-sample results

The results in the previous subsection are based on the whole sample period, such that they focus on ex-post maximum Sharpe ratios. In the presence of estimation risk, these Sharpe ratios will be biased upward and differ from what investors can actually attain in practice. In particular, a potential concern is that the documented substantial SR^2 improvements after accounting for nonlinearities would not hold out-of-sample. To address this concern, we follow Fama and French (2018) and Detzel, Novy-Marx and Velikov (2022) by assessing out-of-sample performance with bootstrap simulations.

Following their methodology, we split the 556 months of our sample period into 278 adjacent pairs: months (1, 2), (3, 4), . . . , (555, 556). A simulation run draws a random sample of 278 pairs with replacement.[16] The in-sample (IS) simulation run chooses a month randomly from each pair in the run, reusing the same month if the pair is drawn more than once. We calculate the IS SR^2 for each factor model and γ on that sample of months. We then apply the respective mimicking portfolio weights in the unused months of the simulation pairs to produce the corresponding out-of-sample (OOS) estimates.

Table 2 reports, for each factor model and for the IS and OOS samples, the mean SR^2 across bootstrap runs of both the linear SDF pricing the factors and the best nonlinear SDF. We also report the percentage of times that the best nonlinear SDF outperforms the linear SDF. Results are based on 1,000 bootstrap runs. It is worth noting that the "optimal" γ that yields the highest SR^2 is allowed to vary across bootstrap runs, that is, we view the optimal γ as random. Focusing first on the IS estimates, results are similar to the previous subsection in that the mean SR^2 of the best nonlinear specification is always higher than that of the linear specification. In fact, for all factor models, the nonlinear SDF beats its linear counterpart for the vast majority of the bootstrap runs, ranging from 88.6% to 99.1% of the times.

As for the OOS estimates, results are similar to in-sample. The main difference is that the maximum Sharpe ratios are generally smaller, which is expected given the upward bias of IS estimates. The mean SR^2 of the best nonlinear SDF is again always higher than that of the linear SDF, but the improvement is relatively smaller compared to in-sample. One potential reason is that the mimicking portfolio of the nonlinear SDF loads on the whole universe of assets, being thus exposed to more estimation uncertainty. Even so, this additional uncertainty is small compared to the overall improvement in performance. The nonlinear SDF still outperforms the linear benchmark in the vast majority of the simulation runs, ranging from 75.6% to 99.5% of the times.

We next focus on the implications of nonlinearities for model comparison. Table 3 shows, for each factor model and for the IS and OOS samples, the mean rank of its

[16] The paired observation approach is meant to guarantee that effects of parameter nonstationarity are similar in-sample and out-of-sample.

linear SDF among the other linear models and the mean rank of its best nonlinear SDF among the other candidate nonlinear models. We also report the percentage of bootstrap runs that the specification of a given factor model was chosen as the best in terms of SR^2. Starting with the IS estimates, under the traditional linear approach only three models are ever classified as best across simulation runs: BS, DHS and FF5*. In contrast, comparison across nonlinear models is more democratic, where q4, BAB and FF6 are more competitive and together beat the other models more than 10% of the times. Allowing for nonlinearities, there is also a clearer distinction between the best factor model overall (BS) and the second best (DHS).

Focusing now on the OOS estimates, nonlinearities again matter for relative performance. First, the CAPM is more competitive than in the linear case, and is even be the best model for a few of the bootstrap runs. In particular, it beats the HKM model, which does not happen in the standard approach. Second, BS is the best model most often than FF5* under the nonlinear specification, while the opposite happens under the linear specification. Third, again model comparison is more democratic based on the best nonlinear SDF of each factor model. The best model overall is now the DHS, rather than BS as in the IS test. Arguably, since DHS is a three-factor model, it is less subject to estimation uncertainty than the six-factor model of BS.

We also consider an alternative out-of-sample exercise that is informative about the time-series performance of the candidate models. For each factor model, we first estimate the mimicking portfolio of the linear SDF pricing the factors using the entire past history of returns. Then, we keep the portfolio weights in the next month to compute the out-of-sample return of selling the mimicking portfolio.[17] We repeat this procedure until the whole sample is exhausted, where we require an estimation window of at least 30 years. For each factor model, we follow the same procedure to compute the out-of-sample returns of selling the mimicking portfolio of the best nonlinear SDF pricing the factors. The best nonlinear SDF is the one that yields the highest SR^2 in each estimation window.

Figure 6 plots, for each factor model, the cumulative return over the out-of-sample window of the mimicking portfolio of the best nonlinear SDF and the linear SDF. Cumulative returns are directly comparable and reflect risk-adjusted performance as all portfolio returns have been scaled to have the same volatility as the market. As can be seen, the

[17] The mimicking portfolio of an SDF has negative mean as it provides insurance against systematic risk, such that one needs to sell it to get a risk premium. This is easily seen in the linear case where the mimicking portfolio of the linear SDF is minus the tangency portfolio (see Equations 10 and 11).

nonlinear SDF outperforms the linear SDF across all factor models. While the difference is small for HKM, FF3 and q4, the improvement is substantial for the remaining seven models. The cumulative return when incorporating nonlinearities can be up to twice as high as the linear approach, as observed for BAB. Interestingly, the cumulative return associated with the nonlinear SDF is nearly always above that of the linear SDF over time. For most factor models, the outperformance of the nonlinear approach appears to be more acute after the 2008 financial crisis. Overall, this provides further evidence that nonlinearities consistently improve upon the linear case.

In terms of model comparison, nonlinearities also lead to different conclusions relative to the traditional approach. Figure 6 shows that the best performing linear model in the out-of-sample period is FF5*. In contrast, under the best nonlinear SDF, the factor model with highest cumulative return is BAB. The DHS comes in third place. In particular, even though the linear SDF of DHS is almost tied with that of BS, the nonlinear specification of the former factor model delivers much stronger performance than that of the latter. Moreover, while q4 outperforms FF5 and FF6 under the linear metric, the opposite is true when taking nonlinearities into account. Finally, the CAPM is preferred over HKM and FF3 under both the linear and nonlinear SDFs.

In sum, we show that the substantial improvements in maximum Sharpe ratios associated with nonlinearities are robust to estimation uncertainty and hold out-of-sample, both in bootstrap and time-series tests. The implications of nonlinearities for model comparison are also not affected by estimation risk and are similar in-sample and out-of-sample.

