Which (Nonlinear) Factor Models?*

Caio Almeida† **Gustavo Freire**‡

This draft: August 12, 2024
First draft: January 20, 2023

Abstract

We show that a nonlinear factor model is closer to the mean-variance frontier the larger the Sharpe ratio of its mimicking portfolio. A linear factor model is a special case for which the mimicking portfolio is the tangency portfolio of the factors. Across a wide range of models, nonlinearities of the factors are priced as they significantly increase the Sharpe ratio metric. The preferred model depends on the test assets, which are relevant for model comparison as they are needed to mimic factor nonlinearities. Momentum anomalies are the most important in the mimicking portfolios, which helps explain why they command a premium.

JEL classification: C52, G11, G12

Keywords: Factor Models, Nonlinearities, Anomalies, Stochastic Discount Factor, Mimicking Portfolio

---

*We would like to thank Svetlana Bryzgalova, Christian Julliard (discussant), Onno Kleen, Erik Kole, Stefan Nagel, Tim Kroencke, Rogerio Oliveira (discussant), Olivier Scaillet (discussant), Alberto Quaini, seminar participants at the Erasmus School of Economics, Princeton University, University of Geneva, Kellogg School of Management, HEC Montreal and University of Luxembourg, and conference participants at the TSE Financial Econometrics Conference 2023 (Toulouse), Research Workshop on Econometric Advances in Macro and Finance (Rotterdam), 15th Annual SoFiE Conference (Seoul), IAAE 2023 Annual Conference (Oslo), 23rd Brazilian Finance Meeting (Sao Paulo), 5th International Workshop in Financial Econometrics (Bahia) and 4th Frontiers of Factor Investing Conference (Lancaster) 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, Tinbergen Institute and Erasmus Research Institute of Management (ERIM).

1. Introduction

Factor models dominate empirical work in asset pricing. They provide tractability by determining observable proxies for the representative agent's unknown marginal rate of substitution, which according to theory (e.g., Lucas 1978), should be the stochastic discount factor (SDF) pricing assets. A seminal example is the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965), which predicts that the SDF should be a function of only one factor: the excess return on the market portfolio. This prediction has been eventually rejected, 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 a plethora of alternative factors beyond the market (Harvey, Liu and Zhang, 2016).

The vast majority of this work imposes that the marginal rate of substitution is linear in the set of factors under consideration. This allows asset pricing tests to be easily implemented through linear regressions, where the metric of pricing performance has a clear economic interpretation. Namely, the linear factor model prices the test assets if and only if the maximum Sharpe ratio attainable from the factors cannot be improved by also trading in the test assets (Gibbons, Ross and Shanken, 1989). In addition, under the linearity assumption, comparing two factor models boils down to comparing the maximum Sharpe ratio attainable from the factors in each model (Barillas and Shanken, 2017). That is, test assets are surprisingly irrelevant for model comparison.

However, enforcing the SDF to be linear in the factors entails important limitations. First, the SDF can attain negative values, which is inconsistent with the marginal utility of a representative agent and the absence of arbitrage. Second, payoffs that are nonlinear functions of the factors cannot be priced by a linear SDF on the factors (Bansal and Viswanathan, 1993). Third, investors care about systematic skewness and kurtosis risk (Harvey and Siddique, 2000; Dittmar, 2002), which is ignored under a linear SDF. In fact, failing to account for preferences for higher moments can give rise to anomalies that would otherwise not exist (Schneider, Wagner and Zechner, 2020). In other words, a rejection of a linear factor model may be a rejection of linearity, rather than a rejection of the factors as proxies for the aggregate marginal rate of substitution.

This raises several important questions. How can we systematically account for nonlinearities in a given factor model? What is the metric of pricing performance under nonlinearities? Are factor nonlinearities priced in the cross-section of returns, i.e., can they significantly improve the explanatory power for the cross-section relative to linear factor models? Do test assets continue to be irrelevant for model comparison? Is it necessary to add more factors beyond the market once we allow for nonlinearities? Can nonlinear factor models achieve the (out-of-sample) mean-variance frontier? In this paper, we aim to answer these questions.

To that end, we extend the linear approach to allow for the SDF to be a nonlinear function of the factors. Our generalization is still easy to implement and has a similar economic interpretation. It relies on first estimating a nonlinear SDF pricing the factors of a given model, and then using this SDF as a single factor in the regression-based tests. The pricing performance metric becomes the Sharpe ratio of the mimicking portfolio of the nonlinear SDF.[1] This framework provides a reason why test assets are irrelevant for model comparison under the linear approach. If the SDF is a linear function of the factors, its mimicking portfolio (i.e., its projection onto the universe of test assets and factor returns) loads only on the factors themselves and has zero weights in the test assets. 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 and factors. That is, the relevance of test assets is restored as they are needed to mimic the nonlinearities.

Knowing how to evaluate a factor model under a nonlinear SDF, the natural question is then *which* nonlinearities to consider. Under no-arbitrage, there exists an infinity of admissible SDFs that price the factors beyond the linear one. Each of these alternative specifications introduce nonlinearities that are irrelevant to price the factors, but that may be relevant to price the extended economy with test assets.[2] While the no-arbitrage set is too large and can contain SDFs that are not economically plausible, focusing on one

[1] It is easy to see why this generalizes the traditional linear approach: if the linear SDF pricing the factors (Hansen and Jagannathan, 1991) is used, its mimicking portfolio is precisely the portfolio of the factors that attains the maximum Sharpe ratio possible when trading in the factors.

[2] In fact, 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 the "true" systematic risk, i.e., it must covary positively with the marginal rate of substitution that prices the whole universe of test assets and factors.

specific nonlinear SDF would require a strong prior on a particular form of nonlinearity. Instead, in the spirit of Cochrane and Saa-Requejo (2000), we propose to restrict the no-arbitrage set based on economic restrictions. From the restricted set, we identify the nonlinear SDF for which the mimicking portfolio has the highest Sharpe ratio.

More specifically, we choose to work with the set of SDFs minimizing Cressie and Read (1984) discrepancy functions, which generalize the variance metric. These SDFs are thus a direct generalization of the linear one and satisfy several desirable properties. First, they are all nonnegative. Second, they map to the marginal utilities of investors with hyperbolic absolute risk aversion (HARA) solving an optimal portfolio problem. As such, they capture a diverse set of preferences for higher order co-moments with an economically meaningful portfolio of the factors. Third, their nonlinearities are indexed by a single parameter, allowing us to interpret how pricing performance depends on it.

Empirically, we analyze the pricing implications of incorporating factor nonlinearities for a wide range of popular models. We consider 10 factor models encompassing 19 unique traded 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 factor nonlinearities are priced as they substantially improve the absolute pricing performance of nearly all factor models considered.[3] In some cases, such as for the CAPM and the BAB, the Sharpe ratio metric can even double. These results are not only economically, but also statistically significant, holding for different sets of test assets. 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

[3]The exception is FF3, for which the linear model is optimal within the family of SDFs we consider.

case, only information on the factors is used, such that improvements come solely from economically meaningful nonlinearities embedded in the SDF. The Cressie-Read SDFs that yield those improvements are usually the ones associated with investors with higher degrees of absolute risk aversion relative to the linear SDF.

Improvements associated with nonlinearities are heterogeneous across factor models. In particular, nonlinearities improve substantially the performance of the CAPM, imposing a stronger hurdle to beat it. We find that the market factor outperforms both the HKM and FF3 when nonlinearities are allowed for the three models, while the opposite happens under the linear specification. This is consistent with previous evidence that nonlinear versions of the CAPM perform well in cross-sectional asset pricing (Harvey and Siddique, 2000) and outperform the FF3 multi-factor model (Dittmar, 2002; Chung, Johnson and Schill, 2006). However, we also show that nonlinearities are not enough to make the CAPM comparable to the other multi-factor models in our analysis, confirming the need to go beyond the market return as the only relevant state variable.

Accounting for nonlinearities leads to different rankings between the remaining factor models as well. Overall, the best performing factor model is the DHS, followed by BS. This is true both under the linear specification and the nonlinear specification for pricing the baseline set of anomalies. However, if we consider different sets of test assets (anomalies from Hou, Xue and Zhang, 2020; the 25 Fama-French size/book-to-market portfolios; or the 49 Fama-French industry portfolios), the opposite conclusion is obtained under the nonlinear case: BS is the best performing model. This highlights the relevance of the test assets once nonlinearities are taken into account and is consistent with the intuitive notion that the preferred model will depend on the set of assets being priced.

Among test assets, we find that momentum strategies are the most important to mimic factor nonlinearities. This provides a potential explanation for the momentum anomaly: it commands a premium because it works as a proxy for priced nonlinearities of systematic sources of risk. In a sense, this helps justify the use of momentum as a factor in linear factor models as a simple way of incorporating factor nonlinearities. On the other hand, low-risk anomalies such as idiosyncratic volatility are among the least important in the mimicking portfolios. The only exception is for the market factor, for which these strategies are especially useful in reproducing nonlinear patterns. This is consistent with the evidence from Schneider, Wagner and Zechner (2020) that low-risk anomalies are closely related to market co-skewness risk.

While our baseline results are out-of-sample in the cross-sectional dimension, in the sense that test assets are not used in the estimation of the nonlinear SDFs, they are based on the whole sample period and thus rely on ex-post Sharpe ratios. To address this concern, we conduct a pricing exercise that is also out-of-sample in the time-series dimension, where we estimate the SDFs and their mimicking portfolios in an expanding training window and compute the out-of-sample portfolio returns. For most of the models, the best nonlinear SDF (chosen ex-ante) continues to deliver higher Sharpe ratios than the linear specification. In particular, the returns of the nonlinear SDF mimicking portfolios are not spanned by those of the linear SDF, while the latter are spanned by the former. Interestingly, some of the factor models are even able to achieve the out-of-sample mean-variance frontier when nonlinearities are contemplated, indicating that the SDF is more likely to be a sparse function of observable factors under nonlinearities.

Finally, we also analyze the implications of nonlinearities for the pricing performance of 17 nontraded factors from Bryzgalova, Huang and Julliard (2023). Results are even more impressive than for traded factors. For 9 out of the 17 models, the squared Sharpe ratio more than doubles under a nonlinear SDF. The performances of the investor sentiment index of Baker and Wurgler (2006) and the dividend yield factor in the nonlinear case are particularly remarkable, as the Sharpe ratios associated with these one-factor models are comparable to those of the best traded multi-factor models we analyze. Furthermore, nonlinearities are especially important for the pricing performance of consumption factors, supporting the idea that higher-order co-moments with consumption are priced.

The remainder of the paper is organized as follows. After a brief review of the related literature, Section 2 summarizes the traditional liner factor model approach. Section 3 presents our extension to allow for nonlinearities, while Section 4 discusses the specific set of nonlinear SDFs we consider. Section 5 contains our empirical analysis on the importance of nonlinearities for the pricing performance of factor models. Section 6 concludes the paper.

1.1. Related literature

Our paper is mainly related to four strands of the literature. The first strand studies asset pricing tests of factor models. 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 pricing errors from cross-sectional regressions, which Kan and Robotti (2008) show is analogous to a modified Hansen and Jagannathan (1997) distance. Kan and Robotti (2009) provide a formal model comparison test using the Hansen-Jagannathan distance. Lewellen, Nagel and Shanken (2010) discuss how to improve empirical tests. Barillas and Shanken (2017) show that, under traditional tests, the preferred model is the one with higher maximum squared Sharpe ratio. Barillas and Shanken (2018) and Bryzgalova, Huang and Julliard (2023) derive Bayesian asset pricing tests, 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 factor models. Kozak and Nagel (2023) discuss the conditions under which different approaches for factor construction span the linear SDF pricing individual stocks. While all these papers focus on linear models, we provide an extension of the standard regression-based tests to allow for nonlinearities in the factors that still has a similar economic interpretation in terms of Sharpe ratios of tradable strategies. Relatedly, Bollerslev, Patton and Quaedvlieg (2024) extend linear factor models to allow factor exposures and risk premia to vary in the return space in cross-sectional regressions. Since our approach conceptually relies on using a nonlinear SDF as the single factor in a cross-sectional regression, both methods could be combined and are thus complementary.

The second strand of the literature proposes nonlinear SDF models for pricing the cross-section of returns. 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 three-moment CAPM 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 the market can explain low-risk anomalies. While these papers motivate our work, we do not share the goal of providing a new nonlinear market model. Instead, we take any set of factors as given and investigate the implications of accounting for nonlinearities in their pricing performance. We show that, for a variety of models, factor nonlinearities are important to explain returns as they substantially improve upon the linear specification.

The third strand of the literature leverages complex machine learning techniques to nonlinearly map the information from a large set of stock characteristics into portfolios entering a linear SDF (Gu, Kelly and Xiu, 2021; Bryzgalova, Pelger and Zhu, 2024; Chen, Pelger and Zhu, 2024; Cong, Feng, He and He, 2023; Fan, Ke, Liao and Neuhierl, 2023; Didisheim, Ke, Kelly and Malamud, 2024). In a sense, we do the opposite: we take a low-dimensional and interpretable set of factors proposed by the literature as given, and obtain from them parsimonious nonlinear SDFs indexed by a single parameter. When projected onto the universe of assets, these SDFs lead to tradable portfolios with Sharpe ratios that are substantially higher than those associated with the linear model. Our approach is not meant as a competitor of the machine learning methods, as we aim to enhance the pricing performance of existing factor models keeping their interpretability, rather than building linear models based on a high-dimensional set of newly constructed factors. Our theoretical motivation is that for any linear factor model that does not achieve the mean-variance frontier, incorporating nonlinearities can bring it closer to the frontier as they might be relevant for pricing the extended economy beyond the factors. Whether applying this idea to the high-dimensional linear SDFs implied by these papers would be fruitful is a question we leave for future research.

