Asset Pricing with Nonlinear Principal Components

Caio Almeida*, René Garcia†, Stéphane N'Dri‡

January 2022
Preliminary Version - Do not circulate

Abstract

This paper studies the role of truly independent nonlinear factors in asset pricing. While the most successful stochastic discount factor (SDF) benchmarks pricing the cross-section of stock returns are obtained from regularized linear principal components of characteristic-based returns we show that allowing for substitution of some linear principal components by independent nonlinear factors consistently improves the SDF's ability to price this cross-section. We use the Fama-French 25 ME/BM-sorted portfolios, fifty anomaly portfolios, and fifty anomalies plus characteristic-based interaction terms to test the effectiveness of the nonlinear dynamic factors. The SDF estimated using a mixture of nonlinear and linear factors outperforms the ones using solely linear factors or raw characteristic returns in terms of out-of-sample R^2 pricing performance. Moreover, the hybrid model –using both nonlinear and linear principal components– requires less risk factors to achieve the highest out-of-sample performance compared to a model using only linear factors.

Keywords: Nonlinear principal components, Cross-section of returns, Stochastic discount factors, Entropy, Predictability, Dimension reduction

*Princeton University, Bendheim Center for Finance, calmeida@princeton.edu

†Université de Montréal and Toulouse School of Economics, rene.garcia@umontreal.ca, (Corresponding Author). Address for correspondence: Department of Economics, Université de Montréal, Pavillon Lionel-Groulx, 3150 rue Jean-Brillant, Montréal QC H3T 1N8, Canada. The second author thanks the NSERC (RGPIN-2018-04891), the SSHRC (Insight 435-2019-1216), and the FQRSC (2020-SE2-269954) research grant agencies for their financial support as well as the ANR for support under the Investissements d'Avenir program, grant ANR-17-EURE-0010 and the project ANR-17-CE03-0010 - LONGTERMISM: Evaluation des investissements ultra-longs. He is a research Fellow of CIRANO and CIREQ.

‡Université de Montréal, kouadio.stephane.ndri@umontreal.ca

1 Introduction

The search for a parsimonious stochastic discount factor (SDF) that can explain a large cross-section of equity returns is central to empirical asset pricing. The recent contribution of Kozak et al. (2020) shows that a large cross-section of characteristics-based portfolios can be shrunk to a few principal components that make up a SDF with a robust predictive out-of-sample pricing performance. This departs from the previous literature, where a limited number of factors predicted the variation of a given cross-section of characteristics-based portfolios. As this cross-section enlarged and new characteristics emerged, the number of factors increased from three in (Fama and Kenneth (1993)) to four in (Hou et al. (2015)), five in (Fama and French (2015)), and six in (Barillas and Shanken (2018)). Kozak et al. (2020) show that restricting the SDF to a few characteristics-based factors does not adequately capture the cross-section of expected returns. A sparse SDF with a few high-variance principal components produces a good and robust out-of-sample fit of the cross-section of expected returns.

The above-mentioned achievements make clear the importance of performing transformations of the original set of raw returns in order to obtain a robust SDF. In this paper, we move one step further by studying how changing the metric in which raw information is transformed in (potential) SDF factors affects this SDF's pricing performance on a fixed cross section of returns. While Kozak et al. (2020) use a Bayesian method based on a quadratic criterion we adopt an entropic criterion, which brings novel statistical properties to the extracted factors. We build on the recent theoretical developments of Gunsilius and Schennach (2021) who using a multivariate additive entropy decomposition generalize the principal component analysis (PCA) to a nonlinear setting. Their nonlinear principal components analysis delivers truly independent factors (as opposed to the uncorrelated factors of PCA) that maximize the percent of entropy information from the original cross-section of raw returns that is explained by the nonlinear factors.

Our main empirical contribution is to show that, for different fixed cross-sections of returns, when a small number of nonlinear principal components is allowed to complement/substitute factors on an SDF based on linear principal components, the nonlinear SDF consistently outperforms the linear specification and with fewer factors. To better understand the additional value of nonlinear principal components in pricing

the cross-section of returns, we apply the new methodology to the Fama-French 25 ME/BM-sorted portfolios. While the Fama-French three-factor linear model has been the workhorse of the asset pricing literature, it did not explain well the returns of extreme portfolios, especially the small-growth one, where nonlinear relations between portfolios and factors may be present. Next, we explore the set of fifty anomaly portfolios built by Kozak et al. (2020) using individual stock characteristics to assess the predictive ability of a stochastic discount factor combining both linear and nonlinear factors. Lastly, we use the big data set of 2600 portfolios proposed by Kozak et al. (2020), which includes the fifty raw characteristic excess returns plus 2550 interaction terms obtained by crossing the characteristics of the stocks (two by two and adding the third power of each characteristic) and aimed at capturing nonlinearities. Adding nonlinear principal components to linear factors increases substantially the out-of-sample R^2's for the different cross-sections of returns under consideration.

The numerical procedure used to extract the nonlinear principal components proceeds in several steps. We start with n variables that are potential predictors of future returns and extract n linear principal components in the usual way. We then select the k linear factors having the largest eigenvalues (variances) and apply the algorithm of Gunsilius and Schennach (2021) to capture nonlinear forms of dependence through truly independent factors. The approach relies on the theory of Brenier maps Brenier (1991), which are a generalization of monotone functions in multivariate settings, and on the use of entropy[1] to determine the principal nonlinear components that capture most of the information content of the data, instead of variance for linear principal components. Another important ingredient is a multivariate additive decomposition of the entropy into one-dimensional contributions.

Entropy as a measure of dispersion has received considerable attention in the asset pricing literature. The main focus has been on extracting SDFs from observed asset prices in the spirit of Hansen and Jagannathan (1991) who minimize the variance of the SDF subject to asset pricing restrictions. Minimizing the entropy involves higher moments of the distribution of asset returns and captures nonlinearities in the pricing kernel or non-Gaussianity in returns. Stutzer (1995) suggests a nonparametric bound to test asset pricing models based on entropy minimization, while Bansal and Lehmann

[1]See Shannon (1948), Kullback (1997), Csiszar (1991), and the other references cited in Gunsilius and Schennach (2021).

(1997) propose a related entropic bound that is obtained by maximizing the growth portfolio. Backus et al. (2011) tests disaster-based models based on this entropic bound. Several papers derive pricing kernels based on entropy. Alvarez and Jermann (2005) provide a decomposition of the pricing kernel into permanent and transitory components, while Ghosh et al. (2016) propose a factorization of the SDF into an observable component (a parametric function of consumption) and an unobservable nonparametric one. Backus et al. (2014) characterize the pricing kernel entropy and its dynamics for several representative agent models. Chen et al. (2020) use a statistical measure of discrepancy that extends relative entropy to recover information about investor beliefs embedded in forward-looking asset prices in conjunction with asset pricing models. Almeida and Garcia (2012, 2017) minimize a large class of divergence measures which includes entropy under asset pricing restrictions and derive information bounds as well as misspecification measures.

Since the nonlinear factors are not tradable by nature, we construct mimicking portfolios to extract tradable factors to be priced alongside with the linear principal components. For robustness purposes, we proceed in three manners. First, we performed a linear regression of the nonlinear factors on the fifty anomaly excess returns including a constant term. Second, we add one more asset, the CRSP value-weighted index, in the previous (first) regression. Third, we add an option on the market, to the previous (second) regression. In this third regression, we approximate the nonlinear principal components using a piecewise function to take into account their nonlinearities (see Glosten and Jagannathan (1994), and Diez De Los Rios and Garcia (2011)). Therefore, the mimicking portfolios are the predicted nonlinear factors from the regressions. We then use these mimicking portfolios along with the linear factors or linear principal components to estimate the stochastic discount factor or to predict future expected returns under the set up of Kozak et al. (2020). In the final step of our methodology, we regress the expected returns of different sets of tradable factors on the covariance matrix of different set of factors under Elastic-Net and Ridge penalties, and then, assess the accuracy of these regressions by computing the out-of-sample and in-sample cross-sectional R^2. We use the 3-fold cross-validation procedure of Kozak et al. (2020) to compute the out-of-sample R^2.

This paper is related to three strands of literature. First, the vast literature on nonlinear principal components. Papers here include Kramer (1991), Schölkopf et al.

