

# Introduction to Machine Learning

Financial Engineering 2025/2026

EDHEC Business School

## Linear vs. Kernel Models in Simulated Returns

SALIOLA BUCELLI Andrea

Prof. LOPEZ Nicolas

November 27, 2025

## Abstract

This report investigates the effectiveness of kernel-based regression techniques in modeling nonlinear relationships in simulated financial return data. Three synthetic datasets are constructed to reflect increasingly complex market dynamics: a quadratic-only specification, a quadratic model with factor interaction, and a regime-switching extension. Linear regression and kernel ridge regression with a radial basis function (RBF) kernel are evaluated across all scenarios. Performance is assessed using standard predictive accuracy measures, including  $R^2$ , RMSE, MAE, bias, and correlation between actual and predicted returns. The results show that linear regression deteriorates rapidly as nonlinear structure increases, while the kernel model consistently captures nonlinearities, achieves near-perfect fit in all cases, and exhibits substantially lower errors. These findings demonstrate the limitations of linear models in nonlinear financial environments and highlight the suitability of kernel methods for asset return modeling.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Data Simulation</b>	<b>1</b>
2.1	Quadratic Nonlinearity	1
2.2	Quadratic Nonlinearity and Interaction	1
2.3	Quadratic, Interaction, and Regime Switching	2
<b>3</b>	<b>Models and Estimation</b>	<b>2</b>
3.1	Linear Regression	2
3.2	Kernel Ridge Regression with RBF Kernel	2
<b>4</b>	<b>Performance Metrics</b>	<b>3</b>
<b>5</b>	<b>Results</b>	<b>3</b>
5.1	Time-Series Behaviour of Simulated Returns	3
5.2	Linear Regression	4
5.3	RBF Gram Matrix	5
5.4	Kernel Ridge Regression	6
5.5	Summary of Performance Metrics	6
<b>6</b>	<b>High-Noise Environment</b>	<b>7</b>
<b>7</b>	<b>Conclusion</b>	<b>8</b>

# 1 Introduction

The modeling of asset returns is a central task in empirical finance and risk management. Standard linear factor models are analytically convenient, but they often fail to capture nonlinear relationships and regime changes that arise in practice. Kernel methods provide a flexible alternative by implicitly mapping the data into a high-dimensional feature space in which nonlinear relationships can be represented linearly.

The objective of this study is to compare the performance of a linear regression model and a kernel ridge regression model in a controlled simulation framework. Three data-generating processes (DGPs) with increasing nonlinear complexity are considered. This design makes it possible to quantify how model performance degrades for the linear specification and to what extent the kernel method can recover the underlying nonlinear structure.<sup>1</sup>

## 2 Data Simulation

Let  $F_t$  denote a market factor and  $V_t$  a volatility indicator at time  $t$ , for  $t = 1, \dots, n$ , with  $n = 500$ . In all scenarios the following basic structure is used:

$$\begin{aligned} F_t &\sim \mathcal{N}(0, 1), \\ Z_t &\sim \mathcal{N}(0, 1), \quad V_t = |Z_t|, \\ \varepsilon_t &\sim \mathcal{N}(0, \sigma^2), \quad \sigma = 0.03, \end{aligned}$$

where  $F_t$ ,  $Z_t$  and  $\varepsilon_t$  are mutually independent. Asset returns  $R_t$  are generated as nonlinear functions of  $(F_t, V_t)$  plus noise.

The parameter vector

$$\theta = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = (0.002, 0.040, -0.030, 0.060, 0.050, 0.080)$$

is kept fixed across scenarios, but some components are set to zero depending on the specification.

### 2.1 Quadratic Nonlinearity

The first scenario is a baseline nonlinear factor model with a quadratic effect in the market factor:

$$R_t = \theta_0 + \theta_1 F_t + \theta_2 V_t + \theta_3 F_t^2 + \varepsilon_t. \quad (1)$$

This DGP introduces curvature in  $F_t$  but preserves a relatively simple structure. A correctly specified linear model would require inclusion of  $F_t^2$ , which is intentionally omitted in the estimation stage.

### 2.2 Quadratic Nonlinearity and Interaction

The second scenario adds an interaction term between the market factor and volatility:

$$R_t = \theta_0 + \theta_1 F_t + \theta_2 V_t + \theta_3 F_t^2 + \theta_4 F_t V_t + \varepsilon_t. \quad (2)$$

---

<sup>1</sup>All results are fully reproducible using the Python code available at [https://github.com/ansabu01/Linear\\_vs\\_Kernel\\_Models\\_in\\_Simulated>Returns](https://github.com/ansabu01/Linear_vs_Kernel_Models_in_Simulated>Returns).

