# Nonlinearities and Pricing Complexity in the Cross-Section of Stock Returns

Fabio Girardi* Lukas Koerber† Christian Schlag‡

December 15, 2025

**Abstract**

We evaluate the pricing performance of a robust SDF spanned by a broad cross-section of factor returns. Methodologically, we combine kernel principal component analysis—which extracts factors as nonlinear functions of a high-dimensional set of firm characteristics—with novel regularization techniques. Allowing for nonlinearities enhances the model's performance in explaining a wide range of prominent cross-sectional stock-return anomalies and reduces pricing errors. We further decompose the mean–variance efficient portfolio into linear and nonlinear components and study their relative contributions across different states of the economy. Out-of-sample, incorporating nonlinearities increases the Sharpe ratio of the mean–variance efficient portfolio by roughly 25% to 2.15.

**Keywords:** stock returns, stochastic discount factor, nonlinearities, uncertainty.

**JEL Codes:** G10, G11, G12, G14, G17

*Finance Department, Vienna University of Economics and Business.
†Finance Department, Goethe University Frankfurt.
‡Finance Department, Goethe University Frankfurt, and Leibniz Institute for Financial Research SAFE.

# 1 Introduction

The characterization of the stochastic discount factor (SDF, hereafter) is a central objective of modern asset pricing research. While the absence of arbitrage opportunities guarantees the existence of a SDF spanned by traded payoffs, identifying the relevant sources of priced risk remains a fundamental and challenging task. Characteristic-based factor models have emerged as a key tool in empirical asset pricing to explain cross-sectional variation in expected returns. Pioneered by Rosenberg (1974) and further developed by Fama and French (1992, 1993, 1996), these models construct risk factors from firm-specific characteristics, such as size (i.e., equity market capitalization) or the book-to-market equity ratio. In the recent decades, financial economists have identified numerous firm characteristics that predict future stock returns (see, e.g., Cochrane, 2011; Harvey et al., 2016); however, which characteristics are truly relevant remains unclear (see, e.g., Hou et al., 2015).

Historically, a strong and popular assumption has been that risk premia are linear functions of these characteristic-based factors. However, although linearity facilitates estimation and interpretation, it also renders models sensitive to outliers and obviously constrains their ability to capture complex, i.e., potentially nonlinear, relationships between firm characteristics and returns, thereby creating the risk of omitting important dimensions in the variation of expected returns across assets.

In this paper, we build on a growing stream of asset pricing literature that leverages statistical learning methods to relax the linearity assumption and to provide more flexible approaches that help to better capture the complex shape of the SDF. We assess the pricing performance of a robust SDF spanned by principal components constructed from a high-dimensional set of characteristic-based risk factors. To construct these factors, we employ kernel principal component analysis (henceforth abbreviated K-PCA) and refer to the resulting components as kernel principal components (K-PCs). K-PCA generalizes classical PCA (Hotelling, 1933) by allowing factors to reflect nonlinear transformations and interactions of firm characteristics.[1]

Our analysis contributes to the asset pricing literature by assessing the pricing performance of the linear and nonlinear components of the SDF. Specifically, we decompose the SDF implied by nonlinear kernels (in our case, polynomial and Gaussian) into (i) a linear component—its projection onto the span of the mean–variance efficient (MVE) portfolio implied by a linear kernel—and (ii) an orthogonal complement. We interpret the orthogonal component as the nonlinear (complex) contribution to the SDF, i.e., the portion omitted when factors are constructed exclusively as linear functions of characteristics. This decomposition induces a

[1] K-PCs was introduced by Schölkopf et al. (1997) and applied to asset pricing by, among others, Kozak (2020).

bivariate factor structure. We then quantify the incremental pricing contribution of the nonlinear component across prominent cross-sectional equity-return anomalies and under different macro-financial conditions.

In order to evaluate the pricing performance of the resulting two-factor SDF, we start by considering an unconditional factor model, and we show that accounting for nonlinearities increases explanatory power across different economic categories. Figure 1 reports the Hansen and Jagannathan (1997) (henceforth abbreviated HJ) distance as a measure of pricing performance across 82 characteristic-based test assets, grouped into 24 economic categories (Chen and Zimmermann, 2022). Results show that constructing the SDF nonlinearly with a Gaussian kernel significantly reduces the HJ distance across nearly all categories. Moreover, on average, test-asset $R^2$ rises from 17% to 25%, while the average cross-sectional pricing error is reduced by half (from 2% to 1%).

On average, we find that the Gaussian kernel is better suited to capture complex, nonlinear relationships, owing to its greater flexibility than a polynomial kernel. Unlike a second-order polynomial, the Gaussian kernel adapts to subtle patterns in high-dimensional data without imposing a rigid global structure. This flexibility reflects the fact that the Gaussian kernel is universal—it can approximate a very broad class of continuous functions—so it is well equipped to model rich and heterogeneous SDF dynamics. This, in turn, translates into improved pricing performance.

We document substantial heterogeneity in the unconditional exposures of anomaly portfolios to the SDF's nonlinear component. For example, the CAPM beta–sorted portfolio and the beta-against-beta (BAB) factor (Frazzini and Pedersen, 2014) exhibit large positive loadings on the nonlinear component, whereas cash-flow-to-price (Lakonishok et al., 1994) and book-to-market portfolios load strongly negatively. Notably, for several portfolios—such as market leverage and equity duration—the exposures to the linear and nonlinear SDF components have large magnitude and opposite signs.

Then, we generalize our analysis introducing a conditional factor model that allows time variation in factor loadings as a function of macro-financial conditions. In particular, we let the loadings depend on state variables such as inflation, industrial production, the corporate credit spread, financial uncertainty, and policy uncertainty. This additional flexibility further increases the average explained variation of anomaly portfolios from 25% to 33%.

To further validate our results, we compare out-of-sample Sharpe ratios for the MVE portfolio implied by nonlinear kernels (polynomial and Gaussian) against a linear kernel, employing the "Constrained Universal Portfolio Shrinkage Approximator" (CUPSA) regularization (Kelly et al., 2024). We find that nonlinear kernels deliver better performance. In particular, the Gaussian kernel generates the best results, achieving

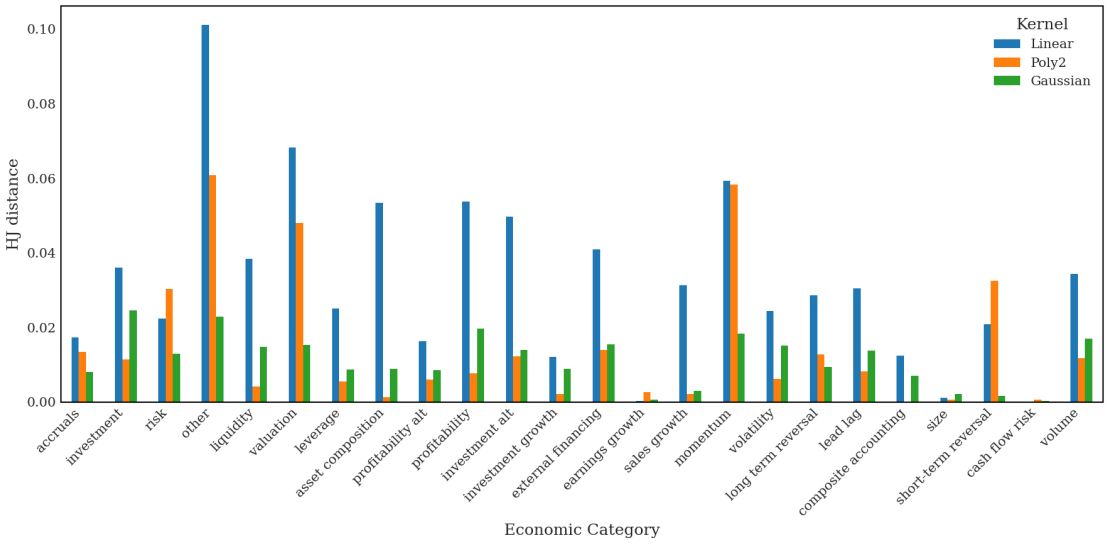



Figure 1: Hansen-Jagannathan Distance. The figure reports the in-sample unconditional Hansen-Jagannathan (HJ) distance for a large set of 82 cross-sectional stock return anomalies. Test assets are divided into 24 groups, based on the economic category definition by Chen and Zimmermann (2022). Underlying SDFs are constructed via K-PCA and KNS regularization. The blue bar reports the pricing performance delivered by the linear kernel, while the orange and green bar report results for the second-order polynomial and the Gaussian kernel, respectively. The estimation sample ranges from 1974 to 2022. Data have monthly frequency.

an out-of-sample Sharpe ratio of 2.15, compared to 1.73 for the linear kernel. Thus, allowing for nonlinearities significantly expands the boundaries of the mean-variance frontier, increasing the Sharpe ratio of the MVE portfolio by approximately 25%. Notably, our findings are consistent with those of Kelly et al. (2024), who argue that the CUPSA estimator performs better out-of-sample when principal components are constructed from a very large set of variables. This suggests that low-variance PCs (which are not penalized under the CUPSA method) also contribute meaningfully to the improved Sharpe ratio.

Methodologically, since we consider a high-dimensional set of 82 characteristics (Chen and Zimmermann, 2022), our analysis is subject to the curse of dimensionality. In fact, simply considering second-order polynomials already increases the overall number of factors to 3,485, and adding higher-order terms of course further intensifies this problem. In this context, proper regularization techniques must be employed to achieve a robust estimation of the weights of different assets or portfolios in the SDF. Besides the purely statistical argument, regularization methods are also economically motivated by the Hansen and Jagannathan (1991) duality between the variance of the SDF and the maximum attainable Sharpe ratio. Motivated by this relationship, Kozak et al. (2018) argue that the absence of near-arbitrage opportunities implies that the SDF can be well approximated by a function of a few dominant sources of return variation, corresponding to the components associated with high eigenvalues.

We compare two recent regularization techniques that operate on the K-PCs constructed from factor returns. These methods differ in the extent to which they penalize low-eigenvalue PCs. Kozak et al. (2020) (KNS) introduce a Ridge estimator that shrinks the contribution of low-variance components in the SDF. While endorsing the theoretical foundation of this approach, Kelly et al. (2024) challenge its empirical validity. They argue that data noise may render low-variance PCs empirically relevant, even if the true SDF is sparse in the space of PCs. As a result, the authors claim that recovering a truly PC-sparse representation is infeasible in practice. To address this limitation, Kelly et al. (2024) propose CUPSA, an estimator which efficiently weights, rather than discards, low-variance PCs, thus retaining potentially valuable pricing information. By comparing these two methods, we investigate whether low-eigenvalue PCs are relevant for SDF characterization in a nonlinear, high-dimensional setting. We find that, in the linear benchmark, performance is concentrated roughly in the first 20 PCs, and adding further low-eigenvalue PCs does not translate into better performance. In contrast, under polynomial and Gaussian kernels, CUPSA's more efficient weighting of low-eigenvalue PCs yields rising out-of-sample Sharpe ratios and positive alphas as $J$ grows—even beyond 50 PCs—indicating incremental pricing information that ridge-based KNS misses.

### 1.1 Literature

Early seminal models, such as the Capital Asset Pricing Model (CAPM) by Sharpe (1964), rely on economic theory to guide factor selection, deriving risk factors from assumptions about investor beliefs and market equilibrium. In contrast, the Arbitrage Pricing Theory (APT) introduced by Ross (1976) marks a shift toward reduced-form models, where factor selection is primarily driven by statistical properties rather than explicit economic foundations. The APT has subsequently inspired a major stream of modern asset pricing research—characteristics-based models—where risk factors are constructed as tradable portfolio returns formed by sorting stocks on firm-level characteristics. These factors are then linearly combined to form the stochastic discount factor (SDF). The most prominent examples include the value and size factors introduced by Fama and French (1993, 2015).

With this paper, we relate to different streams of the financial economics literature. The first concerns dimensionality reduction. In recent years, principal component analysis (PCA) has regained popularity in empirical asset pricing as a tool to address the so-called *factor zoo*—a term coined by Cochrane (2011) to describe the proliferation of proposed return predictors in the literature. Cochrane (2011) argues that many of these factors lack economic justification and are likely the result of data mining, thereby exposing the limitations of conventional approaches used to assess the statistical significance of newly proposed factors.

Researchers have addressed this challenge of dimensionality reduction in different ways.

One set of scholars has aimed at consolidating the large number of potential return predictors through the introduction of new statistical testing frameworks better suited to jointly control for many variables when assessing the predictive power of a newly proposed predictor (see, e.g., Harvey et al., 2016; Feng et al., 2020). The underlying motivation to tame the factor zoo is to derive an efficient characteristics-sparse SDF, since sparse models with only a few observable factors are generally preferable relative to latent factor models because they are easier to interpret and to link to economic theories (Feng et al., 2020). Concrete attempts include Fama and French (1993, 2015), Hou et al. (2015), and Barillas and Shanken (2018).

More recently, a growing body of research has revisited PCA—along with its various generalizations—as a means to extract latent or weak statistical factors from the factor zoo (see, e.g., Kozak et al., 2018; Kelly et al., 2019; Lettau and Pelger, 2020; Giglio and Xiu, 2021; Bryzgalova et al., 2023; Giglio et al., 2025). Contrary to earlier views that advocate a sparse representation of the stochastic discount factor (SDF) in the space of firm characteristics, these studies challenge the existence of an efficient characteristic-sparse SDF. Supporting this perspective, Bryzgalova et al. (2023) conduct an extensive empirical analysis and conclude that large subsets of predictor variables are necessary to capture the full pricing implications of the SDF. Instead, this literature proposes that the SDF admits a *PC-sparse* representation. Kozak et al. (2018) formally show that, under the economically plausible assumption of no near-arbitrage opportunities and the empirical observation that characteristics-managed portfolios exhibit a strong factor structure, the SDF can be well approximated by a function of a few high-eigenvalue principal components of factor returns. Kozak et al. (2018) and Kozak et al. (2020) provide empirical support for this hypothesis, demonstrating that a PC-sparse SDF delivers strong and robust pricing performance across a broad cross-section of traded anomaly portfolios.

Another important dimension in recent research concerns the role of nonlinearities. In practice, expected returns and exposures to risk factors often do not vary linearly with economic variables such as macroeconomic indicators or firm characteristics. This view is supported by both theoretical models of asset pricing (see, e.g., Campbell and Cochrane, 1999; Bansal and Yaron, 2004; He and Krishnamurthy, 2013) and empirical evidence (see, e.g., Hong et al., 2000; Ang et al., 2006; Stambaugh et al., 2012). Nevertheless, the empirical asset pricing literature—particularly the strand focused on identifying the SDF in the context of factor models—has traditionally favored sparsely specified linear models, largely ignoring nonlinearities and interaction effects in predictor variables as a way to regularize the estimation problem.

More recently, machine learning (ML) has provided financial economists with a flexible and powerful set of tools to model complex relationships in high-dimensional settings, where the number of potential predictors

is large relative to the number of observations, and the signal-to-noise ratio is typically low (Nagel, 2021). Several studies emphasize the importance of capturing nonlinearities to explain the cross-section of returns. For example, Moritz and Zimmermann (2016) introduce tree-based conditional portfolio sorts—a method inspired by the random forest literature—to construct return predictors from past returns. Their findings suggest a complex and nonlinear relationship between past and future returns, driven by interactions across different return horizons. Similarly, Gu et al. (2020) perform a comprehensive comparison of traditional linear models and various ML methods, showing that regression trees and neural networks consistently outperform linear alternatives, primarily due to their ability to capture nonlinear interactions among firm characteristics. Freyberger et al. (2020) further underscore the importance of nonlinearities by applying an adaptive group LASSO, although they exclude interaction terms because of the curse of dimensionality. Overall, while empirical support for nonlinear and interactive effects is growing, such evidence has so far remained limited—often restricted to a small subset of variables—mainly due to the computational and statistical challenges associated with modeling interactions at scale (Kozak et al., 2020).[2]

Within this broader class of nonlinear models, our work is closely related to Kozak et al. (2020), who combine nonlinear factors with an economically motivated Ridge estimator to estimate SDF weights. Specifically, they construct factors based on a large set of firm characteristics, their second- and third-order polynomial transformations, and all pairwise interaction terms. The authors show that when appropriate regularization is applied, such nonlinear factors can substantially improve the out-of-sample efficiency of the SDF. Kozak (2020) extend this framework further by introducing K-PCA, which captures arbitrary nonlinear relationships among characteristics. Their findings indicate substantial gains in pricing performance, including a doubling of the Sharpe ratio of the MVE portfolio implied by the SDF.

