Investment Performance of Complex Factors

Kaihe Liu

January 24, 2026

Abstract

Current APT asset pricing theory posits that few pricing factors can adequately explain asset pricing. However, empirical evidence demonstrates that traditional APT pricing models exhibit weak out-of-sample pricing performances. Recent advancements in machine learning propose a novel approach for asset pricing: large models incorporating extensive factor characteristics leverage their high-dimensional approximation advantages to achieve accurate pricing with stochastic discount factor. This paper investigates the performance of complex factors with fixed information set in China's A-share market. Utilizing the A-share dataset, we construct complex factor portfolios to evaluate out-of sample portfolio Sharpe ratios. We employ Random Fourier factor (RFF) for feature augmentation of initial features. Under fixed input, we examine how SDF portfolio performance evolves with complexity and regularization. Our study finds evidence of virtue of complexity as well as benefit of regularization.

1. Introduction

The introduction of Arbitrage Pricing Theory (APT) marked a significant breakthrough in the field of asset pricing. According to this theory, a small number of effective factors can suffice for pricing target assets. These orthogonal factors can explain a major portion of cross-sectional asset returns and are incorporated into models as key indicators for predicting future asset prices. Classical factor models, such as the Fama-French six-factor model (FF6) and the CH-3 model tailored for the Chinese market, typically include only a few uncorrelated factors. Highly complex factor models are often dismissed as redundant and lacking interpretability.

With advancements in modern machine learning technologies, researchers and investors have increasingly adopted large-scale statistical models for data fitting and forecasting—a trend that has yielded substantial results in various domains. Machine learning has been introduced into the domain of asset pricing, prompting a reevaluation of traditional factor models. Under a fixed information set, complex models can better capture the underlying data-generating process. Capital markets involve intricate interactions among a multitude of variables, governed by highly complex internal mechanisms. Machine learning models such as artificial neural networks (ANNs), with their vast parameter space and high complexity, are particularly adept at fitting such intricate relationships.

When selecting models, it is commonly believed that a trade-off exists between bias and variance. Compared with simpler models, complex models often achieve better test-set performance due to their superior fitting capacity, reflecting lower bias. However, these models may also perform poorly on out-of-sample data, manifesting higher variance—a phenomenon referred to as overfitting. Historical financial data is noisy, making complex models inherently susceptible to overfitting. In asset pricing, whether high-complexity models can yield strong out-of-sample investment returns, and the role of regularization in this context, are critical issues for model selection and investment decision-making in practice.

This study seeks to demonstrate that increasing model complexity does not necessarily degrade test-set performance. We employ Random Fourier Features (RFF) to map original features into higher dimensions, thereby increasing model complexity and fitting power. To mitigate overfitting, we incorporate regularization terms and evaluate how models of varying complexity and regularization levels perform in out-of-sample investment scenarios in China's A-share market. Specifically, with a fixed set of initial stock characteristics, we construct complex factor-based out-of-sample investment portfolios and examine their Sharpe ratios to assess the practical viability of complex factor models in trading and their applicability to the Chinese equity market.

Existing literature has extensively discussed the pricing mechanisms of complex factor models. KMZ have analyzed the return prediction mechanisms of machine learning models from a theoretical standpoint. Chinico et al. (2019) demonstrate the success of machine learning in cross-sectional return forecasting, which this study extends to the Chinese financial market. Theoretical investigations by Martin and Nagel (2021), Da et al. (2022), and Fan et al. (2022) delve into the "learning limits" of high-dimensional asset pricing models. Our research, from an empirical perspective, helps explain the emergence of new pricing factors overlooked by traditional models, contributing to the broader academic discourse on the "factor zoo" as discussed by Harvey et al. (2016), McLean and Pontiff (2016), Hou et al. (2020), Feng et al. (2020), Hensen et al. (2023), and Chen and Zimmermann (2021).

Hastie et al. (2022) investigate the favorable out-of-sample investment performance of complex models, employing various parameterized approaches to analyze pricing errors across different complexities. They also provide theoretical results on the asymptotic behavior of complex models under small sample conditions. We apply their analytical framework to the Chinese A-share market using RFF-based methods proposed by Rahimi and Recht (2007) to control model complexity. Kozak et al. (2020) argue that factor returns can be effectively explained by a few principal components, suggesting that a successful asset pricing model might not require many factors. We test this hypothesis by comparing portfolio performance before and after dimensionality reduction. This paper contributes to the literature by addressing the research gap in complex factors in China's A-share market. Through a comprehensive evaluation of investment performance, we provide empirical support for applying complex machine learning models to China's equity markets.