5. Conclusion

We provide a general approach allowing for nonlinearities in asset pricing tests. Traditional regression-based tests can be seen as a particular case of our method that uses the linear SDF pricing a set of factors as the asset pricing model. We propose to use a comprehensive family of nonlinear SDFs pricing the model factors. This family naturally generalizes the linear case and is economically meaningful. Empirically, we investigate the implications of nonlinearities for both the absolute and relative pricing performance of a number of leading factor models. We find that, for nearly all factor models, their nonlinear versions significantly improve upon the linear specification. Furthermore, nonlinearities also affect relative model comparison and often lead to different rankings between models compared to the linear case. Overall, our analysis provides extensive evidence that nonlinearities have deep implications for asset pricing.

A. A general approach for nontraded factors

In this appendix, we take as given a model of nontraded factors f. In this case, our framework can be seen as a four-step procedure. First, the mimicking portfolios of the nontraded factors f_p are calculated by projecting each factor onto the returns of the universe of test assets R_{all}. Second, we identify an SDF m that prices f_p:

$$\mathbb{E}(mf_p) = 0. \tag{A.1}$$

Then, we run the two-step GLS CSR using the SDF m as a single nontraded factor to obtain the pricing errors α. From Equation (6), the following holds:

$$\alpha' V^{-1} \alpha = Sh^2(R_{all}) - Sh^2(m_p), \tag{A.2}$$

where m_p is the mimicking portfolio of the SDF.

We now show that the traditional GLS CSR in Section 2 is the particular case of our approach that uses in the second step the linear SDF that prices f_p:

$$m^* = 1 - b'[f_p - \mu_{f_p}], \quad b = \Sigma_{f_p}^{-1} \mu_{f_p}. \tag{A.3}$$

The SDF above is a linear function of the portfolio of the factors mimicking portfolios $b' f_p$ with maximum squared Sharpe ratio. Therefore, if we run the GLS CSR on m^* and analyze the quadratic form in the pricing errors alphas, we have:

$$\alpha' V^{-1} \alpha = Sh^2(R_{all}) - Sh^2(m_p^*) = Sh^2(R_{all}) - Sh^2(-b' f_p) = Sh^2(R_{all}) - Sh^2(f_p). \tag{A.4}$$

The second equality stems from the fact that the mimicking portfolio of m^* recovers precisely the portfolio $-b' f_p$ since m^* is linear in the mimicking portfolios of the factors.[18] The third equality holds because $b' f_p$ is already the portfolio of the factors mimicking portfolios that yields the maximum squared Sharpe ratio $Sh^2(f_p)$. That is, the GLS CSR approach on the nontraded factors f is equivalent to using the linear SDF pricing f_p in the GLS CSR.

[18] First, note that $f_{p,t} = A_p R_t$, where A_p is obtained from the regression $f_t = a_p + A_p R_t + u_t$. This is such that $1 - b'[f_{p,t} - \mu_{fp}] = 1 - b'[A_p R_t - \mu_{fp}]$. Then, in the regression $1 - b'[A_p R_t - \mu_{f_p}] = a + AR_t + \eta_t$, A is equal to $-b' A_p$, while $a = 1 + b' \mu_{f_p}$.

B. Data sources

Below we describe the sources for the data on the factors and anomalies used in our empirical analysis.

- Market factor, FF3, FF5 and FF6: Kenneth French's Data Library (`https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`).
- HKM intermediary risk factor: Asaf Manela's website (`https://apps.olin.wustl.edu/faculty/manela/data.html`).
- BAB factor: AQR website (`https://www.aqr.com/Insights/Datasets/Betting-Against-Beta-Equity-Factors-Monthly`).
- DHS behavioral factors, FF5*: Kent Daniel's website (`http://www.kentdaniel.net/data.php`).
- q4 factors: Authors' website (`https://global-q.org/factors.html`).
- High-minus-low factor updated monthly: AQR website (`https://www.aqr.com/Insights/Datasets/The-Devil-in-HMLs-Details-Factors-Monthly`).
- 44 anomalies: Serhiy Kozak's website (`https://sites.google.com/site/serhiykozak/data`).

The complete list of anomalies included in our analysis is: size, value, prof, dur, valprof, fscore, nissa, accruals, growth, aturnover, gmargins, divp, ep, cfp, noa, inv, invcap, igrowth, sgrowth, lev, roaa, roea, sp, divg, mom, indmom, valmom, valmomprof, mom12, momrev, lrrev, valuem, nissm, strev, ivol, betaarb, season, indrrev, indrrevlv, indmomrev, ciss, price, age and shvol. For their definitions and original papers, see Kozak, Nagel and Santosh (2020).

C. Additional empirical results

In this section, we report additional results complementing the main empirical analysis. In Figure C.1, we report, for each factor model, the standard deviation of the returns of the mimicking portfolio of each nonlinear SDF pricing the factors. The horizontal line depicts the corresponding volatility for the linear SDF. Figure C.2 contains analogous plots, but showing the mean of the mimicking portfolio instead of its volatility.

References

Almeida, C., Freire, G., 2022. Pricing of index options in incomplete markets. Journal of Financial Economics 144, 174-205.

Almeida, C., Garcia, R., 2012. Assessing misspecified asset pricing models with empirical likelihood estimators. Journal of Econometrics 170, 519–537.

Almeida, C., Garcia, R., 2017. Economic implications of nonlinear pricing kernels. Management Science 63, 3147-3529.

Alvarez, F., Jermann, U. J., 2005. Using asset prices to measure the persistence of the marginal utility of wealth. Econometrica 73, 1977-2016.

Asness, C., Frazzini, A., 2013. The devil in HML's details. Journal of Portfolio Management 39, 49–68.

Backus, D., Chernov, M., Zin, S., 2014. Sources of entropy in representative agent models. The Journal of Finance 69, 51–99.

Bansal, R., Lehmann, B. N., 1997. Growth-optimal portfolio restrictions on asset pricing models. Macroeconomic Dynamics 1, 333-354.

Bansal, R., Viswanathan, S., 1993. No arbitrage and arbitrage pricing: A new approach. The Journal of Finance 48, 1231–1262.

Barillas, F., Kan, R., Robotti, C., Shanken, J., 2020. Model comparison with sharpe ratios. Journal of Financial and Quantitative Analysis 55, 1840-1874.