## 2 Characteristics-Based SDF

We consider an economy populated by a cross-section of individual firms $i \in \{1, \dots, N\}$, and we denote by $R_t^e : \Omega \to \mathbb{R}^N$ the vector of excess returns and by $Z_t : \Omega \to \mathbb{R}^{N \times D}$ the matrix of firm characteristics realized at time $t$. We assume the economy is arbitrage free, therefore there exists a strictly positive stochastic discount factor (SDF) $M_t$ which price all the traded portfolios

$$M_t^* = \mathbb{E}\left[M_t^*\right] - b_{t-1}^\top \left(R_t^e - \mathbb{E}\left[R_t^e\right]\right), \tag{1}$$

[2]See, for example, Asness et al. (2013) for a detailed analysis of the interaction between momentum and value strategies.

which, under no arbitrage, becomes

$$M_t^* = \max\left(\mathbb{E}\left[M_t^*\right] - b_{t-1}^\top \left(R_t^e - \mathbb{E}\left[R_t^e\right]\right), 0\right). \tag{2}$$

To parametrize the loadings of the SDF, the canonical approach adopted by characteristics-based asset pricing models is to set

$$b_{t-1} = Z_{t-1}^\top b, \tag{3}$$

where $b$ is an $D \times 1$ vector of time-invariant coefficients. Under this specification, the SDF in equation (1) is spanned by the characteristics-based factor returns

$$M_t^* = \mathbb{E}\left[M_t^*\right] - b^\top \left(Z_{t-1}^\top R_t^e - \mathbb{E}\left[Z_{t-1}^\top R_t^e\right]\right) \tag{4}$$

$$= \mathbb{E}\left[M_t^*\right] - b^\top \left(F_t - \mathbb{E}\left[F_t\right]\right). \tag{5}$$

To characterize the vector of time-invariant loadings of the SDF, we follow the construction of Hansen and Jagannathan (1991), who derive restrictions on the first two moments of any valid SDF given a set of observed asset payoffs, even in the absence of a risk-free asset to pin down the mean of the SDF, which, following a common choice in the literature, we assume to be equal to one $\mathbb{E}\left[M_t^*\right] = 1$. Specifically, the minimum variance stochastic discount factor that prices all the elements in the space of traded payoffs satisfies

$$\min_{M_t \in L^2} \quad \mathbb{E}\left[(M_t - 1)^2\right] \tag{6}$$

$$\text{s.t.} \quad \mathbb{E}_{t-1}\left[M_t R_t^e\right] = 0, \tag{7}$$

and the problem solution is given by $b = \Sigma_F^{-1}\mathbb{E}\left[F\right]$.[3] Using the eigendecomposition of the matrix $Cov\left(F_t\right) := \Sigma_F = Q\Lambda Q^\top$, equation (5) becomes

$$M_t^* = 1 - \mathbb{E}^\top\left[F_t\right] Q\Lambda^{-1}Q^\top \left(F_t - \mathbb{E}\left[F_t\right]\right) \tag{8}$$

$$= 1 - \mathbb{E}^\top\left[P_t\right] \Lambda^{-1}\left(P_t - \mathbb{E}\left[P_t\right]\right), \tag{9}$$

in which $P_t = Q^\top F_t \in \mathbb{R}^{\dim(F_t)}$ and $\mathbb{E}^\top\left[P_t\right]\Lambda^{-1} = \left[\frac{\mathbb{E}^\top[P_{t,1}]}{\lambda_1}, \ldots, \frac{\mathbb{E}^\top[P_{t,L}]}{\lambda_L}\right]$.

[3] The vector $b$ can be interpreted as the result of a cross-sectional regression $b = \left(\Sigma_F\Sigma_F\right)^{-1}\Sigma_F\mathbb{E}\left[F\right]$ of the expected return of the factor on the pairwise covariances among the factors.

## 2.1 Kernel-Based Risk Factors

In this paper, we want the SDF to be spanned not only by factors constructed from separable effects of firm characteristics, but also by factors that capture interactions among characteristics (see, e.g., Kozak, 2020). Kernel PCA (K-PCA) provides a generalization of standard PCA, which has recently re-gained popularity for the regularization of SDF models built on a large set of characteristics-managed portfolios (see, e.g., Kozak et al., 2018, 2020).

To this end, we define a function $\phi : \mathbb{R}^{N\times K} \to \mathbb{R}^{N\times L}$ that maps that the original data $Z_t$ into a higher-dimensional space, where $L \in \mathbb{N} \cup \{+\infty\}$. Such a function is often referred to as a feature map, and it generalizes the structure of the SDF loadings reported in equation (3), by allowing the construction of risk factors as nonlinear functions of firm characteristics: $F_t = \phi^\top(Z_{t-1})R_t \in \mathbb{R}^L$.

**Assumption 1.** *The unconditional mean of $F$ is zero, i.e., $\mathbb{E}[F_t] = 0$.*

Under Assumption 1, $Cov(F_t) = \mathbb{E}\left[F_t F_t^\top\right] = \Sigma \in \mathbb{R}^{L\times L}$. Let $V = [v_1, \ldots, v_L]$ be the $L \times L$ matrix of the right eigenvectors of $\Sigma$. It follows that

$$\Sigma V = V\Lambda, \tag{10}$$

where $\Lambda = diag(\lambda_1, \ldots, \lambda_L)$ is the diagonal matrix of eigenvalues. Replacing population means with sample averages, we obtain

$$\Sigma = \frac{1}{T}\sum_{t=1}^{T} F_t F_t^\top = \frac{1}{T}F^\top F, \tag{11}$$

where

$$F = \begin{bmatrix} F_1^\top \\ \vdots \\ F_T^\top \end{bmatrix} = \begin{bmatrix} F_{1,1} & \cdots & F_{1,L} \\ \vdots & \ddots & \vdots \\ F_{T,1} & \cdots & F_{T,L} \end{bmatrix} \in \mathbb{R}^{T\times L}. \tag{12}$$

Let $F = QDV^T$ be the singular value decomposition (SVD) of $F$, in which $Q \in \mathbb{R}^{T\times T}$ and $V \in \mathbb{R}^{L\times L}$ are unitary matrices (i.e., with orthonormal rows and columns) of the left and right singular values of $F$, respectively. $D \in \mathbb{R}^{T\times L}$ is a rectangular diagonal matrix $D = diag(\sigma_1, \ldots, \sigma_r)$, $r = \min(T, L)$, containing the singular values of $F$.

Combining the SVD of $F$ with the definition of principal components

$$P = FV = QDV^\top V = QD \in \mathbb{R}^{T\times L}. \tag{13}$$

in which $P = [P_1, \ldots, P_L] \in \mathbb{R}^{T \times L}$. From the equations above, it follows that

$$\sigma_i P_i = \sigma_i F v_i = \sigma_i^2 q_i = \lambda_i q_i = \Omega q_i \tag{14}$$

$$FV = QDV^\top V = QD \Rightarrow Q = FVD^{-1}. \tag{15}$$

where $\Omega = FF^\top \in \mathbb{R}^{T \times T}$ given by[4]

$$\Omega = \begin{bmatrix} \langle F_1, F_1 \rangle, \ldots, \langle F_1, F_T \rangle \\ \vdots \\ \langle F_T, F_1 \rangle, \ldots, \langle F_T, F_T \rangle \end{bmatrix}, \tag{16}$$

is the Gram matrix. In particular, $\Omega$ is a matrix where each entry is an inner product of two $L$-dimensional return vectors, which may be infeasible to compute explicitly. However, due to the properties of the kernel function, we can replace dot product operations of mapped firm characteristics in the feature space by much cheaper kernel evaluations. Observe that we can expand each entry of $\Omega$ in the following way:

$$\Omega_{i,j} = \langle F_i, F_j \rangle = \langle \phi^\top (Z_{i-1}) R_i, \phi^\top (Z_{j-1}) R_j \rangle \tag{17}$$

$$= R_i^\top \langle \phi (Z_{i-1}), \phi (Z_{j-1}) \rangle R_j \underbrace{=}_{\text{kernel trick}} R_i^\top \mathcal{K} (Z_{i-1}, Z_{j-1}) R_j. \tag{18}$$

Additionally, note that the matrix $\mathcal{K} (Z_{i-1}, Z_{j-1})$ has entries

$$\left[\mathcal{K}(Z_{i-1}, Z_{j-1})\right]_{nm} = \phi(z_{n,i-1})^\top \phi(z_{m,j-1}) \underbrace{=}_{\text{kernel trick}} k (z_{n,i-1}, z_{m,j-1})$$

such that each element is a kernel evaluation $k(\cdot, \cdot)$ between firm-level characteristic vectors. Therefore, in practice, we never form $\phi(\cdot)$ explicitly and compute $\mathcal{K}$ directly via the chosen kernel (e.g., polynomial or Gaussian), which is a much cheaper numerical operation.

As a result, we can compute the PC projections of high-dimensional factor return vectors in the following

[4]Since $D$ is not square matrix, $D^{-1}$ is a rectangular diagonal matrix defined through the pseudo-inverse $D^+$.

way:

$$P_i = Fv_i = \frac{1}{\sigma_i}\lambda_i q_i = \frac{1}{\sigma_i}\Omega q_i \tag{19}$$

$$= \begin{bmatrix} \sum_{t=1}^{T} \frac{q_{i,t}}{\sigma_i} R_1^\top \mathcal{K}\left(Z_0, Z_{t-1}\right) R_t \\ \vdots \\ \sum_{t=1}^{T} \frac{q_{i,t}}{\sigma_i} R_T^\top \mathcal{K}\left(Z_{T-1}, Z_{t-1}\right) R_t \end{bmatrix}. \tag{20}$$

### 2.1.1 Out-of-sample K-PCs.

Let $F_{T+k} = \phi(Z_{T+k-1})^\top R_{T+k}$ be a new out-of-sample point. Using the in-sample SVD identity $F = QDV^\top$, its $i$-th PC score can be written as

$$P_{i,T+k} = v_i^\top F_{T+k} = \frac{1}{\sigma_i} q_i^\top \left(F\, F_{T+k}\right) = \frac{1}{\sigma_i} \sum_{t=1}^{T} q_{t,i} \left\langle F_t, F_{T+k} \right\rangle, \qquad i = 1, \ldots, r, \tag{21}$$

where $r = \text{rank}(\Omega) \leq \min(T, L)$. By the kernel trick, we can expand the inner product explicitly:

$$\left\langle F_t, F_{T+k} \right\rangle = \left\langle \phi(Z_{t-1})^\top R_t,\ \phi(Z_{T+k-1})^\top R_{T+k} \right\rangle \tag{22}$$

$$= R_t^\top \left\langle \phi(Z_{t-1}),\ \phi(Z_{T+k-1}) \right\rangle R_{T+k} \tag{23}$$

$$= R_t^\top\, \phi(Z_{t-1})\, \phi(Z_{T+k-1})^\top\, R_{T+k} \tag{24}$$

$$= R_t^\top\, \mathcal{K}(Z_{t-1}, Z_{T+k-1})\ R_{T+k}, \tag{25}$$

where the matrix $\mathcal{K}(Z_{t-1}, Z_{T+k-1})$ has again entries given as

$$\left[\mathcal{K}(Z_{t-1}, Z_{T+k-1})\right]_{nm} = \underbrace{\phi(z_{n,t-1})^\top \phi(z_{m,T+k-1})}_{k(z_{n,t-1},\, z_{m,T+k-1})}.$$

Again, each element of $\mathcal{K}$ is a kernel evaluation $k(\cdot,\cdot)$ between firm-level characteristic vectors. Therefore, we never have to explicitly compute the computationally costly, high-dimensional mapping function $\phi(\cdot)$. By plugging Equation (25) into Equation (21), we obtain the OOS PC scores as an eigenvector-weighted average of similarity evaluations between the new observation and all $T$ training periods (scaled by $1/\sigma_i$):

$$P_{i,T+k} = \frac{1}{\sigma_i} \sum_{t=1}^{T} q_{t,i} \left\langle F_t, F_{T+k} \right\rangle = \frac{1}{\sigma_i} \sum_{t=1}^{T} q_{t,i}\, R_t^\top\, \mathcal{K}(Z_{t-1}, Z_{T+k-1})\ R_{T+k}. \tag{26}$$

Stacking $\{P_{i,T+k}\}_{i=1}^{r}$ across $k$ yields the matrix of OOS projections, $P_{\text{OOS}}$.

## 2.2 Regularization of the K-PCA

Under the Hansen and Jagannathan (1991) characterization of the pricing kernel, the variance of the SDF is:

$$\begin{aligned} \text{Var}\left(M_t^*\right) &= Var\left(b^\top F_t\right) = b^\top \Sigma b & (27) \\ &= \sum_{l=1}^{L} \lambda_l^{-1} \left( \sum_{k=1}^{L} \mathbb{E}\left[F_{k,t}\right] q_{k,l} \right)^2 = \mathbb{E}^\top\left[P_t\right] \Lambda^{-1} \mathbb{E}\left[P_t\right]. & (28) \end{aligned}$$

Among others, Kozak et al. (2018) argue that small eigenvalues can cause instability, potentially leading to an explosive variance of $M_t^*$.

More generally, since the maximum Sharpe ratio attainable in the economy is determined by the volatility of the stochastic discount factor, the potential problem of an economically implausible variance motivates the need for regularization. In the following subsection, we present different regularization methods explored in the literature for a robust estimation of the SDF in high-dimensional settings. Specifically, we will use a ridge-type regularization for our in-sample analysis and the CUPSA regularization proposed by Kelly et al. (2024) for the out-of-sample analysis.

### 2.2.1 Ridge-type regularization

Kozak et al. (2020) (KNS) estimate the SDF with an $L^2$ (ridge) penalty that shrinks coefficients more strongly in directions associated with low-variance principal components (PCs). Let $\bar{\Sigma}$ denote the sample covariance matrix of factor (PC) returns and $\bar{\mu}$ their sample mean. For a given penalty level $z \geq 0$, the ridge SDF weights are

$$\hat{b}_z = (\bar{\Sigma} + zI)^{-1} \bar{\mu}. \tag{29}$$

In the PC basis, if $\bar{\Sigma} = VDV^\top$ with $D = \text{diag}(d_1, \ldots, d_L)$ and $\bar{\mu}_P := V^\top \bar{\mu}$, then

$$\hat{b}_P(f_z) = (D + zI)^{-1} \bar{\mu}_P \quad \Rightarrow \quad \hat{b}_{P,j}(f_z) = \frac{\bar{\mu}_{P,j}}{d_j + z} = \left(\frac{d_j}{d_j + z}\right) \underbrace{\frac{\bar{\mu}_{P,j}}{d_j}}_{\text{unpenalized}},$$

so each PC's coefficient is scaled by the shrinkage factor $d_j/(d_j + z) \in [0, 1]$, with smaller $d_j$ implying stronger shrinkage.

KNS select $z$ from a predefined grid $Z = \{z_\ell\}_{\ell=1}^{M}$ via $K$-fold cross-validation, choosing the value $z^*$ that

maximizes the cross-validated out-of-sample $R^2$. The final KNS estimator is then given by

$$\hat{b}_{\text{KNS}} \;=\; \hat{b}_{z^*} \;=\; (\bar{\Sigma} + z^* I)^{-1}\,\bar{\mu}. \tag{30}$$

### 2.2.2 Constrained Universal Portfolio Shrinkage (CUPSA)

Kelly et al. (2024) replace single-penalty selection with an optimal combination of Ridge portfolios computed over a grid $Z = \{z_i\}_{i=1}^{M}$:

$$\hat{b}_{CUPSA} \;=\; \sum_{i=1}^{M} w_i\,\hat{b}(f_{z_i})\,, \qquad \hat{b}(f_{z_i}) = (\bar{\Sigma} + z_i I)^{-1}\bar{\mu}. \tag{31}$$

The weights $W = [w_1, \ldots, w_M]'$ solve a constrained mean–variance problem using leave-one-out (LOO) out-of-sample moment estimates of the Ridge components, $\bar{\mu}(Z)$ and $\bar{\Sigma}(Z)$:

$$W_{\text{CUPSA}} \;=\; \underset{W \in S_+^M}{\arg\max} \left\{ W'\bar{\mu}(Z) \;-\; \tfrac{1}{2}\, W'\bar{\Sigma}(Z)W \right\}, \tag{32}$$

with

$$S_+^M \;=\; \left\{ W \in \mathbb{R}_+^M : \sum_{i=1}^{M} \tfrac{w_i}{c_{z_i}} = 1 \right\}, \qquad c_{z_i} \;=\; \frac{\frac{1}{N}\text{tr}(\bar{\Sigma}) + z_i}{\frac{1}{N}\text{tr}(\bar{\Sigma})}.$$

The scaling constants $c_{z_i}$ approximately equalize the volatility of the Ridge basis across $z_i$. The construction of $\bar{\mu}(Z)$ and $\bar{\Sigma}(Z)$ is detailed in Appendix A.1.

## 2.3 Linear and Nonlinear Pricing Components

The arbitrary kernel function $\mathcal{K}(\cdot,\cdot)$ allows us to derive an effective mean-variance return by constructing principal components using nonlinear interactions of firm-level characteristics. We denote by $R^*_{l,t}$ the mean variance efficient portfolio implied by a linear kernel, and by $R^*_{nl,t}$ the mean-variance efficient return implied by an arbitrary nonlinear kernel. In order to disentangle the contribution of the linear and nonlinear (or complex) components of the sdf, we use the following orthogonal decomposition

$$R^*_{nl,t} \;=\; \text{Proj}\left(R^*_{nl,t} \mid R^*_{l,t}\right) + R^{*,\perp}_{l,t}, \tag{33}$$

$$\text{Proj}\left(R^*_{nl,t} \mid R^*_{l,t}\right) \;=\; \underbrace{\mathbb{E}^{-1}\left[R^{*2}_{l,t}\right]\mathbb{E}\left[R^*_{l,t}R^*_{nl,t}\right]}_{=:\beta_l} R^*_{l,t} = \beta_l R^*_{l,t}, \tag{34}$$

where $R_{l,t}^{*,\perp}$ is its orthogonal complement of the linear component — i.e., the part of $R_{nl,t}^*$ unexplained by $R_{l,t}^*$. Moreover, under the assumption that $R_{nl,t}^*$ is on the efficient frontier, it follows that

$$M_t = a + bR_{nl,t}^* = a + bR_{l,t}^{*,\perp} + b\beta_l R_{l,t}^* \tag{35}$$

$$= a + bR_{l,t}^{*,\perp} + cR_{l,t}^* \tag{36}$$

with $c = b\beta_l$, and

$$\mathrm{Var}_t\left(M_{t+1}^*\right) = c^2\mathrm{Var}_t\left(R_{l,t+1}^*\right) + b^2\mathrm{Var}_t\left(R_{l,t+1}^{*,\perp}\right) \tag{37}$$

where the last term on the rhs introduces an additional source of variability induced by the nonlinear component of the sdf.

