Fama-MacBeth Regression with Asset Pricing Restriction

Yuanqi Yang

Luoshu Investment

Guofu Zhou

Olin Business School, Washington University in St. Louis

Yifeng Zhu

Central University of Finance and Economics*

This version: August 2024

*Yang, yangyuanqi0936@163.com; Zhou (corresponding author), zhou@wustl.edu; Zhu, suxian2000@gmail.com.
We thank Yufeng Han, Jun Tu, Jianfeng Yu, Bohui Zhang, Yingzi Zhu, and seminar participants at Central University of Finance and Economics, Shanghai Jiatong University, Tsinghua University, Washington University in St. Louis, Xi'an Jiaotong-Liverpool University, and Zhejiang University.

Fama-MacBeth Regression with Asset Pricing Restriction

Abstract

In this paper, we propose a modified Fama-MacBeth regression that incorporates both asset pricing restrictions and factor selection into the estimation. The restrictions require the model to explain both the time series and cross-sectional variations while simultaneously selecting factors for sparsity. Solving the estimation via a least angle regression-type algorithm, we find empirically that the new model outperforms existing factor selection methodologies in terms of predicting better the cross-sectional stock returns and earning a greater Sharpe ratio. In addition, the factors selected through our approach are interpretable and outperform classical factors models.

JEL Classification: G11, G12
Keywords: Fama-MacBeth regression, LASSO, characteristics selection, high-dimensional prediction

1 Introduction

The Fama-MacBeth (FM) regression, introduced by Fama and MacBeth (1973), is one of the major econometric methods for selecting factors, estimating their risk premia, and predicting the cross-section of stock returns. However, when analyzing a large number of characteristics, the FM regression often becomes inadequate and subject to over-fitting. To overcome this problem, various penalized methods have been proposed, such as the pooled LASSO in Freyberger, Neuhierl, and Weber (2020) and the cross-sectional elastic net in Han, He, Rapach, and Zhou (2024). While these methods improve the performance, they ignore the economic constraints between the cross section and time series, and their estimates are often difficult to interpret.

In this paper, we propose a modified Fama-MacBeth (FM) regression that incorporates cross-sectional asset pricing restrictions. In this way, the solution seeks to explain both the time series and cross-sectional variations of the asset returns.[1] Moreover, when dealing with a large set of characteristic factors, it is important to impose additional restrictions for factor selection. Hence, we also impose LASSO into the FM regression, creating a framework that generalizes LASSO by adding the cross-sectional asset pricing restriction. Numerically, the resulting optimization problem is non-standard. To address this, We provide a least angle regression-type algorithm, which produces solutions in the form of characteristics-based factors that are tradable with predetermined loadings. Our factors are easy to interpret. If k characteristics are selected, there will be k corresponding factors, each directly linked to one of the characteristics. For example, selecting characteristics such as size, book-to-market ratio (BM), and momentum will result in new size, BM, and momentum factors. These factors, like classical ones, are tradable, but unlike classical factors, they are designed to optimally explain both time series variation and cross-sectional returns. Furthermore, they maintain consistent sparsity over time, with each factor either consistently shrunk to zero or never shrunk to zero.

Empirically, we use two datasets, one from Lewellen (2015), containing a set of 15 popular characteristics, and another from Chen and Zimmermann (2022), containing a comprehensive set

[1] Lettau and Pelger (2020a,b) impose similar restrictions in PCA analysis of returns, which differs from FM regression as it does not directly use characteristics information. In addition, we add LASSO for factor selection.

of 205 characteristics, with data available from April 1962 to December 2021. In our out-of-sample tests, we find that the traditional FM regression performs poorly, yielding negative cross-sectional R^2 values for both individual stock returns and anomaly returns during the out-of-sample period from April 1977 to December 2021. The pricing errors are also large. The strong evidence indicates that the FM regression is ineffective in high-dimensional settings. In contrast, both the existing cross-sectional LASSO and the pooled LASSO improve the FM regression, yet they still underperform compared to our model, which achieves an impressive R^2 of 0.66%, surpassing all other models. Additionally, our model exhibits smaller pricing errors than those observed in previous models. The results suggest that our model offers superior performance in explaining returns and anomalies.

We also construct standard long-short portfolios based on the prediction of each model. The results show that our model consistently outperforms the other models, yielding higher average returns and a greater Sharpe ratio. After controlling for the CAPM, FF6 factors from Fama and French (2018), and Q5 factors from Hou, Mo, Xue, and Zhang (2020), we find that all the αs remain highly significant.

Since the solutions of our model are in the form of characteristics-based factors, we also conduct a series of factor test against existing factor models. We select our size, value, return on equity (RoE), investment, and momentum factors and compare them to classical factors models, including the FF6 factors from Fama and French (2018) and the Q5 factors from Hou, Mo, Xue, and Zhang (2020). Since RoE is not selected in our model, our RoE factor is shrunk to 0. However, our other factors are all more efficient than their classical counterpart. For example, our size factor has a Sharpe ratio of 0.74, significantly higher than the 0.22 of the SMB factor in FF6 and the 0.28 of the ME factor in Q5. Our value factor has a Sharpe ratio of 1.27, larger than the 0.20 of the HML factor in FF6. Our investment factor has a Sharpe ratio of 0.76, surpassing the 0.43 of the CMA factor in FF6 and the 0.52 of the IA factor in Q5. Lastly, our momentum factor has a Sharpe ratio of 1.13, also much larger than the 0.46 of the MOM factor in FF6. Moreover, the Sharpe ratio of the efficient portfolio spanned by our factors reaches 2.22 in-sample and 1.84 out-of-sample, far exceeding those achieved by FF6 (1.29 in-sample and 0.8 out-of-sample) and Q5 (1.96 in-sample and 1.44 out-of-sample). Additionally, our factors exhibit better predictive power for a wide range of anomalies, with higher R^2s and smaller pricing errors compared to FF6

and Q5.

We also explore the economic implications of the factors derived from the various models. Notably, the risk premia estimated by using the FM regression are unreliable, while both the cross-sectional LASSO (CS-LASSO) and pooled LASSO provide more reliable estimates, although they are still unsatisfactory. Specifically, CS-LASSO, which is designed for predicting time series variation, exhibits a preference for short-term characteristics with high volatility. Conversely, pooled LASSO leans towards long-term characteristics. Interestingly, our proposed model, tailored to predict both time series and cross-sectional variations, strikes a balance between short-term and long-term characteristics. Following Chen and Zimmermann (2022), we categorize all characteristics into eight groups and examine how model predictions vary across these groups. We find that CS-LASSO places greater reliance on Price and Trading-type characteristics, while pooled LASSO leans towards Accounting and Analyst-type characteristics. Our model, as expected, positioning itself between these two, but it places greater emphasis on Event-type characteristics, which are exceptionally short-term and typically exhibit high volatility. This contrasts with CS-LASSO, which tends to overfit on Event-type characteristics, and pooled LASSO, which under-exploits their pricing potential. Notably, replacing the risk premia of Event-type characteristics with 0 leads to a nearly 40% reduction in R^2 for our model, a more significant impact than on pooled LASSO (which experiences no more than a 20% reduction). This highlights our model's heightened reliance on short-term characteristics compared to pooled LASSO. Crucially, CS-LASSO appears not to rely on Event-type characteristics, despite its focus on short-term characteristics. This suggests that while CS-LASSO leverages short-term characteristics for predicting time series variation, it falls short in predicting cross-sectional variation, which is an indicative symptom of overfitting. These findings show the importance of predicting both time series and cross-sectional variations. By achieving this dual prediction, our model effectively emphasizes short-term characteristics while maintaining robust cross-sectional predictions.

Our paper is related and contributes to several strands of literature. First, it aligns with the literature advancing the FM regression. There are growing number of studies using FM regression for cross-sectional forecasting, including Green, Hand, and Zhang (2017), Kelly, Pruitt, and Su (2019), Freyberger, Neuhierl, and Weber (2020), Fama and French (2020), Gu, Kelly, and Xiu (2020), and Han, He, Rapach, and Zhou (2024). However, we distinguish ourselves from

these studies by being the first, to our knowledge, to impose both time series and cross-sectional restrictions.

Second, our paper extends the field of characteristics-based factors. Unlike classical factors model, characteristics-based factors are usually constructed through optimization. The literature includes Back, Kapadia, and Ostdiek (2015), Fama and French (2020), and Kelly, Pruitt, and Su (2019, 2020). Our contribution lies in being the first to propose characteristics-based factors that are both interpretable and robust. Different from the Instrumented Principal Component Analysis (IPCA) from Kelly, Pruitt, and Su (2019), we achieve robustness by constructing factors that are consistently sparse over time instead of integration, making our factors easier to interpret. For instance, if we select characteristics including size, BM, and momentum, our approach produces new size, BM, and momentum factors instead of some principle component.

Third, our paper aligns with the literature on characteristic selection, particularly in a high-dimensional setting. Cochrane (2011) raises the pivotal question of which characteristic provides incremental information after controlling for hundreds of others. Subsequent studies, including Green, Hand, and Zhang (2017), Freyberger, Neuhierl, and Weber (2020), and Gu, Kelly, and Xiu (2020), have explored this question using various techniques, such as FM regressions, additive group LASSO, and machine learning methods. Additionally, numerous studies focus on factor selection, exemplified by works like Barillas and Shanken (2017, 2018) and Kozak, Nagel, and Santosh (2020). Despite the technical differences between factors and characteristics, the overarching goal of identifying driving factors in the cross-section. Our paper provides a new approach that incorporates pricing restrictions.

Fourth, our paper extends the LASSO approach. Initially introduced by Tibshirani (1996), the LASSO model offers sparse estimation, with subsequent extensions such as the elastic net model by Zou and Hastie (2005) and the group LASSO by Yuan and Lin (2006) and Simon and Tibshirani (2012). Numerous other applications and extensions of LASSO have been developed, as highlighted by Hastie, Tibshirani, and Wainwright (2015) and Hastie, Tibshirani, and Friedman (2017). Our paper differs from these works by applying penalties across different samples to construct factors with consistent shrinkage over time. Additionally, we consider both time series and cross-sectional variations, which is crucial in asset pricing.

Fifth, our paper aims to find a piecewise linear solution path for the model with a penalty

over time, thus linking to the piecewise linear method literature. Early work includes McZgee and Carleton (1970), Bookstein (1975), and other papers focus on the selection strategies such as forward stagewise selection, as surveyed by Hastie, Tibshirani, and Friedman (2017). Efron, Hastie, Johnstone, and Tibshirani (2004) propose the least-angle regression (LARS) algorithm, and show that the LASSO can be approximately solved by LARS, with the solution closely resembling LASSO with minor adjustments. Some other LASSO-type models, such as en-LARS and group-LARS proposed by Zou and Hastie (2005) and Yuan and Lin (2006), can also be solved using LARS-type algorithms. While numerous LASSO-type models and algorithms have been proposed, our paper introduces a novel penalty design across different samples rather than predictors. Our algorithm seeks a piecewise linear solution path, ensuring consistent shrinkage of factors across time.

The rest of the paper is as follows. Section 2 introduces the theoretical framework and the LARS-type solution, along with the construction of characteristics-based factors. Section 3 reports the major empirical results, provides a factor test and the interpretation of different characteristics. Section 4 concludes.

2 FM Regression with Asset Pricing Restrictions

2.1 Our methodology

Denote characteristics as $Z_t(N_t \times l)$, excess returns as $r_t(N_t \times 1)$, factors as $f_t(l \times 1)$, and risk premia as $\mu(l \times 1)$. Classical Fama-MacBeth regression is a procedure that runs cross-sectional regression of r_{t+1} on Z_t periods by periods, or by another words, minimize the object function $\|r_{t+1} - Z_t f_{t+1}\|_2^2$, then estimate the risk premia as $\mu = \frac{1}{T-1}\sum_{t=1}^{T-1} f_{t+1}$. We can also sum up all the object function in different periods and rewrite FM regression as follows,

$$f = \underset{f}{\operatorname{argmin}} \sum_{t=1}^{T-1} \|r_{t+1} - Z_t f_{t+1}\|_2^2. \tag{1}$$

However, there are some shortcomings about classical Fama-MacBeth regression. On the one hand, since it is an OLS based procedure, it is obvious that FM regression might be not robust enough when the dimension of characteristics is high. On the other hand, we should point out that

although Fama-MacBeth regression is a popular method for predicting the cross-sectional return, it is optimized by explaining the time series return instead of cross-section. However, to use the characteristics to predict the stock return and then construct the characteristics-based factors, it is important for the factors to explain both the variation of time series and cross-section. To address the issues of Fama-MacBeth regression and those emerging methods, we propose a new model:

$$f = \underset{f}{\operatorname{argmin}} \left\{ \overbrace{\sum_{t=1}^{T-1} ||r_{t+1} - Z_t f_{t+1}||_2^2}^{\text{Optimizing for the time series}} + \phi^2 \overbrace{\sum_{t=1}^{T-1} \|r_{t+1} - Z_t \mu\|_2^2}^{\text{Optimizing for the cross section}} + 2\lambda \overbrace{\sum_{i=1}^{l} \sqrt{\sum_{t=1}^{T-1} f_{i,t+1}^2 + \phi^2 \mu_i^2}}^{\text{Group LASSO penalty over time}} \right\}, \tag{2}$$

where $\mu = \frac{1}{T-1} \sum_{t=1}^{T-1} f_{t+1}$ and $\phi^2 \geq 0$. In the upcoming section, we will demonstrate that our model, in comparison to Fama-MacBeth regression, effectively predicts both time series and cross-sectional variations while generating sparse factors. Equation (2) comprises three components. The first part represents the loss function of FM regressions, aiming to explain the time series variation. The second part corresponds to capturing the cross-sectional variation. The third part introduces a group LASSO penalty on the second moment of each factor and its associated risk premia. With this penalty, our model aims to generate factors with consistent sparsity that once a factor is shrunk to 0, it remains 0, and if not, it stays nonzero.

To gain a deeper understanding of why our model produces sparse factors with consistent sparsity over time, let's assume $\phi^2 = 0$ for simplicity. Equation (2) then simplifies to:

$$\sum_{t=1}^{T-1} ||r_{t+1} - Z_t f_{t+1}||_2^2 + 2\lambda \sum_{i=1}^{l} \sqrt{\sum_{t=1}^{T-1} f_{i,t+1}^2}. \tag{3}$$

Equation (3) comprises the loss function of FM regressions and a group LASSO penalty on each factor across different samples. If we denote $R(i) = \sqrt{\sum_{t=1}^{T-1} f_{i,t+1}^2}$ as the radius of factor i, Equation (3) imposes a LASSO penalty on the radius of each factor, ensuring the sparsity of the solution remains consistent over time. If $R(i) = 0$, factor i will always be shrunk to 0; if $R(i) \neq 0$, factor i will hardly be 0.

To illustrate this idea, consider the simplified case of only two periods. In this case, Equation (3) can be expressed as:

$$\begin{aligned} &\min_{f} \sum_{t=1}^{T-1} ||r_{t+1} - Z_t f_{t+1}||_2^2 \\ &s.t. \sum_{i=1}^{l} R(i) \leq R. \end{aligned} \tag{4}$$

Regard Equation (3) as the Lagrange form, then Equation (4) is the restrict form of exactly the same optimizing problem. For any λ in Equation (3), there exist a R in Equation (4) and the two equations will give the same solution. The solution will not be sparse if $\lambda = 0$, and if λ increase, the solution will be sparse gradually. Correspondingly, the solution will also not be sparse if $R \to +\infty$, and will become sparse gradually when R decrease. We illustrate the optimization problem in Figures 1 and 2. The objective is to minimize the loss function $\sum_{t=1}^{T-1} ||r_{t+1} - Z_t f_{t+1}||_2^2$ within the cone restriction, leading to two possible cases for the solution. In the first case (Figure 1), the solution is at the corner, resulting in factor j being consistently shrunk to 0. In the second case (Figure 2), the solution is not at the corner, and factor j remains nonzero. Although theoretically possible for factor j to be 0 in one period,[2] this situation has negligible probability in practice. The consistency holds for multiple periods. Unlike the LASSO applied independently in each period, our model ensures consistent selection over time, facilitating straightforward interpretation. A factor is either consistently shrunk to 0 or consistently remains nonzero. This contrasts with the traditional group LASSO penalty method (Yuan and Lin, 2006; Simon and Tibshirani, 2012), which penalizes the group variable within one sample, while our penalty restricts the variation of f_t across different samples.

[Insert Figures 1 and 2 about here]

Equation (3) allows us to obtain a sparse solution for our model with consistency over time. However, it alone does not ensure an effective explanation of cross-sectional returns. To enhance the prediction of cross-sectional returns, we introduce a pooled OLS penalty to the loss function, resulting in Equation (2). Obviously, when $\phi^2 > 0$, our model becomes capable of predicting both

[2]Such as $f_{j,2} \neq 0$ and $f_{j,1} = 0$, or in other words, the solution is not at the corner, and the AOB plane coincides with the $f_{j,1} - O - \sum_{i \neq j} |R(i)|$ plane.

time series and cross-sectional variations. The selection of ϕ^2 can be determined through methods like cross-validation. In the empirical section, we will present results with different values of ϕ^2. The outstanding question is how to solve the model. Rather than resorting to numerical solutions, we aim to derive a latent factors-type solution, following the approaches of Fama and French (2020) and Kelly, Pruitt, and Su (2019). We expect our solution to take the form $f_{t+1} = g(Z_t)r_{t+1}$, which can be interpreted as a set of tradable portfolios with predetermined loadings. This representation holds special significance in the context of asset pricing and facilitates an examination of model efficiency by inspecting the Stochastic Discount Factor (SDF). We propose a piecewise linear algorithm to derive the solution for Equation (2). This algorithm, akin to least-angle regression, is designed to explain both time series and cross-sectional variations. The tuning parameter is the number of selected characteristics k instead of λ, rendering it easy to interpret. The solution generated by our algorithm consistently exhibits sparsity over time. A factor is either consistently shrunk to 0 or consistently remains nonzero, making our model an effective choice for characteristic selection and cross-sectional prediction.

2.2 The LARS type algorithm

The Least Angle Regression (LARS) algorithm introduced originally by Efron, Hastie, Johnstone, and Tibshirani (2004), has widespread use in machine learning literature due to its similarity to LASSO and its piecewise linear solution path, which proves to be highly advantageous in our context. To estimate Equation (2), we simply need to extend the traditional LARS algorithm, and the methodology is outlined in the following parts. We present the Karush–Kuhn–Tucker (KKT) conditions for Equation (2):

Proposition 1 *Denote* $f_t = (f_{1,t}, \cdots, f_{l,t})^\top, f_{(i)} = (f_{i,2}, \cdots, f_{i,T})^\top, Z_t = (Z_{1,t}, \cdots, Z_{l,t})$, *a necessary and sufficient condition for the solution of Equation (2) is:*

$$-\begin{pmatrix} Z_{i,1}^\top(r_2 - Z_1 f_2) \\ \vdots \\ Z_{i,T-1}^\top(r_T - Z_{T-1} f_T) \\ \phi \sum_{t=1}^{T-1} Z_{i,t-1}^\top(r_{t+1} - Z_t \mu) \end{pmatrix} + \frac{\lambda}{\sqrt{||f_{(i)}||_2^2 + \phi^2 \mu_i^2}} \begin{pmatrix} f_{i,2} \\ \vdots \\ f_{i,T} \\ \phi \mu_i \end{pmatrix} = 0, \quad \forall ||f_{(i)}||_2^2 > 0, \tag{5}$$

with

$$\left\| \begin{array}{c} Z_{i,1}^{\top}(r_2 - Z_1 f_2) \\ \vdots \\ Z_{i,T-1}^{\top}(r_T - Z_{T-1} f_T) \\ \phi \sum_{t=1}^{T-1} Z_{i,t-1}^{\top}(r_{t+1} - Z_t \mu) \end{array} \right\|_2 \leq \lambda, \quad \forall ||f_{(i)}||_2^2 = 0, \tag{6}$$

where $\mu = \frac{1}{T-1} \sum_{t=1}^{T-1} f_{t+1}$.

The gradient descent aids in designing the algorithm, which proceeds as follows.

Step 1: Start from $f_t^{(0)} = 0, k = 1, r_t^{(0)} = r_t, t = 2, \cdots, T$.

Step 2: Find the current 'most correlated factor':

$$\mathcal{A}_1 = \underset{i}{argmax} \{ \sum_{t=1}^{T-1} (Z_{i,t}^{\top} r_{t+1})^2 + \phi^2 (\sum_{t=1}^{T-1} Z_{i,t}^{\top} r_{t+1})^2 \}$$

Step 3: For $t = 1, \cdots, T-1$, compute the current direction $\gamma_{t+1}^{(k)} = A^{(k)}(Z_t) r_{t+1}^{(k-1)}$, where $A^{(k)}(Z_t)$ is a $l \times N_t$ matrix, with the $\mathcal{A}_k$'th rows to be $(Z_{\mathcal{A}_k,t}^{\top} Z_{\mathcal{A}_k,t})^{-1} Z_{\mathcal{A}_k,t}^{\top}$ and other rows to be 0.[3]

Step 4: Define the angle:

$$C(j,k,\alpha) = \sum_{t=1}^{T-1} [Z_{j,t}^{\top}(r_{t+1}^{(k-1)} - \alpha Z_t \gamma_{t+1}^{(k)})]^2 + \phi^2 [\sum_{t=1}^{T-1} Z_{j,t}^{\top}(r_{t+1}^{(k-1)} - \alpha Z_t \gamma_{t+1}^{(k)})]^2, \tag{7}$$

for each $j \notin \mathcal{A}_k$, solve out the α_j from the following equation:

$$C(j,k,\alpha_j) = C(j',k,\alpha_j), \tag{8}$$

where j' can be selected from $\mathcal{A}_k$ arbitrarily.[4]

Step 5: Let $j^* = \underset{j}{argmin}\{\alpha_j\}, \alpha^{(k)} = \alpha_{j^*}$ and $\mathcal{A}_{k+1} = \mathcal{A}_k \cup \{j^*\}$.

Step 6: For $t = 1, \cdots, T-1$ update $f_{t+1}^{(k)} = f_{t+1}^{(k-1)} + \alpha^{(k)} \gamma_{t+1}^{(k)}, r_{t+1}^{(k)} = r_{t+1}^{(k-1)} - \alpha^{(k)} Z_t \gamma_{t+1}^{(k)}$. If

[3]Note that the first subscript $\mathcal{A}_k$ in $Z_{\mathcal{A}_k,t}$ denotes the row number. For example, if $\mathcal{A}_k = \{1,3\}$, then $Z_{\mathcal{A}_k,t}$ consists of the first and third rows of Z_t.

[4]For computational efficiency, we can apply the following equation in coding:

$$Z_{j,t}^{\top} r_{t+1}^{(k-1)} = \prod_{i=1}^{k-1} [I_l - \alpha^{(i)} Z_t^{\top} Z_t Icrp_t^{(i)}] N_t x_{t+1},$$

where $Icrp_t^{(i)}$ (inverse cross product) is a $l \times l$ matrix, and the $\mathcal{A}_i$th row and column are $(Z_{\mathcal{A}_i,t}^{\top} Z_{\mathcal{A}_i,t})^{-1}$ and $A^{(i)}(Z_t) = Icrp_t^{(i)} Z_t^{\top}$

$k < l$ then go back to step 3.

Solving Equation (8) is straightforward, as it involves solving a univariate quadratic equation. Similar to most LARS algorithms, our algorithm serves as a good approximation of Equation (2). We will proceed with the argument in two ways. First, we demonstrate that in each step of the algorithm, a necessary condition for Equation (2) to be true is maintained:

Proposition 2 *In the $k-1$th round, the k'th most correlated factor has been found and added to $\mathcal{A}_k$. In the kth round, the factors—$\mathcal{A}_k$ in the active set are move together in a way that preserves their "angle" and ensures it is tied and decreasing. Regardless of the choice of α_j, denote $\mathcal{A}_k = \{\mathcal{A}_k^{(i)}\}, i = 1, \cdots, k$, the following holds:*

$$C(\mathcal{A}_k^{(1)}, k, \alpha_j) = \cdots = C(\mathcal{A}_k^{(k)}, k, \alpha_j).$$

This proposition can be proven by mathematical induction, as detailed in Appendix A. Second, the solution of the algorithm will be exactly the same as Equation (2) in some special cases:

Proposition 3 *If the characteristics are orthogonal, i.e., $Z_t^\top Z_t = I_l$, for $t = 1, \ldots, T-1$, then the solution of the LARS-type algorithm will generate the entire solution path of Equation (2).*

The proof is provided in Appendix B.

Proposition 2 reveals that the algorithm facilitates the movement of factors together, ensuring that all selected characteristics share the same "angle" with excess returns. This property is crucial for the piecewise linear solution path. Similar to group LARS, Proposition 2 states that, at each step of the algorithm, the necessary condition of the KKT condition of Equation (5) is consistently satisfied, justifying its effectiveness as a good approximation. It provides insights into how the solution path should be obtained.

The historical approach to variable selection in statistics includes forward stepwise selection, which picks one variable in each round and updates the residual using the OLS residual. However, this method proves too aggressive and lacks global optimality. The forward stagewise selection strategy emerged as a milder alternative, taking small steps towards the OLS residual in each round but incurring higher computational costs. The LARS algorithm, proposed by Efron, Hastie, Johnstone, and Tibshirani (2004), combines the benefits of forward stagewise selection by ensuring that the selected variable maintains the same angle to the residual at each step. It has

demonstrated its effectiveness as a good approximation of LASSO, providing a global optimum efficiently and forming the ideal way to obtain the solution path for LASSO-type problems.[5]

In the application of the LARS-type algorithm, Proposition 2 ensures empirical and computational efficiency. Proposition 3 indicates that the solution of Equation (2) aligns with the LARS-type algorithm in special cases, suggesting that the latter is a reliable approximation for a more general case. Denote $Icrp_t^{(i)}$ (inverse cross product) as a $l \times l$ matrix, where the $\mathcal{A}_i$th row and column are $(Z_{\mathcal{A}_i,t}^\top Z_{\mathcal{A}_i,t})^{-1}$ and $A^{(i)}(Z_t) = Icrp_t^{(i)} Z_t^\top$, then the LARS-type algorithm helps us to obtain a solution:

Proposition 4 *A simple iterative relationship exists:*

$$r_{t+1}^{(k)} = [I_{N_t} - \alpha^{(k)} Z_t A^{(k)}(Z_t)] r_{t+1}^{(k-1)}, \tag{9}$$

and the characteristics-based factors can be computed as:

$$\begin{aligned} f_{t+1}^{(k)} &= \{\sum_{i=1}^{k} \alpha^{(i)} A^{(i)}(Z_t) \prod_{j=1}^{i-1} [I_{N_t} - \alpha^{(j)} Z_t A^{(j)}(Z_t)]\} r_{t+1} \\ &= \{\sum_{i=1}^{k} \alpha^{(i)} Icrp_t^{(i)} \prod_{j=1}^{i-1} [I_l - \alpha^{(j)} Z_t^\top Z_t Icrp_t^{(i)}]\} N_t x_{t+1}. \end{aligned} \tag{10}$$

Equation (10) facilitates out-of-sample forecasting by estimating $\mathcal{A}_k$ and $\alpha^{(k)}$ in-sample, then using the loading—$\{\sum_{i=1}^{k} \alpha^{(i)} A^{(i)}(Z_t) \prod_{j=1}^{i-1} [I_{N_t} - \alpha^{(j)} Z_t A^{(j)}(Z_t)]\}$ to construct sparse, interpretable, and robust out-of-sample factors.