In this setting the marginal effect of  $F_t$  depends on  $V_t$  and vice versa, leading to a curved surface in the  $(F_t, V_t)$  space. The linear model, which only uses  $F_t$  and  $V_t$  as regressors, is misspecified both with respect to the quadratic and interaction terms.

## 2.3 Quadratic, Interaction, and Regime Switching

The third scenario augments the previous structure with a simple volatility-based regime switch. Define the regime indicator

$$H_t = \mathbb{I}\{V_t > 1\},$$

which equals one in periods of high volatility and zero otherwise. The return equation becomes

$$R_t = \theta_0 + \theta_1 F_t + \theta_2 V_t + \theta_3 F_t^2 + \theta_4 F_t V_t + \theta_5 H_t F_t + \varepsilon_t. \quad (3)$$

The coefficient on  $F_t$  is thus regime-dependent, implying different linear sensitivities to the factor in low-volatility and high-volatility states. This specification combines smooth nonlinearities with discrete regime changes.

## 3 Models and Estimation

### 3.1 Linear Regression

For each simulated dataset, a standard linear regression model is estimated of the form

$$R_t = \beta_0 + \beta_1 F_t + \beta_2 V_t + u_t, \quad (4)$$

where  $u_t$  denotes the regression residual. The model is intentionally kept linear in  $F_t$  and  $V_t$  to mimic a typical linear factor approach that ignores higher-order terms and interactions. Parameters  $(\beta_0, \beta_1, \beta_2)$  are estimated by ordinary least squares.

### 3.2 Kernel Ridge Regression with RBF Kernel

Kernel ridge regression (KRR) is used as the nonlinear benchmark. Let  $X_t = (F_t, V_t)^\top$ , and let  $K \in \mathbb{R}^{n \times n}$  denote the Gram matrix with entries

$$K_{ij} = k(X_i, X_j) = \exp(-\gamma \|X_i - X_j\|^2), \quad (5)$$

corresponding to the radial basis function (RBF) kernel with scale parameter  $\gamma > 0$ . Given the vector of returns  $y = (R_1, \dots, R_n)^\top$ , KRR solves

$$\alpha = (K + \lambda I_n)^{-1} y, \quad (6)$$

where  $\lambda > 0$  is a regularization parameter and  $I_n$  is the  $n \times n$  identity matrix. The fitted values are then

$$\hat{y} = K\alpha. \quad (7)$$

The same kernel hyperparameters  $(\gamma, \lambda)$  are used across all three cases to isolate the impact of data complexity rather than tuning choices.

## 4 Performance Metrics

Model performance is evaluated using the following measures, computed in-sample for each case and each method. Each metric is accompanied by a brief explanation of what aspect of predictive quality it captures.

- Coefficient of determination:

$$R^2 = 1 - \frac{\sum_t (R_t - \hat{R}_t)^2}{\sum_t (R_t - \bar{R})^2}.$$

This statistic measures the proportion of total variance in returns that is explained by the model; higher values indicate greater explanatory power.

- Root mean squared error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_t (R_t - \hat{R}_t)^2}.$$

RMSE quantifies the average magnitude of prediction errors, with larger errors penalized more heavily due to the square; lower values indicate better accuracy.

- Mean absolute error (MAE):

$$\text{MAE} = \frac{1}{n} \sum_t |R_t - \hat{R}_t|.$$

MAE represents the typical size of prediction errors without overweighting large deviations, providing a robust measure of average error magnitude.

- Bias (mean error):

$$\text{Bias} = \frac{1}{n} \sum_t (\hat{R}_t - R_t).$$

The bias captures systematic over- or underestimation by the model; values near zero indicate the absence of directional distortion.