Barillas, F., Shanken, J., 2017. Which alpha? The Review of Financial Studies 30, 1316–1338.

Barillas, F., Shanken, J., 2018. Comparing asset pricing models. The Journal of Finance 73, 715–754.

Breeden, D. T., 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. Journal of Financial Economics 7, 265–296.

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

Cerny, A., 2003. Generalized Sharpe ratios and asset pricing in incomplete markets. Review of Finance 7, 191–233.

Chapman, D. A., 1997. Approximating the asset pricing kernel. The Journal of Finance 52, 1383-1410.

Cochrane, J., 2001. Asset pricing. Princeton University Press, Princeton.

Cressie, N., Read, T., 1984. Multinomial goodness-of-fit tests. Journal of the Royal Statistical Society Series B (Methodological) 46, 440-464.

Daniel, K., Hirshleifer, D., Sun, L., 2020. Short- and long-horizon behavioral factors. The Review of Financial Studies 33, 1673–1736.

Daniel, K., Mota, L., Rottke, S., Santos, T., 2020. The cross-section of risk and returns. The Review of Financial Studies 33, 1927–1979.

Detzel, A., Novy-Marx, R., Velikov, M., 2022. Model comparison with transaction costs. The Journal of Finance, forthcoming.

Dittmar, R., 2002. Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. The Journal of Finance 57, 369-403.

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

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

Fama, E. F., French, K. R., 2018. Choosing factors. Journal of Financial Economics 128, 234–252.

Frazzini, A., Pedersen, L., 2014. Betting against beta. Journal of Financial Economics 111, 1–25.

Ghosh, A., Julliard, C., Taylor, A., 2017. What is the consumption-CAPM missing? An information-theoretic framework for the analysis of asset pricing models. The Review of Financial Studies 30, 442–504.

Ghosh, A., Julliard, C., Taylor, A., 2019. An information-theoretic asset pricing model.

Unpublished working paper.

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

Hansen, L. P., Jagannathan, R., 1991. Implications for security market data for models of dynamic economies. Journal of Political Economy 99, 225–262.

Hansen, L. P., Jagannathan, R., 1997. Assessing specification errors in stochastic discount factor models. The Journal of Finance 52, 557-589.

Harvey, C., Liu, Y., Zhu, H., 2016. ... and the cross-section of expected returns. The Review of Financial Studies 29, 5–68.

Harvey, C., Siddique, A., 2000. Conditional skewness in asset pricing tests. The Journal of Finance 55, 1263-1295.

He, Z., Kelly, B., Manela, A., 2017. Intermediary asset pricing: New evidence from many asset classes. Journal of Financial Economics 126, 1–35.

Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: An investment approach. The Review of Financial Studies 28, 650–705.

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

Kan, R., Robotti, C., 2008. Specification tests of asset pricing models using excess returns. Journal of Empirical Finance 15, 816–838.

Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross-section. Journal of Financial Economics 135, 271-292.

Kraus, A., Litzenberger, R. H., 1976. Skewness preference and the valuation of risk assets. The Journal of Finance 31, 1085-1100.

Lewellen, J. W., Nagel, S., Shanken, J., 2010. A skeptical appraisal of asset pricing tests. Journal of Financial Economics 96, 175–194.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.

Liu, Y., 2021. Index option returns and generalized entropy bounds. Journal of Financial Economics 139, 1015-1036.

Sandulescu, M., Trojani, F., Vedolin, A., 2021. Model-free international stochastic discount factors. The Journal of Finance 76, 935-976.

Schneider, P., Wagner, C., Zechner, J., 2020. Low-risk anomalies? The Journal of Finance 75, 2673-2718.

Shanken, J., 1985. Multivariate tests of the zero-beta CAPM. Journal of Financial Economics 14, 327-348.

Sharpe, W. F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance 19, 425–442.

Snow, K., 1991. Diagnosing asset pricing models using the distribution of assets returns. The Journal of Finance 46, 955–883.

Stutzer, M., 1995. A Bayesian approach to diagnosis of asset pricing models. Journal of Econometrics 68, 367–397.

Stutzer, M., 1996. A simple nonparametric approach to derivative security valuation. The Journal of Finance 51, 1633–1652.

Vanden, J. M., 2006. Option coskewness and capital asset pricing. The Review of Financial Studies 19, 1279–1320.

Table 1: **Summary statistics for factor returns**

Factor	Mean (%)	Std (%)	t-stat	SR
MKT	0.55	4.49	2.90	0.42
FIRFT	1.01	6.69	3.55	0.52
BAB	0.89	3.44	6.14	0.90
FIN	0.75	3.83	4.61	0.67
PEAD	0.61	1.87	7.76	1.14
SMB	0.18	3.01	1.44	0.21
HML	0.35	2.91	2.85	0.41
ME	0.25	3.07	1.97	0.29
IA	0.37	1.86	4.70	0.69
ROE	0.53	2.54	4.99	0.73
RMW	0.27	2.30	2.78	0.40
CMA	0.31	1.94	3.86	0.56
MKT*	0.54	3.13	4.08	0.59
SMB*	0.14	1.94	1.69	0.24
HML*	0.24	1.67	3.38	0.49
RMW*	0.24	1.45	4.02	0.59
CMA*	0.23	1.20	4.49	0.65
UMD	0.65	4.36	3.53	0.51
HMLm	0.33	3.55	2.22	0.32

This table reports summary statistics (mean, standard deviation, t-statistic of the mean and annualized Sharpe ratio) for each of the factors in our analysis. The sample ranges from July 1972 to October 2018 (556 months).

