NBER WORKING PAPER SERIES

FORWARD SELECTION FAMA-MACBETH REGRESSION WITH HIGHER ORDER ASSET-PRICING FACTORS

Nicola Borri
Denis Chetverikov
Yukun Liu
Aleh Tsyvinski

Working Paper 33663
http://www.nber.org/papers/w33663

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2025

We are grateful for comments by Stefano Giglio, Whitney Newey, and Dacheng Xiu. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Forward Selection Fama-MacBeth Regression with Higher Order Asset-Pricing Factors
Nicola Borri, Denis Chetverikov, Yukun Liu, and Aleh Tsyvinski
NBER Working Paper No. 33663
April 2025
JEL No. G0

ABSTRACT

We show that the higher-orders and their interactions of the common sparse linear factors can effectively subsume the factor zoo. We propose a forward selection Fama-MacBeth procedure as a method to estimate a high-dimensional stochastic discount factor model, isolating the most relevant higher-order factors. Applying this approach to terms derived from six widely used factors (the Fama-French five-factor model and the momentum factor), we show that the resulting higher-order model with only a small number of selected higher-order terms significantly outperforms traditional benchmarks both in-sample and out-of-sample. Moreover, it effectively subsumes a majority of the factors from the extensive factor zoo, suggesting that the pricing power of most zoo factors is attributable to their exposure to higher-order terms of common linear factors.

Nicola Borri
Dipartimento di Economia e Finanza
LUISS Guido Carli
Viale Romania 32
00197 Rome (ITALY)
nborri@luiss.it

Yukun Liu
Department of Finance
Simon Business School
University of Rochester
Rochester, NY 14620
yliu229@simon.rochester.edu

Denis Chetverikov
UCLA, Department of Economics
315 Portola Plaza
Bunche Hall, Room 8283
Los Angeles, CA 90095-1477
chetverikov@econ.ucla.edu

Aleh Tsyvinski
Department of Economics
Yale University
Box 208268
New Haven, CT 06520-8268
and NBER
a.tsyvinski@yale.edu

1. INTRODUCTION

The asset pricing literature has proposed hundreds of candidate factors that could potentially explain the cross-section of asset returns. Cochrane (2011), in his presidential address, refers to this extensive set of candidate factors as the "factor zoo." Sparse linear factor models, such as those introduced by Fama and French (2015) and Carhart (1997), struggle to adequately price this expansive factor zoo, implying they may not fully span the stochastic discount factor (SDF). In this paper, we show that the interactions of the higher-order terms of the common sparse linear factors can effectively subsume the factor zoo. We do so by determining important interactions via a new factor-selection procedure, which we call the *forward selection Fama-MacBeth regression procedure* (FS-FMB).

The prevalent focus in the literature on sparse linear factor models is primarily driven by considerations of parsimony and is motivated by the first order Taylor expansion of the SDF. However, there is no theoretical guarantee that a linear approximation sufficiently captures the variation of the SDF. We show instead that higher-order factors and their interactions, motivated by the higher-order Taylor expansion, can significantly better capture the SDF.

The main difficulty of studying the higher-order terms is the dimensionality issue. Consider, for example, the standard 6-factor model that includes the Fama-French 5 factors (Fama and French, 2015) and the momentum factor (Carhart, 1997). The higher-order interactions up to degree of 3 (4) already have 63 (114) factors. Therefore, we develop a forward selection procedure to choose factors that explain the stochastic discount factor that is suitable for the case when the number of potentially important factors is large. The method adds factors one by one into the set of selected factors, each time maximizing the R^2 of the cross-sectional regression (second step) of the Fama-MacBeth procedure. After the set of factors is selected, we estimate the SDF loadings by running the cross-sectional regression of sample average returns on sample covariances between returns and selected factors. We derive the rate of convergence of the proposed estimator and also propose a method to debias it, leading to simple inference procedures. We refer to our method as the forward selection Fama-MacBeth regression procedure.

Our method provides an alternative to the Lasso-based method proposed in Feng et al. (2020). Both forward selection and Lasso are among the most important machine learning techniques suitable for estimation of high-dimensional models (Hastie et al., 2009). Advantages of the forward selection method are the following. First, it is invariant to scaling of factors, which is important because Lasso-based methods yield results that depend non-trivially on how factors are scaled (see Nagel (2021) for a discussion of why potentially unequal scaling of factors is important in empirical asset pricing). Second, forward selection is highly interpretable: not only it gives the set of factors explaining the SDF, but it also gives the order in which factors enter the model. Third, forward selection is very easy to implement from the computational point of view, and thus can be easily adapted for desired modifications, e.g. ensuring that certain variables are guaranteed to enter the model.

We show that the performance of the common factors is dramatically improved with their higher-order interactions. We start with 6 of the most common factors in the asset pricing literature. These are the Fama-French 5 factors of Fama and French (2015) and the momentum factor of Jegadeesh and Titman (1993), and we refer to this model as the FF5M. We consider higher-orders as well as interactions of these 6 factors up to degree of 3, leading to 57 candidate higher-order factors. For the test assets, we use a large cross-section of equities including 484 characteristic-managed portfolios from Kozak et al. (2020). We apply our FS-FMB to select the higher-order factors that are most important in pricing the cross-section of these 484 asset returns. The procedure selects 7 higher-order factors, including 2 second degree higher-orders and 5 interactions. We show that the inclusion of the higher-order factors leads to dramatic improvement in pricing the cross-section of asset returns – the adjusted cross-sectional R^2 increases from 31.2% for the baseline FF5M model to 59% for the selected higher-order factor model. Moreover, the intercept alpha of the higher-order factor model is not statistically significant at the 5% level.

We then examine the SDF loadings of the higher-order factors selected by the FS-FMB using both the non-debiased estimates based on the standard Fama-MacBeth (FMB) two-pass regression and the debiased estimates. Estimates from the standard FMB two-pass procedure are likely biased because of the omitted variable bias (see, e.g., Chernozhukov et al. (2018) and Giglio and Xiu (2021)). The directions of the point estimates are the same and the magnitudes are similar with both procedures. The estimates based on the non-debiased procedure are all significant at the 5% level, while the estimates based on the debiased procedure are all significant at the 10% level with 5 of the 7 SDF loading estimates being significant at the 5% level. Overall, the results show that the higher-order factors are important to empirically capture the SDF and significantly price assets in the cross-section.

In our baseline analysis, we demonstrate that higher-order factors are important components of the SDF and price cross-sectional asset returns in-sample. We further provide evidence of the out-of-sample performance of the selected higher-order factor model, where we conduct two types of out-of-sample tests. Firstly, we present an out-of-sample test in the asset space. We select higher-order factors and estimate SDF loadings from the training assets and then evaluate the out-of-sample performance using the remaining assets. Our results show that the selected higher-order factors and the performance of the higher-order factor model are stable in this out-of-sample exercise.

Secondly, we conduct an out-of-sample test in the time series, where we split the sample into two equal subsamples based on the time series. We use the first half of the sample as the training sample and the second half of the sample as the test sample. Then, we estimate the model in the training sample and evaluate the out-of-sample performance in the test sample. Compared with the benchmark factor models including the CAPM, the FF3 factor model, the FF5 factor model, and the FF5M factor model, the out-of-sample R^2 of the higher-order factor model is significantly larger.

Next, we examine whether the selected higher-order factor model can subsume the pricing power of the factors in the factor zoo. We employ two empirical procedures. First, we perform standard FMB regressions to test the significance of the SDF loading of each factor in the factor zoo, controlling for the higher-order factor model. Second, we construct factor mimicking portfolios for the seven higher-order factors and conduct time-series asset pricing tests on the zoo factors.

Under the first approach, we estimate each zoo factor's SDF loading while controlling for our higher-order factor model. We find that only 7 our of the 148 zoo factors from Jensen et al. (2023) remain significant at the 5% level. The second approach is based on factor-mimicking portfolios. The benefit is that we are able to convert the non-tradable higher-order factors to tradable factors but the projection inevitably loses information from the original factors. The results using the second approach confirm that the higher-order factor model significantly reduces the pricing power of the zoo factors. Overall, these results demonstrate that the higher-order factors selected by the forward Fama-MacBeth procedure effectively subsume the zoo factors, although the linear benchmark factors cannot.

We then examine the relationship between the selected higher-order factors and macroeconomic risks. First, we examine the correlations between the selected higher-order factors and two primary macroeconomic indicators – U.S. equity market returns and the CBOE VIX index – in the full sample and during the recession periods. For the first two higher-order factors (i.e., SMB2 and SMB2*Mom), the magnitude of the correlations with U.S. equity market returns are markedly higher during the NBER recession periods relative to the full sample. For the last four higher-order factors (i.e., Mkt-RF2, Mkt-RF2*RMW, Mkt-RF*SMB, and HML2*Mkt-RF), we find they tend to have high correlation in magnitude with the VIX index, especially during the NBER recession periods. These results show that the selected higher-order factors capture different macroeconomic risks, especially during economic downturns. Additionally, we examine the exposures of the 7 selected higher-order factors to a large set of common macroeconomic factors. We find that the these higher-order factors explain a sizable fraction of variations for 5 of the 7 higher-order factors, but not for SMB2*Mom or Mom2*RMW. Of the macroeconomic factors, factors related to uncertainties appear to be more important in accounting for the higher-order factors.

Additionally, we provide robustness results for our empirical analyses. First, we show that the performance of the higher-order factor model is extremely unlikely to be driven by randomness by simulating white-noise factors. Second, we use the selected higher-order factor model to subsume an alternative factor zoo of Chen and Zimmermann (2022) and reach similar conclusions. Third, we experiment with alternative sets of candidate higher-order factors: (1) only include higher-orders, (2) only interactions, (3) up to degree 2, and (4) up to degree 4. We find that both higher-orders and interactions are important components of the SDF, especially interactions. Moreover, we show that the selected higher-order factor model up to degree 2 significantly underperform, while the selected higher-order factor model up to degree 4 has similar performance as our baseline model up to degree 3.

Our paper relates to a vast literature on factor models in asset pricing. A major part of the literature focuses on linear factor models, such as the characteristic-based models (e.g., Fama and French (1993, 1996, 2015); Carhart, 1997; Koijen et al., 2017) and statistical-based models (e.g., Connor and Korajczyk, 1986). A small subset of papers emphasizes non-linearity in explaining asset returns, such as Bansal and Viswanathan (1993), Harvey and Siddique (2000), and Dittmar (2002). We contribute to this literature by showing that higher-orders and higher-order interactions of common linear factors are important components of the SDF and price the cross-sectional asset returns.

The paper also contributes to the factor zoo literature. Cochrane (2011) identifies the factor zoo issue and calls for research to identify factors that are actually important for the SDF. Harvey et al. (2016) argue that, based on multiple hypothesis framework, many zoo factors are not significant. Harvey and Liu (2021) provide a bootstrap model selection framework that accounts for multiple hypothesis testing. McLean and Pontiff (2016) show that most of the zoo factors suffer from post publication bias. Green et al. (2017) uses FMB regression to evaluate 94 firm characteristics and argue that less than ten provide significant independent information to expected returns. Feng et al. (2020) propose a rigorous econometric framework to evaluate whether a new factor contributes to the SDF in light of the existing factor zoo. Our paper demonstrates that the higher-orders of the common linear factors are able to subsume the majority of the factor zoo.

Finally, our paper relates to the machine learning literature. Gu et al. (2020) show that machine learning methods are better in predicting future stock returns out-of-sample than OLS methods. Freyberger et al. (2020) propose a group LASSO method to choose firm characteristics that are most related to future returns. Kozak et al. (2020) show that machine learning methods applied to the factor zoo can better approximate the SDF. Lettau and Pelger (2020) propose a method for estimating latent asset pricing factors that fit both the time-series and the cross-section of asset returns. Giglio and Xiu (2021) propose a method to estimate risk premia of factors even if there are missing factors. Borri et al. (2024) show that a new non-linear single-factor asset pricing model motivated by the Kolmogorov-Arnold representation theorem is able to price assets from different asset classes. In this paper, we propose a forward selection method to estimate a high-dimensional stochastic discount factor model, isolating the most relevant higher-order factors and showing their importance.

The rest of the paper is organized as follows. Section 2 discusses the forward selection Fama-MacBeth procedure. Section 3 presents the data and the main empirical results. Section 4 presents additional robustness analyses. The Online Appendix contains technical derivations and further empirical results.

2. METHODOLOGY

In this section, we set up the model. We introduce its main components – returns, factors, and the stochastic discount factor – and explain the relationship between them. We then present a forward selection Fama-MacBeth regression procedure as a method for selecting the factors that drive the stochastic discount factor. Using this selection procedure, we propose an estimator of the stochastic discount factor loadings. As we will see, the proposed estimator is biased. We therefore also explain how to debias it and how to perform inference on the stochastic discount factor loadings using the debiased estimator.

Notation. For any integer $p \geq 1$, we denote $[p] = \{1, \ldots, p\}$ and $\mathbf{0}_p = (0, \ldots, 0)^\top \in \mathbb{R}^p$. For any set $\mathrm{S} \subset [p]$, we define $\mathrm{S}^c = [p] \setminus \mathrm{S}$, and we denote by $|\mathrm{S}|$ the number of elements in S. For any vector $x = (x_1, \ldots, x_p)^\top$ and any set $\mathrm{S} \subset [p]$, we use $x_\mathrm{S} = (x_i)_{i \in \mathrm{S}}^\top$ to denote the sub-vector of x consisting of all its components with indices in S. Similarly, for any vector $x_t = (x_{t,1}, \ldots, x_{t,p})^\top$ with an extra index t (or some other index), we denote by $x_{t,\mathrm{S}} = (x_{t,i})_{i \in \mathrm{S}}^\top$ the corresponding sub-vector. For any matrix $A = (A_{i,j})_{i \in [N], j \in [p]}$ and any set $\mathrm{S} \subset [p]$, we use $A_\mathrm{S} = (A_{i,j})_{i \in [N], j \in \mathrm{S}}$ to denote the sub-matrix of A consisting of all its columns with indices in S. For any vector $x = (x_1, \ldots, x_p)^\top \in \mathbb{R}^p$, we let $\|x\|_0 = \sum_{j \in [p]} \mathbf{I}\{x_j \neq 0\}$ be its ℓ_0-"norm" and let $\|x\|_1 = \sum_{j \in [p]} |x_j|$, $\|x\|_2 = (\sum_{j \in [p]} |x_j|^2)^{1/2}$, and $\|x\|_\infty = \max_{j \in [p]} |x_j|$ be its ℓ_1-, ℓ_2-, and ℓ_∞-norms. For any matrix $A = (A_{i,j})_{i \in [N], j \in [p]} \in \mathbb{R}^{N \times p}$, we let $\|A\|_2 = \sup_{v \in \mathbb{R}^p : \|v\|_2 = 1} \|Av\|_2$ and $\|A\|_{\infty,1} = \max_{i \in [N]} \sum_{j \in [p]} |A_{i,j}|$ be its spectral and $\ell_{\infty,1}$—norms. For any symmetric matrix $A \in \mathbb{R}^{p \times p}$, we let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ be its smallest and largest eigenvalues. We use the abbreviation "w.p.a.1" to denote "with probability approaching one." For non-random sequences $\{a_T\}_{T \geq 1}$ and $\{b_T\}_{T \geq 1}$ in $\mathbb{R}$, we write $a_T \lesssim b_T$ if there exists a bounded non-random sequence

$\{C_T\}_{T\geq 1}$ such that $a_T = C_T b_T$ for all $T \geq 1$. For random sequences $\{a_T\}_{T\geq 1}$ and $\{b_T\}_{T\geq 1}$ in $\mathbb{R}$, we write $a_T \lesssim_P b_T$ if there exists a bounded in probability random sequence $\{C_T\}_{T\geq 1}$ such that $a_T = C_T b_T$ for all $T \geq 1$.