The remainder of this paper is organized as follows. Section 2 outlines the theoretical framework of complex factor models and Random Fourier Features, as well as our data selection, parameter estimation methodology, and performance evaluation criteria. Section 3 presents empirical findings, including analyses of model complexity and regularization, single-feature tests, PLS-based dimensionality reduction, and alternative nonlinear transformations. Section 4 concludes.

2. Complex Factor Models

2.1. Factor Models and Stochastic Discount Factor

The origin of factor models can be traced to the Capital Asset Pricing Model (CAPM) proposed by Sharpe (1964). CAPM posits that, under market equilibrium, asset excess returns are linearly related to the excess return of the market portfolio, implying that asset returns arise from systematic risk. In essence, the market factor can explain asset returns. As asset pricing theory evolved, the Arbitrage Pricing Theory (APT) developed by Ross (1976) offered a formal foundation for factor models. APT suggests that the cross-sectional variation in asset returns can be captured by a linear combination of multiple uncorrelated factors:

$$R_i = \alpha_i + \sum_{j=1}^{N} \beta_j^i F_j + \varepsilon_i, \tag{1}$$

where F_j denotes the latent factors driving asset returns, and the idiosyncratic error ε_i satisfies the following conditions:

$$\mathbb{E}[\varepsilon_i] = 0, \tag{2}$$

$$\mathrm{Cov}(\varepsilon_i, F_j) = 0, \tag{3}$$

$$\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \text{for } i \neq j. \tag{4}$$

Based on APT, factor data can be utilized to construct out-of-sample Markowitz mean-variance optimized portfolios. The stochastic discount factor (SDF), which prices risky assets, can be expressed as a tradable portfolio:

$$M_{t+1} = 1 - w(Z_t)' R_{t+1}, \tag{5}$$

where R_{t+1} denotes the excess returns of risky assets and $w(Z_t)'$ represents their corresponding weights in the SDF. These weights are obtained through mean-variance optimization to construct the Markowitz optimal portfolio. This framework reveals that classical factor models serve as a linear approximation to the SDF.

Prominent examples of classical models include the Fama-French three-factor model (1993), which incorporates the market portfolio, firm size, and book-to-market ratio as explanatory factors. Over time, many variants have emerged, including the CH-3 model tailored for the Chinese market.

Inspired by recent developments in machine learning, we introduce the concept of *complex factor models* to describe models that incorporate a large number of conditioning variables and nonlinear functional forms. A key aspect of these models is that they do not assume a pre-specified functional form but instead use flexible, nonparametric approximations to the unknown SDF. These models can be represented as:

$$w(Z_t) = \sum_{p=1}^{P} \theta_p S_p(Z_t) = S_t \theta, \tag{6}$$

where Z_t is an $N_t \times D$ matrix representing D features for N_t assets, $S_p(Z_t)$ denotes a nonlinear basis function, and θ_p is the corresponding scalar coefficient optimized via mean-variance portfolio selection. As p increases, both the number of parameters and model flexibility grow due to the incorporation of nonlinearities.

In this formulation, the SDF portfolio can be rewritten as a linear combination of factor-mimicking portfolios:

$$F_{t+1} = S_t' R_{t+1}, \tag{7}$$

where each F_{t+1} represents a characteristic-based portfolio weighted by a nonlinear transformation $S_p(Z_t)$ of asset characteristics.

2.2. Random Fourier Features

Random Fourier Features (RFF), introduced by Rahimi (2007), constitute a specific nonlinear transformation technique for feature generation. RFF employs trigonometric functions to create a desired number of new features through nonlinear combinations of original inputs. Although initially proposed for regression prediction tasks, RFF can be seamlessly embedded within the stochastic discount factor (SDF) framework. The transformed features are generated using sine and cosine functions:

$$[S_{2p-1,t}, S_{2p,t}] \in \mathbb{R}^{N_t \times 2} = [\sin(\gamma Z_t \omega_p), \cos(\gamma Z_t \omega_p)]^\top, \tag{8}$$

where ω_p is a $D \times K$ random matrix sampled from a normal distribution, used to construct K random linear combinations of the D original features (i.e., the RFFs). The parameter γ controls the bandwidth of the Gaussian kernel used in generating the random Fourier features. Specifically, for each ω_p, it is randomly selected from a scalar range between 0 and 1. This configuration allows control over the level of nonlinearity introduced during feature generation.

The random linear combinations of the original features are passed through trigonometric activation functions, thereby inducing nonlinear transformations. Notably, this process does not operate on individual features in isolation; instead, it captures general multi-dimensional interactions among all features for a given stock i.