3 Empirical Results

3.1 Data description

In our empirical analysis, we primarily use two datasets. The first dataset, cited from Lewellen (2015), includes 15 characteristics. Meanwhile, the second dataset, derived from Chen and Zim-

[5]Indicating how much each step should be taken so that we do not have to take a tiny step because of the fear of being too aggressive.

mermann (2022), consists of 205 characteristics[6]. These datasets are crucial in asset pricing due to their high quality. The Lewellen (2015) dataset is used to demonstrate our model's effectiveness in a moderately dimensional setting, highlighting its ability to predict variations across both time series and cross-sectional dimensions. In contrast, the Chen and Zimmermann (2022) dataset is employed to validate the robustness of our model in scenarios with higher dimensionality.

We have standardized all characteristics within the Chen and Zimmermann (2022) dataset to ensure unit variance and a mean of zero. Due to the presence of outliers in several characteristics from the Chen and Zimmermann (2022) dataset, we have applied rank normalization, mapping them to an interval of $[-1, 1]$. Our stock return data is sourced from the Center for Research in Security Prices (CRSP) monthly stock file, focusing on stocks from the NYSE, Amex, or Nasdaq for analysis. This data covers the period from April 1962 to December 2021.

Additionally, we have incorporated a section on factor testing to elucidate the efficacy of our factors, as denoted by Equation (10). In this segment, we leverage 205 anomalies constructed from the Chen and Zimmermann (2022) dataset's characteristics. We show the predicting power of each factors model on these 205 characteristics and readers can see that our factors performs the best. Besides, we also compare the Sharpe ratio of our factors to classical factors, together with their factors spanned pricing kernel and again our factors outperforms all the other factors.

To prevent information leakage, we've implemented lagging across all characteristics. Following standard practices, we utilize a one-month lag for trading characteristics (e.g., Mom12m), a six-month lag for annual accounting data availability, and a one-quarter lag for quarterly accounting data.

3.2 Out-of-sample-test

In this subsection, we utilize the in-sample data to estimate our model, including parameters such as $\mathcal{A}_k$ and $\alpha^{(k)}$, alongside other models. We then use Equation (10) for out-of-sample prediction. Our approach involves conducting rolling out-of-sample tests. Initially, we utilize the data from the first 15 years for training and predict the next 5 years. Subsequently, we expand the training sample to the first 20 years and predict the subsequent 5 years, repeating this process. In essence,

[6]The data from Chen and Zimmermann (2022) is accessible at www.openassetpricing.com. Chen and Zimmermann (2022) provides a comprehensive summary. The dataset we utilize represents an earlier version, and current versions may contain additional characteristics.

we refit all the models every 5 years, progressively increasing the training sample by 5 years and predicting the next 5 years.

We will show the out of sample performance of FM with Restriction from different aspect. Our empirical results show that FM with Restriction performs the best among all the models we have tested from different aspects. For example, while predicting individual stock return or anomalies return, our model have the largest R^2 out of sample. We can also construct long-short portfolios based on different predictions and we show in later part that the portfolios based on our model performs the best, with largest Sharpe ratio and most significant α after controlling FF6 factors in Fama and French (2018) and Q5 factors in Hou, Mo, Xue, and Zhang (2020). What's more, while constructing the characteristics-based factor as in Equation (10), we show that our factors achieve a high Sharpe ratio of factor spanned tangency portfolio, or pricing kernel out of sample. Considering that the risk premia of characteristics might vary across time, following Lewellen (2015), we apply a smoothing average technique for estimating the risk premia. Once we get the factors f_{t+1}, then instead of using all the historical data to calculate mean value as the estimated risk premia, we can just use the data in the recent years:

$$\hat{\mu}_t = \frac{1}{t - MA} \sum_{s=t-MA}^{t} f_{s+1}, \tag{11}$$

where MA is the moving average length and μ_t denote the risk premia estimated by using all the data before time t. After getting μ_t, we can use it to predict out of sample.

We aim to compare our model with traditional Fama-MacBeth regression and other emerging methods. Fama-MacBeth regression is proposed by Fama and MacBeth (1973) and has a profound influence to the following researches. However, since it is a OLS based procedure, it is not robust when dimension is large. Lewellen (2015) and Green, Hand, and Zhang (2017) have shown that the prediction of Fama-MacBeth regression can be a good signal for constructing the long-short portfolio, and in this paper, we will show a more comprehensive results of its performance to be used as a basic benchmark for comparing different models. It is shown that while the prediction is indeed a useful signal for constructing the long-short portfolio, the explanatory power is weak, for example, R^2 can be negative due to a severe overfitting. Following Lewellen (2015), to avoid the case that some outliers in the history might have a huge influence to the estimation, we also apply a moving average methods for Fama-MacBeth regression.

We will also compare our model to some other emerging methods. The first method is to pooled the panel data first and then use LASSO for prediction and selection, which we called pooled LASSO. While pooled LASSO is more robust in high-dimensional settings and is a logical choice for characteristic selection, it might fall short in explaining time series variation, potentially lacking efficiency from an asset pricing perspective. Pooled LASSO may emphasize characteristics with long-term and low volatility, exhibiting lower reliance on high-volatility characteristics compared to other models. Given that our characteristics are standardized, higher absolute risk premia indicate greater importance for model prediction. Therefore, absolute risk premia of high-volatility characteristics estimated by pooled LASSO might be relatively smaller than those estimated by other methods. Following Han, He, Rapach, and Zhou (2024), the second method is to use LASSO to replace the OLS procedure in the Fama-MacBeth regression. We refer to it as cross-sectional LASSO (CS-LASSO). It aims at enhancing Fama-MacBeth regression, focuses on explaining time series variation rather than the cross-section. The potential drawback is that it might excessively rely on high-volatility characteristics for time series prediction, potentially struggling with cross-sectional prediction. In addition, its selection is not consistent over time, leading to some degree of inefficiency. In contrast, our model predicts both time series and cross-sectional variation while maintaining consistent selection over time, establishing it as the most efficient choice. In subsequent sections, we empirically demonstrate the superiority of our model. To simplify pooled LASSO and CS-LASSO estimation, we employ the LARS procedure proposed by Efron, Hastie, Johnstone, and Tibshirani (2004). This procedure provides an approximate solution for LASSO-type models, converting the tuning parameter λ into k, the number of selected characteristics. This is more economically meaningful and significantly more computationally efficient. The LARS procedure is a piecewise linear algorithm, where one characteristic is selected and added to the active set in each round, updating the residual for the subsequent round. Stopping at the k'th round results in k selected characteristics. This procedure is analogous to deducing λ from $+\infty$ to 0, with the number of selected predictors decreasing at breakpoints. In pooled LASSO, we use LARS to select k characteristics and estimate the risk premia, while in CS-LASSO, we use LARS to select k characteristics in each period. It's important to note that the solution of pooled LASSO and CS-LASSO lacks a latent factors structure, and no Stochastic Discount Factor (SDF) can be constructed for these methods.

In this section, we focus on tuning the parameter using the Bayesian Information Criterion (BIC). To apply the BIC criterion, it is essential to determine the degrees of freedom for the LARS-type algorithm. Efron, Hastie, Johnstone, and Tibshirani (2004) demonstrates that the degrees of freedom for LARS are approximately the number of selected predictors. Consequently, we can employ the following equation as an approximation of the BIC for each model:

$$\text{BIC} = \frac{\sum_{t=1}^{T-1} \|r_{t+1} - Z_t\hat{\mu}\|_2^2}{\hat{\sigma}^2} + kln(\sum_{t=1}^{T-1} N_t). \tag{12}$$

Additionally, our model involves another hyperparameter, ϕ^2. We report the results of $\phi^2 = 3$ for Lewellen (2015) data and $\phi^2 = 12$ for Chen and Zimmermann (2022) data in most of the parts in our empirical section. Readers interested in the performance of other choice of ϕ^2 can read the section of robustness check.

3.2.1 The predictive capability of different models

In this subsection, we compare the explanatory power of different models using several metrics. Denote the prediction of r_t as $\hat{r}_{t+1} = Z_t\mu$, the first is R^2 for r_t, defined as

$$R^2(r_t) = 1 - \frac{\sum_{t=1}^{T-1} ||r_{t+1} - \hat{r}_{t+1}||_2^2}{\sum_{t=1}^{T-1} ||r_{t+1}||_2^2}. \tag{13}$$

The second is R^2 for $x_{t+1} = \frac{1}{N_t} Z_t^\top r_{t+1}$, defined as

$$R^2(x_t) = 1 - \frac{\sum_{t=1}^{T-1} ||x_{t+1} - \frac{1}{N_t} Z_t^\top \hat{r}_{t+1}||_2^2}{\sum_{t=1}^{T-1} ||x_{t+1}||_2^2} \tag{14}$$

These two metrics are utilized in Kelly, Pruitt, and Su (2019) and Gu, Kelly, and Xiu (2021) to demonstrate the models' ability in explaining cross-sectional returns. Naturally, higher R^2 values indicate better predictions. The third metric is the predictive coefficient. Following the prediction $\hat{r}_{t+1} = Z_t\mu$, another measure of prediction accuracy is β_{slope}, derived from regressing r_t on $\hat{r}_t$ to obtain a predictive coefficient β_t, which is then averaged across time to yield $\beta slope$:

$$
\begin{aligned}
r_{t+1} &= \beta_t^\top \hat{r}_{t+1} + \epsilon_{t+1}, \quad t = 1, \cdots, T-1 \\
\beta_{slope} &= \frac{1}{T-1} \sum_{t=1}^{T-1} \beta_t.
\end{aligned} \tag{15}
$$

Lewellen (2015) demonstrated the regression slope of individual characteristics, whereas our approach emphasizes the predictive power of each model using all characteristics collectively rather than independently. Additionally, we calculate the pricing error of x_t by

$$
\alpha(x) = \frac{1}{T-1} \sum_{t=1}^{T-1} \left(x_{t+1} - \frac{1}{N_t} Z_t^\top r_{t+1} \right). \tag{16}
$$

The fourth metric is the mean absolute pricing error. Referring to the risk premia of x_t as $\mu(x)$, as per Fama and French (2015), we derive:

$$
\overline{|\alpha|} / \overline{|\mu|} = \frac{\|\alpha(x)\|_1}{\|\mu(x)\|_1}. \tag{17}
$$

The fifth metric is the mean square pricing error. Following Fama and French (2015), then we obtain:

$$
\overline{\alpha^2} / \overline{\mu^2} = \frac{\|\alpha(x)\|_2^2}{\|\mu(x)\|_2^2}. \tag{18}
$$

We report the out-of-sample results of different models in Table 1. Here we use the recursive window starting from April 1962 to March 1977 to train each model and predict the next 5 years, then we add the next 5 years data into the training set and predict the 5 future years and repeat the procedure to obtain the whole out-of-sample prediction. Panels A and B show the results based on 15 characteristics data provided by Lewellen (2015) and 205 characteristics data provided by Chen and Zimmermann (2022), respectively. In each panel, we present the outcomes obtained using various lengths of smoothing averages for estimating the risk premia. It's evident that regardless of the datasets or the specific moving average lengths used, our model consistently delivers superior performance. The selection of k is determined by the BIC rule in this context. For Panel A, we set $\phi^2 = 3$, while for Panel B, $\phi^2 = 12$. [7]

[7] Our empirical results remain qualitatively similar with different ϕ^2 (shown in Figures 15, 16, 17, 18, and 19).

[Insert Table 1 about here]

In Panel A of Table 1, we conduct analysis with the Lewellen (2015) dataset, revealing the nuanced capabilities of each model within a moderate-dimensional context. Our model stands out as the frontrunner, particularly evidenced by its performance in the expanding moving average scenario, where it achieves the most commendable $R^2(r_t)$ at 0.08% and $R^2(x_t)$ at 0.94%. It demonstrates pricing errors of 0.4 and 0.21 for $\overline{|\alpha|}/\overline{|\mu|}$ and $\overline{\alpha^2}/\overline{\mu^2}$, respectively, using an expanding window. The results for other choices of moving average lengths are similarly favorable, consistently demonstrate our model's superiority over others. On the contrary, the Fama-MacBeth regression struggles, displaying negative R^2 values that point to challenges in predictive accuracy. The Cs-lasso and Pooled lasso models show moderate improvements, with Cs-lasso achieving a notable β_{slope} of 0.93, indicating some alignment with market movements but falling short of matching the comprehensive excellence of our model.

Panel B shifts to the high-dimensional realm using the Chen and Zimmermann (2022) dataset, where complexity substantially increases. Here, our model's performance is notably strong, with an $R^2(r_t)$ of 0.62% and an $R^2(x_t)$ of 2.52% in the expanding window. Results with a moving average length of 120 months are particularly impressive, as the β_{slope} becomes exactly 1 and most of other metrics also come to their best. This stands in stark contrast to the Fama-MacBeth model, which encounters significant challenges with notably negative R^2 values, highlighting its limitations in analyzing high-dimensional data. While Cs-lasso shows improvement from its performance in moderate-dimensional scenarios, it still falls short in certain moving average window and metric selections. Similarly, the Pooled lasso model displays some resilience with positive R^2 values but lacks the comprehensive efficiency and predictive accuracy of our model.

Through these evaluations, our model stands out as a symbol of adaptability, precision, and efficiency. Its consistent outperformance in both moderate and high-dimensional scenarios, marked by high out-of-sample R^2 values, minimized pricing errors, and the highest Sharpe ratios, highlights its potential as an exceptional predictive tool for individual stock-level returns. Readers might be interested in how each model performs across diverse time periods. To address this, we segment the data into distinct periods, and the outcomes are detailed in the robustness check sections. In line with Lewellen (2015), our focus will be on displaying a rolling β_{slope} to illustrate the evolving predictive prowess of various models.

To better illustrate our predictive ability in constructing long-short portfolios, we present the out-of-sample results for different models in Figure 3. To ensure clarity, we selected the moving average length for each model based on its optimal performance, as outlined in Table 1, and only include these top-performing results in Figure 3. For instance, we opted for the expanding window for CS-LASSO using Lewellen (2015) data, reporting it here due to its β_{slope} of 0.93, closer to 1 compared to other options. In the figure, we observe that the performance of CS-LASSO, pooled LASSO, and our model are quite similar on Lewellen (2015) data for the majority of the time. However, the performance of pooled LASSO sharply declines after 2015, likely contributing to its poor performance in Table 1. On Chen and Zimmermann (2022) data, CS-LASSO and our model consistently exhibit the best performance. Meanwhile, Fama-MacBeth regression consistently shows poor performance in Figure 3, suggesting it may lack robustness even with a moderate number of characteristics.

[Insert Figure 3]

3.2.2 Long-short portfolios

In this subsection, we aim to show the performance of long-short portfolios based on the predictions of each model. At the end of each month, we compute out-of-sample stock return predictions—$\hat{r}_{t+1}$ for the following month using each model. Subsequently, we categorize stocks into deciles according to the forecasts generated by each model. Finally, we construct a zero-cost investment portfolio that purchasing (longing) stocks with the highest expected returns (decile 10) and selling (shorting) those with the lowest expected returns (decile 1). The results, displayed in Table 2, show our findings using the optimal moving average length for β_{slope} performance, as determined in Table 1. For instance, we utilize an expanding window for FM with Restriction for Lewellen (2015) data, due to its β_{slope} is closer to 1 compared to alternative choices.

[Insert Table 2 about here]

Table 2 features two panels: Panel A for Lewellen (2015) data and Panel B for Chen and Zimmermann (2022) data. The long-short portfolios are value-weighted, and we present the mean return, standard deviation, and Sharpe ratio for each. Notably, our model demonstrates

superior performance. In the case of Lewellen (2015) data, with 15 characteristics, our model achieves a noteworthy mean return of 1.31 and a Sharpe ratio of 0.76. Likewise, on Chen and Zimmermann (2022) data, our model delivers a mean return of 2.17 and a Sharpe ratio of 1.65, underscoring its robust predictive capabilities. These outcomes underscore our model's ability to explain cross-sectional returns. Although other models, like Fama-MacBeth regression, exhibit weaker performance across various metrics, their long-short portfolios still present positive mean returns and Sharpe ratios significantly above zero. While CS-LASSO and pooled LASSO perform better, they fall short of matching our model's predictive prowess. As detailed in Table 1, we set $\phi^2 = 3$ for Panel A and $\phi^2 = 12$ for Panel B.[8]

To further illustrate the efficiency of our prediction, we also consider the α after controlling classical factor models including the CAPM in Sharpe (1964); the six-factor model of Fama and French (2018), denoted FF6; and the five-factor q model of Hou, Mo, Xue, and Zhang (2020), denoted Q5.[9]. We report the results in Tables 3 and 4.

[Insert Tables 3 and 4 about here]

In Tables 3 and 4, it's evident that these models exhibit strong performance, with their long-short portfolios cannot be explained by classical factor models. These findings contribute to the ongoing discussion surrounding 'Beta versus Characteristics.' Works such as Kelly, Pruitt, and Su (2019) and Fama and French (2020) offer comparative analyses between characteristics-based factor models and traditional beta pricing models. They consistently demonstrate the superior performance of characteristics-based models, attributed to their optimization-focused design, contrasting with the predominant focus on beta pricing models in asset pricing research. The results presented here further substantiate this superiority. Notably, even after accounting for classical factors like FF6, Q5, and CAPM, the Fama-MacBeth regression's long-short portfolios continue to yield significant α values, reinforcing the efficacy of characteristics-based factor models.

Our model stands out as the top performer among all these models. The α values are notably

[8]Readers interested in alternative ϕ^2 choices can refer to Figure 20 and 21. Additionally, the moving average lengths for each model are selected based on their optimal β_{slope} performance from Table 1, prioritizing closeness to 1 for better predictions. For Lewellen (2015) data, Fama-MacBeth regression employs a 120-month moving average, while CS-LASSO, pooled LASSO, and FM with Restriction utilize an expanding window. Conversely, for Chen and Zimmermann (2022) data, Fama-MacBeth regression employs a 120-month moving average, CS-LASSO uses a 60-month moving average, pooled LASSO adopts an expanding window, and FM with restriction employs a 120-month moving average.

[9]Our FF6 data is obtained from Kenneth French's library and Q5 data from https://global-q.org/factors.html

substantial, accompanied by highly significant t values. For instance, utilizing 205 characteristics from Chen and Zimmermann (2022) data, the t value for α remains at a remarkable 8.31 even after adjusting for FF6 factors.

3.3 Factor test

3.3.1 The characteristics-based factors and factor test

In addition to robust predictive capabilities, this paper also aims to introduce a set of interpretable and efficient factors. These factors are detailed in Equation (10), serving as an approximate solution to Equation (2) and designed to capture both time series and cross-sectional variations. The preceding subsection focuses on out-of-sample stock-level predictions in a high-dimensional setting. In this subsection, we are going to evaluate our factors using classical factor tests and then dissect their construction methodology.

Before the analysis, let's first introduce how the factors of FM with Restriction are constructed. There are two versions of these factors: an in-sample version and an out-of-sample version. The in-sample version is derived by training FM with Restriction on all available data from Chen and Zimmermann (2022) (the full time period is from April 1962 to December 2021), followed by constructing the factors using Equation (10). Conversely, the out-of-sample version is generated through a recursive procedure as previously described: utilizing data from the first 15 years (April 1962 to March 1977) to train our model and construct factors using Equation (10) for the subsequent 5 years. This process continues by incorporating the next 5 years' data into the training set, refitting the model, and generating factors for the following years, ensuring no information leakage occurs. These out-of-sample factors align with those utilized for out-of-sample prediction in the preceding section. Additionally, our factors boast high interpretability. For instance, we can employ our model to address the following problem:

$$\begin{aligned} r_{t+1} =& size_t \times f_{t+1}(size) + BMdec_t \times f_{t+1}(BMdec) + RoE_t \times f_{t+1}(RoE) \\ &+ AssetGrowth_t \times f_{t+1}(AssetGrowth) + Mom12m \times f_{t+1}(Mom12m) \\ &+ \sum_{i \notin \mathcal{A}} Z_i f_{t+1}(i) + \epsilon_{t+1}, \end{aligned} \tag{19}$$

where $\mathcal{A} = \{size, BMdec, RoE, AssetGrowth, Mom12m\}$. Here, $size$, $BMdec$, and others are characteristics in Chen and Zimmermann (2022) data. In Equation (19), if we view $size_t$ as a dynamic loading for market size, then $f(size)$ represents a novel size factor; similarly, if we see $BMdec_t$ as a dynamic loading for book-to-market value, then $f(BMdec)$ becomes a new BMdec factor. In a broader context, for any characteristic i, when employing FM with Restriction to address a model like Equation (19), we can consistently utilize Equation (10) to derive the corresponding factor $f(i)$. Additionally, the fundamental concept is that we leverage characteristics to deduce factors $f(i)$, which should effectively capture both temporal and cross-sectional variations. These factors maintain consistent sparsity across time, bolstering their robustness in high-dimensional scenarios. Once a factor is shrunk to zero, it remains zero; conversely, if it's non-zero, it persists as such.

Given the interpretability of our factors, we can directly utilize $f(size)$ against classic factors like the SMB factor in FF6, anticipating our factor to demonstrate greater efficacy. Similarly, we can contrast $f(BMdec)$ with HML or $f(RoE)$ with RMW, a comparison we'll show shortly. Notably, Chen and Zimmermann (2022) data comprises both BM and BMdec, suggesting that using $f(BMdec)$ for comparison with HML might yield better insights than using $f(BM)$. [10] However, an intriguing observation emerges: when considering BM or BMdec in isolation, their risk premia may be positive, aligning with conventional beliefs. Yet, when analyzing all 205 characteristics simultaneously, BM's risk premia remains positive while BMdec's turns negative, both exhibiting remarkable significance. Although we will compare $f(size)$ to SMB, $f(BMdec)$ or $f(BM)$ to HML, and other factors, our factors are estimated by leveraging information from all 205 characteristics simultaneously. We will provide detailed insights and results in the upcoming sections.

In this section, we will benchmark our factors against FF6, Q5 factors, and x_t. While we will delve deeper into the relationship between our factors and x_t later on, let's briefly introduce the notation of x_t. Recall that $x_{t+1} = \frac{1}{N_t} Z_t^\top r_{t+1}$, where Z_t represents the characteristics after rank normalization and mapped into the [-1,1] range. In other words, x_t constitutes a set of long-short portfolios, with a positive position on stocks with characteristics above the median and a negative

[10] The authors of Chen and Zimmermann (2022) provide code on their website for replicating the HML factor using their data, utilizing BMdec instead of BM in their R code. https://github.com/OpenSourceAP/CrossSectionDemos/blob/main/FF1993_style_implementation.r

position on those below the median. For instance, if the first row of Z_t pertains to size, the initial elements of x_t form a portfolio with percent rank serving as the weight. This can be seen as a size factor. To create a corresponding market factor, we can introduce a constant 1_{N_t} into the characteristics without rank normalization. Subsequently, we calculate $x_{t+1} = \frac{1}{N_t} 1_{N_t}^\top r_{t+1}$ to obtain an equal-weighted market factor.

For instance, imagine a scenario with 5 stocks sorted by market size from smallest to largest. When focusing solely on the market and size factors within x_t, we would observe the following:

$$x_t = \frac{1}{5} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & -0.5 & 0 & 0.5 & 1 \end{pmatrix} r_t = \begin{pmatrix} \frac{1}{5} \sum_{i=1}^{5} r_{i,t} \\ \frac{1}{5}(-r_{1,t} - 0.5r_{2,t} + 0.5r_{4,t} + r_{5,t}) \end{pmatrix}.$$

The first element represents the return of the equal-weighted market factor, and the second element is a size factor, or at least a portfolio that is long on stocks with a large market size and short on those with a small market size. In the following section, we will use the notation $x(.)$ ito refer to the corresponding factors in in x_t. For instance, $x(size)_{t+1}$ denotes $\frac{1}{N_t} Z_t^\top r_{t+1}$, where Z_t is the size characteristic, and $x(BM)_{t+1}$ denotes $\frac{1}{N_t} Z_t^\top r_{t+1}$, where Z_t is the BM characteristic. Specifically, $x(Mkt)$ refers to the market factor in x_t with Z_t as a constant number equal to 1. To demonstrate the performance of factors in x_t, we select the market (Mkt), size, BM, RoE, AssetGrowth, and Mom12m factors in x_t and report their cumulative log returns in 6, along with the cumulative log returns of FF6 factors and Q5 factors.

[Insert Figure 6 about here]

Note that for the results to be comparable, we must standardize each factor since the appropriate scale of x_t is unknown. Factors like SMB in FF6 and ME in Q5 are derived from conventional characteristic-sorting methods, making their construction quite similar. Therefore, while we don't need to standardize the scale or standard deviation of SMB and ME, we can assume that their standard deviations are reasonably comparable, eliminating the need for standardization to compare SMB and ME. However, we lack even a basic understanding of the scale of $x(size)$ since its weight is not based on sorting. Thus, to compare $x(size)$ to SMB and ME, we must standardize each factor, along with the others. After standardization, we compute the log return using $log\left(1 \pm \frac{r_t}{100}\right)$ and compare the cumulative log return across different factors.[11]

[11] Here, we use $\pm$ instead of $+$ in the formula of the log return formula because we also don't know the sign of

Referring to Figure 6, we can compare the cumulative log return of x_t to its counterparts in the FF6 factors. For instance, we can compare $x(size)$ with SMB, given that both factors are constructed based on size characteristics albeit through different methods. $x(size)$ is constructed by $\frac{1}{N_t} Z_t^\top r_{t+1}$ while SMB involves multiple sorting processes. Surprisingly, we observe that x_t exhibits high efficiency, with many of its factors showing even greater cumulative log returns than their counterparts in the FF6 factors. Another noteworthy observation is that $x(RoE)$ appears to lack a significant cumulative log return, suggesting that the RoE characteristic might be redundant, or at the very least, less relevant in the Chen and Zimmermann (2022) dataset.