2.1. *Model*

Let $r_{i,t} \in \mathbb{R}$ be the excess return of asset $i \in [N]$ at time period $t \in [T]$ and let $f_t = (f_{t,1}, \ldots, f_{t,p})^\top \in \mathbb{R}^p$ be the vector of factors. Throughout the paper, we assume that the returns and the factors are stationary over time. In addition, we assume that the vector of factors is able to span the stochastic discount factor (SDF):

$$m_t = 1 - \psi^\top (f_t - \mathrm{E}[f_t]), \quad \text{for all } t \in [T], \tag{1}$$

where m_t is the SDF and $\psi = (\psi_1, \ldots, \psi_p)^\top \in \mathbb{R}^p$ is the vector of SDF loadings.

By definition, since we work with the *excess* returns, m_t is orthogonal to the returns in the sense that $\mathrm{E}[m_t r_{i,t}] = 0$ for all $i \in [N]$. Using this equation in combination with (1), it then follows from Section 6.3 in Cochrane (2005) that for each asset $i \in [N]$,

$$r_{i,t} = \beta_i^\top (\gamma - \mathrm{E}[f_t]) + \beta_i^\top f_t + \varepsilon_{i,t}, \quad \text{for all } t \in [T], \tag{2}$$

where $\beta_i = (\beta_{i,1}, \ldots, \beta_{i,p})^\top \in \mathbb{R}^p$ is the vector of factor exposures corresponding to the factors f_t, $\gamma = (\gamma_1, \ldots, \gamma_p)^\top \in \mathbb{R}^p$ is the vector of corresponding factor risk premia, and $\varepsilon_{i,t} \in \mathbb{R}$ is the residual satisfying $\mathrm{E}[\varepsilon_{i,t}] = 0$ and $\mathrm{E}[\varepsilon_{i,t} f_t] = \mathbf{0}_p$. The vector of SDF loadings ψ is related to the vector of factor risk premia γ via the following equation:

$$\psi = (\mathrm{E}[(f_t - \mathrm{E}[f_t])(f_t - \mathrm{E}[f_t])^\top])^{-1} \gamma, \tag{3}$$

where we assumed that the matrix $\mathrm{E}[(f_t - \mathrm{E}[f_t])(f_t - \mathrm{E}[f_t])^\top]$ is invertible. In addition, by using standard least-squares projection formulas together with (2) and (3), one can show that

$$\mathrm{E}[r_{i,t}] = \sum_{j \in [p]} C_{i,j} \psi_j, \tag{4}$$

where $\mathrm{C}_{i,j} = \mathrm{E}[(f_{t,j} - \mathrm{E}[f_{t,j}])(r_{i,t} - \mathrm{E}[r_{i,t}])]$ is the covariance between $r_{i,t}$ and $f_{t,j}$.

In this section, our task is to propose a new estimator of the vector of SDF loadings ψ that can be used in the case when the number of factors p is large, potentially much larger than the number of assets N and/or the number of time periods T. As we will see, the classic estimation techniques may fail in this case. In fact, one can show that the SDF loadings cannot be consistently estimated without further assumptions in this case. Therefore, to make progress, and following the literature on estimation of high-dimensional models, we assume that the model is sparse, in the sense that most of the variation in the SDF in equation (1) can be explained by a relatively small number of factors. In other words, only a relatively small number of coefficients in the vector ψ are sufficiently different from zero.

2.2. *Estimation*

To develop intuition for the proposed method, let us first recall the classic Fama-MacBeth procedure that can be used to estimate the vector of risk premia γ in the low-dimensional case, i.e. when p is small. The procedure consists of two steps. First, using equation (2), for each asset $i \in [N]$, we run a time series OLS regression of $r_{i,t}$ on f_t, with intercept included, to obtain estimates of factor exposures $\widehat{\beta}_i$. Second, we run a cross-sectional OLS regression of average return $\bar{r}_i = T^{-1}\sum_{t=1}^{T} r_{i,t}$ on $\widehat{\beta}_i$ (without intercept). The vector of slope estimates in this regression is then an estimator of the vector of risk premia γ. In turn, once the vector of risk premia is estimated, the vector of SDF loadings ψ can be estimated using an empirical version of equation (3).

The procedure described above is consistent when the number of factors p is small. When the number of factors p is large, however, all three steps may yield large estimation errors or fail as they require inverting estimates of poorly-conditioned/singular matrices. To deal with this problem, we propose a forward selection algorithm that starts with no factors and then adds factors iteratively one by one with the goal of achieving the highest possible R^2 in the cross-sectional regression of the Fama-MacBeth procedure. We stop the algorithm once a desired number of factors is achieved or the growth in the R^2 becomes sufficiently small. Once a set of factors is selected, we perform the Fama-MacBeth procedure and use an empirical version of equation (3) to estimate ψ as described above on selected factors. We set an estimate of the SDF loadings for non-selected factors to be zero.

To define our method more precisely, for each subset S of $[p]=\{1,\ldots,p\}$, consider the Fama-MacBeth procedure using only factors in the set S. In particular, define

$$\widehat{\beta}_{i,\mathrm{S}}=\left(\sum_{t\in[T]}(f_{t,\mathrm{S}}-\bar{f}_{\mathrm{S}})(f_{t,\mathrm{S}}-\bar{f}_{\mathrm{S}})^{\top}\right)^{-1}\left(\sum_{t\in[T]}(f_{t,\mathrm{S}}-\bar{f}_{\mathrm{S}})r_{i,t}\right), \tag{5}$$

where $\bar{f}=(\bar{f}_1,\ldots,\bar{f}_p)^{\top}=T^{-1}\sum_{t=1}^{T}f_t$ and the inverse of the matrix is understood as the Moore-Penrose generalized inverse if the matrix is singular.[1] Also, let $R^2_{FM}(\mathrm{S})$ be the R^2 in the cross-sectional OLS regression of $\bar{r}_i$ on $\widehat{\beta}_{i,\mathrm{S}}$. We apply the forward selection algorithm with the aim to maximize $R^2_{FM}(\mathrm{S})$. For a desired number of factors $\widehat{s}$, the algorithm takes the following form:

ALGORITHM 1—Forward Selection Fama-MacBeth Procedure:

1. *Set* $\mathrm{S}=\emptyset$.
2. *Set* $\widehat{j}=\arg\max_{j\in[p]\setminus\mathrm{S}}R^2_{FM}(\mathrm{S}\cup j)$.
3. *Set* $\mathrm{S}=\mathrm{S}\cup\widehat{j}$.
4. *If* $|\mathrm{S}|=\widehat{s}$*, then stop. Otherwise, proceed to Step 2.*

Alternatively, the algorithm can be stopped when adding an extra factor does not significantly increase the R^2, i.e. $R^2_{FM}(\mathrm{S}\cup\widehat{j})-R^2_{FM}(\mathrm{S})\le\epsilon$ for some small value of the tolerance parameter ϵ and $\widehat{j}$ appearing in Step 2 of the algorithm. Also, some factors that are deemed to be important, e.g. Fama-French factors, can be included in the algorithm in the first step, i.e. we can set $\mathrm{S}=\mathrm{S}_0$ for some pre-specified set of factors S_0 in Step 1 of the algorithm instead of setting $\mathrm{S}=\emptyset$.

The result of Algorithm 1 is the set of factors S, which we denote as $\widehat{\mathrm{S}}$ throughout the paper. Using this set, we estimate the vector of SDF loadings by $\widehat{\psi}=(\widehat{\psi}_1,\ldots,\widehat{\psi}_p)^{\top}$, where we set

$$\widehat{\psi}_{\widehat{\mathrm{S}}}=\left(\frac{1}{T}\sum_{t\in[T]}(f_{t,\widehat{\mathrm{S}}}-\bar{f}_{\widehat{\mathrm{S}}})(f_{t,\widehat{\mathrm{S}}}-\bar{f}_{\widehat{\mathrm{S}}})^{\top}\right)^{-1}\left(\sum_{i\in[N]}\widehat{\beta}_{i,\widehat{\mathrm{S}}}\widehat{\beta}_{i,\widehat{\mathrm{S}}}^{\top}\right)^{-1}\left(\sum_{i\in[N]}\widehat{\beta}_{i,\widehat{\mathrm{S}}}\bar{r}_i\right) \tag{6}$$

[1] Note that the vector $\widehat{\beta}_{i,\mathrm{S}}$ is not a sub-vector of some p-dimensional vector $\widehat{\beta}_i$. In fact, even if we have $j\in\mathrm{S}_1\cap\mathrm{S}_2$ for some j, it is clearly possible that components of the vectors $\widehat{\beta}_{i,\mathrm{S}_1}$ and $\widehat{\beta}_{i,\mathrm{S}_2}$ corresponding to the factor j may differ.

and $\widehat{\psi}_j = 0$ for all $j \in \widehat{\mathrm{S}}^c$. We will prove consistency and derive the rate of convergence of this estimator in Section 2.4 below.

REMARK 1—Equivalent Form of Estimator $\widehat{\psi}$: For each $i \in [N]$ and $j \in [p]$, let $\widehat{\mathrm{C}}_{i,j} = T^{-1}\sum_{t\in[T]}(f_{t,j} - \bar{f}_j)r_{i,t}$ be the time series sample covariance between the factor $f_{t,j}$ and the return $r_{i,t}$. Also, let $\widehat{\mathrm{C}} = (\widehat{\mathrm{C}}_{i,j})_{i\in[N],j\in[p]}$ be the matrix of all sample covariances. Under conditions to be imposed in Section 2.4, with probability approaching one, the matrices $T^{-1}\sum_{t\in[T]}(f_{t,\mathrm{S}} - \bar{f}_{\mathrm{S}})(f_{t,\mathrm{S}} - \bar{f}_{\mathrm{S}})^\top$ will be non-singular for all S used in Algorithm 1. However, whenever these matrices are non-singular, substituting (5) into (6), it follows that $\widehat{\psi}_{\widehat{\mathrm{S}}}$ can be equivalently rewritten as

$$\widehat{\psi}_{\widehat{\mathrm{S}}} = \left(\widehat{\mathrm{C}}_{\widehat{\mathrm{S}}}^\top \widehat{\mathrm{C}}_{\widehat{\mathrm{S}}}^\top\right)^{-1} \widehat{\mathrm{C}}_{\widehat{\mathrm{S}}}^\top \bar{r}, \tag{7}$$

where $\bar{r} = (\bar{r}_1, \ldots, \bar{r}_N)^\top$ is the vector of sample average returns, which means that $\widehat{\psi}_{\widehat{\mathrm{S}}}$ can be computed by the cross-sectional OLS regression of $\bar{r}_i$ on $\widehat{\mathrm{C}}_{i,\widehat{\mathrm{S}}}$. ▪

REMARK 2—Equivalent Forms of Forward Selection Fama-MacBeth Procedure: As in the previous remark, under our conditions, the matrices $T^{-1}\sum_{t\in[T]}(f_{t,\mathrm{S}} - \bar{f}_{\mathrm{S}})(f_{t,\mathrm{S}} - \bar{f}_{\mathrm{S}})^\top$ will be non-singular for all S used in Algorithm 1 with probability approaching one. However, whenever these matrices are non-singular, it follows from equation (5) that $R^2_{FM}(\mathrm{S})$ will be numerically equal to the R^2 in the cross-sectional OLS regression of $\bar{r}_i$ on $\widehat{\mathrm{C}}_{i,\mathrm{S}}$ as multiplying the matrix of covariates by a non-singular matrix does not change the R^2 in OLS regressions. Thus, we can equivalently maximize the R^2 in the latter regression using Algorithm 1 to obtain the same set of factors $\widehat{\mathrm{S}}$. This equivalent form of Algorithm 1, together with equation (4), explains why the proposed algorithm is expected to select factors whose SDF loadings are sufficiently different from zero.

Another equivalent form of Algorithm 1 can be obtained by replacing the multivariate betas (5) in the cross-sectional OLS regression by univariate betas. Indeed, for each $i \in [N]$ and $j \in [p]$, let $\widetilde{\beta}_{i,j}$ be the slope coefficient in the univariate time series OLS regression of $r_{i,t}$ on $f_{t,j}$, with intercept included:

$$\widetilde{\beta}_{i,j} = \left(\sum_{t\in[T]}(f_{t,j} - \bar{f}_j)^2\right)^{-1}\left(\sum_{t\in[T]}(f_{t,j} - \bar{f}_j)r_{i,t}\right).$$

Also, let $\widetilde{\beta}_i = (\widetilde{\beta}_{i,1}, \dots, \widetilde{\beta}_{i,p})^\top$ be the vector of all univariate betas corresponding to the asset i. Under conditions in Section 2.4, we will have $T^{-1}\sum_{t\in[T]}(f_{t,j} - \bar{f}_j)^2 > 0$ for all $j \in [p]$ with probability approaching one, and whenever this happens, the R^2 in the cross-sectional OLS regression of $\bar{r}_i$ on $\widetilde{\beta}_{i,\mathrm{S}}$ will be equal to the R^2 in the cross-sectional OLS regression of $\bar{r}_i$ on $\widehat{\mathrm{C}}_{i,\mathrm{S}}$, yielding another equivalent form of Algorithm 1. ▪

REMARK 3—Relation to Literature: Forward selection is one of the most popular machine learning algorithms; see Hastie et al. (2009). In the context of high-dimensional linear regression models, it was analyzed, for example, by Zhang (2009), Das and Kempe (2018), and Kozbur (2020). Elenberg et al. (2018) studied forward selection in the context of a general criterion function optimization and showed that certain guarantees can be provided as long as the criterion function satisfies the so-called restricted strong convexity conditions. However, there are no available results in the literature in the context of the Fama-MacBeth procedure, which is a central estimation technique in finance. Importantly, existing results can not be directly applied in this context, with the main challenge stemming from the multi-step nature of the Fama-MacBeth procedure, where the regression covariates evolve as new factors are added, complicating the analysis. Our paper thus contributes to the literature by establishing the forward selection algorithm and its econometric properties in the context of the Fama-MacBeth procedure as a method to select factors explaining the SDF variation when the number of factors is large, potentially much larger than the number of time periods. In the context of factor selection, our algorithm is related to but substantially different from that in Harvey and Liu (2021), who proposed maximizing the significance of the added factors instead of the regression R^2 but did not provide theoretical guarantees for the proposed algorithm. ▪

2.3. *Debiased Estimation*

Here, we explain how to debias the estimator $\widehat{\psi} = (\widehat{\psi}_1, \dots, \widehat{\psi}_p)^\top$ proposed in the previous section. To develop intuition for the debiasing procedure, fix any component of the vector $\widehat{\psi}$, say $\widehat{\psi}_j$, and observe that even though Algorithm 1 is expected to select factors whose SDF loadings are sufficiently different from zero, it may not select factors whose SDF loadings are close to zero. The estimator $\widehat{\psi}_j$ may thus be biased for two reasons. First, ψ_j may be close to zero itself, in which case Algorithm 1 may not be able to select it, leading to the bias in $\widehat{\psi}_j$ as long as $\psi_j \neq 0$. Second, even if ψ_j is not close to zero and the factor j is selected, Algorithm 1 may fail to select other factors k whose covariance with returns $\mathrm{C}_{i,k}$ is cross-sectionally correlated with $\mathrm{C}_{i,j}$, leading to the omitted variable bias in $\widehat{\psi}_j$ as long as $\psi_k \neq 0$, which follows from the OLS characterization of the estimator $\widehat{\psi}_{\widehat{\mathrm{S}}}$ in equation (7).

Two reasons for the bias of $\widehat{\psi}_j$ suggest that we can debias it by appropriately expanding the set $\widehat{\mathrm{S}}$. Specifically, we need to add to the set $\widehat{\mathrm{S}}$ the factor j itself and also all factors k whose covariance with returns $\mathrm{C}_{i,k}$ is cross-sectionally correlated with $\mathrm{C}_{i,j}$. In practice, since neither $\mathrm{C}_{i,j}$ nor $\mathrm{C}_{i,k}$ is observed, we work with sample covariances $\widehat{\mathrm{C}}_{i,j}$ and $\widehat{\mathrm{C}}_{i,k}$.