Our decision to employ RFF is motivated by several considerations. First, the RFF structure resembles a neural network layer in machine learning, wherein new features are constructed via weighted summation and nonlinear activation. This enables us to establish a connection between our study and practical machine learning-based investment scenarios. Second, compared to neural networks, RFF benefits from enhanced computational efficiency due to its inherent randomness. It facilitates nonparametric approximations of the SDF and enables closed-form estimation procedures. Lastly, RFF allows flexible feature expansion and reduction to any desired complexity level while keeping the initial information set fixed. This isolates the influence of input data, permitting us to evaluate model performance based solely on complexity.

2.3. Parameter Estimation

As previously noted, the vector θ represents the Markowitz mean-variance optimal portfolio weights for the constructed factors. It is derived by solving the following optimization problem:

$$\theta = \arg\max_{\theta} \left\{ \hat{\mathbb{E}}[\theta' F_t] - \frac{1}{2}\hat{\mathbb{E}}[(\theta' F_t)^2] \right\} = \hat{\mathbb{E}}[F_t F_t']^{-1}\hat{\mathbb{E}}[F_t]. \tag{9}$$

There are multiple approaches to estimate $\hat{\mathbb{E}}$. In this study, we adopt the sample mean from training data, as it is a commonly used and computationally efficient method in mean-variance optimization. Accordingly, the estimated SDF and the associated portfolio returns are represented as:

$$\hat{M}_t = 1 - \hat{R}_t^M, \quad \text{where} \quad \hat{R}_t^M = \hat{\theta}' F_t. \tag{10}$$

Given the presence of noise in historical financial data, complex models are prone to overfitting—a situation where models exhibit superior in-sample performance but significantly deteriorate out-of-sample. This issue is especially prevalent in finance due to the inherently low signal-to-noise ratio in stock markets, which poses substantial challenges for the practical implementation of asset pricing models. Consequently, regularization becomes essential.

To explore the effects of regularization alongside model complexity, we introduce a ridge penalty term in the optimization objective:

$$\theta(z) = \arg\max_{\theta} \left\{ \hat{\mathbb{E}}[\theta' F_t] - \frac{1}{2}\widehat{\mathrm{Var}}[\theta' F_t] - \lambda \theta'\theta \right\} = (\lambda I + \hat{\mathbb{E}}[F_t F_t'])^{-1}\hat{\mathbb{E}}[F_t], \tag{11}$$

where λ denotes the ridge regularization parameter. We employ the ℓ_2-norm ridge penalty because it effectively mitigates issues such as rank deficiency in $\hat{\mathbb{E}}[F_t F_t']^{-1}$ when the model size becomes large. Additionally, it addresses multicollinearity and reduces overfitting by shrinking parameter magnitudes, thereby enhancing portfolio performance.

2.4. Data and Evaluation

We employ daily trading data from China's A-share market covering the period from 2017 to 2024, along with 58 stock characteristic variables sourced from the JoinQuant database. Stock returns are computed based on daily closing prices. These 58 features are categorized into six groups—valuation, price-volume, profitability, growth, and dividend—ensuring a comprehensive representation of characteristic types. We restrict our investment universe to the CSI 300 Index constituents (the 300 largest market capitalization stocks) due to their lower noise and greater representativeness.

Some features are derived from quarterly or annual financial reports. To facilitate processing, these are converted into structured daily frequency data. All selected features have missing value rates below 5%, and observations with missing entries are excluded from the analysis.

Further, we standardize all features. Since several features are already in cross-sectional rank format, we avoid additional rank normalization. To address the issue of differing magnitudes across features, we apply z-score normalization to all variables, ensuring consistent mean and variance. The final input is an $N_t \times D$ matrix containing the stock features for each trading day. These features are subsequently passed through the Random Fourier Feature (RFF) transformation to generate the desired number of complex factors.

Model training follows a rolling window approach: for each trading day that satisfies the training sample requirement, we fit the model using data from the preceding year (252 trading days). Using the generated factor data and expected returns, we perform Markowitz optimization to construct out-of-sample Stochastic Discount Factor (SDF) portfolios. Each day's SDF portfolio is trained using the previous year's data.

To evaluate investment performance, we compute the annualized Sharpe ratio of the out-of-sample portfolio returns, given that our SDF portfolios are constructed using mean-variance optimization. The Sharpe ratio is defined as:

$$\hat{SR} = \frac{\hat{\mathbb{E}}_{OS}[\hat{R}_t^M]}{\hat{\sigma}_{OS}[\hat{R}_t^M]}, \tag{12}$$

where $\hat{\mathbb{E}}_{OS}$ and $\hat{\sigma}_{OS}$ denote the annualized sample mean and standard deviation of out-of-sample portfolio returns, respectively. A higher Sharpe ratio indicates superior portfolio performance by achieving higher returns with lower risk.