Table 2: **Best nonlinear model vs. linear model - Bootstrap analysis**

	In-Sample			Out-of-Sample		
	Mean-SR^2 Lin	Mean-SR^2 Nonlin	% SR^2 Nonlin > SR^2 Lin	Mean-SR^2 Lin	Mean-SR^2 Nonlin	% SR^2 Nonlin > SR^2 Lin
CAPM	0.27	0.82	99.1	0.25	0.63	99.5
HKM	0.42	0.93	97.5	0.24	0.57	97.5
BAB	1.33	2.89	98.0	1.05	1.67	94.8
DHS	3.31	4.32	90.6	2.75	3.15	75.6
FF3	0.76	1.23	88.6	0.41	0.71	91.2
q4	2.36	3.59	95.1	1.68	2.23	87.7
FF5	1.72	2.54	91.8	0.95	1.28	85.1
FF5*	3.13	3.97	88.8	2.16	2.50	77.3
FF6	2.26	3.35	94.1	1.25	1.68	86.1
BS	3.39	4.62	91.3	2.20	2.63	80.0

This table reports statistics of our bootstrap analysis as detailed in Section 4.3 based on 1,000 bootstrap runs. For both the in-sample and out-of-sample samples, and for each factor model, we report the mean SR^2 across bootstrap runs of the linear SDF pricing the factors and of the best nonlinear SDF pricing the factors. We also report the percentage of runs for which the SR^2 of the best nonlinear SDF is higher than that of the linear SDF. The complete sample from which we draw random samples ranges from July, 1972 to October, 2018.

Table 3: **Factor model comparison - Bootstrap analysis**

	In-Sample				Out-of-Sample			
	Linear		Nonlinear		Linear		Nonlinear	
	Mean-Rank	% Best	Mean-Rank	% Best	Mean-Rank	% Best	Mean-Rank	% Best
CAPM	9.99	0	9.23	0	9.16	0	8.62	0.2
HKM	8.86	0	8.98	0	9.06	0	8.78	0
BAB	6.75	0	5.39	3.8	5.93	0.6	5.44	4.3
DHS	2.25	34.6	2.79	28.6	1.71	56.3	2.03	47.1
FF3	7.94	0	8.36	0	8.32	0	8.38	0
q4	4.25	0	4.04	5.2	3.91	1.3	3.85	7.0
FF5	6.05	0	6.15	0.9	6.45	0	6.63	0.1
FF5*	2.58	29.0	3.34	20.3	2.73	22.6	3.14	19.9
FF6	4.36	0	4.48	3.7	5.26	0.1	5.37	1.1
BS	1.92	36.4	2.19	37.5	2.41	19.1	2.73	20.3

This table reports statistics of our bootstrap analysis as detailed in Section 4.3 based on 1,000 bootstrap runs. For both the in-sample and out-of-sample samples, for each factor model, and for both the linear SDF and the best nonlinear SDF of each factor model, we report its mean rank in terms of SR^2 across bootstrap runs and the percentage of times this specification was chosen as the best across factor models. The complete sample from which we draw random samples ranges from July, 1972 to October, 2018.

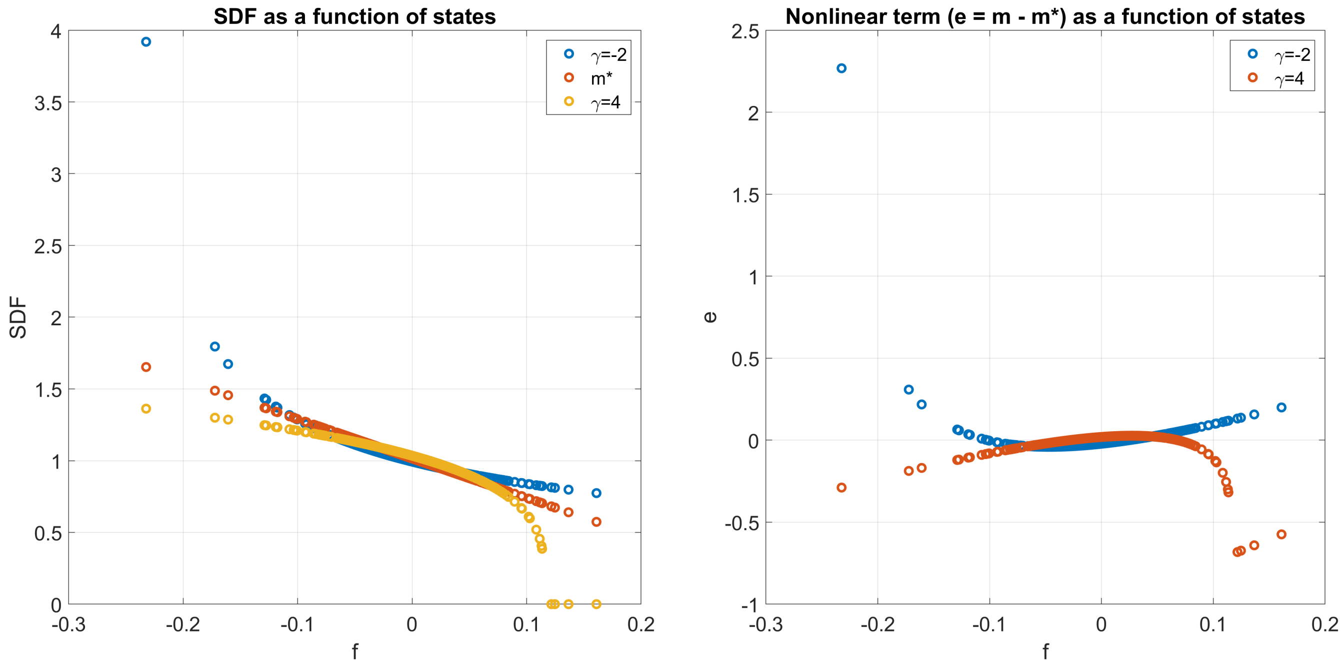

Fig. 1: **CAPM minimum discrepancy SDFs.** This figure plots in the left panel three SDFs pricing the market factor ($\gamma = -2$, m^* and $\gamma = 4$) as a function of the factor returns. The right panel plots the difference between each nonlinear SDF ($\gamma = -2$ and $\gamma = 4$) and the linear SDF m^*. The sample ranges from July, 1972 to October, 2018.