The relation between the additional variability of the sdf and the maximum Sharpe ratio attainable in the economy is given as

$$\frac{\mathbb{E}_t\left[R_{i,t+1} - R_{f,t}\right]}{\sigma_t\left(R_{i,t+1}\right)} \leq \frac{\sigma_t\left(cR_{l,t+1}^* + bR_{l,t+1}^{*,\perp}\right)}{\mathbb{E}_t\left[M_{t+1}\right]}. \tag{38}$$

or, using Jensen and the second order approximation of $\log M_t^*$, the associated entropy bound is given by

$$\mathbf{L}(M_{t+1}^*) := \log\mathbb{E}_t[M_{t+1}^*] - \mathbb{E}_t[\log M_{t+1}^*] \geq \frac{1}{2}\left(\frac{\mathrm{Var}_t(M_{t+1}^*)}{\mathbb{E}_t^2[M_{t+1}^*]}\right) \tag{39}$$

$$= \frac{1}{2}\left(\frac{c^2\mathrm{Var}_t\left(R_{l,t+1}^*\right) + b^2\mathrm{Var}_t\left(R_{l,t+1}^{*,\perp}\right)}{\mathbb{E}_t^2[M_{t+1}^*]}\right) \geq \frac{1}{2}\frac{\mathbb{E}_t^2\left[R_{i,t+1} - R_{f,t}\right]}{\mathrm{Var}_t\left(R_{i,t+1}\right)}, \tag{40}$$

where the term on the right-hand side is one-half times the square of the conditional Sharpe ratio.

### 2.3.1 Types of Kernels and Nonlinearities

The choice of kernel determines the flexibility and contribution of the nonlinear pricing component $R_{nl,t}^*$, and thus also that of $R_{l,t}^{*,\perp}$.

**Linear kernel.** The linear kernel isolates the contribution of the purely linear pricing component $R_{l,t}^*$ and is defined by

$$k(z_i, z_j) = \langle z_i, z_j\rangle, \tag{41}$$

where $\langle\cdot,\cdot\rangle$ is the Euclidean inner product. The implicit mapping $\phi(\cdot)$ associated with the linear kernel equals the identity mapping, which leaves the original characteristics unchanged. Unlike more complex kernels, the linear kernel does not capture complex patterns in data. This choice is hyperparameter-free and constitutes

the simplest baseline: using it in a kernel PCA recovers ordinary principal components.

**Polynomial kernel.** The polynomial kernel of degree $d$ is given by:

$$k(z_i, z_j) = (c + \langle z_i, z_j \rangle)^d, \tag{42}$$

where the hyperparemeter $c$ governs the relative influence of higher-order versus lower-order terms (see, e.g., Schölkopf and Smola, 2018; Kozak, 2020). In our empirical analysis, we consider a second-order polynomial kernel, i.e., $d = 2$. The implicit mapping $\phi(\cdot)$ is best illustrated by considering a simple example in which for each firm we observe a bi-variate vector of characteristics. Let $z_i$ and $z_j$ contain two characteristics for each of the two stocks. Then, the second-order polynomial kernel applied on the two vectors is given by:

$$\begin{aligned} (c + \langle z_i, z_j \rangle)^2 &= (c + z_i' z_j)^2 = (c + z_{i1} z_{j1} + z_{i2} z_{j2})^2 \\ &= c^2 + 2c z_{i1} z_{j1} + 2c z_{i2} z_{j2} + 2 z_{i1} z_{i2} z_{j1} z_{j2} + z_{i1}^2 z_{j1}^2 + z_{i2}^2 z_{j2}^2 = \langle \phi(z_i), \phi(z_j) \rangle, \end{aligned} \tag{43}$$

thus, the mapping function is given by $\phi(z) \to \{c, \sqrt{2c} z_1, \sqrt{2c} z_2, z_1^2, z_2^2, \sqrt{2} z_1 z_2\}$. Observe now that $c$ determines the relative influence of second-order terms $(z_{i1}^2, z_{i1} z_{i2}, z_{i2} z_{i1}, z_{i2}^2)$ relative to the original characteristics $(z_{i1}, z_{i2})$. For $c = 0$, the feature space consists exclusively of second-order interaction terms, with the original (linear) features entirely omitted. Conversely, as $c \to \infty$, the contribution of the original features increasingly dominates the kernel evaluation, while the second-order interactions become negligible. Consequently, in the limit of large $c$, the second-order polynomial kernel asymptotically reduces to the linear kernel.

**Gaussian kernel** The Gaussian kernel

$$\kappa(x, x') = \exp\big( - \|x - x'\|^2 / (2\sigma^2) \big) = \exp\big( - c \|x - x'\|^2 \big) \tag{44}$$

is a popular choice because it captures highly flexible patterns in the data. It is also stable and reliable, as small changes in the input lead to smooth, gradual changes in the kernel values, making it a robust choice in financial applications.

**Theorem** (Universal Gaussian kernel)**.** *Let $d \in \mathbb{N}$ and $\gamma > 0$. For any compact set $K \subset \mathbb{R}^d$, the RKHS $\mathcal{H}_k$ implied by the Gaussian kernel is dense in $C(K)$ w.r.t. the sup norm. Therefore, for every $f \in C(K)$ and*

*every* $\varepsilon > 0$*, there exists* $g \in \mathcal{H}_k$ *such that*

$$\sup_{x \in K} \|f(x) - g(x)\| < \varepsilon.$$

*Proof.* See Steinwart (2001) □

The above theorem shows that the Gaussian kernel induces a universal RKHS, meaning that the associated Hilbert space is dense in the space of continuous functions on the domain. Consequently, the corresponding kernel principal components (KPCs) form a basis that is sufficiently rich to approximate any smooth variation in the data, including highly complex functional relationships. Practically, RBF provides greater functional richness (often increasing $\text{Var}_t(R^{*,\perp}_{l,t+1})$ and the upper bound for conditional Sharpe improvements but requires careful bandwidth $\sigma$ selection and regularization to avoid overfitting.

## 2.4 Factor structure

Our goal is to explain the differences in average returns for a large cross-section of anomaly portfolios. Let $R_{t+1,i}$ denote the return of asset $i$ at time $t+1$. For any return in excess of the risk-free rate $R^e_{t+1,i} = R_{t+1,i} - R^f_{t+1}$, it holds the following beta pricing model

$$\mathbb{E}_t\left[M_{t+1}R^e_{t+1,i}\right] = 0 \quad \Leftrightarrow \quad \mathbb{E}_t\left[R^e_{t+1,i}\right] = \underbrace{\left(-\frac{\text{Cov}_t(R^e_{t+1,i}, M_{t+1})}{\text{Var}_t(M_{t+1})}\right)}_{\beta_{i,t}} \cdot \underbrace{\frac{\text{Var}_t(M_{t+1})}{\mathbb{E}_t[M_{t+1}]}}_{\lambda_t},$$

where $\beta_{i,t}$ is the exposure to systematic risk and $\lambda_t$ is the market price of risk. Exploiting the two factor characterization of the SDF described in equation (35), it follows that for any excess return

$$\mathbb{E}_t[R^e_{t+1,i}] = -c\frac{\text{Cov}_t(R^*_{l,t+1}, R^e_{t+1,i})}{\mathbb{E}_t[M_{t+1}]} - b\frac{\text{Cov}_t(R^{*,\perp}_{l,t+1}, R^e_{t+1,i})}{\mathbb{E}_t[M_{t+1}]} \tag{45}$$

$$= \underbrace{\left(-c\frac{\text{Cov}_t(R^e_{t+1,i}, R^*_{l,t+1})}{\text{Var}_t(R^*_{l,t+1})}\right)}_{\beta^l_{i,t}} \cdot \underbrace{\frac{\text{Var}_t(R^*_{l,t+1})}{\mathbb{E}_t[M_{t+1}]}}_{\lambda^l_t} - \underbrace{b\left(\frac{\text{Cov}_t(R^e_{t+1,i}, R^{*,\perp}_{l,t+1})}{\text{Var}_t(R^{*,\perp}_{l,t+1})}\right)}_{\beta^{nl}_{i,t}} \cdot \underbrace{\frac{\text{Var}_t(R^{*,\perp}_{l,t+1})}{\mathbb{E}_t[M_{t+1}]}}_{\lambda^{nl}_t}, \tag{46}$$

where $\lambda_t^{nl}$ and $\lambda_t^{nl}$ are the linear and nonlinear component of the market price of risk, respectively. Moreover, since $R_{l,t}^*$ and $R_{l,t}^{*,\perp}$ are uncorrelated excess returns, it follow that

$$\mathbb{E}_t[R_{l,t+1}^*] = -c\frac{\text{Var}_t(R_{l,t+1}^*)}{\mathbb{E}_t[M_{t+1}]}, \tag{47}$$

$$\mathbb{E}_t[R_{l,t+1}^{*,\perp}] = -b\frac{\text{Var}_t(R_{l,t+1}^{*,\perp})}{\mathbb{E}_t[M_{t+1}]}, \tag{48}$$

so that

$$\mathbb{E}_t[R_{t+1,i}^e] = \beta_{i,t}^l \cdot \mathbb{E}_t[R_{l,t+1}^*] + \beta_{i,t}^{nl} \cdot \mathbb{E}_t[R_{l,t+1}^{*,\perp}]. \tag{49}$$

Ultimately, our formulation of the SDF implies the following factor model

$$R_{t+1,i}^e = \alpha(z_t) + \beta_i^l(z_t) \cdot R_{l,t+1}^* + \beta_i^{nl}(z_t) \cdot R_{l,t+1}^{*,\perp} + \varepsilon_{i,t+1}, \tag{50}$$

and we instrument the time-variation of the intercept $\alpha(z_t)$ and slopes coefficients $\beta_i^l(z_t)$, $\beta_i^l(z_t)$ using a vector of macro-variables $z \in \mathbb{R}^m$, i.e.,

$$\begin{bmatrix} \alpha(z_t) \\ \beta_i^l(z_t) \\ \beta_i^{nl}(z_t) \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \beta_0^l \\ \beta_0^{nl} \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \beta_1(I_2 \otimes I_m) \end{bmatrix} z_t \tag{51}$$

where $\otimes$ denotes the Kronecker product and $I_2$ is an identity matrix. Excess stock return can then be rewritten as

$$R_{i,t+1}^e = \alpha_0 + \alpha_1 z_t + \beta_0^l R_{l,t+1}^* + \beta_0^{nl} R_{l,t+1}^{*,\perp} + \beta_1(I_2 \otimes z_t). \begin{bmatrix} R_{l,t+1}^* \\ R_{nl,t+1}^* \end{bmatrix} + \varepsilon_{i,t+1}. \tag{52}$$

With $\alpha_1$ and $\beta_1$ equal to zero, the model in (52) reduces to a specification with time-invariant parameters.

# 3 Empirical Analysis

## 3.1 Data

We use the Open Source Asset Pricing dataset by Chen and Zimmermann (2022), which comprises 212 monthly US stock-level characteristics grouped into different categories including accounting data, analyst forecasts, or stock prices. The data release we employ in this study stems from August 2023 and spans a sample period from January 1925 to December 2022. We lag the monthly stock characteristics by one month and match them with monthly stock returns, volume and price data, sourced from the Center for Research in Security Prices (CRSP).

We impose three filters on the characteristics. First, we truncate the sample to begin in January 1974 and exclude any characteristic with an inception date thereafter, ensuring sufficient data availability for accounting variables (Simon et al., 2022; Freyberger et al., 2025). Second, following Freyberger et al. (2025), we retain only one characteristic from each cluster of highly similar variables, while preserving key asset pricing predictors such as size and book-to-market.[5] Third, we exclude binary and categorical characteristics, as well as those with more than 30% missing values. These filters yield a final set of 82 characteristics.

Moreover, we restrict our analysis to securities with a CRSP share code of 10 and 11 (common stocks). To mitigate the influence of micro caps, we also filter out stocks with a market capitalization below 0.01% of aggregate stock market capitalization at each point in time, following Kozak (2020).

Finally, we rank-transform and normalize each characteristic cross-sectionally following the recent literature (Gu et al., 2020; DeMiguel et al., 2020; Kozak et al., 2020; Simon et al., 2022). For each month, stocks are ranked by characteristic values, mapped to the unit interval, and then normalized by subtracting the cross-sectional mean and dividing by the average absolute deviation. Missing values are imputed with the unconditional mean, which is zero by construction. The resulting normalized characteristics are used as portfolio weights in constructing zero-cost long-short factor portfolios. Lastly, we scale those weights so that the long and short legs of each factor portfolio have equal dollar exposure. As Kozak (2020) highlights, the described transformation achieves three things. First, it allows us to exclusively focus on the cross-sectional aspect of return predictability, second, it removes the influence of outliers, and, third, it keeps leverage across all portfolios constant.

[5] Examples of highly similar characteristics include the share of zero-trading days, computed over different lookback periods.

## 3.2 Model Estimation

Assume access to data covering stock characteristics $Z_t$ and excess returns $R_{t+1}$, with the time index ranging from $t \in \{1, \ldots, T, T+1, \ldots, T+T_{\text{OOS}}\}$. Split the data at index $T$ into an in-sample and an out-of-sample period.

**Step 1. K-PCA**

1. Select a kernel $\kappa$ and a kernel parameter $c$.[6]

2. Construct the $T \times T$ matrix $\Omega$ in (16) using the selected kernel and the in-sample data, and compute the scaled eigenvectors of $\Omega$.

   This procedure recovers the $T$ dominant principal components $P \in \mathbb{R}^{T \times T}$ from a PCA in the high-dimensional space of factor returns $F_{t+1} = \phi(Z_t)' R_{t+1}$, where $\phi(\cdot)$ is the feature map induced by the kernel.

3. (Optional) Compute the out-of-sample principal components $P_{\text{OOS}} \in \mathbb{R}^{T_{\text{OOS}} \times T}$ using Equation (26).

**Step 2. MVE Portfolio Construction**

1. Use the in-sample principal components $P$ as input to an MVE portfolio estimation algorithm.

   - **KNS.** Compute the SDF loadings $\hat{b}_{KNS}$ according to Equation (30), with the Ridge penalty $z^*$ selected via cross-validation.
   - **CUPSA.** Compute the SDF loadings $\hat{b}_{CUPSA}$ according to Equation (31).

Finally, the fitted MVE portfolio estimators can be used to construct a time series of MVE portfolio returns given as

$$\hat{R}^*_{\kappa,t} = \hat{b}'_{\kappa,i} P_{\kappa,t}, \quad t \in \{1, \ldots, T, T+1, \ldots, T+T_{\text{OOS}}\}, \tag{53}$$

where $i \in \{\text{KNS}, \text{CUPSA}\}$.

## 3.3 Model Selection and In-Sample Estimation

We begin by evaluating, in an in-sample setting, how allowing for nonlinearities in the SDF affects pricing performance. Using the full sample from January 1974 through December 2022, we estimate—separately for each kernel specification—the kernel PCA and, from the resulting components, the SDF via the KNS

[6] In the case of the linear kernel, no kernel hyperparameter is selected.

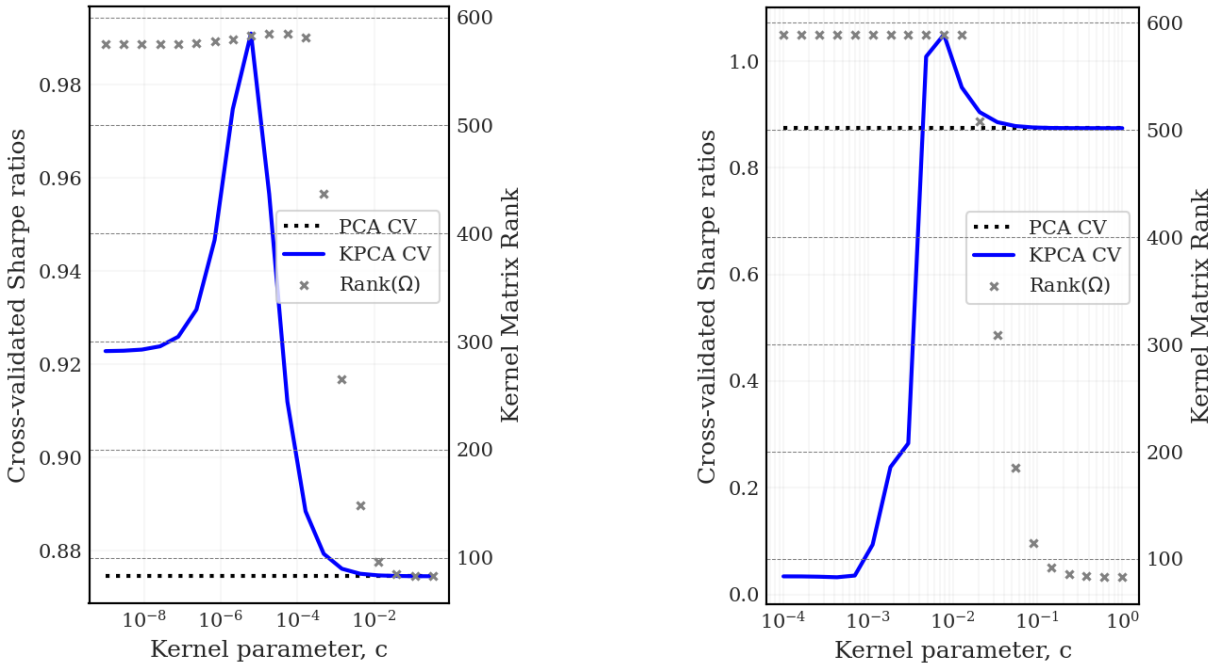
estimator across a wide range of model specifications. We select free hyperparameters (the kernel-specific parameter $c$ and the Ridge penalty $z$) via cross-validation (CV). The final estimates for the K-PCs and KNS portfolio weights are then applied period by period to construct the time series of MVE portfolio returns for each kernel. We deliberately do not use CUPSA in this exercise, as the algorithm is designed primarily for out-of-sample evaluation. Our implementation follows Kozak (2020).

The kernel hyperparameter $c$, which we have to determine for both nonlinear kernels (polynomial and Gaussian), governs the effective degree of nonlinearity considered in the K-PCA. As discussed above in Section 2.3.1, $c$ controls the balance between base (linear) features and higher–order transformations in the kernel PCs: as $c \to \infty$, the feature map approaches the identity and K-PCA collapses to standard PCA on the original characteristics; as $c$ becomes small, the representation is dominated by second-order terms (for the polynomial) or by complex, higher-order transformations (for Gaussian). Intermediate values of $c$ yield a mixture in which linear and nonlinear components jointly shape the K-PCs. This interpretation motivates selecting $c$ empirically rather than fixing it a priori.