3. Empirical Results

3.1. Virtue of Complexity and Regularization

The primary objective of this section is to examine the effects of factor complexity and regularization on portfolio performance. Specifically, we vary the number of factors P and the ridge regularization parameter λ to plot the out-of-sample Sharpe ratios of the Stochastic Discount Factor (SDF) across different settings.

We vary the number of Random Fourier Features (RFFs) from 1 to 200 in increments of 2. The ridge parameter λ is set to five distinct values: 1×10^{-5}, 0.001, 0.1, 1, and 10, to compare model performance under different degrees of regularization.

Our empirical findings show that when the ridge coefficient is small, the Sharpe ratio experiences a sharp initial rise, followed by a rapid decline and another subsequent increase, eventually stabilizing. This pattern reveals what we term a "virtue of complexity" However, when we increase and fix the ridge coefficient at a higher level, a different pattern emerges: even with lower model complexity, we observe significant performance improvements. While there may be a brief initial rise, the Sharpe ratio tends to stabilize at a level that exceeds the corresponding level observed under low-regularization scenarios.

Notably, this result diverges from the conventional bias-variance tradeoff theory. Given a fixed initial information set, the out-of-sample performance of complex factor models does not deteriorate compared to those with fewer factors.

Our findings provide valuable insights for model design in investment contexts. When constructing a complex factor model, applying regularization is essential. If employing a mild ridge penalty, it is advisable to include more factors, thereby leveraging complex structures to enhance the model's fitting capability. On the other hand, setting a higher ridge penalty can compensate for model deficiencies and improve portfolio performance. Under high regularization, increasing the number of factors appears to yield limited additional benefits to the SDF's performance.

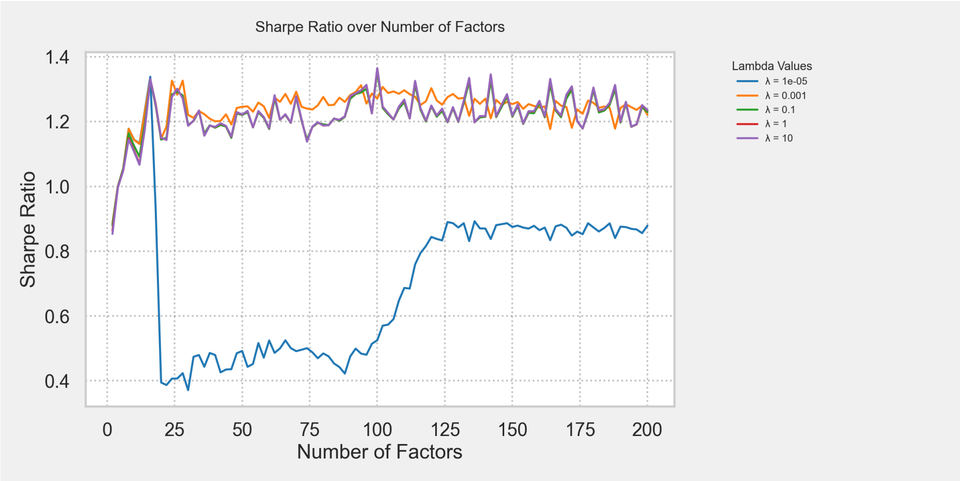

Figure 1: Sharpe Ratio vs. Model Complexity and Regularization

3.2. Single Feature Analysis

In the preceding section, we fixed the initial information set to include 58 features. However, holding other variables constant, variations in the initial information set can lead to significantly different outcomes. This is attributable to the intricate multi-directional interactions among features in the information set, where each individual feature may offer a distinct perspective in explaining asset returns. While such analysis is inherently extensive, it is vital to the comprehensiveness of the study. In this section, we focus on initial information sets comprising single stock features and evaluate portfolio performance after dimensional expansion into numerous Random Fourier Features (RFFs).

We selected seven well-performing and commonly used features in the Chinese A-share market from five different categories and constructed models using the previously established methodology.