We now aim to compare our factors with those of FF6 and Q5, as well as x_t. It's important to note that our characteristics data lacks a constant, which means there is no market factor in FM with Restriction[12]. To facilitate a more meaningful comparison, we include $x(Mkt)$ alongside them. Similar to the FF6 factors, we can also propose a new six-factor model:

$$x(Mkt) + f(size) + f(BMdec) + f(RoE) + f(AssetGrowth) + f(Mom12m). \tag{20}$$

The Factor model (20) is constructed using the same characteristics as FF6, allowing us to directly compare our $f(size)$ to SMB, $f(BMdec)$ to HML and so forth. However, ours also incorporates information from all 205 characteristics, and the factors are optimized to explain both the time series and cross-sectional variations. Therefore, we might anticipate our factors to be more effective. As we will discuss in the following section, one potential source of frustration is that while the risk premium of $BMdec$ might be positive if considered in isolation, it becomes significantly negative when all characteristics are considered simultaneously, which goes against classical beliefs. To address this issue, we can replace $BMdec$ with BM, resulting in the following six-factor model:

$$x(Mkt) + f(size) + f(BM) + f(RoE) + f(AssetGrowth) + f(Mom12m). \tag{21}$$

x_t either. A factor in x_t might yield a negative return, but it can still be effective if the risk premium is significant; in such cases, we simply reverse the sign. For a factor in x_t, if the cumulative log return is positive, we calculate the log return as $log\left(1 + \frac{r_t}{100}\right)$; if negative, we compute it as $log\left(1 - \frac{r_t}{100}\right)$.

[12] Adding a constant to characteristics and training FM with Restriction is not recommended, as constants cannot be standardized. Consequently, the scale of the constant would influence the results of FM with Restriction. For instance, if we add a constant of 2 to the dataset, the outcomes would differ from adding a constant of 1 or 0, and it would also affect the selection of characteristics.

Just as Fama and French (2018) have pointed out concerns about the underlying theory of the Momentum factor, we can similarly use the following two five-factor models for comparison with the FF5 factors proposed by Fama and French (2015):

$$x(Mkt) + f(size) + f(BMdec) + f(RoE) + f(AssetGrowth), \tag{22}$$

$$x(Mkt) + f(size) + f(BM) + f(RoE) + f(AssetGrowth). \tag{23}$$

To begin, we utilize the factor model (20) to compute correlations with other factors. Classical factors such as FF6 or Q5 factors are constructed using a similar approach — multiple sorting — which can lead to a potential issue: many of these factors may be redundant. For instance, although the SMB in FF3 factors proposed by Fama and French (1992) and Fama and French (1993) differs from the SMB in FF6 factors proposed by Fama and French (2015), they exhibit nearly identical historical returns, resulting in a high correlation that suggests redundancy. This problem arises from only considering a small subset of characteristics among the larger dataset. When we don't examine all characteristics simultaneously and instead analyze only a few at a time, there are infinite combinations and thus infinite ways to construct factors. However, most of these constructions will be redundant since all infinite factors are derived from a finite set of characteristics.

In contrast, our approach involves considering all characteristics simultaneously and constructing factors through optimization only once. Therefore, we avoid similar redundancy issues. None of the selected factors are redundant, and any redundant factors are not chosen by FM with Restriction. Since all selected factors represent their own unique incremental information according to our model, we anticipate that our factors will have very low correlation with FF6 factors.

In Figure 7, we present the correlation between different factors, featuring 8 sub-figures. The first two figures display the correlation among FF6, Q5 factors, and x_t, while the subsequent six figures show the correlation among FF6, Q5 factors, x_t, and our factors, both in-sample (IS) and out-of-sample (OOS).[13] For brevity, we report only the results for $\phi^2 = 12$ here, as the results for other choices of ϕ^2 are similar.

[13]Readers can refer back to the beginning of this section for the definition of our in-sample and out-of-sample factors (refer to 3.3.1).

[Insert Figure 7 about here]

From Figure 7, we observe that the correlation between FF6 and Q5 factors can be relatively high regardless of their signs. However, the correlation between x_t and our factors is also high, whereas the correlation between FF6, Q5 factors, and our factors is notably low. This indicates that our characteristics-based factors differ significantly from classical factors. This finding is intriguing, as many contemporary factors are often similar or redundant when considered together, providing minimal additional information, unlike ours.

In Figure 8, we display the cumulative log return of our factors. The figure contains two panels: Panel A presents the in-sample results, while Panel B shows the out-of-sample results. The in-sample results are derived from training our model and calculating the factors using data from April 1962 to December 2021, whereas the out-of-sample results are obtained through a recursive procedure.[14] For comparability, we only show results from May 1977 to December 2021 and standardize each factor to have unit variance. It's worth noting that some factors, like $f(size)$, exhibit a significant negative risk premium, but for clarity, we reverse the sign in Figure 8.

As evident from Figure 8, our factors demonstrate good performance, and intuitively, all of them should have a high absolute Sharpe ratio. Moreover, the out-of-sample results, while slightly inferior to the in-sample results, still show strong performance. We notice that $BMdec$ is not selected for some small ϕ^2 in the early years. However, overall, all the factors perform quite well throughout this entire period.

To demonstrate the effectiveness of our factors, we present the Sharpe ratio of the factor-spanned pricing kernel in Table 5. Comparing the efficiency of factor-spanned pricing kernels is a critical method for evaluating factor models. For instance, Gibbons, Ross, and Shanken (1989) introduced the well-known GRS test for assessing the significance of differences between Sharpe ratios of factor-spanned pricing kernels. Table 5 shows both in-sample and out-of-sample results. Pricing kernel is one of the most important object in asset pricing, and given a group of factors f_t, the pricing kernel they spanned can be denoted as:

[14]This procedure involves training the model with data from April 1962 to April 1977 initially, constructing the factors using Equation (10) from May 1977 to April 1982, adding this data to the training sample, and repeating the process.

$$
\begin{aligned}
b &= Cov_t(f_t, f_t^\top)^{-1} E_t(f_t) \\
SDF_t &= b^\top f_t
\end{aligned}
\tag{24}
$$

The in-sample results are derived from using all historical data to calculate b and construct the Stochastic Discount Factor (SDF) in Equation (24). The out-of-sample results are obtained by first estimating the model using data from the initial 10 years, then using Equation (24) to construct the factors for the subsequent 5 years, adding this data to the training sample, repeating the process for the next 5 years, and calculating the Sharpe ratio after completing all estimations. It's worth noting that although we present both in-sample and out-of-sample results, they serve to illustrate our method for testing factor-spanned pricing kernels. Furthermore, the factors in the FM with Restriction are exclusively the out-of-sample factors, ensuring no information leakage.

[Insert Figure 8 and Table 5 about here]

Table 5 contains 3 panels. Panel A presents the classical factor models, which include FF6, FF5, Q5 factors, along with Q4 factors from Hou, Xue, and Zhang (2015). Panel B is dedicated to x_t, while Panel C focuses on factors in the FM with Restriction. Within Panel C, there are three sub-panels representing different choices of ϕ^2, and each sub-panel reports the results of factor models (20), (21), (22), and (23).

We can directly compare our factor models to classical factor models in Panel A. For instance, if we set $\phi^2 = 1$, factor model (20) exhibits an in-sample Sharpe ratio of 2.09 and an out-of-sample Sharpe ratio of 1.84, significantly outperforming FF6 with an in-sample Sharpe ratio of 1.29 and an out-of-sample Sharpe ratio of 0.8. Similar comparisons can be made between factor model (21) and FF6, or between factor models (22) and (23) and FF5, yielding consistent conclusions. We can also compare our factor models with Q5 factors. While the Q5 is indeed highly competitive, boasting an in-sample Sharpe ratio of 1.96 and an out-of-sample Sharpe ratio of 1.44, which appears superior to FF6 factors and even comparable to our factors, dropping the EG factor results in a significant performance decline. The in-sample Sharpe ratio reduces to 1.35, and the out-of-sample Sharpe ratio drops to 0.89, only marginally better than FF6.

Thus, Expected growth seems pivotal, yet since we lack this characteristic in our datasets,

a direct comparison between our factors and Q5 may not be entirely fair. Our perspective is that leveraging available characteristics allows us to construct the best factors according to those characteristics, ensuring robustness in high-dimensional settings. We emphasize that our "best" factors are contingent on the available characteristics, and while they may not encompass characteristics like expected growth, our six-factor model still outperforms Q5, and the five-factor model fares slightly worse than Q5 but significantly better than FF6 and Q4, highlighting the efficiency of our factors.

Additionally, as we utilize all 205 characteristics to derive the six or five-factor models, we can demonstrate how our factors perform when considering all of them rather than selecting just six. In section 3.2, we detail these results, show the predictive power of using all factors to forecast out-of-sample outcomes and their effectiveness in predicting individual stock behavior.

To further look into the performance of each factor, we present the risk premia along with their corresponding t values and the Sharpe ratio of different factors in Table 6. This table contains two panels: Panel A includes FF6, Q5 factors, and x_t, while Panel B covers out-of-sample factors in FM with Restrictions. Here, we directly compare our factors with classical factors, and most of the time, our factors demonstrate higher efficiency. For instance, assuming $\phi^2 = 1$, regardless of the sign, $f(size)$ exhibits a higher Sharpe ratio than SMB, $f(BMdec)$ and $f(BM)$ surpass HML, $f(AssetGrowth)$ outperforms CMA, and $f(Mom12m)$ exceeds Mom in terms of Sharpe ratio.

[Insert Table 6 about here]

As previously mentioned, an intriguing observation arises regarding the risk premia of $BMdec$, which appears negative when considering information from 205 characteristics simultaneously. It's important to note that since $x_{t+1} = \frac{1}{N_t} Z_t^\top r_{t+1}$, x_t is akin to the covariance between Z_t and r_{t+1}, thereby necessitating that x_t aligns in sign with the risk premia when considering characteristics independently. Given that $x(BMdec)$ exhibits a positive risk premia, we deduce that when considered independently, its risk premia is indeed positive. Consequently, the negative risk premia observed in FM with Restrictions stems from high-dimensional inference.

We also conduct a factor test using the traditional beta pricing model and present the results in Table 7. In this analysis, we aim to assess the explanatory power of each factor model concerning x_t. The time series regression we run is as follows:

$$x_t = \beta f_t + \epsilon_t, \tag{25}$$

where x_t is a $l \times 1$ vector, f_t is a $k \times 1$ vector (for example, if the factor model is FF6, then $k = 6$), and β is a $l \times k$ matrix. We employ time series regression to estimate β, followed by utilizing the mean of f_t, denoted as μ, as the estimated risk premia. Subsequently, we predict $\hat{x}_t$ using $\hat{\beta}\mu$ and assess its predictive capability. To quantify this predictive power, we define R^2 as:

$$R^2 = 1 - \frac{\sum_{t=1}^{T} \|x_t - \hat{\beta}\mu\|_2^2}{\sum_{t=1}^{T} \|x_t\|_2^2}. \tag{26}$$

The definition provided here differs from the one in Equation (14) as we are now testing using the traditional beta pricing model. We can also define pricing error metrics $\overline{|\alpha|}/\overline{|\mu|}$ and $\overline{\alpha^2}/\overline{\mu^2}$ in the same manner as in Equations (17) and (18). Unlike Equation (16), the key distinction is that these pricing errors now correspond to the beta pricing model version:

$$\alpha(x) = \frac{1}{T} \sum_{t=1}^{T} \left(x_t - \hat{\beta}\mu \right). \tag{27}$$

In Table 7, we present both in-sample and out-of-sample results for R^2, $\overline{|\alpha|}/\overline{|\mu|}$, and $\overline{\alpha^2}/\overline{\mu^2}$. It's important to note that a higher R^2, or smaller values of $\overline{|\alpha|}/\overline{|\mu|}$ and $\overline{\alpha^2}/\overline{\mu^2}$, indicate better predictive performance. The in-sample results are derived from using all available data (from April 1977 to December 2021) to estimate the model, while the out-of-sample results are obtained through a recursive procedure.[15]

[Insert Table 7 about here]

Table 7 comprises two panels. Panel A presents the results of classical factor models, while Panel B shows the outcomes of our out-of-sample factors. Within Panel B, we provide two sub-panels: B1 displays the results of the six factors model in Equation (20), and B2 presents the results of high-dimensional factors. As observed in Table 7, our factors consistently outperform all classical factor models. The R^2 values of our six factors model are higher, and the pricing

[15]Similar as previously, this procedure involves using the data from the first five years to run the time series regression in Equation (25), estimating risk premia, and predicting the subsequent five years. This process is then iterated by adding the data from the next five years into the training sample and predicting the following five years, continuing in this manner.

errors are smaller, indicating a superior predictive power for x_t. Additionally, the performance of the high-dimensional factors is notably better, with higher R^2 values and significantly smaller pricing errors both in the in-sample and out-of-sample analyses.

$\overline{\alpha^2}/\overline{\mu^2}$ of our high-dimensional factors appears as 0 in the in-sample test, but in reality, it is not exactly 0. Due to rounding to two decimal places, it is displayed as 0. Addtionally, although we utilize all 205 characteristics simultaneously, many of our factors are shrunk to 0. Consequently, we are not employing all 205 factors to predict 205 anomalies (x_t) here. In reality, as detailed in Table 7, we only select 133 of them for $\phi^2 = 0$, 112 for $\phi^2 = 1$, and 131 for $\phi^2 = 12$.

3.3.2 Decomposing the characteristics-based factors

Now, let's turn to decompose our factors. It's worth noting that in Equation (10), our factors can be expressed as:

$$f_{t+1}^{(k)} = B_t x_{t+1}, \tag{28}$$

where

$$B_t = \{\sum_{i=1}^{k} \alpha^{(i)} A^{(i)}(Z_t) \prod_{j=1}^{i-1} [I_{N_t} - \alpha^{(j)} Z_t A^{(j)}(Z_t)]\} N_t. \tag{29}$$

It occurs to us that understanding how FM with Restriction rotates x_t to construct efficient factors might be a beneficial approach to comprehending our factors. In this section, we will illustrate what B_t looks like and demonstrate how our factors evolve through the rotation of x_t.

Before explore that, it might be beneficial to exploit deeper into the essence of x_t. Given that our characteristics are already rank-normalized and scaled into the range [-1,1], consider selecting just one characteristic denoted as $Z_{i,t}(N_t \times 1)$. If we were to run a cross-sectional regression of r_{t+1} on $Z_{i,t}$, the resulting estimator would be:

$$\beta_{i,t+1} = (Z_{i,t}^\top Z_{i,t})^{-1} Z_{i,t}^\top r_{t+1} \overset{N_t \to +\infty}{\longrightarrow} \frac{2}{3} x_{i,t+1}.$$

Therefore, we can observe that x_t is essentially proportional to the estimator of univariate cross-sectional regression. It can interpreted as the estimator when considering each characteristic independently, as opposed to simultaneously like in FM with Restriction. The results derived from

FM with Restriction and x_t can be quite different, whereas our factors represent a time-varying rotation of x_t as per Equation (28). For instance, as previously mentioned, $x(BMdec)$ exhibits a significant positive risk premia, whereas $f(BMdec)$ shows a significant negative risk premia, emphasizing the disparity between low-dimensional inference and high-dimensional inference.

After discussing x_t, let's move on to the statistics concerning B_t. We present in Figures 9, 10, and 11 the average loading $\frac{1}{T}\sum_{t=1}^{T} B_t$ for $\phi^2 = 0$, 1, and 2, respectively. To save space, we only report the results for $f(size)$, $f(BMdec)$, $f(AssetGrowth)$, and $f(Mom12m)$. The labels at the top denote the names of our factors, while those on the left indicate the factors in x_t. The height of the bar represents the absolute value of $\frac{1}{T}\sum_{t=1}^{T} B_t$, with blue bars indicating positive loadings and red bars indicating negative loadings.

[Insert Figures 9, 10, and 11 about here]

Before interpreting those three figures, it's important to clarify two points. First, recall that if we run a cross-sectional univariate regression independently, we will obtain $\beta_{i,t+1} = \frac{2}{3}x_{i,t+1}$. So, if we were to create a similar figure for univariate cross-sectional regression, it would essentially show that $f(size)$ places a weight of exactly $\frac{2}{3}$ on $x(size)$, with the weight on other factors in x_t being exactly 0. This figure would appear straightforward, and the difference between this simple figure and Figures 9, 10, and 11 arises from considering numerous characteristics simultaneously. Second, the only distinction among Figures 9, 10, and 11 is the value of ϕ^2. For instance, comparing Figure 9 with Figure 11allows us to explore how we can rotate x_t differently to simultaneously explain the variation in time series and cross-section.

From Figures 9, 10, and 11, it's evident that the outcomes of considering numerous characteristics can differ significantly from considering them independently. For instance, while $f(BMdec)$ places a large positive weight on $x(BMdec)$, it also assigns a substantial negative weight to $x(BM)$. Referring back to Table 6, where both $x(BMdec)$ and $x(BM)$ exhibit significant positive risk premia, the significant risk premia of $f(BMdec)$ and the substantial negative weight on $x(BM)$ might explain this. This finding is intriguing because, to our knowledge, the book-to-market ratio in much classical literature refers to $BMdec$ instead of BM here.[16] Since these studies do not

[16]While Exploring the exact definition of all characteristics might exceed the scope of this paper, we refer to the website of Chen and Zimmermann (2022) again. They provide code on their website illustrating how to use $BMdec$ to replicate HML in FF3 factors.

consider both $BMdec$ and BM, along with many other characteristics, they conclude that the risk premia of $BMdec$ is positive.[17]

Comparing Figures 9, 10, and 11 allows us to discern the difference between solely considering time series variation and also taking into account cross-sectional variation. For instance, when $\phi^2 = 0$, we observe that $f(size)$ places a large positive weight on $x(size)$ and $x(DolVol)$, with much smaller weights on other factors in x_t. However, if we opt for $\phi^2 = 1$, we notice that $x(DolVol)$ is no longer prominent in the figure, and $x(Illiquidity)$ replaces its position. Now, $f(size)$ begins to allocate a large positive weight on $x(Illiquidity)$, suggesting that incorporating $x(Illiquidity)$ into the size factor could be advantageous in predicting cross-sectional returns.

3.4 Interpreting the risk premia

In this subsection, we present the risk premia associated with all characteristics estimated using FM with Restrictions. For the data from Lewellen (2015), we set ϕ^2 to 3, while for data from Chen and Zimmermann (2022), we use ϕ^2 of 12. We will begin by comparing our interpretations with those of other models.

Before looking at the results, it's worth noting some expectations for each model. First, regarding Fama-MacBeth (FM), we should temper our expectations as it may not align perfectly with prior beliefs due to potential overfitting issues. Several studies advocate for incorporating penalties in cross-sectional regression, such as Han, He, Rapach, and Zhou (2024), Gu, Kelly, and Xiu (2020), and Martin and Nagel (2021). We won't reiterate those arguments here. Second, as discussed in Section 2, although we can use LASSO (specifically CS-LASSO) instead of OLS for robustness, it primarily targets explaining time series variation. In contrast, pooled LASSO, which pools data first before applying the algorithm, is expected to capture cross-sectional variation better. Consequently, CS-LASSO may favor short-term characteristics due to their volatility and potential to explain time series variation, whereas pooled LASSO may lean towards long-term characteristics. Third, FM with Restrictions aims to predict both time series and cross-sectional variation. It occupies a middle ground between CS-LASSO and pooled LASSO, likely displaying

[17] We do not aim to argue whether $BMdec$ or BM is superior for constructing the value factor or whether the risk premia should be positive, as we understand that historically, researchers might have preferred $BMdec$ for computational ease due to limited computing power in the past. However, our insight is that if the results suggest a positive weight on $BMdec$ and a negative weight on BM, it might imply that changes in BM are also meaningful.

a nuanced preference for long-term versus short-term characteristics.

Additionally, it's important to note that the preferences we discussed earlier refer to selection preferences, which may differ from the importance when predicting the cross-section. The calculation of risk premia in this section follows the same procedure as in Section 3.2 or Equation (11), with the difference being the use of an expanding window for simplicity and omitting the subscript t. For instance, methods like CS-LASSO may favor selecting short-term characteristics, such as event types, for optimization. However, when calculating risk premia to predict the cross-section, these characteristics may not prove as crucial. This highlights the distinction between predicting time series and cross-sectional outcomes. Using a method optimized for time series prediction to predict the cross-section may lead to discrepancies and potential violations.

In Figures 12 and 13, we present the estimated risk premia for the data from Lewellen (2015) and Chen and Zimmermann (2022), respectively. For brevity, we display the results for the 15 characteristics with the largest absolute risk premia in Figure 13. Given that all data are already rank-normalized, higher absolute risk premia indicate greater importance of a characteristic. Here, we employ the Bayesian Information Criterion (BIC) rule, as done previously, to determine the number of selected characteristics.

[Insert Figures 12 and 13 about here]

The results depicted in Figures 12 and 13 align with our anticipated outcomes. Regardless of the dataset used, the estimated risk premia derived from the Fama-MacBeth model appear quite strange; the magnitude is disproportionately large compared to other models. For instance, the Fama-MacBeth regression estimates the risk premia of DivOmit in Chen and Zimmermann (2022) to be approximately 50% per month. Given that our characteristics have already been rank normalized and constrained to the [-1,1] range, this would imply an anticipated equal-weighted long-short return for DivOmit of 90% per month, a figure that seems unrealistically high. Such an inflated scale in estimation often signifies overfitting. To address this, methods like ridge regression or LASSO are commonly employed to shrink the estimator and mitigate these unrealistic magnitudes.

Additionally, we observe that FM with Restriction and CS-LASSO exhibit a preference for short-term characteristics over pooled LASSO. For instance, based on data from Lewellen (2015),

the top two characteristics for CS-LASSO and FM with Restriction are turn and mom12m, whereas for pooled LASSO, turn ranks second and mom12m ranks fourth. In contrast, long-term characteristics like Logsize rank fourth for CS-LASSO, third for FM with Restriction, and first for pooled LASSO.

The scenario with Chen and Zimmermann (2022) data is similar. Chen and Zimmermann (2022) have categorized all 205 characteristics into 8 groups based on data sources: 13F, Accounting, Analyst, Event, Options, Other, Price, and Trading. While we won't delve into definitions here, their website provides detailed explanations.[18] Instinctively, Event-type characteristics are extremely short-term. For instance, consider defining a consistent payer as a company that has paid dividends quarterly or semi-annually for the past 18 months, or annually for the past 2 years. Then, the Event-type characteristic, DivOmit indicates if there were omissions for such consistent payers. Omitting dividends could raise investor suspicions, likely leading to negative risk premia. These types of characteristics often signal rare events and are undoubtedly extremely short-term.

Characteristics falling under Price and Trading categories are also short-term but not as fleeting as Event-type ones. Price-type examples include Mom12m, IndRetBig, MaxRet, and so forth, while Trading-type examples encompass Illiquidity, DolVol, std_turn, and similar metrics. Accounting-type characteristics may be even more distinct, including BM, BMdec, and others. Analyst-type characteristics are derived from analyst forecasts and recommendations, mostly representing forecasted accounting values. Accounting and Analyst types tend to be relatively long-term. It is worth noting that since CS-LASSO aims to explain time series variation comprehensively, it might favor short-term characteristics like those in Event or Price categories. Conversely, pooled LASSO might lean towards Accounting-type characteristics.

Based on Chen and Zimmermann (2022) data, Figure 13 illustrates that the top three characteristics for CS-LASSO are IndIPO, DivOmit, and ExchSwitch, all falling under the Event type category. Following these are IO_Shortinterest, IndRetBig, and MaxRet, which also belong to short-term categories. In contrast, the top characteristic for pooled LASSO is IndRetBig, a short-term metric, but the subsequent two are BM and ChNAnalyst, which are relatively long-term. For FM with Restriction, the top characteristic is DivOmit, followed by AM, representing the total asset to market size and notably a long-term metric. MaxRet and IndRetBig, both short-term

[18] https://www.openassetpricing.com/faq/#q-datacatwebsite

indicators, occupy the third and fourth positions. Comparing these model results, CS-LASSO shows a strong preference for short-term characteristics, while pooled LASSO leans least towards them, with FM with Restriction falling between the two in terms of preference for short-term attributes.

Given that Chen and Zimmermann (2022) categorized characteristics, we can assess the importance of each category across different models. After obtaining the estimated risk premia μ for any model, we can make predictions using $\hat{r}t+1 = Z_t\mu$, where Z_t represents the characteristic values at time t. We define $R^2 = 1 - \frac{\sum t=1^{T-1}|r_{t+1}-\hat{r}t+1|2^2}{\sum t=1^{T-1}|rt+1|_2^2}$ to measure prediction accuracy. To evaluate a category's importance, we can isolate it and measure the resulting drop in prediction accuracy. In other words, to assess the significance of category i ($i \in 13F, Accounting, Analyst, Event, Options, Other, Price, Trading$), we set the risk premia of category i to 0, denoted as $\mu(i)$, and predict using $\hat{r}t+1(i) = Z_t\mu(i)$. This leads us to define the new R^2 as $R^2(i) = 1 - \frac{\sum t=1^{T-1}|r_{t+1}-\hat{r}t+1(i)|2^2}{\sum t=1^{T-1}|rt+1|_2^2}$. Then, we define:

$$L(i) = 1 - \frac{R^2(i)}{R^2}. \tag{30}$$

$L(i)$ represents the percentage loss of R^2 after excluding category i. A larger $L(i)$ signifies the greater importance of category i in predicting cross-sectional returns. Conversely, a smaller $L(i)$ suggests lower importance of category i in predicting the cross section accurately. However, it's worth noting that in practice, $L(i)$ can often be smaller than 0. If $L(i) < 0$, it implies that excluding category i improves the prediction, indicating that category i is not only redundant but may even lead to misleading predictions.

In Figure 14, we observe the significance of each category for each model. Our initial assumptions hold true for the most part, except for Event-type characteristics. Short-term characteristics, such as those in Price or Trading types, are heavily relied upon by CS-LASSO, followed by FM with Restriction and then pooled LASSO, which places the least reliance on them. Conversely, long-term characteristics like Accounting and Analyst types are most crucial for pooled LASSO, followed by FM with Restriction and then CS-LASSO.

[Insert Figure 14 about here]

The intriguing aspect is the behavior of Event-type characteristics. As mentioned earlier,

CS-LASSO prioritizes short-term characteristics to explain time series variation but may not effectively explain cross-sectional returns. Figure 13 confirms that CS-LASSO's top three characteristics are Event type. However, in Figure 14, we notice that removing Event type characteristics improves R^2, suggesting CS-LASSO's extreme preference for Event type characteristics leads to misleading cross-sectional return predictions. CS-LASSO optimizes its estimator for time series returns, resulting in inflated risk premia for short-term characteristics that may not contribute significantly to cross-sectional predictions. In contrast, Our model — FM with Restriction balances both time series and cross-sectional variation. While it also favors short-term characteristics, its preference is more tempered. It avoids excessively large risk premia for Event type characteristics that could misguide predictions. Instead, we provide moderate estimations and leverage short-term characteristics efficiently, ensuring comprehensive explanation of both time series and cross-sectional variations. Notably, $L(Event)$ for our model is even twice as large as that of pooled LASSO, indicating our successful utilization of these predictors.

The estimated risk premia and associated t values for FM with Restriction are presented in Tables 8 and 9. Since our model offers a latent factors solution, we can easily compute the t value following the methodologies outlined in Newey and West (1987, 1994).