(1998), Roweis and Saul (2000), Lee and Verleysen (2007), Chen et al. (2009), Lawrence (2012), and Damianou et al. (2021) among others. Recall that traditional principal components are extracted under the assumption of independence and stationarity of the raw random variables, which are then rotated to obtain uncorrelated linear factors chosen to maximize the explained variance of the original variables. While all the above-mentioned methods use variations and / or generalizations of the traditional principal components method which go from relaxing independence (Chen et al. (2009)) to applying traditional PCA to an augmented features' space via "the kernel trick" (Schölkopf et al. (1998)), they all obtain factors that maximize the explained variance of the original raw data using a quadratic criterion. In contrast, we build on Gunsilius and Schennach (2021) who find factors that are truly independent by construction and that maximize an aggregate measure of the entropy of the original raw random variables. In this context, our paper is the first to empirically test these entropic dynamic factors in an asset pricing application involving the identification of a SDF that prices the cross-section of stocks.

Second, this paper is related to the new growing literature on machine learning asset pricing models (Feng et al. (2018), Nakagawa et al. (2019), Chen et al. (2020), and Fang and Taylor (2021)). It provides a test for the effectiveness of using an alternative dimension reduction technique (based on entropy as a metric) to price the cross-section of stock returns. Third, this paper is also related to the strand of literature on the stochastic discount factor estimation using a given set of factors. Papers here include Fama and Kenneth (1993), Hou et al. (2015), Fama and French (2015), Barillas and Shanken (2018) and Kozak et al. (2018). While all these models bet on linear factor models we show that obtaining nonlinear dynamic factors can bring additional valuable information to price the cross-section of stock returns.

Our paper is also closely related to Gunsilius and Schennach (2021) and Kozak et al. (2020) in terms of methodology. However, there are key differences between this paper and theirs. Compared to Gunsilius and Schennach (2021) –which is mainly theoretical and did a simple application to predict bond excess returns using directly non tradable factors (nonlinear principal components) conjointly with tradable factors–, our paper sheds light on how to better adapt empirically the theory of truly independent nonlinear factors to an asset pricing context. We use the stock market as opposed to the bond market and the (tradable) mimicking portfolios analysis as opposed to the non tradable

portfolios analysis in Gunsilius and Schennach (2021). Compared to Kozak et al. (2020), our analysis uses a set of truly independent nonlinear factors and linear factors. This hybrid method delivers compelling out-of-sample performance measured by the R^2.

The remainder of the paper is organized as follows. Section 2 explains the methodology to extract the independent nonlinear components and to construct the stochastic discount factor to price the various sets of portfolios. Section 3 describes the construction of the data. We report the results of our analysis in Section 4 and conclude in Section 5.

2 Methodology

We first describe the steps to follow to construct the nonlinear factors, then we specify the estimation procedure of the stochastic discount factor.

2.1 Nonlinear Principal Components

Let us consider n portfolios whose excess returns are stacked in a $1 \times n$ vector $r = (r_1, r_2, ..., r_n)$. One would like to reduce the dimensionality of the space of these n portfolios to $k < n$ factors. The most common way is to take the first k linear principal components. This dimension reduction operates by finding successively the linear combination of the portfolio returns that explains the largest share of the variance-covariance matrix of the original set of portfolios, with the condition that each successive principal component is uncorrelated with the previous ones. The intuition is that we search for a line along which the points are the most dispersed, with the variance as a measure of dispersion. However, the principal component analysis can be generalized to explore nonlinear data representations. Gunsilius and Schennach (2021) propose to use entropy, as another concept of dispersion, to determine the most informative principal nonlinear components. Moreover, the method delivers independent instead of uncorrelated factors.

Let us suppose that r has a density function denoted $g(r)$. The idea of extracting truly independent nonlinear factors is to find a map $T : \mathbb{R}^n \longmapsto \mathbb{R}^n$ transforming $g(r)$ into a target density $\Phi(\tilde{r})$ where $\tilde{r} = T(r)$. The change of variable formula yields an expression of the original density function in terms of the target density function and the Brenier map as follows :

$$g(r) = \Phi(T(r))det(\frac{\partial T(r)}{\partial r'}) \tag{1}$$

First, we need to estimate the mapping function T by minimizing the distance between a nonlinear transformation of the data $T(r)$ and the original data r. We want to take into account a possible nonlinear relationship between the portfolios and yet, we do not want to depart too much from the original portfolios. Hence, T minimizes

$$\int \|T(r) - r\|^2 g(r)dr \tag{2}$$

The known solution of this minimization problem (Monge-Kantorovich-Brenier optimal transportation problem) is $T(r) = \frac{\partial C(r)}{\partial r}$, where C is a convex function. The function C can be estimated using a grid point procedure or an approximation procedure. This paper uses the grid point procedure. Second, we extract k eigenvectors $e = (e_1, e_2, ..., e_k)$ corresponding to the k largest eigenvalues of $\tilde{J}$ defined by :

$$\tilde{J} = -\int g(r) log \frac{\partial T(r)}{\partial r'} dr \tag{3}$$

Finally, the i^{th} nonlinear principal component is defined by : $\tilde{f}_i = T(r)e_i$ (versus $f_i = r \times \tilde{e}_i$ in the linear case). This is a nonlinear transformation of the original portfolio returns because T is a nonlinear function unless r is a Gaussian process. The nonlinear transformation of the portfolio returns implies that the new factors, $\tilde{F} = (\tilde{f}_1, \tilde{f}_2, ..., \tilde{f}_k)$ are not tradable in the sense that their returns cannot be obtained as a linear combination of the original portfolios of assets. We therefore resort to the usual procedure of forming a mimicking portfolio for each nonlinear factor -in order to see which asset will be long and short - before the estimation of the stochastic discount factor. We explain in Section 3.2 the various regression methods we used to obtain the mimicking portfolios.

2.2 Stochastic Discount Factor estimation procedure

We assume that the Stochastic Discount Factor (SDF) is an affine function of the factors, as follows :

$$SDF_t = 1 - \lambda'(F_t - \mu) \tag{4}$$

where λ is a $K \times 1$ vector of factor loadings, F_t is a $K \times 1$ vector of risk factors and $\mu = \mathbb{E}(F_t)$ is a $K \times 1$ vector of factors' mean. We make two additional assumptions about the factors: first, we suppose that μ is random and follows a normal distribution, and second, we consider that $\Sigma = Cov(F_t)$ is known.

$$\mu \sim \mathcal{N}(0, \frac{\kappa^2}{\tau}\Sigma^\eta) \quad , \ \tau = tr[\Sigma] \tag{5}$$

where κ is a scale parameter and η is a shape parameter which we set its value to 2. Because we also suppose that there is no near-arbitrage opportunities, we want the Sharpe ratio of high-eigenvalue principal components (PCs) to be higher than the Sharpe ratio of the low-eigenvalue PCs (which is economically plausible since the latter do not bring much risk premium, as emphasized in **?**). The stochastic discount factor should satisfy the law of one price:

$$\mathbb{E}(SDF_t \times F_t) = 0 \tag{6}$$

The equation 6 is in line with the characteristics-based factor model literature. We follow Kozak et al. (2020) procedure to estimate the stochastic discount factor. Basically, λ solves the following minimization problem:

$$\hat{\lambda} = arg\min_{\lambda}(\mu - \Sigma\lambda)'\Sigma^{-1}(\mu - \Sigma\lambda) + \gamma_1 \sum_{i=1}^{k} |\lambda_i| + \gamma_2 \lambda'\lambda, \tag{7}$$

that is we minimize the HJ-distance subject to an Elastic-net/Ridge constraint. Under the assumption that $\eta = 2$, $\gamma_1 = 0$, and therefore $\gamma_2 = \dfrac{\tau}{\kappa^2 T}$, the expected maximum squared Sharpe ratio is equal to the squared scale parameter :

$$\mathbb{E}(\mu\Sigma^{-1}\mu) = \kappa^2 \tag{8}$$

Under the ridge shrinkage hypothesis $\gamma_1 = 0$, we obtain

$$\hat{\lambda} = (\Sigma + \gamma_1 \mathbb{I})^{-1}\mu$$

. Under the Elastic-Net shrinkage hypothesis, we use the Least Angle Regression (LAR-EN) algorithm[2] to estimate λ.

The shrinkage parameters γ_2 or (γ_1 and γ_2) are optimally chosen using the out-of-sample R-squared constructed by cross-validation as in Kozak et al. (2020). First, we set a grid on the shrinkage parameters γ_1, and γ_2. Second, we divide the sample into H equal subsamples. Third, for each possible pair of γ_1 and γ_2, we compute $\hat{\lambda}$ by using $H - 1$ of these subsamples. Then, we evaluate the out-of-sample fit of the resulting model on the single withheld subsample by computing the out-of-sample R-squared

[2]See the appendix for the algorithm.