Table 1: Stock Features Used in Single-Feature Analysis

Feature Category	**Feature Name**	**Variable Code**
Valuation	Inverse PE Ratio (TTM), Ranked Inverse PE (TTM)	EP_TTM, REP_TTM
Volume-Price	40-Day Return, 20-Day Average Turnover Rate	RT_2M, TO_1M
Profitability	Return on Equity (ROE)	ROE
Growth	Net Profit Growth Rate	E_Growth2
Dividend	1-Year Dividend Yield	DivR1

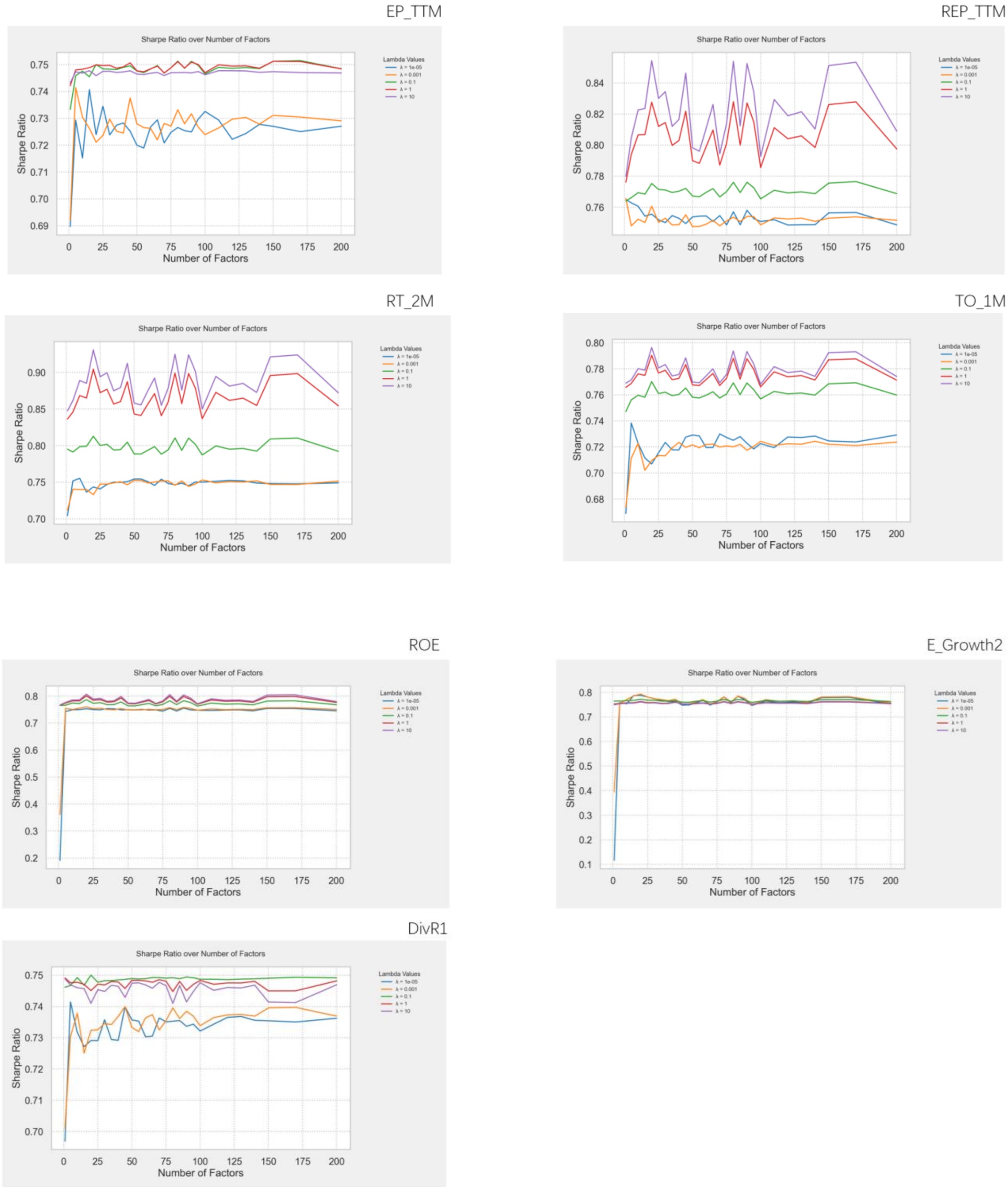

Figure 2: Single Feature Sharpe Ratios by Complexity and Regularization

Our empirical results indicate that different features yield varying performances. Importantly, complex factor models built on single-feature information sets do not exhibit deteriorated out-of-sample performance. As expected, due to the absence of broader information, the Sharpe ratio drops to approximately 0.8 (lower than the previous 1.3). Furthermore, compared to models using all 58 features, Sharpe ratio curves derived from single-feature inputs exhibit greater volatility and lack a clear relationship with model complexity. This suggests that in large-scale information sets with intricate interactions, the "complexity dividend" becomes more pronounced. Complex factor models, given their large parameter space, can capture deep and nuanced patterns. However, when constrained to a single feature, such cross-feature interactions vanish, and the underlying structure becomes simpler, diminishing the advantage of increased complexity.