The fourth strand uses Cressie-Read discrepancies for different purposes in asset pricing. 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.[4] Stutzer (1996) and Almeida and Freire (2022) analyze the option pricing im-

[4]More specifically, these papers have analyzed whether the pricing kernel of a candidate equilibrium

plications of the minimum entropy SDF and the whole family of Cressie-Read SDFs, respectively. Almeida, Ardison and Garcia (2020) derive performance measures for hedge funds based on minimum discrepancy SDFs. Korsaye, Quaini and Trojani (2019) estimate minimum discrepancy SDFs under general constraints on pricing errors. Ghosh, Julliard and Taylor (2019) show that the minimum entropy SDF estimated from test assets outperforms popular factor models out-of-sample, while Sandulescu, Trojani and Vedolin (2021) examine ratios of minimum entropy SDFs from international markets. Our paper is unique in that we use the whole family of Cressie-Read SDFs obtained from a given set of factors to price the extended economy with test assets. More recently, Almeida, Masini and Schneider (2023) and Sandulescu and Schneider (2024) identify the smallest polynomial modification to a linear factor model to be consistent with no-arbitrage. Each SDF we consider also imposes no-arbitrage, but adds a different nonlinear term to the linear model determined by the maximization of a HARA utility function.

2. Asset pricing tests of linear factor models

In this section, we briefly describe asset pricing tests of linear factor models, which we later extend to allow for nonlinearities. Consider N test 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. Traditional tests evaluate whether the unconditional expected return-beta relation is satisfied:[5]

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

where β is $N \times K$ and contains the covariances between the test 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 characterize pricing errors, or "alphas": $\alpha \equiv \mathbb{E}(R) - \beta\lambda$.

model is able to generate enough dispersion as measured by a particular Cressie-Read loss function.

[5] Conditional asset pricing tests as in, e.g., Lewellen and Nagel (2006), do not fall in this category.

The traditional approach to evaluate 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 test assets in addition to the factors:[6]

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

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 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:

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

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

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.[7] 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.[8]

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).

Considering the alpha mispricing metric associated with the asset pricing tests above, Barillas and Shanken (2017) provide a surprising result for the comparison of linear factor models. 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

[7] Kan and Robotti (2008) show that this test is equivalent to a modified Hansen-Jagannathan distance when the zero-beta rate is constrained to equal the risk-free rate.

[8] 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 $f_t = a + AR_t + \eta_t$.

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

In this section, we extend the traditional linear approach of evaluating factor models to allow for nonlinearities. In this context, we discuss the implications of nonlinearities for the pricing performance of a set of factors. In particular, we provide a sufficient condition for nonlinearities to improve performance relative to the linear case. Then, instead of focusing on a particular form of nonlinearity, we propose to address performance under nonlinearities by considering the highest Sharpe ratio one can obtain from nonlinear SDFs within an economically meaningful set.

3.1. Extending traditional asset pricing tests

To incorporate nonlinearities of factors when evaluating a given model, we propose a simple generalization of the traditional methods described in Section 2. Our approach can be conceptually 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 factors from competing models).

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 using m^* as the single factor 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.[9] 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.

By looking through the lens of the space of SDFs, our framework provides additional insights into the test asset irrelevance result of Barillas and Shanken (2017) for comparing factor models. 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

[9] 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 $a = 1 + b'\mu_f$ and η_t is zero.

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.

While we mostly 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 when the linear SDF pricing the mimicking portfolios of the factors is used.

In sum, we propose to incorporate nonlinearities by using a nonlinear SDF that prices the factors f 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.[10] 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.

3.2. Implications of nonlinearities for pricing performance

We start from the following decomposition that any admissible SDF m that prices the factors f 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 linear function m^* of the factors f does not fully capture the systematic risk in the extended economy with all test assets and factors. In this case, the nonlinearity *can* improve the pricing performance in the extended economy relative to the linear SDF.

[10]This is true under incomplete markets, which is the realistic case where the number of states is larger than the number of assets in a one-period problem.

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^* where the variance of e is small relative to that of m^*, we can write $\sigma^2_{m_p^*} + \sigma^2_{e_p} \approx \sigma^2_{m_p^*}$ and obtain the following approximation:[11]

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

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 portfolios of these assets) and is associated with $Sh^2(R_{all})$.[12] More specifically, from

[11] Such reasonable distortions would be those consistent with good-deal bounds (Cochrane and Saa-Requejo, 2000) or, equivalently, the absence of near-arbitrage opportunities (Kozak, Nagel and Santosh, 2020). Moreover, if σ^2_e is small relative to $\sigma^2_{m^*}$, $\sigma^2_{e_p}$ is even smaller relative to $\sigma^2_{m_p^*}$. This is because m_p^* has the same variance as m^*, as m_p^* is equal to m^* plus a constant, while e_p has a smaller variance than e, as $e_t = a + e_{p,t} + \eta_t$ and $Var(\eta_t)$ is nonzero.

[12] Note that, *for the extended 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}. Note also that m^*_{all} is the projection of the "true" unobserved marginal rate of substitution onto the test assets and factors R_{all}.

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 positively with the extended economy SDF. Arguably, we should expect that any sensible factor model f produces an SDF m^* that covaries positively with the economy-wide SDF. In fact, this is the case empirically for all the factor models we consider in Section 5. 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 and has a small enough variance, 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^*.

3.3. Which nonlinearities?

Given that nonlinearities can have important implications for the pricing performance of factor models, a natural question is then *which* nonlinear SDFs to consider from the no-arbitrage admissible set. On the one hand, this set is too large and may contain SDFs that are not economically meaningful. On the other hand, focusing on one specific nonlinear SDF would require a strong prior on a particular form of nonlinearity. Instead, in the spirit of Cochrane and Saa-Requejo (2000), we propose to restrict the no-arbitrage set based on economic restrictions. From the restricted set, we identify the nonlinear SDF for which the mimicking portfolio has the highest Sharpe ratio.

More specifically, as the metric of pricing performance of a factor model f, we propose to consider the following:

$$\max_{m} \;\; Sh^2(m_p) \qquad \text{s.t. } m \in \mathcal{M}, \tag{18}$$

where m_p is the mimicking portfolio of m when projected onto the universe of assets in the extended economy and $\mathcal{M}$ imposes that the factors in the model are priced, i.e., $\mathbb{E}(mf) = 0$. If $\mathcal{M}$ additionally requires that $m \geq 0$, we obtain the no-arbitrage admissible set. If, instead, $\mathcal{M}$ imposes that the SDF minimizes variance, this set contains only the linear SDF pricing the factors, such that (18) boils down to the traditional metric of linear factor models: $Sh^2(f)$.

Cochrane and Saa-Requejo (2000) apply a similar idea to obtain the maximum (or minimum) price for an option from SDFs pricing the underlying asset returns.[13] They propose to restrict the admissible set by eliminating good-deals, i.e., by avoiding SDFs with high variance that would imply too-high Sharpe ratios in an economy where assets are exactly priced by these SDFs.[14] This approach requires taking a stand on the threshold determining the maximum variance to be allowed for the SDF. Moreover, the solution for the bounds in Cochrane and Saa-Requejo (2000) cannot be directly applied for problem (18). Instead, we propose to work with a set $\mathcal{M}$ defined by an economically meaningful and tractable family of nonlinear SDFs that naturally generalize the linear case, which we discuss in further detail in the next section.

4. Minimum discrepancy SDFs

As we have shown, the traditional linear factor model approach is equivalent to using the linear SDF pricing the factors f as the single factor in asset pricing tests. The linear SDF is the projection of any admissible SDF on the space of factors returns, such that m^* has the minimum variance among all candidate SDFs (see decomposition 12). We propose to use nonlinear SDFs that naturally generalize the minimum variance one. More specifically, we consider as our set $\mathcal{M}$ in (18) SDFs minimizing the Cressie and

[13] In their context, the underlying asset is the "factor", the option is the asset to be priced in the extended economy and the metric of interest is the highest price of the option consistent with the set $\mathcal{M}$. In our case, we take the factors from a model, the universe of test assets and factors as the assets to be priced in the extended economy, and the highest Sharpe ratio obtained from mimicking portfolios of the SDFs within $\mathcal{M}$ as the metric of interest.

[14] Note that the Sharpe ratio bound implied by such a high-variance SDF is different from the Sharpe ratio of its mimicking portfolio, as test assets are not necessarily priced by this SDF. In other words, simply taking SDFs with high variance would not translate into high $Sh^2(m_p)$, but rather in principle smaller $Sh^2(m_p)$ as the variance would hurt the Sharpe ratio of the mimicking portfolio.

Read (1984) family of discrepancies:

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

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.[15] This constraint is important to guarantee that the nonlinear SDFs we identify are consistent with no-arbitrage in the extended economy. This is not necessarily satisfied by m^* as it can reach negative values. Whenever the nonnegativity constraint is not binding, $m_{\gamma=1}$ and m^* are equivalent.

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:

$$\mathbb{E}\left[\phi_\gamma\left(m\right)\right] = \frac{1}{2}\mathbb{E}\left(m-1\right)^2 + \frac{(\gamma-1)}{3!}\mathbb{E}\left(m-1\right)^3 + \frac{(\gamma-1)(\gamma-2)}{4!}\mathbb{E}\left(m-1\right)^4 + \ldots \tag{20}$$

For $\gamma = 1$, the minimum discrepancy problem (19) minimizes variance as higher-order 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, except for γ between 1 and 2.

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 $\gamma < 0$:[16]

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

[16] For $\gamma > 0$, the problem is unconstrained with an indicator function in the objective function:

$$\theta_\gamma = \underset{\{\theta \in R^K : (1-\gamma\theta' f)>0\}}{arg\ max} \ \mathbb{E}\left[-\frac{1}{\gamma+1}\left(1-\gamma\,\theta' f\right)^{\left(\frac{\gamma+1}{\gamma}\right)}\right], \tag{21}$$

where the minimum discrepancy SDF can be recovered from the first-order condition of (21) with respect to θ, evaluated at θ_γ:

$$m_\gamma = \left(1-\gamma\,\theta_\gamma' f\right)^{\frac{1}{\gamma}}. \tag{22}$$

Mathematically, θ_γ is the vector of Lagrange multipliers associated with the Euler equations for the factors in (19). Economically, (21) 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 allocation of wealth in the factors f. We discuss this interpretation in more detail in the next subsection.

4.1. Economic interpretation

Consider a standard optimal portfolio problem for an investor with HARA utility:

$$u^\gamma(W) = -\frac{1}{\gamma+1}(b - a\gamma W)^{\frac{\gamma+1}{\gamma}}, \tag{23}$$

where $a > 0$ and $b - a\gamma W > 0$, which guarantees that the function u^γ is well-defined, concave and strictly increasing. The investor distributes her initial wealth W_0 by investing $\tilde{\theta}$ units of wealth on the factors with excess returns f, such that the end-of-period wealth is given by $W(\tilde{\theta}) = W_0 R_f + \tilde{\theta}' f$, where R_f is the risk-free rate. The optimal allocation is chosen as to maximize expected utility:

$$\tilde{\theta}_\gamma = \max_{\tilde{\theta} \in R^K} \mathbb{E}\left[u^\gamma(W(\tilde{\theta}))\right]. \tag{24}$$

Almeida and Freire (2022) show that there is a one-to-one mapping between problem (24) and the dual problem (21) for a given γ. This is such that the SDF m_γ is proportional to

---

$\mathbb{E}\left[-\frac{1}{\gamma+1}\left(1-\gamma\theta' f\right)^{\left(\frac{\gamma+1}{\gamma}\right)} I_{\Theta_\gamma(f)}(\theta)\right]$, 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: $\mathbb{E}\left[-e^{-\theta' f}\right]$.

the marginal utility of the HARA investor with concavity parameter γ. Moreover, it holds that $\theta_\gamma = \tilde{\theta}_\gamma a/(b - a\gamma W_0 R_f)$, i.e., the optimal Lagrange multipliers θ_γ are proportional to the optimal portfolio weights $\tilde{\theta}_\gamma$. Importantly, the parameter γ also indexes the attitude towards risk of the investor: the higher the γ, the higher the coefficient of absolute risk aversion $-u^{\gamma\prime\prime}(W)/u^{\gamma\prime}(W) = a/(b - \gamma a W)$.