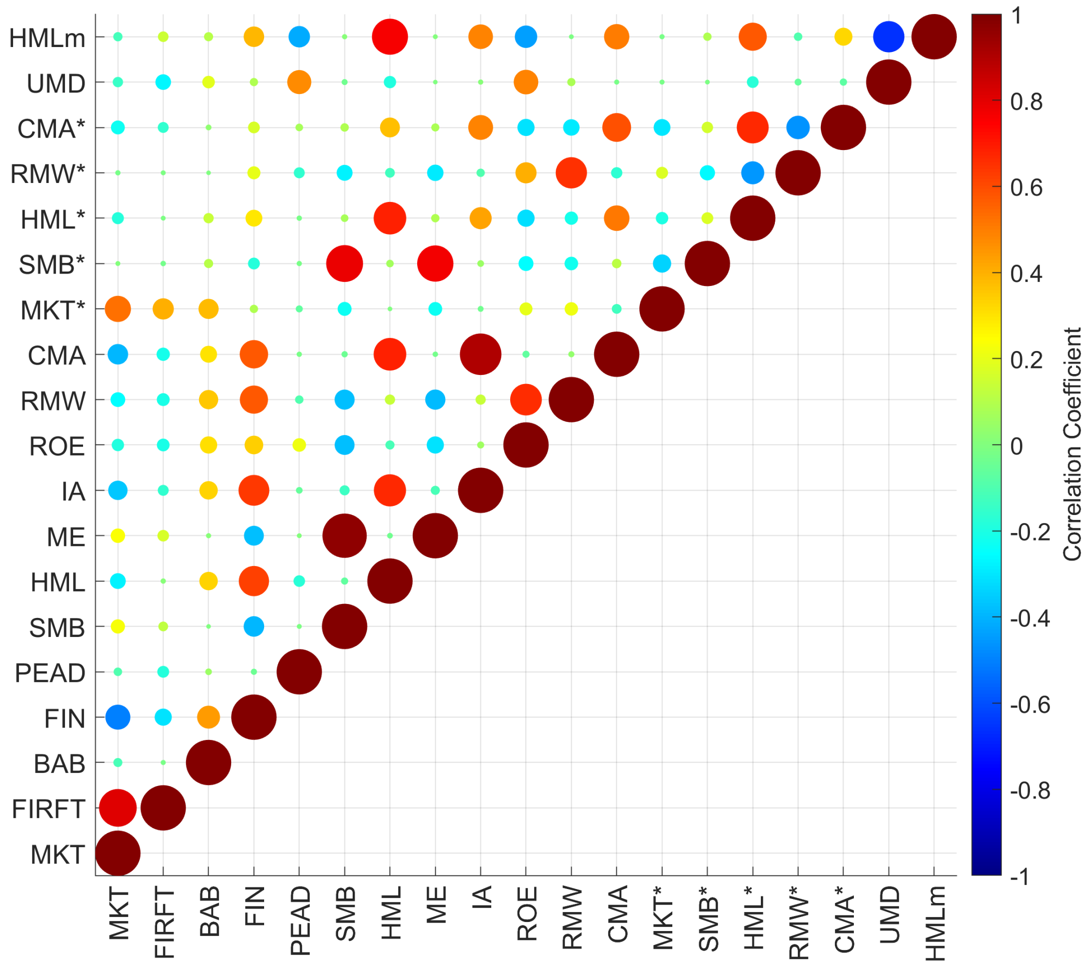

Fig. 2: **Factor correlations.** This figure depicts a heatmap plot of the correlation matrix of the factors. The sample ranges from July, 1972 to October, 2018.

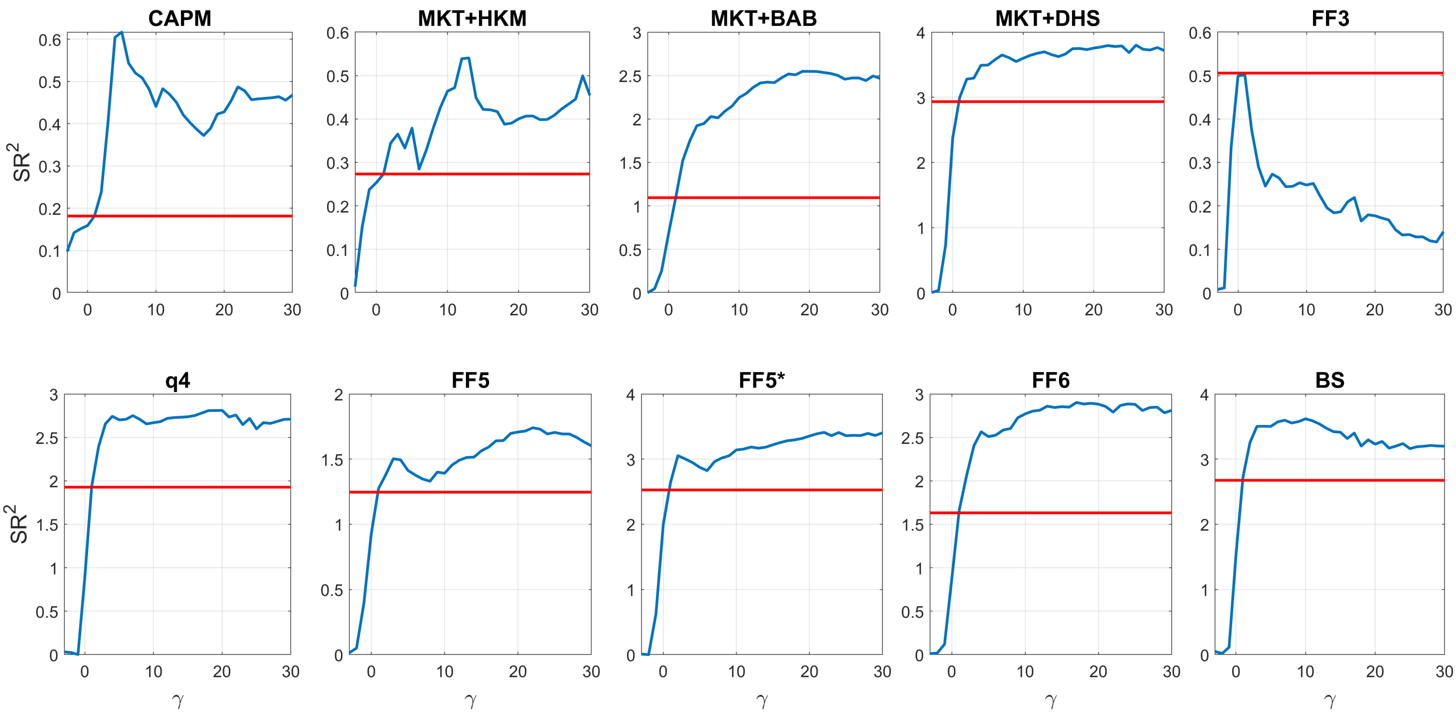

Fig. 3: **Maximum squared Sharpe ratio across γ.** This figure plots, for each factor model, the maximum squared Sharpe ratio (SR^2) across γ in blue. For each γ, the corresponding minimum discrepancy SDF pricing the factor model is obtained. Then, the SR^2 of its mimicking portfolio is reported. We consider $\gamma \in [-3, 30]$, with a grid with spacing of 1. The red horizontal line depicts the SR^2 of the linear SDF. Sharpe ratios are annualized. The sample ranges from July, 1972 to October, 2018.