[Insert Tables 8 and 9 about here]

Tables 8 and 9 arrange characteristics according to the absolute value of their risk premia. Therefore, a higher rank indicates greater importance for a characteristic. It's evident that almost all top-ranked factors exhibit a significant t value, even after considering Newey-West adjustment.

3.5 Robustness checks

In earlier sections, we selected $\phi^2 = 3$ for Lewellen (2015) data and $\phi^2 = 12$ for Chen and Zimmermann (2022) data. Now, we'll present the out-of-sample results across a broad range of ϕ^2 values in this subsection.

In Figures 15, 16, 17, 18, 19, 20, and 21, we present the out-of-sample $R^2(r_t)$, $R^2(x_t)$, β_{slope}, $\overline{|\alpha|}/\overline{|\mu|}$, $\overline{\alpha^2}/\overline{\mu^2}$, mean return, and Sharpe ratio of long-short portfolios, respectively, for different choices of ϕ^2.

We won't delve into a detailed analysis of each figure here, but it's important to note that

our selection of ϕ^2 is not biased to favor our model's performance. In fact, while the conclusions drawn from different metrics may vary, our model consistently performs well across a wide range of ϕ^2 values. It's worth mentioning that $\phi^2 = 0$ generally yields the poorest performance in most cases, and many metrics show optimized results with a suitable choice of ϕ^2. This indicates the importance of considering both time series and cross-sectional variations rather than focusing solely on one aspect.

[Insert Figures 15, 16, 17, 18, 19, 20, and 21 about here]

In Online Appendix, we present the out-of-sample results on the dataset from Gu, Kelly, and Xiu (2020), which contains 94 characteristics. The detailed description of these characteristics is available in Gu, Kelly, and Xiu (2020). Our model's performance on this dataset is highlighted, reaffirming its superiority. The out-of-sample test mirrors the format of Tables 1, 2, and 3.[19]

4 Conclusion

To make the Fama-MacBeth regression applicable in high-dimension settings, we introduce two additional constraints. The first constraint ensures that the model effectively captures both time series and cross-sectional variations. The second constraint mandates consistent sparsity of factors over time. We propose a piecewise linear algorithm to obtain the solution of the FM regression under these constraints. The solution is, as usual, a tradable portfolio with predetermined loadings.

Empirically, we find that the modified FM regression exhibits superior performance in predicting cross-sectional returns at both the individual stock level and for anomalies. It consistently achieves the highest out-of-sample R^2 and the smallest pricing errors. Additionally, its long-short portfolio has the highest mean return and Sharpe ratio compared with existing models.

Furthermore, the modified FM regression produces the most significant alphas even after controlling for classical factors such as CAPM, FF6, and Q5. Additionally, the resulting characteristics-based factors show remarkable performance compared with their counterparts, and the estimated efficient portfolio achieves a notably higher Sharpe ratio compared to other models.

[19]Note that we also conduct a recursive out-of-sample test using data from April 1962 to December 2021). Readers seeking more information can refer to the preceding section for additional insights.

The Fama-MacBeth regression has long been a cornerstone in empirical studies, for its clear economic interpretation. By extending it with economically and econometrically meaningful constraints, we demonstrate that the modified framework can continue to play a central role in empirical asset pricing in the era of machine learning in analyzing high-dimensional data.

References

Back, K., Kapadia, N., Ostdiek, B., 2015. Testing factor models on characteristic and covariance pure plays (july 16, 2015). Available at SSRN: https://ssrn.com/abstract=2621696 or http://dx.doi.org/10.2139/ssrn.2621696 .

Barillas, F., Shanken, J., 2017. Which alpha? Review of Financial Studies 30, 1316–1338.

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

Bookstein, F. L., 1975. On a form of piecewise linear regression. The American Statistician 29, 116–117.

Chen, A. Y., Zimmermann, T., 2022. Open source cross-sectional asset pricing. Critical Finance Review 27, 207–264.

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

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. The Annals of statistics 32, 407–499.

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

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

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

Fama, E. F., French, K. R., 2018. Choosing factors. Journal of Financial Economics 128, 234–252.

Fama, E. F., French, K. R., 2020. Comparing cross-section and time-series factor models. Review of Financial Studies 33, 1891–1926.

Fama, E. F., MacBeth, J. D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81, 607–636.

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

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

Green, J., Hand, J. R., Zhang, X. F., 2017. The characteristics that provide independent information about average us monthly stock returns. Review of Financial Studies 30, 4389–4436.

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

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

Han, Y., He, A., Rapach, D. E., Zhou, G., 2024. Cross-sectional expected returns: new Fama–MacBeth regressions in the era of machine learning. Review of Finance p. rfae027.

Hastie, T., Tibshirani, R., Friedman, J., 2017. The elements of statistical learning: data mining, inference, and prediction. Springer, New York, NY.

Hastie, T., Tibshirani, R., Wainwright, M., 2015. Statistical learning with sparsity. Chapman & Hall/CRC.

Hou, K., Mo, H., Xue, C., Zhang, L., 2020. An Augmented q-Factor Model with Expected Growth*. Review of Finance 25, 1–41.

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

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

Kelly, B. T., Pruitt, S., Su, Y., 2020. Instrumented principal component analysis. Available at SSRN 2983919 .

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

Lettau, M., Pelger, M., 2020a. Estimating latent asset-pricing factors. Journal of Econometrics 218, 1–31.

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

Lewellen, J., 2015. The cross-section of expected stock returns. Critical Finance Review 4, 1–44.

Martin, I. W., Nagel, S., 2021. Market efficiency in the age of big data. Journal of Financial Economics .

McZgee, V. E., Carleton, W. T., 1970. Piecewise regression. Journal of the American Statistical Association 65, 1109–1124.

Newey, W. K., West, K. D., 1987. A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55, 703–708.

Newey, W. K., West, K. D., 1994. Automatic Lag Selection in Covariance Matrix Estimation. The Review of Economic Studies 61, 631–653.

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

Simon, N., Tibshirani, R., 2012. Standardization and the group lasso penalty. Statistica Sinica 22, 983.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58, 267–288.

Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68, 49–67.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology) 67, 301–320.

Table 1: Out-of-sample performance with different moving average length

In this table, we report the out-of-sample performance including R^2, β_{slpoe}, the mean absolute pricing error ($\overline{|\alpha|}/\overline{|\mu|}$) and the mean square pricing error ($\overline{\alpha^2}/\overline{\mu^2}$). The results based on the data provided by Lewellen (2015) and Chen and Zimmermann (2022) are shown in Panel A and Panel B, respectively. The out-of-sample period is from April 1977 through December 2021. The R^2 here is in percentile. Here, we let $\phi^2 = 3$ for Panel A and $\phi^2 = 12$ for Panel B. The definitions of the metrics are provided in subsection 3.2.1.

(1) MA length	(2) Model	(3) $R^2(r_t)$	(4) $R^2(x_t)$	(5) β_{slpoe}	(6) $\overline{\lvert\alpha\rvert}/\overline{\lvert\mu\rvert}$	(7) $\overline{\alpha^2}/\overline{\mu^2}$
Panel A: Lewellen (2015) data with 15 characteristics						
Expanding	Fama-MacBeth	-0.8	-4.4	0.15	2.18	6.46
	Cs-lasso	0.07	0.88	**0.93**	0.47	0.24
	Pooled lasso	0.06	**0.97**	0.68	0.5	0.29
	FM with Restriction	**0.08**	0.94	0.89	**0.4**	**0.21**
120 months	Fama-MacBeth	-1.53	-6.93	0.22	1.19	1.83
	Cs-lasso	0.03	0.59	**0.87**	0.41	0.17
	Pooled lasso	0.05	0.92	0.61	0.44	0.23
	FM with Restriction	**0.07**	**0.92**	0.82	**0.3**	**0.1**
60 months	Fama-MacBeth	-2.71	-15.68	0.22	1.38	2.37
	Cs-lasso	-0.01	0.36	**0.61**	0.56	0.29
	Pooled lasso	0.01	0.5	0.43	0.43	0.21
	FM with Restriction	**0.04**	**0.57**	0.57	**0.32**	**0.12**
Panel B: Chen and Zimmermann (2022) data with 205 characteristics						
Expanding	Fama-MacBeth	-541.06	-507.46	-0	14.62	110.84
	Cs-lasso	-0.15	-0.64	**1.04**	0.56	0.19
	Pooled lasso	0.09	-0.6	0.62	0.68	0.8
	FM with Restriction	**0.62**	**2.52**	1.1	**0.31**	**0.06**
120 months	Fama-MacBeth	-8.84	-1.02	0.65	0.58	0.24
	Cs-lasso	0.63	2.25	1.06	**0.24**	0.04
	Pooled lasso	0.18	0.73	0.62	0.6	0.68
	FM with Restriction	**0.66**	**2.27**	**1**	0.25	**0.04**
60 months	Fama-MacBeth	-12.34	-3.8	0.56	0.44	0.15
	Cs-lasso	0.56	1.96	**0.96**	0.27	0.06
	Pooled lasso	-2.3	-15.32	0.44	1.12	2.71
	FM with Restriction	**0.58**	**2.11**	0.89	**0.23**	**0.04**

Table 2: The out-of-sample performance of long-short portfolios with value-weighting

In this table, we report the out-of-sample performance of long-short portfolios. If we regard $\hat{r}_{t+1}$ as a signal measurable at time t, we can construct decile portfolios every month from April 1977 to December 2021 by sorting stocks based on $\hat{r}_{t+1}$. Portfolio 1(10) is the portfolio of stocks with the lowest (highest) $\hat{r}_{t+1}$. The table reports the value-weighted next-month averages excess return, standard deviation (SD), as well as the annualized Sharpe ratio (SR), for each decile. H-L denotes the 10-1 return spread and its associated standard deviation (SD) together with the annualized Sharpe ratio (SR). The results based on the data provided by Lewellen (2015) and Chen and Zimmermann (2022) are shown in Panel A and Panel B, respectively. Here, we let $\phi^2 = 3$ for Panel A and $\phi^2 = 12$ for Panel B.

	Fama-MacBeth			Cs-lasso			Pooled lasso			FM with Restriction		
	mean(%)	SD(%)	SR	mean(%)	SD(%)	SR	mean(%)	SD(%)	SR	mean(%)	SD(%)	SR
Panel A: Lewellen (2015) data with 15 characteristics												
1(L)	0.44	6.58	0.23	0.32	5.60	0.20	0.36	5.16	0.24	0.25	5.56	0.16
2	0.57	5.49	0.36	0.52	4.99	0.36	0.67	4.91	0.47	0.55	4.67	0.41
3	0.66	5.15	0.44	0.67	4.91	0.47	0.78	5.02	0.54	0.75	4.59	0.56
4	0.79	5.23	0.52	0.92	4.94	0.64	0.97	5.14	0.65	0.81	4.61	0.61
5	0.86	5.1	0.59	0.87	5	0.61	0.99	5.13	0.67	1.02	4.88	0.72
6	0.95	5.08	0.65	0.99	5.23	0.66	1.01	5.35	0.65	0.96	5.03	0.66
7	0.84	5.07	0.58	1.15	5.4	0.74	1.21	5.61	0.75	1.13	5.37	0.73
8	0.98	5.39	0.63	1.17	5.9	0.69	1.25	5.62	0.77	1.09	5.55	0.68
9	0.98	5.35	0.63	1.27	6.37	0.69	1.4	6.08	0.8	1.29	6.24	0.71
10(H)	1.23	5.67	0.75	1.52	7.13	0.74	1.49	6.84	0.76	1.57	7.84	0.69
H-L	0.78	5.13	0.53	1.19	5.37	0.77	1.13	4.88	**0.8**	**1.31**	5.96	0.76
Panel B: Chen and Zimmermann (2022) data with 205 characteristics												
1(L)	0.12	5.58	0.08	-0.11	6.46	-0.06	-0.12	6.35	-0.06	-0.1	5.79	-0.06
2	0.52	4.97	0.36	0.3	5.16	0.2	0.17	5.06	0.11	0.3	4.91	0.21
3	0.55	4.57	0.42	0.57	4.87	0.4	0.49	4.75	0.36	0.54	4.54	0.41
4	0.62	4.74	0.45	0.64	4.59	0.49	0.57	4.6	0.43	0.59	4.59	0.45
5	0.83	4.65	0.62	0.77	4.52	0.59	0.79	4.49	0.61	0.8	4.57	0.6
6	0.92	4.77	0.67	0.97	4.63	0.73	0.87	4.54	0.67	0.98	4.57	0.74
7	1.02	4.78	0.74	1.22	4.81	0.88	1.02	4.59	0.77	1.14	4.7	0.84
8	1.11	4.85	0.79	1.27	4.9	0.9	1.22	4.81	0.88	1.33	5.09	0.91
9	1.17	4.86	0.84	1.36	5.34	0.88	1.38	4.98	0.96	1.49	5.37	0.96
10(H)	1.35	5.76	0.81	1.91	6.06	1.09	1.62	5.68	0.98	2.07	5.96	1.2
H-L	1.23	4.39	0.97	2.02	5.24	1.34	1.73	4.88	1.23	**2.17**	4.56	**1.65**

Table 3: Alphas for long-short portfolios based on Lewellen (2015)'s data

In this table, we report the alphas and their associated t-values of long-short portfolios for different models. The alphas are computed using three different factor models: the CAPM; the six-factor model of Fama and French (2018), denoted FF6; and the five-factor q model of Hou, Mo, Xue, and Zhang (2020), denoted Q5. If we regard $\hat{r}_{t+1}$ as a signal measurable at time t, we can construct decile portfolios every month from April 1977 to December 2021 by sorting stocks based on $\hat{r}_{t+1}$. Portfolio 1(10) is the portfolio of stocks with the lowest (highest) $\hat{r}_{t+1}$. H-L denotes the differences in monthly returns and the associated t-values.

	Fama-MacBeth						Cs-lasso					
	α_{FF6}	$t(\alpha_{FF6})$	α_Q	$t(\alpha_Q)$	α_{CAPM}	$t(\alpha_{CAPM})$	α_{FF6}	$t(\alpha_{FF6})$	α_Q	$t(\alpha_Q)$	α_{CAPM}	$t(\alpha_{CAPM})$
1(L)	0.38	1.27	0.43	1.34	0.41	1.42	0.06	0.22	0.2	0.69	0.18	0.7
2	0.53	2.11	0.64	2.39	0.53	2.19	0.4	1.77	0.54	2.24	0.49	2.23
3	0.56	2.38	0.71	2.81	0.6	2.66	0.58	2.6	0.7	2.91	0.62	2.91
4	0.74	3.1	0.95	3.73	0.75	3.26	0.85	3.75	0.98	4.04	0.87	4.05
5	0.79	3.4	0.94	3.78	0.8	3.6	0.8	3.53	0.96	3.89	0.83	3.81
6	0.9	3.92	1.08	4.35	0.92	4.16	0.94	3.92	1.17	4.56	0.97	4.26
7	0.75	3.22	0.92	3.69	0.79	3.59	1.1	4.44	1.32	4.98	1.11	4.72
8	0.97	3.95	1.14	4.3	0.94	4.01	1.11	4.08	1.32	4.53	1.13	4.38
9	0.92	3.75	0.99	3.78	0.93	3.99	1.17	4.01	1.45	4.61	1.22	4.4
10(H)	1.18	4.54	1.31	4.71	1.19	4.79	1.47	4.43	1.69	4.76	1.45	4.58
H-L	0.8	3.37	0.88	3.47	0.78	3.46	1.41	5.41	1.49	5.35	1.27	5.13
	Pooled lasso						FM with Restriction					
	α_{FF6}	$t(\alpha_{FF6})$	α_Q	$t(\alpha_Q)$	α_{CAPM}	$t(\alpha_{CAPM})$	α_{FF6}	$t(\alpha_{FF6})$	α_Q	$t(\alpha_Q)$	α_{CAPM}	$t(\alpha_{CAPM})$
1(L)	0.26	1.1	0.36	1.41	0.31	1.38	0.11	0.42	0.19	0.7	0.2	0.83
2	0.58	2.6	0.75	3.14	0.64	2.96	0.46	2.18	0.58	2.56	0.51	2.48
3	0.66	2.89	0.84	3.43	0.72	3.3	0.68	3.25	0.84	3.74	0.71	3.52
4	0.89	3.81	1.05	4.17	0.93	4.16	0.73	3.48	0.92	4.05	0.79	3.91
5	0.89	3.81	1.08	4.3	0.96	4.29	0.97	4.37	1.11	4.65	0.98	4.59
6	0.92	3.77	1.12	4.27	0.98	4.17	0.91	3.95	1.1	4.45	0.94	4.27
7	1.08	4.2	1.29	4.66	1.19	4.83	1.04	4.22	1.26	4.78	1.09	4.63
8	1.13	4.4	1.37	4.95	1.25	5.07	0.99	3.91	1.22	4.45	1.06	4.38
9	1.25	4.51	1.45	4.85	1.38	5.2	1.27	4.43	1.55	5.05	1.25	4.58
10(H)	1.27	4.06	1.42	4.19	1.45	4.86	1.61	4.47	1.82	4.69	1.54	4.5
H-L	1.01	4.53	1.06	4.36	1.14	**5.34**	**1.51**	**5.49**	**1.62**	**5.52**	**1.34**	5.15

Table 4: Alphas for long-short portfolios based on Chen and Zimmermann (2022)'s data

In this table, we report the alphas and their associated t-values of long-short portfolios for different models. The alphas are computed using three different factor models: the CAPM; the six-factor model of Fama and French (2018), denoted FF6; and the five-factor q model of Hou, Mo, Xue, and Zhang (2020), denoted Q5. If we regard $\hat{r}_{t+1}$ as a signal measurable at time t, we can construct decile portfolios every month from April 1977 to December 2021 by sorting stocks based on $\hat{r}_{t+1}$. Portfolio 1(10) is the portfolio of stocks with the lowest (highest) $\hat{r}_{t+1}$. H-L denotes the differences in monthly returns and the associated t-values.

	Fama-MacBeth						Cs-lasso					
	α_{FF6}	$t(\alpha_{FF6})$	α_Q	$t(\alpha_Q)$	α_{CAPM}	$t(\alpha_{CAPM})$	α_{FF6}	$t(\alpha_{FF6})$	α_Q	$t(\alpha_Q)$	α_{CAPM}	$t(\alpha_{CAPM})$
Low(L)	-0.33	-3.15	-0.17	-1.46	-0.68	-6.01	-0.74	-5.68	-0.41	-3	-1.02	-7.38
2	-0.19	-2.54	-0.01	-0.18	-0.24	-3.21	-0.4	-4.75	-0.23	-2.56	-0.47	-5.68
3	-0.2	-3.23	-0.16	-2.37	-0.16	-2.54	-0.2	-2.8	-0.04	-0.51	-0.17	-2.49
4	-0.16	-2.59	-0.19	-2.8	-0.11	-1.89	-0.09	-1.58	-0.08	-1.22	-0.07	-1.23
5	-0.02	-0.29	-0.06	-0.81	0.12	1.88	0.01	0.23	0	-0.02	0.07	1.24
6	0.12	1.71	0.1	1.26	0.2	2.86	0.2	2.81	0.13	1.73	0.27	3.92
7	0.19	2.51	0.12	1.43	0.3	3.95	0.44	5.55	0.31	3.75	0.5	6.64
8	0.24	3	0.13	1.55	0.38	4.82	0.4	4.41	0.23	2.38	0.55	6.19
9	0.34	3.91	0.23	2.42	0.45	5.33	0.52	5.04	0.39	3.51	0.58	5.74
High(H)	0.49	3.82	0.41	2.91	0.55	4.28	1.03	6.67	0.81	4.97	1.12	7.21
H-L	0.82	4.29	0.58	2.78	1.23	6.39	**1.76**	7.55	**1.22**	5	2.14	9.42
	Pooled lasso						FM with Restriction					
	α_{FF6}	$t(\alpha_{FF6})$	α_Q	$t(\alpha_Q)$	α_{CAPM}	$t(\alpha_{CAPM})$	α_{FF6}	$t(\alpha_{FF6})$	α_Q	$t(\alpha_Q)$	α_{CAPM}	$t(\alpha_{CAPM})$
Low(L)	-0.57	-5.79	-0.4	-3.63	-1	-8.59	-0.51	-4.73	-0.26	-2.23	-0.92	-7.74
2	-0.45	-6.02	-0.42	-5.1	-0.6	-7.92	-0.31	-4.08	-0.14	-1.76	-0.44	-5.94
3	-0.22	-3.51	-0.17	-2.48	-0.24	-4.06	-0.17	-2.8	-0.1	-1.56	-0.16	-2.74
4	-0.21	-3.48	-0.2	-3.08	-0.15	-2.51	-0.14	-2.44	-0.12	-1.96	-0.12	-2.33
5	0.03	0.54	0.03	0.42	0.1	1.77	0.02	0.29	0.02	0.35	0.09	1.57
6	0.1	1.79	0.05	0.72	0.17	3.05	0.13	2.08	0.04	0.55	0.27	4.44
7	0.24	3.67	0.17	2.41	0.32	5.01	0.29	3.76	0.13	1.65	0.44	5.79
8	0.32	4	0.22	2.58	0.51	6.14	0.44	4.83	0.26	2.67	0.58	6.32
9	0.48	5.26	0.43	4.33	0.65	7.02	0.56	5.32	0.43	3.8	0.71	6.82
High(H)	0.71	5.41	0.71	4.98	0.81	6.14	1.1	8.24	0.89	6.22	1.25	9.08
H-L	1.28	6.92	1.12	5.43	1.81	9.3	1.61	**8.31**	1.15	**5.62**	**2.17**	**10.88**

Table 5: The Sharpe ratios for factors spanned pricing kernel by different factor models

In this table, we report the Sharpe ratios for factors spanned pricing kernel for each factor model. Panel A reports the result of four different factor models including: the five-factor model of Fama and French (2015), denoted FF5; the six-factor model of Fama and French (2018), denoted FF6; the four-factor q model of Hou, Xue, and Zhang (2015), denoted Q4; and the five-factor q model of Hou, Mo, Xue, and Zhang (2020), denoted Q5. Panel B reports the results of x_t and we choose 4 set of factors to report. Note that Chen and Zimmermann (2022) data contains both BM and BMdec, we report the results of both of them. Panel C reports the results of out-of-sample factors for FM with Restriction, while we use three sub-panel to report the results for different ϕ^2. In each sub-panel, we also report the result of four set of factors. The name of factors of FM with Restriction and x_t here are the same as the original data from Chen and Zimmermann (2022), so readers might find that we use RoE instead of ROE, Mom12m instead of Mom.

	In sample	Out of sample
Panel A: Classical factors models		
FF5: Mkt + SMB+ HML + RMW + CMA	1.20	0.75
FF6: Mkt + SMB+ HML + RMW + CMA + Mom	1.29	0.8
Q4: Mkt + ME + ROE + IA	1.35	0.89
Q5: Mkt + ME + ROE + IA + EG	1.96	1.44
Panel B: $x_{t+1} = \frac{1}{N_t} Z_t^\top r_{t+1}$		
x(Mkt) + x(size) + x(BMdec) + x(RoE) + x(AssetGrowth) + x(Mom12m)	1.87	1.19
x(Mkt) + x(size) + x(BM) + x(RoE) + x(AssetGrowth) + x(Mom12m)	2.29	1.54
x(Mkt) + x(size) + x(BMdec) + x(RoE) + x(AssetGrowth)	1.79	1.14
x(Mkt) + x(size) + x(BM) + x(RoE) + x(AssetGrowth)	1.82	1.20
Panel C: Out-of-sample factors for FM with Restriction		
Panel C1: Result for $\phi^2 = 0$		
x(Mkt) + f(size) + f(BMdec) + f(RoE) + f(AssetGrowth) + f(Mom12m)	1.78	1.40
x(Mkt) + f(size) + f(BM) + f(RoE) + f(AssetGrowth) + f(Mom12m)	**2.06**	**1.46**
x(Mkt) + f(size) + f(BMdec) + f(RoE) + f(AssetGrowth)	1.46	1.11
x(Mkt) + f(size) + f(BMdec) + f(RoE) + f(AssetGrowth)	1.78	1.18
Panel C2: Result for $\phi^2 = 1$		
x(Mkt) + f(size) + f(BMdec) + f(RoE) + f(AssetGrowth) + f(Mom12m)	2.09	**1.84**
x(Mkt) + f(size) + f(BM) + f(RoE) + f(AssetGrowth) + f(Mom12m)	**2.22**	1.65
x(Mkt) + f(size) + f(BMdec) + f(RoE) + f(AssetGrowth)	1.55	1.36
x(Mkt) + f(size) + f(BMdec) + f(RoE) + f(AssetGrowth)	1.73	1.18
Panel C3: Result for $\phi^2 = 12$		
x(Mkt) + f(size) + f(BMdec) + f(RoE) + f(AssetGrowth) + f(Mom12m)	2.12	**1.69**
x(Mkt) + f(size) + f(BM) + f(RoE) + f(AssetGrowth) + f(Mom12m)	**2.18**	1.46
x(Mkt) + f(size) + f(BMdec) + f(RoE) + f(AssetGrowth)	1.64	1.25
x(Mkt) + f(size) + f(BMdec) + f(RoE) + f(AssetGrowth)	1.72	1.04

Table 6: Summary statistics for different factor models

In this table, we report the risk premia, its t value and the annualized Sharpe ratio for each factor model. There are 2 panels in the table. Panel A reports the results of the six-factor model of Fama and French (2018) (FF6), the five-factor q model of Hou, Mo, Xue, and Zhang (2020) (Q5), and $x_{t+1} = \frac{1}{N_t} Z_t^\top r_{t+1}$. Panel B reports the results of the our of sample factors for FM with Restriction. Note that since Chen and Zimmermann (2022) data contains both BM and BMdec, we report the results including both of them. Here, the name of factors of FM with Restriction and x_t are the same as the original data from Chen and Zimmermann (2022), thus we use RoE instead of ROE and Mom12m instead of Mom.