- Correlation between true and predicted returns:

$$\rho = \text{corr}(R_t, \hat{R}_t).$$

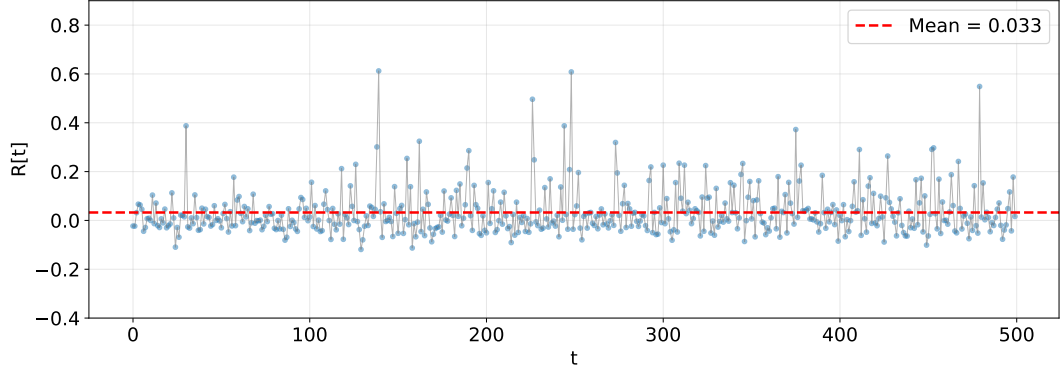
This metric assesses how well the model preserves the relative ranking and co-movements of returns; higher correlation implies more faithful reproduction of the underlying structure.

## 5 Results

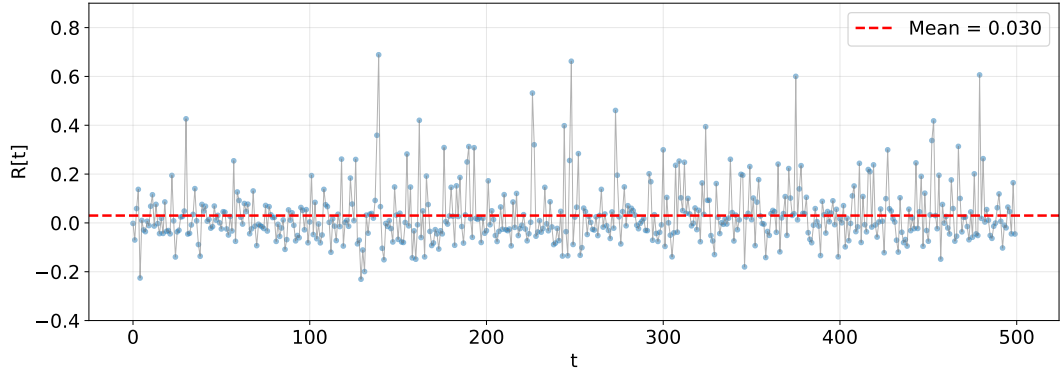
### 5.1 Time-Series Behaviour of Simulated Returns

Figure 1 displays the simulated return paths for the three cases. All series exhibit similar unconditional mean levels, but the dispersion of returns increases from the quadratic-

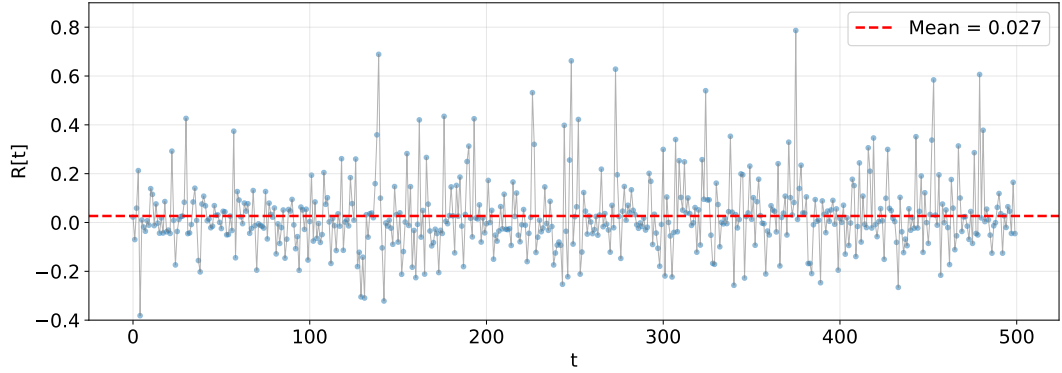
only case to the interaction and switching specifications, reflecting the added nonlinear components.<sup>2</sup>



(a) Quadratic



(b) Interaction



(c) Switching

Figure 1: Simulated returns  $R_t$  for the three data-generating processes. The red dashed line denotes the sample mean in each panel.