More formally, we proceed as follows. First, as before, let $\widehat{\mathrm{S}}$ be the set S produced by Algorithm 1. Second, let $\widehat{\mathrm{S}}_j$ be the set produced by Algorithm 1 with $R^2_{FM}(\mathrm{S})$ replaced by the R^2 in the cross-sectional OLS regression of $\widehat{\mathrm{C}}_{i,j}$ on $\widehat{\mathrm{C}}_{i,\mathrm{S}}$ (without intercept) and with $[p]$ replaced by $[p]\backslash\{j\}$. Third, set $\widehat{\mathrm{S}}_{D,j} = \widehat{\mathrm{S}} \cup \widehat{\mathrm{S}}_j \cup \{j\}$. Fourth, let $\widehat{\psi}$ be the estimator defined in Section 2.2 but with the set $\widehat{\mathrm{S}}$ there replaced by the set $\widehat{\mathrm{S}}_{D,j}$ here. The corresponding component $\widehat{\psi}_j$ of the vector $\widehat{\psi}$ will be a debiased estimator of ψ_j, which we denote as $\widehat{\psi}_{D,j}$. By repeating this procedure for all $j \in [p]$, we obtain the full vector $\widehat{\psi}_D = (\widehat{\psi}_{D,1}, \ldots, \widehat{\psi}_{D,p})^\top$. In the next section, we will prove that for each $j \in [p]$, the estimator $\widehat{\psi}_{D,j}$ is $\sqrt{T}$-consistent and $\sqrt{T}(\widehat{\psi}_{D,j} - \psi_j)$ converges to a centered normal distribution.

REMARK 4—Relation to Literature: The debiasing procedure proposed here is similar in flavor to that proposed in the context of Lasso variable selection for high-dimensional linear regression models by Belloni et al. (2014). Although our setting is different, the intuition underlying the debiasing procedure is the same. Related procedures, also in the context of Lasso variable selection for high-dimensional linear regression models, were proposed by Zhang and Zhang (2014), Javanmard and Montanari (2014), and de Geer et al. (2014). In the context of SDF loadings estimation, our debiasing procedure is related to but different from that in Feng et al. (2020), who relied upon Lasso factor selection. ∎

2.4. *Asymptotic Theory*

Throughout the rest of the paper, we consider asymptotics where T gets large, i.e. $T\to\infty$. We assume that the number of assets N and the number of factors p, as well as all coefficients and distributions of random vectors in the model, are allowed to vary with T. In most cases, however, we keep this dependence implicit.

Let $k,K>0$ be constants such that $k\le K$. Also, let $\bar s_T\ge 1$ be an integer that is allowed to depend on T (and on N and p through T). In addition, let $\mathrm{C}=(\mathrm{C}_{i,j})_{i\in[N],j\in[p]}$ be the matrix of covariances between returns $r_{i,t}$ and factors $f_{t,j}$ and let $\widehat{\mathrm{C}}=(\widehat{\mathrm{C}}_{i,j})_{i\in[N],j\in[p]}$ be the corresponding matrix of sample covariances. Moreover, for brevity of notation, for all $t\in[T]$, let $v_t=(v_{t,1},\dots,v_{t,p})^\top=f_t-\mathrm{E}[f_t]$ be the demeaned version of the vector of factors f_t. Finally, to state some of the assumptions below in a compact form, let $u_{1,t}=1$ and $u_{2,t}=\psi^\top v_t$ for all $t\in[T]$ and $u_{i+2,t}=\varepsilon_{i,t}$ and $u_{i+2+N,t}=\beta_i^\top v_t$ for all $i\in[N]$ and $t\in[T]$. Similarly, let $w_{t,1}=1$ and $w_{t,2}=\psi^\top v_t$ for all $t\in[T]$ and $w_{t,j+2}=v_{t,j}$ for all $t\in[T]$ and $j\in[p]$.

ASSUMPTION 1—Uniform Law of Large Numbers, I: *We have*

$$\max_{i\in[2N+2]}\max_{j\in[p+2]}\left|\frac{1}{T}\sum_{t\in[T]}u_{i,t}w_{t,j}-\mathrm{E}[u_{i,t}w_{t,j}]\right|\lesssim_P\sqrt{\frac{\log(Np)}{T}}.$$

Assumption 1 is a quantitative version of the uniform law of large numbers. For example, given that $\mathrm{E}[\varepsilon_{i,t}v_{t,j}]=0$, we expect from the classical central limit theorems for time series data that $T^{-1}\sum_{t=1}^T\varepsilon_{i,t}v_{t,j}\lesssim_P\sqrt{1/T}$. Provided that the random variables $\varepsilon_{i,t}v_{t,j}$ have sufficiently light tails, we then also expect that $T^{-1}\sum_{t\in[T]}\varepsilon_{i,t}v_{t,j}\lesssim_P\sqrt{\log(Np)/T}$ uniformly over $(i,j)\in[N]\times[p]$. Assumption 1 imposes conditions of this type on various random variables appearing in our analysis below. Conditions like Assumption 1 are very common in the literature on high-dimensional estimation.

ASSUMPTION 2—SDF Loadings Sparsity: *There exists a set* $\mathrm{S}_0\subset[p]$ *such that*

$$|\mathrm{S}_0|\le\bar s_T,\quad \|\mathrm{C}\psi-\mathrm{C}_{\mathrm{S}_0}\psi_{\mathrm{S}_0}\|_2^2\lesssim\frac{N\log(Np)}{T},\quad \textit{and}\quad \|\psi_{\mathrm{S}_0^c}\|_1^2\lesssim\frac{|\mathrm{S}_0|\log(Np)}{T}.$$

Assumption 2 states that the vector of SDF loadings ψ is approximately sparse. In particular, it requires that a sub-vector ψ_{S_0} of the vector ψ of size at most $\bar s_T$ is sufficient to explain most of the cross-sectional variation of mean returns; see equation (4) for the relationship between SDF loadings and mean returns. Some researchers, e.g. Nagel (2021), argue against sparsity in finance but in different contexts, e.g. in the context of explaining cross-sectional variation of mean returns by past returns. We emphasize that the type of sparsity we impose is different. Indeed, *all* cross-sectional variation of mean returns is explained by just one factor: the stochastic discount factor. Thus, if it happens that one of the factors in the vector f_t is equal to the stochastic discount factor, Assumption 2 will be satisfies with a singleton set S_0. Also, although we allow the set S_0 to depend on T (and on N and p through T), we omit this dependence for brevity of notation.

ASSUMPTION 3—Sparse Eigenvalues, I: *We have*

$$\lambda_{\min}\left(\frac{1}{T}\sum_{t\in[T]}(f_{t,\mathrm{S}}-\bar{f}_{\mathrm{S}})(f_{t,\mathrm{S}}-\bar{f}_{\mathrm{S}})^{\top}\right)>0$$

for all $\mathrm{S}\subset[p]$ *such that* $|\mathrm{S}|\le\bar{3}s_T+1$ *w.p.a.1.*

Provided $\lambda_{\min}(\mathrm{E}[v_t v_t^{\top}])$ is bounded away from zero and $\bar{s}_T$ does not grow too fast, Assumption 3 can be proven to hold if the factors f_t exhibit relatively weak time series dependence and have sufficiently light tails; e.g. see Belloni and Chernozhukov (2009).

ASSUMPTION 4—Sparse Eigenvalues, II: *We have*

$$k\le\lambda_{\min}\left(\frac{\mathrm{C}_{\mathrm{S}}^{\top}\mathrm{C}_{\mathrm{S}}}{N}\right)\le\lambda_{\max}\left(\frac{\mathrm{C}_{\mathrm{S}}^{\top}\mathrm{C}_{\mathrm{S}}}{N}\right)\le K$$

for all $\mathrm{S}\subset[p]$ *such that* $|\mathrm{S}|\le 3\bar{s}_T+1$.

The lower bound in Assumption 4 is similar in flavor to that in Assumption 3. It means that there is no multicollinearity in the vectors of covariances $\mathrm{C}_{\{j\}}$ when we restrict attention to a relatively small number of factors j. The upper bound should be viewed as a mild technical condition. Note also that Assumption 4 is our key identification condition. Indeed, Lewellen et al. (2010) showed that without this assumption, it is possible that factors with zero SDF loadings can yield high R^2 in the cross-sectional

OLS regression as long as they are correlated with the factors explaining the SDF. Assumption 4 ensures that this phenomenon does not occur.

ASSUMPTION 5—Estimator: *We have* $|\widehat{\mathrm{S}}| \le \bar{s}_T$ *w.p.a.1. In addition, for some constant* $c > 0$*, we have* $|\widehat{\mathrm{S}}| \ge (1+c)(K/k)|\mathrm{S}_0|\log T$ *w.p.a.1. Moreover,* $|\widehat{\mathrm{S}}| \lesssim_P |\mathrm{S}_0|\log T$.

The first part of this assumption means that we carry out at most $\bar{s}_T$ rounds in our forward selection Fama-MacBeth procedure, Algorithm 1, so that we select at most $\bar{s}_T$ factors. This condition is rather natural given the assumed sparsity of the true vector ψ. The second part of Assumption 5 implies that the number of selected factors $|\widehat{\mathrm{S}}|$ should be larger than the number of important factors $|\mathrm{S}_0|$ at least by a multiplicative constant proportional to $\log T$. We impose this condition because the forward selection algorithm may occasionally select unimportant factors by chance. The third part of Assumption 5 gives an upper bound on the number of selected factors but is less essential than the first two parts. Indeed, selecting more factors than the number in the upper bound will undermine the convergence rate of our estimator $\widehat{\psi}$ but will not make it inconsistent unless way too many factors are selected. We leave an important question of how to choose the number of selected factors $|\widehat{\mathrm{S}}|$ in such a way that Assumption 5 is satisfied to future work.

ASSUMPTION 6—Expected Returns: *We have* $\max_{i\in[N]} \mathrm{E}[r_{i,t}^2] \lesssim 1$.

ASSUMPTION 7—Growth Condition, I: *We have* $\bar{s}_T \log(NTp) = o(T)$.

Assumption 6 is a very mild regularity condition that is expected to hold in most cases. Assumption 7 restricts how fast $\bar{s}_T$, N, and p are allowed to grow relative to T. Most importantly, it allows the number of factors p to be much larger than the number of time periods T. Together with Assumption 2, it also requires that the size of the approximating vector ψ_{S_0} is small relative to the number of time periods T.

THEOREM 1—Convergence Rate: *Suppose that Assumptions 1 – 7 are satisfied. Then*

$$\|\widehat{\psi} - \psi\|_2 \lesssim_P \sqrt{\frac{|\mathrm{S}_0|\log^2(NTp)}{T}} \quad \textit{and} \quad \|\widehat{\psi} - \psi\|_1 \lesssim_P \sqrt{\frac{|\mathrm{S}_0|^2\log^3(NTp)}{T}}.$$

REMARK 5—Consistency of $\widehat{\psi}$: This theorem implies that the proposed estimator $\widehat{\psi}$ is consistent in the ℓ_2-norm if $|\mathrm{S}_0|\log^2(NTp) = o(T)$ and is consistent in the ℓ_1-norm if $|\mathrm{S}_0^2|\log^3(NTp) = o(T)$. Thus, the estimator remains consistent even if the number of factors p and the number of assets N are growing much faster than the number of time periods T. Also, none of the assumptions of Theorem 1 requires that $N \to \infty$ as $T \to \infty$, which implies that the estimator is consistent even if the number of assets N remains bounded as T gets large. In addition, note that the bounds derived in Theorem 1 are $\sqrt{\log(NTp)}$ worse than the bounds typically expected for the Lasso-based estimators. We do not know if this is an artifact of our proof technique or this is because forward selection generally requires more selected variables to guarantee the same fit as the Lasso-based selection. ∎

To formulate the asymptotic normality result for the debiased estimator $\widehat{\psi}_D$, for all $j \in [p]$, let $\eta_j = (\eta_{j,1}, \dots, \eta_{j,p})^\top \in \mathbb{R}^p$ be the vector defined by $\eta_{j,j} = 0$ and

$$\eta_{j,\{j\}^c} = \Big(\mathrm{E}[v_{t,\{j\}^c} v_{t,\{j\}^c}^\top]\Big)^{-1} \mathrm{E}[v_{t,\{j\}^c} v_{t,j}], \tag{8}$$

so that $\eta_{j,\{j\}^c}$ is the vector of coefficients of the time-series least-squares projection of the demeaned factor $v_{t,j}$ on other demeaned factors $v_{t,\{j\}^c}$. Also, for all $t \in [T]$, let $z_{t,j} = v_{t,j} - \eta_{j,\{j\}^c}^\top v_{t,\{j\}^c}$ be the corresponding residual and let $\sigma_{z,j}^2 = \mathrm{E}[z_{t,j}^2]$ be the residual variance. The asymptotic normality result requires several additional assumptions but, for brevity of the main text, we have placed them in Appendix A.

THEOREM 2—Asymptotic Normality: *Suppose that Assumptions 1 – 6 above and Assumptions 8 – 14 listed in Appendix A are satisfied. Then for all $j \in [p]$,*

$$\sqrt{T}(\widehat{\psi}_{D,j} - \psi_j) \to_d N(0, \sigma_{\psi,j}^2),$$

where

$$\sigma_{\psi,j}^2 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathrm{E}\left[\frac{(z_{t,j} m_t - \mathrm{E}[z_{t,j} m_t])(z_{s,j} m_s - \mathrm{E}[z_{s,j} m_s])}{\sigma_{z,j}^4} \right].$$

REMARK 6—Comparison of $\widehat{\psi}$ and $\widehat{\psi}_D$: It is possible to show that the individual components $\widehat{\psi}_j$ of the estimator $\widehat{\psi}$ converge to the corresponding true values ψ_j with the rate that is slower than $\sqrt{1/T}$. It is therefore fair to say that for each $j \in [p]$, the estimator $\widehat{\psi}_{D,j}$ outperforms $\widehat{\psi}_j$. However, it does not mean that the vector-valued estimator $\widehat{\psi}_D$, as a whole, outperforms $\widehat{\psi}$. For example, because of the convergence result in Theorem 2, it is expected that the estimation error $\|\widehat{\psi}_D - \psi\|_2$ is typically of order $\sqrt{p/T}$, which could be much worse than the estimation error $\|\widehat{\psi} - \psi\|_2$ derived in Theorem 1. It is therefore important to emphasize that both estimators have their advantages: the estimator $\widehat{\psi}$ is useful when we are interested in the values of all SDF loadings, whereas $\widehat{\psi}_D$ is useful when we are interested in the SDF loading of a particular factor. ∎

REMARK 7—Estimating $\sigma^2_{\psi,j}$: Observe that $\sigma^2_{\psi,j}$ is the long-run variance of the normalized sample average $T^{-1/2}\sum_{t\in[T]}(z_{t,j}m_t - \mathrm{E}[z_{t,j}m_t])/\sigma^2_{z,j}$, which suggests the following plug-in estimator thereof. First, our assumptions in Appendix A imply that the vector $\eta_{j,\{j\}^c}$ is bounded in the ℓ_1-norm. We thus can estimate $z_{t,j}$ as $\widehat{z}_{t,j} = (f_{t,j} - \bar{f}_j) - \widehat{\eta}^\top_{j,\{j\}^c}(f_{t,\{j\}^c} - \bar{f}_{\{j\}^c})$, where $\widehat{\eta}_{j,\{j\}^c}$ is the Lasso estimator of $f_{t,j} - \bar{f}_j$ on $f_{t,\{j\}^c} - \bar{f}_{\{j\}^c}$. Second, because of the consistency of $\widehat{\psi}$ implied by Theorem 1, we can estimate m_t as $\widehat{m}_t = 1 - \widehat{\psi}^\top(f_t - \bar{f})$. Third, we can estimate $\sigma^2_{z,j}$ as $\widehat{\sigma}^2_{z,j} = T^{-1}\sum_{t\in[T]}\widehat{z}^2_{t,j}$. Finally, we can estimate $\sigma^2_{\psi,j}$ by applying the standard long-run variance estimation formulas to $T^{-1/2}\sum_{t\in[T]}(\widehat{z}_{t,j}\widehat{m}_t - \bar{z}_j\bar{m}_t])/\widehat{\sigma}^2_{z,j}$, where we set $\bar{z}_j = T^{-1}\sum_{t\in[T]}\widehat{z}_{t,j}$ and $\bar{m}_t = T^{-1}\sum_{t\in[T]}\widehat{m}_t$. It is straightforward to prove that the estimator $\widehat{\sigma}^2_{\psi,j}$ is consistent for $\sigma^2_{\psi,j}$ under our conditions. ∎

REMARK 8—Relation to Literature: The asymptotic variance $\sigma^2_{\psi,j}$ appearing in Theorem 2 is the same as the asymptotic variance in Theorem 1 of Feng et al. (2020). In fact, it follows from the proof of Theorem 2 that our debiased forward-selection-based estimator $\widehat{\psi}_{D,j}$ is asymptotically equivalent to the corresponding Lasso-based estimator in Feng et al. (2020), which can be viewed as a consequence of the results on equivalence of semiparametric estimators in Newey (1994). Our Theorem 2 thus complements Theorem 1 in Feng et al. (2020) by expanding the set of machine learning methods that are theoretically proven to work in the context of inference on SDF loadings.