The SDF m_γ is economically meaningful as it reflects the return on the wealth of a risk averse investor who is evaluating whether any asset would add value to her portfolio of factors. While the interpretation is not necessarily that of an equilibrium asset pricing model, m_γ can be mapped to popular models for specific values of γ and specifications of f. To see that, let $W_\gamma = \tilde{\theta}_\gamma' f$ be the endogenous wealth of the investor and Taylor expand the marginal utility $u^{\gamma\prime}(W(\tilde{\theta}))$ around the initial wealth $w_0 = W_0 R_f$ to obtain:[17]

$$u^{\gamma\prime}(W(\tilde{\theta})) = u^{\gamma\prime}(w_0) + u^{\gamma\prime\prime}(w_0) W_\gamma + \frac{1}{2} u^{\gamma\prime\prime\prime}(w_0) W_\gamma^2 + \frac{1}{3!} u^{\gamma\prime\prime\prime\prime}(w_0) W_\gamma^3 + ... \qquad (25)$$

If $\gamma = 1$, the utility function is quadratic and the weights to all nonlinearities are zero, such that the linear SDF characterizing the CAPM is obtained when f equals the market factor. If $\gamma = 1/2$, higher terms beyond W_γ^2 are set to zero, and a three-moment CAPM (Harvey and Siddique, 2000) is obtained. If instead $\gamma = 1/3$, higher terms beyond W_γ^3 are zero and the SDF is consistent with a four-moment CAPM (Dittmar, 2002). In general, however, the minimum discrepancy SDFs will depend on all nonlinearities of the optimal portfolio returns W_γ in different ways. This is desirable given the importance of higher moments other than the third and the fourth to capture investors' preferences towards tail probabilities (Chung, Johnson and Schill, 2006).

Relatedly, the minimum discrepancy SDFs account for a diverse set of preferences towards higher co-moments with the endogenous wealth when evaluating the abnormal performance of a generic asset with excess return R_i. To show that, we Taylor expand the risk-neutralized excess return $(1 - \gamma W_\gamma)^{\frac{1}{\gamma}} R_i$ around $\mathbb{E}[W_\gamma]$ and take expectations:

[17] The derivatives are given by $u^{\gamma\prime}(w_0) = a(b - a\gamma w_0)^{\frac{1}{\gamma}}$, $u^{\gamma\prime\prime}(w_0) = -a^2(b - a\gamma w_0)^{-1+\frac{1}{\gamma}}$, $u^{\gamma\prime\prime\prime}(w_0) = a^3(1-\gamma)(b - a\gamma w_0)^{-2+\frac{1}{\gamma}}$, $u^{\gamma\prime\prime\prime\prime}(w_0) = -a^4(1-\gamma)(1-2\gamma)(b - a\gamma w_0)^{-3+\frac{1}{\gamma}}$ and so on.

$$
\begin{aligned}
\mathbb{E}[(1-\gamma\, W_\gamma)^{\frac{1}{\gamma}}\, R_i] = {} & (1-\gamma\mathbb{E}[W_\gamma])^{\frac{1}{\gamma}}\mathbb{E}[R_i] - (1-\gamma\mathbb{E}[W_\gamma])^{\frac{1-\gamma}{\gamma}}\mathbb{E}[R_i(W_\gamma - \mathbb{E}[W_\gamma])] \\
& + \frac{1}{2}(1-\gamma)(1-\gamma\mathbb{E}[W_\gamma])^{\frac{1-2\gamma}{\gamma}}\mathbb{E}[R_i(W_\gamma - \mathbb{E}[W_\gamma])^2] \\
& - \frac{1}{6}(1-\gamma)(1-2\gamma)(1-\gamma\mathbb{E}[W_\gamma])^{\frac{1-3\gamma}{\gamma}}\mathbb{E}[R_i(W_\gamma - \mathbb{E}[W_\gamma])^3] + ...
\end{aligned} \tag{26}
$$

The expression above reveals that the pricing error $\alpha_i \equiv \mathbb{E}[(1-\gamma\, W_\gamma)^{\frac{1}{\gamma}}\, R_i]$, or the abnormal performance of asset i with respect to the factors, depends on a particular combination of co-moments with W_γ. Since $(1-\gamma W_\gamma)$ is nonnegative by construction, $(1-\gamma\mathbb{E}[W_\gamma])$ is also nonnegative and the signs of the weights given to co-moments are determined only by γ. All SDFs imply a negative weight to the covariance: any expected return that is earned by covarying with W_γ (and thus negatively covarying with the SDF) gets discounted, leading to a smaller α_i.

Preferences for coskewness, on the other hand, depend on whether γ is below or above one. For $\gamma < 1$, investors value assets that offer protection against extreme deviations of wealth from its mean (i.e., asset returns that go up when W_γ is volatile), such that positive coskewness increases α_i. For $\gamma > 1$, assets compensating for intermediate states of wealth are preferred (i.e., returns that go up when W_γ is not volatile), such that negative coskewness increases the α_i. As for cokurtosis, its weight is negative for nearly all γ's (with exception of $1/2 < \gamma < 1$). This means that positive cokurtosis decreases abnormal performance as investors discount asset returns that do not help make the tails of the wealth distribution thinner. In the particular linear case with $\gamma = 1$, higher co-moments are given zero weights and only factor covariance risk is accounted for.

4.2. Illustration

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.[18] Each market return state corresponds to one time-series observation in our data. The minimum

[18] As can be seen in Equation (22), 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 plotted as a function of $\theta_\gamma' f$.

variance SDF m^* is a linear function of the market. Since it never attains negative values for these returns, 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 compared to m^*. 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 given to extreme returns compared to the linear case. For the largest returns, which represent "good" states of nature with low marginal utility, the nonnegativity constraint is binding and 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. This also helps understand the preferences for coskewness embedded in the Cressie-Read SDFs. Investors associated with $\gamma < 1$ have higher marginal utility for extreme factor returns relative to the linear case, such that they value assets with returns that go up during those states. In contrast, for $\gamma > 1$, marginal utility is higher, compared to m^*, for intermediate factor returns, such that asset returns offering protection for those states are preferred.

4.3. 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, which satisfy a number of important properties. More specifically, 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 assess 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.[19] Our results do not depend on this particular choice as this interval is broad enough to capture the main effects of nonlinearities. In practice, 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, and then compute (18) from these SDFs.[20]

4.4. How restrictive is the Cressie-Read family?

As detailed above, the Cressie-Read SDFs are hyperbolic functions of a linear combination of the factors, which can be traced back to the marginal utilities of HARA investors solving a one-period optimal portfolio problem. The HARA class is a considerably large class of risk averse investors. In fact, Almeida and Freire (2022) show that this class comes close to generating the same pricing implications as the entire set of SDFs compatible with risk aversion, with the advantage that the Cressie-Read SDFs are indexed by a single parameter. Therefore, going beyond the set of nonlinear SDFs we propose would essentially mean considering SDFs inconsistent with risk aversion in a one-period problem. In this sense, any improvements we document coming from the Cressie-Read SDFs relative to the linear SDF can be seen as a lower bound to the added value of nonlinearities coming from meaningful preferences.

Even so, it is worth discussing how our approach compares to popular alternatives

[19] 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 those for positive γ's.

[20] This spacing is sufficient as the minimum discrepancy SDFs change continuously with γ.

in the literature for introducing nonlinearities. One such alternative, as in Harvey and Siddique (2000) and Dittmar (2002), has been to Taylor expand the marginal rate of substitution to get the SDF as a polynomial of a state variable, usually the market return. While this can work well for obtaining a nonlinear CAPM, it quickly becomes unfeasible as the number of factors increases and the powers of each factor need to be included. The Cressie-Read SDFs overcome this issue by introducing nonlinearities in a meaningful portfolio of the factors. In fact, as shown in (25), these SDFs can be seen as a polynomial with all powers of the optimal portfolio of the factors in a given model.

One could also be tempted to employ machine learning techniques, such as neural networks as in Bansal and Viswanathan (1993), to estimate a general nonlinear SDF. These methods could be well-suited if one wanted to find a nonlinear function of the factors that maximizes the Sharpe ratio of its mimicking portfolio when projected onto test assets. However, this would require using information from the test assets, thus not fulfilling our goal of extending traditional tests of factor models, where only data on the factors can be used to obtain the SDF. In this case, a neural network is uninteresting: if the criterion is to maximize the Sharpe ratio obtainable from the factors, this is already done by the linear SDF that minimizes variance. Similarly, if the criterion is to minimize different discrepancy functions, this is already achieved by the Cressie-Read SDFs.

5. Empirical analysis

In this section, we describe the factor models and test assets we consider in our analysis and discuss the empirical results. First, we study how the nonlinear models compare to the linear model for a given factor model f, i.e., the implications of nonlinearities for absolute pricing performance. Second, 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. Third, we analyze which test assets are relevant to reproduce factor nonlinearities in the mimicking portfolios. Fourth, we consider a pricing exercise that is out-of-sample not only in the cross-sectional, but also in the time-series dimension.

5.1. *Data on factor models and test assets*

We consider 10 traded 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).[21] 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 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 tenth model is composed of the six factors statistically selected by Barrilas and Shanken (BS, 2018) using a Bayesian method. The model includes the market, the q4 investiment and profitability factors, the small-minus-big of FF3, the high-minus-low

[21] 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.

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.[22] 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 it 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 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 and Santosh (KNS, 2020) that are available for the same sample period as the factors. The complete list of anomalies is provided in Appendix B. The entire universe of test assets R_{all} in our baseline 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. In additional tests, we alternatively consider a different set of 118 anomalies from Hou, Xue and Zhang (2020) and traditional test assets such as the 25 size/book-to-market portfolios of Fama and

[22] In Appendix B, we describe our data sources.

French (1993) and 49 industry portfolios.

We also analyze the pricing performance of 17 nontraded factors from Bryzgalova, Huang and Julliard (2023): LIQNT, the liquidity factor of Pastor and Stambaugh (2003); INTERMCAPRATIO, innovations to the intermediaries' capital ratio of He, Kelly and Manela (2017); measures of financial (FINUNC), real economic activity (REALUNC) and macroeconomic (MACROUNC) uncertainty of Jurado, Ludvigson and Ng (2015); the term spread and default spread, TERM and DEFAULT, respectively; DIV, the dividend yield; UNRATE, unemployment rate; PE, Price-earnings ratio; the investor sentiment measures from Baker and Wurgler (2006) and Huang et al. (2015), BWISENT and HJTZISENT, respectively; the growth rate of nondurable consumption (NONDUR), service expenditure (SERV), industrial production (IPGrowth) and the producer price index for crude petroleum (OIL); and the slope of the yield curve (DeltaSLOPE). The nontraded factors are available from October 1973 to December 2016.

5.2. Baseline analysis

In Section 3, we show how to incorporate nonlinearities into asset pricing tests of factor models. The relevant metric of pricing performance becomes the 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 first compute $Sh^2(m_p^*)$, which is simply the maximum SR^2 attainable from the factors, that is, the usual metric $Sh^2(f)$ from the traditional linear approach. Then, for each γ in the grid we consider, we obtain m_γ from (22), compute the SR^2 of its mimicking portfolio and then take the highest SR^2 over γ. This analysis uses the whole sample from July 1972 to October 2018.

Figure 3 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. As can be seen, nonlinearities substantially improve pricing performance for nearly all factor models considered. 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 pricing performance comes from economically meaningful nonlinearities in the factors embedded in $m_\gamma(f)$. The only exception is the FF3, for which the linear specification is already the optimal one within the set of SDFs we consider.

To assess the statistical significance of the improvements coming from nonlinearities, the left panel of Figure 4 contains the SR^2 difference together with bootstrapped 95% confidence intervals for each factor model.[23] For 7 out of the 9 models (excluding FF3), the SR^2 difference is statistically significant. The right panel of Figure 4 sheds further light on how large are the relative improvements. In some cases, such as for the CAPM and the BAB, the SR^2 can even double relative to the linear specification. Interestingly, the model that benefits the most from nonlinearities is the CAPM, such that there is a stronger hurdle to beat it. Going back to Figure 3, while the linear approach implies that both the HKM and FF3 models outperform the CAPM, the opposite is true when we allow 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; Chung, Johnson and Schill, 2006). However, nonlinearities are not enough to make the CAPM comparable to the other multi-factor models in our analysis, confirming the need to go beyond the market return as the only relevant state variable.

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 can be observed for 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 models benefit similarly from nonlinearities, the ranking between them is the same compared to the linear case. In this sense, such ranking is robust within the Cressie-Read family of nonlinearities when

[23]For each bootstrap sample, we estimate the Cressie-Read SDFs, project them into the universe of assets, compute the squared Sharpe ratios of the mimicking portfolios, and compare the highest Sharpe ratio with the one coming from the linear model in that sample. This takes into account the uncertainty associated with the estimation of the SDFs, the mimicking portfolios and their Sharpe ratios. For a fair comparison, for each bootstrap sample we exclude from the set of Cressie-Read SDFs the $\gamma = 1$ SDF, such that the linear SDF is never a particular case in this set.

considering the KNS anomalies as test assets.

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 ex-post SR^2 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.

We next exploit the fact that the Cressie-Read SDFs are indexed by a single parameter to analyze and interpret which nonlinearities lead to the improvements we document. Figure 5 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, pricing performance increases substantially for $\gamma > 1$ relative to the linear SDF. To help understand that, we report in Figures OA.1 and OA.2 in the Online Appendix 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 4, 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 already 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. Therefore, higher degrees of absolute risk aversion embedded in nonlinear SDFs indexed by $\gamma > 1$ increase the covariance with the true marginal rate of substitution.

Finally, we investigate what are the most important anomalies to mimic factor nonlinearities. To facilitate interpretation, we group the 44 anomalies into seven categories, following similar definitions as Lettau and Pelger (2020): momentum, reversals, value, investment, profitability, frictions and other. Appendix B contains the detailed definitions. As our measure of importance for each anomaly category, we consider the average across all factor models of the maximum t-stat across the anomalies on a given category in the regression of the best nonlinear SDF of the factor model on the universe of test assets. The maximum t-stat within the category is a natural choice as the effect among similar strategies may be dominated by a single anomaly.