## 5.2 Linear Regression

Figure 2 shows the relationship between true and predicted returns for the linear regression benchmark across the three data-generating processes.

<sup>2</sup>The return distributions could in principle be further differentiated by increasing the scaling of the coefficients associated with the nonlinear terms (i.e.,  $\theta_3$ ,  $\theta_4$ , and  $\theta_5$ ).

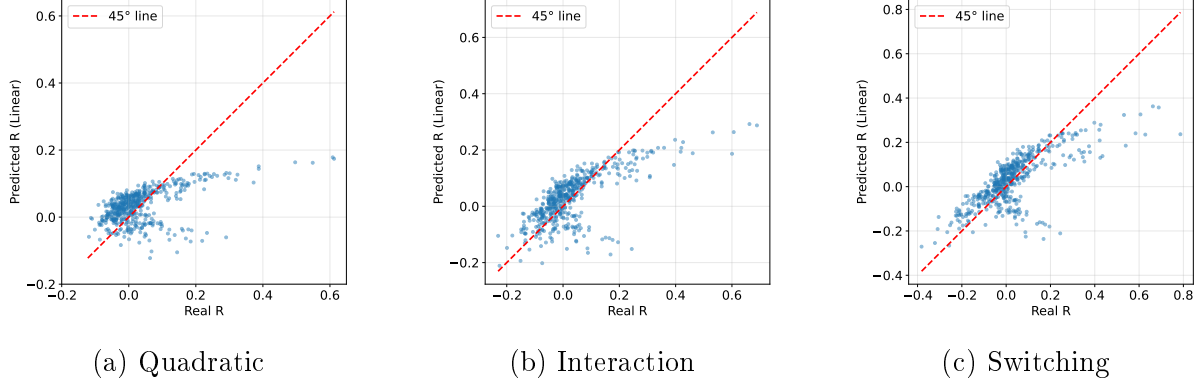


Figure 2: Real vs. predicted returns for the linear regression model. The red dashed line indicates the 45° line corresponding to perfect predictions.

In all cases, the linear model is unable to capture the nonlinear structure embedded in the simulated returns, producing systematic deviations from the 45° line. Under the quadratic specification, the scatter exhibits a characteristic hyperbolic pattern, reflecting the curvature induced by the  $F_t^2$  term, which a linear model cannot approximate. In the interaction case, the dispersion becomes asymmetric and more diffuse, as the joint effect of  $F_t V_t$  alters the shape of the conditional mean in a way that is incompatible with linearity. Finally, in the switching specification, two distinct hyperbolic clusters appear, corresponding to low- and high-volatility regimes, further emphasizing the inability of a single linear predictor to accommodate regime-dependent nonlinear behavior.

### 5.3 RBF Gram Matrix

Before analyzing the kernel regression results, it is useful to inspect the structure of the RBF Gram matrix, shown in Figure 3. The Gram matrix encodes pairwise similarity between observations and determines how information is propagated in kernel ridge regression. Its block patterns reflect the geometry of the input space  $(F_t, V_t)$ .

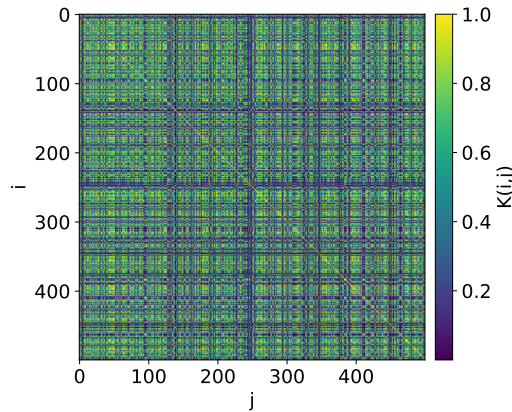


Figure 3: RBF Gram matrix  $K$  computed from the input features  $(F_t, V_t)$ . Brighter values correspond to higher similarity between observations.



The figure reveals a dense and smoothly varying similarity structure, with no clear clustering among observations. The gradual transitions in color intensity indicate that the input variables generate a continuum of local neighborhoods rather than sharply separated groups, which is consistent with the continuous nature of  $(F_t, V_t)$ . The diagonal band of high similarity reflects the fact that each observation is most similar to itself and to nearby points in feature space, a typical signature of RBF kernels.<sup>3</sup>

## 5.4 Kernel Ridge Regression

Figure 4 reports the real vs. predicted returns obtained from kernel ridge regression with an RBF kernel.