(R^2_{oos}) as

$$R^2_{oos} = 1 - \frac{\left(\mu_2 - \tilde{\Sigma}_2\hat{\lambda}\right)'\left(\mu_2 - \tilde{\Sigma}_2\hat{\lambda}\right)}{\mu_2'\mu_2} \tag{9}$$

where μ_2 and $\tilde{\Sigma}_2$ are respectively the sample mean and covariance of the factors from the withheld subsample. We do this exercise H times and for each time, we treat a different subsample as the out-of-sample data. Finally, we define the cross-validated R^2_{cv-oos} as the average of the R^2_{oos} across these H estimates and choose γ_1 and γ_2 that generates the highest R^2_{cv-oos} as the optimal values.

3 Data

As mentioned in the introduction, the empirical part of this paper employs three data sets. First, we use the Fama-French 25 ME/BM-sorted (FF25) portfolios dowloaded from Kenneth R. French website. The two other data sets are built using individual stock characteristics. We start by all the firms' stocks available on the Center for Research in Security Prices (CRSP) and take the accounting data from Compustat. The data from CRSP are monthly/daily and the one from Compustat are quarterly. Our final data is a set of daily/monthly returns spanning the period from November 1973 to December 2019. For each date t and from each stock $s \in 1, 2, ..., n_t$, we build fifty stock-characteristic portfolios (see Table 1) following the common definition of anomalies (Novy-Marx and Velikov (2016), Kozak et al. (2020)) : $\left(x^i_{s,t}\right)_{s\in 1,2,...,n_t;\ i\in 1,2,...,50;\ t\in 1,2,...,T}$, where n_t is the number of stocks at time t for which we can calculate the anomaly variable. Following Freyberger et al. (2020) and Kozak et al. (2020), we perform a rank-transformation denoted $rx^i_{s,t}$ before normalizing that rank-transformed characteristic to obtain the final zero-investment long-short portfolios denoted $z^i_{s,t}$. First, we rank all stocks for which data are available based on $x^i_{s,t}$ for each i, t. Second, we compute $rx^i_{s,t}$ and $z^i_{s,t}$ as follows :

$$rx^i_{s,t} = \frac{rank(x^i_{s,t})}{1+n_t} \tag{10}$$

$$z^i_{s,t} = \frac{rx^i_{s,t} - \frac{1}{n_t}\sum_{s=1}^{n_t} rx^i_{s,t}}{\sum_{s=1}^{n_t} |rx^i_{s,t} - \frac{1}{n_t}\sum_{s=1}^{n_t} rx^i_{s,t}|} \tag{11}$$

The anomalies are monthly for anomalies using CRSP variables and quarterly for anomalies whose calculations use accounting variables. To obtain daily factor returns, we assumed that the anomalies take the same values for each day within the month or the quarter. Third, the raw characteristic factors are obtained by : $RC_t = Z'_{t-1}R_t$ where Z_t is a n_t-by-50 matrix containing the $z^i_{s,t}$ for all s, i and R_t is a n_t-by-1 vector of daily returns from CRSP. Before doing this interaction, we remove all small capitalization stocks with capitalization under 0.01% of the aggregate market capitalization for each t from the data. Furthermore, we orthogonalized each factor returns with respect to the CRSP value-weighted index return using β's estimated with the full sample. Then, we rescaled the portfolio returns to have their standard deviations equal to the in-sample standard deviation of the excess returns on the CRSP value-weighted index returns which we took as the market index. We computed the excess returns using the one-month Treasury bill rate. The first data is the Fama-French 25 portfolios. Our second data set is the time series of the fifty raw characteristic excess returns we just computed.

3.1 Interactions

We add interaction terms to the second data set to constitute the third data set. The aim of the use of this database is to compare the nonlinearity introduced by Kozak et al. (2020), which requires a very high-dimensional data set (2,600 factors), to the empirical performance of the nonlinearity introduced by Gunsilius and Schennach (2021), which require less nonlinear factors. The interaction-term weights on the individual stocks are constructed as follows:

$$z_{s,t}^{ij} = \frac{z_{s,t}^i z_{s,t}^j - \frac{1}{n_t}\sum_{s=1}^{n_t} z_{s,t}^i z_{s,t}^j}{\sum_{s=1}^{n_t} |z_{s,t}^i z_{s,t}^j - \frac{1}{n_t}\sum_{s=1}^{n_t} z_{s,t}^i z_{s,t}^j|} \tag{12}$$

$$z_{s,t}^{ii} = \frac{(z_{s,t}^i)^2 - \frac{1}{n_t}\sum_{s=1}^{n_t} (z_{s,t}^i)^2}{\sum_{s=1}^{n_t} |(z_{s,t}^i)^2 - \frac{1}{n_t}\sum_{s=1}^{n_t} (z_{s,t}^i)^2|} \tag{13}$$

$$z_{s,t}^{iii} = \frac{(z_{s,t}^i)^3 - \frac{1}{n_t}\sum_{s=1}^{n_t} (z_{s,t}^i)^3}{\sum_{s=1}^{n_t} |(z_{s,t}^i)^3 - \frac{1}{n_t}\sum_{s=1}^{n_t} (z_{s,t}^i)^3|} \tag{14}$$

3.2 Nonlinear principal components

As stated in the methodology section, this paper uses a grid-point procedure to estimate the Brenier map needed to compute the nonlinear factors. We extract the nonlinear factors using either the fifty raw characteristics data or the Fama-French 25 portfolios. To reduce the computation burden, we adopt a hybrid procedure. First, one extracts the linear principal components, henceforth LF from the fifty raw characteristics (or the Fama-French 25 portfolios). Second, we compute k nonlinear principal components, hereafter NLF from the first k linear principal components $(LF_1, LF_2, ..., LF_k)$[3]. The rationale for this approach rests on the premise that the nonlinearity comes from the linear factors with the highest eigenvalues, the ones that capture the most information.

As mentioned before, we need to reconcile the statistical factor extraction and the financial factors extraction by constructing portfolios mimicking NLF. For robustness purposes, we consider three regression strategies for constructing the mimicking portfolios. The first two regressions estimate the mimicking portfolios in the usual way. It consists in projecting the factors on a set of basis assets. We add a third regression to account for the nonlinearity by adding piecewise linear functions (see Glosten and Jagannathan (1994) and Diez De Los Rios and Garcia (2011)). Henceforth, we will denote the mimicking portfolios MP.

[3]See appendix for the computation details.

$$NLF_{j,t} = \beta_{0,j} + \beta'_{c,j} RC_t + \epsilon_{j,t} \quad t = 1, ..., T \tag{15}$$

$$NLF_{j,t} = \beta_{0,j} + \beta_{1,j} r_{mkt,t} + \beta'_{c,j} RC_t + \epsilon_{j,t} \quad t = 1, ..., T \tag{16}$$

$$NLF_{j,t} = \beta_{0,j} + \beta_{1,j} r_{mkt,t} + \beta'_{c,j} RC_t + \delta_j max(r_{mkt,t} - l_j, 0) + \epsilon_{j,t} \quad t = 1, ..., T \tag{17}$$

where the βs and δs are the regression coefficients, RC_t stands for the raw characteristics excess returns, $r_{mkt,t}$ for the market excess returns, and $\epsilon_{j,t}$ for the error terms. The returns of the mimicking portfolios will be the predicted values of these regressions $MP = N\hat{L}F$. In the following section 4, we will present the basic case and we put the robustness checking analysis in the appendix 8.

Let LF_{-k} be a set of $50 - k$ linear principal components, excluding the first k linear PCs and let $MP^{(k)}$ and $NLF^{(k)}$ be respectively a set of k mimicking portfolios from the third regression and nonlinear PCs. In the basic case, we price $[LF_{-k}, MP^{(k)}]$ using risk factors derived from $[LF_{-k}, MP^{(k)}]$. Namely, $\mu = \mathbb{E}([LF_{-k}, MP^{(k)}]), \Sigma = Cov([LF_{-k}, MP^{(k)}])$. To check the robustness of our results, we use directly the nonlinear principal components in the risk factors instead of their mimicking portfolios. Thereby, $\mu = \mathbb{E}([LF_{-k}, MP^{(k)}]), \Sigma = Cov([LF_{-k}, NLF^{(k)}])$

4 Results

4.1 Fama-French 25 ME/BM-sorted portfolios

To acquire an intuition about the potential for capturing nonlinearities with the NLPC methodology, we start our analysis with the Fama-French 25 ME/BM-sorted portfolios. The properties of this set of test assets are well-known from previous studies, but most of the time in a linear factor-model context.[4] We consider five specifications for the stochastic discount factor and present the optimal model in Table 2 in terms of out-of-sample R^2, Sharpe ratio of the mean-variance efficient portfolio, and scale parameter

[4]Ghosh et al. (2019) build a one-factor SDF from a large cross-section of equity portfolios based on entropy and show that it delivers smaller out-of-sample pricing errors and a better cross-sectional fit than leading factor models, in particular the three-factor Fama-French model that we consider in our analysis. The so-called information theoretic SDF is highly positively skewed and leptokurtic, and therefore captures nonlinearities in the test assets that imply compensation in the observed risk premia.