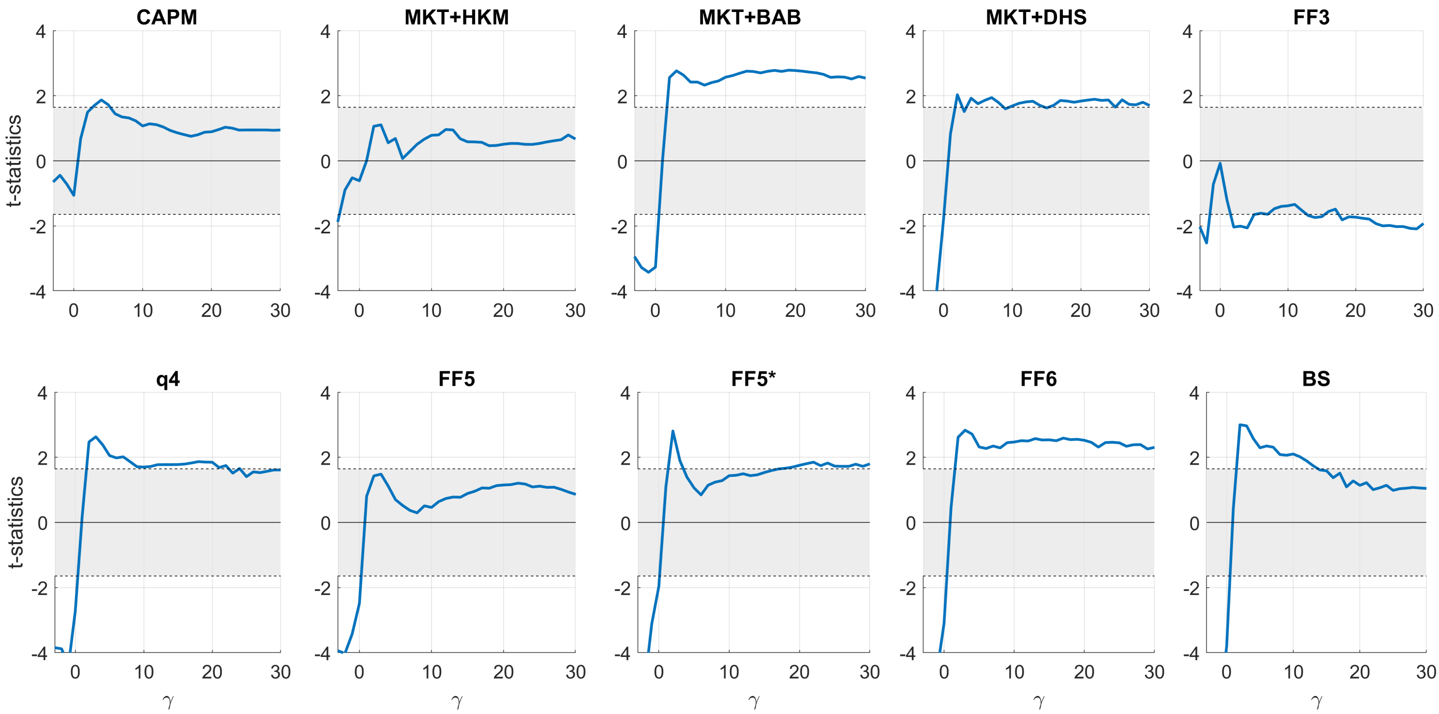

Fig. 4: **Statistical significance of Sharpe ratio difference across γ.** This figure plots, for each factor model f and each γ, the t-statistics for the difference between $Sh^2(m_{\gamma,p})$ and $Sh^2(m_p^*) = Sh^2(f)$. The t-statistics is derived using the asymptotic test of Barillas et al. (2020). We consider $\gamma \in [-3, 30]$, with a grid with spacing of 1. The gray area denotes the region of statistical insignificance at the 5% level. The sample ranges from July, 1972 to October, 2018.