Panel A: Classical factors model

	Fama and French 6 factors				Q5 factors				$x_{t+1} = \frac{1}{N_t} Z_t^\top r_{t+1}$			
	risk premia	t	SR		risk premia	t	SR		risk premia	t	SR
Mkt	0.72	3.63	0.56	Mkt	0.72	3.63	0.56	x(Mkt)	0.01	3.41	0.59
SMB	0.18	1.44	0.22	ME	0.24	1.83	0.28	x(size)	-0.16	-2.98	-0.46
HML	0.17	1.13	0.20	ROE	0.56	4.60	0.74	x(BMdec)	0.06	2.01	0.32
RMW	0.37	3.32	0.55	IA	0.29	3.41	0.52	x(BM)	0.21	5.41	0.89
CMA	0.24	2.61	0.43	EG	0.75	7.96	1.27	x(RoE)	0.01	0.16	0.03
Mom	0.58	2.97	0.46					x(AssetGrowth)	0.15	6.17	1.03
								x(Mom12m)	0.12	2.36	0.36

Panel B: Out of sample factors for FM with Restriction

	$\phi^2 = 0$			$\phi^2 = 1$			$\phi^2 = 12$		
	risk premia	t	SR	risk premia	t	SR	risk premia	t	SR
f(size)	-0.49	-3.64	-0.66	-0.60	-4.80	-0.74	-0.65	-4.80	-0.74
f(BMdec)	-0.08	-5.75	-0.93	-0.50	-8.60	-1.34	-0.57	-9.33	-1.42
f(BM)	0.34	7.51	1.13	0.68	8.90	1.30	0.71	8.69	1.27
f(RoE)	0.02	2.51	0.37	0			0		
f(AssetGrowth)	0.11	6.28	0.94	0.19	4.83	0.74	0.20	4.92	0.76
f(Mom12m)	0.41	7.34	1.08	0.52	7.42	1.12	0.54	7.58	1.13

Table 7: Factor test on $x_{t+1} = \frac{1}{N_t} Z_t^\top r_{t+1}$

In this table, we report both the in-sample and out-of-sample performance of each factor model for predicting 205 anomalies. Panel A reports the results of the six-factor model of Fama and French (2018) (FF6), the five-factor q model of Hou, Mo, Xue, and Zhang (2020) (Q5). Panel B reports results of FM with Restriction. We use all the 205 characteristics to train the model and in Panel B1 we report the results of only using size, BM, RoE, AssetGrowth, and Mom12m factors to explain the anomalies, while in Panel B2 reporting the results of using all 205 factors (many of them are shrunk to zero since not being selected) for explaining.

	In sample			Out of sample		
	$R^2(\%)$	$\overline{\lvert\alpha\rvert}/\overline{\lvert\mu\rvert}$	$\overline{\alpha^2}/\overline{\mu^2}$	R^2	$\overline{\lvert\alpha\rvert}/\overline{\lvert\mu\rvert}$	$\overline{\alpha^2}/\overline{\mu^2}$
Panel A: Classical factors model						
FF6: Mkt + SMB + HML + RMW + CMA + MOM	2.01	0.57	0.16	1.75	0.53	0.13
Q5: Mkt + ME + IA + ROE + EG	2.01	0.54	0.16	1.75	0.51	0.14
Panel B: Factors in FM with Restriction (out of sample factors estimated in the rolling test)						
Panel B1: Results with x(Mkt) + f(size) + f(BMdec) + f(RoE) + f(AssetGrowth) + f(Mom12m)						
$\phi^2 = 0$	2.07	0.54	0.14	1.98	0.45	0.1
$\phi^2 = 1$ (RoE not selected)	2.05	0.52	0.14	1.95	0.46	0.11
$\phi^2 = 12$ (RoE not selected)	2.08	0.52	0.13	1.96	0.45	0.1
Panel B2: Results with all 205 factors (Not all the characteristics or factors are selected) and $x(Mkt)$						
$\phi^2 = 0$ (Select 133 in sample, 121 out of sample totally)	2.39	0.07	0.00	2.04	0.26	0.04
$\phi^2 = 1$ (Select 112 in sample, 145 out of sample totally)	2.39	0.07	0.00	2.05	0.25	0.04
$\phi^2 = 12$ (Select 131 in sample, 154 out of sample totally)	2.4	0.06	0.00	2.05	0.25	0.04

Table 8: Estimated risk premia of 11 selected out of 15 characteristics provided by Lewellen (2015)

In this table, we report estimated risk premia and associated t values of 11 selected out of 15 characteristics provided by Lewellen (2015). The order is ranked by the absolute risk premia. The sample period is from April 1962 through December 2021. here, we let $\phi^2 = 3$.

rank	characteristics	risk premia	t value
1	mom12m	0.27	5.55***
2	turn	-0.27	-6.86***
3	LogSize	-0.22	-4.66***
4	Accruals	-0.16	-4.18***
5	sp	0.16	3.29***
6	DY	-0.14	-3.63***
7	mom36m	-0.13	-3.5***
8	ROA	0.11	3.65***
9	beta	0.11	1.68
10	StdDev	-0.04	-0.57
11	LogBM	-0.01	-0.32

Table 9: Estimated risk premia of 131 selected out of 205 characteristics provided by Chen and Zimmermann (2022)

In this table, we report estimated risk premia and associated t values of 131 selected out of 205 characteristics provided by Chen and Zimmermann (2022). The order is ranked by the absolute risk premia. The sample period is from April 1962 through December 2021. here, we let $\phi^2 = 12$.

rank	characteristics	risk premia	t value
1	DivOmit	-1	-3.65***
2	AM	0.99	8.64***
3	MaxRet	0.83	13.92***
4	IndRetBig	0.71	13.13***
5	BMdec	-0.67	-10.89***
6	Mom12m	0.63	10.03***
7	BM	0.62	8.1***
8	size	-0.6	-4.81***
9	TrendFactor	0.57	12.83***
10	FirmAgeMom	0.48	8.44***
11	roaq	0.48	9.01***
12	VolMkt	-0.46	-4.67***
13	AnnouncementReturn	0.45	11.02***
14	RD	0.4	6.47***
15	RDIPO	-0.39	-5.03***
16	ChTax	0.35	10***
17	std_turn	0.34	8.55***
18	Cash	0.29	8.76***
19	IndMom	0.29	7.4***
20	BookLeverage	0.29	6.27***
21	Spinoff	0.29	9.03***
22	ShortInterest	0.25	6.58***
23	zerotradeAlt12	0.24	3.49***
24	High52	-0.24	-2.62**
25	CompEquIss	0.23	5.91***
26	IdioRisk	-0.23	-4.77***
27	OperProf	0.23	5.92***
28	MomSeason	0.22	8.04***
29	IntanCFP	0.22	5.84***
30	Frontier	0.21	4.38***
31	AssetGrowth	0.21	5.14***
32	DivYieldST	0.2	14.16***
33	PS	-0.19	-4.8***
34	SmileSlope	0.19	5.93***
35	RevenueSurprise	0.18	7.86***
36	EarningsStreak	0.18	6.69***
37	ResidualMomentum	-0.15	-6.02***
38	CustomerMomentum	0.15	3.51***
39	MomOffSeason	0.14	3.24***
40	AnalystRevision	0.14	7.72***
41	EarnSupBig	0.14	4.39***
42	OPLeverage	0.13	2.22*
43	MomSeason06YrPlus	0.13	5.45***

Table 9 (continued)

rank	characteristics	risk premia	t value
44	ChangeInRecommendation	0.13	6.96***
45	retConglomerate	0.13	5.23***
46	MomOffSeason06YrPlus	0.13	4.75***
47	zerotradeAlt1	0.13	2.16*
48	OrgCap	0.12	4.15***
49	ShareIss5Y	-0.12	-5.44***
50	EP	-0.11	-4.25***
51	Beta	0.11	2.25*
52	MeanRankRevGrowth	-0.1	-4.47***
53	CashProd	-0.1	-3.6***
54	VolSD	-0.1	-2.38*
55	MomSeasonShort	0.09	3.08***
56	tang	0.09	3.19***
57	OperProfRD	0.09	1.92
58	HerfBE	0.09	3.78***
59	SurpriseRD	-0.09	-3.79***
60	BPEBM	-0.09	-4.47***
61	IntanBM	-0.08	-2.83**
62	Mom12mOffSeason	0.08	1.86
63	ChNAnalyst	0.07	1.83
64	Illiquidity	0.07	1.15
65	PctAcc	0.07	2.44*
66	ChInvIA	0.07	3.3***
67	NetPayoutYield	-0.07	-2.87**
68	NetDebtPrice	0.07	1.57
69	MRreversal	0.07	2.01*
70	EntMult	0.07	2
71	VolumeTrend	0.07	2.95**
72	BidAskSpread	-0.06	-2.31*
73	AgeIPO	0.06	1.86
74	BetaTailRisk	0.06	2.25*
75	Coskewness	0.06	2.22*
76	betaVIX	0.06	2.96**
77	AnalystValue	-0.06	-1.96
78	IdioVolAHT	0.06	2.09*
79	EquityDuration	-0.06	-1.59
80	InvestPPEInv	0.06	2.4*
81	GrSaleToGrInv	0.05	2.67**
82	FEPS	0.05	1.07
83	NumEarnIncrease	0.05	3.03***
84	DownRecomm	-0.05	-5.45***
85	GrLTNOA	0.05	2.26*
86	Tax	-0.05	-2.84**

Table 9 (continued)

rank	characteristics	risk premia	t value
87	EarningsConsistency	0.04	2.75**
88	MomSeason11YrPlus	0.04	2.77**
89	BetaFP	0.04	0.96
90	ChNNCOA	0.04	2.22*
91	RDcap	0.04	1.65
92	RIO_Disp	-0.04	-3.53***
93	CitationsRD	-0.04	-0.83
94	CBOperProf	0.04	0.94
95	Accruals	0.04	1.39
96	hire	-0.04	-2.61**
97	EarningsForecastDisparity	0.03	2.45*
98	IntanEP	0.03	2.1*
99	SP	-0.03	-0.38
100	GrSaleToGrOverhead	0.03	1.88
101	ConsRecomm	-0.03	-2.05*
102	ForecastDispersion	0.03	1.62
103	fgr5yrLag	-0.02	-1.77
104	AdExp	0.02	1.16
105	AOP	0.02	2.68**
106	PayoutYield	-0.02	-1.49
107	MomOffSeason11YrPlus	0.02	1.89
108	MomSeason16YrPlus	0.02	1.94
109	IntMom	-0.02	-2.24*
110	DolVol	0.02	1.02
111	GP	0.02	0.38
112	PriceDelaySlope	-0.02	-1.76
113	NetDebtFinance	0.02	0.65
114	realestate	0.01	1.65
115	UpRecomm	-0.01	-3.85***
116	iomom_cust	-0.01	-0.77
117	ReturnSkew3F	-0.01	-2.59**
118	ReturnSkew	-0.01	-0.4
119	BetaLiquidityPS	-0.01	-1.2
120	XFIN	0.01	0.27
121	FR	0.01	2.02*
122	DivInit	0.01	3.02***
123	FirmAge	0.01	1.37
124	DelBreadth	0.01	1.11
125	DelFINL	0.01	0.29
126	grcapx	0.01	0.59
127	ChAssetTurnover	0	0.72
128	NOA	0	0.12
129	sinAlgo	0	1.69
130	skew1	0	2.08*
131	PatentsRD	0	0.05

Table 10: Out-of-sample performance in different sub-samples

In this table, we report the out-of-sample performance including R^2, β_{slpoe}, the mean absolute pricing error ($\overline{|\alpha|}/\overline{|\mu|}$) and the mean square pricing error ($\overline{\alpha^2}/\overline{\mu^2}$). The results based on the data provided by Lewellen (2015) and Chen and Zimmermann (2022) are shown in Panel A and Panel B, respectively. The out-of-sample period is from April 1977 through December 2021. The R^2 here is in percentile. Here, we let $\phi^2 = 3$ for Panel A and $\phi^2 = 12$ for Panel B. The definitions of the metrics are provided in subsection 3.2.1.

(1) Sample	(2) Model	(3) $R^2(r_t)$	(4) $R^2(x_t)$	(5) β_{slpoe}	(6) $\overline{\lvert\alpha\rvert}/\overline{\lvert\mu\rvert}$	(7) $\overline{\alpha^2}/\overline{\mu^2}$
Panel A: Lewellen (2015) data with 15 characteristics						
Before 2003	Fama-MacBeth	-0.77	-3.78	0.19	2.06	6.52
	Cs-lasso	0.1	1.34	1.26	0.55	0.34
	Pooled lasso	**0.13**	**1.71**	**0.97**	**0.19**	**0.04**
	FM with Restriction	0.11	1.43	1.18	0.45	0.28
After 2003	Fama-MacBeth	-0.84	-5.49	0.08	2.61	7.39
	Cs-lasso	-0.01	0.05	0.47	0.98	1.02
	Pooled lasso	-0.08	-0.34	0.28	1.26	1.68
	FM with Restriction	**-0.01**	**0.08**	**0.49**	**0.96**	**1.01**
Top 20% size	Fama-MacBeth	-0.15	-0.76	0.1	2.1	5.88
	Cs-lasso	-0.02	-0.2	**0.35**	**0.95**	0.94
	Pooled lasso	-0.06	-0.25	0.13	1.13	1.5
	FM with Restriction	**-0.01**	**-0.16**	0.22	0.96	**0.9**
Bottom 20% size	Fama-MacBeth	0.03	1.19	0.92	0.59	0.43
	Cs-lasso	0.08	1.41	2.33	0.73	0.52
	Pooled lasso	**0.15**	**2.91**	**0.96**	**0.36**	**0.1**
	FM with Restriction	0.1	1.74	1.57	0.57	0.37
Panel B: Chen and Zimmermann (2022) data with 205 characteristics						
Before 2003	Fama-MacBeth	-13.02	-4.26	0.55	0.93	0.68
	Cs-lasso	0.68	1.78	**1.18**	**0.28**	**0.04**
	Pooled lasso	0.17	-1.79	0.68	0.77	1.22
	FM with Restriction	**0.73**	**2.23**	1.2	0.32	0.05
After 2003	Fama-MacBeth	0.53	3.54	0.76	0.45	0.11
	Cs-lasso	0.54	3.06	**0.94**	**0.44**	**0.11**
	Pooled lasso	0.01	1.24	0.49	0.78	0.44
	FM with Restriction	**0.54**	**3.19**	0.87	0.5	0.12
Top 20% size	Fama-MacBeth	-2.68e6	-7.36e6	0.03	724.08	4.31e5
	Cs-lasso	**0.5**	**1.6**	1.74	0.44	0.17
	Pooled lasso	-0.71	-2.43	-0.04	1.17	1.41
	FM with Restriction	0.48	1.58	**1.14**	**0.44**	**0.17**
Bottom 20% size	Fama-MacBeth	-2.59e7	-2.62e8	0	3.03e3	8.37e6
	Cs-lasso	0.36	0.22	1.23	**0.11**	0.01
	Pooled lasso	-2.15	-34.8	4.53	0.5	0.24
	FM with Restriction	**0.69**	**4.76**	**1.1**	0.12	**0.01**

Figure 1: The case that factor j is shrunk to 0 forever: corner solution

Here, we assume that there are only two periods, and then the restriction can be written as $\sqrt{f_{j,1}^2 + f_{j,2}^2} + \sum_{i \neq j} R(i) \leq R$. The circular cone shows the restriction in the left figure and we shows the projection on the AOB plane in the right figure where the AOB plane is selected arbitrarily.

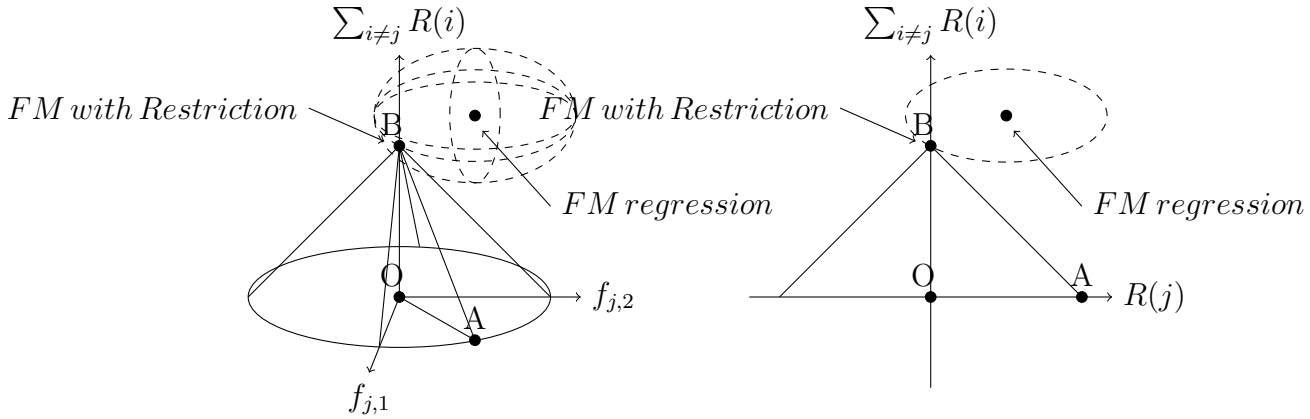


Figure 2: The case that factor j is never 0

Here, we assume that there are only two periods, and then the restriction can be written as $\sqrt{f_{j,1}^2 + f_{j,2}^2} + \sum_{i \neq j} R(i) \leq R$. The circular cone shows the restriction in the left figure and we show the projection on the AOB plane in the right figure where the object meets the restriction on the AOB plane.

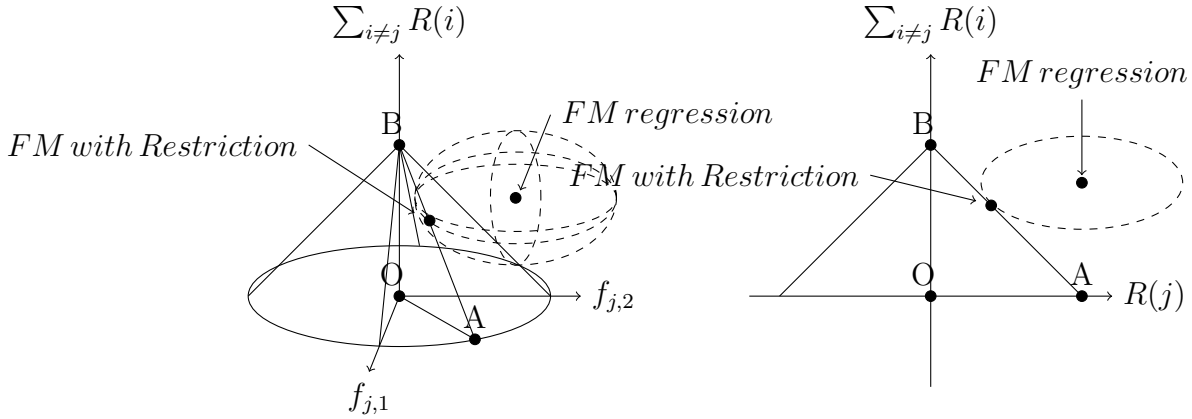
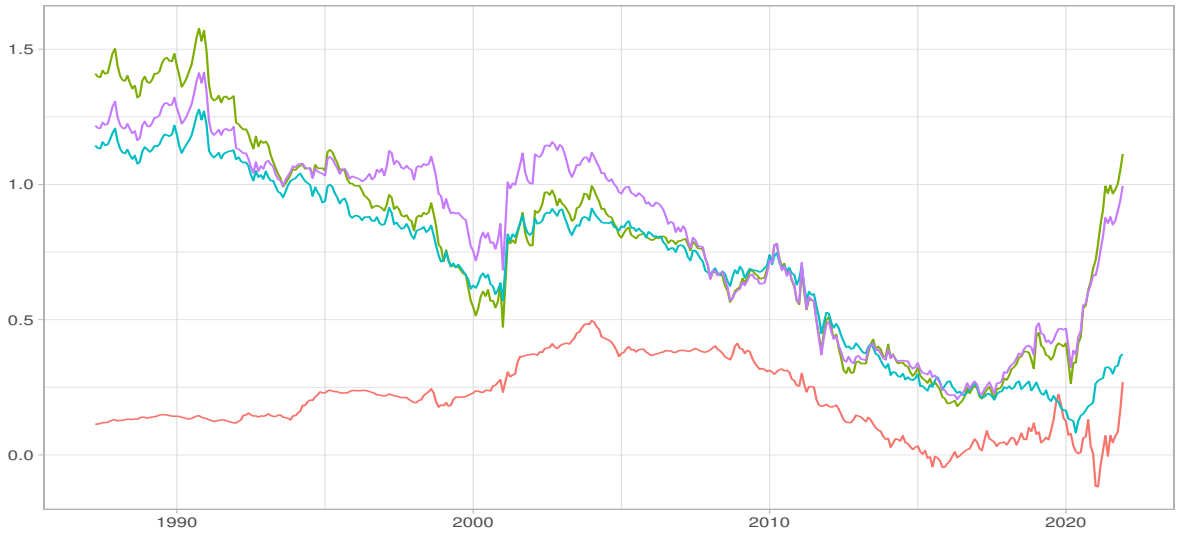
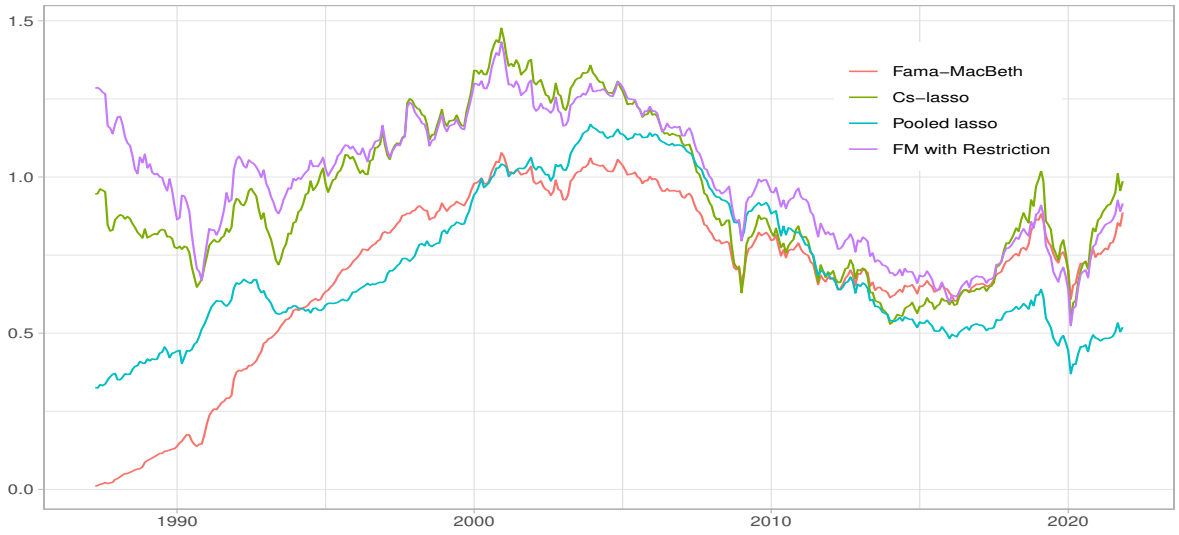


Figure 3: Out-of-sample β_{slope} with 120 months moving average estimation

In this figure, we report the out-of-sample β_{slope} from April 1987 through December 2021. β_{slope} is estimated through 120 months moving average. The figure at the top shows the result for Lewellen (2015) data and the one at the bottom is for Chen and Zimmermann (2022) data. The horizontal axis denotes the time and vertical axis denotes the β_{slope} value.



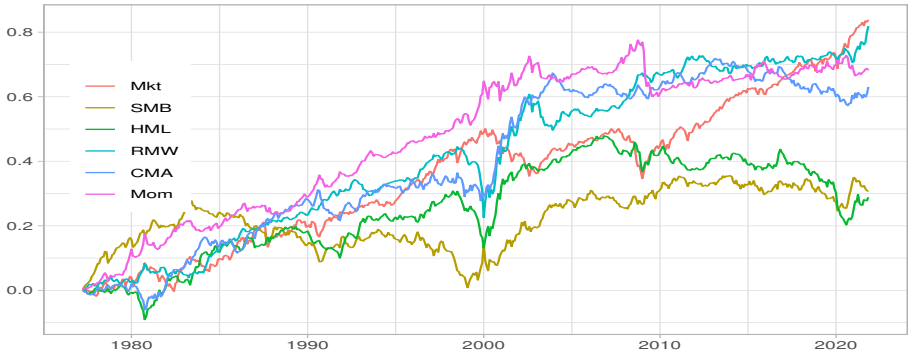
(a) Lewellen (2015) data



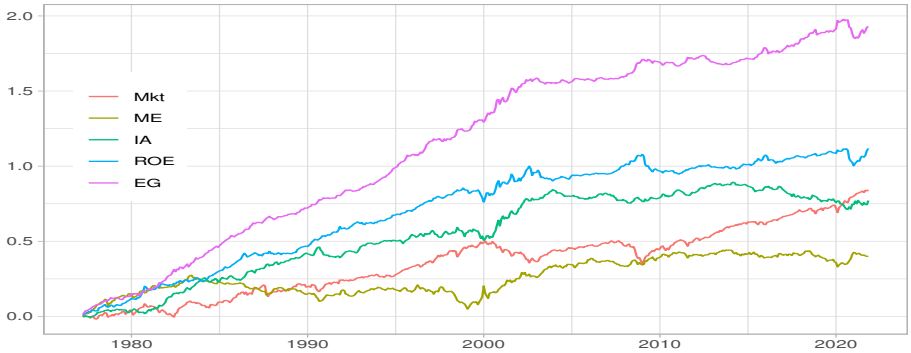
(b) Chen and Zimmermann (2022) data

Figure 6: Cumulative log factors returns of FF6, Q5, and x_t

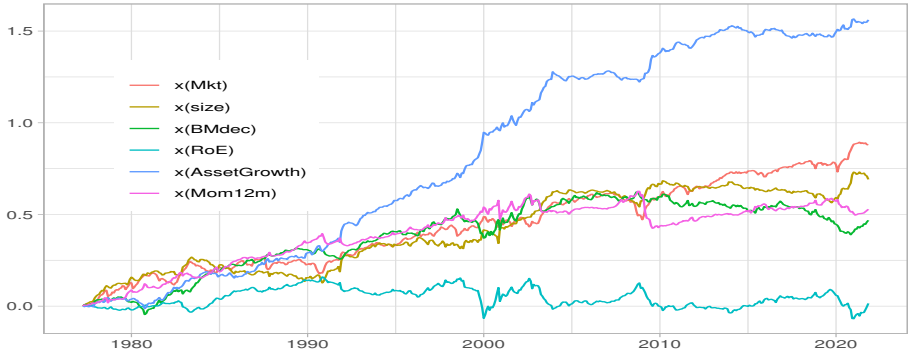
In this figure, we report the cumulative log returns for the six factors (FF6) proposed by Fama and French (2018), the five factors (Q5) proposed by Hou, Mo, Xue, and Zhang (2020) , and $x_{t+1} = \frac{1}{N_t} Z_t^\top r_{t+1}$ from April 1977 through December 2021. To make the factors comparable, we standardize each factor to make the standard deviation equals 1 and calculate the log return as $log\left(1 + \frac{r_t}{100}\right)$.



(a) Fama and French 6 factors



(b) Q5 factors



(c) $x_{t+1} = \frac{1}{N_t} Z^\top r_{t+1}$

Figure 7: Correlation of FF6 factors, Q5 factors, $x_{t+1} = \frac{1}{N_t} Z_t^\top r_{t+1}$, and factors in FM with Restriction

In this figure, we report the correlation between the six factors (FF6) proposed by Fama and French (2018), the five factors (Q5) proposed by Hou, Mo, Xue, and Zhang (2020), $x_{t+1} = \frac{1}{N_t} Z_t^\top r_{t+1}$, and factors in FM with Restriction. The name of factors in x_t and FM with Restriction is the same with the original characteristics in Chen and Zimmermann (2022) data. We let $\phi^2 = 12$ here. IS denotes in-sample while OOS means out-of-sample.