κ.

First, we estimate the SDF using all the 25 portfolio excess returns as factors (column 1). Second, we do the same exercise but using the 25 principal components extracted from the Fama-French 25 ME/BM-sorted portfolios as factors (column 2). Third, we estimate an hybrid model where we replace the first two linear principal components by two mimicking portfolios of the nonlinear factors extracted from the first two linear factors. This hybrid specification is reasonable because the nonlinearity dwell in the principal component with the highest variances or eigenvalues as mentioned before. Fourth, we estimate Fama-French 3-factor model, and finally we estimate the SDF using just 2-nonlinear factors. From the comparison of the performance of these specifications in Table 2, we draw three main observations. Compared to the linear models, the nonlinear specifications lead to a higher out-of-sample R^2_{cv-oos}, a higher Sharpe ratio of the MVE portfolio, and a lower κ, which means that we impose more L2 shrinkage when we consider nonlinear factors.

4.2 Fifty Anomaly Portfolios

We now apply the same methodology to a set of fifty portfolios sorted according to different stock characteristics as in Kozak et al. (2020). We will proceed in a similar way as with the Fama-French 25 ME/BM-sorted portfolios by comparing linear specifications and parsimonious nonlinear specifications. We gather our results about the stochastic discount factor specifications in Figures 5, 6, 7, and 8. In the first two figures, we show the out-of-sample R^2's of a model using fifty raw characteristic excess returns (Figure 5), and a model using fifty linear principal components (Figure 6). We can see the difference between those two specifications as in Kozak et al. (2020). The left panel shows the R^2_{oos} in color map under the dual penalty. The right panel shows in red the R^2_{oos} pattern for different values of the tuning parameter or equivalently the Sharpe Ratio. From these graphs, we conclude that the projection of the SDF on the linear principal components space requires less factors to attain the maximum R^2_{oos} compared to the projection into the raw characteristics space, and that the difference between the two approaches in terms of the highest R^2_{oos} is small.

We now look at the projection of the stochastic discount factor into the hybrid space where we replace the first k linear factors ($k = 2$ for Figure 7 and $k = 3$ for Figure 8)

by k mimicking portfolios of the nonlinear factors. In these graphs, we are pricing fifty factors (made of k mimicking portfolios and $50-k$ linear factors. For both k values, we observe that we need fewer factors compared to the linear case in 5 to reach the maximum R^2_{oos} under the elastic net penalty. With about 5 factors, the R^2_{oos} is now around 0.5. The LARS-EN algorithm adds the factors starting by the model with all the mimicking portfolios of the nonlinear factors. The mimicking portfolios are always kept among the selected factors in the optimal model. We can conclude that thry capture more information than the linear factors that they replace since less factors are selected in the optimal model. In the right-nad panels of Figures 7 and 8, which feature the R^2_{oos} under the ridge penalty ($\gamma_1 = 0$), we also report values higher than 50% compared to a bit more than 20% with the linear-factor analysis in Figure 5.

4.3 Linear factors versus mimicking portfolios

We have put forward the importance of introducing nonlinearity in the stochastic discount factor estimation by replacing the linear factors with the highest variances by the mimicking portfolios of the nonlinear factors. As a matter of fact, the out-of-sample R^2 in average increases from 0.49 to 0.65 for the Fama-French 25 ME/BM-sorted portfolios data and from 0.22 to 0.55 for the fifty anomaly portfolios data. To better understand where this improvement comes from, we provide in Figures 9 and 10 a plot of the respective weights $(\frac{w_i}{\sum |w_i|})$ of the first two linear factors and the first two mimicking portfolios for the fifty anomalies and the twenty-five Fama-French portfolios. Let us first look at the characteristics portfolios. For the first factor, the main differences in exposures appear for idiosyncratic volatility, beta arbitrage, composite issuance, price and share volume. The weights for the other portfolios remain very similar between the linear and the mimicking portfolios. For the second factor, we observe differences for most factors, albeit with varied magnitudes. The large ones occur mainly for characteristics linked to momentum. The differences are much less apparent for the 25 FF portfolios, which is consistent with the fact that the portfolios are built with two characteristics, size and book-to-market, but we observe small differences for most of the portfolios for the first factor. There are relatively no significant differences between the linear and the mimicking portfolios for the second factor.

4.4 Interactions : Very high-dimensional data

In order to compare the nonlinearity introduced in this paper with the one of Kozak et al. (2020), we include in Figure 11 the R^2_{oos} for the dual penalty both in the raw characteristics space and in the linear principal components space with the full set of 2600 portfolios of raw characteristics built by Kozak et al. (2020). It is important to emphasize that the optimal achievable R^2_{oos}s are not comparable with the results from our approach since we are not pricing the same assets. However, we can see in the right-hand panel that the optimal performance is obtained for a small number of PCs but that the maximum R^2_{oos} in the vicinity of 35%. It tells at least that our more parsimonious approach to construct nonlinear factors achieves a competitive performance.

5 Conclusion

This paper shows how truly independent nonlinear factors alongside linear principal components improve the prediction of future expected returns . We use the Fama-French 25 ME/BM-sorted portfolios and fifty anomaly portfolios built using individual stock characteristics to reveal the strengths of the truly independent nonlinear principal components. Then, we estimate the expected returns or equivalently the stochastic discount factor using risk factors derived from raw characteristic excess returns, linear principal component portfolio returns and nonlinear principal component (mimicking) portfolio returns. The hybrid model –using both nonlinear and linear principal components– requires less risk factors to achieve the highest out-of-sample performance compared to a model using only linear factors or a model projected into the raw characteristic returns. Plotting the weights of the anomalies on the linear principal component portfolios and the portfolios mimicking the nonlinear factors, we find a weight shifting on some anomalies. Our results that nonlinear principal components should be considered when the SDF is built with many anomalies since nonlinearities are likely to appear.

References

Almeida, C. and R. Garcia (2012). Assessing misspecified asset pricing models with empirical likelihood estimators. *Journal of Econometrics 170*(2), 519–537.

Almeida, C. and R. Garcia (2017). Economic implications of nonlinear pricing kernels. *Management Science 63*(10), 3361–3380.

Alvarez, F. and U. J. Jermann (2005). Using asset prices to measure the persistence of the marginal utility of wealth. *Econometrica 73*(6), 1977–2016.

Backus, D., M. Chernov, and I. Martin (2011). Disasters implied by equity index options. *The Journal of Finance 66*(6), 1969–2012.

Backus, D., M. Chernov, and S. Zin (2014). Sources of entropy in representative agent models. *The Journal of Finance 69*(1), 51–99.

Bansal, R. and B. N. Lehmann (1997, June). Growth-Optimal Portfolio Restrictions On Asset Pricing Models. *Macroeconomic Dynamics 1*(2), 333–354.

Barillas, F. and J. Shanken (2018). Comparing asset pricing models. *The Journal of Finance 73*(2), 715–754.

Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics 44*(4), 375–417.

Chen, L., M. Pelger, and J. Zhu (2020). Deep learning in asset pricing. *Available at SSRN 3350138*.

Chen, X., L. P. Hansen, and P. G. Hansen (2020). Robust identification of investor beliefs. *Proceedings of the National Academy of Sciences 117*(52), 33130–33140.

Chen, X., L. P. Hansen, and J. Scheinkman (2009). Nonlinear principal components and long-run implications of multivariate diffusions. *The Annals of Statistics*, 4279–4312.

Csiszar, I. (1991). Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems. *Annals of Statistics 19*, 2032–2066.

Damianou, A., N. D. Lawrence, and C. H. Ek (2021). Multi-view learning as a nonparametric nonlinear inter-battery factor analysis. *Journal of Machine Learning Research 22*(86), 1–51.

Diez De Los Rios, A. and R. Garcia (2011). Assessing and valuing the nonlinear structure of hedge fund returns. *Journal of Applied Econometrics 26*(2), 193–212.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of financial economics 116*(1), 1–22.

Fama, E. F. and R. Kenneth (1993). French, 1993, common risk factors in the returns on stocks and bonds. *Journal of financial economics 33*(1), 3–56.

Fang, M. and S. Taylor (2021). A machine learning based asset pricing factor model comparison on anomaly portfolios. *Economics Letters*, 109919.

Feng, G., J. He, and N. G. Polson (2018). Deep learning for predicting asset returns. *arXiv preprint arXiv:1804.09314*.