∎

3. EMPIRICAL ANALYSIS

In this section, we present the data used in the empirical analysis and the main results of the selection of the high-order factors by the forward selection Fama-MacBeth procedure.

3.1. *Data*

We present a description of the data used for the estimation of the FS-FMB procedure. In order to estimate the model and perform cross-sectional asset pricing tests, we use data from various sources. Test assets are 484 characteristic-managed portfolios from Kozak et al. (2020).[2] These portfolios are constructed by weighting stocks based on their characteristic signals, which are based on cross-sectional ranks. Small firms (below 0.01% of market cap) are excluded. Baseline factors are the Fama and French (2015)'s five factors and the momentum factor of Jegadeesh and Titman (1993), which we obtain from Kenneth French's data library. We refer to these six baseline factors as the FF5M model. Furthermore, we consider factors for 153 characteristics in 13 themes, using data from the US equity market, constructed by Jensen, Kelly, and Pedersen (2023).[3] We refer to these 153 factors as the factor zoo (see, e.g., Feng, Giglio, and Xiu (2020)). The sample starts in October 1973 and ends in December 2019.

We construct factors based on the FF5M factors. We denote both the higher-orders and the interactions of FF5M as higher-order factors. Specifically, for any factor f_i in the FF5M, we first form all the f_i^2 and f_i^3 as candidate factors. Moreover, we form all the pairwise interactions of degree 2 and 3. That is, for any two factors f_i and f_j in the FF5M model, we form the pairwise interactions of the type $f_i \times f_j$ and $f_i \times f_j^2$ as candidate factors. Therefore, the total number of candidate higher-order factors is 57, where there are 6 f_i^2, 6 f_i^3, 15 $f_i \times f_j$, and 30 $f_i \times f_j^2$.

Table I presents summary statistics of the test assets and factors. For the mean, the standard deviation, and the quantiles, we report averages across assets. All returns are monthly, in excess of the US risk-free rate, and reported in percentage. The average mean return for the test assets is 0.65% monthly, with a standard deviation of 4.96%.

---

[2] Portfolios are from Serhiy Kozak's web page (https://serhiykozak.com/#/data-code).

[3] Factors are capped value weighted from Global Factor Data (https://jkpfactors.com/).

TABLE I

SUMMARY STATISTICS

This table reports summary statistics of the tests assets and factors. For the mean, the standard deviation and the quantiles, we report averages across assets. Returns are monthly and in excess of the US risk-free rate. Test assets are from Kozak et al. (2020). The Fama-French five factors and the momentum factor (FF5M) are from Ken French's data library. "Interactions" correspond to pairwise interactions of the FF5M factors of degree 2 (i.e., $f_i \times f_j$) and degree 3 (i.e., $f_i^2 \times f_j$). "Powers" correspond to the power of the individual FF5M factors of degree 2 (i.e., f_i^2) and degree 3 (i.e., f_i^3). The factor zoo includes 153 factors based on US equity data constructed by Jensen et al. (2023). The sample period is October 1973 to December 2019.

$R_{i,t}$ (%)	Mean	Std	$q_{5\%}$	$q_{50\%}$	$q_{95\%}$	N	T
Test Assets	0.65	4.97	-7.58	0.9	8.1	484	555
FF5M	0.39	3.16	-4.63	0.4	5.13	6	555
Interactions degree 2	-0.48	16.47	-17.82	-0.02	14.12	15	555
Interactions degree 3	0.53	20.14	-6.6	0.02	8.44	30	555
Powers degree 2	11.1	28.62	0.03	3.34	42.75	6	555
Powers degree 3	-9.83	581.32	-147.67	0.23	167.67	6	555
Factor zoo	0.23	2.71	-3.5	0.19	4.14	153	555

The average return for the FF5M factors is lower and equal to 0.39% monthly, with a standard deviation of 3.16%. The average return for the factor zoo is even lower and equal to 0.23% monthly, with a standard deviation of 2.71%. The higher-order factors have a distribution characterized by more extreme realizations, as exemplified by the 95% quantile of the powers of degree 2 and 3 equal, respectively, to 42.75% and 167.67%. We note that the higher order factors are generally non-tradable. Figure E.1 in the Appendix plots the distributions of the returns of the high-order factors.

3.2. *Selected Factors by Forward Selection Fama-MacBeth Regressions*

Our goal is to study whether and which of the higher-order factors constructed from the six factors in the baseline FF5M are important in explaining the SDF. To select the most relevant higher-order factors to the SDF, we use the FS-FMB procedure that we have discussed in detail in Section 2 above. We start with the six baseline FF5M factors and add the higher-order interactions one at a time.

We present the main estimation results from the second step of the FS-FMB procedure in Table II. Panel A shows the performance of the baseline models using the 484 characteristic-managed portfolios from Kozak et al. (2020). The baseline models include the standard CAPM, the Fama and French (1993)'s 3-factor model (FF3), the Fama and French (2015)'s 5-factor model (FF5) and the 5-factor model augmented with the momentum factor of Jegadeesh and Titman (1993) (FF5M). The adjusted cross-sectional R-squared for the CAPM is just 0.032, and the estimated intercept (0.009) is significantly different from zero with a t-statistic of 3.6. The adjusted cross-sectional R-squared reaches 0.312 for the FF5M model, while the estimate for the intercept (0.004) remains significantly different from zero with a t-statistic of 2.2.

Panel B refers to the FS-FMB procedure. At each step, we augment the FF5M by one of the additional higher-order factors, which we select in order to maximize the adjusted cross-sectional R-squared. The selection procedure stops when the improvement in the R-squared in a given step is smaller than 1 pp. The forward selection FMB procedure selects 7 higher-order factors: the SMB factor squared (SMB2), the interaction between SMB2 and the momentum factor (SMB2*Mom), the interaction between the momentum factor squared and the profitability factor (Mom2*RMW), the market factor squared (Mkt-RF2), the interaction between Mkt-RF2 and the profitability factor (Mkt-RF2*RMW), the interaction between the market and size factors (Mkt-RF*SMB) and the interaction between the value factor squared and the market factor (HML2*Mkt-RF). Of these 7 higher-order factors that are selected, 2 are higher-orders (SMB2 and Mkt-RF2) and 5 are interactions (SMB2*Mom, Mom2*RMW, Mkt-RF2*RMW, Mkt-RF*SMB, and HML2*Mkt-RF). Interestingly, none of the higher-orders of degree 3 are selected.

The adjusted cross-sectional R-squared increases from 31.2% for the baseline FF5M model to 41% in the first step when SMB2 is included, and gradually increases to 59%, in the last step when all 7 factors are included and the procedure stops. Across the seven steps, the estimate for the intercept is around 0.003, with a t-statistic in the final step of 1.819, which is below the threshold for significance at the 5% level.

TABLE II

CROSS-SECTIONAL PERFORMANCE

This table reports the results from the second step of the FMB method. Panel A refers to baseline models: the standard CAPM, the Fama-French 3-factor, the Fama-French 5-factor and the Fama-French 5-factor and momentum models. The table reports the adjusted cross-sectional R-squared, the estimate for the intercept (α) and associated t-statistic. The t-statistics is based on Newey-West corrected standard errors. Panel B refers to the FS-FMB procedure. At each step, we augment the FF5M by one additional higher-order factor (h_j) which maximizes the adjusted cross-sectional R-squared. The method stops when the improvement in the R-squared in a given step is smaller than 1 pp.

Panel A: Baseline Models				
#	Model	Adj. R-squared	α	t-stat (α)
1	CAPM	0.032	0.009	3.635
2	FF3	0.126	0.01	4.701
3	FF5	0.275	0.006	2.899
4	FF5M	0.312	0.004	2.201
Panel B: FS-FMB procedure				
Step	h_j	Adj. R-squared	α	t-stat (α)
1	SMB2	0.41	0.003	1.821
2	SMB2*Mom	0.467	0.004	2.02
3	Mom2*RMW	0.495	0.003	1.324
4	Mkt-RF2	0.529	0.003	1.769
5	Mkt-RF2*RMW	0.552	0.003	1.636
6	Mkt-Rf*SMB	0.572	0.004	2.161
7	HML2*Mkt-RF	0.587	0.003	1.819

We then examine the SDF loadings of the higher-order factors selected by the FS-FMB procedure. The results are summarized in Table III. Panel A shows the non-debiased estimates based on the standard FMB two-pass regression, while Panel B presents the debiased estimates as outlined above. The directions of the point estimates based on the non-debiased procedure and the debiased procedure are the same.

TABLE III

SDF LOADINGS

This table reports the estimates for the SDF loadings (b_j) associated with the selected higher-order factors. Panel A reports non-debiased estimates while Panel B reports debiased estimates. The t-statistics is based on Newey-West corrected standard errors. High-order factors are selected by the FS-FMB procedure (see Table II).

	Panel A: Non-Debiased Estimates		Panel B: Debiased Estimates	
h_j	b_j	$t(b_j)$	b_j	$t(b_j)$
SMB2	1.158	2.43	1.403	1.837
SMB2*Mom	-1.193	-4.074	-1.004	-1.812
Mom2*RMW	-1.38	-4.105	-1.178	-2.882
Mkt-RF2	1.637	5.619	1.878	4.606
Mkt-RF2*RMW	-3.53	-4.169	-3.87	-4.301
Mkt-RF*SMB	-1.404	-3.032	-2.012	-3.636
HML2*Mkt-RF	-1.385	-2.823	-1.606	-2.039

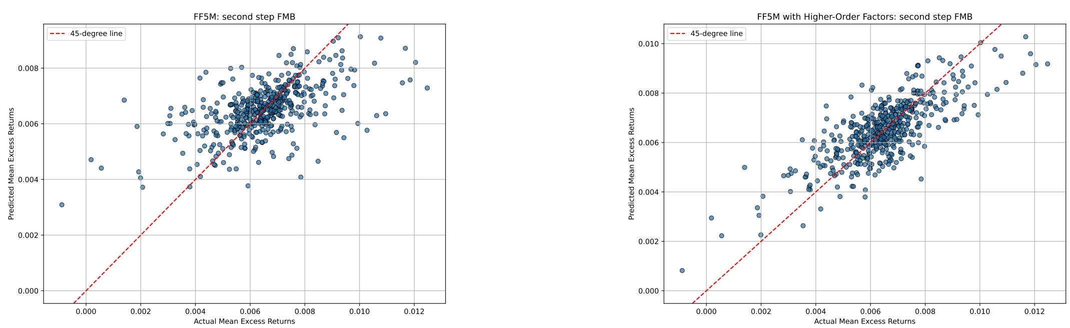
Moreover, the estimates from the two procedures are similar in magnitudes. The estimates based on the non-debiased procedure are all significant at the 5% level, while the estimates based on the debiased procedure are all significant at the 10% level with 5 of the 7 SDF loading estimates being significant at the 5% level. Overall, the results highlight that the higher-order factors are important components of the SDF and significantly price assets in the cross-section.

Figure 1 plots predicted against realized returns for all the test assets. Panel A shows the performance of the FF5M model and panel B presents the higher-order model. The FF5M has a limited ability in explaining the average returns (Panel A), in line with the relatively low R-squared of the model. On the other hand, there is a strong positive relationship between the average realized returns and the predicted returns for the higher-order model selected by the FS-FMB procedure.

FIGURE 1.—**Actual vs. Predicted Returns**

This figure plots actual mean returns of the test assets against the mean predicted returns from the FF5M model (subplot a) and the FF5M model augmented with the higher-order factors selected by the FS-FMB procedure (subplot b, see Table II). The predicted return is the OLS estimate of the betas times the estimated prices of risk.

(a) **FF5M model** (b) **FF5M + higher-order factors**



We note that the predicted return used in the graph is the OLS estimate of the betas from the first step of the two-pass FMB procedure times the estimated prices of risk. That is, we do not impose the no-arbitrage condition for the tradable factors that pins down the price of risk by the sample mean of the factor. Doing so would, by construction, result in an even worse fit by construction for the baseline FF5M models where all the factors are tradable. We present the results in Table E.I in the Online Appendix where we impose the restriction for the tradable factors. The results confirm the importance of higher-order factors in pricing assets in the cross-section.

3.3. *Out-of-Sample*

The main results presented in Table II are based on in-sample estimation. We further provide evidence of the out-of-sample performance of the higher-order model selected by the FS-FMB procedure. We consider two types of out-of-sample tests: (1) out-of-sample in terms of asset space and (2) out-of-sample in terms of time period estimations.

3.3.1. *Asset space*

In Table IV, we present an additional set of results based on cross-validation applied to the FS-FMB procedure. The cross-validation is applied to the second step of the FS-FMB procedure. This is an out-of-sample exercise in terms of asset space, where we select higher-order factors and estimate SDF loadings from the training assets, and then evaluate the out-of-sample performance in the out-of-sample assets.

Specifically, at each step of the FS-FMB procedure, we divide the sample into 5 equal-sized, non-overlapping folds. For each fold k, we use the remaining 4 folds as the training set and the left-out fold as the test set. In each training step, we estimate the cross-sectional regression of average returns on the betas and obtain estimates for the risk prices. Next, we use the estimated risk prices to compute the adjusted R-squared on the test set. This process is repeated 5 times and the adjusted cross-sectional R-squared associated with a given higher-order factor is the average, at each step, of the adjusted cross-sectional R-squareds obtained in the 5-folds cross-validation. Table IV reveals that the FS-FMB procedure with cross-validation (CV) selects a total of 8 higher-order factors, with a cross-sectional R-squared in the final step of 0.596. Moreover, the first six selected factors are the same as in the baseline method, while the last selected factor corresponds to the last from the baseline method. Furthermore, the cross-validated adjusted cross-sectional R-squareds, which represent out-of-sample measures of fit of the cross-sectional model, are close to the in-sample R-squareds at 0.535. Overall, these results show that the selected higher-order factor model has stable out-of-sample performance.

TABLE IV

FS-FMB METHOD WITH CROSS-VALIDATION

This table reports the results from the second step of the FMB regression procedure for the FS-FMB procedure with 5-fold cross-validation. At each step, we include one additional higher-order factor which maximizes the adjusted cross-sectional R-squared computed with cross-validation using 5-folds. The procedure stops when the improvement in the R-squared in a given step is smaller than 1 pp. For each step, we report the selected factor (h_j), the adjusted cross-sectional R-squared, the adjusted cross-validated (CV) cross-sectional R-squared, the estimate for the intercept (α) and associated t-statistic. The t-statistic is based on Newey-West corrected standard errors. The adjusted cross-validated cross-sectional R-squared is the average at each step of the adjusted R-squareds obtained in the 5-folds cross-validation.