Figure 6 reports the results. Overall, momentum strategies are the most important test assets to mimic factor nonlinearities, with an average maximum t-stat across models of 2.35. This provides a potential explanation for the momentum anomaly: it generates a premium because it works as a proxy for nonlinearities of priced sources of risk. The only other category with comparable maximum t-stat is investment. On the other hand, anomalies related to frictions or low-risk, such as idiosyncratic volatility, are among the least important in reproducing factor nonlinearities. This is also illustrated in Figure OA.3 in the Online Appendix, showing the maximum t-stat per category for each model. The exception is for the market factor, where these anomalies are useful in reproducing nonlinear patterns. This is consistent with the evidence from Schneider, Wagner and Zechner (2020) that low-risk anomalies are closely related to market co-skewness risk.

In sum, we show that factor nonlinearities are important to explain the cross-section of returns. For nearly all factor models, squared Sharpe ratios improve substantially by considering nonlinear SDFs. This means that economically meaningful nonlinear versions of these factor models come closer to spanning the mean-variance frontier of the extended economy that includes all factors and test assets. Furthermore, model comparison is also affected, as it is often the case that the ranking between two factor models changes if we incorporate nonlinearities. Among test assets, momentum strategies are the most important to mimic factor nonlinearities. In the next subsection, we provide further evidence from different sets of test assets.

5.3. Different test assets

In this subsection, we conduct our analysis considering three alternative sets of test assets: 118 anomalies from Hou, Xue and Zhang (2020); the 25 size/book-to-market portfolios of Fama and French (1993); 49 industry portfolios; and only the factors themselves (in this case, R_{all} consists only of the 19 unique factors across the 10 factor models). In particular, we investigate whether the patterns of absolute pricing performance across γ observed for the KNS anomalies are similar for other sets of test assets. Moreover, while test assets are irrelevant for model comparison under the traditional linear approach (Barillas and Shanken, 2017), we show in Section 3 that they are relevant in the case of nonlinear SDFs as they are needed to mimic nonlinearities. Therefore, we analyze how relative pricing performance varies with the set of test assets under consideration.

Figure OA.4 depicts, for each factor model and for each set of test assets, the SR^2 across γ. Again, the horizontal line denotes the performance of the linear SDF, which does not depend on the test assets. For most factor models, patterns are similar across the different test assets. Namely, the squared Sharpe ratio is usually below the linear case for $\gamma < 1$, while it is above for $\gamma > 1$. This reinforces our finding that nonlinear SDFs associated with higher degrees of risk aversion tend to increase co-movement with the economy-wide SDF. One interesting exception is the CAPM, for which SDFs associated with γ's below 1 lead to the highest Sharpe ratios for the alternative sets of test assets. This means that, for pricing those sets of assets, accounting for positive coskewness with the market factor is relatively more important than higher absolute risk aversion, which is consistent with Harvey and Siddique (2000). Overall, with the exception of FF3, incorporating nonlinearities often leads to improvements in absolute pricing performance compared to the linear SDF when considering other sets of test assets.

Figure 7 further plots the SR^2 associated with the linear SDF and the best nonlinear SDF for each factor model and each set of test assets. As can be seen, the magnitude of the improvements in pricing performance coming from nonlinearities depends on the set of test assets being priced. In particular, the most important implication is for defining the best performing factor model. While under the linear case and the nonlinear case with KNS anomalies the DHS outperforms BS, the opposite is true for all other sets of test assets, where BS attains the highest SR^2. The SR^2 of the best nonlinear model of BS can be even 35% higher than that of the DHS when pricing the 49 industry portfolios. This highlights the relevance of test assets for model comparison once nonlinearities are contemplated.

5.4. Nontraded factors

In this subsection, we investigate the implications of incorporating nonlinearities when evaluating one-factor models composed of each of the 17 nontraded factors obtained from Bryzgalova, Huang and Julliard (2023). The universe of test assets is the same as that of the baseline analysis, i.e., the 19 traded factors and the 44 KNS anomalies. As we show in Appendix A, for a given SDF specification (linear or nonlinear), the pricing performance metric of a nontraded factor boils down to the squared Sharpe ratio of the mimicking portfolio of the SDF that prices the mimicking portfolio of the nontraded factor.

Figure 8 plots, for each nontraded factor, the squared Sharpe ratio associated with the linear SDF and the best nonlinear SDF within the Cressie-Read family.[24] Similarly to the results for traded factors, nonlinearities substantially improve pricing performance relative to the linear case. In fact, for 9 out of the 17 models, the squared Sharpe ratio more than doubles when nonlinearities are taken into account. There is one important difference, however: for nontraded factors, the mimicking portfolios of both the linear and nonlinear SDF load on all assets, while for traded factors the mimicking portfolio of the linear SDF loads only on the factors in the model. Therefore, the evidence that nonlinearities also improve performance for nontraded factors indicates that the relevance of

[24] In the Online Appendix, Figure OA.5 reports the SR^2 across γ, showing that for most factor models the usual pattern of Sharpe ratios being higher (lower) than the linear case for $\gamma > 1$ ($\gamma < 1$) holds.

nonlinearities in the traded factor case cannot be explained by the fact that the mimicking portfolio of the nonlinear SDF trades on more assets than that of the linear SDF.

Figure 8 further shows that nonlinearities have strong implications for comparing non-traded factors. This is evident from the notable shifts in rankings across different models. For instance, among the 17 models, UNRATE jumps from the 12th position under linearity to the 5th one under the nonlinear case. Moreover, while BWISENT is by far the best overall model under the linear metric, it is matched by DIV when nonlinearities are incorporated. This is because DIV benefits substantially from the nonlinear specification, which is aligned with the fact that dividends are nonlinearly related to returns (see, e.g., Giglio, Kelly and Kozak, 2023). The performances of BWISENT and DIV under the nonlinear case are remarkable, as the Sharpe ratios associated with these one-factor models are comparable to those of the best traded multi-factor models in Figure 3.

Nonlinearities are especially important for the pricing performance of consumption factors: the squared Sharpe ratios of NONDUR and SERV more than triplicate with the best nonlinear SDF. Breeden, Gibbons and Litzenberger (1989) discuss how the traditional consumption-CAPM (CCAPM) relies on a linear approximation that ignores higher-order co-moments with consumption. Our results suggest that such approximation comes with a high cost. In particular, Breeden, Gibbons and Litzenberger (1989) show that the empirical pricing performance of the CCAPM is similar to that of the CAPM. We find the same for the CAPM and SERV under the linear metric. However, under the nonlinear specification, SERV leads to higher Sharpe ratios, supporting the idea of a nonlinear CCAPM.

5.5. Out-of-sample in the time-series

The results we discussed so far are out-of-sample in the cross-sectional dimension, in the sense that we investigate the performance of SDFs obtained from a set of factors in pricing test assets that were not used in the estimation. However, the results 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 this subsection, we address this concern with a pricing exercise that is also out-of-sample in the time-series dimension. For these results, we again consider the baseline set of 10 traded factor models and the 44 anomalies of Kozak, Nagel and Santosh (2020) as test assets.

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.[25] We repeat this procedure until the whole sample is exhausted, where we require an estimation window of at least 20 or 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 (expanding) estimation window.

Table 2 reports, for each factor model, the squared Sharpe ratio of the out-of-sample mimicking portfolios of the best nonlinear SDF and the linear SDF. As would be expected, ex-ante Sharpe ratios are generally smaller than their ex-post counterparts in Figure 3. However, the relative comparison between the nonlinear and linear specifications is similar. With the 20-year expanding estimation window, nonlinearities improve pricing performance out-of-sample for 8 out of the 10 factor models. This improvement is statistically significant at the 10% level or better for half of these models.[26] When considering a 30-year expanding window instead, the nonlinear SDFs always outperform the linear ones, albeit with reduced statistical significance due to a smaller sample.

We also consider spanning regressions of the nonlinear models on the linear models and vice-versa. For both estimation windows, the alpha from regressing the nonlinear SDF out-of-sample mimicking portfolio on the linear one is statistically significant for 6 out of the 10 factor models. In contrast, the alpha from regressing the linear SDF mimicking portfolio on the nonlinear one is virtually always insignificant, that is, the linear specification is spanned by the nonlinear one out-of-sample. These results reinforce the idea that factor nonlinearities are priced, as they are important to explain the cross-

[25] 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).

[26] We assess statistical significance with the Barillas et al. (2022) test for differences in squared Sharpe ratios.

section of returns beyond the information contained in linear factor models.

The factor models in our analysis come from a long tradition in asset pricing attempting to explain cross-sectional variation in returns with a small number of factors that are observable (i.e., defined by sorting on a given observed stock characteristic). The idea that the SDF is a linear function of a few observable factors has been recently questioned by Kozak, Nagel and Santosh (2020). Considering a large number of characteristics-based portfolios, they use model selection techniques to show that an SDF that is a sparse linear function of those portfolios performs worse in pricing the cross-section than a sparse linear SDF on high-variance principal components (PCs) summarizing information from all the portfolios. We now analyze how nonlinearities affect the comparison between low-dimensional observable factor models and low-dimensional latent factor models based on PCs of the test assets, and how they fare relative to the out-of-sample mean-variance efficient (MVE) portfolio. We also include models with risk-premium (RP) PCs following the method of Lettau and Pelger (2020), which extracts latent factors that maximize explained return variation while minimizing cross-sectional pricing errors.

Table 2 contains the out-of-sample SR^2 associated with linear models based on different numbers of PCs (and RPs) and the MVE portfolio, all estimated on the training window.[27] With the 20-year expanding estimation window, the best PC model outperforms 4 of the linear factor models, but only 2 when nonlinearities are incorporated. The RP models improve substantially relative to the PC ones. The only factor model that beats the best RP model is the DHS under nonlinearities. With the 30-year expanding window, the SR^2 associated with the latent factor models are substantially smaller, such that they are consistently outperformed by the nonlinear observable factor models. Two of the factor models (BAB and FF5*) are even able to achieve the out-of-sample MVE frontier under the nonlinear specification, with SR^2 higher than the MVE portfolio. These results suggest that the SDF is more likely to be a low-dimensional function of observable factors if nonlinearities are contemplated.

[27]That is, both the PCs (or RPs) of the universe of assets and the portfolio of PCs (or RPs) with maximum Sharpe ratio for a given number of factors are estimated in the training window and kept fixed in the following month. The MVE portfolio is simply the portfolio of all assets that maximizes Sharpe ratio in the training window, which is kept in the following month to compute its out-of-sample return. We have also considered a regularized MVE portfolio with Ridge and results are very similar.

6. Conclusion

We extend the linear factor model approach to allow for the SDF to be a nonlinear function of the factors. The pricing performance metric of the nonlinear model becomes the Sharpe ratio of the mimicking portfolio of the nonlinear SDF. To investigate the role of factor nonlinearities in explaining the cross-section of returns, we propose to use a family of nonlinear SDFs that generalizes the linear case and is economically meaningful. Empirically, we find that for a wide range of popular factor models, nonlinearities are priced as they significantly improve the Sharpe ratio relative to the linear specification. Furthermore, nonlinearities affect model comparison and often lead to different rankings between models relative to the linear case. The preferred model depends on the test assets, which are relevant for model comparison as they are needed to mimic factor nonlinearities. Among test assets, momentum strategies are the most important to mimic factor nonlinearities, which suggests that momentum generates a premium because it proxies for nonlinearities of priced sources of risk.

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.[28] 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

[28] 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 $1 - b'[f_{p,t} - \mu_{f_p}] = 1 - b'[A_p R_t - \mu_{f_p}]$. 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 $a = 1 + b' \mu_{f_p}$.

approach on the nontraded factors f is equivalent to using the linear SDF pricing f_p in the GLS CSR.

B. Data sources

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

- Market factor, FF3, FF5, FF6, 25 size/book-to-market portfolios and 49 industry portfolios: 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 and FF5*: Kent Daniel's website (`http://www.kentdaniel.net/data.php`).
- q4 factors and 118 anomalies from Hou, Xue and Zhang (2020): 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 from Kozak, Nagel and Santosh (2020): Serhiy Kozak's website (`https://sites.google.com/site/serhiykozak/data`).
- 17 nontraded factors from Bryzgalova, Huang and Julliard (2023): Jiantao Huang's website (`https://sites.google.com/view/jiantaohuang/home`).