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies 33*(5), 2326–2377.

Ghosh, A., C. Julliard, and A. P. Taylor (2016, 10). What Is the Consumption-CAPM Missing? An Information-Theoretic Framework for the Analysis of Asset Pricing Models. *The Review of Financial Studies 30*(2), 442–504.

Ghosh, A., C. Julliard, and A. P. Taylor (2019). An Information-Theoretic Asset Pricing Model. Technical report, Desautels Faculty of Management, McGill University.

Glosten, L. R. and R. Jagannathan (1994). A contingent claim approach to performance evaluation. *Journal of Empirical Finance 1*(2), 133–160.

Gunsilius, F. and S. Schennach (2021). Independent nonlinear component analysis. *Journal of the American Statistical Association 0*(0), 1–14.

Hansen, L. and R. Jagannathan (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy 99*(2), 225–62.

Hou, K., C. Xue, and L. Zhang (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies 28*(3), 650–705.

Kozak, S., S. Nagel, and S. Santosh (2018). Interpreting factor models. *The Journal of Finance 73*(3), 1183–1223.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics 135*(2), 271–292.

Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE journal 37*(2), 233–243.

Kullback, S. (1997). *Information theory and statistics*. Courier Corporation.

Lawrence, N. D. (2012). A unifying probabilistic perspective for spectral dimensionality reduction: Insights and new models. *The Journal of Machine Learning Research 13*(1), 1609–1638.

Lee, J. A. and M. Verleysen (2007). *Nonlinear dimensionality reduction*. Springer Science & Business Media.

Nakagawa, K., T. Ito, M. Abe, and K. Izumi (2019). Deep recurrent factor model: interpretable non-linear and time-varying multi-factor model. *arXiv preprint arXiv:1901.11493*.

Novy-Marx, R. and M. Velikov (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies 29*(1), 104–147.

Roweis, S. T. and L. K. Saul (2000). Nonlinear dimensionality reduction by locally linear embedding. *science 290*(5500), 2323–2326.

Schölkopf, B., A. Smola, and K.-R. Müller (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation 10*(5), 1299–1319.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal 27*(3), 379–423.

Stutzer, M. (1995). A bayesian approach to diagnosis of asset pricing models. *Journal of Econometrics 68*(2), 367–397.

6 Tables

Table 1: **List of anomalies**

Abbreviation	Name of the anomaly
accruals	Accruals Follows Sloan (1996).
agrowth	Asset Growth Follows Cooper et al. (2008)
aturnover	Asset Turnover Follows Soliman (2008).
cfp	Cash Flow / Market Value of Equity Follows Lakonishok et al. (1994).
ciss	Composite Issuance Follows Daniel and Titman (2006).
ep	Earnings/Price Follows Basu (1977).
gltnoa	Growth in LTNOA Follows Fairfield et al. (2003).
gmargins	Gross Margins Follows Novy Marx (2013a).
inv	Investment Follows Chen et al. (2011); Lyandres et al. (2007).
igrowth	Invetment Growth Follows Xing (2008).
invcap	Investment-to-Capital Follows Xing (2008).
indmomrev	Industry Momentum-Reversal Follows Moskowitz and Grinblatt (1999).
indrrev	Industry Relative Reversals Follows Da et al. (2013)
indrrevlv	Industry Relative Reversals (Low Volatility) Follows Da et al. (2013).
lrrev	Long-term Reversals Follows DeBondt and Thaler (1985).
mom11	Momentum (11m) Follows Jegadeesh and Titman (1993)
mom6	Momentum (6m) Follows Jegadeesh and Titman (1993)
indmom	Industry Momentum Follows Moskowitz and Grinblatt (1999).
valmom	Value-Momentum Follows Novy Marx (2013b).
momrev	Momentum-Reversal Follows Jegadeesh and Titman (1993).
nissa	Share Issuance (annual) Follows Pontiff and Woodgate (2008)
noa	Net Operating Assets Follows Hirshleifer et al. (2004).
noaa	Net Operating Assets Follows Kozak et al paper
price	Follows Blume and Husic (1973).
roa	Return on Assets Follows Chen et al. (2011).
roaa	Return on Assets (annual) Follows Chen et al. (2011).
roe	Return on Book Equity Follows Chen et al. (2011)
season	Seasonality Follows Heston and Sadka (2008).

Continued on next page

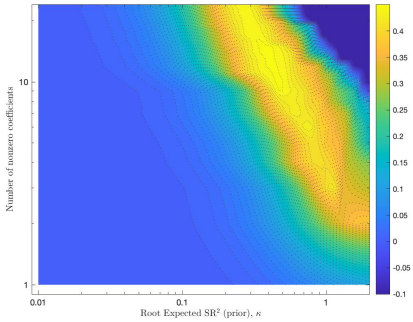
Table 1 – continued from previous page

Abbreviation	Name of the anomaly
sgrowth	Sales Growth Follows Lakonishok et al. (1994)
shvol	Share Volume Follows Datar et al. (1998).
size	Follows Fama and French (1993b)
strev	Short-term Reversal Follows Jegadeesh (1990).
sue	Standardized Unexpected Earning Follows Foster et al. (1984).
value	Follows Fama and French (1993b)
valuem	Value (monthly) Follows Asness and Frazzini (2013).
prof	Gross Profitability Follows Novy Marx (2013a).
valprof	Value-Profitability Follows Novy Marx (2013b).
F-score	Piotroski?s F -score Follows Piotroski (2000).
debtiss	Debt Issuance Follows Spiess and Affleck-Graves (1999).
repurch	Share Repurchases Follows Ikenberry et al. (1995).
divp	Dividend Yield Follows Naranjo et al. (1998).
lev	Leverage Follows Bhandari (1988).
sp	Sales-to-Price Follows Barbee Jr et al. (1996).
valmomprof	Value-Momentum-Profitability Follows Novy Marx (2013b).
shortint	Short Interest Follows Dechow et al. (1998).
nissm	Share Issuance (monthly) Follows Pontiff and Woodgate (2008).
rome	Return on Market Equity Follows Chen et al. (2011).
ivol	Idiosyncratic Volatility Follows Ang et al. (2006).
beta	Beta Arbitrage Follows Cooper et al. (2008).
ciss	Composite Issuance Follows Daniel and Titman (2006).
age	Firm Age Follows Barry and Brown (1984).

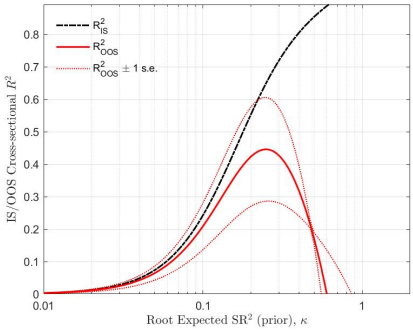
Table 2: **Comparison of Linear and Nonlinear Specifications**

	25RR	25LF	HM	3FF	2NL
R^2_{oos}	0.4403	0.4948	0.6765	0.5387	0.6184
κ	0.4011	0.2913	0.3149	0.1706	0.3430
SR	0.501	0.4758	0.7788	0.4610	0.3553

7 Figures

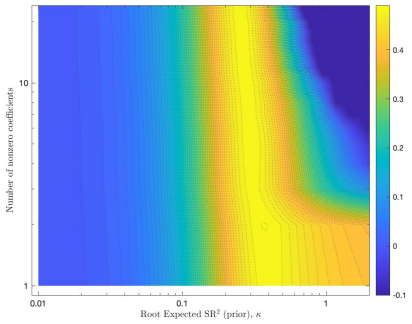


(a) Dual penalty

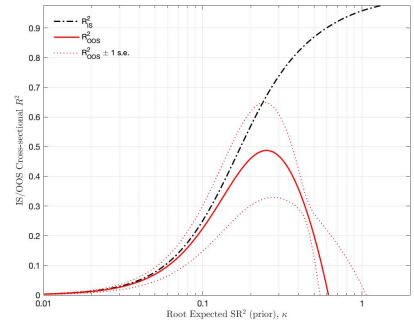


(b) L2 penalty

Figure 1: **Raw 25 anomaly portfolios** In the left panel, we plot in a color map the out-of-sample cross-sectional R^2 (R^2_{oos}) -calculated using 3-fold cross-validation- of the regression of the expected returns on the covariance matrix (risk factors) under the Elastic-Net penalty. In the right panel, we plot the R^2_{oos} (solid red) -calculated using 3-fold cross-validation- and the in-sample cross-sectional R^2_{is} (dashed black) of the regression of the expected returns on the covariance matrix (risk factors) under the Ridge penalty ($\gamma_1 = 0$). The confidence interval of the R^2_{oos}, i.e. $R^2_{oos} \pm 1 s.e.$ is drawn with dotted lines.