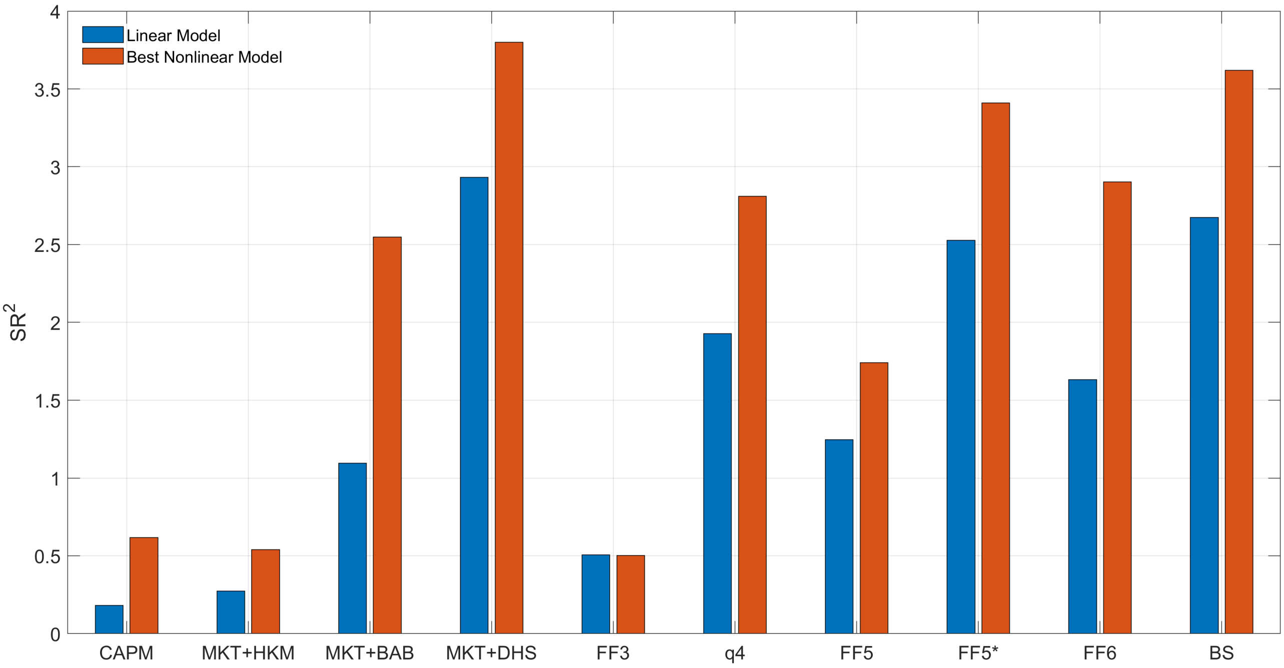

Fig. 5: **Maximum squared Sharpe ratio of linear vs. best nonlinear model.** This figure plots, for each factor model, the maximum squared Sharpe ratio (SR^2) coming from the best nonlinear SDF and from the linear SDF. The best nonlinear model is the one within $\gamma \in [-3, 30]$, with a grid with spacing of 1, that yields the highest SR^2. Sharpe ratios are annualized. The sample ranges from July, 1972 to October, 2018.

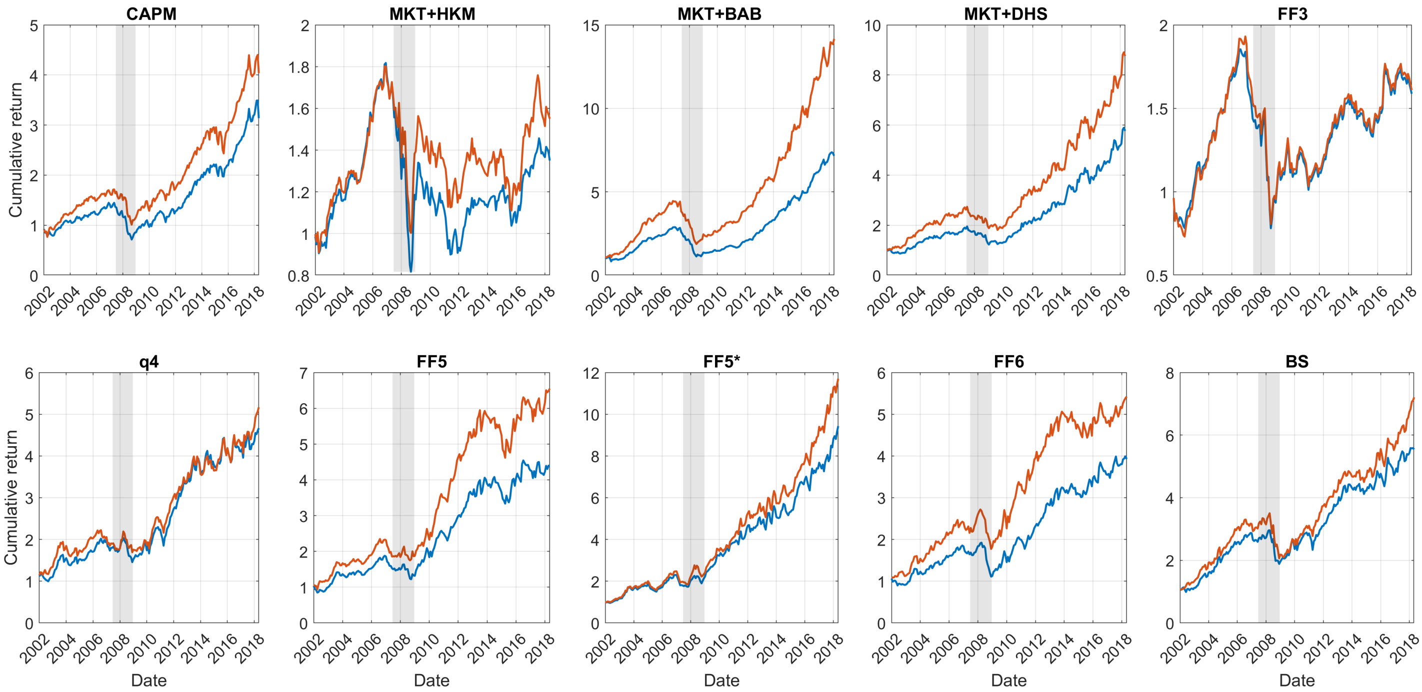

Fig. 6: **Out-of-sample performance of linear vs. best nonlinear model.** This figure plots, for each factor model, the cumulative return of selling the out-of-sample mimicking portfolio of its best nonlinear SDF (in red) and its linear SDF (in blue). We consider an expanding estimation window, with a minimum of thirty years to estimate the models. The best nonlinear model is the one within $\gamma \in [-3, 30]$, with a grid with spacing of 1, that yields the highest SR^2 in the estimation window. The out-of-sample returns of each model have been scaled to have the same volatility as the market, such that plotted cumulative returns reflect risk-adjusted performance and are directly comparable. Shaded areas depict NBER recession dates. The sample ranges from July, 1972 to October, 2018.