While the impact of complexity on portfolio performance remains ambiguous in this context, the effect of regularization is considerably more significant. For each stock feature, an increase in the ridge parameter consistently leads to a higher Sharpe ratio, exhibiting a degree of monotonicity within the examined range. This highlights the potent role of regularization in enhancing model performance—even when only a single feature is available. Accordingly, we interpret our earlier findings as follows: within a large-scale information set consisting of 58 features, regularization proves effective since the averaged characteristics yield a consistent positive contribution to performance. When the majority of features exhibit a "regularization dividend," the overall model benefits substantially from the inclusion of a regularization term.

3.3. Model Compression

In this section, we validate the existence of virtue of complexity through model compression. If large-scale factor models containing numerous factors indeed exhibit a complexity dividend, then, all else being equal, reducing the dimensionality of the model's random Fourier factor inputs should lead to a decline in portfolio Sharpe ratios. Keeping the initial information set unchanged, we apply Partial Least Squares (PLS) to carry out the dimensionality reduction.

PLS is a common dimensionality reduction algorithm that constructs principal components by maximizing the covariance between independent variables (X) and dependent variables (Y). For the k-th component, PLS iteratively computes the weight vector $\boldsymbol{w}_k$ through the following optimization problem:

$$\boldsymbol{w}_k = \arg\max_{\|\boldsymbol{w}\|=1} \mathrm{Cov}(\boldsymbol{X}_{k-1}\boldsymbol{w}, \boldsymbol{Y}) \tag{13}$$

where $\boldsymbol{X}_{k-1}$ denotes the residual matrix of $\boldsymbol{X}$ after subtracting the variance explained by previous components. The final regression model is:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}, \quad \boldsymbol{B} = \boldsymbol{W}(\boldsymbol{P}^T\boldsymbol{W})^{-1}\boldsymbol{C}^T \tag{14}$$

where $\boldsymbol{W}$ is the matrix of weights, $\boldsymbol{P}$ is the loading matrix of $\boldsymbol{X}$, and $\boldsymbol{C}$ is the loading matrix of $\boldsymbol{Y}$.

PLS is a supervised regression method suitable for dimensionality reduction. As a supervised learning algorithm, PLS selects K components that are most correlated with the target variable. These K reduced features are orthogonal, meaning that correlations among the original features are eliminated, while also preserving their relationship with the response variable.

The empirical results show that applying PLS to models with different levels of complexity and regularization causes a modest decline in performance. This finding is consistent across both $K = 2$ and $K = 10$ compressed components, as well as across models with various ridge parameters. Our findings indicate that when large-scale factor models are compressed into smaller ones, their virtue of complexity diminishes. This supports the claim that virtue of complexity exists in large factor models.

Even though there is some performance degradation after dimensionality reduction, it is not substantial. This suggests that compressed complex factor models still retain a significant degree of explanatory power. In practice, it is feasible to construct portfolios with comparable performance to complex models using only a small number of factors, which can be derived from dimensionality reduction over a large set of features.

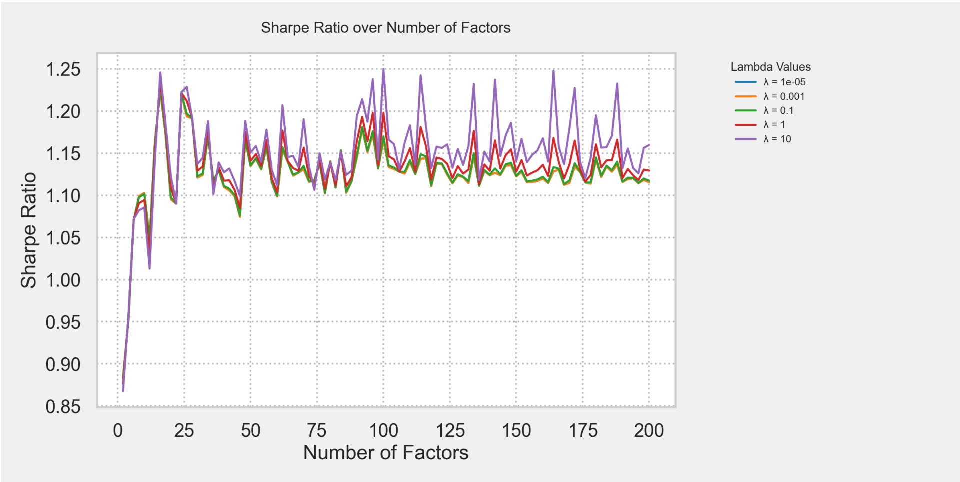

Figure 3: PLS Compressed Model Performance ($K = 2$)