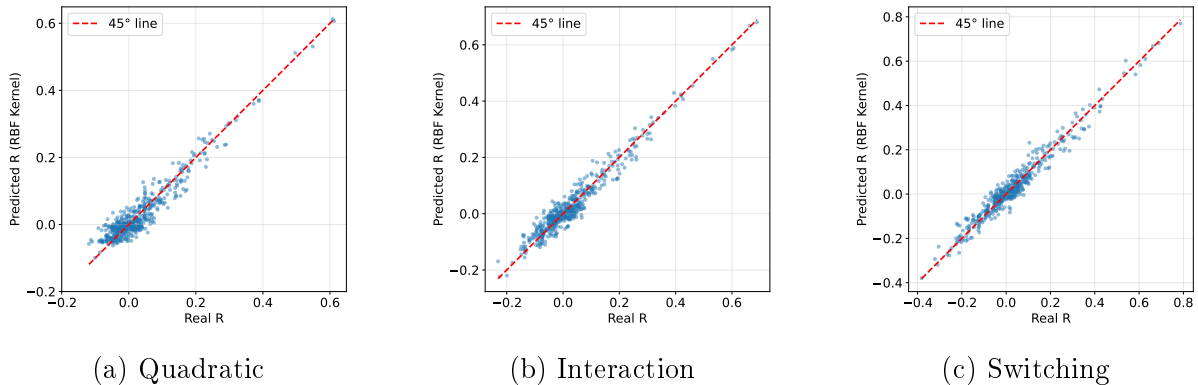


Figure 4: Real vs. predicted returns for kernel ridge regression with an RBF kernel. Points closely follow the 45° line in all cases, indicating a high-quality nonlinear fit.

Unlike the linear model, the kernel estimator successfully captures the nonlinear structure of the three data-generating processes. The points lie tightly along the 45° line even in the interaction and switching cases, indicating near-perfect in-sample fit. This reflects the ability of the RBF kernel to represent flexible nonlinear mappings without explicitly specifying polynomial or interaction terms.

## 5.5 Summary of Performance Metrics

Tables 1–3 report the numerical performance metrics for the two models across the three simulation scenarios. For each case, the best-performing value (between linear regression and kernel ridge regression) is highlighted in bold.

Table 1: Quadratic

Model	$R^2$	RMSE	MAE	Bias	$\text{Corr}(y, \hat{y})$
Linear	0.261338	0.082031	0.059006	0.000000	0.511212
Kernel	<b>0.906657</b>	<b>0.029161</b>	<b>0.023096</b>	-0.000018	<b>0.952189</b>

<sup>3</sup>The Gram matrix is identical across the three simulation scenarios because all cases share the same input features  $(F_t, V_t)$ . The differences between the data-generating processes arise only in the construction of the target variable  $R_t$ , which does not affect the kernel similarity structure.

Table 2: Interaction

Model	$R^2$	RMSE	MAE	Bias	$\text{Corr}(y, \hat{y})$
Linear	0.500116	0.085491	0.060026	0.000000	0.707189
Kernel	<b>0.941819</b>	<b>0.029166</b>	<b>0.023116</b>	-0.000017	<b>0.970475</b>

Table 3: Switching

Model	$R^2$	RMSE	MAE	Bias	$\text{Corr}(y, \hat{y})$
Linear	0.545199	0.099767	0.070334	0.000000	0.738376
Kernel	<b>0.952246</b>	<b>0.032328</b>	<b>0.025232</b>	-0.000016	<b>0.975834</b>

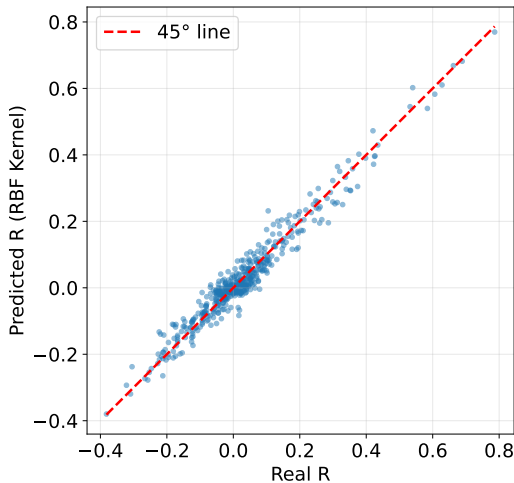
Across all scenarios, kernel ridge regression with an RBF kernel consistently outperforms linear regression, particularly in the interaction and switching cases where nonlinear effects become more pronounced. The linear model provides a reasonable approximation only in settings with mild nonlinearity, while the kernel method remains accurate even under strong curvature and regime-dependent behavior.