The regularization parameter $z$ in the KNS step controls the amount of shrinkage applied to the K-PC portfolio weights. Increasing $z$ dampens exposures to on low-variance K-PCs — trading in-sample fit for lower variance and effective model complexity. In the extreme cases, $z = 0$ recovers OLS (no shrinkage), while $z \to \infty$ drives the weights toward zero.

Empirically, we gauge the optimal degree of nonlinearities and regularization in a data-driven way that optimizes the out-of-sample fit, via CV. We partition the data into four contiguous folds; for each fold we estimate the model on the other three folds and evaluate on the held-out fold, iterating over all four splits. From each held-out block, we compute the out-of-sample $R^2$ implied by the KNS objective and the Sharpe ratio of the MVE return. Averaging these across the four folds yields the cross-validated $R^2_{\text{OOS}}$ and the cross-validated Sharpe ratio. Hyperparameters are chosen in two steps. First, for each candidate $c$ on a predefined grid $C = \{c_\ell\}_{\ell=1}^{L}$, we select the optimal shrinkage $z^*(c)$ as the value that maximizes $R^2_{\text{OOS}}$ under our four-fold CV procedure. Second, holding $z = z^*(c)$, we select the kernel parameter $c^*$ as the value that maximizes the cross-validated Sharpe ratio of the implied MVE return.

Figure 2 plots the cross-validated Sharpe ratio as a function of the kernel parameter $c$. The solid blue line shows the KNS-implied Sharpe ratio for the two nonlinear kernels for each value of $c$. The dotted black line is the linear benchmark, which does not depend on $c$ and is therefore flat. In both cases, the objective attains its maximum at interior values of $c$, indicating that the best in-sample specification is achieved when the K-PCs blend linear and nonlinear structure rather than collapsing to either extreme. For the quadratic




(a) Polynomial kernel ($2^{nd}$ order)

(b) Gaussian kernel

Figure 2: Cross-validated Sharpe ratios for nonlinear kernels. The plot depicts the cross-validated Sharpe ratios for the polynomial and Gaussian kernel and varying kernel hyperparameters. Kernel principal components are estimated using the full sample from January 1974 to December 2022. We use the KNS estimator and a four-fold cross-validation procedure to select the regularization parameter $z$.

kernel (left), allowing second-order terms raises the cross-validated Sharpe from 0.87 (linear benchmark) to 0.99, an increase of about 14%. For the Gaussian kernel, the improvement is even larger with about 21%, going from a level of 0.87 to 1.05.

Following Kozak (2020), we also report the empirical rank of the Gram matrix $\Omega$, depicted as grey crosses. Effectively, the rank of $\Omega$ summarizes the in-sample dimensionality of the features space—that is, the number of distinct characteristics-managed portfolios considered in the PCA. When the kernel parameter $c$ is large, both the polynomial and Gaussian kernel approach the linear case. Therefore, the feature map collapses to the 82 base characteristics, and the rank correspondingly tends toward 82. As $c$ decreases, higher-order terms receive more weight and nonlinear combinations of characteristics enter, prompting an increase in the rank of $\Omega$. With a second-order polynomial kernel, the implied number of features rises to $\binom{82}{2} + 2 \times 82 = 3{,}485$, covering all pairwise interactions, quadratic terms, and the original characteristics. By contrast, the Gaussian kernel induces an infinite-dimensional feature space. In all cases, however, the empirical rank of $\Omega$ is bounded above by the number of observations in the training sample, which in our application is $T = 589$.

Equipped with the cross-validated kernel and regularization hyperparameter, we proceed to recover the full-sample time series of MVE portfolio returns. We denote by $\hat{R}^*_{l,t}$ the MVE return implied by the linear kernel and by $\hat{R}^*_{nl,t} := \hat{R}^*_{\kappa,t}$ the MVE return implied by a nonlinear kernel ($\kappa \in \{poly2, gaussian\}$):

$$\hat{R}^*_{l,t} \;=\; \hat{b}'_{l,KNS} P_{l,t}, \qquad \hat{R}^*_{nl,t} \;=\; \hat{b}_{\kappa,KNS} * P_{\kappa,t}, \quad t = 1, \ldots, T. \tag{54}$$

The rolling annualized returns of these portfolios are displayed in Figure 3. To ensure comparability across models, all return series are volatility-scaled to match the realized volatility of the CRSP value-weighted market index. Visually, the return paths are similar over time, with the polynomial kernel (upper graph) showing its largest departures from the linear benchmark in the pre-2000 period. In contrast, the Gaussian kernel (lower graph) tends to post higher annualized returns in most periods, implying higher realized Sharpe ratios than the linear and polynomial specifications.

**Pricing Performance.** We next assess the incremental efficiency of the SDF when moving from the linear to a nonlinear specification. To isolate the contribution of nonlinearities to pricing performance, we orthogonalize the nonlinear MVE return $\hat{R}^*_{nl,t}$ with respect to the linear benchmark $\hat{R}^*_{l,t}$ via

$$\hat{R}^*_{nl,t} = \alpha + \beta_l \, \hat{R}^*_{l,t} + \varepsilon_t. \tag{55}$$

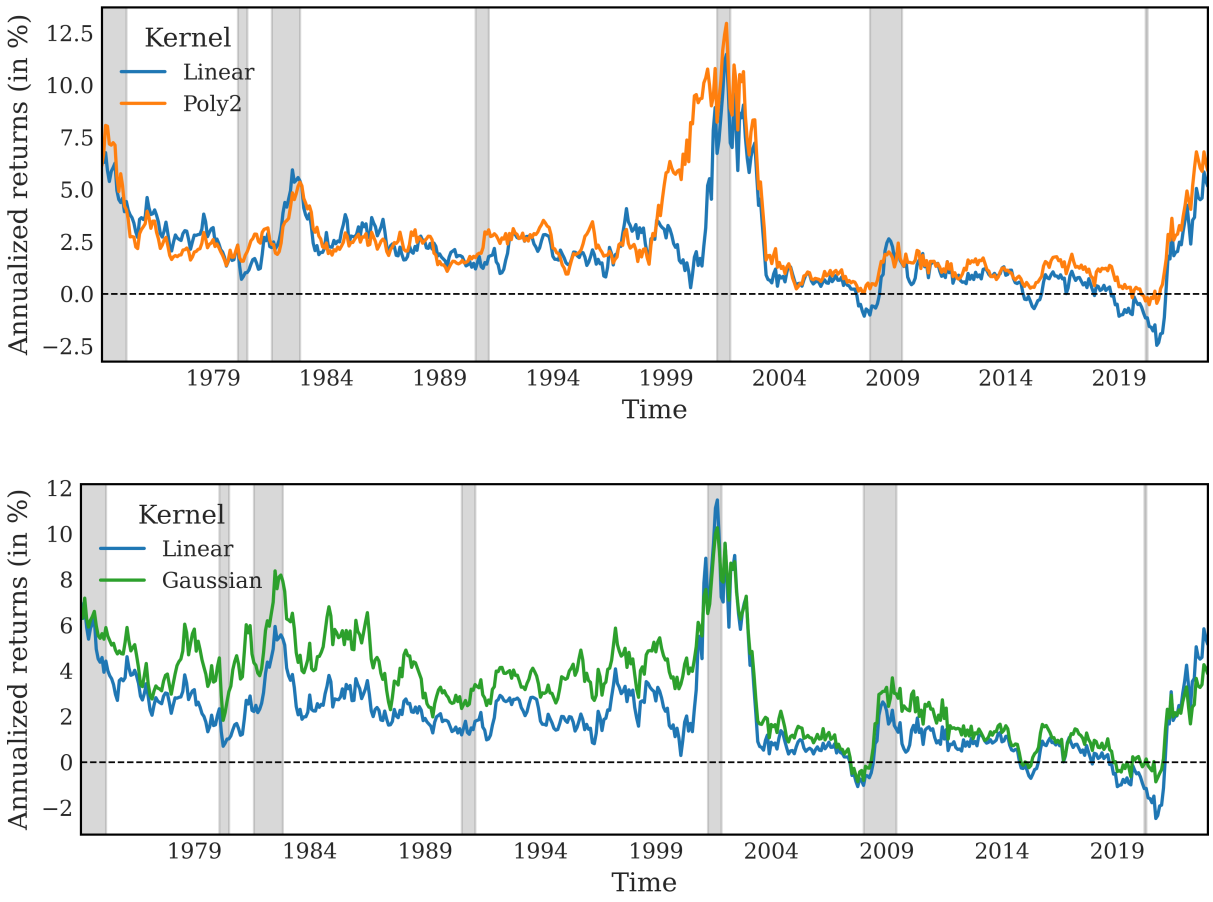



Figure 3: Annualized MVE Portfolio Returns by Kernel. The figure plots the annualized returns of MVE portfolios constructed from kernel principal components. Panel A (upper) reports results for the linear and polynomial kernels; Panel B (lower) reports results for the linear and Gaussian kernels. Kernel PCs are estimated using the full sample from January 1974 to December 2022. The KNS algorithm is applied to construct the optimal portfolio from the resulting factors. The kernel hyperparameter $c$ is selected via 4-fold cross-validation to maximize the Sharpe ratio. Finally, all return series are scaled to match the volatility of the aggregate market.

The results of regression (55) are presented in Table 1. While the nonlinear MVE returns load significantly on the linear return, both nonlinear SDF models deliver a large, positive, and statistically significant $\alpha$. Based on the estimated coefficients, we compute the orthogonal nonlinear component as

$$\hat{R}^{*,\perp}_{l,t} = \hat{R}^{*}_{nl,t} - \hat{\beta}_l \, \hat{R}^{*}_{l,t} = \hat{R}^{*}_{nl,t} - \text{Proj}\Big(\hat{R}^{*}_{nl,t} \mid \hat{R}^{*}_{l,t}\Big) \,. \tag{56}$$

Table 1: Orthogonolization of MVE portfolio returns for nonlinear kernels. The table depicts the $\alpha$ (annualized, in %) along with the $\beta_{linear}$ coefficient from time-series regression in (55). Standard errors are in parentheses. The significance levels of both coefficients are indicated by stars: * p<0.1, ** p<0.05, *** p<0.01.

| | **Poly2** | **Gaussian** |
|---|---|---|
| $\hat{\alpha}$ (annualized, in %) | 11.36*** | 16.22*** |
| | (1.94) | (1.59) |
| $\hat{\beta}_{\text{linear}}$ | 0.81*** | 0.90*** |
| | (0.05) | (0.06) |
| $R^2$ (in %) | 65.39 | 80.50 |

We next conduct an asset–pricing test using the returns of 82 portfolios constructed from the long–short characteristic signals that serve as inputs to K–PCA. Let $F_t$ denote the vector of time-$t$ test-asset returns with $F_{i,t}$ its $i$-th element. We estimate three time-series regressions. First, we project each test asset's return on the linear MVE return:

$$F_{i,t} = \alpha_i + \beta_{l,i} \, \hat{R}^{*}_{l,t} + \epsilon_{i,t}. \tag{57}$$

Second, we augment the specification by adding the orthogonal nonlinear component:

$$F_{i,t} = \alpha_i + \beta_{l,i} \, \hat{R}^{*}_{l,t} + \beta_{nl,i} \, \hat{R}^{*,\perp}_{t} + \epsilon_{i,t}. \tag{58}$$

Regression (58) is then run for the polynomial and the Gaussian kernel.

For all three regressions setups, we collect the absolute values of the estimated coefficients for $\alpha$, $\beta_l$, and $\beta_{nl}$, along with the corresponding $R^2$ values. We then aggregate the 82 test assets into 24 groups based on the *economic category* classification proposed by Chen and Zimmermann (2022), and compute the average coefficient estimates within each group. Figure 4a reports the average absolute values of $\alpha$ across the different groups. The blue bar indicates the baseline regression with only the linear return, while the orange and green bars incorporate the orthogonal nonlinear components implied by the polynomial and Gaussian kernel, respectively. As one can see, the addition of the nonlinear components substantially reduces the pricing errors. Specifically, the inclusion of the polynomial kernel component reduces average $\alpha$ values in

20 out of 24 categories, whereas the Gaussian kernel achieves reductions in 22 out of 24. Moreover, the average reduction in alphas across all test assets amounts to roughly 37% for the polynomial and 50% for the Gaussian kernel. A consistent pattern emerges for the explained variation, measured by $R^2$, as shown in Figure 4b, with increases in average $R^2$ of 26% (22%) for the Gaussian (polynomial) kernel.

In addition to Sharpe ratios and $\alpha$'s from the time–series regressions, we assess pricing efficiency using the Hansen and Jagannathan (1991) (HJ) distance. For each kernel $\kappa \in \{linear, poly2, gaussian\}$, we form the SDF from the MVE return implied by the K–PCA/KNS fit:

$$M_{t+1}(\kappa) \;=\; 1 - h_\kappa \, \hat{R}^*_{\kappa,t+1}, \qquad t = 1, \ldots, T,$$

where the scale factor

$$h_\kappa \;=\; \frac{\bar{E}\left[\hat{R}^*_{\kappa,t+1}\right]}{\bar{E}\left[\left(\hat{R}^*_{\kappa,t+1}\right)^2\right]}, \qquad \bar{E}[X] \;=\; \frac{1}{T}\sum_{t=1}^{T} X_t,$$

follows the standard normalization in Cochrane (2009), enforcing the conventional self-pricing normalization, $\bar{E}[M_{t+1}\hat{R}^*_{\kappa,t+1}] = 0$.

Let $F_{t+1}$ again denote the return vector of our 82 test assets. The pricing error for asset $i$ under kernel $\kappa$ is then given as

$$PE_i(\kappa) \;=\; \bar{E}[F_{i,t+1}\, M_{t+1}(\kappa)]\,.$$

Collecting the pricing errors in the vector

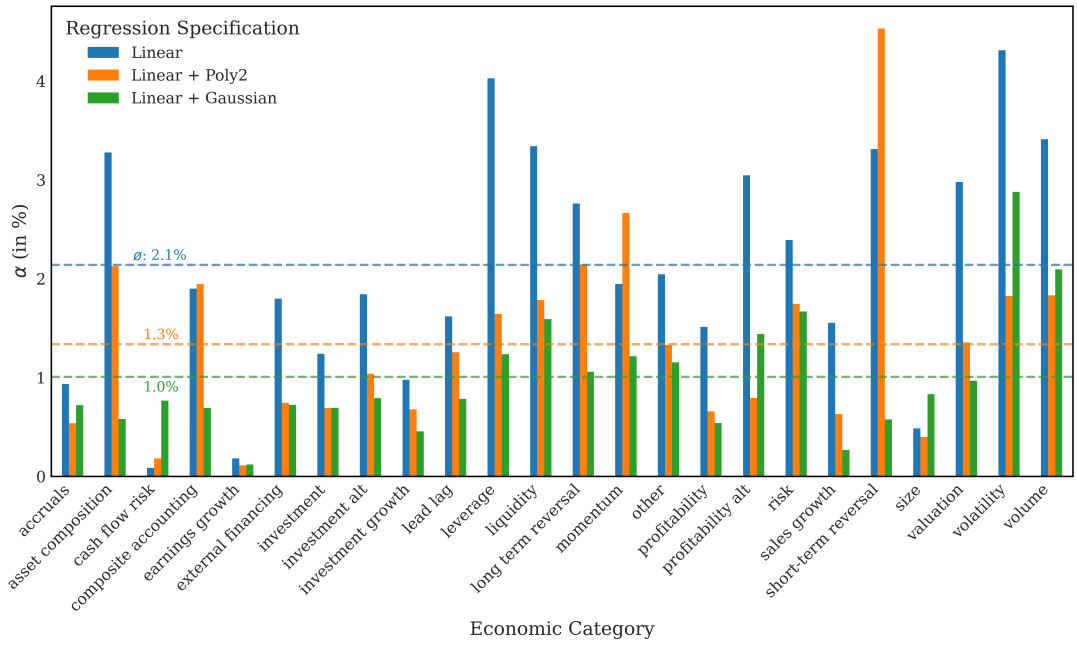
$$PE(\kappa) \;=\; \big(PE_i(\kappa)\big)_{i=1}^{82},$$

the HJ distance is obtained as

$$D^{HJ}(\kappa) \;=\; PE(\kappa)'\, A^{-1}\, PE(\kappa), \text{ with} \qquad A \;=\; \bar{E}[FF']\,.$$

Figure 1 above reports $D^{HJ}(\kappa)$ aggregated across all test assets within each economic category group, as defined by Chen and Zimmermann (2022). Relative to the linear kernel, both nonlinear specifications deliver materially lower in-sample HJ distance, mirroring the improvements seen in cross-validated Sharpe ratios and in the time-series regressions. As one can see, these gains are broad-based rather than confined to a handful of categories, with the Gaussian kernel typically yielding the largest declines.