Step	h_j	Adj. R-squared	Adj. CV R-squared	α	t-stat (α)
1	SMB2	0.41	0.366	0.003	1.821
2	SMB2*Mom	0.467	0.422	0.004	2.02
3	Mom2*RMW	0.495	0.443	0.003	1.324
4	Mkt-RF2	0.529	0.472	0.003	1.769
5	Mkt-RF2*RMW	0.552	0.495	0.003	1.636
6	Mkt-Rf*SMB	0.572	0.511	0.004	2.161
7	Mom2*HML	0.582	0.527	0.004	2.145
8	HML2*Mkt-RF	0.596	0.535	0.003	1.833

3.3.2. *Time periods*

We next evaluate the out-of-sample performance in the time-series of the higher-order factor model against the benchmark factor models (i.e., CAPM, FF3, FF5, and FF5M). Specifically, we first split the sample into two 50-50 subsamples based on the time series, where we denote the first half of the sample as the training sample and the second half as the out-of-sample. Then, we estimate the covariances and the SDF loadings in the training sample based on the different factor models. We denote the R^2 from these estimations as R^2_{train}. Lastly, we compare the average asset returns in the out-of-sample to the predicted asset returns using covarainces and the SDF loadings estimated from the training sample to obtain the out-of-sample R^2, or R^2_{oos}. We note that both the R^2_{train} and the R^2_{oos} are likely to be lower than the full sample R^2 for any model due to estimation errors, especially related to average returns. It is well-known that estimations of the average equity returns are noisy and require a long time series. Therefore, our focus is comparing R^2_{oos} for the higher-order factor model against those of the benchmark factor models.

Table V presents the results. The R^2_{oos}s of CAPM and the FF3 factor model are close to zero. The R^2_{oos}s of the FF5 factor model and the FF5M factor model are 7% and 8.9%, representing an improvement from the CAPM and the FF3 factor model. The R^2_{oos} of higher-order factor model selected from the FS-FMB procedure is significantly larger than the benchmark factor models at 15.8%, which is about twice as high as the R^2_{oos}s of FF5 and FF5M factor models. Additionally, in untabulated result, we repeat the exercise but select the higher-order factors in the training sample. The magnitude of the R^2_{oos} is close to Column (5) at 14.8%.

One concern is that the 50-50 split where the first half of the sample is the training sample and the second half of the sample is the out-of-sample may be arbitrary. As a robustness check, we repeat the analyses by randomly selecting, for 1,000 times, 50% of the sample as the training sample and the remaining as the out-of-sample. We report the results in Table E.II in the Online Appendix. Table E.II delivers the same message – theR^2_{oos} of the higher-order factor model is significantly higher than those of the benchmark factor models.

3.4. *Factor Zoo*

In this subsection, we show that the baseline FF5M model, augmented with the higher-order factors selected by the FS-FMB procedure, effectively subsumes the majority of the factors in the factor zoo, or zoo factors.

We employ two empirical methods to support this argument. First, we perform FMB regressions to test the significance of the SDF loading of each factor in the factor zoo, controlling for our higher-order factors. Second, we construct factor mimicking portfolios for the seven higher-order factors to conduct time-series asset pricing tests on the zoo factors.

TABLE V

COMPARISON OF R-SQUAREDS

This table presents the in-sample and out-of-sample R^2 under different factor models. The factor models include the CAPM, the Fama-French 3-factor model (FF3), the Fama-French 5-factor model (FF5), the Fama-French 5-factor plus momentum factor model (FF5M), and the higher-order factor model selected by the FS-FMB procedure. We first split the sample into two 50-50 subsamples based on the time series, where we denote the first half of the sample as the training sample and the second half as the out-of-sample. Then, we estimate the covariances and the SDF loadings in the training sample based on the different factor models. We denote the R^2 from these estimations as R^2_{train}. Lastly, we compare the average asset returns in the out-of-sample to the predicted asset returns using covariances and the SDF loadings estimated from the training sample to obtain the out-of-sample R^2, or R^2_{oos}. The table reports the R^2_{train} and R^2_{oos} for each model.

	(1) CAPM	(2) FF3	(3) FF5	(4) FF5M	(5) Higher-Order
R^2_{train}	0.003	0.204	0.273	0.312	0.449
R^2_{oos}	0.001	0.012	0.07	0.089	0.158

3.4.1. *Cross-sectional method*

The first empirical strategy to evaluate whether the high-order factors can subsume the zoo factors is based on the two-pass Fama-MacBeth procedure, motivated by the analysis in Feng et al. (2020). Specifically, for each zoo factor, we estimate its SDF loading while controlling for our higher-order factor model, which includes the FF5M and the selected higher-order factors.

Figure 2 summarizes the results. Panel A plots the distribution of the absolute values of the estimates for the intercept from the cross-sectional regressions. Panel B displays the distribution of the absolute t-statistics associated with the SDF loading of the factor zoo, using Newey and West (1987) corrected standard errors. In both panels, the blue histogram represents the estimates obtained by controlling for the FF5M factors, while the green histogram represents the estimates obtained by additionally controlling for the higher-order factors selected by the forward selection FMB procedure. Vertical dashed lines indicate the median values of each distribution – a vertical blue dashed-line for estimates obtained controlling for the FF5M factors and a vertical green dashed-line for estimates obtained controlling for the higher-order factor model. Additionally, we also include a vertical red dashed-line which corresponds to the median value with no controls, or the factor model including only the factor in the factor zoo.

Panel A shows that the intercept estimates tend to be highly statistically significant, with a median value of 4.00, when the factor model includes only the factor in the factor zoo. When the factor model includes both FF5M and the factor in the factor zoo, the median t-statistic decreases substantially but remains significant at the 5% level at 2.23. When the factor model further includes the higher-order factors, the median t-statistic is no longer statistically significant at the 5% level at 1.84.

Panel B shows that, when the factor model includes only the factor in the factor zoo, the median SDF loading estimate for the factor zoo is not statistically different from zero, with a t-statistic of 0.86. Note that these estimates are likely biased because of omitted factors (see, e.g., Feng et al. (2020)). Controlling for the FF5M, the median t-statistic increases substantially to 1.38. 34% of zoo factors become statistically significant at the 5% level with a t-statistic greater than 2 suggesting that FF5M helps to remove some bias in the factor estimates. Importantly, when we additionally control for the high-order factors, we find that for 95% of the zoo factors the risk price estimate is not statistically different from zero, with the median t-statistic dropping from 1.38 to only 0.71. Specifically, we find that only 7 out of 148 zoo factors remain significant at the 5% level. Of these seven factors, three belong to the "low leverage" category of Jensen et al. (2023) (*cash_at*, *rd5_at*, and *rd_sale*). The remaining significant factors are in the categories "size" (*rd_me*), "seasonality" (*seas_6_10an*), "accruals" (*cowc_grla*) and "quality" (*qmj*). Overall, these results reveal that the high-order factors selected by the forward selection FMB procedure effectively subsume the majority of the zoo factors.

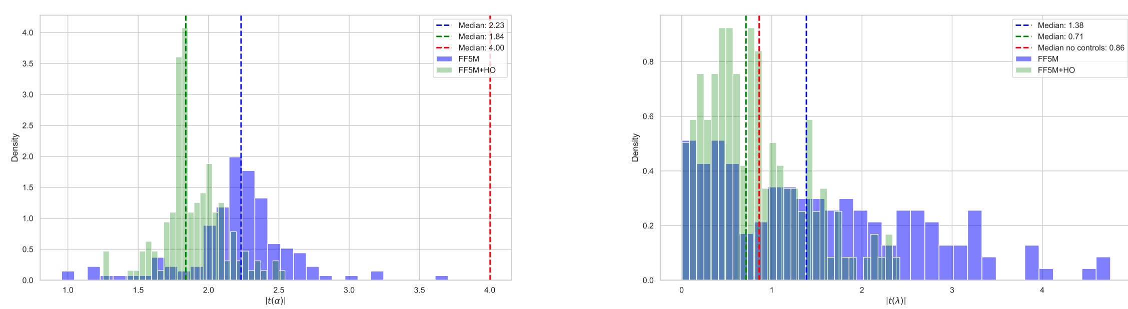
3.4.2. *Time-series method*

The second empirical strategy we employ to evaluate whether the high-order factors can subsume the factor zoo is based on time-series regressions, where we study whether the zoo factors can be spanned by the FF5M and the selected high-order factors.

FIGURE 2.—**Reducing the Factor Zoo in the Cross-Section**

This figure plots the distribution of absolute t-statistics associated with the risk prices ($t(\lambda)$) and the intercepts ($t(\alpha)$) of the factor zoo from Jensen et al. (2023) based on US equity data. For each zoo factor, we use the FMB procedure to estimate its risk price controlling for the FF5M (blue histogram) and the FF5M with the high-order factors selected by the FS-FMB procedure (green histogram). The dashed lines in blue and green correspond to the median absolute t-statistic for the FF5M and for the FF5M with high-order factor models, respectively. The dashed red line corresponds to the median absolute t-statistic with no controls. The t-statistics are based on Newey-West corrected standard errors.

(a) **Distribution of** $|t(\alpha)|$ (b) **Distribution of** $|t(\lambda)|$



Note that although the FF5M factors are tradable, the high-order factors are generally not directly tradable. Therefore, in order to estimate time-series spanning regressions, we must first construct factor mimicking portfolios of the higher-order factors based on tradable assets. In order to do so, we project each high-order factors onto the entire factor zoo and obtain its factor-mimicking portfolio as the regression's predicted values. This factor-mimicking portfolios only capture the information in the factor zoo that are related to the higher-order factors. The benefit of the factor-mimicking portfolio approach is that we are able to convert the non-tradable higher-order factors to tradable factors, but we note that the factor-mimicking portfolios of the higher-order factors inevitably lose information from the original factors. Figure E.2 shows the adjusted R^2 of projecting each selected higher-order factor onto the factor zoo. Mkt-RF*SMB has the lowest adjusted R^2 which is less than 40% and SMB2*Mom has the highest adjusted R^2 at more than 80%, suggesting the projections capture importance fractions, though to various degrees, of variations in the higher-order factors. Nevertheless, we use them to check if they capture pricing information of each factor.

To do so, we estimate a time-series regression for each factor in the factor zoo on the FF5M and the factor-mimicking portfolios of the high-order factors. The object of interest is the intercept. If the higher-order factors indeed span a given factor in the factor zoo, we expect to find a small and statistically insignificant intercept.

Figure 3 summarizes the results of the time-series spanning regressions. Panel A plots the distribution of the absolute t-statistics associated with the intercepts, and Panel B shows the distribution of the absolute magnitudes for the intercepts. The latter are annualized and expressed in percentage points.

Panel A shows that the median of the absolute t-statistic of the average returns for the zoo factors is statistically significant at the 5% level at 2.08. Controlling for the FF5M brings the median absolute t-statistic from 2.06 to 1.87, a decline of about 9%. Controlling for the high-order factors brings the median absolute t-statistic further down to 1.49, a decline of about 21% from 1.87. Panel B reveals a similar pattern. The median of the absolute intercept estimate of the zoo factors drops from 2.92 pp to 1.69 pp controlling for the FF5M model. It further decreases to 1.34 pp when the higher-order factors are included.

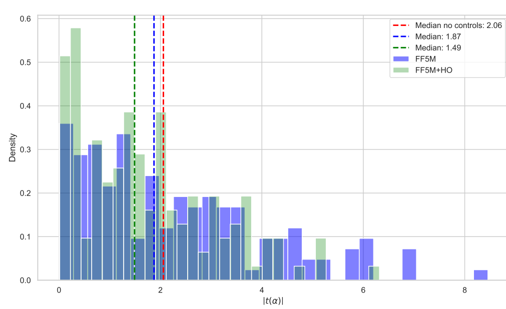
3.4.3. *Relationship between factor zoo and higher-order factors*

In the previous subsections, we show that the higher-order factor model selected by the FS-FMB procedure is successful in subsuming the majority of the zoo factors. Mechanically, the ability of the selected higher-order factor model to subsume the factor zoo has to come from correlation between the zoo factors and some or all of the factors in the higher-order factor model. In this subsection, we provide information on the loadings of the zoo factors on the factors of the selected higher-order factor model, which sheds light on the relative importance of the higher-order factors in subsuming the zoo factors.

FIGURE 3.—**Reducing the Factor Zoo in the Time-Series**

This figure presents the results of factor spanning time-series regressions for each zoo factor controlling for the FF5M factors (blue histogram) and the FF5M factors with the mimicking portfolios corresponding to the higher-order factors selected by the forward selection FMB procedure (green histogram). Subplot (a) plots the distribution of the absolute t-statistics for the intercept (α). Subplot (b) plots the distribution of the absolute αs. We denote with dashed vertical lines the median value when controlling for the FF5M factors (blue dashed line), for the FF5M with the mimicking portfolios corresponding to the higher-order factors selected by the FS-FMB procedure (green dashed line), and when we include no controls (red dashed line). The factors are 153 factors from the factor zoo constructed in Jensen et al. (2023). The t-statistic is based on Newey-West corrected standard errors.

(a) **Distribution of** $|t(\alpha)|$



(b) **Distribution of** $|\alpha|$

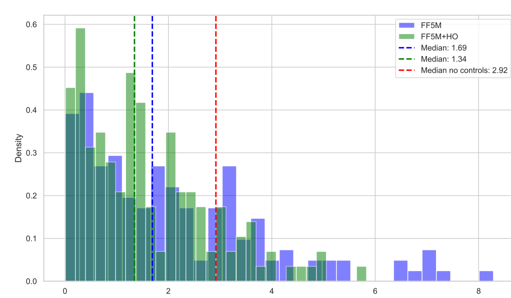


Table VI documents the fraction of zoo factors with loading significantly different from zero at the 5% confidence level on each of the FF5M factors and the selected higher-order factors selected from the forward selection FMB procedure. Column (1) shows the fraction of significance for the factors in FF5M model only and Column (2) reports the fraction of significance for the factors in the full higher-order factor model. Column (1) shows that the fractions of significance range from 45.9% for Mkt-RF to 75% for RMW for the 6 factors in FF6M. Column (2) shows that the fractions of significance remain stable for the 6 factors in FF6M, when we further include the selected higher-order factors. The fractions of significance for the higher-order factors range from 27% for Mkt-RF2 to 55.4% for HML2*Mkt-RF, highlighting that it is not the case a specific higher-order factor fully subsumes the zoo factors.

TABLE VI

LOADINGS OF THE ZOO FACTORS ON HIGHER-ORDER FACTOR MODEL

This table reports the fraction of zoo factors with loading significantly different from zero at the 5% confidence level on each of the FF5M factors and higher-order factors selected by the FS-FMB procedure. Significance is evaluated by t-statistics based on Newey-West corrected standard errors. The zoo factors are 148 factors constructed in Jensen et al. (2023).

	(1)	(2)
Factor	Frac Sig 5%	Frac Sig 5%
Mkt-RF	0.459	0.459
SMB	0.568	0.628
HML	0.642	0.709
RMW	0.75	0.669
CMA	0.601	0.595
Mom	0.486	0.405
SMB2		0.446
SMB2*Mom		0.331
Mom2*RMW		0.338
Mkt-RF2		0.27
Mkt-RF2*RMW		0.358
Mkt-RF*SMB		0.304
HML2*Mkt-RF		0.554
# zoo factors	148	148

3.5. *Higher-Order Factors and Macro Risk*

In this subsection, we conduct analyses to better understand the economic forces the higher-order factors capture. In the first exercise, we examine the correlations between the selected higher-order factors and two important macroeconomic risks as proxied by the U.S. equity market returns and the CBOE VIX index.

In Table VII, we present the results. Panel A documents the correlations with U.S. equity market returns and Panel B shows the correlations with CBOE VIX index. Beyond correlations over the full sample, we also calculate the correlations during the NBER recession periods and the correlations corresponding to the bottom 10%, the middle 10%-90%, and the top 10% of higher-order factors' return distribution.

For the first two higher-order factors (i.e., SMB2 and SMB2*Mom), the magnitude of the correlations with U.S. equity market returns are markedly higher during the NBER recession periods relative to the full sample. For example, the correlation between SMB2 and U.S. equity market returns is only 0.016 for the full sample, but it jumps to 0.205 during the NBER recession periods. For the last four higher-order factors (i.e., Mkt-RF2, Mkt-RF2*RMW, Mkt-RF*SMB, and HML2*Mkt-RF), they tend to have high correlation in magnitude with the VIX index, especially during the NBER recession periods. For example, the correlation between Mkt-RF*RMW and VIX is 0.32 for the full sample and increases to 0.515 during the NBER recession periods. These results show that the selected higher-order factors capture different macroeconomic risks, especially during economic downturns.