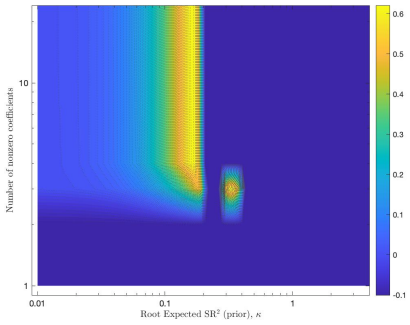


(a) Dual penalty

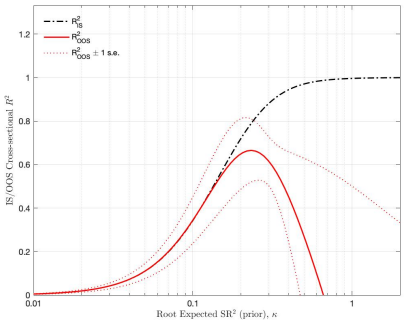


(b) L2 penalty

Figure 2: **PCs 25 anomaly portfolios** In the left panel, we plot in a color map the out-of-sample cross-sectional R^2 (R^2_{oos}) -calculated using 3-fold cross-validation- of the regression of the expected returns on the covariance matrix (risk factors) under the Elastic-Net penalty. In the right panel, we plot the R^2_{oos} (solid red) -calculated using 3-fold cross-validation- and the in-sample cross-sectional R^2_{is} (dashed black) of the regression of the expected returns on the covariance matrix (risk factors) under the Ridge penalty ($\gamma_1 = 0$). The confidence interval of the R^2_{oos}, i.e. $R^2_{oos} \pm 1 s.e.$ is drawn with dotted lines.

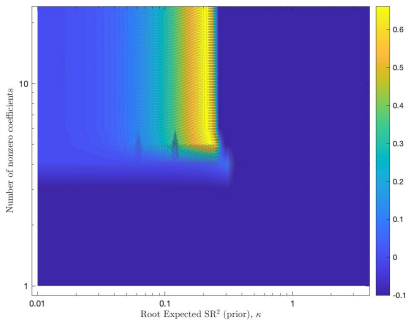


(a) Dual penalty

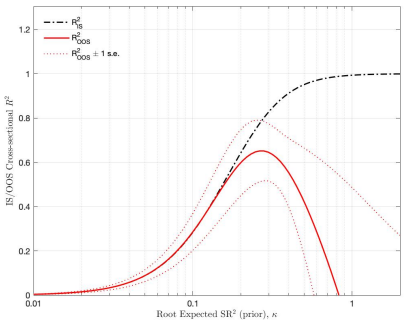


(b) L2 penalty

Figure 3: **Hybrid FF25P using 2 nonlinear factors +23 linear factors** In the left panel, we plot in a color map the out-of-sample cross-sectional R^2 (R^2_{oos}) -calculated using 3-fold cross-validation- of the regression of the expected returns on the covariance matrix (risk factors) under the Elastic-Net penalty. In the right panel, we plot the R^2_{oos} (solid red) -calculated using 3-fold cross-validation- and the in-sample cross-sectional R^2_{is} (dashed black) of the regression of the expected returns on the covariance matrix (risk factors) under the Ridge penalty ($\gamma_1 = 0$). The confidence interval of the R^2_{oos}, i.e. $R^2_{oos} \pm 1 s.e.$ is drawn with dotted lines.



(a) Dual penalty

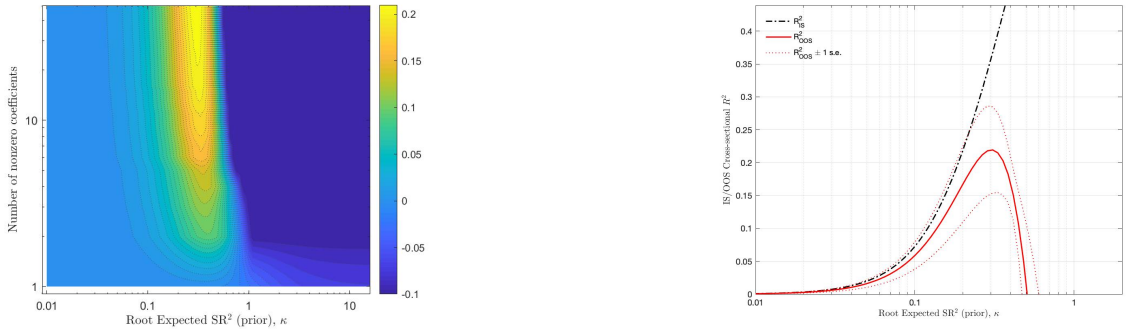


(b) L2 penalty

Figure 4: **Hybrid FF25P using 3 nonlinear factors +22 linear factors** In the left panel, we plot in a color map the out-of-sample cross-sectional R^2 (R^2_{oos}) -calculated using 3-fold cross-validation- of the regression of the expected returns on the covariance matrix (risk factors) under the Elastic-Net penalty. In the right panel, we plot the R^2_{oos} (solid red) -calculated using 3-fold cross-validation- and the in-sample cross-sectional R^2_{is} (dashed black) of the regression of the expected returns on the covariance matrix (risk factors) under the Ridge penalty ($\gamma_1 = 0$). The confidence interval of the R^2_{oos}, i.e. $R^2_{oos} \pm 1 s.e.$ is drawn with dotted lines.

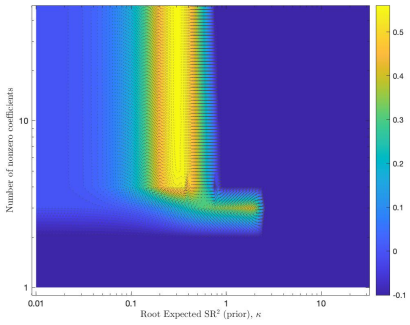
(a) Dual penalty (b) L2 penalty

Figure 5: **Raw 50 anomaly portfolios**. In the left panel, we plot in a color map the out-of-sample cross-sectional R^2 (R^2_{oos}) -calculated using 3-fold cross-validation- of the regression of the expected returns on the covariance matrix (risk factors) under the Elastic-Net penalty. In the right panel, we plot the R^2_{oos} (solid red) -calculated using 3-fold cross-validation- and the in-sample cross-sectional R^2_{is} (dashed black) of the regression of the expected returns on the covariance matrix (risk factors) under the Ridge penalty ($\gamma_1 = 0$). The confidence interval of the R^2_{oos}, i.e. $R^2_{oos} \pm 1s.e.$ is drawn with dotted lines.

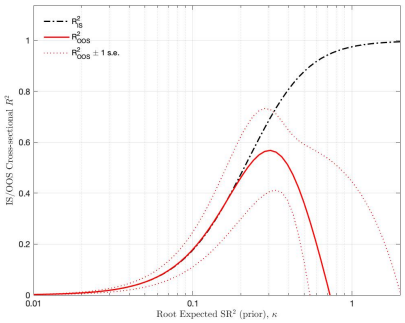


(a) Dual penalty (b) L2 penalty

Figure 6: **PCs 50 anomaly portfolios**. In the left panel, we plot in a color map the out-of-sample cross-sectional R^2 (R^2_{oos}) -calculated using 3-fold cross-validation- of the regression of the expected returns on the covariance matrix (risk factors) under the Elastic-Net penalty. In the right panel, we plot the R^2_{oos} (solid red) -calculated using 3-fold cross-validation- and the in-sample cross-sectional R^2_{is} (dashed black) of the regression of the expected returns on the covariance matrix (risk factors) under the Ridge penalty ($\gamma_1 = 0$). The confidence interval of the R^2_{oos}, i.e. $R^2_{oos} \pm 1s.e.$ is drawn with dotted lines.