(a) Absolute pricing error $\alpha$




(b) In-sample $R^2$

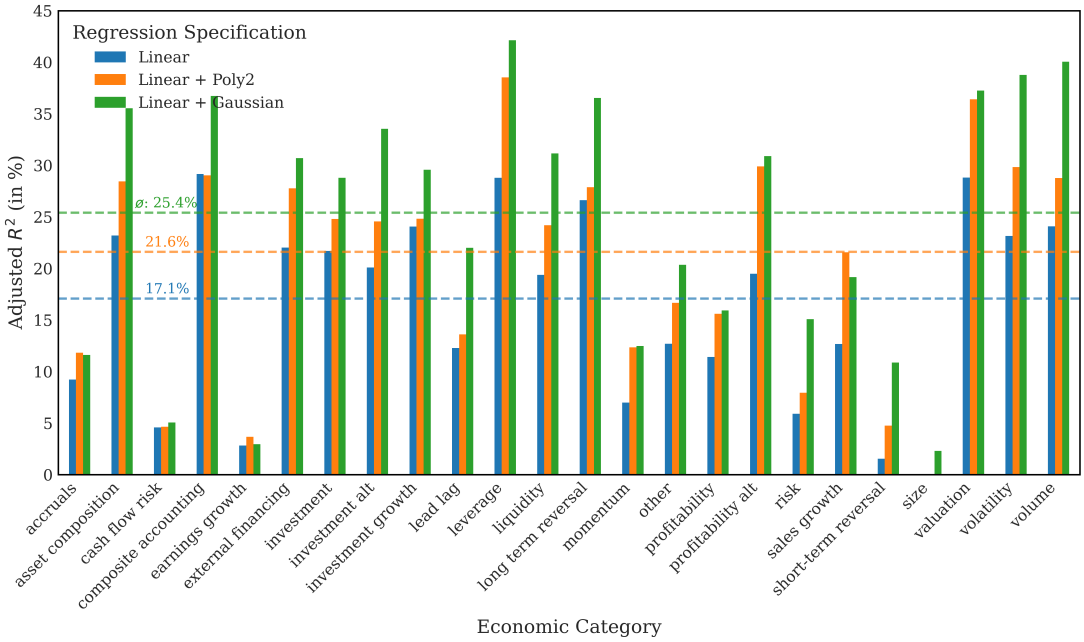



Figure 4: In-sample $R^2$ and absolute $\alpha$ coefficients grouped by economic theme. The plot displays the remaining alpha along the $R^2$ from regression the test assets' returns on the linear (plus nonlinear) part in Equation 57 (Equation 58). The absolute alphas and $R^2$ for the single test assets are averaged per group based on the economic category, as defined in Chen and Zimmermann (2022). Horizontal lines depict averages across all individual test assets. Underlying MVE portfolio returns for the linear, polynomial and Gaussian kernel are estimated in-sample via K-PCA and KNS regularization. Hyperparameters are selected through a four-fold CV procedure. The estimation sample ranges from 1974 to 2022.

### 3.4 Time-Varying Dynamics

In this section, we report results of our conditional factor model in which the factor loadings and pricing errors are allowed to vary with observable macro-financial states. For each test asset $i$, we estimate the macro-instrumented (expanded) time-series regression

$$F_{i,t+1} = \alpha_0 + \alpha_1 z_t + \beta_0^l R_{l,t+1}^* + \beta_0^{nl} R_{l,t+1}^{*,\perp} + \beta_1 \left(I_2 \otimes z_t\right) \begin{bmatrix} R_{l,t+1}^* \\ R_{l,t+1}^{*,\perp} \end{bmatrix} + \varepsilon_{i,t+1}, \tag{59}$$

where $R_l^*$ is the MVE return from the linear kernel and $R_l^{*,\perp}$ is the orthogonal nonlinear return obtained by projecting the nonlinear MVE return on $R_l^*$ and retaining the residual.[7] The term $\alpha_1 z_t$ delivers a state-dependent intercept, while the interaction block $\beta_1(I_2 \otimes z_t)\left[R_l^*,\, R_l^{*,\perp}\right]^\top$ makes both loadings, i.e., $\beta_i^l(z_t)$ and $\beta_i^{nl}(z_t)$, time-varying as a function of the state vector $z_t$.

In our empirical analysis, the state vector $z_t$ comprises six variables selected to capture macro/business-cycle conditions, financial conditions, and sentiment. Specifically, we include (i) inflation proxied by the CPI (CPIAUCSL), (ii) real activity proxied by industrial production (INDPRO), and (iii) the corporate credit spread constructed as BAA$-$AAA, all from FRED–MD. We then complete the set with (iv) investor sentiment from Baker and Wurgler (2006), (v) the financial uncertainty index of Jurado et al. (2015), and (vi) U.S. economic policy uncertainty from Baker et al. (2016).[8] All instruments are then shifted so that $z_t$ reflects information available at $t$ for forecasting $t{+}1$ returns. Finally, each series is standardized having mean zero and unit variance.

Given the state variables $z_t$ and the two traded returns, we estimate the empirical model in (59) asset by asset via OLS. [9] The regressors are the constant, $z_t$, $R_{l,t+1}^*$, $R_{l,t+1}^{*,\perp}$, and the interaction block $(I_2 \otimes z_t)\,[R_{l,t+1}^*,\, R_{l,t+1}^{*,\perp}]'$ so that both $\beta_i^l(z_t)$ and $\beta_i^{nl}(z_t)$ may vary with the state. From the estimated coefficients, we can recover the conditional objects from Equation (51):

$$\hat{\beta}_i^{nl}(z_t) = \hat{\beta}_{0,i}^{nl} + \sum_j \hat{\gamma}_{i,j}^{nl} z_{j,t}, \qquad \hat{\alpha}_i(z_t) = \hat{\alpha}_{0,i} + \hat{\alpha}_{1,i}' z_t,$$

where $\{\hat{\gamma}_{i,j}^{nl}\}$ are the coefficients on $R_{l,t+1}^{*,\perp} \times z_{j,t}$ in the interaction block. We compare the performance of the

[7] Given the superior pricing performance of the Gaussian kernel, we henceforth focus this nonlinear specification.

[8] FRED–MD series are transformed using the file tcodes and aligned to month end; slow-release indicators (CPI, INDPRO) enter with a two-month publication lag, while market-based spreads are contemporaneous.

[9] In this excercise, we use heteroskedasticity-adjusted standard errors with 6 lags

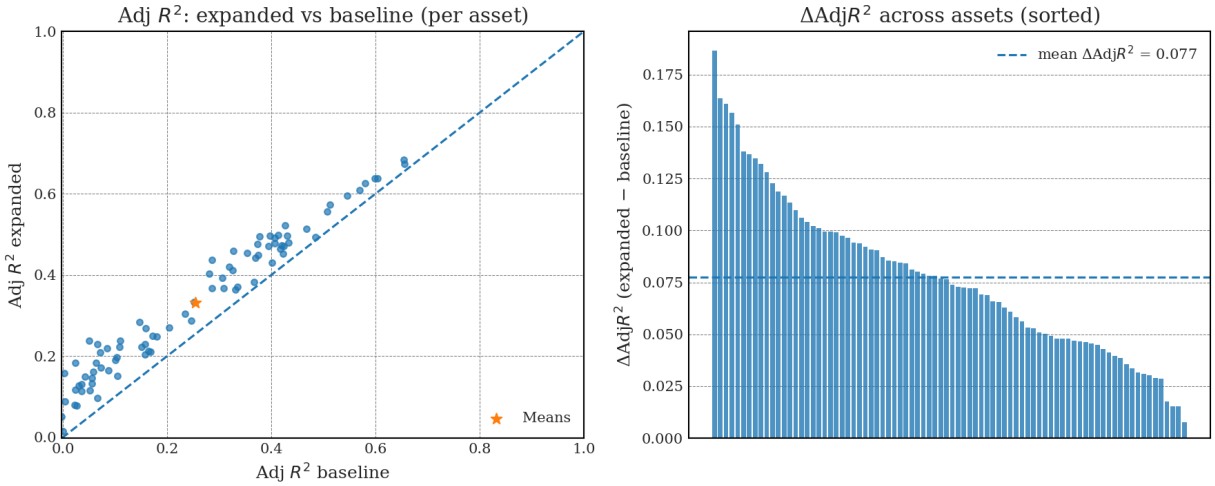
conditional factor model with a baseline, time-invariant specification in which all coefficients are constant over time.

Panel (a) in Figure 5 shows the results for the unrestricted specification, in which the state vector $z_t$ enters directly ($\alpha_1 \neq 0$). Panel (b) imposes the restriction $\alpha_1 = 0$, so that the pricing errors are time-invariant, and $z_t$ influences returns only through its interactions with the linear and nonlinear components of the SDF. The two panels deliver very similar gains in explanatory power (the average $\Delta \text{Adj}R^2$ is nearly identical at about 7.8%), indicating that the direct contribution of $z_t$ is limited in explaining cross-sectional anomaly portfolios. This is a good result: pricing errors (alphas) remain small even when $z_t$ is added directly, and in a correctly specified—or "perfect"—model these pricing errors should be indistinguishable from zero. Hence, $z_t$ primarily provides conditioning information rather than introducing additional priced sources of risk beyond the linear and nonlinear MVE components. Instead, most of the improvement arises from state-dependent exposures to the two MVE components. This is reassuring from a spanning perspective: the linear and nonlinear MVE returns capture the conditional SDF, while $z_t$ serves primarily as conditioning information rather than as an additional priced factor.

Figure 6 displays the change in adjusted $R^2$ for each of the 24 economic categories relative to the time-invariant baseline. Nearly all categories experience a meaningful improvement in explanatory power, with a cross-sectional average increase of about 8%. The largest increases occur for cash-flow risk, size, and risk groups, followed by short-term reversal and volatility; categories such as long-term reversal and composite recommendations display smaller, yet still positive, gains. Because improvements are diffuse rather than concentrated in a few groups, the figure suggests that state dependence—and, by construction, the nonlinear structure embedded in the Gaussian kernel—adds explanatory content broadly across anomalies. Moreover, characteristics tied to cash-flow fundamentals and market frictions appear especially sensitive to macro conditions, which the conditional SDF captures.

Figure 7 shows that the 82 test assets exhibit a smooth, heterogeneous loading profile: when assets are sorted by their signed loading, exposures decay gradually from positive to negative values, with no sharp breaks. This smooth gradient appears for both the linear component $R_t^l$ and the orthogonal nonlinear component $R_t^{\perp,l}$, indicating a broad spectrum of sensitivities across assets. This pattern is confirmed in Figure 8, which reports, by economic category, the average unconditional loadings on the two components (Panel a for $R_t^l$; Panel b for $R_t^{\perp,l}$). The linear component has a positive cross-category mean of about 0.08, with relatively high exposures in categories linked to volatility and trading frictions, and negative exposures in cash-flow risk, size, and risk. On the other hand, the nonlinear component has a negative cross-category

(a) Without restrictions




(b) Imposing $\alpha_1 = 0$

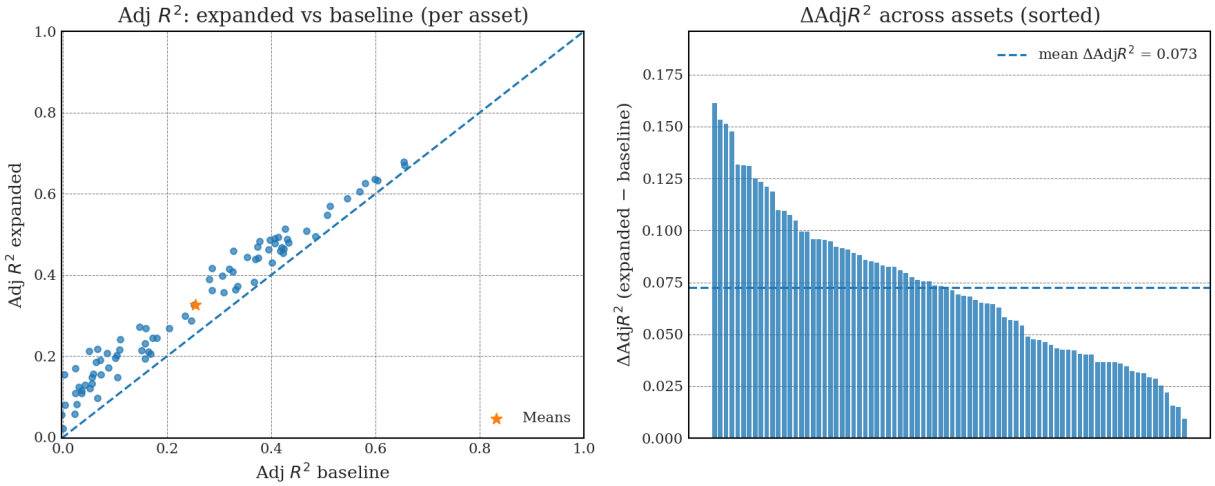



Figure 5: ADJUSTED $R^2$ COMPARISON. Panel A plots per-asset adjusted $R^2$ from the expanded, macro-instrumented model—using the linear and orthogonal nonlinear MVE returns and their interactions with $z_t$—against the baseline (time-invariant) model; the 45° line marks parity and the star denotes cross-asset means. Panel B shows $\Delta\text{Adj}R^2$ (expanded − baseline) by asset, sorted, with the dashed line indicating the cross-asset average. Improvements indicate that allowing state dependence in $\beta_i^l(z_t)$ and $\beta_i^{nl}(z_t)$ increases explanatory power relative to the linear-only specification. Underlying MVE portfolio returns for the linear and Gaussian kernel are estimated in-sample via K-PCA and KNS regularization. Hyperparameters are selected through a four-fold CV procedure. The estimation sample ranges from 1974 to 2022.

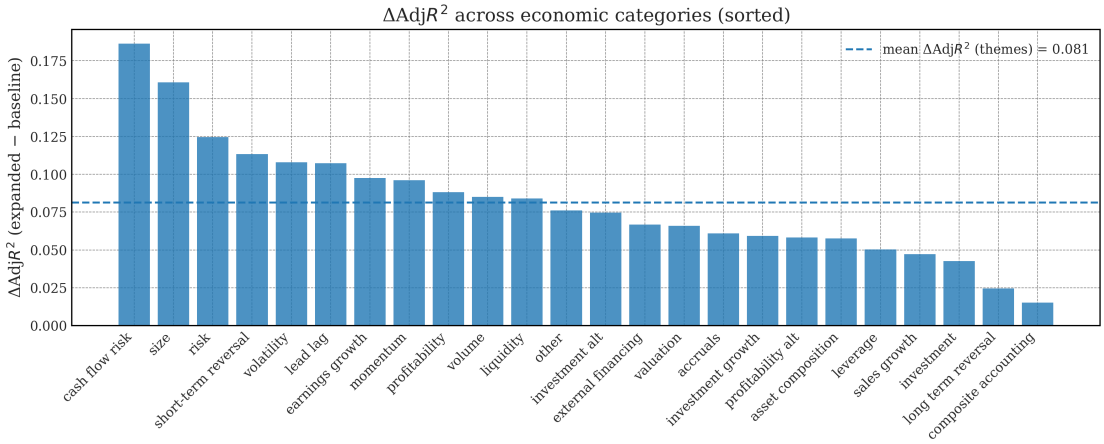


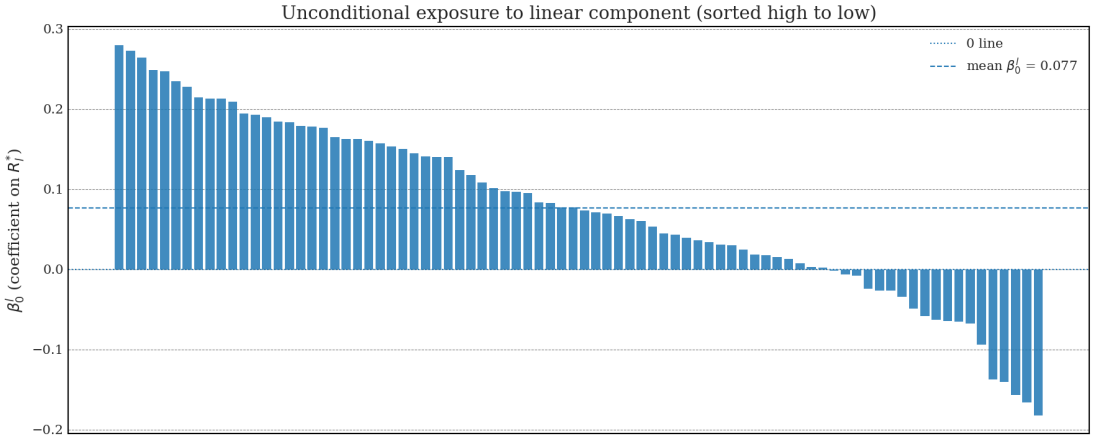

Figure 6: Adjusted $R^2$ improvements by economic category. Bars plot, for each economic category, the change in the category-level mean adjusted $R^2$ (average across assets within the category) from the expanded, macro-instrumented model relative to the baseline (time-invariant) specification. Categories are sorted high to low by $\Delta\text{Adj}R^2$; the dashed line marks the cross-category mean improvement. Underlying MVE portfolio returns for the linear and Gaussian kernel are estimated in-sample via K-PCA with KNS regularization. Hyperparameters are selected through a four-fold CV procedure. The estimation sample ranges from 1974 to 2022.

mean of about $-0.06$, with positive loadings for short-term reversal and risk, and sizable negative loadings for volatility-, profitability-, and size-related themes. Several categories display opposite signs across the two components, indicating that the nonlinear structure adds information beyond the linear span rather than just replicating it. The pattern reveals pronounced heterogeneity across single test assets and economic groups and suggests that both the linear and nonlinear components are needed to capture the cross-sectional structure of expected returns.