Additionally, in Table VIII, we examine the exposures of the selected higher-order factors to the common macroeconomic factors. The adjusted R^2s are relatively higher and above 20% for 5 of the 7 higher-order factors, but the adjusted R^2s for the SMB2*Mom and the Mom2*RMW factors are low at 9% and 1%, respectively. In particular, Mom2*RMW is not significantly exposed to any of the macroeconomic factors considered. Of the macroeconomic factors, factors related to uncertainties appear to be more important in accounting for the higher-order factors, especially financial uncertainty.

4. ADDITIONAL RESULTS

In this section, we provide additional results for our empirical analyses, and we discuss extensions of our baseline setup.

4.1. *Simulation Result*

The selected higher-order model incorporates 7 higher-order factors beyond the traditional FF5M model. Our results indicate that the model performance signifi-

TABLE VII

CORRELATIONS WITH MACROECONOMIC RISK

This table reports the correlation coefficients for each high-order factor selected by the forward selection FMB procedure with the U.S. equity market return (Panel A) and with the CBOE VIX index. We present both unconditional (Full) correlation coefficients and conditional correlation coefficients for specific subsamples: one containing U.S. NBER recession periods, and others corresponding to the bottom and top 10% of each factor's return distribution, as well as the remaining 10%–90% quantile. Correlation coefficients involving the VIX index are estimated starting in January 1990, reflecting the index's availability from that date onward.

Factor (h_j)/Sample	Full	NBER Recessions	$q_{\rightarrow 0.1}(h_j)$	$q_{0.1\rightarrow 0.9}(h_j)$	$q_{0.9\rightarrow}(h_j)$
Panel A: Correlation with US Equity Market					
SMB2	0.016	0.205	-0.062	-0.01	0.1
SMB2*Mom	-0.032	-0.345	-0.119	-0.018	0.085
Mom2*RMW	-0.113	-0.109	-0.085	-0.077	-0.001
Mkt-RF2	-0.164	-0.108	0.388	0.199	-0.267
Mkt-RF2*RMW	-0.372	-0.287	0.118	0.027	-0.479
Mkt-RF*SMB	-0.206	-0.136	-0.293	-0.009	-0.299
HML2*Mkt-RF	0.524	0.636	0.357	0.64	0.607
Panel B: Correlation with VIX Index					
SMB2	0.134	0.03	0.076	0.189	0.005
SMB2*Mom	0.001	-0.07	-0.233	0.043	-0.053
Mom2*RMW	0.015	0.116	-0.07	0.195	0.236
Mkt-RF2	0.603	0.697	0.141	0.284	0.515
Mkt-RF2*RMW	0.32	0.515	-0.255	0.138	0.376
Mkt-RF*SMB	0.257	0.486	-0.298	0.197	0.438
HML2*Mkt-RF	-0.239	-0.437	-0.266	-0.334	0.104

cantly improves with the 7 higher-order factors where the R-squared of the second-pass regression increases from 31.2% to 58.7%. However, as noted by Lewellen et al. (2010), spurious factors can artificially inflate model fit in the second-pass of the Fama-MacBeth procedure, particularly when the test assets exhibit a strong factor structure.

TABLE VIII: **Exposure to Macroeconomic Factors**

This table reports the results of regressing each selected higher-order factor on a list of common macroeconomic factors. Significance of the coefficients is based on Newey-West corrected standard errors. *** denotes significance at 1%. ** denotes significance at 5%. * denotes significance at 10%. To each factor, the table associate a reference, a category, and the number of high-order factors with a significant exposure at the 10% confidence level.

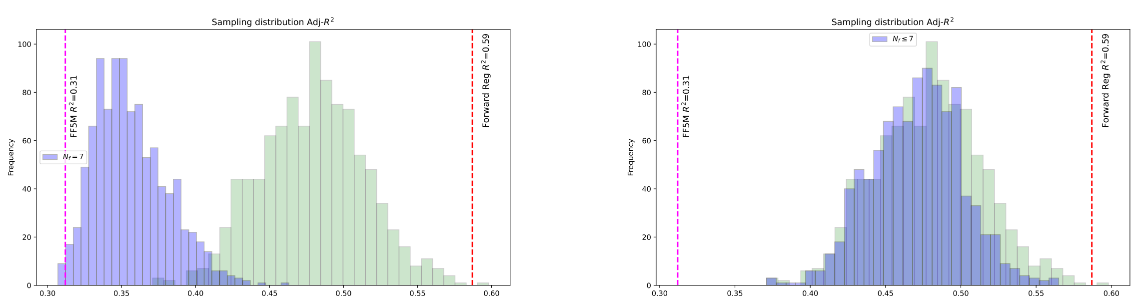
Factor	Reference	Category	(1) SMB2	(2) SMB2*Mom	(3) Mom2*RMW	(4) Mkt-RF2	(5) Mkt-RF2*RMW	(6) Mkt-RF*SMB	(7) HML2*Mkt-RF	# sig
Common idio-vol shock	Herskovic et al. (2016)	Volatility	0.31	0.77***	0.31	0.44	-0.16	0.2	0.42**	2
Consum growth	Chapman (1997)	Consum	1.06	0.25	-0.2	-0.25	0.05	-0.04	0.01	0
Consum growth squared	Chapman (1997)	Consum	3.56*	-3.08	1.46	-1.44	-1.46***	-2.99**	1.04	3
Intermediary capital return factor	He et al. (2017)	Inter	1.15	-5.29**	1.99	2.02*	0.27	1.26	1.92**	3
Leverage factor	Adrian et al. (2014)	Inter	-0.03	-0.23	0.11	0.73**	-0.11	0.3	0.23	1
Intermediary capital risk factor	He et al. (2017)	Inter	-0.75	0.56	1.97	0.09	0.01	-0.32	-1.50**	1
Equity market return	Fama and French (1993)	Fin Markets	-0.35	4.39	-5.34	-0.77	-1.33**	0.08	1.78***	2
Liquidty	Pástor and Stambaugh (2003)	Liquidity	-0.02	-0.37	0.12	-0.43	0.04	-0.78*	0.09	1
Polynomial of labor income growth	Dittmar (2002)	Macro	-69.94	353.57	-167.24	145.62	-88.83	-73.58	-121	0
Macro uncertainty	Bali et al. (2017)	Macro	-0.92*	0.02	0.55	-0.33	-0.29*	-0.25	-0.24	2
Financial uncertainty	Bali et al. (2017)	Macro	1.36**	0.38	-0.92	0.78***	0.19**	0.30**	0.13	4
Real uncertainty	Bali et al. (2017)	Macro	0.23	-1.69	0.59	1.64***	0.19	1	0.07	1
Labor income growth	Campbell (1996)	Macro	144.2	-703.78	329.56	-290.04	177.76	146.78	244.27	0
Change long-term govt yield	Sweeney and Warga (1986)	Macro	-5.21	-24.69**	13.85	9.34*	-3.21	10.21**	4.18	3
News implied volatility	Manela and Moreira (2017)	Macro	-0.02**	-0.01	0.01	0.01**	0	0	-0.00**	3
Investor sentiment (orth macro)	Baker and Wurgler (2006)	Sentiment	0.02	-0.03	0	0.12***	0.02*	0.03	-0.03	2
Investor sentiment (orth macro)	Huang et al. (2015)	Sentiment	-0.01	0.05	-0.05	-0.09***	0	-0.05**	-0.01	2
		Adj. R^2	0.21	0.09	0.01	0.42	0.21	0.27	0.46	

Lewellen et al. (2010) illustrate this concern using the Fama-French 25 Size-BM portfolios which have a particularly strong factor structure. While our analysis mitigates this issue by employing a broad set of 484 test assets and demonstrating the superior out-of-sample performance, we further conduct simulation exercises to confirm that our results are not driven by spurious factors.

FIGURE 4.—**Sampling Distribution of R-squareds**

This figure plots the sampling distribution (green histogram) of adjusted cross-sectional R-squareds obtained applying the FS-FMB procedure to the FF5M factors and randomly generated higher-order factors. First, we randomly generate 57 factors $i.i.d. \sim N(0, \sigma(\text{Mkt-RF}))$. Next, at each step of the FS-FMB procedure, we select one factor to maximize the adjusted cross-sectional R-squared until $\Delta\left(\text{adj } R^2\right) \leq 1\%$. The procedure is repeated 1,000 times to obtain the sampling distribution of R-squareds. In subplot (a), we include the distribution obtained when the number of additionally selected factors is constrained to be ≤ 7 (blue histogram). In subplot (b), we include the distribution obtained when the FF5M model is augmented with 7 randomly generated factors, repeating the procedure 1,000 times. The pink-dashed line corresponds to the adjusted R-squared of the FF5M model. The red-dashed line corresponds to the R-squared of the FF5M model with higher-order factors selected by the forward selection FMB procedure (see Table II).

(a) (b)



Therefore, we estimate the sampling distribution of cross-sectional R-squareds obtained using randomly generated factors, following the approach advocated by Lewellen et al. (2010). Specifically, in each simulation, we generate a set of 57 random factors, matching the number of candidate higher-order factors. These random factors are assumed to be $i.i.d. \sim N(0,\sigma^2)$, and we set σ^2 equal to the sample variance of the market factor (Mkt-RF). We then apply a forward selection FMB procedure, sequentially selecting factors that maximize the adjusted cross-sectional R-squared when added to the FF5M factors. We repeat the procedure 1,000 times by generating 1,000 independent sets of factors.

The green histogram in figure 4 summarizes the sampling distribution of the resulting R-squareds from the 1,000 simulations. In the simulation exercise, we obtain a cross-sectional R-squared larger than 0.59 only in 0.1% of the simulations, meaning that the adjusted R-squared from the higher-order model is significant at the 0.1% level. This simulation exercise highlights the fact that it is extremely unlikely that our results are obtained by chance.

Note that the procedure can select a number of factors different than 7, which is the number of higher-order factors presented in Table II. In fact, in the simulation exercise we conduct, the maximum number of random factors selected is 12. In Panel A, we compare the sampling distribution of R-squareds with that obtained with the additional constraint that the number of selected factors cannot be larger than 7 (blue histogram). In this case, we never observe R-squareds larger than 0.59 (maximum adjusted cross-sectional R-squared of 0.565).

Finally, in Panel B, we further compare the sampling distribution of R-squareds with that obtained by directly adding 7 randomly generated factors, $i.i.d. \sim N(0,\sigma^2)$, to the FF5M factors, repeating the procedure 1,000 times (blue histogram). In this case, the maximum adjusted cross-sectional R-squared is only 0.463.

4.2. *An alternative factor zoo*

In order to investigate the robustness of our results in terms of subsuming the factor zoo, we replicate the analysis based on the two-pass Fama-MacBeth procedure using an alternative set of zoo factors, based on the factors constructed by Chen and Zimmermann (2022). Specifically, for each factor in Chen and Zimmermann (2022), we estimate its SDF loading while controlling for our higher-order factor model, which includes the FF5M and the selected higher-order factors. Figure 5 summarizes the results. Panel A shows that the intercept estimates tend to be highly statistically significant, with a median value of 3.97, when the factor model includes only the factor in the factor zoo. When the factor model includes both FF5M and the factor in the factor zoo, the median t-statistic decreases substantially but remains significant at the 5% level at 2.23. When the factor model further includes the higher-order factors, the median t-statistic is no longer statistically significant at the 5% level at 1.87. Panel B reveals that, when the factor model includes only the factor in the factor zoo, the median SDF loading estimate for the factor zoo is not statistically different from zero, with a t-statistic of 0.80. Note that these estimates are likely biased because of omitted factors (see, e.g., Feng et al. (2020)). Controlling for the FF5M, the median t-statistic increases substantially to 1.91. 46% of zoo factors become statistically significant at the 5% level with a t-statistic greater than 2 suggesting that FF5M helps to remove some bias in the factor estimates. Importantly, when we additionally control for the high-order factors, we find that for 93% of the zoo factors the risk price estimate is not statistically different from zero, with the median t-statistic dropping from 1.79 to only 0.72. Specifically, we find that only 12 out 159 zoo factors remain significant at the 5% level. These results confirm our findings based on the factors in Jensen et al. (2023).[4]

Table E.III in the Online Appendix presents the fraction of zoo factors in Chen and Zimmermann (2022) with loading significantly different from zero at the 5% confidence level on each of the FF5M factors and the selected higher-order factors selected from the forward selection FMB procedure. Column (1) shows the fraction of significance for the factors in FF5M model only and Column (2) reports the fraction of significance for the factors in the full higher-order factor model. Column (1) shows that the fractions of significance range from 42.8% for CMA to 57.2% for SMB and RMW for the 6 factors in FF6M. Column (2) shows that the fractions of significance remain stable for the 6 factors in FF6M, when we further include the selected higher-order factors. The fractions of significance for the higher-order factors range from 20.8% for Mkt-RF2*RMW to 47.8% for SMB2*Mom. These results are consistent with the findings we

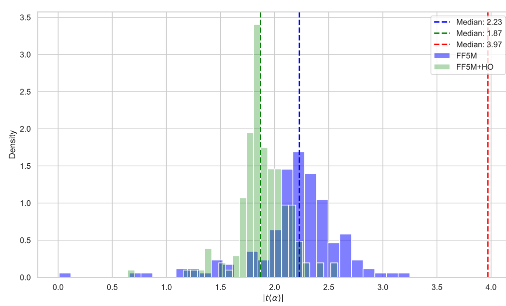
[4]The factors in Chen and Zimmermann (2022) that remain significant at the 5% confidence level are: MomSeason06YrPlus, CBOperProf, ChInv, STreversal, DelFINL, DivYieldST, InvGrowth, GrSaleToGrInv, MomOffSeason16YrPlus, DebtIssuance, CompositeDebtIssuance, OrderBacklog.

obtain using the zoo factors in Jensen et al. (2023), further highlighting that it is not the case a specific higher-order factor fully subsumes the zoo factors.

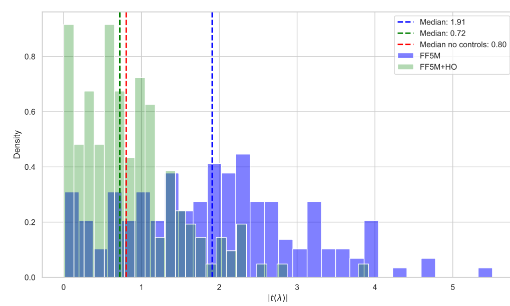
FIGURE 5.—**Reducing Alternative Zoo Factors in the Cross-Section**

This figure plots the distribution of absolute t-statistics associated with the risk prices ($t(\lambda)$) and the intercepts ($t(\alpha)$) of the alternative zoo factors from Chen and Zimmermann (2022) based on US equity data. For each zoo factor, we use the FMB procedure to estimate its risk price controlling for the FF5M (blue histogram) and the FF5M with the high-order factors selected by the FS-FMB procedure (green histogram). The dashed lines in blue and green correspond to the median absolute t-statistic for the FF5M and for the FF5M with high-order factor models, respectively. The dashed red line corresponds to the median absolute t-statistic with no controls. The t-statistics are based on Newey-West corrected standard errors.

(a) **Distribution of** $|t(\alpha)|$



(b) **Distribution of** $|t(\lambda)|$



4.3. *Alternative sets of candidate factors*

We start considering alternative sets of higher-order factors formed using the FF5M factors. Table E.IV presents the estimation results of the forward selection FMB procedure with four alternative sets of higher-order factors. Specifically, Panel A presents estimation results using a set of high-order factors of degree 2. The latter is a lower degree than the one from the baseline analysis (i.e., degree 3). These high-order factors include all pairwise interactions of the FF5M, and the FF5M factors squared. In this case, the FS-FMB procedure selects four high-order factors, reaching an adjusted cross-sectional R-squared of 0.517 in the final step, approximately 7 pp smaller than in the baseline setup (see Table II). Panel B presents estimation results using a set of high-order factors of higher degree, and specifically of degree up to 4. These factors, augment the high-order factors of degree 3 of the baseline setup with the FF5M to the fourth power (i.e., f_i^4), and pairwise interactions of the FF5M factors of degree 4, i.e., $f_i^2 \times f_j^2, f_i^3 \times f_j, f_i \times f_j^3$. In this case, the FS-FMB procedure selects 9 high-order factors, reaching an adjusted cross-sectional R-squared of 0.6 in the final step. The latter is a R-squared value very similar to that obtained in the baseline setup, with high-order factors up to degree 3. These results suggest that higher-order factors up to degree 3 are most relevant in explaining cross-section asset returns, corroborating our choice of using high-order factors of degree 3 in the baseline analysis.