The complete list of anomalies from Kozak, Nagel and Santosh (2020) considered in our analysis is (we include those available for our sample period): `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,shvol. For their definitions and original papers where they first appeared, see Kozak, Nagel and Santosh (2020). Following a similar definition as Lettau and Pelger (2020), we groups these anomalies in the following categories:

- momentum: mom,indmom,valmom,valmomprof,mom12,momrev
- reversals: lrrev,strev,indrrev,indrrevlv,indmomrev
- value: value,dur,divp,ep,cfp,sgrowth,lev,sp,divg,valuem
- investment: nissa,accruals,growth,noa,inv,invcap,igrowth,nissm,ciss
- profitability: prof,valprof,fscore,aturnover,gmargins,roaa,roea
- frictions: size,ivol,betaarb,shvol
- other: season,price,age

The complete list of anomalies from Hou, Xue and Zhang (2020) considered in our analysis is (we include those available for our sample period): cim_12,cim_1,cim_6,ile_1,ilr_12,ilr_1,ilr_6,im_12,im_1,im_6,p52w_12,p52w_6,r11_12,r11_1,r11_6,r6_12,r6_1,r6_6,resid11_12,resid11_1,resid11_6,resid6_12,resid6_6,rs_1,sim_12, sim_1,sue_1,sue_6,bmj,bmq_12,bm,cpq_12,cpq_1,cpq_6,cp,dp,dur,ebp,em,epq_12, epq_1,epq_6,ep,ir,rev_12,rev_1,rev_6,spq_12,spq_1,spq_6,sp,vhp,aci,cei,dac, dbe,dcoa,dfin,dfnl,dii,dlno,dnca,dnco,dnoa,dpia,dwc,ia,ig2,ig,ivc,ivg,noa, nsi,oa,pda,poa,pta,ta,ato,cla,cop,cto,droe_12,droe_1,droe_6,eg_12,eg_1,eg_6,gpa,opa,ope,roe_1,roe_6,sgq_1,tbiq_12,tbiq_6,eprd,etl,etr,hs,ioca,oca,ol, r10a,r10n,r15a,r1a,r1n,r20a,r5a,r5n,beta_1,dtv_12,isff_1,ivff_1,me,srev,tv_1. For their definitions and original papers where they first appeared, see Hou, Xue and Zhang (2020).

References

Almeida, C., Ardison, K., Garcia, R., 2020. Nonparametric assessment of hedge fund performance. Journal of Econometrics 214, 349-378.

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.

Almeida, C., Masini, R., Schneider, P., 2023. Constrained Polynomial Likelihood. Journal of Business and Economic Statistics, forthcoming.

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.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. The Journal of Finance 61, 1645–1680.

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.

Bollerslev, T., Patton, A. J., Quaedvlieg, R., 2024. Granular betas and risk premium functions. Working paper, Duke University.

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

Breeden, D. T., Gibbons, M. R., Litzenberger, R. H., 1989. Empirical test of the consumption-oriented CAPM. The Journal of Finance 44, 231-262.

Bryzgalova, S., Huang, J., Julliard, C., 2023. Bayesian solutions for the factor zoo: We just ran two quadrillion models. The Journal of Finance 78, 487-557.

Bryzgalova, S., Pelger, M., Zhu, J., 2024. Forest through the trees: Building cross-sections of stock returns. The Journal of Finance, forthcoming.

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

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

Chen, L., Pelger, M., Zhu, J., 2024. Deep-learning in asset pricing. Management Science, forthcoming.

Chung, Y. P., Johnson, H., Schill, M. J., 2006. Asset pricing when returns are nonnormal: Fama-French factors versus higher-order systematic comoments. Journal of Business 79, 923-940.

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

Cochrane, J., Saa-Requejo, J., 2000. Beyond arbitrage: Good-deal asset price bounds in incomplete markets. Journal of Political Economy 108, 79-119.

Cong, L. W., Feng, G., He, J., He, X., 2023. Asset pricing with panel tree under global split criteria. Working paper, Cornell University.

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.

Didisheim, A., Ke, S., Kelly, B., Malamud, S., 2024. Complexity in factor pricing models. Working paper, Yale University.

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.

Fan, J., Ke, Z., Liao, Y., Neuhierl, A., 2023. Structural deep learning in conditional asset pricing. Working paper, Princeton University.

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.

Giglio, S., Kelly, B., Kozak, S., 2023. Equity term structures without dividend strips data. The Journal of Finance, forthcoming.

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

Gu, S., Kelly, B., Xiu, D., 2021. Autoencoder asset pricing models. Journal of Econometrics 222, 429-450.

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.

Harrison, M., Kreps, D., 1979. Martingales and arbitrage in multi-period securities markets. Journal of Economic Theory 20, 381–408.

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.

Huang, D., Jiang, F., Tu, J., Zhou, G., 2015. Investor sentiment aligned: A powerful predictor of stock returns. The Review of Financial Studies 28, 791–837.

Jurado, K., Ludvigson, S., Ng, S., 2015. Measuring uncertainty. American Economic Review 105, 1177–1215.

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

Kan, R., Robotti, C., 2009. Model comparison using the Hansen-Jagannathan distance. The Review of Financial Studies 22, 3449–3490.

Korsaye, S. A., Quaini, A., Trojani, F., 2019. Smart SDFs. Working paper, University of Geneva.

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

Kozak, S., Nagel, S., 2023. When do cross-sectional asset pricing factors span the stochastic discount factor? Unpublished working paper.

Lettau, M., Pelger, M., 2020. Factors that fit the time series and cross-section of stock returns. The Review of Financial Studies 33, 2274–2325.

Lewellen, J. W., Nagel, S., 2006. The conditional CAPM does not explain asset-pricing anomalies. Journal of Financial Economics 82, 289–314.

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.

Lucas, R., 1978. Asset prices in an exchange economy. Econometrica 46, 1429–1445.

Pastor, L., Stambaugh, R. F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.

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

Sandulescu, M., Schneider, P., 2024. International arbitrage premia. Working paper, University of Michigan.

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: **Out-of-sample pricing performance**

Panel A: 20-year expanding estimation window

	CAPM	HKM	BAB	DHS	FF3	q4	FF5	FF5*	FF6	BS
SR^2 Lin	0.26	0.21	0.76	1.81	0.28	0.97	0.57	1.45	0.81	0.98
SR^2 Nonlin	0.87	0.12	1.50	2.19	0.27	1.11	0.96	1.74	0.93	1.02
SR^2 Diff p-value	0.00	–	0.00	0.06	–	0.21	0.04	0.13	0.21	0.41
Nonlin-Lin α p-value	0.00	0.66	0.00	0.00	0.41	0.08	0.00	0.07	0.18	0.24
Lin-Nonlin α p-value	0.04	0.29	0.18	0.79	0.32	0.82	0.42	0.83	0.98	0.49

	MVE	1PC	1RP	3PC	3RP	5PC	5RP	7PC	7RP
SR^2 Other Models	2.55	0.02	0.54	0.54	1.94	0.73	1.59	0.72	1.60

Panel B: 30-year expanding estimation window

	CAPM	HKM	BAB	DHS	FF3	q4	FF5	FF5*	FF6	BS
SR^2 Lin	0.36	0.04	0.79	0.69	0.08	0.48	0.48	1.08	0.40	0.61
SR^2 Nonlin	0.52	0.07	1.39	1.04	0.09	0.54	0.74	1.28	0.62	0.82
SR^2 Diff p-value	0.16	0.30	0.05	0.09	0.47	0.37	0.11	0.13	0.13	0.21
Nonlin-Lin α p-value	0.05	0.45	0.00	0.04	0.82	0.34	0.03	0.07	0.08	0.14
Lin-Nonlin α p-value	0.73	0.70	0.69	0.41	0.96	0.73	0.71	0.63	0.68	0.97

	MVE	1PC	1RP	3PC	3RP	5PC	5RP	7PC	7RP
SR^2 Other Models	1.26	0.01	0.16	0.01	0.37	0.23	0.35	0.17	0.30

This table reports, in Panel A (Panel B), statistics of our out-of-sample analysis as detailed in Section 5.5 based on a 20-year (30-year) expanding estimation window. For each factor model, we report the annualized squared Sharpe ratio of the out-of-sample mimicking portfolio of the best nonlinear SDF (SR^2 Nonlin) and the linear SDF (SR^2 Lin). The best nonlinear SDF 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. We also report p-values for: (i) a one-sided test (Barillas et al., 2022) of the null hypothesis that the SR^2 difference between the nonlinear and the linear model is below zero against the alternative that it is positive (SR^2 Diff p-value); (ii) the alpha of regressing the nonlinear mimicking portfolio returns on the linear ones (Nonlin-Lin α p-value); (iii) the alpha of regressing the linear mimicking portfolio returns on the nonlinear ones (Lin-Nonlin α p-value). Finally, we report the SR^2 of the out-of-sample MVE portfolio, PC and RP portfolios with 1, 3, 5 and 7 principal components. The universe of test assets is composed of the 19 factors and the 44 anomalies from Kozak, Nagel and Santosh (2020). The sample ranges from August, 1972 to October, 2018.

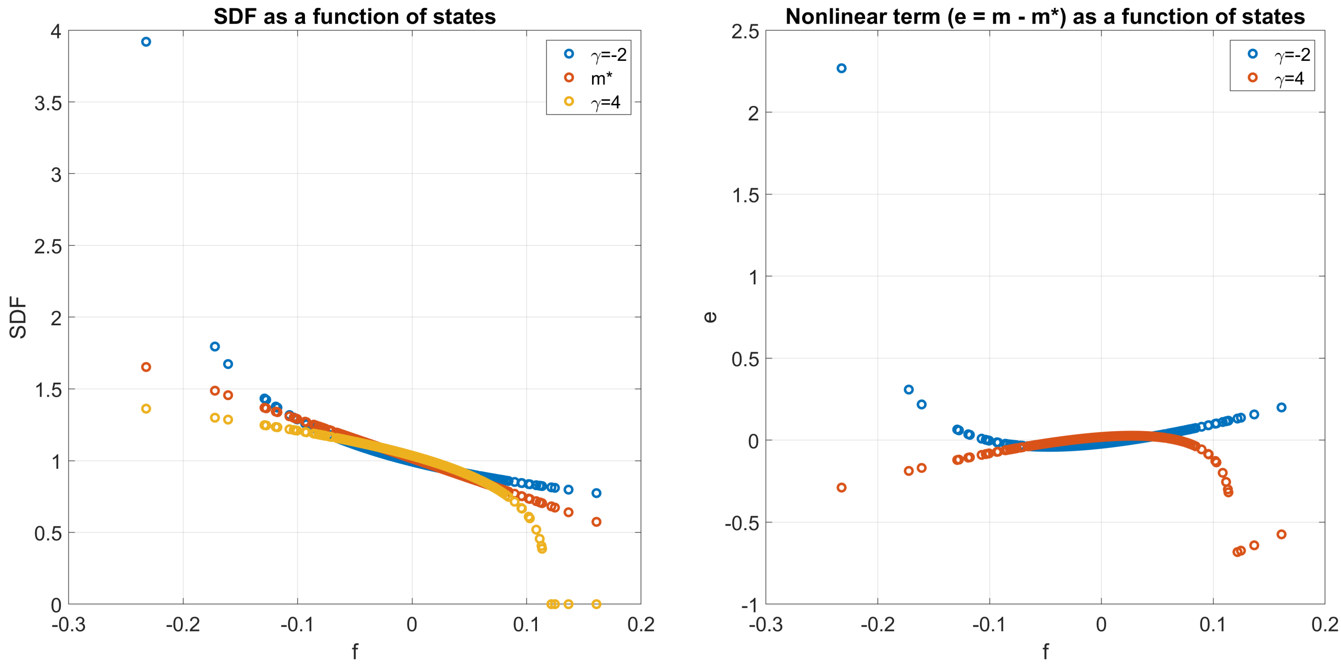

Fig. 1: **CAPM minimum discrepancy SDFs.** This figure plots in the left panel three SDFs pricing the market factor ($\gamma = -2$, m^* and $\gamma = 4$) as a function of the factor returns. The right panel plots the difference between each nonlinear SDF ($\gamma = -2$ and $\gamma = 4$) and the linear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

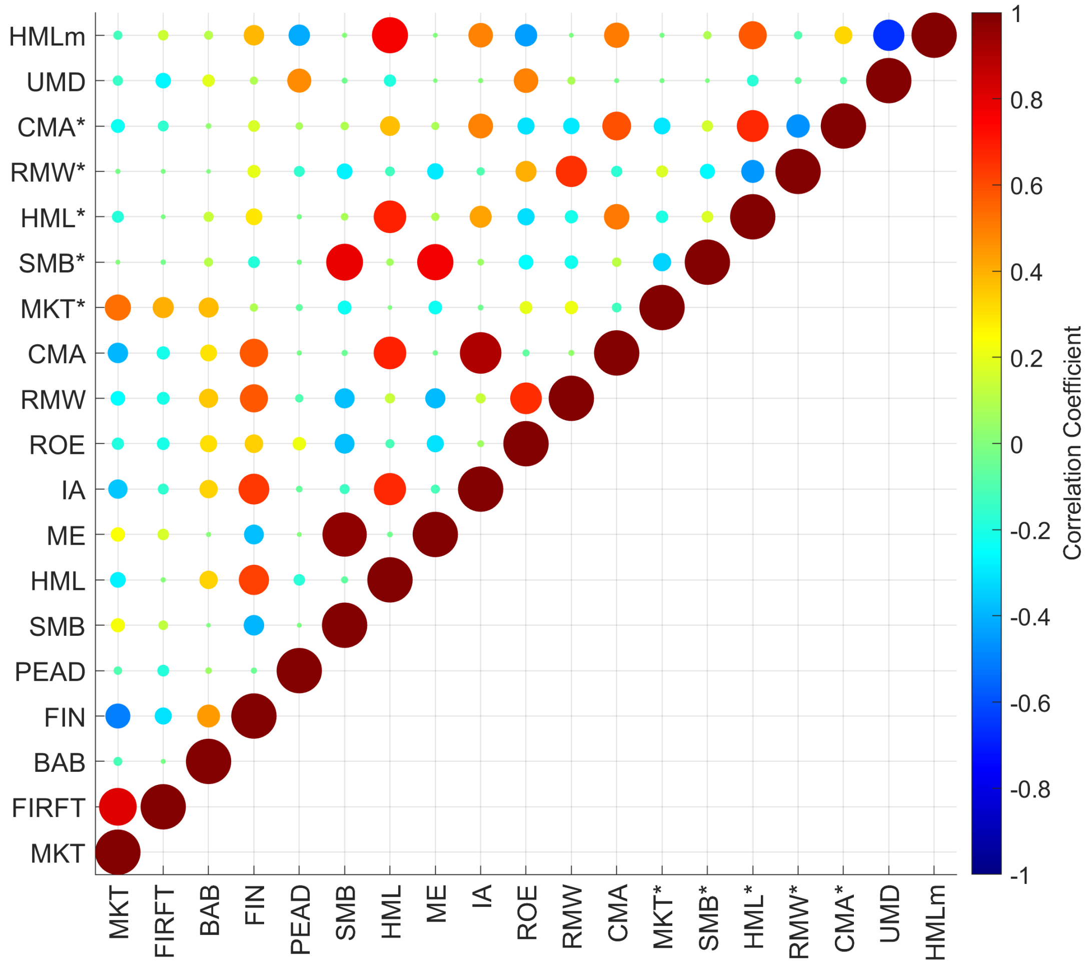

Fig. 2: **Factor correlations.** This figure depicts a heatmap plot of the correlation matrix of the factors. The sample ranges from July, 1972 to October, 2018.