## 3.5 Out-of-sample Analysis

To assess robustness, we extend the analysis to a fully out-of-sample setting with iterative re-estimation. This serves two purposes. First, it verifies whether the in-sample improvements delivered by nonlinear kernels also translates into superior OOS performance. Second, it allows us to examine whether appropriately shrunk low-variance K-PCs improve the characterization of the SDF out-of-sample. The relevance of low-variance K-PCs is inherently an out-of-sample issue: in sample, their contribution is hard to separate from overfitting, whereas out-of-sample evaluation reveals whether additional, lower-variance directions contain stable pricing information. Accordingly, we test (a) whether polynomial and Gaussian kernels attain higher out-of-sample Sharpe ratios and positive intercepts in a horse-race against the linear kernel; and (b) whether an estimator that retains low-variance K-PCs with disciplined shrinkage (CUPSA) outperforms a standard ridge-type

(a) Loading on the linear component $R^*_{l,t}$




(b) Loading on nonlinear component $R^{*,\perp}_{l,t}$

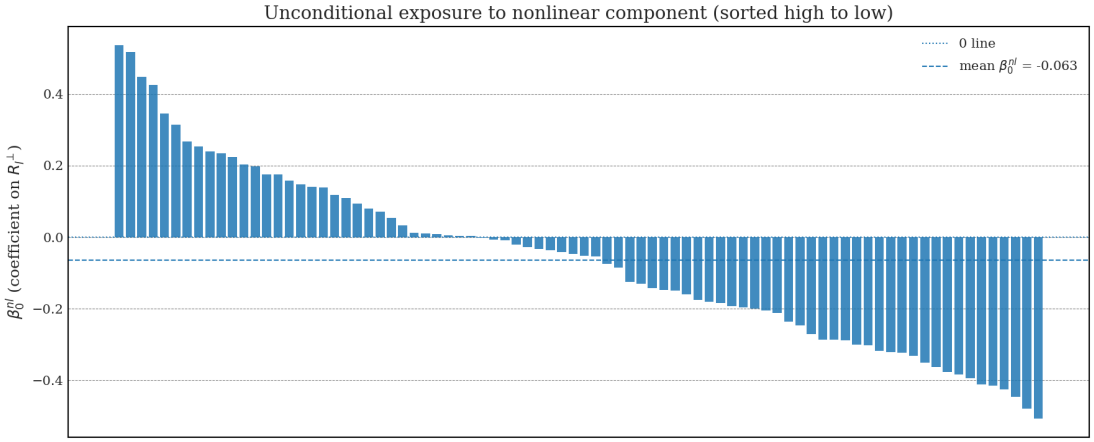
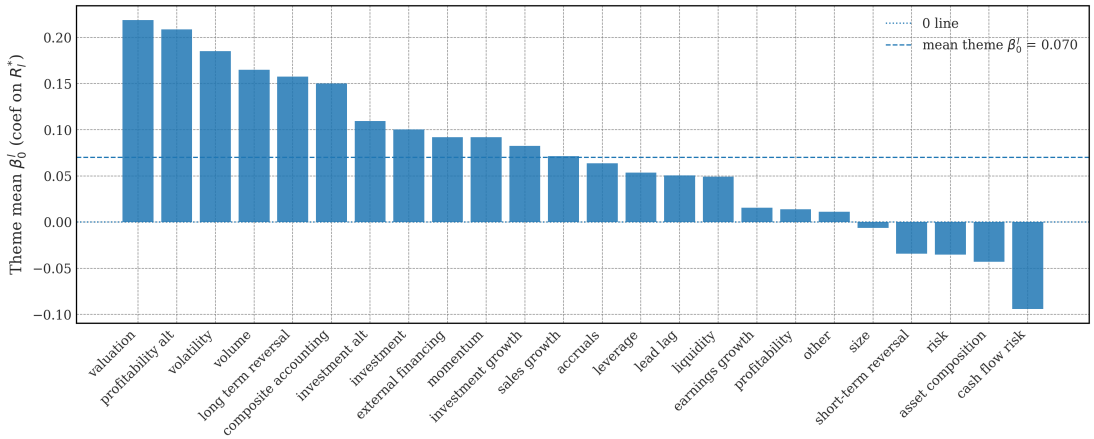



Figure 7: Unconditional exposure to the linear and nonlinear components. Bars plot per-asset unconditional loadings on the orthogonal nonlinear MVE return $R^{*,\perp}_{l,t}$, i.e., the constant slope $\beta^{nl}_{0,i}$ from the expanded (macro-instrumented) specification before interactions with $z_t$. Assets are sorted high to low by the signed $\beta^{nl}_{0,i}$; the dotted line marks zero and the dashed line reports the cross-asset mean exposure. Large magnitudes indicate assets whose returns co-move more strongly, on average, with the nonlinear SDF component. Underlying MVE portfolio returns for the linear and Gaussian kernel are estimated in-sample via K-PCA and KNS regularization. Hyperparameters are selected through a four-fold CV procedure. The estimation sample ranges from 1974 to 2022.

(a) Loading on the linear component $R^*_{l,t}$




(b) Loading on nonlinear component $R^{*,\perp}_{l,t}$

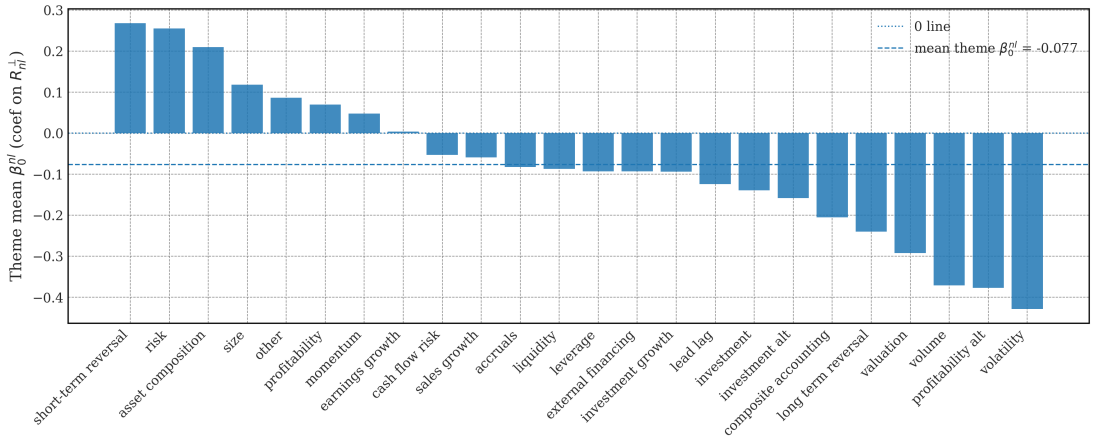



Figure 8: Unconditional linear and nonlinear exposure by economic category. Bars plot economic category-level averages of the unconditional loadings on the linear and orthogonal component of the MVE return, i.e., $R^*_{l,t}$ and $R^{*,\perp}_{l,t}$ from the expanded (macro-instrumented) specification before interactions with $z_t$. Categories are sorted high to low by the signed mean loading. The dashed line reports the cross-theme mean. Underlying MVE portfolio returns for the linear and Gaussian kernel are estimated in-sample via K-PCA and KNS regularization. Hyperparameters are selected through a four-fold CV procedure. The estimation sample ranges from 1974 to 2022.

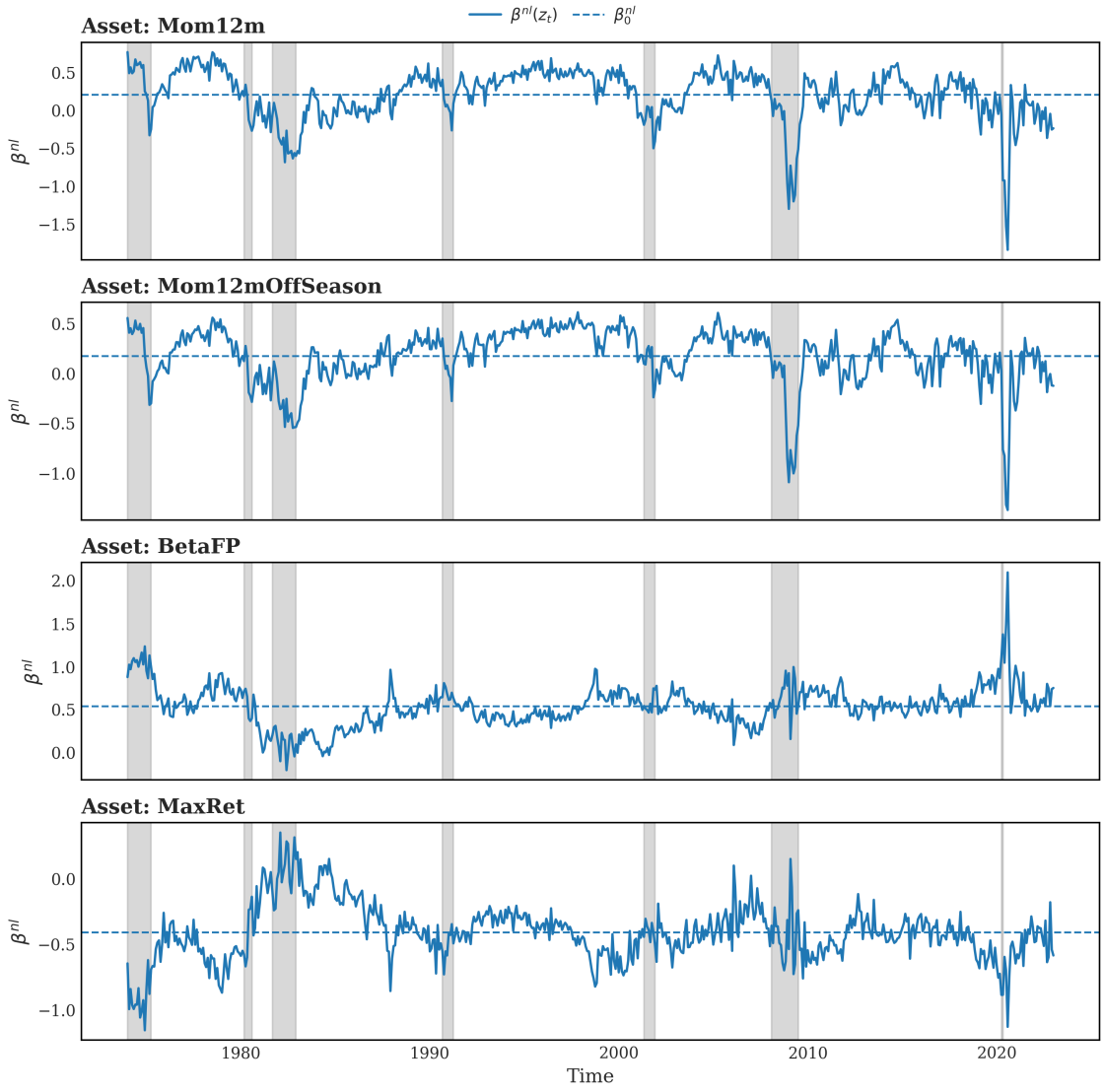



Figure 9: Time-varying nonlinear exposures. Each panel plots the conditional loading on the orthogonal nonlinear MVE return for each test asset, $\beta_i^{nl}(z_t)$ (solid), and the corresponding unconditional coefficient, $\beta_0^{nl}$, on $R_{l,t+1}^{*,\perp}$ (dashed). Shaded regions denote U.S. recessions. Underlying MVE portfolio returns for the linear and Gaussian kernel are estimated in-sample via K-PCA and KNS regularization. Hyperparameters are selected through a four-fold CV procedure. The estimation sample ranges from 1974 to 2022.

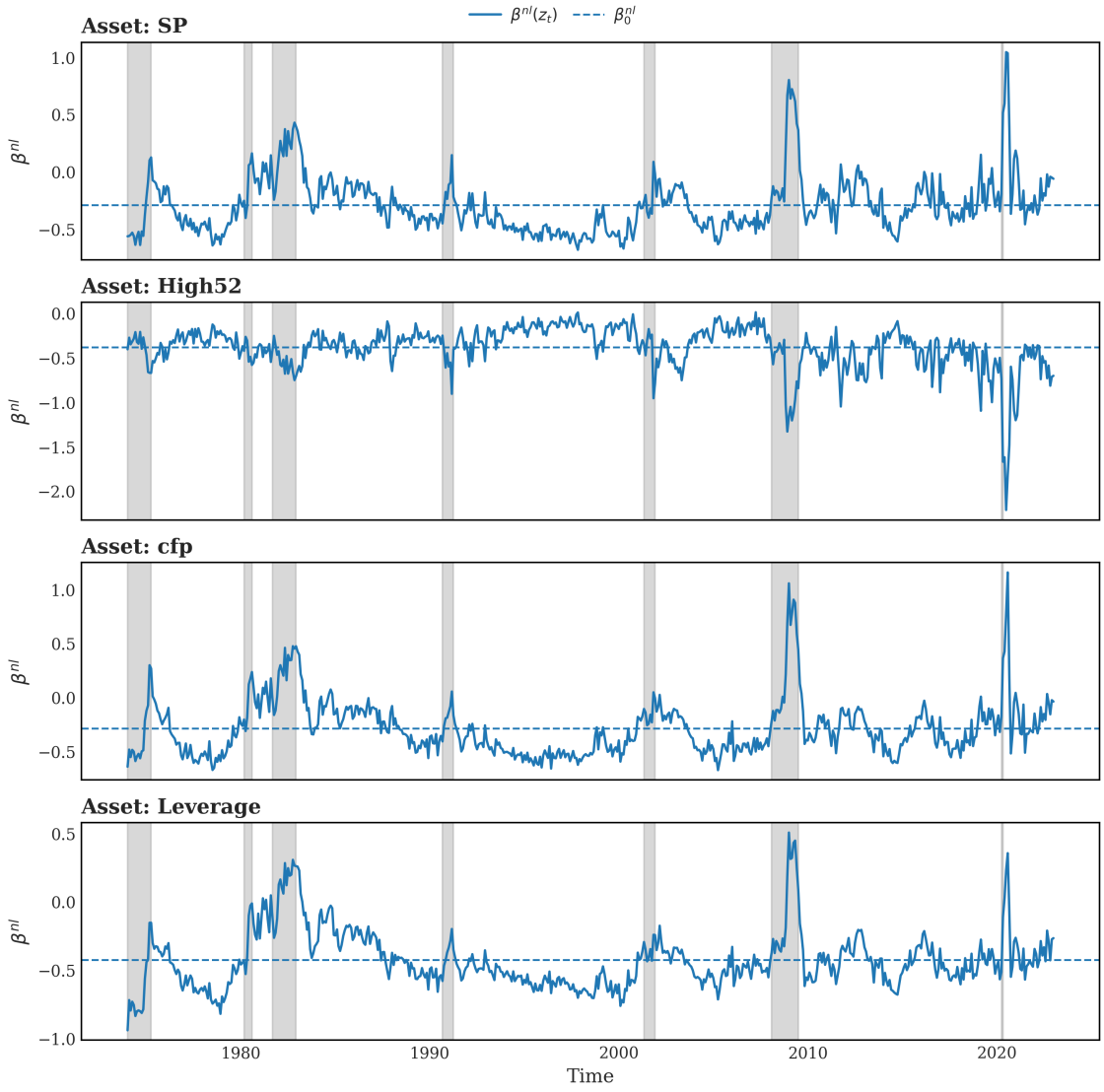



Figure 10: Time-varying nonlinear exposures. Each panel plots the conditional loading on the orthogonal nonlinear MVE return for each test asset, $\beta_i^{nl}(z_t)$ (solid), and the corresponding unconditional coefficient, $\beta_0^{nl}$, on $R_{l,t+1}^{*,\perp}$ (dashed). Shaded regions denote U.S. recessions. Underlying MVE portfolio returns for the linear and Gaussian kernel are estimated in-sample via K-PCA and KNS regularization. Hyperparameters are selected through a four-fold CV procedure. The estimation sample ranges from 1974 to 2022.

alternative (KNS) as the set of retained PCs is expanded.

Our analysis builds on a the full time series of out-of-sample portfolio returns for each model specification. We implement a monthly expanding-window design with a 20-year warm-up (first out-of-sample observation January 1994). At each month $t$, using data up to $t$, we compute K-PCs for the different types of kernels. We then select the optimal hyperparameter $c$ via 4-fold cross-validation. For KNS, we collect cross-validated Sharpe ratios in the same way as detailed in the previous section. For CUPSA, we split the training PC observations into four equal folds and iteratively use three folds to estimate the portfolio weights, evaluating the fitted portfolio by the OOS Sharpe ratio it achieves on the left-out fold. By rotating the left-out fold and averaging the resulting Sharpe ratios, we obtain the cross-validated Sharpe ratio for each kernel-hyperparameter combination. The optimal kernel parameter $c^*$ is then selected as the one that maximizes this metric. Fixing $\kappa$ and the selected hyperparameter, we re-estimate the weights on the full training sample up to $t$ and compute the one-step-ahead realization of the out-of-sample return

$$\hat{R}^{*,\text{OOS}}_{i,\kappa,t+1} = \hat{b}'_{i,\kappa,t} P^{\text{OOS}}_{\kappa,t+1},$$

where $P^{\text{OOS}}_{\kappa,t+1}$ are the out-of-sample return K-PCs at $t$+1 and $i \in \{KNS, CUPSA\}$ . Iterative re-estimation yields a time series of out-of-sample returns for each specification of estimator and kernel $(i, \kappa)$. The annualized average return paths are shown in Figure 11 and Figure 12.

To benchmark nonlinear against linear specifications we run, for each estimator $i$, the time–series regression

$$\hat{R}^{*,\text{OOS}}_{i,nl,t+1} = \alpha_i + \beta_i \, \hat{R}^{*,\text{OOS}}_{i,l,t+1} + \varepsilon_{t+1}, \tag{60}$$

and report annualized intercepts, heteroskedasticity-adjusted standard errors (five lags), and OOS Sharpe ratios in Table 2. Before estimation we rescale each OOS return series to match the volatility of the CRSP value-weighted market index, which places coefficients on a comparable scale without altering the relative performance rankings. The evidence is consistent with our previous findings, as moving from a linear to a nonlinear model specification significantly improves portfolio performance across both estimators. For KNS, the OOS Sharpe ratio rises from 1.31 (linear) to 1.70 (Gaussian), an increase of about 30%; the Gaussian specification also delivers a large and statistically significant annualized intercept of 12.02% with $\hat{\beta} = 0.72$. CUPSA exhibits a higher baseline Sharpe ratio but similarly strong nonlinear gains: Sharpe ratios increase from 1.71 (linear) to 2.04 (poly2, +19%) and 2.15 (Gaussian, +25%), accompanied by significant intercepts of 11.12% and 14.19%, respectively, and slope estimates $\hat{\beta}$ below one. Positive intercepts together with $\beta < 1$


Figure 11: Out-of-sample MVE returns by kernel specification. The graph displays the annualized out-of-sample returns of MVE portfolios constructed from kernel principal components, evaluated across different kernel choices. We re-estimate optimal portfolio for each month with an expanding window approach. The kernel parameter $c^*$ is selected via 4-fold cross-validation based on maximizing the out-of-sample Sharpe ratio. Given $c^*$, we apply the CUPSA algorithm to derive optimal portfolio weights using the corresponding kernel PC factors. All return series are volatility-matched to the CRSP value-weighted market index. Our data has monthly frequency with an out-of-sample period from 1994 to 2022.