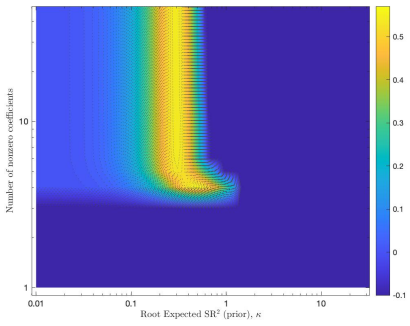


(a) Dual penalty

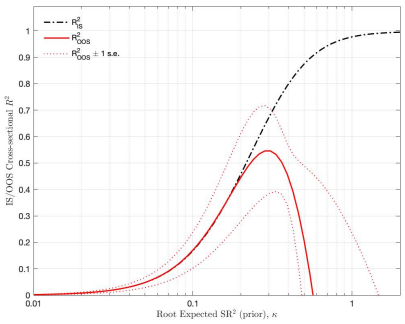


(b) L2 penalty

Figure 7: **Hybrid 50 factors using 2 nonlinear factors +48 linear factors**. In the left panel, we plot in a color map the out-of-sample cross-sectional R^2 (R^2_{oos}) -calculated using 3-fold cross-validation- of the regression of the expected returns on the covariance matrix (risk factors) under the Elastic-Net penalty. In the right panel, we plot the R^2_{oos} (solid red) -calculated using 3-fold cross-validation- and the in-sample cross-sectional R^2_{is} (dashed black) of the regression of the expected returns on the covariance matrix (risk factors) under the Ridge penalty ($\gamma_1 = 0$). The confidence interval of the R^2_{oos}, i.e. $R^2_{oos} \pm 1 s.e.$ is drawn with dotted lines.

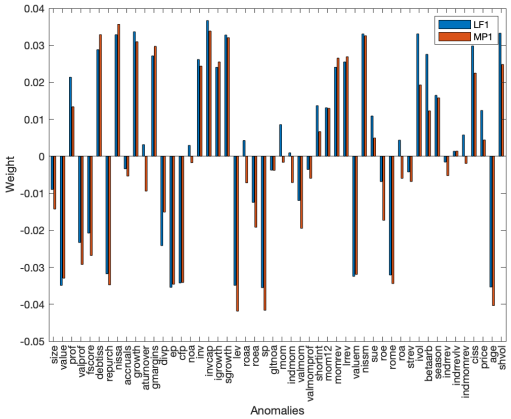


(a) Dual penalty

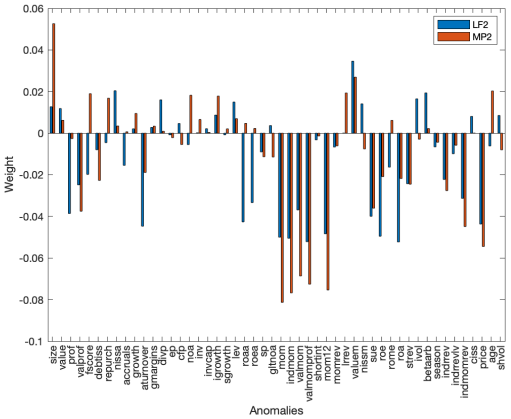


(b) L2 penalty

Figure 8: **Hybrid 50 factors using 3 nonlinear factors +47 linear factors**. In the left panel, we plot in a color map the out-of-sample cross-sectional R^2 (R^2_{oos}) -calculated using 3-fold cross-validation- of the regression of the expected returns on the covariance matrix (risk factors) under the Elastic-Net penalty. In the right panel, we plot the R^2_{oos} (solid red) -calculated using 3-fold cross-validation- and the in-sample cross-sectional R^2_{is} (dashed black) of the regression of the expected returns on the covariance matrix (risk factors) under the Ridge penalty ($\gamma_1 = 0$). The confidence interval of the R^2_{oos}, i.e. $R^2_{oos} \pm 1 s.e.$ is drawn with dotted lines.

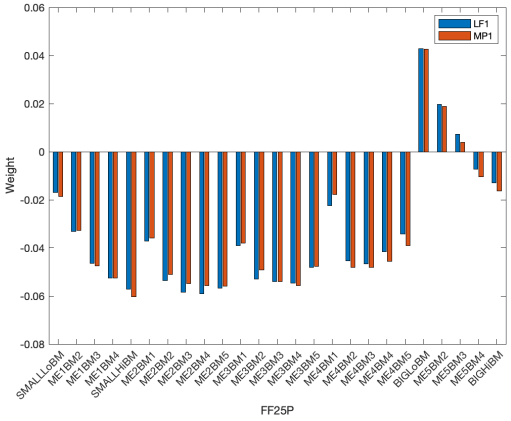


(a) LF1 versus MP1

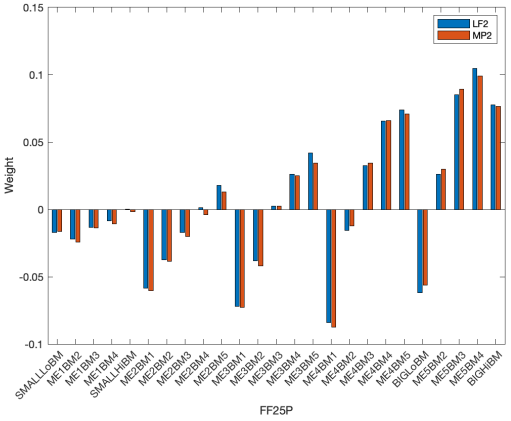


(b) LF2 versus MP2

Figure 9: **50 anomaly data**. These graphs plot the weight of the anomalies on the first two linear factors and mimicking portfolios

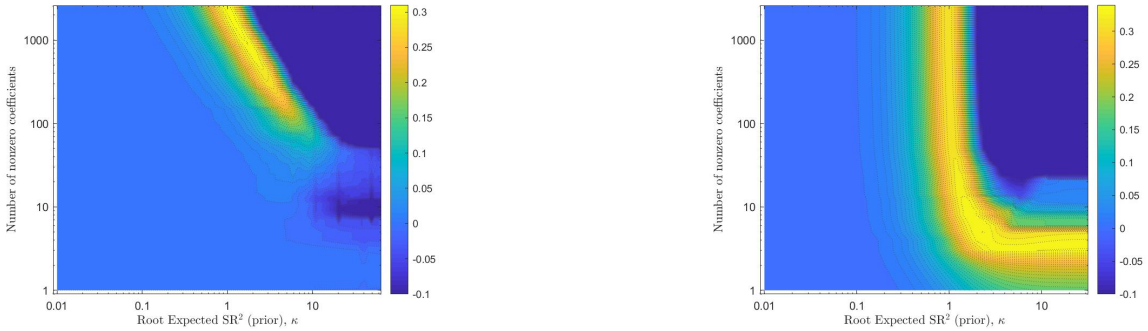


(a) LF1 versus MP1



(b) LF2 versus MP2

Figure 10: **FF25P data**. These graphs plot the weight of the 25 Fama-French portfolios on the first two linear factors and mimicking portfolios



(a) 2600 raw characteristics

(b) PCs of 2600 raw characteristics

Figure 11: **50 anomaly + interaction terms data**. We plot in a color map the out-of-sample cross-sectional R^2 (R^2_{oos}) -calculated using 3-fold cross-validation- of the regression of the expected returns on the covariance matrix (risk factors) under the Elastic-Net penalty. In the left panel, we rotate the stochastic discount factor in the raw characteristics space and in the right panel, we rotate the stochastic discount factor in the linear principal components space.

8 Appendix

8.1 Robustness check

We put in this section additional figures (12, 13, 14, 15) concerning the robustness case scenario for the Fama-French 25 ME/BM-sorted portfolios and the fifty anomaly portfolios. For robustness purposes, we replace the first k linear factors directly by the nonlinear principal components in the risk factors instead of their mimicking portfolios. And price a set of portfolios containing n factors where we replace the first k linear principal components by k mimicking portfolios. Moreover, we let the LARS-EN algorithm adds the factors starting by the model with no risk factors instead of starting by the model with our k mimicking portfolios as risk factors. The figures confirm the results presented in the section 4. We checked the robustness of the mimicking portfolios (nonlinear) factors are always in the optimal model giving the highest R^2_{oos}. Our hybrid factors models outperform the linear models.