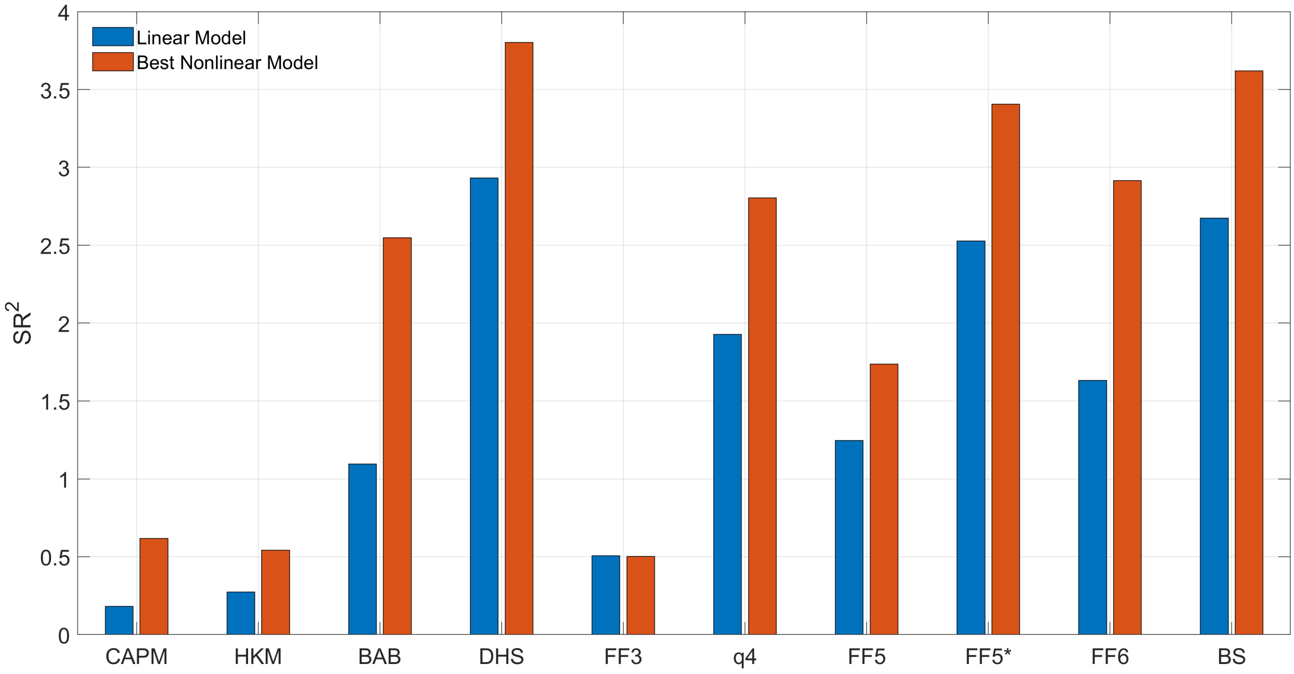

Fig. 3: **Squared Sharpe ratio of linear vs. best nonlinear model.** This figure plots, for each factor model, the squared Sharpe ratio (SR^2) coming from the best nonlinear SDF and from the linear SDF. The best nonlinear model is the one within $\gamma \in [-3, 30]$, with a grid with spacing of 1, that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

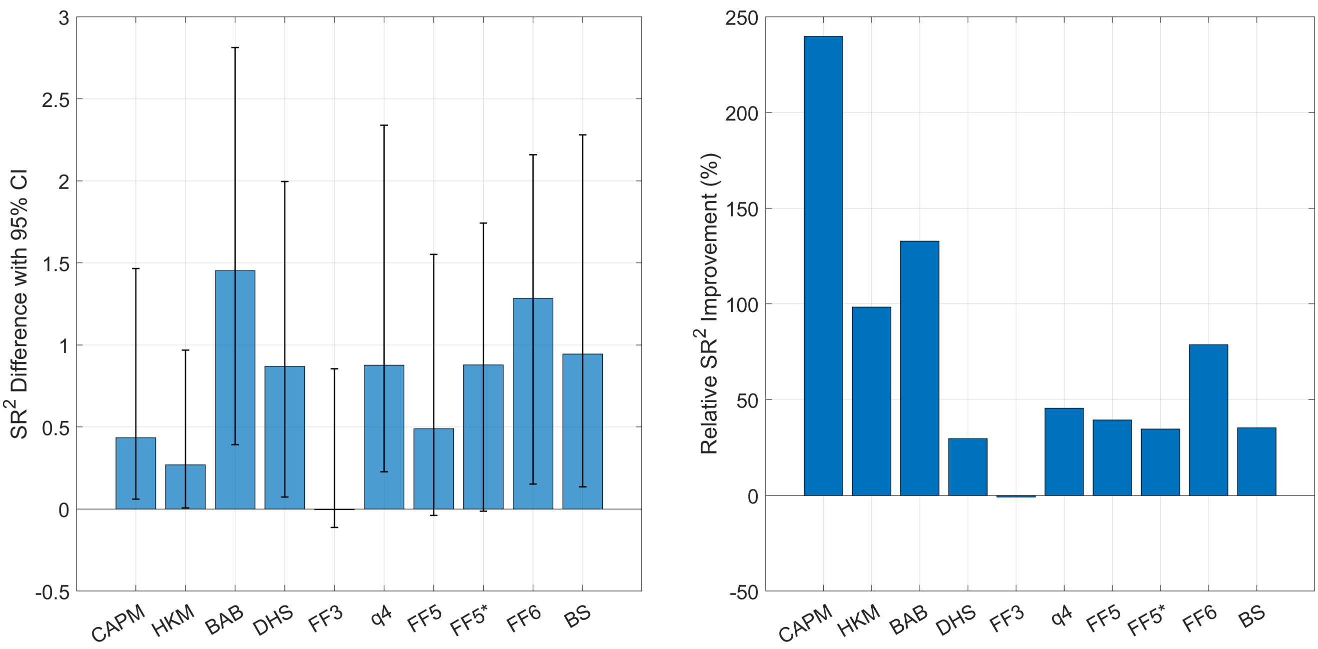

Fig. 4: **Sharpe ratio difference of linear vs. best nonlinear model.** The left panel of this figure plots, for each factor model, the difference between the squared Sharpe ratio (SR^2) coming from the best nonlinear SDF and from the linear SDF, together with bootstrapped 95% confidence intervals based on 1,000 bootstrap samples. 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 right panel plots the relative improvement coming from nonlinearities for each model, that is, the SR^2 ratio of the nonlinear and linear models minus one. The sample ranges from July, 1972 to October, 2018.

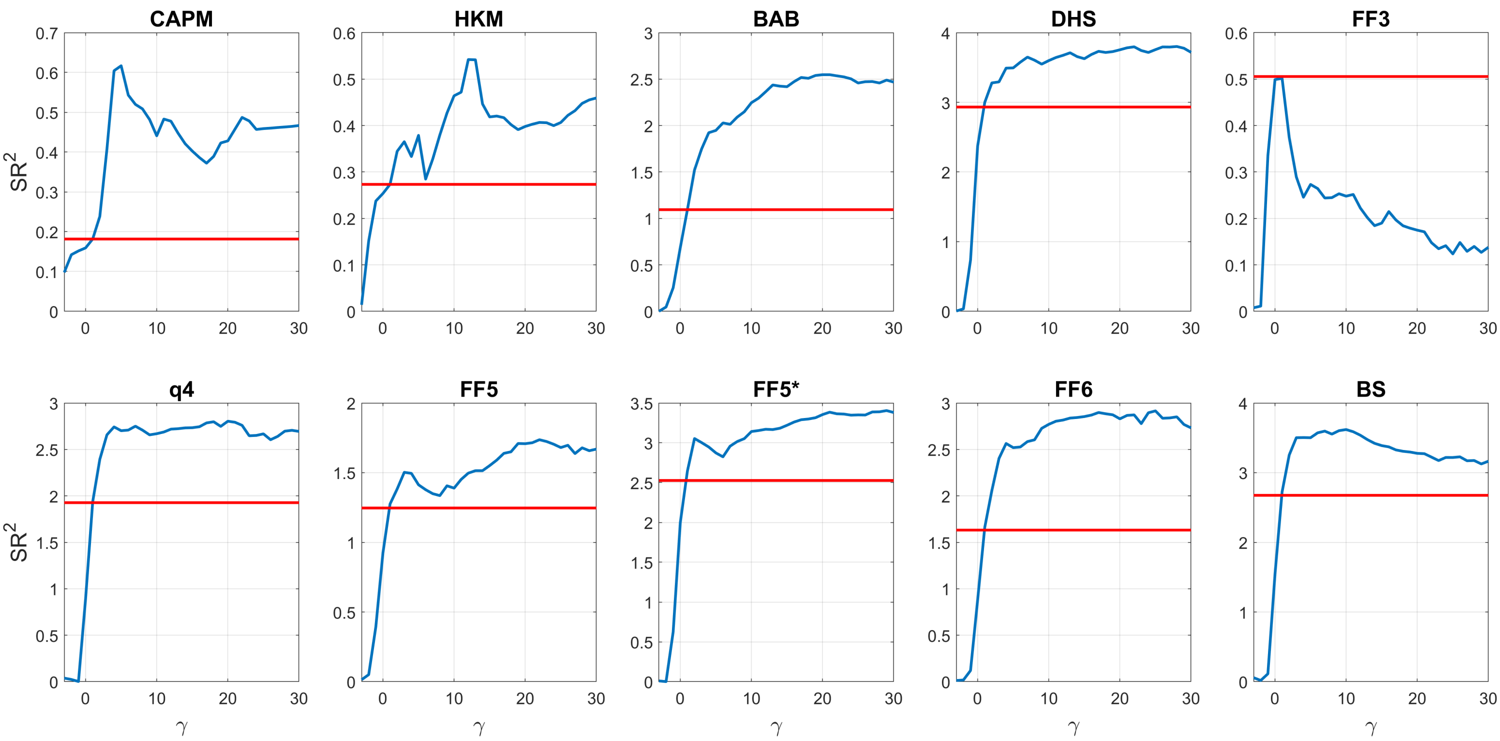

Fig. 5: **Squared Sharpe ratio across γ.** This figure plots, for each factor model, the squared Sharpe ratio (SR^2) across γ in blue. For each γ, the corresponding minimum discrepancy SDF pricing the factor model is obtained. Then, the SR^2 of its mimicking portfolio is reported. We consider $\gamma \in [-3, 30]$, 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.