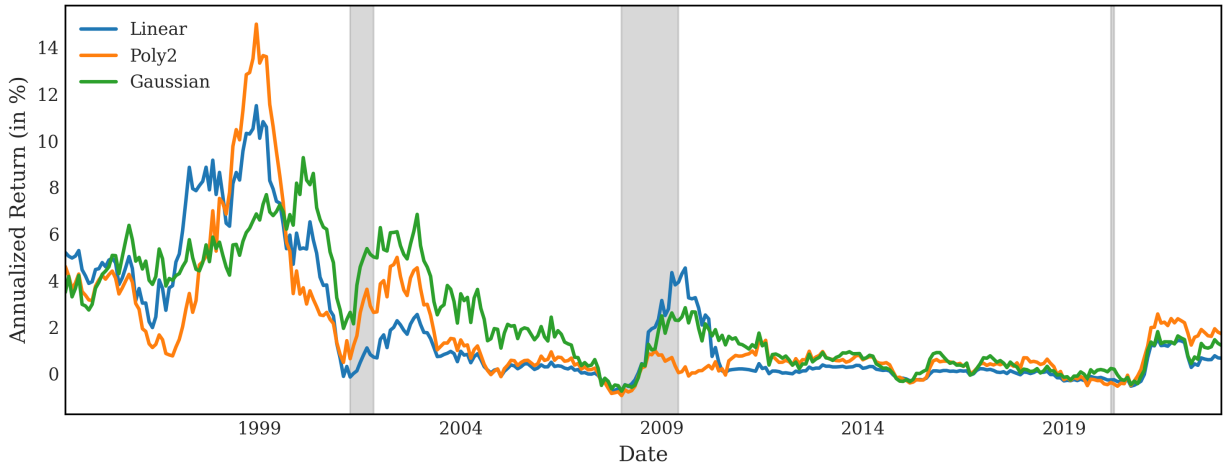



Figure 12: Out-of-sample MVE returns by kernel specification. The graph displays the annualized out-of-sample returns of MVE portfolios constructed from kernel principal components, evaluated across different kernel choices. We re-estimate optimal portfolio for each month with an expanding window approach. The kernel parameter $c^*$ is selected via 4-fold cross-validation based on maximizing the out-of-sample Sharpe ratio. Given $c^*$, we apply the KNS algorithm to derive optimal portfolio weights using the corresponding kernel PC factors. All return series are volatility-matched to the CRSP value-weighted market index. Our data has monthly frequency with an out-of-sample period from 1994 to 2022.

indicate that the nonlinear MVE return contains a component not spanned by the linear kernel.

Table 2: OOS OUTPERFORMANCE OF NON-LINEAR KERNELS. Each panel reports the $\alpha$ (annualized, in %) and $\beta_{\text{linear}}$ from the time-series regression in (60) for MVE returns built with different kernels. We also report annualized OOS Sharpe ratios. Standard errors are in parentheses. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

| | | **Linear** | **Poly2** | **Gaussian** |
|---|---|---|---|---|
| *Panel A: KNS* | $\hat{\alpha}$ (annualized, in %) | - | 4.17*** | 12.02*** |
| | | | (1.43) | (2.29) |
| | $\hat{\beta}_{\text{linear}}$ | - | 0.82*** | 0.72*** |
| | | | (0.12) | (0.08) |
| | OOS Sharpe ratio (annualized) | 1.31 | 1.33 | 1.70 |
| *Panel B: CUPSA* | $\hat{\alpha}$ (annualized, in %) | - | 11.12*** | 14.19*** |
| | | | (1.69) | (2.02) |
| | $\hat{\beta}_{\text{linear}}$ | - | 0.78*** | 0.73*** |
| | | | (0.05) | (0.04) |
| | OOS Sharpe ratio (annualized) | 1.71 | 2.04 | 2.15 |

We next study how nonlinearities contribute as the information set expands from dominant to progressively lower–variance KPCs. For each kernel $k$ and estimator $i$, we repeat the full OOS construction for a fixed number of KPCs $J = J_{\min}, J_{\min}+\Delta, \ldots, J_{\max}$ used in each monthly estimation step:

$$\hat{R}^{*,\text{OOS}}_{i,J,\kappa,t+1} = \hat{b}'_{i,J,\kappa,t} P^{\text{OOS}}_{J,\kappa,t+1},$$

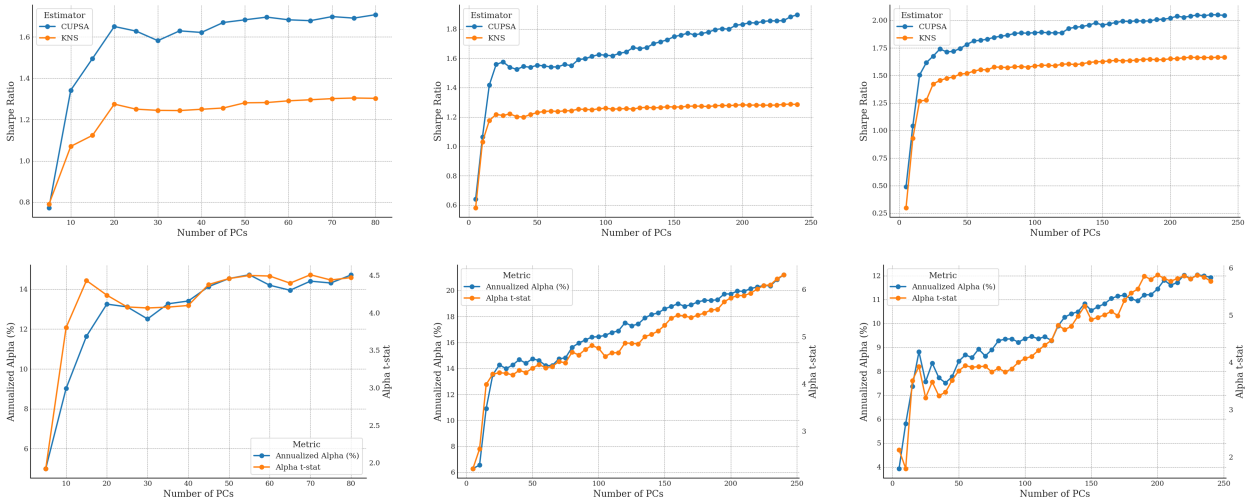
This yields a sequence of OOS return series $\{\hat{R}^{*,\text{OOS}}_{i,\kappa,J,t}\}_J$. To compare estimators' performance as $J$ increases, we run for each tuple of $(\kappa, J)$ the following regression:

$$\hat{R}^{*,\text{OOS}}_{\text{CUPSA},\kappa,J,t+1} = \alpha_J^{(\kappa)} + \beta_J^{(\kappa)} \hat{R}^{*,\text{OOS}}_{\text{KNS},\kappa,J,t+1} + \epsilon_{t+1}. \tag{61}$$

In Figure 13, we report the performance of both estimation setups as $J$ grows, with the out-of-sample Sharpe ratios displayed in the upper panel, along with the annualized $\hat{\alpha}_J^{(\kappa)}$ and its $t$-statistic in the lower panel.

The linear kernel serves as a benchmark consistent with the prevailing literature. CUPSA's Sharpe ratio advantage over KNS is realized almost entirely within the first 20 principal components, after which both estimators exhibit only muted improvements. The corresponding $\hat{\alpha}_J^{(\text{lin})}$ levels off once $J$ exceeds roughly 20. This pattern indicates that, in the linear (tradable-return) space, the SDF is effectively sparse and can be approximated out of sample with a small set of leading variance components—aligning with near-arbitrage arguments that concentrate pricing power in high-variance PCs.

For the polynomial and Gaussian kernels, the picture changes markedly. As $J$ increases beyond 50,




(a) Linear Kernel (b) Polynomial kernel ($2^{\text{nd}}$ order) (c) Gaussian Kernel

Figure 13: CUPSA OUTPERFORMANCE FOR INCREASING NUMBER OF PCs. This plot reports the performance spread between CUPSA and KNS as the number of PCs used in the estimation increases. MVE returns are estimated expanding window, given a fixed maximum number of $J$ top-eigenvalue PCs for each monthly estimation step. The first row of charts in each panel reports the Sharpe ratios delivered by the CUPSA and KNS, given $J$. The second row reports the $\alpha$ and $\alpha$ t-stat from regression in 61. We run CK-PCA on the data covering 82 characteristics by Chen and Zimmermann (2022).

CUPSA continues to translate additional KPCs into economically and statistically significant gains relative to KNS: Sharpe ratios keep rising with $J$, and both $\hat{\alpha}_J^{(k)}$ and its $t$-statistic trend upward. The performance gap is widest under the polynomial kernel, where higher-order interaction structure is spread across many directions, but it is also pronounced under the Gaussian kernel, where the richer (universal) feature space generates incremental improvements well past the leading components.

The results presented in this section imply the following takeaways. (1) Low-variance PCs can, when shrunk efficiently, contribute significantly to an enhanced performance of the empirical SDF. The performance gains survive out-of-sample and cannot be attributed to an increased risk exposure. Therefore, it appears infeasible to derive a close empirical approximation of the true SDF when restricting the estimation to a sparse set of high-eigenvalue PCs. (2) The relevance of low-variance PCs increases with the degree to which non-linearities are considered in the model. The second statement follows from the observation that the CUPSA–KNS performance gap widens most significantly when constructing PCs via the polynomial and Gaussian kernel. Therefore, we provide additional evidence for nonlinear kernels uncovering a meaningful orthogonal component which improves the out-of-sample approximation quality of the SDF model.

# 4 Conclusion

This paper shows that incorporating risk factors constructed from nonlinear interactions of firm characteristics—via kernel principal component analysis—improves the estimation of the SDF and its pricing performance. By decomposing the kernel-implied SDF into a linear and an orthogonal nonlinear component, we obtain a two-factor representation which allows us to focus on the pricing contribution of the nonlinear component of the SDF.

Empirically, nonlinear modeling delivers several performance improvements: Hansen–Jagannathan distances reduce across nearly all economic categories, the average test-asset $R^2$ rises from 17% to 25% in the unconditional model (with pricing errors halved from 2% to 1%), and a conditional specification tied to macro-financial states increases the average $R^2$ to 33%. The Gaussian kernel dominates both linear and fixed-order polynomial alternatives, consistent with its greater flexibility and universality in approximating continuous functions. With CUPSA regularization, retaining low-variance principal components enhances pricing accuracy and out-of-sample performance; the MVE Sharpe ratio increases by roughly 25% (2.15 versus 1.73), expanding the mean–variance frontier.

We also uncover heterogeneous exposures to the nonlinear component—beta/BAB load positively, while cash-flow-to-price and book-to-market load negatively—and, for some portfolios, the linear and nonlinear loadings have opposite signs. Taken together, these results indicate that nonlinear structure is a first-order feature of the SDF in rich, high-dimensional settings and that exploiting it yields robust improvements in both fit and performance.

# A Appendix

## A.1 OOS Moment Estimates

Leave-One-Out (LOO) is a special case of cross-validation, which iteratively treats each data point in the sample as an OOS observation. The CUPSA estimator builds on that methodology in order to compute empirical estimates of the OOS return and covariance structure of returns. To describe the construction of those estimates, let $F_t$ denote the realization of factor returns at time $t$ and assume access to a sample ranging from $t \in \{1, \dots, T\}$. Then we can define the LOO first and second moment estimates for any $t$ as the empirical moments computed on all observations except for $t$:

$$\bar{\Sigma}_{T,t} = \frac{1}{T} \sum_{\substack{\tau \neq t \\ 1 \leq \tau \leq T}} F_\tau F_\tau' \tag{62}$$

$$\bar{\mu}_{T,t} = \frac{1}{T} \sum_{\substack{\tau \neq t \\ 1 \leq \tau \leq T}} F_\tau \tag{63}$$

We can next use these moment estimates to compute the Ridge portfolio for a given shrinkage level $z$:

$$\hat{b}_{T,t}(f_z) = \left(\bar{\Sigma}_{T,t} + zI\right)^{-1} \bar{\mu}_{T,t} \tag{64}$$

and its OOS return at $t$:

$$\hat{R}_{T,t}^{MVE}(f_z) = \hat{b}_{T,t}(f_z)' F_t. \tag{65}$$

By iterating over all observations in the sample, we can construct a time series of OOS returns for each Ridge portfolio in the grid $Z = \{z_\ell\}_{\ell=1}^M$. These time series serve as the basis for computing the OOS estimates for the mean and covariance matrix of the Ridge portfolios using:

$$\bar{\mu}(Z) = \left( \frac{1}{T} \sum_{t=1}^{T} \hat{R}_{T,t}^{MVE}(f_{z_i}) \right)_{i=1}^{M} \in \mathbb{R}^M, \tag{66}$$

and

$$\bar{\Sigma}(Z) = \left( \frac{1}{T} \sum_{t=1}^{T} \hat{R}_{T,t}^{MVE}(f_{z_i}) \hat{R}_{T,t}^{MVE}(f_{z_j}) \right)_{i,j=1}^{M} \in \mathbb{R}^{M \times M}. \tag{67}$$

Table 3: Asset pricing test results for individual test assets. This table reports coefficient estimates from time-series regressions (57) and (58) for all 82 test assets. Reported $R^2$ values are in percent. The intercept $\hat{\alpha}$ is annualized and reported in percent per year.

| | Linear | | | Linear + Poly2 | | | | Linear + Gaussian | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\hat{\alpha}$ | $R^2$ | $\hat{\beta}_l$ | $\hat{\alpha}$ | $R^2$ | $\hat{\beta}_l$ | $\hat{\beta}_{nl}$ | $\hat{\alpha}$ | $R^2$ | $\hat{\beta}_l$ | $\hat{\beta}_{nl}$ |
| Test Asset | | | | | | | | | | | |
| Accruals | 0.96 | 0.07 | -0.00 | 0.99 | 0.08 | -0.00 | -0.00 | 0.86 | 0.10 | -0.00 | 0.01 |
| AssetGrowth | 2.75 | 48.00 | 0.18 | 1.69 | 52.31 | 0.18 | -0.09 | 1.31 | 65.55 | 0.18 | -0.25 |
| Beta | 4.93 | 13.05 | -0.20 | 2.40 | 18.53 | -0.20 | 0.22 | 3.23 | 28.85 | -0.20 | 0.50 |
| BetaFP | 4.33 | 13.17 | -0.20 | 2.01 | 18.03 | -0.20 | 0.20 | 3.46 | 28.28 | -0.20 | 0.48 |
| BetaTailRisk | 2.58 | 1.37 | -0.04 | 1.88 | 2.28 | -0.04 | 0.06 | 2.63 | 15.02 | -0.04 | 0.32 |
| BidAskSpread | 4.82 | 22.17 | -0.22 | 2.44 | 29.07 | -0.22 | 0.21 | 1.96 | 37.61 | -0.22 | 0.42 |
| BMdec | 4.00 | 40.60 | 0.22 | 2.36 | 46.74 | 0.22 | -0.14 | 0.73 | 54.77 | 0.22 | -0.29 |
| BookLeverage | 3.37 | 28.76 | -0.17 | 1.50 | 37.60 | -0.17 | 0.16 | 1.01 | 42.15 | -0.17 | 0.27 |
| Cash | 6.12 | 26.33 | -0.20 | 3.63 | 36.81 | -0.20 | 0.22 | 1.14 | 50.90 | -0.20 | 0.45 |
| CashProd | 3.05 | 19.63 | 0.16 | 0.79 | 30.13 | 0.16 | -0.20 | 1.44 | 31.13 | 0.16 | -0.28 |
| CBOperProf | 2.08 | 1.14 | 0.02 | 0.74 | 13.34 | 0.02 | 0.12 | 0.25 | 7.43 | 0.02 | 0.11 |
| CF | 3.23 | 26.96 | 0.23 | 0.74 | 35.34 | 0.23 | -0.22 | 2.02 | 37.20 | 0.23 | -0.32 |
| cfp | 3.49 | 26.68 | 0.21 | 1.10 | 35.93 | 0.21 | -0.21 | 0.97 | 35.52 | 0.21 | -0.28 |
| ChEQ | 2.41 | 41.58 | 0.17 | 1.48 | 45.02 | 0.17 | -0.08 | 1.75 | 60.54 | 0.17 | -0.26 |
| ChInv | 0.73 | 12.10 | 0.06 | 0.44 | 12.96 | 0.06 | -0.03 | 0.68 | 17.44 | 0.06 | -0.09 |
| ChInvIA | 0.61 | 12.83 | 0.05 | 0.37 | 13.56 | 0.05 | -0.02 | 0.35 | 16.08 | 0.05 | -0.06 |
| CompEquIss | 2.42 | 5.09 | -0.07 | 0.78 | 13.20 | -0.07 | 0.14 | 1.37 | 17.12 | -0.07 | 0.23 |
| CompositeDebtIssuance | 0.32 | 14.14 | 0.05 | 0.06 | 15.19 | 0.05 | -0.02 | 0.44 | 16.66 | 0.05 | -0.05 |
| Coskewness | 1.40 | 0.31 | 0.01 | 2.07 | 2.74 | 0.01 | -0.06 | 0.72 | 6.89 | 0.01 | 0.13 |
| DelCOA | 2.44 | 27.81 | 0.13 | 1.05 | 36.00 | 0.13 | -0.12 | 1.19 | 43.28 | 0.13 | -0.22 |
| DelCOL | 3.27 | 29.01 | 0.14 | 1.88 | 36.90 | 0.14 | -0.12 | 0.72 | 46.97 | 0.14 | -0.25 |
| DelFINL | 0.58 | 6.27 | 0.03 | 0.05 | 11.80 | 0.03 | 0.05 | 0.09 | 8.76 | 0.03 | 0.04 |
| DelLTI | 0.09 | 2.33 | 0.02 | 0.27 | 2.94 | 0.02 | 0.02 | 0.14 | 2.62 | 0.02 | -0.01 |