## 6 High-Noise Environment

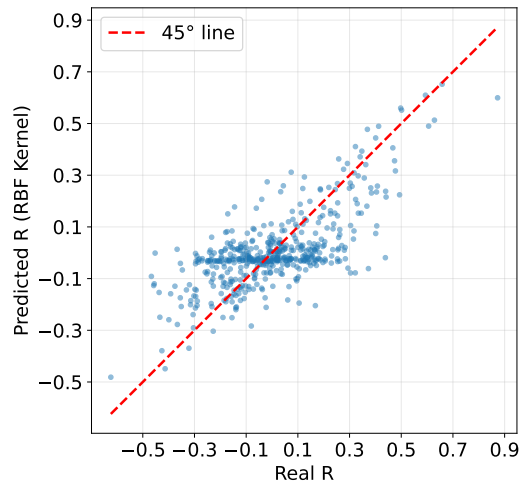
To assess the sensitivity of kernel methods to purely random variation, an additional scenario is considered in which the nonlinear structure of the switching specification—as defined in Eq. (3)—is kept unchanged, but the noise variance is substantially increased.

$$\varepsilon_t \sim \mathcal{N}(0, \sigma^2), \quad \sigma : 0.03 \rightarrow 0.15.$$

Figure 5 compares the fitted values for kernel ridge regression in the low-noise switching case and in the high-noise environment.



(a) Switching



(b) High-noise

Figure 5: Kernel ridge regression: real vs. predicted returns under the switching specification (left) and under an increased noise level (right).

Although the kernel model still captures a positive relationship in the high-noise scenario, the cloud of points becomes substantially more dispersed around the  $45^\circ$  line, signalling a sharp reduction in predictive precision.

Table 4: Kernel ridge regression: switching vs. high-noise.

Case	Model	$R^2$	RMSE	MAE	Bias	$\text{Corr}(y, \hat{y})$
Switching	Kernel	<b>0.952246</b>	<b>0.032328</b>	<b>0.025232</b>	-0.000016	<b>0.975834</b>
Noise	Kernel	0.500826	0.145495	0.115271	-0.000040	0.707750

As Table 4 shows, when moving from the switching case to the high-noise environment,  $R^2$  drops dramatically, RMSE and MAE more than quadruple, and the correlation between true and predicted values declines as well. These results highlight that kernel methods are highly effective at capturing nonlinear factor relationships, but they cannot compensate for a substantial increase in idiosyncratic randomness when the signal-to-noise ratio becomes low.

## 7 Conclusion

The simulation results confirm that linear factor models are not robust to nonlinearities and regime changes of the type considered here. Even in the relatively simple quadratic-only case, the omission of  $F_t^2$  leads to a noticeable decline in  $R^2$  and an increase in prediction errors, while the addition of interaction effects and regime-dependent coefficients further deteriorates linear performance.

Kernel ridge regression with an RBF kernel, by contrast, is able to approximate the true nonlinear mapping from  $(F_t, V_t)$  to  $R_t$  without explicitly specifying higher-order terms. The Gram matrix reveals a rich similarity structure in the input space, and the resulting kernel predictions closely track the simulated returns across all scenarios. The consistently high correlations and low RMSE and MAE values illustrate the ability of kernel methods to capture complex factor dynamics in a flexible yet regularized manner.

The additional high-noise experiment further clarifies the limitations of kernel methods: although they continue to recover the underlying nonlinear relationship, their predictive accuracy deteriorates markedly when the variance of the idiosyncratic component becomes large. This highlights that kernel models, while flexible, remain ultimately constrained by the signal-to-noise ratio of the data.

Overall, the findings show that kernel-based approaches provide substantial improvements over linear models when returns depend nonlinearly on underlying factors or exhibit regime-switching behaviour. While linear models remain useful as simple benchmarks, the evidence suggests that kernel methods are better suited for modeling nonlinear factor effects in asset returns. Future work could extend the analysis to out-of-sample evaluation, alternative kernels, and higher-dimensional factor structures, as well as applications to empirical financial data.