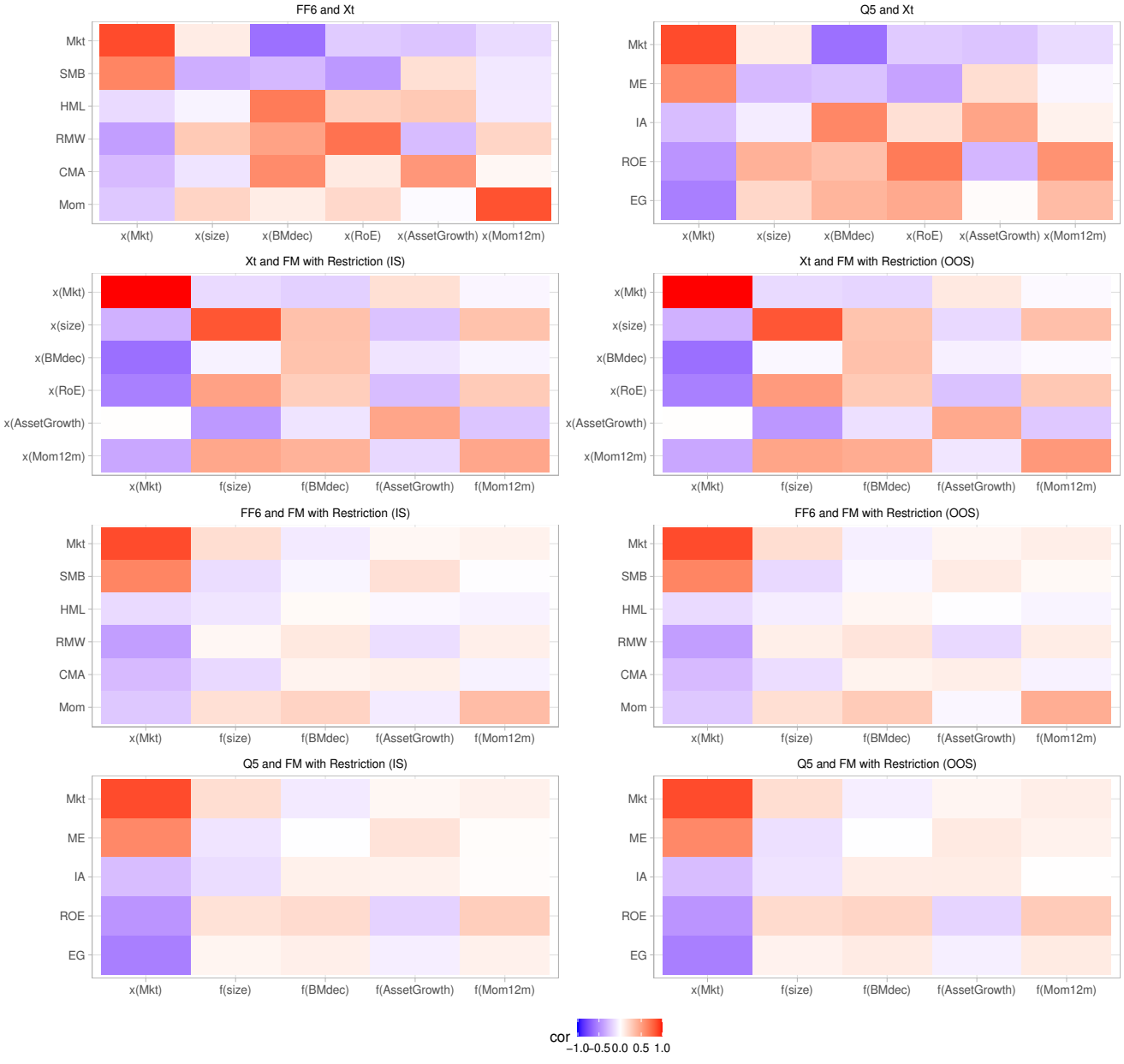
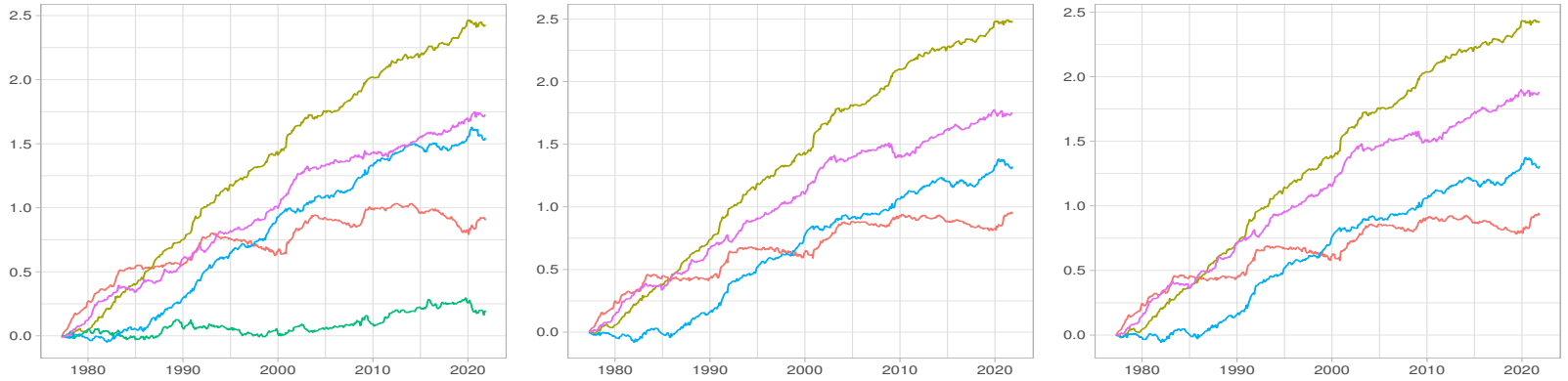


Figure 8: In-sample and out-of-sample cumulative log return for each factor

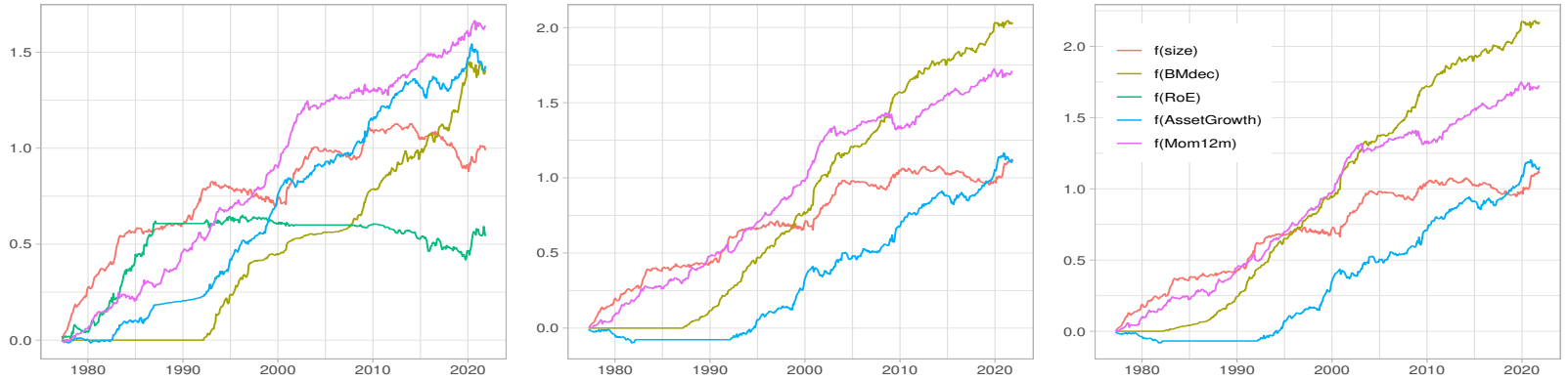
In this figure, we report the cumulative log returns for $f(size)$, $f(BMdec)$, $f(RoE)$, $f(AssetGrowth)$, and $f(Mom12m)$. We use all the 205 characteristics in Chen and Zimmermann (2022) data to train the model and only show the results of 5 factors here. Panel A reports the in-sample results. We use all the data to train FM with Restriction and use Equation 10 to construct factors. Panel B reports the out-of-sample results. RoE is missing because it was not chosen for the in-sample test when $\phi^2 = 1$ or 12.

Panel A: In-sample cumulative log factor returns from April 1977 to December 2021



(a) $\phi^2 = 0$ (b) $\phi^2 = 1$ (c) $\phi^2 = 12$

Panel B: Out-of-sample cumulative log factor returns from April 1977 to December 2021



(d) $\phi^2 = 0$ (e) $\phi^2 = 1$ (f) $\phi^2 = 12$

Figure 9: Average loading ($\frac{1}{T}\sum_{t=1}^{T} B_t$) for out-of-sample factors when $\phi^2 = 0$

In this figure, we show the average loading $\frac{1}{T}\sum_{t=1}^{T-1} B_t$ of $f(size)$, $f(BMdec)$, $f(AssetGrowth)$, and $f(Mom12m)$ with $\phi^2 = 0$ and to save some space, we only report the results of 15 factors with the largest absolute loading. The label on the top denotes our factor. For example, $f(size)$ denotes our size factor, and the label on the left, such as $x(size)$ denotes the size factor in x_t. The height of the corresponding bar denotes how much average loading does our factor put on x_t. The blue bar means the loading is positive while the red bar means the loading is negative.

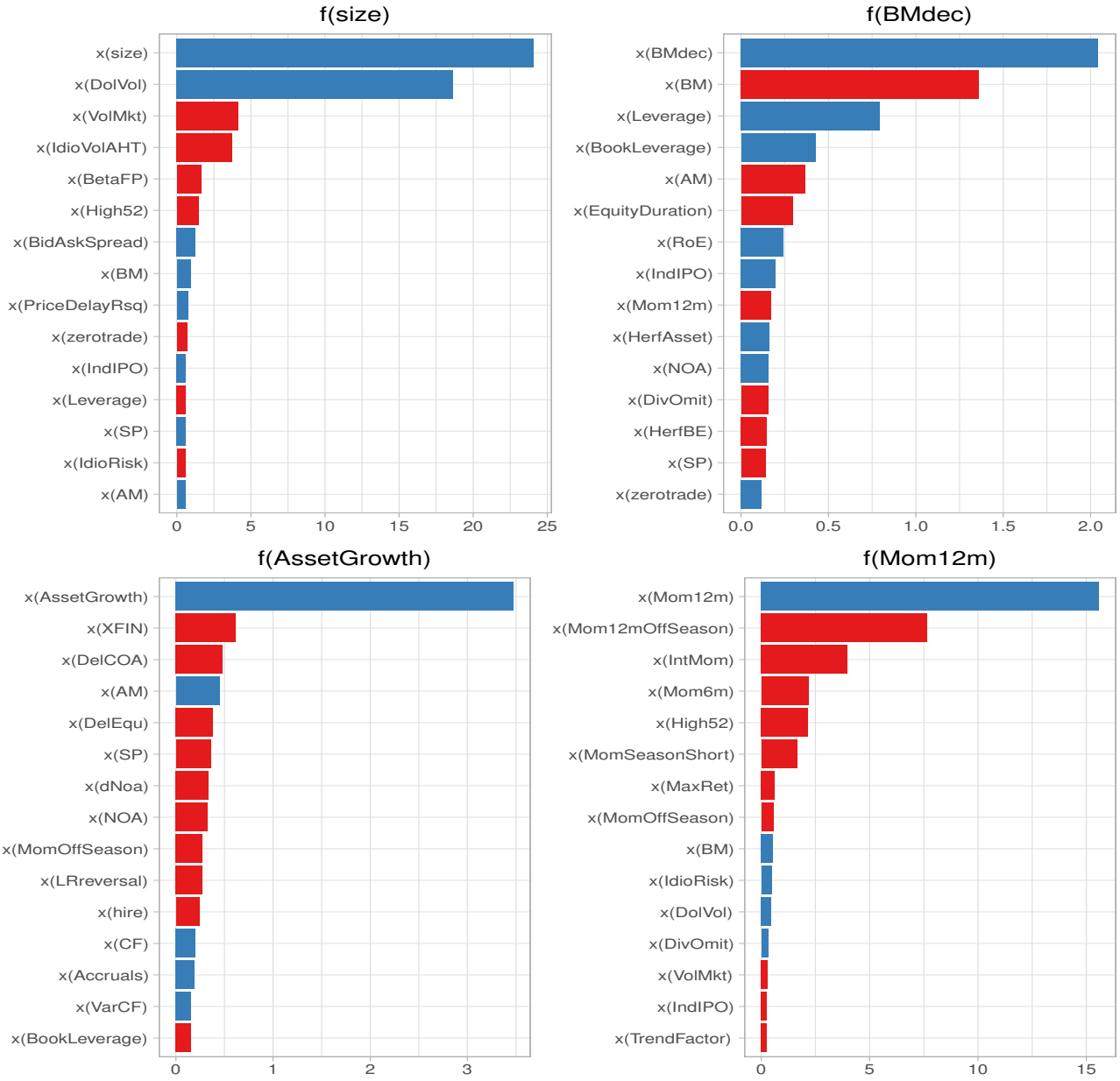


Figure 10: Average loading ($\frac{1}{T}\sum_{t=1}^{T} B_t$) for out-of-sample factors when $\phi^2 = 1$

In this figure, we show the average loading $\frac{1}{T}\sum_{t=1}^{T-1} B_t$ of $f(size)$, $f(BMdec)$, $f(AssetGrowth)$, and $f(Mom12m)$ with $\phi^2 = 1$ and to save some space, we only report the results of 15 factors with the largest absolute loading. The label on the top denotes our factor. For example, $f(size)$ denotes our size factor, and the label on the left, such as $x(size)$ denotes the size factor in x_t. The height of the corresponding bar denotes how much average loading does our factor put on x_t. The blue bar means the loading is positive while the red bar means the loading is negative.

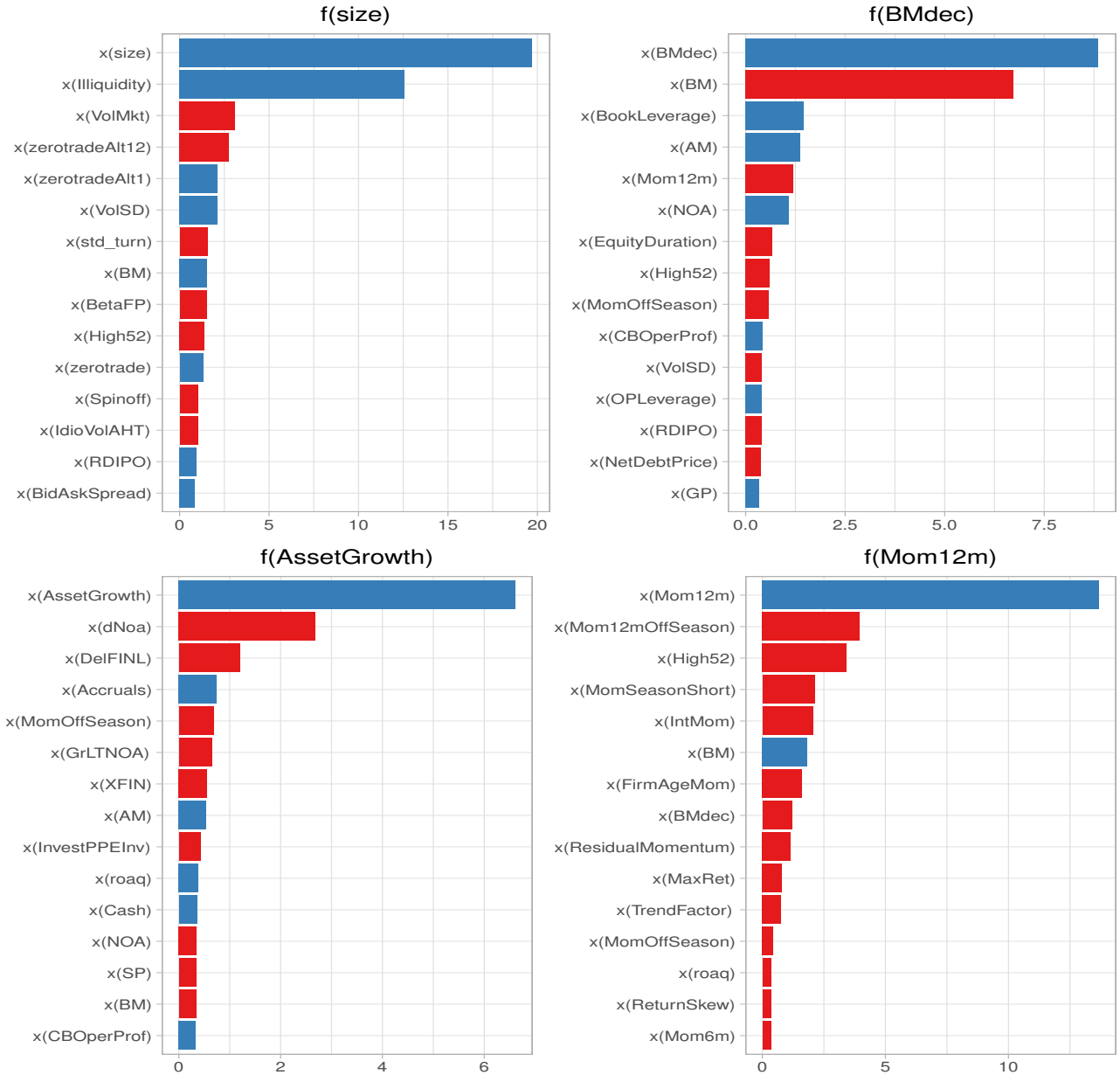


Figure 11: Average loading ($\frac{1}{T}\sum_{t=1}^{T} B_t$) for out-of-sample factors when $\phi^2 = 12$

In this figure, we show the average loading $\frac{1}{T}\sum_{t=1}^{T-1} B_t$ of $f(size)$, $f(BMdec)$, $f(AssetGrowth)$, and $f(Mom12m)$ with $\phi^2 = 1$ and to save some space, we only report the results of 15 factors with the largest absolute loading. The label on the top denotes our factor. For example, $f(size)$ denotes our size factor, and the label on the left, such as $x(size)$ denotes the size factor in x_t. The height of the corresponding bar denotes how much average loading does our factor put on x_t. The blue bar means the loading is positive while the red bar means the loading is negative.

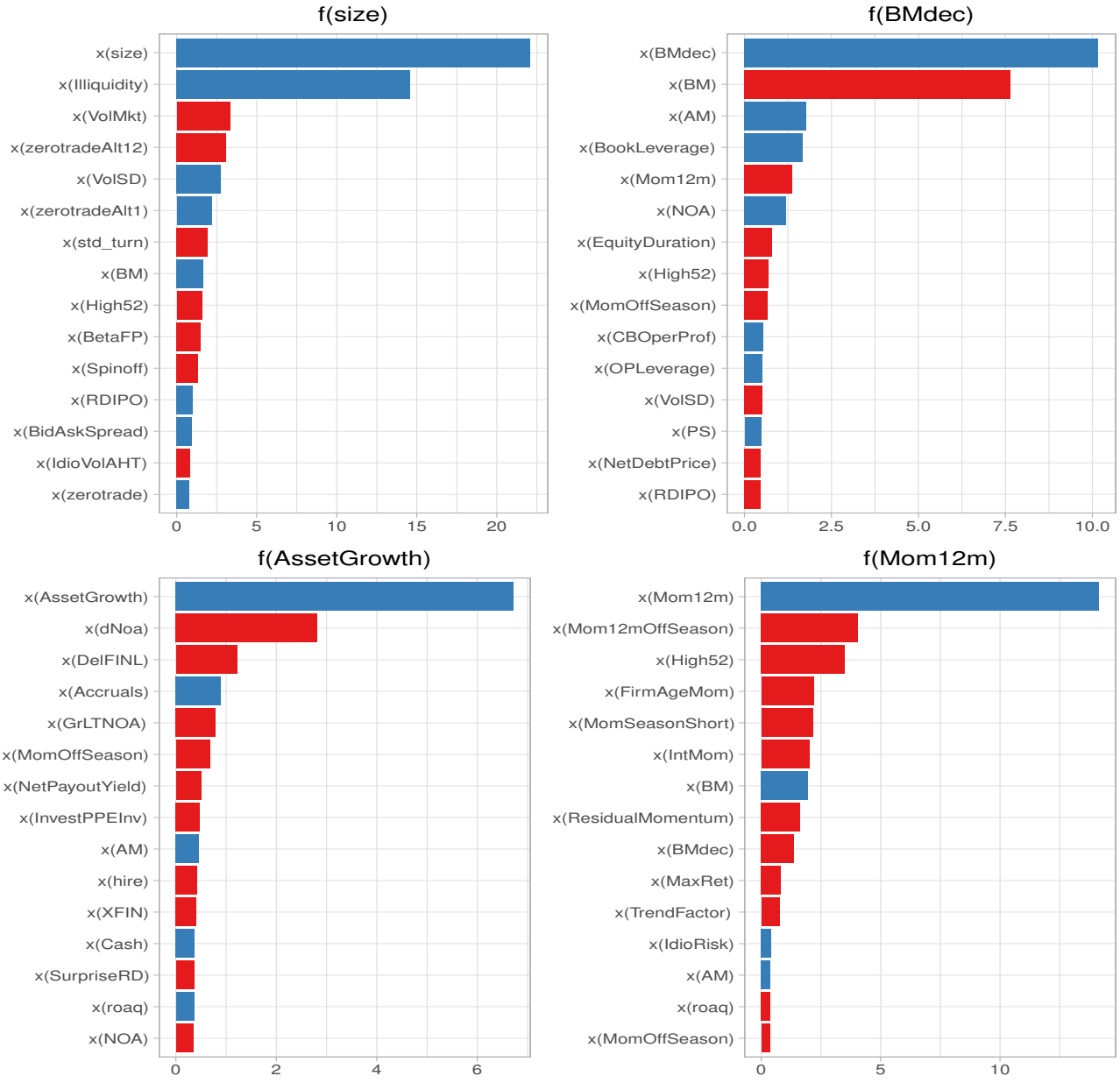


Figure 12: Estimated risk premia for Lewellen (2015) data

In this figure, we report the estimated risk premia for Lewellen (2015) data. The height of the bars indicates the absolute value of risk premia. The blue bar means it is positive while the red bar means negative. The time period is from April 1962 through December 2021. The risk premia here is in percentile. Let $\phi^2 = 3$ for FM with Restriction and all the k is selected by BIC rule.

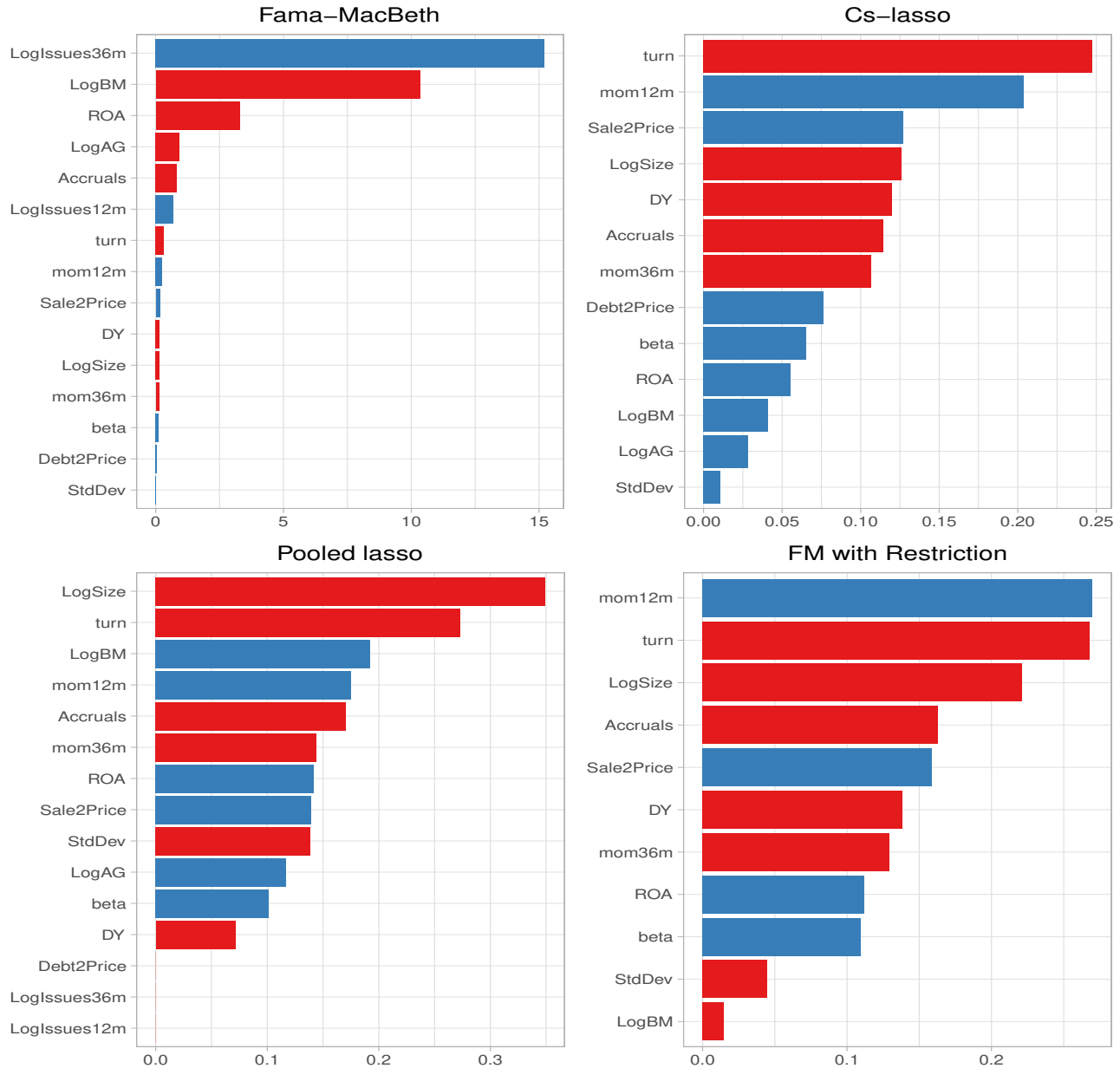


Figure 13: Estimated risk premia for Chen and Zimmermann (2022) data

In this figure, we report the estimated risk premia for Chen and Zimmermann (2022) data. The height of the bars indicates the absolute value of risk premia. The blue bar means it is positive while the red bar means negative. The time period is from April 1962 through December 2021. The risk premia here is in percentile. Our characteristics are normalized. To save some place, we only report the result of 15 characteristics with largest absolute risk premia. Let $\phi^2 = 12$ for FM with Restriction and all the k is selected by BIC rule.