Continued on next page

Table 3: Asset pricing test results for individual test assets. This table reports coefficient estimates from time-series regressions (57) and (58) for all 82 test assets. Reported $R^2$ values are in percent. The intercept $\hat{\alpha}$ is annualized and reported in percent per year.

| | Linear | | | Linear + Poly2 | | | | Linear + Gaussian | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\hat{\alpha}$ | $R^2$ | $\hat{\beta}_l$ | $\hat{\alpha}$ | $R^2$ | $\hat{\beta}_l$ | $\hat{\beta}_{nl}$ | $\hat{\alpha}$ | $R^2$ | $\hat{\beta}_l$ | $\hat{\beta}_{nl}$ |
| Test Asset | | | | | | | | | | | |
| DelNetFin | 1.48 | 1.83 | -0.02 | 0.99 | 5.74 | -0.02 | 0.04 | 0.32 | 16.18 | -0.02 | 0.11 |
| EarningsSurprise | 0.18 | 3.00 | 0.03 | 0.11 | 4.01 | 0.03 | -0.03 | 0.12 | 3.29 | 0.03 | -0.02 |
| EBM | 0.22 | 2.79 | 0.04 | 1.20 | 7.44 | 0.04 | -0.09 | 0.56 | 2.95 | 0.04 | -0.02 |
| EntMult | 1.97 | 23.92 | 0.19 | 0.55 | 35.06 | 0.19 | -0.22 | 1.23 | 28.89 | 0.19 | -0.20 |
| EP | 3.53 | 30.86 | 0.20 | 1.69 | 37.90 | 0.20 | -0.16 | 0.84 | 41.88 | 0.20 | -0.27 |
| EquityDuration | 4.02 | 45.68 | 0.24 | 2.34 | 51.56 | 0.24 | -0.15 | 0.43 | 57.08 | 0.24 | -0.27 |
| GP | 1.47 | 0.10 | 0.01 | 0.39 | 4.47 | 0.01 | 0.10 | 1.22 | 7.56 | 0.01 | 0.17 |
| grcapx | 1.34 | 35.57 | 0.11 | 0.98 | 36.60 | 0.11 | -0.03 | 0.56 | 43.55 | 0.11 | -0.12 |
| GrLTNOA | 0.13 | 0.20 | -0.01 | 0.11 | 0.20 | -0.01 | 0.00 | 0.28 | 0.29 | -0.01 | -0.01 |
| GrSaleToGrInv | 0.66 | 0.74 | 0.01 | 0.11 | 9.32 | 0.01 | 0.07 | 0.43 | 5.50 | 0.01 | 0.07 |
| Herf | 2.66 | 19.98 | -0.11 | 1.28 | 28.49 | -0.11 | 0.12 | 0.49 | 32.21 | -0.11 | 0.19 |
| High52 | 4.48 | 20.57 | 0.21 | 5.01 | 20.93 | 0.21 | 0.05 | 1.46 | 32.96 | 0.21 | -0.37 |
| hire | 2.72 | 39.20 | 0.14 | 1.68 | 44.61 | 0.14 | -0.09 | 0.98 | 58.21 | 0.14 | -0.23 |
| IdioVol3F | 4.70 | 23.71 | 0.24 | 2.11 | 30.84 | 0.24 | -0.23 | 2.74 | 39.95 | 0.24 | -0.46 |
| Illiquidity | 1.14 | 13.93 | 0.08 | 1.25 | 14.00 | 0.08 | 0.01 | 0.17 | 15.37 | 0.08 | -0.06 |
| IndMom | 1.78 | 5.32 | 0.09 | 4.17 | 14.38 | 0.09 | 0.21 | 0.68 | 5.85 | 0.09 | -0.07 |
| IntMom | 2.27 | 0.29 | 0.02 | 0.16 | 8.75 | 0.02 | 0.19 | 2.15 | 10.57 | 0.02 | 0.27 |
| Investment | 0.68 | 2.10 | 0.02 | 0.04 | 8.86 | 0.02 | 0.06 | 0.25 | 2.75 | 0.02 | 0.03 |
| InvestPPEInv | 1.38 | 36.80 | 0.13 | 0.56 | 41.02 | 0.13 | -0.07 | 0.43 | 42.43 | 0.13 | -0.11 |
| Leverage | 4.69 | 29.08 | 0.25 | 1.79 | 39.91 | 0.25 | -0.26 | 1.46 | 42.52 | 0.25 | -0.38 |
| LRreversal | 3.67 | 36.26 | 0.20 | 3.48 | 36.35 | 0.20 | -0.02 | 0.64 | 48.67 | 0.20 | -0.27 |
| MaxRet | 3.93 | 22.85 | 0.23 | 1.54 | 29.33 | 0.23 | -0.21 | 3.02 | 38.01 | 0.23 | -0.43 |
| MeanRankRevGrowth | 2.45 | 24.92 | 0.11 | 1.15 | 34.34 | 0.11 | -0.11 | 0.11 | 33.38 | 0.11 | -0.14 |

Continued on next page

Table 3: Asset pricing test results for individual test assets. This table reports coefficient estimates from time-series regressions (57) and (58) for all 82 test assets. Reported $R^2$ values are in percent. The intercept $\hat{\alpha}$ is annualized and reported in percent per year.

| | Linear | | | Linear + Poly2 | | | | Linear + Gaussian | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\hat{\alpha}$ | $R^2$ | $\hat{\beta}_l$ | $\hat{\alpha}$ | $R^2$ | $\hat{\beta}_l$ | $\hat{\beta}_{nl}$ | $\hat{\alpha}$ | $R^2$ | $\hat{\beta}_l$ | $\hat{\beta}_{nl}$ |
| Test Asset | | | | | | | | | | | |
| Mom12m | 0.35 | 5.69 | 0.11 | 2.66 | 16.35 | 0.11 | 0.27 | 1.43 | 6.72 | 0.11 | 0.11 |
| Mom12mOffSeason | 0.68 | 2.33 | 0.07 | 2.03 | 11.26 | 0.07 | 0.24 | 1.44 | 3.83 | 0.07 | 0.13 |
| Mom6m | 1.46 | 5.86 | 0.11 | 3.84 | 13.62 | 0.11 | 0.21 | 0.70 | 6.09 | 0.11 | -0.05 |
| MomOffSeason | 4.23 | 45.68 | 0.25 | 2.91 | 49.04 | 0.25 | -0.12 | 1.91 | 65.65 | 0.25 | -0.38 |
| MomOffSeason06YrPlus | 1.12 | 27.06 | 0.12 | 0.05 | 32.74 | 0.12 | -0.09 | 1.08 | 33.75 | 0.12 | -0.14 |
| MomSeason | 2.68 | 8.37 | -0.06 | 2.66 | 8.37 | -0.06 | 0.00 | 0.66 | 15.98 | -0.06 | 0.12 |
| MomSeason06YrPlus | 2.26 | 0.31 | -0.01 | 1.82 | 2.03 | -0.01 | 0.04 | 0.61 | 6.88 | -0.01 | 0.10 |
| MomSeasonShort | 1.47 | 0.05 | -0.01 | 0.37 | 3.83 | -0.01 | 0.10 | 0.64 | 3.91 | -0.01 | 0.13 |
| MRreversal | 1.85 | 17.22 | 0.14 | 0.81 | 19.93 | 0.14 | -0.09 | 1.48 | 24.85 | 0.14 | -0.21 |
| NetDebtFinance | 0.60 | 1.60 | 0.01 | 0.05 | 7.69 | 0.01 | 0.05 | 0.28 | 5.82 | 0.01 | 0.05 |
| NetEquityFinance | 2.81 | 30.11 | 0.17 | 1.03 | 38.55 | 0.17 | -0.16 | 1.02 | 40.81 | 0.17 | -0.24 |
| NOA | 0.43 | 20.33 | 0.09 | 0.62 | 20.56 | 0.09 | 0.02 | 0.02 | 20.64 | 0.09 | -0.03 |
| OPLeverage | 0.60 | 0.90 | 0.02 | 0.04 | 2.85 | 0.02 | 0.06 | 1.07 | 4.58 | 0.02 | 0.10 |
| PriceDelayRsq | 2.32 | 12.99 | 0.09 | 2.32 | 12.99 | 0.09 | 0.00 | 0.83 | 25.40 | 0.09 | -0.19 |
| PriceDelaySlope | 0.46 | 0.51 | -0.01 | 0.06 | 1.77 | -0.01 | -0.04 | 0.55 | 0.53 | -0.01 | 0.01 |
| PriceDelayTstat | 2.07 | 23.78 | 0.10 | 1.38 | 26.95 | 0.10 | -0.06 | 0.98 | 40.90 | 0.10 | -0.19 |
| RDS | 1.90 | 29.27 | 0.14 | 1.95 | 29.28 | 0.14 | 0.00 | 0.69 | 36.94 | 0.14 | -0.16 |
| ResidualMomentum | 0.11 | 8.72 | 0.08 | 0.40 | 9.68 | 0.08 | 0.04 | 0.39 | 8.98 | 0.08 | 0.03 |
| ReturnSkew | 0.66 | 9.50 | 0.04 | 0.63 | 9.51 | 0.04 | -0.00 | 0.10 | 10.74 | 0.04 | -0.03 |
| roaq | 1.13 | 0.43 | 0.02 | 0.02 | 5.37 | 0.02 | 0.10 | 0.66 | 0.68 | 0.02 | 0.03 |
| RoE | 0.73 | 9.59 | 0.07 | 0.53 | 9.82 | 0.07 | -0.02 | 0.02 | 10.37 | 0.07 | -0.04 |
| ShareIss1Y | 2.20 | 49.27 | 0.18 | 1.08 | 54.36 | 0.18 | -0.10 | 0.89 | 60.04 | 0.18 | -0.19 |
| Size | 0.49 | 0.01 | 0.00 | 0.40 | 0.05 | 0.00 | 0.01 | 0.83 | 2.64 | 0.00 | 0.08 |

Continued on next page

Table 3: Asset pricing test results for individual test assets. This table reports coefficient estimates from time-series regressions (57) and (58) for all 82 test assets. Reported $R^2$ values are in percent. The intercept $\hat{\alpha}$ is annualized and reported in percent per year.

| | Linear | | | Linear + Poly2 | | | | Linear + Gaussian | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\hat{\alpha}$ | $R^2$ | $\hat{\beta}_l$ | $\hat{\alpha}$ | $R^2$ | $\hat{\beta}_l$ | $\hat{\beta}_{nl}$ | $\hat{\alpha}$ | $R^2$ | $\hat{\beta}_l$ | $\hat{\beta}_{nl}$ |
| Test Asset | | | | | | | | | | | |
| SP | 3.37 | 33.95 | 0.25 | 0.87 | 43.09 | 0.25 | -0.22 | 0.96 | 41.51 | 0.25 | -0.27 |
| STreversal | 3.31 | 1.71 | -0.05 | 4.53 | 5.09 | -0.05 | -0.11 | 0.58 | 11.17 | -0.05 | 0.24 |
| Tax | 0.43 | 10.51 | 0.06 | 0.13 | 12.88 | 0.06 | -0.05 | 0.19 | 11.30 | 0.06 | -0.04 |
| TotalAccruals | 0.91 | 18.68 | 0.08 | 0.09 | 24.20 | 0.08 | -0.07 | 0.59 | 23.74 | 0.08 | -0.09 |
| TrendFactor | 3.18 | 3.64 | 0.07 | 2.44 | 4.87 | 0.07 | 0.07 | 1.69 | 18.26 | 0.07 | 0.30 |
| VarCF | 0.09 | 4.74 | -0.06 | 0.18 | 4.97 | -0.06 | -0.02 | 0.77 | 5.39 | -0.06 | -0.05 |
| VolMkt | 5.48 | 21.95 | 0.23 | 3.17 | 28.11 | 0.23 | -0.20 | 2.60 | 42.80 | 0.23 | -0.50 |
| VolSD | 2.59 | 21.23 | 0.14 | 1.47 | 24.68 | 0.14 | -0.10 | 1.34 | 32.87 | 0.14 | -0.24 |
| VolumeTrend | 1.34 | 26.49 | 0.12 | 0.49 | 29.90 | 0.12 | -0.07 | 1.59 | 37.73 | 0.12 | -0.18 |
| XFIN | 2.19 | 41.71 | 0.18 | 1.02 | 46.50 | 0.18 | -0.10 | 0.98 | 51.41 | 0.18 | -0.20 |
| zerotrade6M | 4.81 | 20.74 | 0.22 | 1.97 | 30.04 | 0.22 | -0.25 | 2.90 | 39.68 | 0.22 | -0.48 |
| InvGrowth | 1.58 | 33.16 | 0.13 | 0.90 | 35.79 | 0.13 | -0.06 | 0.56 | 40.37 | 0.13 | -0.13 |
| OperProf | 2.07 | 25.03 | 0.15 | 1.36 | 26.61 | 0.15 | -0.06 | 0.52 | 30.88 | 0.15 | -0.16 |

# References


Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. Journal of Finance 61(1), 259–299.

Asness, C. S., T. J. Moskowitz, and L. H. Pedersen (2013). Value and momentum everywhere. Journal of Finance 68(3), 929–985.

Baker, M. and J. Wurgler (2006). Investor sentiment and the cross-section of stock returns. Journal of Finance 61(4), 1645–1680.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. The Quarterly Journal of Economics 131(4), 1593–1636.

Bansal, R. and A. Yaron (2004). Risks for the long run: A potential resolution of asset pricing puzzles. Journal of Finance 59(4), 1481–1509.

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

Bryzgalova, S., V. DeMiguel, S. Li, and M. Pelger (2023). Asset-pricing factors with economic targets. Working Paper.

Bryzgalova, S., J. Huang, and C. Julliard (2023). Bayesian solutions for the factor zoo: We just ran two quadrillion models. Journal of Finance 78(1), 487–557.

Campbell, J. Y. and J. H. Cochrane (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107(2), 205–251.

Chen, A. Y. and T. Zimmermann (2022). Open source cross-sectional asset pricing. Critical Finance Review 11, 207–264.

Cochrane, J. (2009). Asset pricing: Revised edition. Princeton University Press.

Cochrane, J. H. (2011). Presidential address: Discount rates. Journal of Finance 66(4), 1047–1108.

DeMiguel, V., A. Martin-Utrera, F. J. Nogales, and R. Uppal (2020). A transaction-cost perspective on the multitude of firm characteristics. Review of Financial Studies 33(5), 2180–2222.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. Journal of Finance 47(2), 427–465.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33(1), 3–56.

Fama, E. F. and K. R. French (1996). The capm is wanted, dead or alive. Journal of Finance 51(5), 1947–1958.

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

Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. Journal of Finance 75(3), 1327–1370.

Frazzini, A. and L. H. Pedersen (2014). Betting against beta. Journal of financial economics 111(1), 1–25.

Freyberger, J., B. Höppner, A. Neuhierl, and M. Weber (2025). Missing data in asset pricing panels. Review of Financial Studies 38, 760–802.

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

Giglio, S. and D. Xiu (2021). Asset pricing with omitted factors. Journal of Political Economy 129(7), 1947–1990.

Giglio, S., D. Xiu, and D. Zhang (2025). Test assets and weak factors. Journal of Finance 80(1), 259–319.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. Review of Financial Studies 33(5), 2223–2273.

Hansen, L. P. and R. Jagannathan (1991). Implications of security market data for models of dynamic economies. Journal of Political Economy 99(2), 225–262.

Hansen, L. P. and R. Jagannathan (1997). Assessing specification errors in stochastic discount factor models. Journal of Finance 52(2), 557–590.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. Review of Financial Studies 29(1), 5–68.

He, Z. and A. Krishnamurthy (2013). Intermediary asset pricing. American Economic Review 103(2), 732–770.

Hong, H., T. Lim, and J. C. Stein (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. Journal of Finance 55(1), 265–295.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology 24(6), 417.

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

Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring uncertainty. American Economic Review 105(3), 1177–1216.

Kelly, B. T., S. Malamud, M. Pourmohammadi, and F. Trojani (2024). Universal portfolio shrinkage. Working Paper.

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics 134(3), 501–524.

Kozak, S. (2020). Kernel trick for the cross-section. Working Paper.

Kozak, S., S. Nagel, and S. Santosh (2018). Interpreting factor models. 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.

Lakonishok, J., A. Shleifer, and R. W. Vishny (1994). Contrarian investment, extrapolation, and risk. The journal of finance 49(5), 1541–1578.

Lettau, M. and M. Pelger (2020). Factors that fit the time series and cross-section of stock returns. Review of Financial Studies 33(5), 2274–2325.

Moritz, B. and T. Zimmermann (2016). Tree-based conditional portfolio sorts: The relation between past and future stock returns. Working Paper.

Nagel, S. (2021). Machine learning in asset pricing, Volume 1. Princeton University Press.

Rosenberg, B. (1974). Extra-market components of covariance in security returns. Journal of Financial and Quantitative Analysis 9(2), 263–274.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. Journal of Economic Theory 13(3), 341–360.

Schölkopf, B., A. Smola, and K.-R. Müller (1997). Kernel principal component analysis. In International conference on artificial neural networks, pp. 583–588. Springer.

Schölkopf, B. and A. J. Smola (2018). Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance 19(3), 425–442.

Simon, F., S. Weibels, and T. Zimmermann (2022). Deep parametric portfolio policies. Working Paper.

Stambaugh, R. F., J. Yu, and Y. Yuan (2012). The short of it: Investor sentiment and anomalies. Journal of Financial Economics 104(2), 288–302.

Steinwart, I. (2001). On the influence of the kernel on the consistency of support vector machines. Journal of machine learning research 2(Nov), 67–93.