Furthermore, Panel C and D report the results when augmenting the FF5M with only pairwise interactions of degree 2 and 3 (Panel C) and only the powers of degree 2 and 3 (Panel D). The results from these two panels reveal that incorporating pairwise interactions between FF5M is crucial. In fact, the model with interactions of degree 2 and 3 reaches an adjusted cross-sectional R-squared of 0.542, selecting 7 high-order factors. In contrast, the model with powers of degree 2 and 3 selects 3 high-order factors, for an adjusted cross-sectional R-squared of 0.467 in the final step. These results highlight that both higher-orders of FF5M and higher-order interactions of FF5M are important components of the SDF, especially higher-order interactions.

REFERENCES

ADRIAN, TOBIAS, ERKKO ETULA, AND TYLER MUIR (2014): "Financial intermediaries and the cross-section of asset returns," *Journal of Finance*, 69 (6), 2557–2596. [36]

BAKER, MALCOLM AND JEFFREY WURGLER (2006): "Investor sentiment and the cross-section of stock returns," *The journal of Finance*, 61 (4), 1645–1680. [36]

BALI, TURAN G, STEPHEN J BROWN, AND YI TANG (2017): "Is economic uncertainty priced in the cross-section of stock returns?" *Journal of Financial Economics*, 126 (3), 471–489. [36]

BANSAL, RAVI AND SALIM VISWANATHAN (1993): "No arbitrage and arbitrage pricing: A new approach," *Journal of Finance*, 48 (4), 1231–1262. [6]

BELLONI, ALEXANDER AND VICTOR CHERNOZHUKOV (2009): "High-dimensional sparse econometric models, an Introduction," *Springer Lecture Notes*. [15]

BELLONI, ALEXANDRE, VICTOR CHERNOZHUKOV, AND CHRISTIAN HANSEN (2014): "Inference on treatment effects after selection among high-dimensional controls," *Review of Economic Studies*, 81, 608–650. [13]

BORRI, NICOLA, DENIS CHETVERIKOV, YUKUN LIU, AND ALEH TSYVINSKI (2024): "One Factor to Bind the Cross-Sectional of Returns," *NBER Working Paper No. 32365*. [6]

CAMPBELL, JOHN Y (1996): "Understanding risk and return," *Journal of Political Economy*, 104 (2), 298–345. [36]

CARHART, MARK M (1997): "On persistence in mutual fund performance," *Journal of Finance*, 52 (1), 57–82. [2, 6]

CHAPMAN, DAVID A (1997): "The cyclical properties of consumption growth and the real term structure," *Journal of Monetary Economics*, 39 (2), 145–172. [36]

CHEN, ANDREW Y. AND TOM ZIMMERMANN (2022): "Open Source Cross-Sectional Asset Pricing," *Critical Finance Review*, 27 (2), 207–264. [5, 38, 39, 40, 65]

CHERNOZHUKOV, VICTOR, DENIS CHETVERIKOV, MERT DEMIRER, ESTHER DUFLO, CHRISTIAN HANSEN, WHITNEY NEWEY, AND JAMES ROBINS (2018): "Double/debiased machine learning for treatment and structural parameters," *Econometrics Journal*, 21, C1–C68. [3]

COCHRANE, JOHN (2005): *Asset Pricing*, Princeton University Press. [8]

COCHRANE, JOHN H (2011): "Presidential address: Discount rates," *The Journal of Finance*, 66 (4), 1047–1108. [2, 6]

CONNOR, GREGORY AND ROBERT A KORAJCZYK (1986): "Performance measurement with the arbitrage pricing theory: A new framework for analysis," *Journal of Financial Economics*, 15 (3), 373–394. [6]

DAS, ABHIMANYU AND DAVID KEMPE (2018): "Approximate submodularity and its applications: subset selection, sparse approximation, and dictionary selection," *Journal of Machine Learning Research*, 19, 1–34. [12]

DE GEER, VAN, PETER SARA, BUHLMANN, YAKOV RITOV, AND RUBEN DEZEURE (2014): "On asymptotically optimal confidence regions and tests for high-dimensional models," *Annals of Statistics*, 42, 1166–1202. [13]

DITTMAR, ROBERT F (2002): "Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns," *The Journal of Finance*, 57 (1), 369–403. [6, 36]

ELENBERG, ETHAN, RAJIV KHANNA, ALEXANDROS DIMAKIS, AND SAHAND NEGAHBAN (2018): "Restricted strong convexity implies weak submodularity," *Annals of Statistics*, 46, 3539–3568. [12, 50, 51, 54]

FAMA, EUGENE F. AND KENNETH R. FRENCH (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33 (1), 3–56. [6, 21, 36]

FAMA, EUGENE F AND KENNETH R FRENCH (1996): "Multifactor explanations of asset pricing anomalies," *Journal of Finance*, 51 (1), 55–84. [6]

——— (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116 (1), 1–22. [2, 3, 6, 19, 21]

FENG, GUANHAO, STEFANO GIGLIO, AND DACHENG XIU (2020): "Taming the factor zoo: A test of new factors," *Journal of Finance*, 75 (3), 1327–1370. [3, 6, 13, 18, 19, 28, 29, 39]

FREYBERGER, JOACHIM, ANDREAS NEUHIERL, AND MICHAEL WEBER (2020): "Dissecting characteristics nonparametrically," *The Review of Financial Studies*, 33 (5), 2326–2377. [6]

GIGLIO, STEFANO AND DACHENG XIU (2021): "Asset pricing with omitted factors," *Journal of Political Economy*, 129 (7), 1947–1990. [3, 6]

GREEN, JEREMIAH, JOHN RM HAND, AND X FRANK ZHANG (2017): "The characteristics that provide independent information about average US monthly stock returns," *The Review of Financial Studies*, 30 (12), 4389–4436. [6]

GU, SHIHAO, BRYAN KELLY, AND DACHENG XIU (2020): "Empirical asset pricing via machine learning," *Review of Financial Studies*, 33 (5), 2223–2273. [6]

HARVEY, CAMPBELL R AND YAN LIU (2021): "Lucky factors," *Journal of Financial Economics*, 141 (2), 413–435. [6, 12]

HARVEY, CAMPBELL R, YAN LIU, AND HEQING ZHU (2016): "... and the cross-section of expected returns," *The Review of Financial Studies*, 29 (1), 5–68. [6]

HARVEY, CAMPBELL R AND AKHTAR SIDDIQUE (2000): "Conditional skewness in asset pricing tests," *Journal of Finance*, 55 (3), 1263–1295. [6]

HASTIE, TREVOR, ROBERT TIBSHIRANI, AND JEROME FRIEDMAN (2009): *The Elements of Statistical Learning*, Springer. [3, 12]

HE, ZHIGUO, BRYAN KELLY, AND ASAF MANELA (2017): "Intermediary asset pricing: New evidence from many asset classes," *Journal of Financial Economics*, 126 (1), 1–35. [36]

HERSKOVIC, BERNARD, BRYAN KELLY, HANNO LUSTIG, AND STIJN VAN NIEUWERBURGH (2016): "The common factor in idiosyncratic volatility: Quantitative asset pricing implications," *Journal of Financial Economics*, 119 (2), 249–283. [36]

HUANG, DASHAN, FUWEI JIANG, JUN TU, AND GUOFU ZHOU (2015): "Investor sentiment aligned: A powerful predictor of stock returns," *The Review of Financial Studies*, 28 (3), 791–837. [36]

JAVANMARD, ADEL AND ANDREA MONTANARI (2014): "Confidence intervals and hypothesis testing for high-dimensional regression," *Journal of Machine Learning Research*, 15, 2869–2909. [13]

JEGADEESH, NARASIMHAN AND SHERIDAN TITMAN (1993): "Returns to buying winners and selling losers: Implications for stock market efficiency," *Journal of Finance*, 48 (1), 65–91. [3, 19, 21]

JENSEN, THEIS INGERSLEV, BRYAN KELLY, AND LASSE HEJE PEDERSEN (2023): "Is there a replication crisis in finance?" *Journal of Finance*, 78 (5), 2465–2518. [4, 19, 20, 29, 30, 32, 33, 39, 40, 63]

KOIJEN, RALPH SJ, HANNO LUSTIG, AND STIJN VAN NIEUWERBURGH (2017): "The cross-section and time series of stock and bond returns," *Journal of Monetary Economics*, 88, 50–69. [6]

KOZAK, SERHIY, STEFAN NAGEL, AND SHRIHARI SANTOSH (2020): "Shrinking the cross-section," *Journal of Financial Economics*, 135 (2), 271–292. [3, 6, 19, 20, 21]

KOZBUR, DAMIAN (2020): "Analysis of testing-based forward model selection," *Econometrica*, 88, 2147–2173. [12]

LETTAU, MARTIN AND MARKUS PELGER (2020): "Factors that fit the time series and cross-section of stock returns," *Review of Financial Studies*, 33 (5), 2274–2325. [6]

LEWELLEN, JONATHAN, STEFAN NAGEL, AND JAY SHANKEN (2010): "A skeptical appraisal of asset pricing tests," *Journal of Financial Economics*, 96 (2), 175–194. [15, 35, 37]

MANELA, ASAF AND ALAN MOREIRA (2017): "News implied volatility and disaster concerns," *Journal of Financial Economics*, 123 (1), 137–162. [36]

McLean, R David and Jeffrey Pontiff (2016): "Does academic research destroy stock return predictability?" *The Journal of Finance*, 71 (1), 5–32. [6]

Nagel, Stefan (2021): *Machine learning in asset pricing*, Princeton University Press. [3, 15]

Newey, Whitney (1994): "The asymptotic variance of semiparametric estimators," *Econometrica*, 62, 1349–1382. [18]

Newey, Whitney K and Kenneth D West (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55 (3), 703–708. [28]

Pástor, L'uboš and Robert F Stambaugh (2003): "Liquidity risk and expected stock returns," *Journal of Political Economy*, 111 (3), 642–685. [36]

Sweeney, Richard J and Arthur D Warga (1986): "The pricing of interest-rate risk: evidence from the stock market," *The Journal of Finance*, 41 (2), 393–410. [36]

Zhang, Cun-Hui and Stephanie Zhang (2014): "Confidence intervals for low dimensional parameters in high dimensional linear models," *Journal of the Royal Statistical Society, Series B*, 76, 217–242. [13]

Zhang, Tony (2009): "On the consistency of feature selection using greedy least squares regression," *Journal of Machine Learning Research*, 10, 555–568. [12]

ONLINE APPENDICES

This paper has 5 online appendices. In Appendix A, we describe additional assumptions for Theorem 2. In Appendix B, we give the proof of Theorem 1. In Appendix C, we give the proof of Theorem 2. In Appendix D, we provide several technical lemmas that are useful for proving the theorems. In Appendix E, we present additional tables and figures.

APPENDIX A: ASSUMPTIONS FOR THEOREM 2

In this Appendix, we describe additional assumptions required for the asymptotic normality result for the debiased estimator $\widehat{\psi}_D$ in Theorem 2.

ASSUMPTION 8—First Stage Sparsity: *For all* $j \in [p]$*, there exists a set* $\mathrm{S}_{0,j} \subset \{j\}^c$ *and a vector* $\varphi_j = (\varphi_{j,1}, \ldots, \varphi_{j,p})^\top \in \mathbb{R}^p$ *such that (i)* $\mathrm{supp}(\varphi_j) = \mathrm{S}_{0,j}$*, (ii)* $|\mathrm{S}_{0,j}| \leq \bar{s}_T$*, and (iii) the vector* $R_j = (R_{j,1}, \ldots, \mathbb{R}_{j,N})^\top = \mathrm{C}_{\{j\}} - \mathrm{C}\varphi_j$ *satisfies* $\|\mathrm{C}_{\{j\}^c}^\top R_j\|_\infty^2 \lesssim N^2 \log(Np)/T$.

This assumption means that for each $j \in [p]$, there exists a sparse approximate projection of the vector $\mathrm{C}_{\{j\}}$ on $\mathrm{C}_{\{j\}^c}$: most cross-sectional variation in the components of the vector $\mathrm{C}_{\{j\}}$ can be explained by at most $\bar{s}_T$ columns in the matrix $\mathrm{C}_{\{j\}^c}$ in the sense that the residual of the linear projection of $\mathrm{C}_{\{j\}}$ on these columns is nearly orthogonal to all columns of the matrix.

ASSUMPTION 9—Uniform Law of Large Numbers, II: *We have*

$$\max_{i \in [2N+2]} \max_{j \in [p]} \left| \frac{1}{T} \sum_{t \in [T]} u_{i,t} \varphi_j^\top v_t - \mathrm{E}[u_{i,t} \varphi_j^\top v_t] \right| \lesssim_P \sqrt{\frac{\log(Np)}{T}}.$$

This assumption complements Assumption 1 by imposing the condition that a quantitative version of the law of large numbers applies to the random variables $u_{i,t} \varphi_j^\top v_t$ uniformly over $i \in [2N+2]$ and $j \in [p]$.

ASSUMPTION 10—First Stage Estimator: *For all* $j \in [p]$ *and some constant* $c > 0$*, we have (i)* $|\widehat{\mathrm{S}}_j| \leq \bar{s}_T$ *w.p.a.1, (ii)* $|\widehat{\mathrm{S}}_j| \geq (1+c)(K/k)|\mathrm{S}_{0,j}| \log T$ *w.p.a.1, and (iii)* $|\widehat{\mathrm{S}}_j| \lesssim_P |\mathrm{S}_{0,j}| \log T$.

This assumption mirrors Assumption 5 for the estimation of a different object. In particular, it will be clear from the proofs in Appendix C that for all $j \in [p]$, $\widehat{\mathrm{S}}_j$ serves as an estimator of $\mathrm{S}_{0,j}$. The first part of Assumption 10 means that we carry out at most $\bar{s}_T$ rounds in Algorithm 1 when we choose the set $\widehat{\mathrm{S}}_j$, which is thus rather natural given the existence of a sparse approximate projection with at most $\bar{s}_T$ components in Assumption 8. The second part of Assumption 8 requires that the number of selected factors $|\widehat{\mathrm{S}}_j|$ should be larger than the number of important factors $|\mathrm{S}_{0,j}|$ at least by a multiplicative constant proportional to $\log T$. The third part gives an upper bound on the number of selected factors.

ASSUMPTION 11—No Strong Omitted Factors: *For all $j \in [p]$, we have*

$$\frac{1}{N\sqrt{T}} \sum_{i\in[N]} \sum_{t\in[T]} R_{j,i}\varepsilon_{i,t} m_t = o_P(1).$$

Observe that for each $i \in [N]$ and $j \in [p]$, we expect $T^{-1/2}\sum_{t\in[T]} R_{j,i}\varepsilon_{i,t} m_t = O_P(1)$ given that $\mathrm{E}[R_{j,i}\varepsilon_{i,t} m_t] = 0$. Thus, Assumption 11 means that by taking the average over $i \in [N]$, we can further wash out the variability from $T^{-1/2}\sum_{t\in[T]} R_{j,i}\varepsilon_{i,t} m_t$, which is only possible if the random variables $\varepsilon_{i,t}$ do not have a strong cross-sectional dependence and is thus equivalent to the condition that there are no strong omitted factors in the model.