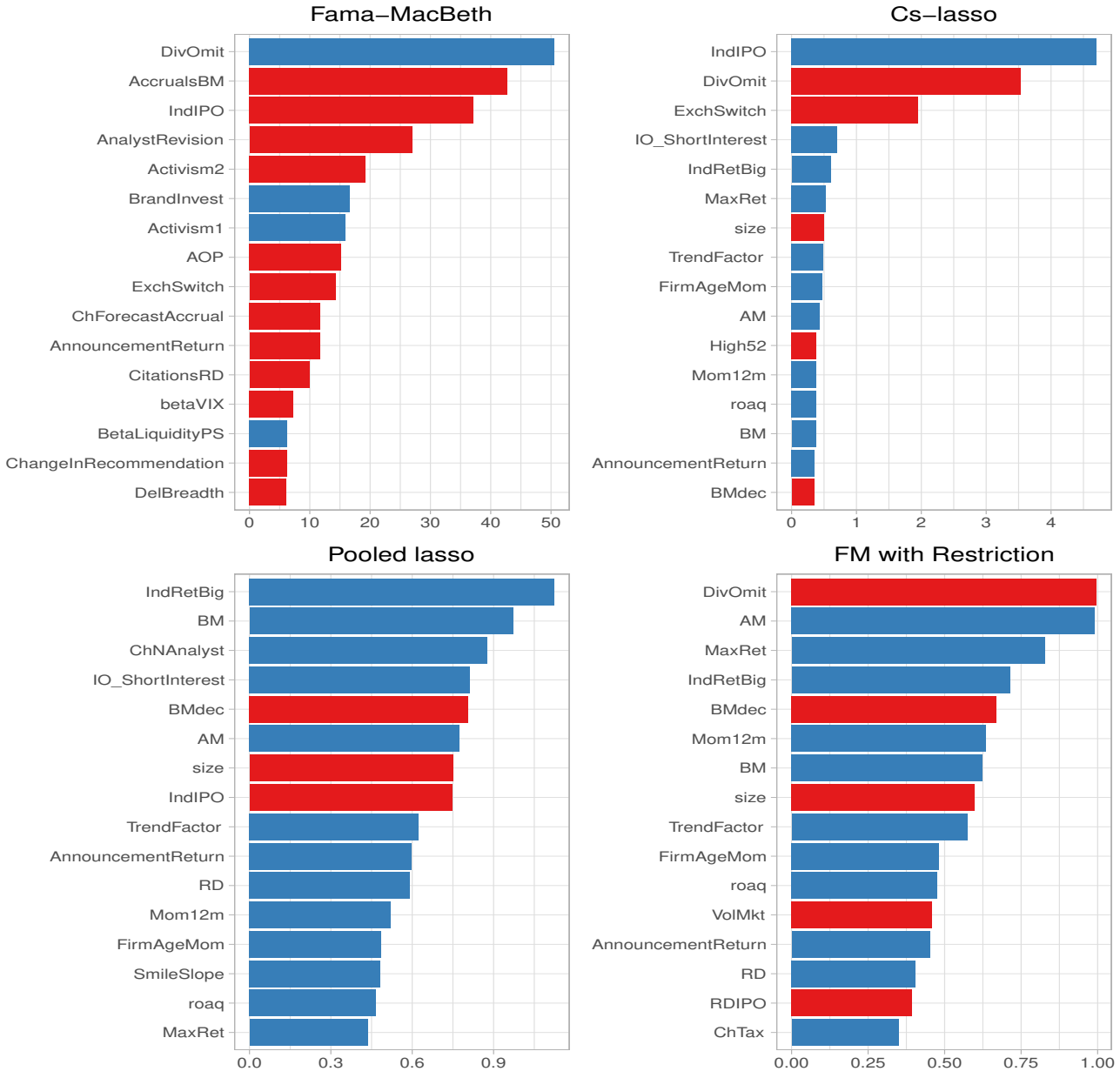


Figure 14: $L(i)$ (percentage loss of $R^2(r_t)$ if withdraw category i)

In this figure, we report $L(i)$ for different models based on Chen and Zimmermann (2022) data. The height of the bar indicates the value of $L(i)$ and different colors indicate different models. The time period is from April 1962 through December 2021. $L(i)$ here is in percentile. $L(i)$ is the percentage loss of $R^2(r_t)$ after withdrawing category i and the definition can be found in Equation (30). We choose $\phi^2 = 12$ here for FM with Restriction.

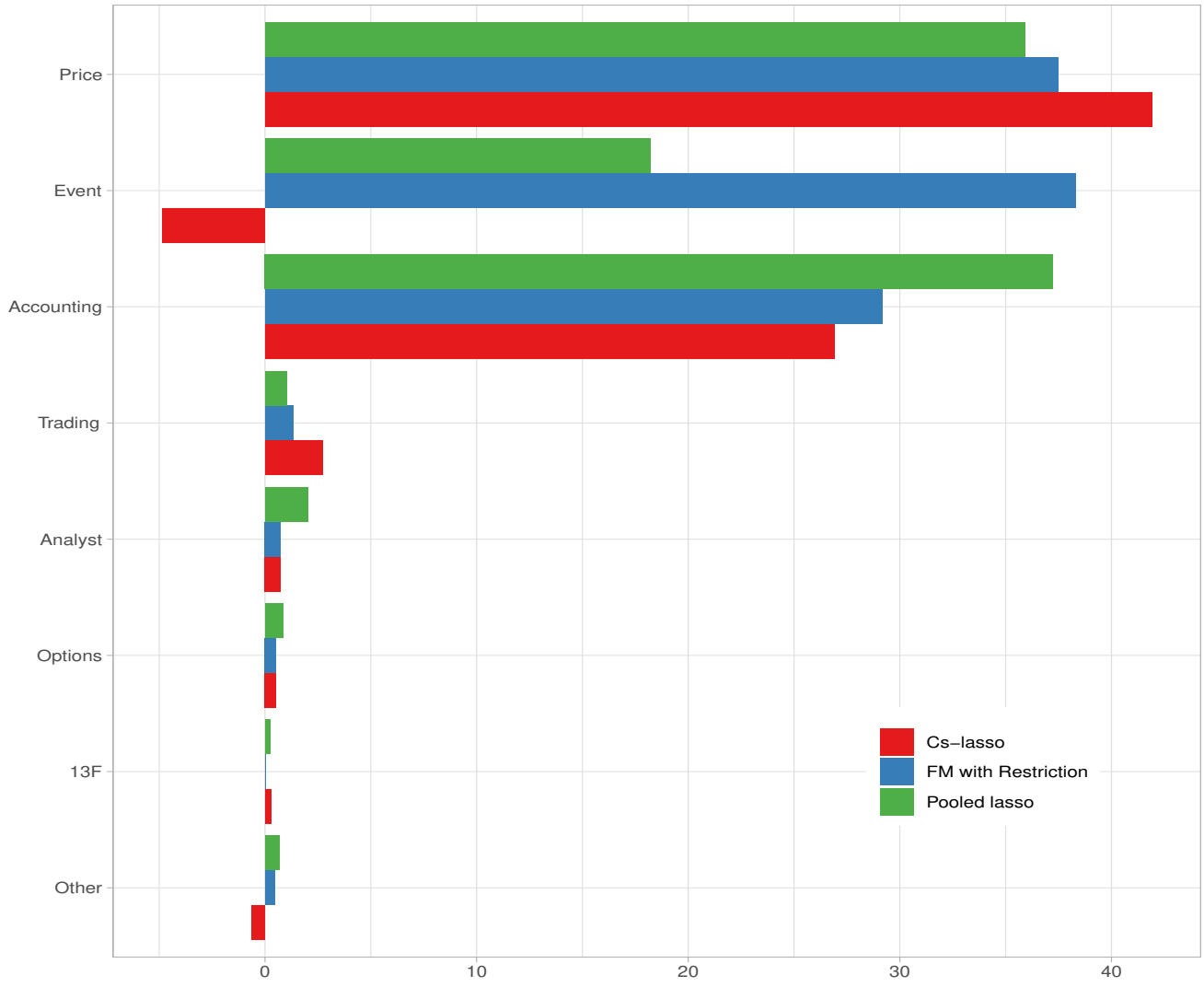
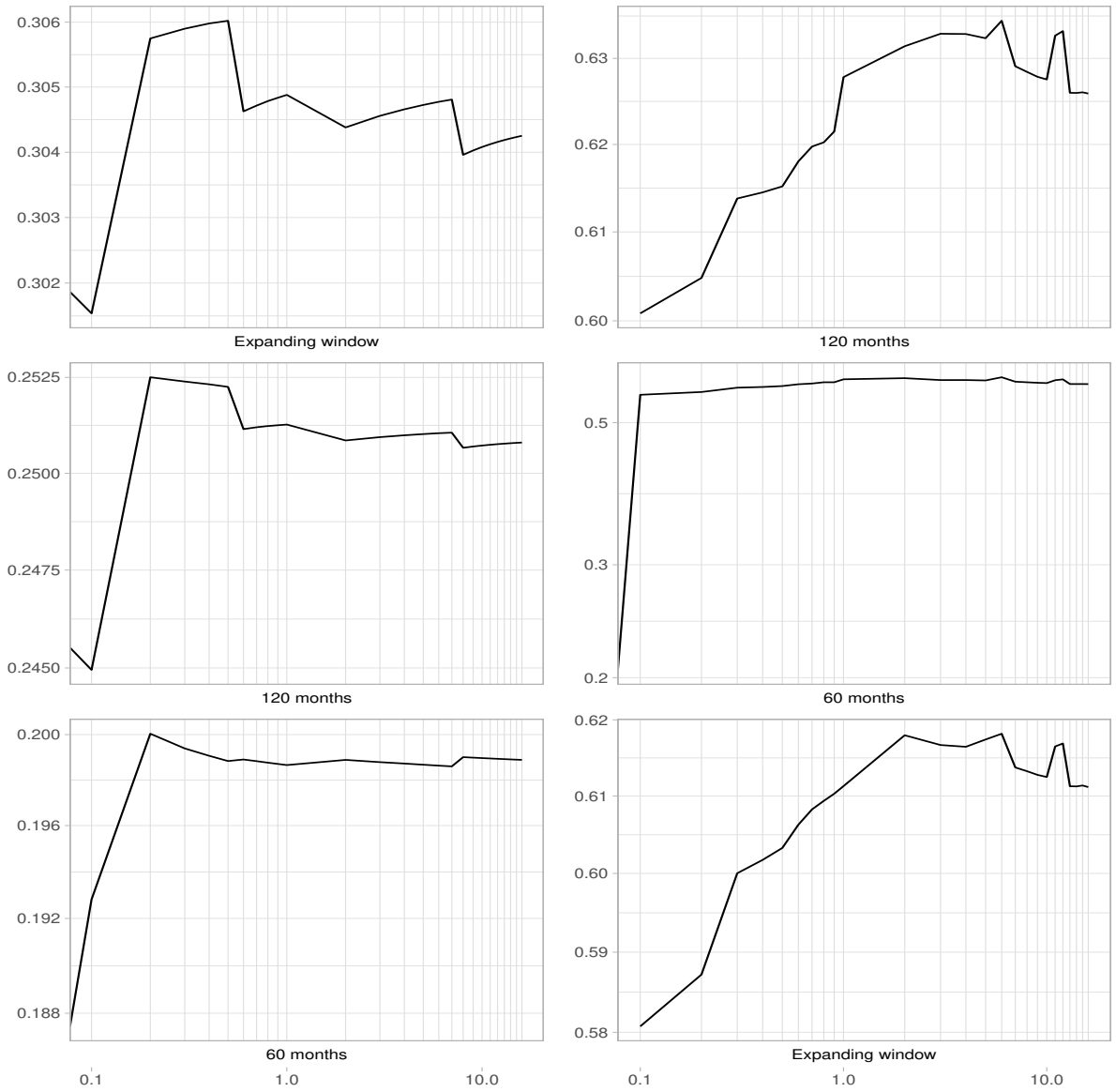


Figure 15: Out-of-sample $R^2(r_t)$ with different ϕ^2 and moving-average length

In this figure, we report the out-of-sample $R^2(r_t)$ for different ϕ^2 and moving-average length. The figures at the left shows the result for Lewellen (2015) data and ones at the right for Chen and Zimmermann (2022) data. The horizontal axis denotes the value of ϕ^2, and figures in each row denote the results for different moving-average length. The out-of-sample time period is from April 1977 through December 2021. The $R^2(r_t)$ here is in percentile.

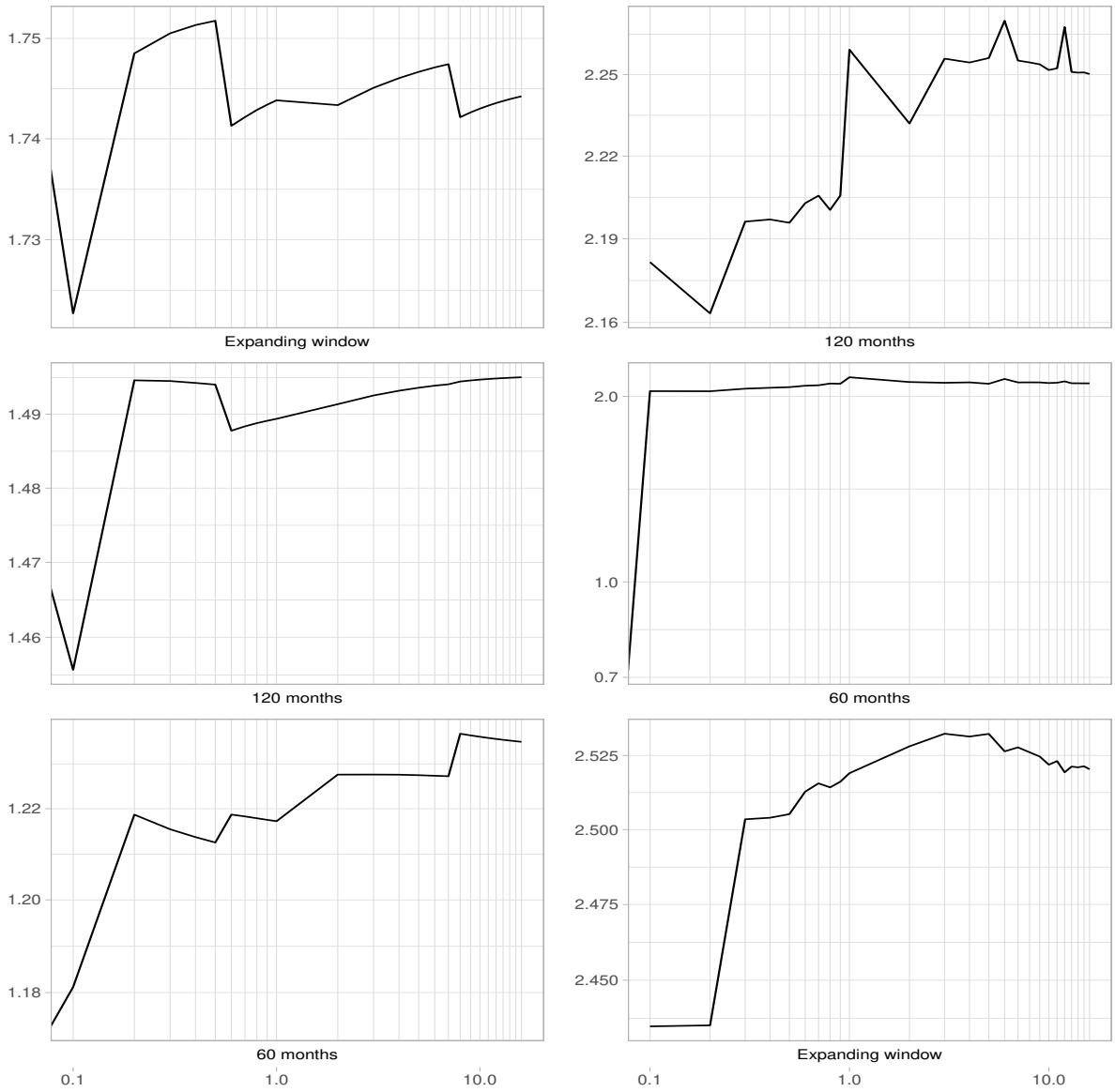


(a) Lewellen (2015) data

(b) Chen and Zimmermann (2022) data

Figure 16: Out-of-sample $R^2(x_t)$ with different ϕ^2 and moving-average length

In this figure, we report the out-of-sample $R^2(x_t)$ for different ϕ^2 and moving-average length. The figures at the left shows the result for Lewellen (2015) data and ones at the right for Chen and Zimmermann (2022) data. The horizontal axis denotes the value of ϕ^2, and figures in each row denote the results for different moving-average length. The out-of-sample time period is from April 1977 through December 2021. The $R^2(r_t)$ here is in percentile.

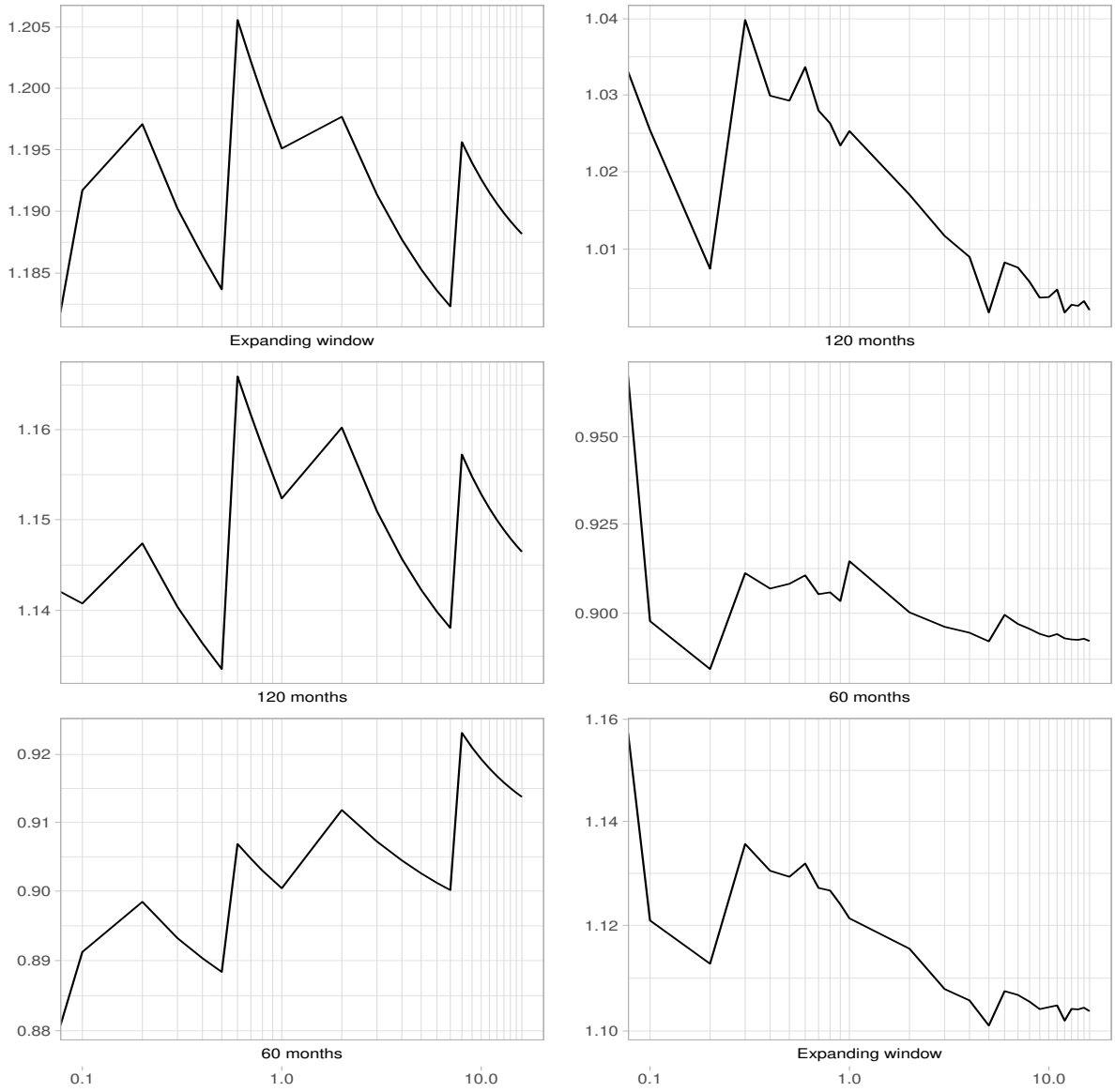


(a) Lewellen (2015) data

(b) Chen and Zimmermann (2022) data

Figure 17: Out-of-sample β_{slope} with different ϕ^2 and moving-average length

In this figure, we report the out-of-sample result for β_{slope} for different ϕ^2 and moving-average length. The figures at the left shows the result for Lewellen (2015) data and ones at the right for Chen and Zimmermann (2022) data. The horizontal axis denotes the value of ϕ^2, and figures in each row denote the results for different moving-average length. The out-of-sample time period is from April 1977 through December 2021.

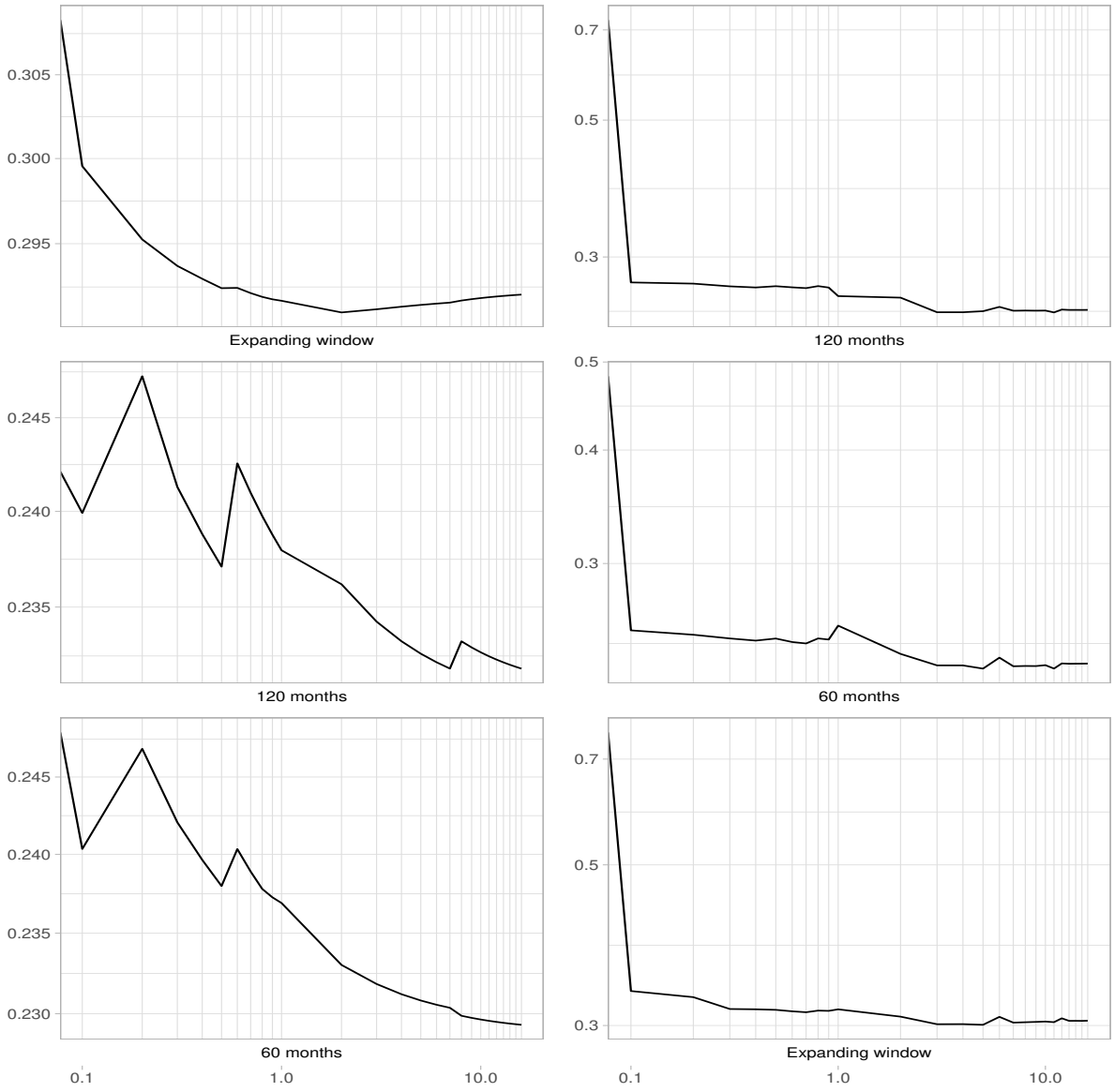


(a) Lewellen (2015) data

(b) Chen and Zimmermann (2022) data

Figure 18: Out-of-sample $\overline{|\alpha|}/\overline{|\mu|}$ with different ϕ^2 and moving-average length

In this figure, we report the out-of-sample result for $\overline{|\alpha|}/\overline{|\mu|}$ for different ϕ^2 and moving-average length. The figures at the left shows the result for Lewellen (2015) data and ones at the right for Chen and Zimmermann (2022) data. The horizontal axis denotes the value of ϕ^2, and figures in each row denote the results for different moving-average length. The out-of-sample time period is from April 1977 through December 2021.