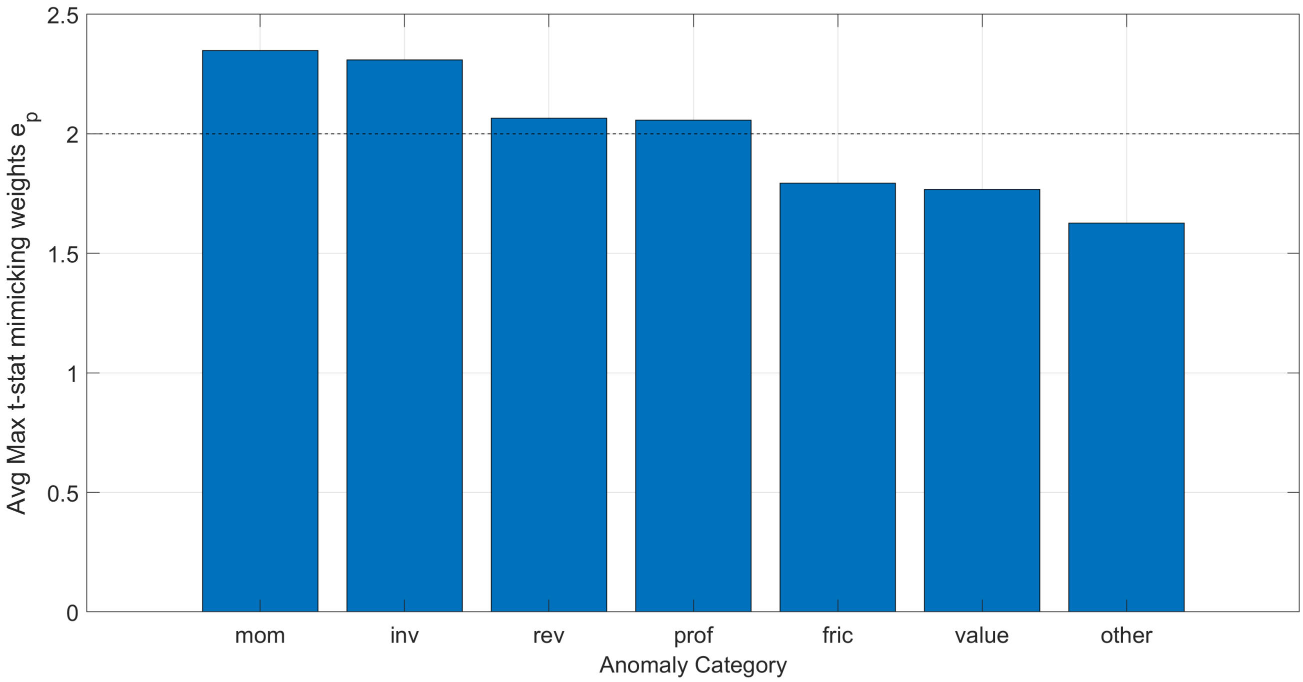

Fig. 6: **Anomaly importance in the mimicking portfolio.** This figure plots, as aggregated measure of importance of each anomaly category in mimicking factor nonlinearities, the average across factor models of the maximum t-stat of the anomalies within the category from the regression of the best nonlinear SDF nonlinear component on the universe of test assets. The sample ranges from July, 1972 to October, 2018.

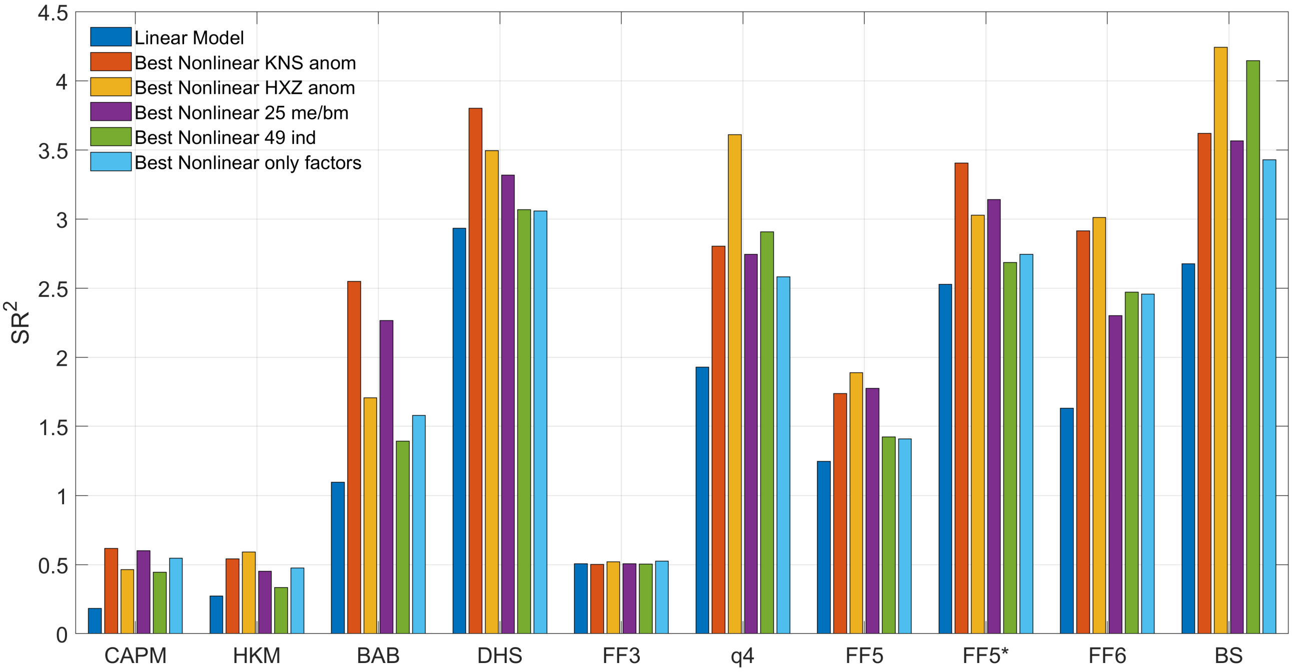

Fig. 7: **Squared Sharpe ratio of linear vs. best nonlinear model for different test assets.** This figure plots, for each factor model, the squared Sharpe ratio (SR^2) coming from the best nonlinear SDF, for each set of test assets, and from the linear SDF. The best nonlinear model is the one within $\gamma \in [-3, 30]$, 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.

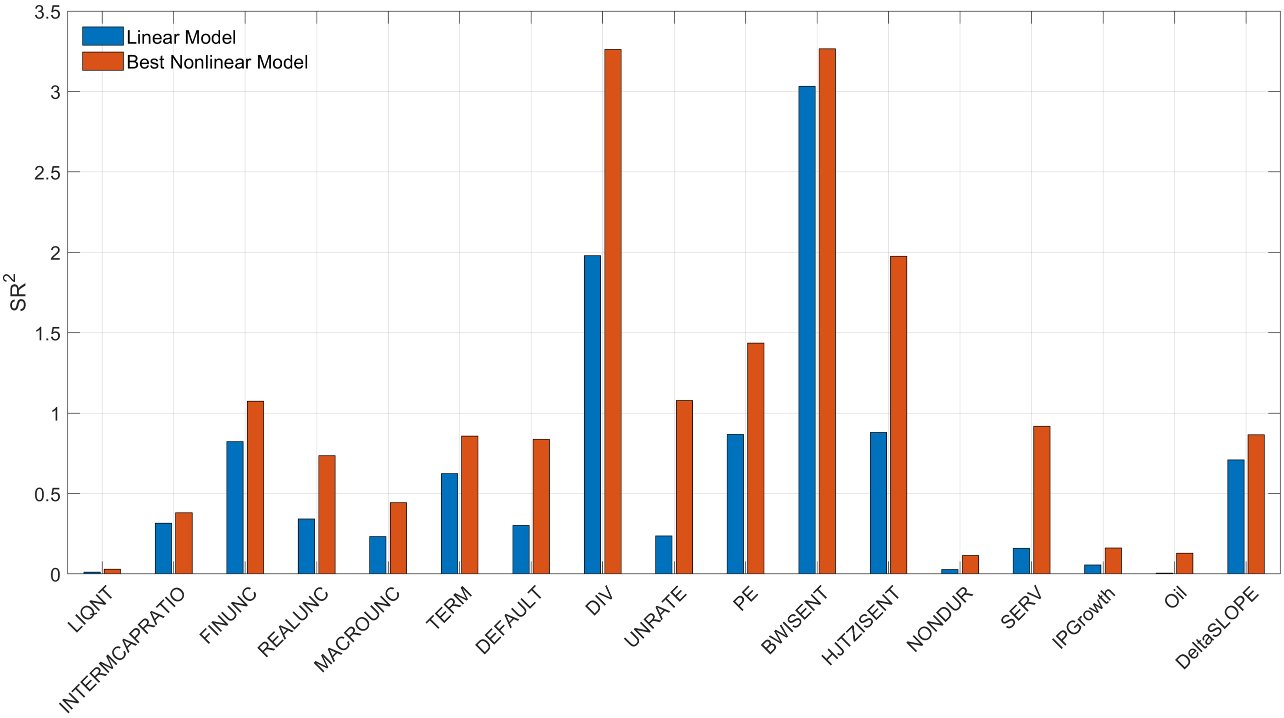

Fig. 8: **Squared Sharpe ratio of linear vs. best nonlinear model for nontraded factors.** This figure plots, for each nontraded one-factor model, the squared Sharpe ratio (SR^2) coming from the best nonlinear SDF and from the linear SDF. The best nonlinear model is the one within $\gamma \in [-3, 30]$, with a grid with spacing of 1, that yields the highest SR^2. Sharpe ratios are annualized. The sample ranges from October, 1973 to December, 2016.

Online Appendix to

Which (Nonlinear) Factor Models?

Caio Almeida[†] **Gustavo Freire**[‡]

August 12, 2024

Abstract

This appendix collects additional empirical results supporting the main paper.

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

[‡]E-mail: freire@ese.eur.nl, Erasmus School of Economics - Erasmus University Rotterdam, Tinbergen Institute and Erasmus Research Institute of Management (ERIM).

OA.1. Additional Empirical Results

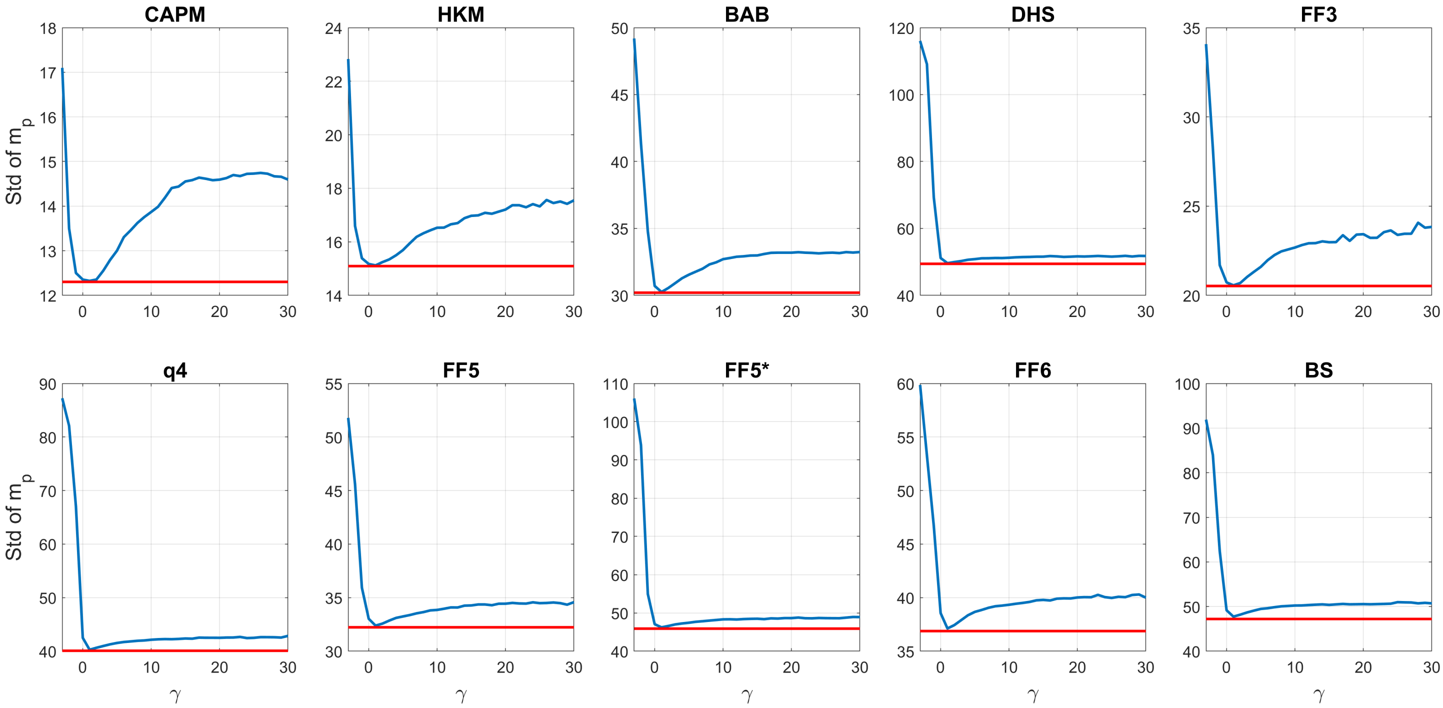

Fig. OA.1: **Volatility of SDF mimicking portfolio across γ.** This figure plots, for each factor model, the standard deviation (in %) of the mimicking portfolio of m_γ across γ in blue. For each γ, the corresponding minimum discrepancy SDF pricing the factor model is obtained. Then, the standard deviation 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.