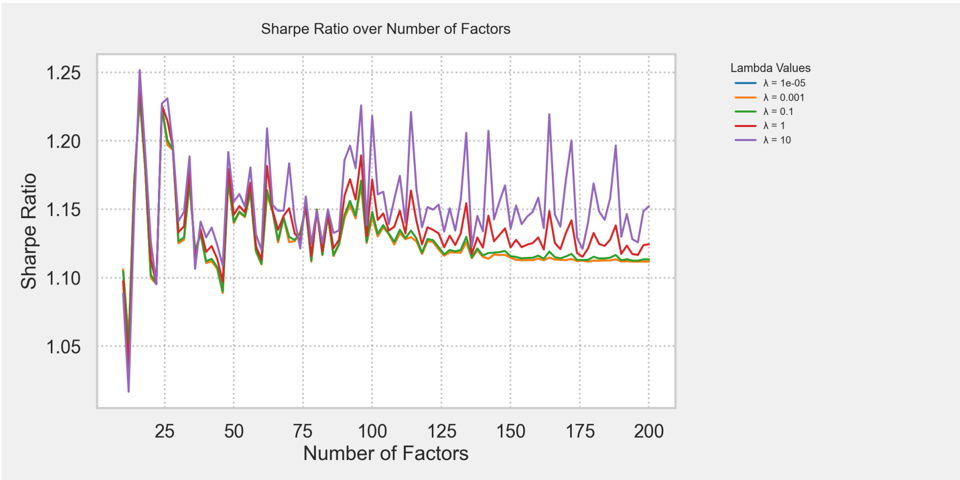

Figure 4: PLS Compressed Model Performance ($K = 10$)

3.4. Alternative Nonlinearities

In the previous section, we introduced nonlinearity via trigonometric functions through the Random Fourier Features (RFF) method proposed by Rahimi (2007), increasing factor complexity. However, in real-world scenarios, the mechanisms by which features exert nonlinear effects may vary significantly and are not necessarily restricted to sine and cosine transformations. Modern neural network models—widely used in the investment domain—employ alternative activation functions such as sigmoid, tanh, and ReLU, and have demonstrated strong performance across numerous financial applications.

Motivated by this, we replace the original RFF-based nonlinearity with commonly used sigmoid and tanh functions. We generate random linear combinations of the original features by sampling weights from a standard normal distribution and then apply these nonlinear transformations to obtain factors for portfolio construction.