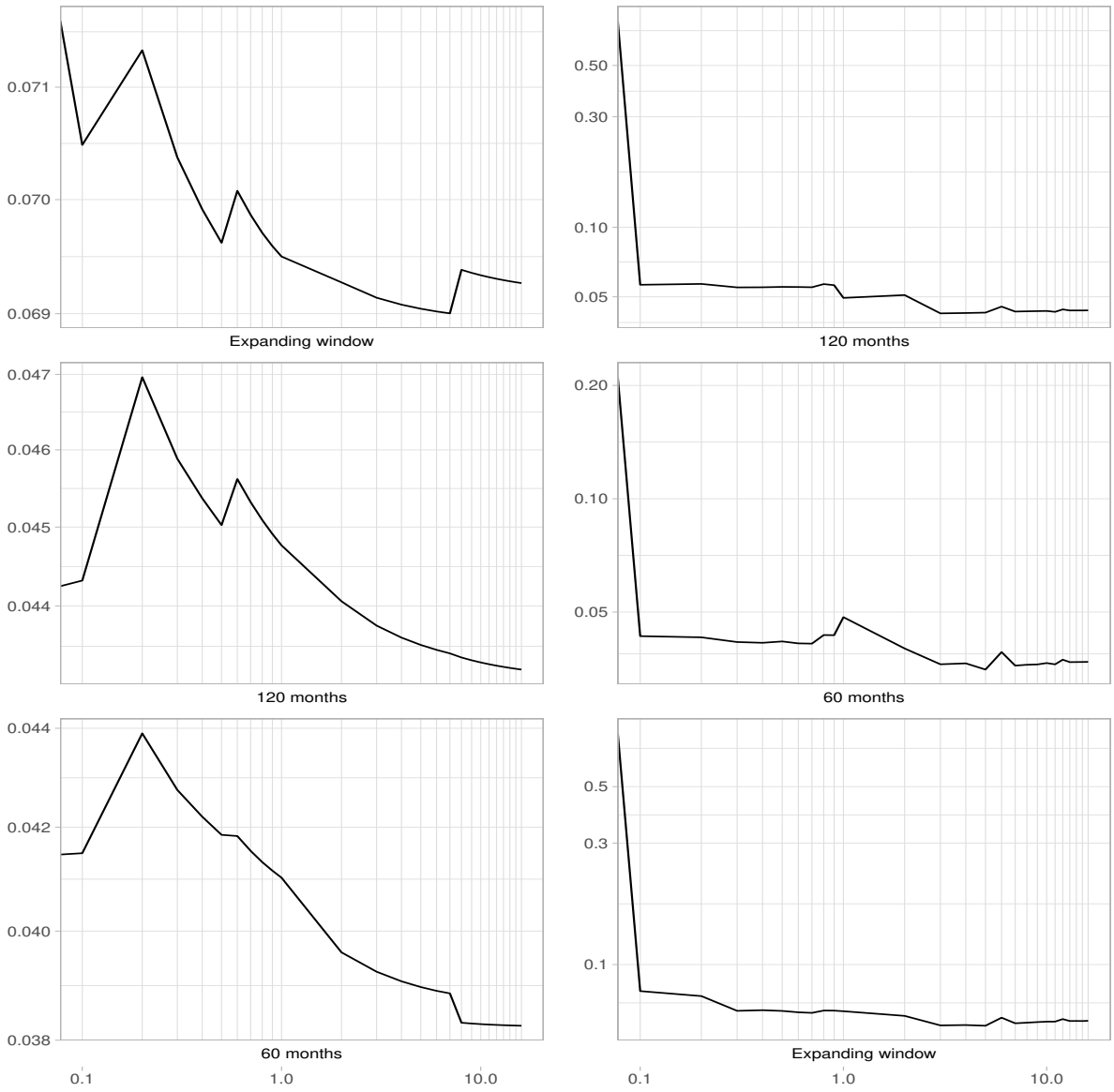


(a) Lewellen (2015) data

(b) Chen and Zimmermann (2022) data

Figure 19: Out-of-sample $\overline{\alpha^2}/\overline{\mu^2}$ with different ϕ^2 and moving-average length

In this figure, we report the out-of-sample result for $\overline{\alpha^2}/\overline{\mu^2}$ for different ϕ^2 and moving-average length. The figures at the left shows the result for Lewellen (2015) data and ones at the right for Chen and Zimmermann (2022) data. The horizontal axis denotes the value of ϕ^2, and figures in each row denote the results for different moving-average length. The out-of-sample time period is from April 1977 through December 2021.

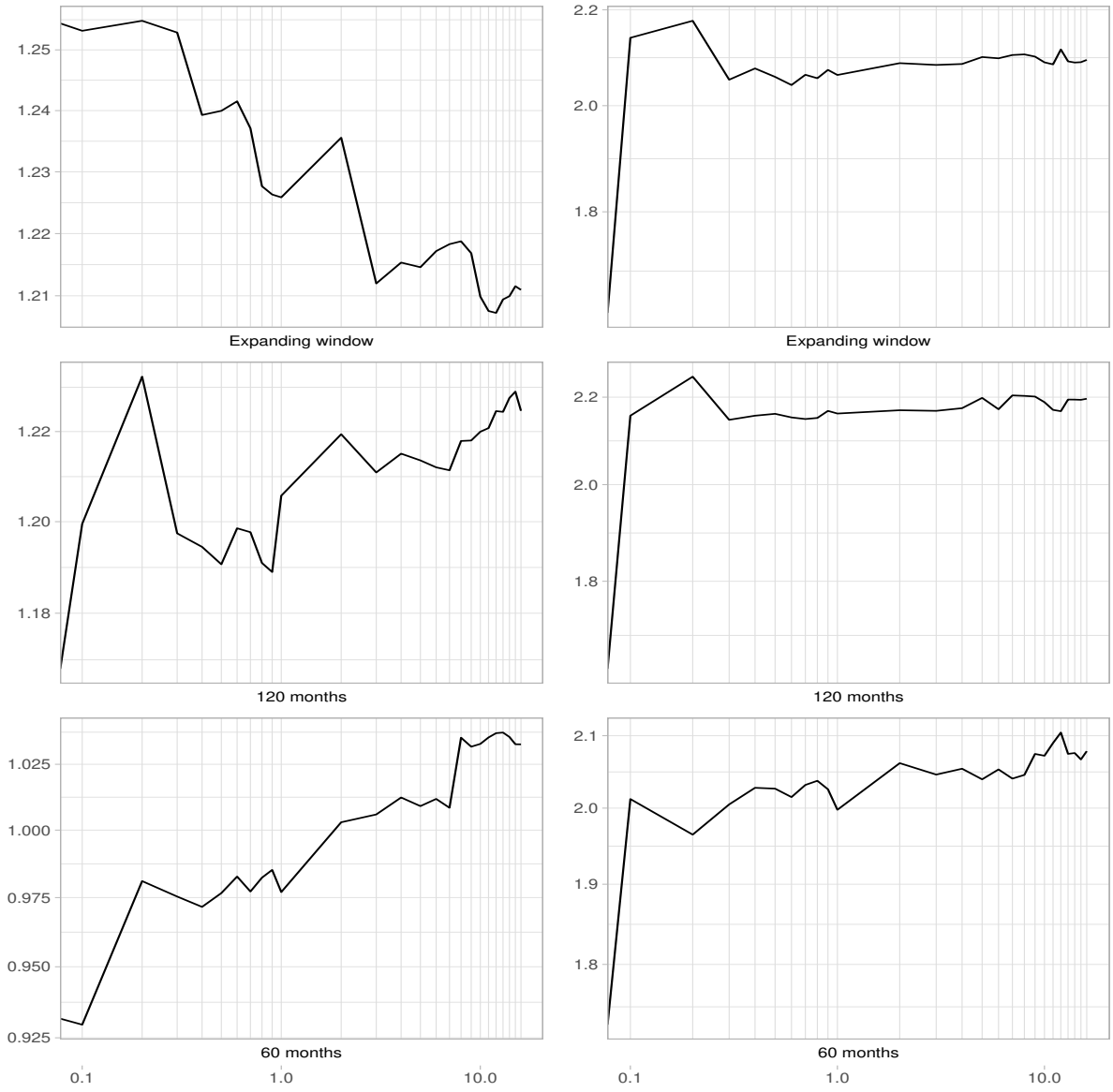


(a) Lewellen (2015) data

(b) Chen and Zimmermann (2022) data

Figure 20: Out-of-sample mean returns for long-short portfolios with different ϕ^2 and moving-average length

In this figure, we report the out-of-sample mean returns for long-short portfolios with different ϕ^2 and moving-average length. The figures at the left shows the result for Lewellen (2015) data and ones at the right for Chen and Zimmermann (2022) data. The horizontal axis denotes the value of ϕ^2, and figures in each row denote the results for different moving-average length. The out-of-sample time period is from April 1977 through December 2021. The mean return here is in percentile.

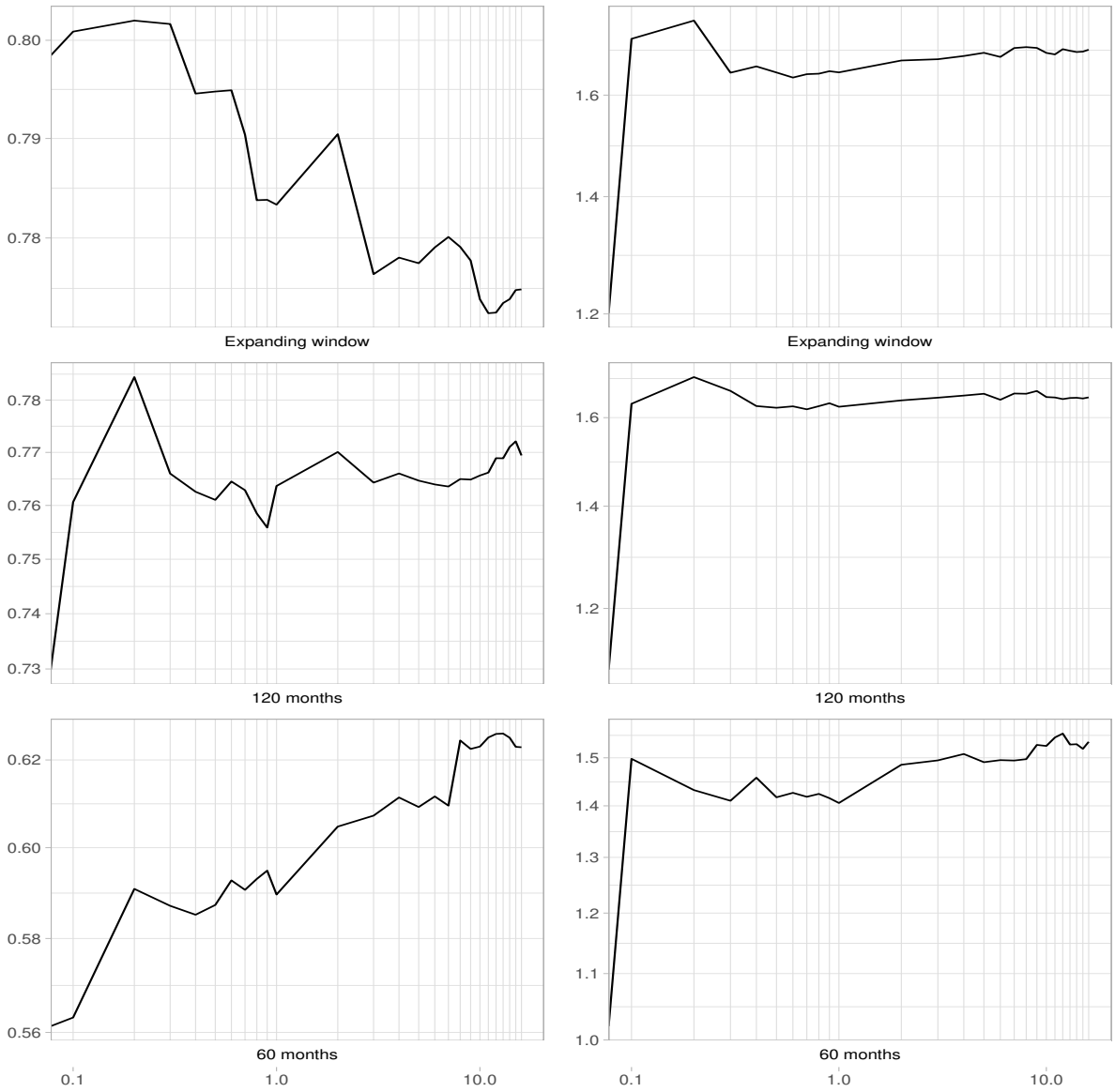


(a) Lewellen (2015) data

(b) Chen and Zimmermann (2022) data

Figure 21: Out-of-sample Sharpe ratios for long-short portfolios with different ϕ^2 and moving-average length

In this figure, we report the Out-of-sample annualized Sharpe ratios for long-short portfolios for different ϕ^2 and moving-average length. The figures at the left shows the result for Lewellen (2015) data and ones at the right for Chen and Zimmermann (2022) data. The horizontal axis denotes the value of ϕ^2, and figures in each row denote the results for different moving-average length. The out-of-sample time period is from April 1977 through December 2021.



(a) Lewellen (2015) data

(b) Chen and Zimmermann (2022) data

A Appendix A: The Proof for Proposition 2

If $k=1$, the conclusion is correct since we only obtain one factor. Assume that for the $k-1$th round the conclusion is correct, denote $\gamma^{(k)}_{\mathcal{A}_k,t+1}=(Z^\top_{\mathcal{A}_k,t}Z_{\mathcal{A}_k,t})^{-1}Z^\top_{\mathcal{A}_k,t}r^{(k-1)}_{t+1}$ as the $\mathcal{A}_k$'th row of $\gamma^{(k)}_{t+1}$, then in the kth round, there will be:

$$\begin{aligned}
\begin{pmatrix} Z^\top_{\mathcal{A}^{(1)}_k,t} \\ \vdots \\ Z^\top_{\mathcal{A}^{(k)}_k,t} \end{pmatrix}(r^{(k-1)}_{t+1}-\alpha_j Z_t\gamma^{(k)}_{t+1}) &= \begin{pmatrix} Z^\top_{\mathcal{A}^{(1)}_k,t} \\ \vdots \\ Z^\top_{\mathcal{A}^{(k)}_k,t} \end{pmatrix}(r^{(k-1)}_{t+1}-\alpha_j Z_{\mathcal{A}_k,t}\gamma^{(k)}_{\mathcal{A}_k,t+1}) \\
&= Z^\top_{\mathcal{A}_k,t}[r^{(k-1)}_{t+1}-Z_{\mathcal{A}_k,t}\gamma^{(k)}_{\mathcal{A}_k,t+1}+(1-\alpha_j)Z_{\mathcal{A}_k,t}\gamma^{(k)}_{\mathcal{A}_k,t+1}] \\
&= (1-\alpha_j)Z^\top_{\mathcal{A}_k,t}Z_{\mathcal{A}_k,t}\gamma^{(k)}_{\mathcal{A}_k,t+1} \\
&= (1-\alpha_j)\begin{pmatrix} Z^\top_{\mathcal{A}^{(1)}_k,t} \\ \vdots \\ Z^\top_{\mathcal{A}^{(k)}_k,t} \end{pmatrix} r^{(k-1)}_{t+1} \\
&= (1-\alpha_j)\begin{pmatrix} Z^\top_{\mathcal{A}^{(1)}_k,t} \\ \vdots \\ Z^\top_{\mathcal{A}^{(k)}_k,t} \end{pmatrix}(r^{(k-2)}_{t+1}-\alpha^{(k-1)}Z_t\gamma^{(k-1)}_{t+1}).
\end{aligned}$$

Therefore:

$$Z^\top_{\mathcal{A}^{(i)}_k,t}(r^{(k-1)}_{t+1}-\alpha_j Z_t\gamma^{(k)}_{t+1})=(1-\alpha_j)Z^\top_{\mathcal{A}^{(i)}_k,t}(r^{(k-2)}_{t+1}-\alpha^{(k-1)}_j Z_t\gamma^{(k-1)}_{t+1}), i=1,\cdots,k.$$

In $k-1$th round it is assumed that $C(\mathcal{A}^{(1)}_{k-1},k-1,\alpha_j)=\cdots=C(\mathcal{A}^{(k-1)}_{k-1},k-1,\alpha_j)$. Since $\mathcal{A}^{(i)}_k=\mathcal{A}^{(i)}_{k-1}, i=1,\cdots,k-1$, according to the equation above, there we have:

$$C(\mathcal{A}^{(1)}_k,k,\alpha_j)=\cdots=C(\mathcal{A}^{(k-1)}_k,k,\alpha_j).$$

Besides, in the $k-1$th round we find that the $\mathcal{A}^{(1)}_k$ so that $C(\mathcal{A}^{(k)}_k,k,\alpha_j)=C(j',k,\alpha_j)$, where j' is chosen arbitrarily from $\mathcal{A}_{k-1}$, and then the conclusion is proved for the kth round.

B Appendix B: The Proof for Proposition 3

If $Z_t^\top Z_t = I_l, t = 1, \cdots, T-1$, let $f_{(i)} = (f_{i,2}, \cdots, f_{i,T})^\top$, then the KKT condition of Equation (2) will be:

$$(1 + \frac{\lambda}{\sqrt{||f_{(i)}||_2^2 + \phi^2\mu_i^2}}) \begin{pmatrix} f_{i,2} \\ \vdots \\ f_{i,T} \\ \phi\mu_i \end{pmatrix} = \begin{pmatrix} Z_{i,1}^\top r_2 \\ \vdots \\ Z_{i,T-1}^\top r_T \\ \phi \sum_{t=1}^{T-1} Z_{i,t-1}^\top r_t \end{pmatrix}, \quad \forall ||f_{(i)}||_2^2 > 0, \tag{31}$$

$$\left\| \begin{matrix} Z_{i,1}^\top r_2 - f_{i,2} \\ \vdots \\ Z_{i,T-1}^\top r_T - f_{i,T} \\ \phi \sum_{t=1}^{T-1} (Z_{i,t}^\top r_{t+1} - f_{i,t+1}) \end{matrix} \right\|_2 \leq \lambda, \quad \forall ||f_{(i)}||_2^2 = 0. \tag{32}$$

After taking the l_2 norm of equation 31, denote $C(i) = \sqrt{\sum_{t=1}^{T-1} (Z_{i,t}^\top r_{t+1})^2 + \phi^2 \sum_{t=1}^{T-1} Z_{i,t}^\top r_{t+1}^2}$, we will know that:

$$\sqrt{||f_{(i)}||_2^2 + \phi^2\mu_i^2} = C(i) - \lambda,$$

so the solution is:

$$f_{i,t+1} = (1 - \frac{\lambda}{C(i)}) Z_{i,t}^\top r_{t+1}, \quad \forall C(i) > \lambda, \tag{33}$$

$$f_{i,t+1} = 0, \quad \forall C(i) \leq \lambda. \tag{34}$$

Obviously, when λ becomes smaller from $+\infty$ to 0, factors become nonzero one by one, and the first one will be the one with largest $C(i)$, second one with the second largest $C(i)$ and so on. And since characteristics are orthogonal, the selection of FM with Restriction will be exactly the same. Besides, if we select k characteristics, both of the factors of Equation (2) and FM with Restriction will be proportional to $Z_{\mathcal{A}_k}^\top r_{t+1}$.

Online Appendix

Fama-MacBeth Regression with Asset Pricing Restriction

This appendix provides complete results for the robustness checks discussed in the paper. Below, we briefly describe the contents of the appendix tables.

Table A1: The out-of-sample performance of different models based on on Gu, Kelly, and Xiu (2020) data discussed in Section 3.5 of the paper.

Table A2: The out-of-sample performance of long-short portfolios with value-weighting on Gu, Kelly, and Xiu (2020) data discussed in Section 3.5 of the paper.

Table A3: the alphas and their associated t-values of long-short portfolios for different models on Gu, Kelly, and Xiu (2020) data discussed in Section 3.5 of the paper.

Table A1: Out-of-sample performance with different moving average length on Gu, Kelly, and Xiu (2020) data

In this table, we report the out-of-sample performance including R^2, β_{slpoe}, the mean absolute pricing error ($\overline{|\alpha|}/\overline{|\mu|}$) and the mean square pricing error ($\overline{\alpha^2}/\overline{\mu^2}$). The out-of-sample period is from April 1977 through December 2021. The R^2 here is in percentile. Here, we set $\phi^2 = 4$. The definitions of the metrics are provided in subsection 3.2.1.

(1) MA length	(2) Model	(3) $R^2(r_t)$	(4) $R^2(x_t)$	(5) β_{slpoe}	(6) $\overline{\lvert\alpha\rvert}/\overline{\lvert\mu\rvert}$	(7) $\overline{\alpha^2}/\overline{\mu^2}$
Expanding	Fama-MacBeth	-0.15	-0.13	0.47	0.98	0.63
	Cs-lasso	0.28	1.78	1.37	0.46	0.16
	Pooled lasso	0.32	1.34	0.7	0.19	0.02
	FM with Restriction	0.37	1.68	1	0.34	0.06
120	Fama-MacBeth	0.03	-0.07	0.66	0.33	0.06
	Cs-lasso	0.29	1.3	0.74	0.16	0.01
	Pooled lasso	0.25	1.2	0.68	0.16	0.01
	FM with Restriction	0.38	1.59	1	0.22	0.03
60	Fama-MacBeth	0.17	0.8	0.65	0.14	0.01
	Cs-lasso	0.18	0.88	0.7	0.16	0.01
	Pooled lasso	0.01	0.87	0.54	0.15	0.01
	FM with Restriction	0.27	1.21	0.83	0.16	0.01

Table A2: The out-of-sample performance of long-short portfolios with value-weighting on Gu, Kelly, and Xiu (2020) data

In this table, we report the out-of-sample performance of long-short portfolios. If we regard $\hat{r}_{t+1}$ as a signal measurable at time t, we can construct decile portfolios every month from April 1977 to December 2021 by sorting stocks based on $\hat{r}_{t+1}$. Portfolio 1(10) is the portfolio of stocks with the lowest (highest) $\hat{r}_{t+1}$. The table reports the value-weighted next-month averages excess return, standard deviation (SD), as well as the annualized Sharpe ratio (SR), for each decile. H-L denotes the 10-1 return spread and its associated standard deviation (SD) together with the annualized Sharpe ratio (SR). Here, we set $\phi^2 = 4$.

	Fama-MacBeth			Cs-lasso			Pooled lasso			FM with Restriction		
	mean(%)	SD(%)	SR	mean(%)	SD(%)	SR	mean(%)	SD(%)	SR	mean(%)	SD(%)	SR
1(L)	0.20	5.72	0.12	0.20	5.79	0.12	0.02	5.90	0.01	0.30	5.82	0.18
2	0.43	5.05	0.29	0.44	5.07	0.30	0.37	5.15	0.25	0.52	4.97	0.36
3	0.56	4.76	0.41	0.62	4.76	0.45	0.58	4.72	0.43	0.64	4.60	0.48
4	0.60	4.73	0.44	0.63	4.62	0.47	0.63	4.54	0.48	0.72	4.64	0.53
5	0.76	4.64	0.56	0.73	4.52	0.56	0.72	4.53	0.55	0.74	4.52	0.57
6	0.74	4.73	0.54	0.76	4.76	0.56	0.90	4.76	0.65	0.78	4.80	0.56
7	0.81	4.92	0.57	0.81	5.06	0.56	0.87	5.04	0.60	0.97	5.18	0.65
8	0.91	5.49	0.58	0.94	5.54	0.59	0.96	5.29	0.63	0.99	5.49	0.62
9	1.06	5.95	0.62	1.07	6.03	0.62	1.11	5.98	0.64	1.07	6.55	0.57
10(H)	1.49	7.41	0.70	1.60	7.55	0.73	1.65	7.81	0.73	1.92	8.26	0.81
H-L	1.29	4.78	0.94	1.40	5.09	0.95	1.63	5.00	1.13	1.62	5.69	0.99

Table A3: Alphas for long-short portfolios on Gu, Kelly, and Xiu (2020) data

In this table, we report the alphas and their associated t-values of long-short portfolios for different models. The alphas are computed using three different factor models: the CAPM; the six-factor model of Fama and French (2018), denoted FF6; and the five-factor q model of Hou, Mo, Xue, and Zhang (2020), denoted Q5. If we regard $\hat{r}_{t+1}$ as a signal measurable at time t, we can construct decile portfolios every month from April 1977 to December 2021 by sorting stocks based on $\hat{r}_{t+1}$. Portfolio 1(10) is the portfolio of stocks with the lowest (highest) $\hat{r}_{t+1}$. H-L denotes the differences in monthly returns and the associated t-values.

	Fama-MacBeth						Cs-lasso					
	α_{FF6}	$t(\alpha_{FF6})$	α_Q	$t(\alpha_Q)$	α_{CAPM}	$t(\alpha_{CAPM})$	α_{FF6}	$t(\alpha_{FF6})$	α_Q	$t(\alpha_Q)$	α_{CAPM}	$t(\alpha_{CAPM})$
Low(L)	-0.25	-2.56	-0.13	-1.19	-0.51	-5.03	-0.21	-2.28	-0.06	-0.56	-0.52	-5.13
2	-0.05	-0.62	0.03	0.37	-0.21	-2.78	-0.04	-0.57	0.03	0.4	-0.21	-2.78
3	-0.06	-1.09	-0.09	-1.42	-0.07	-1.16	-0.01	-0.14	-0.04	-0.7	-0.01	-0.21
4	-0.05	-0.91	-0.04	-0.7	-0.02	-0.38	-0.04	-0.75	-0.03	-0.58	0.02	0.3
5	0.05	0.81	0.05	0.83	0.15	2.58	0.02	0.35	-0.01	-0.14	0.14	2.31
6	-0.03	-0.49	-0.03	-0.41	0.13	1.98	-0.05	-0.79	-0.1	-1.33	0.15	2.16
7	0.08	1.1	0.11	1.26	0.19	2.36	0.02	0.31	0.03	0.34	0.18	2.09
8	0.13	1.39	0.17	1.64	0.24	2.35	0.1	1.06	0.13	1.25	0.27	2.47
9	0.19	1.86	0.22	1.95	0.33	2.96	0.15	1.44	0.14	1.2	0.35	2.9
10(H)	0.74	5.13	0.79	4.97	0.64	3.79	0.75	5.11	0.73	4.56	0.75	4.18
H-L	0.98	5.38	0.92	4.56	1.15	5.96	0.96	5.14	0.79	3.88	1.28	6.17
	Pooled lasso						FM with Restriction					
	α_{FF6}	$t(\alpha_{FF6})$	α_Q	$t(\alpha_Q)$	α_{CAPM}	$t(\alpha_{CAPM})$	α_{FF6}	$t(\alpha_{FF6})$	α_Q	$t(\alpha_Q)$	α_{CAPM}	$t(\alpha_{CAPM})$
Low(L)	-0.32	-3.78	-0.28	-2.58	-0.71	-6.82	0.02	0.2	0.01	0.08	-0.43	-4.36
2	-0.07	-1.1	0.02	0.21	-0.3	-4.15	-0.01	-0.12	0	-0.07	-0.14	-2.27
3	0	0.07	0.02	0.35	-0.04	-0.76	-0.02	-0.47	-0.07	-1.31	0.02	0.52
4	-0.06	-1.22	-0.04	-0.69	0.03	0.56	-0.05	-0.96	-0.01	-0.24	0.11	1.87
5	-0.01	-0.12	-0.07	-1.09	0.12	2.23	-0.02	-0.34	-0.01	-0.17	0.16	2.46
6	0.11	1.66	0.12	1.5	0.29	4.07	-0.06	-0.8	0.06	0.66	0.17	2.18
7	0.07	0.84	0.08	0.84	0.24	2.77	0.07	0.74	0.21	2.07	0.33	3.42
8	0.11	1.26	0.1	1.07	0.3	3.21	0.11	1.15	0.16	1.57	0.32	2.99
9	0.18	1.91	0.3	2.86	0.39	3.34	0.17	1.64	0.26	2.28	0.3	2.18
10(H)	0.83	5.55	0.89	5.52	0.77	4.19	1.09	6.19	1.25	6.44	1.06	4.79
H-L	1.16	6.35	1.18	5.79	1.48	7.36	1.07	5.39	1.24	5.6	1.49	6.43