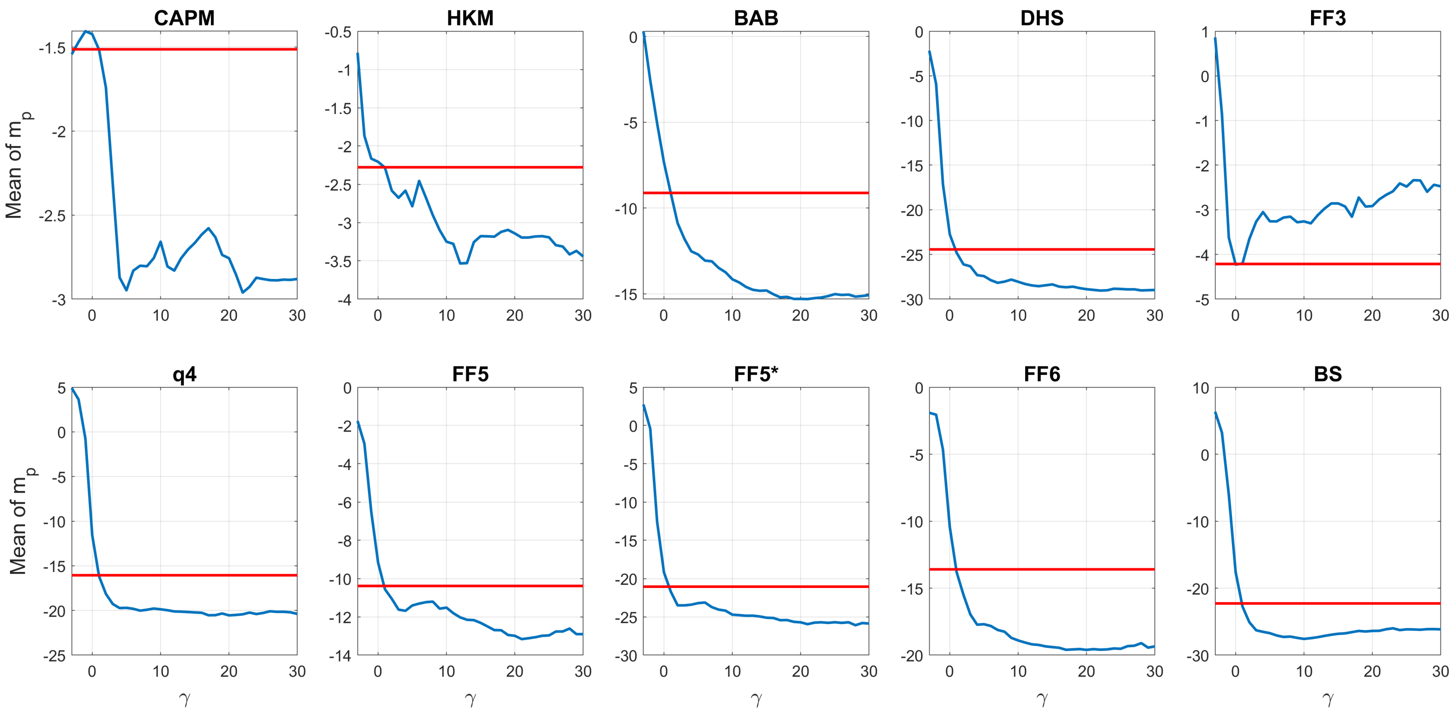

Fig. OA.2: **Mean of SDF mimicking portfolio across γ.** This figure plots, for each factor model, the mean (in %) of the mimicking portfolio of m_γ across γ in blue. For each γ, the corresponding minimum discrepancy SDF pricing the factor model is obtained. Then, the mean of its mimicking portfolio 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.

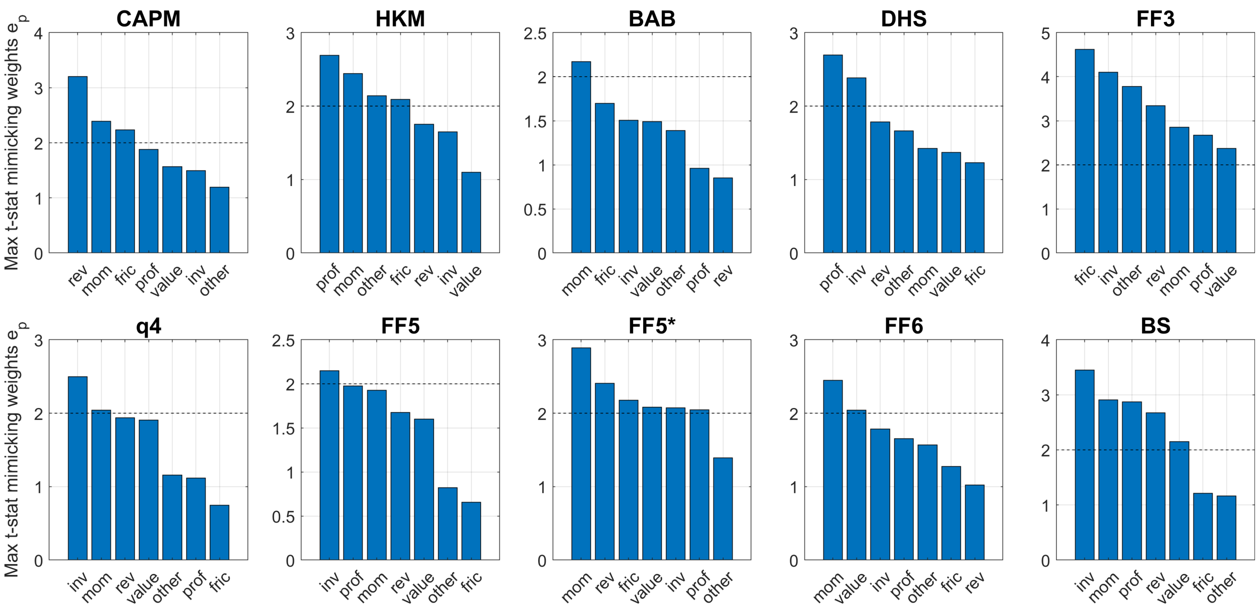

Fig. OA.3: **Anomaly importance in the mimicking portfolio of each factor model.** This figure plots, as measure of importance of each anomaly category in mimicking the factor nonlinearities of each model, the maximum t-stat of the anomalies within the category from the regression of the best nonlinear SDF nonlinear component on the universe of test assets. The sample ranges from July, 1972 to October, 2018.

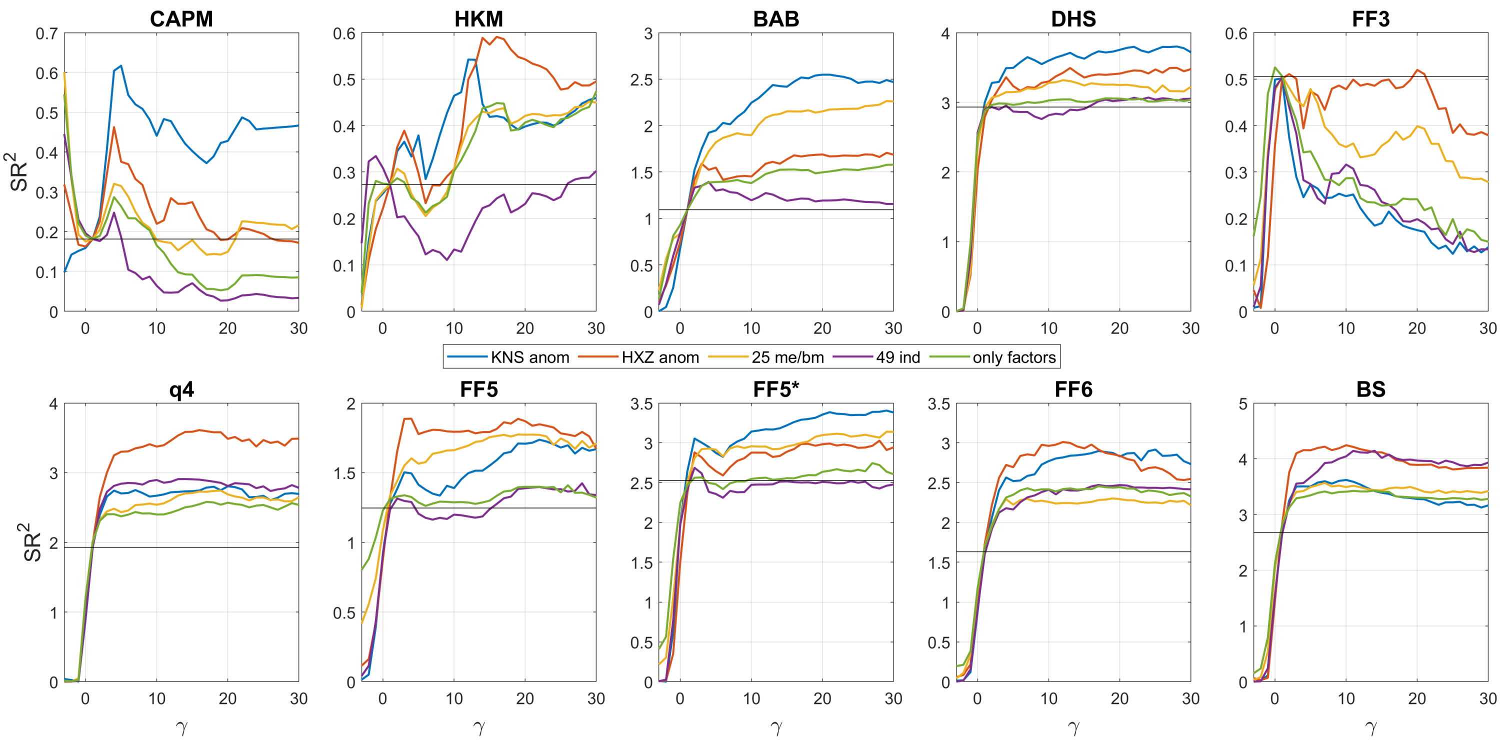

Fig. OA.4: **Squared Sharpe ratio across γ for different test assets.** This figure plots, for each factor model, the squared Sharpe ratio (SR^2) across γ for different sets of test assets. For each γ, the corresponding minimum discrepancy SDF pricing the factor model is obtained. Then, the SR^2 of its mimicking portfolio is reported. We consider $\gamma \in [-3, 30]$, with a grid with spacing of 1. The black horizontal line depicts the SR^2 of the linear SDF. Sharpe ratios are annualized. The sample ranges from July, 1972 to October, 2018.

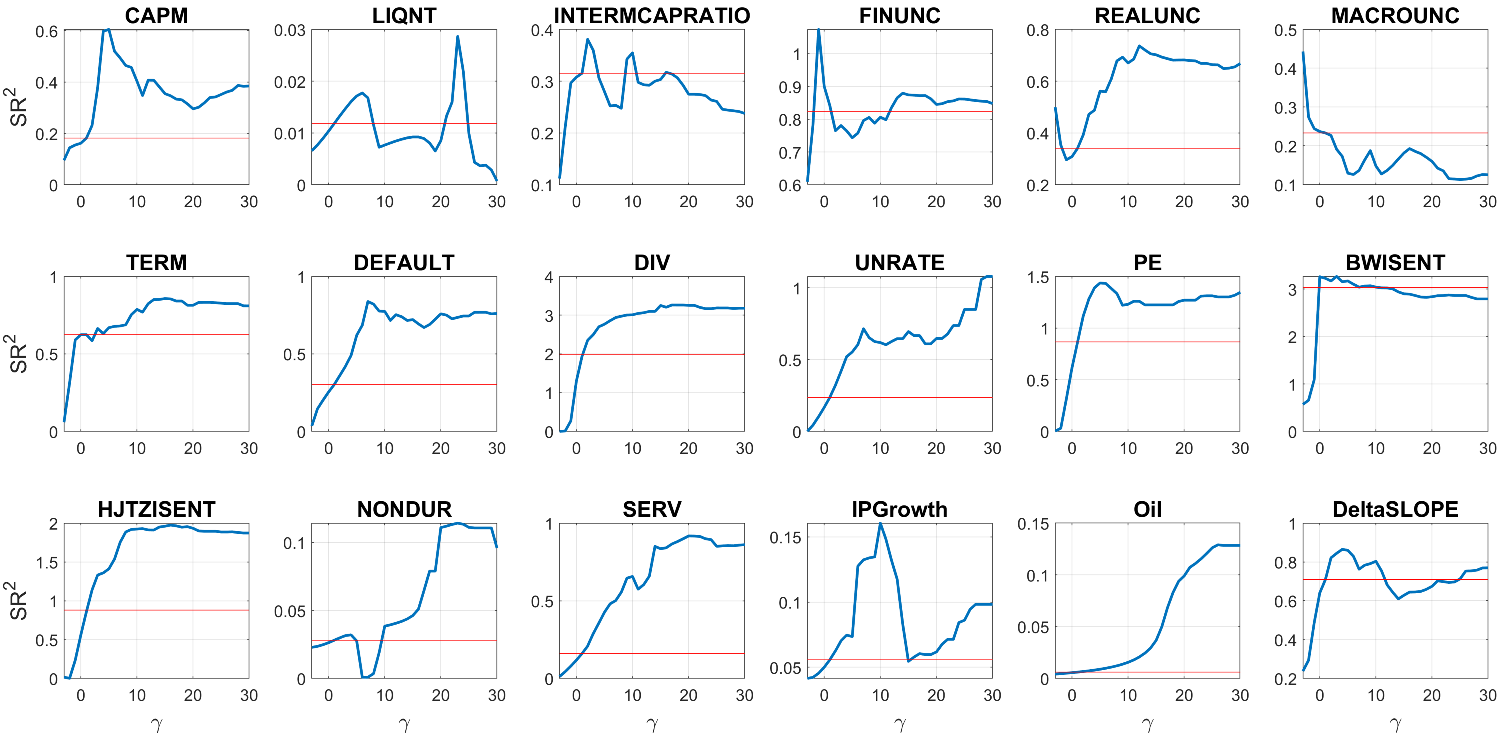

Fig. OA.5: **Squared Sharpe ratio across γ for nontraded factors.** This figure plots, for each nontraded factor model, the squared Sharpe ratio (SR^2) across γ in blue. For each γ, the corresponding minimum discrepancy SDF pricing the mimicking portfolio of the nontraded factor is obtained. Then, the 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. The first plot depicts the results for the traded CAPM as a benchmark. Sharpe ratios are annualized. The sample ranges from October, 1973 to December, 2016.