8.2 Least Angle Regression

1. Initialize $\hat{\lambda}^{(0)} = 0$, $\mathcal{A} = argmax_j|\Sigma_j'\mu|$, $\nabla\hat{\lambda}^{(0)}_{\mathcal{A}} = -sign(\Sigma'_{\mathcal{A}}\mu)$,$\nabla\hat{\lambda}^{(0)}_{I} = 0$, $n = 0$.
2. While $\mathcal{I} \neq \emptyset$ do ;
3. $\delta_j = \min^+_{j\in\mathcal{A}} -\frac{\hat{\lambda}^{(n)}}{\nabla\hat{\lambda}^{(n)}_j}$
4. $\delta_i = \min^+_{i\in\mathcal{I}} \left\{ \frac{(\Sigma_i+\Sigma_j)'(\mu - X\hat{\lambda}^{(n)})}{(\Sigma_i+\Sigma_j)'(\Sigma\nabla\hat{\lambda}^{(n)})}, \frac{(\Sigma_i-\Sigma_j)'(\mu - \Sigma\hat{\lambda}^{(n)})}{(\Sigma_i-\Sigma_j)'(\Sigma\nabla\hat{\lambda}^{(n)})} \right\}$ where j is any index in $\mathcal{A}$.
5. $\delta = \min(\delta_j, \delta_i)$
6. if $\delta = \delta_j$ then move j from $\mathcal{A}$ to $\mathcal{I}$ else move i from $\mathcal{I}$ to $\mathcal{A}$.
7. $\hat{\lambda}^{(n+1)} = \hat{\lambda}^{(n)} + \delta\nabla\hat{\lambda}^{(n)}$
8. $\nabla\hat{\lambda}^{(n+1)}_{\mathcal{A}} = -\frac{1}{2}\left(\Sigma_{\mathcal{A}} + \gamma_2 I\right)^{-1}.sign(\hat{\lambda}^{(n+1)}_{\mathcal{A}})$
9. Update the value of n=n+1
10. end while
11. Output the series of coefficients $\Lambda = (\hat{\lambda}^{(0)}, \hat{\lambda}^{(1)}, ..., \hat{\lambda}^{(k)})$

8.3 Nonlinear principal components estimation

We normalize each linear factor to have 0 mean and 1 standard deviation LF0=(lf_{01},...,lf_{0k}) and set a grid on the standardized data lf_m, $m \in \{1...ng\}^k$ where

$$lf_{mi} \in min(lf_{0i}) : \frac{max(lf_{0i}) - min(lf_{0i})}{ng - 1} : max(lf_{0i}), i = 1, ..., k$$

The estimation of the density function $g(lf_m)$ is done by kernel smoothing. Then the convex function C is computed via gradient descent algorithm using

$$C_{n+1}(lf_m) = C_n(lf_m) + \tau(g(lf_m) - \Phi(\frac{\partial C_n(lf_m)}{\partial lf_m})det(\frac{\partial^2 C_n(lf_m)}{\partial lf_m \partial lf_m'}) \tag{18}$$

All the derivatives are computed using centered finite differences :

$$T_j(lf_m) \approx \frac{\partial C_n(lf)}{\partial lf}_j = \frac{C_n(lf_{m+\Delta_j}) - C_n(lf_{m-\Delta_j})}{2 * step}, \quad j = 1, 2, ..., k$$

and

$$\frac{\partial^2 C_n(lf_m)}{\partial lf_i \partial lf_j'} \approx \frac{C_n(lf_{m+\Delta_j+\Delta_i}) - C_n(lf_{m-\Delta_j+\Delta_i}) - C_n(lf_{m-\Delta_i+\Delta_j}) + C_n(lf_{m-\Delta_j-\Delta_i})}{||lf_{m+\Delta_i} - lf_{m-\Delta_i}||||lf_{m+\Delta_j} - lf_{m-\Delta_j}||}$$

For any boundary points, we use an appropriate noncentered finite differences version that is second-order accurate. Let us denote C^* the optimal convex function and by $x_m^* = T(lf_m)$ the transformed data. The nonlinear principal components are obtained by diagonalizing the matrix $\overline{J}$ defined by :

$$\overline{J} = \sum_{m \in \{1...ng\}^k} \Phi(x_m^*) lnJ(lf_m) \prod_{j=1}^{d} \frac{||x_{m+\Delta_j}^* - x_{m-\Delta_j}^*||}{2}$$

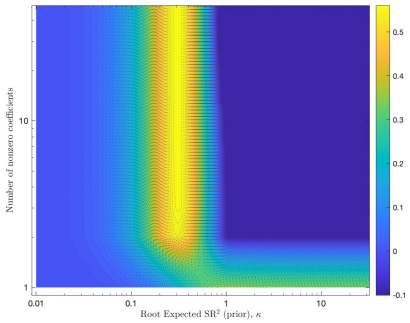
where $J(lf) = \frac{\partial^2 C^*(lf)}{\partial lf \partial lf'}$. The eigenvectors of $\overline{J}$ associated to the k highest eigenvalues : $e_1...e_k$. Finally, we interpolate the Brenier map to have the full nonlinear transformation of the original data $T(LF_1, LF_2, ..., LF_k)$. Therefore, the i^{th} nonlinear factor is $NLF_i = T(LF_1, LF_2, ..., LF_k)e_i$.

(a) Dual penalty (b) L2 penalty

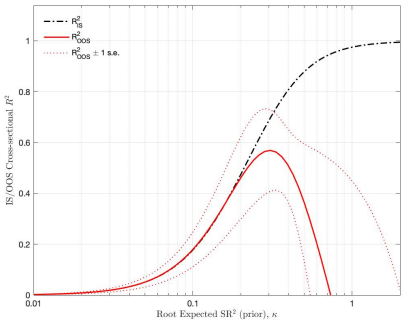
Figure 12: **Hybrid FF25P using 2 nonlinear factors +23 linear factors**. In the left panel, we plot in a color map the out-of-sample cross-sectional R^2 (R^2_{oos}) -calculated using 3-fold cross-validation- of the regression of the expected returns on the covariance matrix (risk factors) under the Elastic-Net penalty. In the right panel, we plot the R^2_{oos} (solid red) -calculated using 3-fold cross-validation- and the in-sample cross-sectional R^2_{is} (dashed black) of the regression of the expected returns on the covariance matrix (risk factors) under the Ridge penalty ($\gamma_1 = 0$). The confidence interval of the R^2_{oos}, i.e. $R^2_{oos} \pm 1 s.e.$ is drawn with dotted lines.

(a) Dual penalty (b) L2 penalty

Figure 13: **Hybrid FF25P using 3 nonlinear factors +22 linear factors** In the left panel, we plot in a color map the out-of-sample cross-sectional R^2 (R^2_{oos}) -calculated using 3-fold cross-validation- of the regression of the expected returns on the covariance matrix (risk factors) under the Elastic-Net penalty. In the right panel, we plot the R^2_{oos} (solid red) -calculated using 3-fold cross-validation- and the in-sample cross-sectional R^2_{is} (dashed black) of the regression of the expected returns on the covariance matrix (risk factors) under the Ridge penalty ($\gamma_1 = 0$). The confidence interval of the R^2_{oos}, i.e. $R^2_{oos} \pm 1 s.e.$ is drawn with dotted lines.

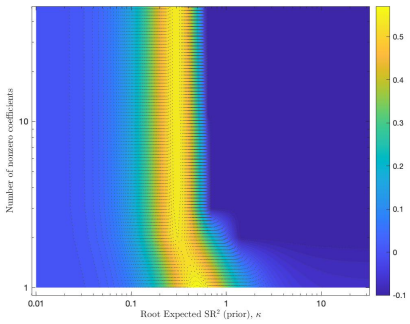


(a) Dual penalty

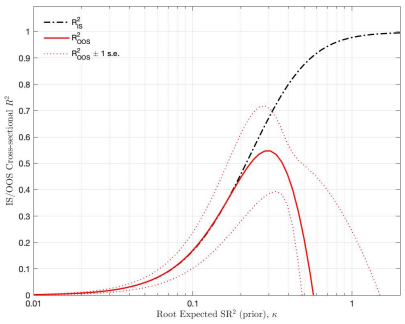


(b) L2 penalty

Figure 14: **Hybrid 50 anomaly portfolios using 2 nonlinear factors +48 linear factors** In the left panel, we plot in a color map the out-of-sample cross-sectional R^2 (R^2_{oos}) -calculated using 3-fold cross-validation- of the regression of the expected returns on the covariance matrix (risk factors) under the Elastic-Net penalty. In the right panel, we plot the R^2_{oos} (solid red) -calculated using 3-fold cross-validation- and the in-sample cross-sectional R^2_{is} (dashed black) of the regression of the expected returns on the covariance matrix (risk factors) under the Ridge penalty ($\gamma_1 = 0$). The confidence interval of the R^2_{oos}, i.e. $R^2_{oos} \pm 1 s.e.$ is drawn with dotted lines.



(a) Dual penalty



(b) L2 penalty

Figure 15: **Hybrid 50 anomaly portfolios using 3 nonlinear factors +47 linear factors** In the left panel, we plot in a color map the out-of-sample cross-sectional R^2 (R^2_{oos}) -calculated using 3-fold cross-validation- of the regression of the expected returns on the covariance matrix (risk factors) under the Elastic-Net penalty. In the right panel, we plot the R^2_{oos} (solid red) -calculated using 3-fold cross-validation- and the in-sample cross-sectional R^2_{is} (dashed black) of the regression of the expected returns on the covariance matrix (risk factors) under the Ridge penalty ($\gamma_1 = 0$). The confidence interval of the R^2_{oos}, i.e. $R^2_{oos} \pm 1 s.e.$ is drawn with dotted lines.