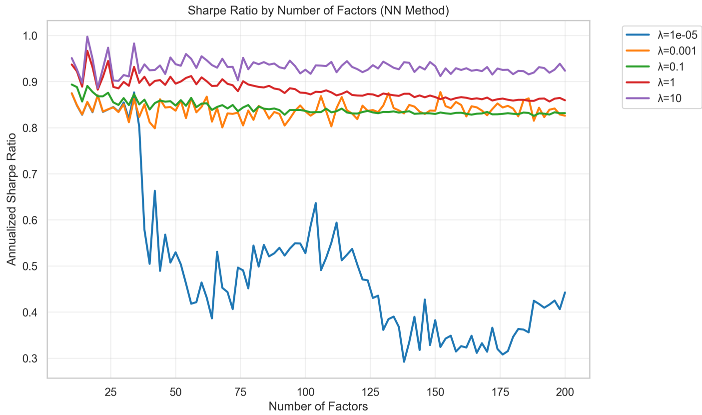

Figure 5: Sharpe Ratios Using Sigmoid Nonlinearity

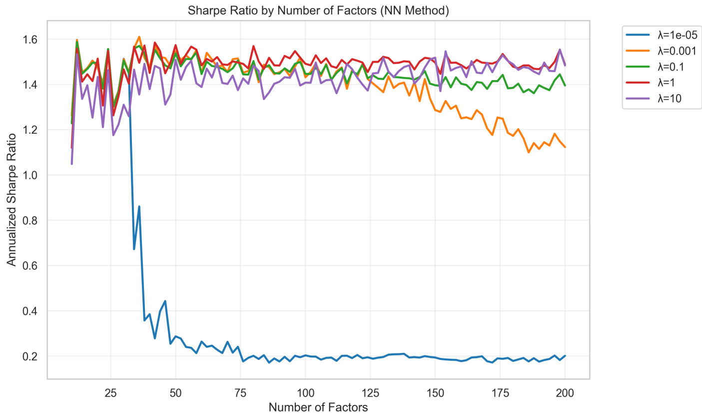

Figure 6: Sharpe Ratios Using Tanh Nonlinearity

Our results indicate that, within a certain range of regularization parameters, using sigmoid or tanh functions does not degrade the out-of-sample investment performance of the complex factor models. This finding enhances the robustness of our research framework, suggesting that the observed performance patterns are not confined to specific RFF formulations but rather extend across different nonlinear modeling paradigms. It also helps explain why highly complex neural network models can still achieve strong investment performance despite their intricate structures.

4. Conclusion

This paper analyzes the portfolio performance of large-scale factor models under varying levels of complexity and regularization. Using Random Fourier Features (RFF), we construct Stochastic Discount Factor (SDF) portfolios by altering model complexity and regularization strength, while holding the initial information set fixed. We provide empirical evidence from 2017 to 2024 on the complexity and regularization dividends observed in the CSI 300 constituents of the Chinese A-share market.

Our results demonstrate that, under low ridge penalty parameters, the Sharpe ratio in the A-share market exhibits a double-peak pattern, indicating the presence of a complexity dividend. When the ridge parameter is large, we observe consistently higher Sharpe ratios, reflecting a regularization dividend. Notably, increased complexity does not degrade portfolio performance. When the information set is restricted to a single feature, the relationship between Sharpe ratio and model complexity becomes unstable, though performance does not deteriorate. Meanwhile, regularization consistently enhances portfolio Sharpe ratios.

To further substantiate our findings, we apply the supervised Partial Least Squares (PLS) method to compare performance across different levels of model complexity. The results reveal a moderate decline in Sharpe ratios after dimensionality reduction of RFFs, consistent with our prior conclusions. Finally, we evaluate alternative nonlinear functions beyond RFFs. Using neural network-style activation functions, we observe similar results: increased complexity does not worsen portfolio performance.

Despite the new insights, this study faces certain limitations. First, due to computational constraints, we are unable to extend our results to extremely complex factor models with a higher number of parameters. Second, estimating expected returns using sample averages may deviate significantly from realized returns, potentially distorting the representation of portfolio performance and affecting portfolio construction.

Nevertheless, our findings offer several key takeaways. Compared to the existing literature focused on U.S. equity markets, our study reveals distinct complexity patterns in the Chinese stock market. Through single-factor analysis, we highlight the importance of the input feature set and provide foundational insight into why certain complex factors, derived from limited initial features, display systematic behaviors. By incorporating supervised PLS dimensionality reduction and linking it to investment outcomes, we validate both the presence of a complexity dividend and the practicality of constructing portfolios using a limited number of principal component factors.

References

1. Avramov, D., Cheng, S., & Metzker, L. (2023). Machine Learning vs. Economic Restrictions: Evidence from Stock Return Predictability. *Management Science*, 69(5), 2587–2619.

2. Bartlett, P. L., Long, P. M., Lugosi, G., & Tsigler, A. (2020). Benign overfitting in linear regression. *Proceedings of the National Academy of Sciences*, 117(48), 30063–30070.

3. Chen, A. Y., & Zimmermann, T. (2021). Open source cross-sectional asset pricing. *Critical Finance Review*, Forthcoming.

4. Didisheim, A., Ke, S. B., Kelly, B. T., et al. (2024). APT or "AIPT"? The Surprising Dominance of Large Factor Models. *NBER Working Papers*.

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

6. Hansen, L. P., & Singleton, K. J. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica*, 50(5), 1269–1286.

7. Hastie, T., Montanari, A., Rosset, S., & Tibshirani, R. J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *The Annals of Statistics*, 50(2), 949–986.

8. Jensen, T. I., Kelly, B., & Pedersen, L. H. (2023). Is there a replication crisis in finance? *The Journal of Finance*, 78(5), 2465–2518.

9. Kelly, B. T., Malamud, S., & Zhou, K. (2022). The virtue of complexity everywhere. Available at SSRN.

10. Kelly, B., & Xiu, D. (2023). Financial machine learning. *Foundations and Trends in Finance*, 13(3–4), 205–363.

11. Kozak, S., & Nagel, S. (2023). When do cross-sectional asset pricing factors span the stochastic discount factor? *Working Paper*.

12. Rahimi, A., & Recht, B. (2007). Random Features for Large-Scale Kernel Machines. *NIPS*, Vol. 3, Citeseer, p. 5.