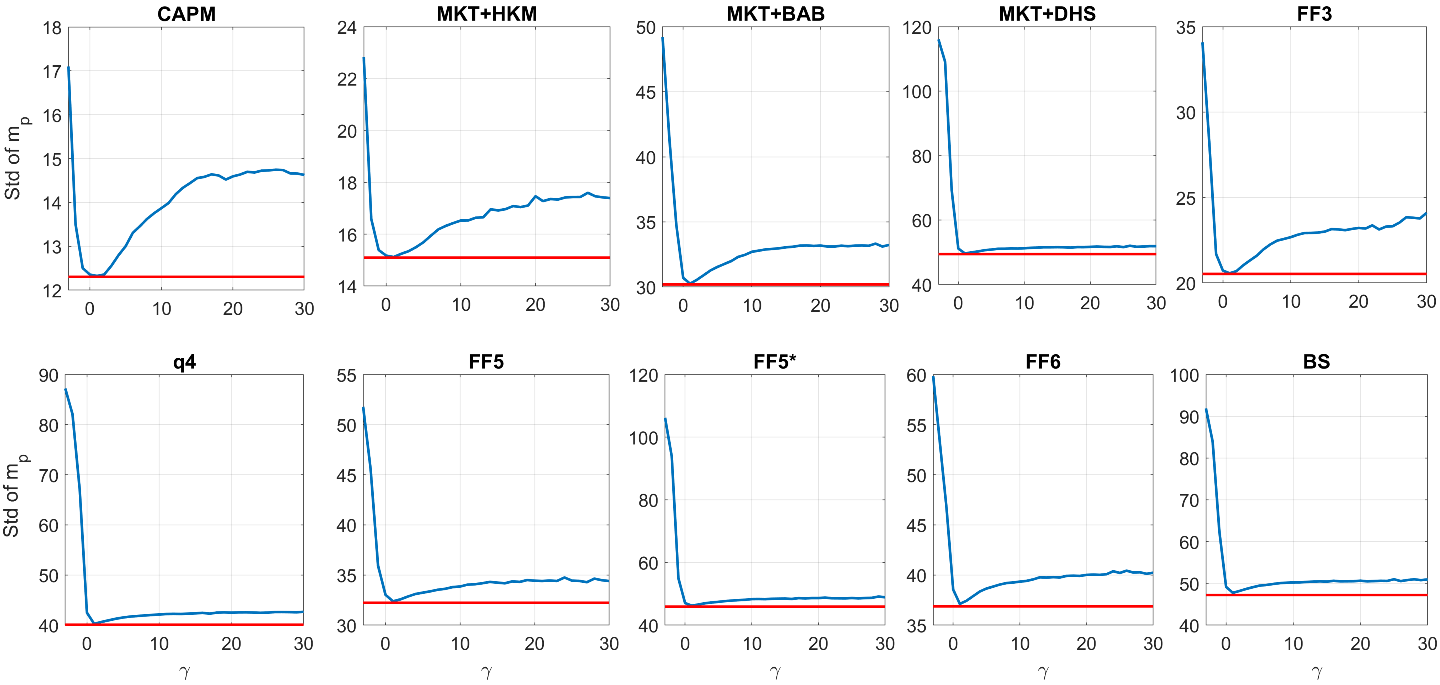

Fig. C.1: **Volatility of SDF mimicking portfolio across γ.** This figure plots, for each factor model, the standard deviation (in %) of the mimicking portfolio of m_γ across γ in blue. For each γ, the corresponding minimum discrepancy SDF pricing the factor model is obtained. Then, the standard deviation of its mimicking portfolio is reported. We consider $\gamma \in [-3, 30]$, with a grid with spacing of 1. The red horizontal line depicts the standard deviation of the mimicking portfolio of the linear SDF. The sample ranges from July, 1972 to October, 2018.

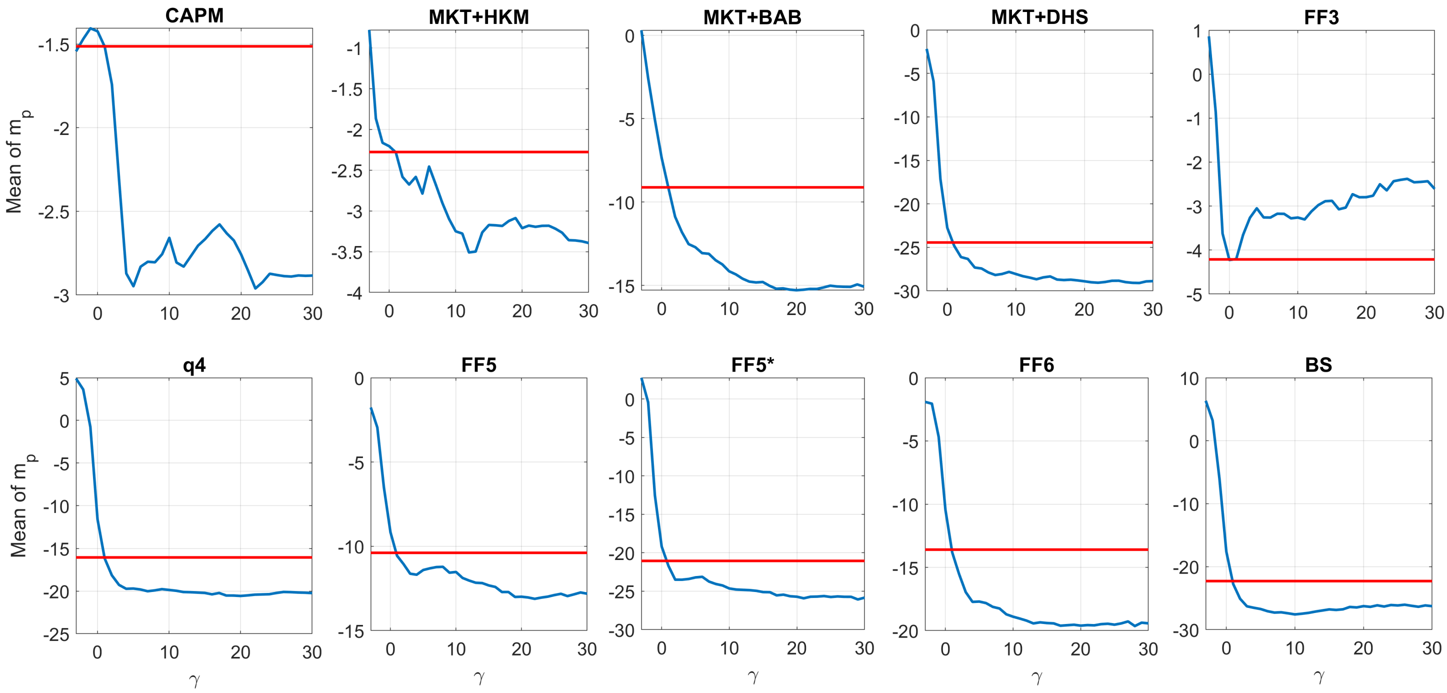

Fig. C.2: **Mean of SDF mimicking portfolio across γ.** This figure plots, for each factor model, the mean (in %) of the mimicking portfolio of m_γ across γ in blue. For each γ, the corresponding minimum discrepancy SDF pricing the factor model is obtained. Then, the mean of its mimicking portfolio is reported. We consider $\gamma \in [-3, 30]$, with a grid with spacing of 1. The red horizontal line depicts the mean of the mimicking portfolio of the linear SDF. The sample ranges from July, 1972 to October, 2018.