ASSUMPTION 12—Factors: *For some constant $C > 0$, we have*

$$1/C \le \lambda_{\min}(\mathrm{E}[v_t v_t^\top]) \le \lambda_{\max}(\mathrm{E}[v_t v_t^\top]) \le C \quad \textit{and} \quad \left\|(\mathrm{E}[v_t v_t^\top])^{-1}\right\|_{\infty,1} \le C.$$

The lower bound in the first part of this assumption means that there is no multicollinearity between factors in the population, and the upper bound is a technical regularity condition. To understand the second part of this assumption, for each $j \in [p]$, let a_j be the jth column of the matrix $\mathrm{E}[(v_t v_t^\top)^{-1}]$. It is then possible to show that $\|\eta_j\|_1 = \sigma^2_{z,j}\|a_j\|_1 - 1$ for η_j appearing in (8); see Lemma 14 in Appendix D. Thus, given that $\sigma^2_{z,j} \le C$, which follows from the first part of the assumption, the condition $\|(\mathrm{E}[v_t v_t^\top])^{-1}\|_{\infty,1} \le C$ means that the vectors η_j are uniformly bounded in the ℓ_1-norm.

Next, for all $j \in [p]$, let $\phi_j = (\phi_{j,1}, \ldots, \phi_{j,p})^\top \in \mathbb{R}^p$ be the vector defined by $\phi_{j,j} = 0$ and

$$\phi_{j,\{j\}^c} = \Big(\mathrm{E}[v_{t,\{j\}^c} v_{t,\{j\}^c}^\top]\Big)^{-1} \mathrm{E}\left[v_{t,\{j\}^c} \left(\frac{1}{N} \sum_{i \in [N]} r_{i,t} R_{j,i} \right)\right], \tag{9}$$

so that $\phi_{j,\{j\}^c}$ is the vector of coefficients of the time-series least-squares projection of the cross-sectional sample average $N^{-1} \sum_{i \in [N]} r_{i,t} R_{j,i}$ on $v_{t,\{j\}^c}$.

ASSUMPTION 13—Extra Sparsity: *For all* $j \in [p]$, *there exists a set* $\mathrm{S}_{0,j,\phi} \subset \{j\}^c$ *such that* $|\mathrm{S}_{0,j,\phi}| \le \bar{s}_T$ *and* $\|\phi_{\mathrm{S}^c_{0,j,\phi}}\|_1 = o(1/\sqrt{\log(Np)})$.

This assumption means that the vectors ϕ_j are approximately sparse. Note, however, that the amount of sparsity imposed in this assumption is rather small: the ℓ_1-norm of the "remainder" $\phi_{\mathrm{S}^c_{0,j,\phi}}$ is required to vanish asymptotically but the convergence rate may be very slow.

ASSUMPTION 14—Growth Condition, II: *We have* $\bar{s}_T^2 \log^4(Np) = o(T)$.

ASSUMPTION 15—Central Limit Theorem: *For all* $j \in [p]$,

$$\frac{1}{\sqrt{T}} \sum_{t \in [T]} \frac{z_{t,j} m_t - \mathrm{E}[z_{t,j} m_t]}{\sigma^2_{z,j}} \to_d N(0, \sigma^2_{\psi,j}),$$

where $\sigma^2_{\psi,j} = \lim_{T\to\infty} T^{-1} \sum_{t=1}^T \sum_{s=1}^T \mathrm{E}\Big[(z_{t,j} m_t - \mathrm{E}[z_{t,j} m_t])(z_{s,j} m_s - \mathrm{E}[z_{s,j} m_s)]/\sigma^4_{z,j}\Big]$.

Assumption 14 strengthens Assumption 7 but, like Assumption 7, it allows the number of factors p to be much larger than the number of time periods T. Assumption 15 is a version of the central limit theorem for time series data.

APPENDIX B: PROOF OF THEOREM 1

Throughout this Appendix, denote $\mathrm{Y} = (\mathrm{E}[r_{1,t}], \ldots, \mathrm{E}[r_{N,t}])^\top$ and $\widehat{\mathrm{Y}} = (\bar{r}_1, \ldots, \bar{r}_N)^\top$. Also, let $\ell : \mathbb{R}^p \to \mathbb{R}$ be the function defined by $\ell(x) = -\|\widehat{\mathrm{Y}} - \widehat{\mathrm{C}}x\|_2^2/N$ for all $x \in \mathbb{R}^p$. In addition, for any vector $x = (x_1, \ldots, x_p)^\top \in \mathbb{R}^p$ and any set $\mathrm{S} \subset [p]$, let $x[\mathrm{S}] = (\overline{x}_1, \ldots, \overline{x}_p)^\top \in \mathbb{R}^p$ be the vector defined by $\overline{x}_j = x_j$ for all $j \in \mathrm{S}$ and $\overline{x}_j = 0$ for all $j \in \mathrm{S}^c$. Observe that the notation $x[\mathrm{S}]$ is different from the notation x_{S}, which we use throughout the paper.

Further, for all $j \in [p]$, let $\widetilde{\mathrm{S}}_{D,j} = \widehat{\mathrm{S}}_{D,j} \setminus \{j\} \subset [p]$ and let $\widehat{\varphi}_j = (\widehat{\varphi}_{j,1}, \dots, \widehat{\varphi}_{j,p})^\top \in \mathbb{R}^p$ be the vector defined by $\widehat{\varphi}_{\widetilde{\mathrm{S}}_{D,j}} = (\widehat{\mathrm{C}}^\top_{\widetilde{\mathrm{S}}_{D,j}} \widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}_{D,j}})^\top \widehat{\mathrm{C}}^\top_{\widetilde{\mathrm{S}}_{D,j}} \widehat{\mathrm{C}}_{\{j\}}$ and $\widehat{\varphi}_l = 0$ for all $l \in \widetilde{\mathrm{S}}^c_{D,j}$.

Before proving Theorem 1, we state and prove four auxiliary lemmas.

LEMMA 1: *Suppose that Assumptions 1 and 7 are satisfied. Then*

$$\max_{i \in [N]} \max_{j \in [p]} \left| \widehat{\mathrm{C}}_{i,j} - \mathrm{C}_{i,j} \right| \lesssim_P \sqrt{\frac{\log(Np)}{T}}.$$

PROOF: For all $(i,j) \in [N] \times [p]$, we have

$$\begin{aligned} \widehat{\mathrm{C}}_{i,j} &= \frac{1}{T} \sum_{t \in [T]} \beta_i^\top v_t (f_{t,j} - \bar{f}_j) + \frac{1}{T} \sum_{t \in [T]} \varepsilon_{i,t} (f_{t,j} - \bar{f}_j) \\ &= \frac{1}{T} \sum_{t \in [T]} \beta_i^\top v_t v_{t,j} + \frac{1}{T} \sum_{t \in [T]} \beta_i^\top v_t (\mathrm{E}[f_{t,j}] - \bar{f}_j) \\ &\quad + \frac{1}{T} \sum_{t \in [T]} \varepsilon_{i,t} v_{t,j} + \frac{1}{T} \sum_{t \in [T]} \varepsilon_{i,t} (\mathrm{E}[f_{t,j}] - \bar{f}_j). \end{aligned} \tag{10}$$

Subtracting $\mathrm{C}_{i,j} = \beta_i^\top \mathrm{E}[v_t v_{t,j}]$ from both sides and applying Assumptions 1 and 7 yields the asserted claim. *Q.E.D.*

LEMMA 2: *Suppose that Assumptions 1 and 7 are satisfied. Then*

$$\|(\widehat{\mathrm{C}} - \mathrm{C})\psi\|_2 \lesssim_P \sqrt{\frac{N \log(Np)}{T}}.$$

PROOF: The claim follows from combining (10) and Assumptions 1 and 7. *Q.E.D.*

LEMMA 3: *Suppose that Assumptions 1, 4, and 7 are satisfied. Then*

$$\frac{k}{2} \le \lambda_{\min}\left(\frac{\widehat{\mathrm{C}}_\mathrm{S}^\top \widehat{\mathrm{C}}_\mathrm{S}}{N} \right) \le \lambda_{\max}\left(\frac{\widehat{\mathrm{C}}_\mathrm{S}^\top \widehat{\mathrm{C}}_\mathrm{S}}{N} \right) \le 2K$$

for all $\mathrm{S} \subset [p]$ *such that* $|\mathrm{S}| \le 3\bar{s}_T + 1$ *w.p.a.1.*

PROOF: For any matrix A, let $s_{\min}(A) = \sqrt{\lambda_{\min}(A^\top A)}$ and $s_{\max}(A) = \sqrt{\lambda_{\max}(A^\top A)}$ be its smallest and largest singular values. Then uniformly over $\mathrm{S} \subset [p]$ such that $|\mathrm{S}| \le 3\bar{s}_T$, we have

$$|s_{\min}(\widehat{\mathrm{C}}_{\mathrm{S}})-s_{\min}(\mathrm{C}_{\mathrm{S}})|^2 \le \|\widehat{\mathrm{C}}_{\mathrm{S}}-\mathrm{C}_{\mathrm{S}}\|_2^2 \le \sum_{i\in[N]}\sum_{j\in\mathrm{S}}|\widehat{\mathrm{C}}_{i,j}-\mathrm{C}_{i,j}|^2$$
$$\lesssim_P \frac{|S|N\log(Np)}{T} \le \frac{\bar{s}_T N\log(Np)}{T} = o(N),$$

where the first inequality follows from Weyl's inequality, the second from the definition of ℓ_2-norm, the third from Lemma 1, and the last bound from Assumption 7. Combining this bound with Assumption 4 yields the lower bound in the asserted claim. To obtain the upper bound, we proceed similarly by noting that Weyl's inequality also gives $|s_{\max}(\widehat{\mathrm{C}}_{\mathrm{S}})-s_{\max}(\mathrm{C}_{\mathrm{S}})|^2 \le \|\widehat{\mathrm{C}}_{\mathrm{S}}-\mathrm{C}_{\mathrm{S}}\|_2^2$. *Q.E.D.*

LEMMA 4: *Suppose that Assumptions 1, 4, and 7 are satisfied. Then for any $c>0$,*

$$-\frac{k}{\sqrt{1+c}}\|y-x\|_2^2 \ge \ell(y)-\ell(x)-\nabla\ell(x)^\top(y-x) \ge -\sqrt{1+c}\,K\|y-x\|_2^2$$

for all $x,y\in\mathbb{R}^p$ satisfying $\|y-x\|_0 \le 3\bar{s}_T+1$ w.p.a.1.

PROOF: Observe that for all $x\in\mathbb{R}^p$,

$$\ell(x) = -\frac{1}{N}\widehat{\mathrm{Y}}^\top\widehat{\mathrm{Y}} + \frac{2}{N}\widehat{\mathrm{Y}}^\top\widehat{\mathrm{C}}x - \frac{1}{N}x^\top\widehat{\mathrm{C}}^\top\widehat{\mathrm{C}}x \quad \text{and} \quad \nabla\ell(x) = \frac{2}{N}\widehat{\mathrm{C}}^\top\widehat{\mathrm{Y}} - \frac{2}{N}\widehat{\mathrm{C}}^\top\widehat{\mathrm{C}}x. \tag{11}$$

Thus, for all $x,y\in\mathbb{R}^p$,

$$\ell(y)-\ell(x)-\nabla\ell(x)^\top(y-x) = \frac{2}{N}\widehat{\mathrm{Y}}^\top\widehat{\mathrm{C}}(y-x) - \frac{1}{N}y^\top\widehat{\mathrm{C}}^\top\widehat{\mathrm{C}}y + \frac{1}{N}x^\top\widehat{\mathrm{C}}^\top\widehat{\mathrm{C}}x$$
$$-\frac{2}{N}\widehat{\mathrm{Y}}^\top\widehat{\mathrm{C}}(y-x) + \frac{2}{N}x^\top\widehat{\mathrm{C}}^\top\widehat{\mathrm{C}}(y-x) = -\frac{1}{N}(y-x)^\top\widehat{\mathrm{C}}^\top\widehat{\mathrm{C}}(y-x).$$

The asserted claim follows from combining this equality with Lemma 3. *Q.E.D.*

PROOF OF THEOREM 1: For all $\mathrm{S}\subset[p]$, denote

$$\mathscr{F}(\mathrm{S}) = \max_{x\in\mathbb{R}^p:\,\mathrm{supp}(x)\subset\mathrm{S}} \ell(x).$$

Observe that under Assumption 3, with probability approaching one, the set $\widehat{\mathrm{S}}$ is equal to the set S produced by Algorithm 1 with $R^2_{FM}(\cdot)$ replaced by $\mathscr{F}(\cdot)$ and

$$\widehat{\psi}_{\widehat{\mathrm{S}}} \in \operatorname*{arg\,max}_{x\in\mathbb{R}^p:\ \mathrm{supp}(x)\subset\widehat{\mathrm{S}}} \ell(x).$$

Thus, to prove the asserted claim, we can apply Theorem 6 in Elenberg et al. (2018). To do so, we first derive three auxiliary bounds.

First, using (11), we have $\nabla\ell(\psi[\mathrm{S}_0]) = 2\widehat{\mathrm{C}}^\top(\widehat{\mathrm{Y}} - \widehat{\mathrm{C}}_{\mathrm{S}_0}\psi_{\mathrm{S}_0})/N$. Thus,

$$\|\nabla\ell(\psi[\mathrm{S}_0])\|_\infty \le \frac{2}{N}\max_{j\in[p]}\|\widehat{\mathrm{C}}_{\{j\}}\|_2 \times \|\widehat{\mathrm{Y}} - \widehat{\mathrm{C}}_{\mathrm{S}_0}\psi_{\mathrm{S}_0}\|_2$$

Here, $\max_{j\in[p]}\|\widehat{\mathrm{C}}_{\{j\}}\|_2 \lesssim_P \sqrt{N}$ by Lemma 3. Also, by (4),

$$\|\widehat{\mathrm{Y}} - \widehat{\mathrm{C}}_{\mathrm{S}_0}\psi_{\mathrm{S}_0}\|_2 \le \|\widehat{\mathrm{Y}} - \mathrm{Y}\|_2 + \|\mathrm{C}\psi - \mathrm{C}_{\mathrm{S}_0}\psi_{\mathrm{S}_0}\|_2 + \|(\mathrm{C}_{\mathrm{S}_0} - \widehat{\mathrm{C}}_{\mathrm{S}_0})\psi_{\mathrm{S}_0}\|_2.$$

In addition, $\|\widehat{\mathrm{Y}} - \mathrm{Y}\|_2 \lesssim_P \sqrt{N\log(Np)/T}$ by Assumption 1 and $\|\mathrm{C}\psi - \mathrm{C}_{\mathrm{S}_0}\psi_{\mathrm{S}_0}\|_2 \lesssim_P \sqrt{N\log(Np)/T}$ by Assumption 2. Moreover,

$$\begin{aligned}
\|(\mathrm{C}_{\mathrm{S}_0} - \widehat{\mathrm{C}}_{\mathrm{S}_0})\psi_{\mathrm{S}_0}\| &\le \|(\mathrm{C} - \widehat{\mathrm{C}})\psi\|_2 + \|(\mathrm{C}_{\mathrm{S}_0^c} - \widehat{\mathrm{C}}_{\mathrm{S}_0^c})\psi_{\mathrm{S}_0^c}\|_2 \\
&\le \|(\mathrm{C} - \widehat{\mathrm{C}})\psi\|_2 + \max_{j\in[p]}\|\mathrm{C}_{\{j\}} - \widehat{\mathrm{C}}_{\{j\}}\|_2 \times \|\psi_{\mathrm{S}_0^c}\|_1 \lesssim_P \sqrt{\frac{N\log(Np)}{T}}
\end{aligned}$$

by the triangle inequality, Lemmas 1 and 2, and Assumptions 2 and 7. Thus,

$$\|\nabla\ell(\psi[\mathrm{S}_0])\|_\infty \lesssim_P \sqrt{\frac{\log(Np)}{T}}, \tag{12}$$

which is the first bound we need.

Second,

$$|\ell(\psi[\mathrm{S}_0]) - \ell(\mathbf{0}_p)| \le \frac{1}{N}\|\widehat{\mathrm{Y}} - \widehat{\mathrm{C}}\psi[\mathrm{S}_0]\|_2^2 + \frac{1}{N}\|\widehat{\mathrm{Y}}\|_2^2 \lesssim_P 1 \tag{13}$$

by the bounds above and Assumption 6.

Third, given that $|\mathrm{S}_0| + |\widehat{\mathrm{S}}| \le 2\bar{s}_T$ w.p.a.1 by Assumptions 2 and 5 and $|\widehat{\mathrm{S}}| \ge (1 + c)(K/k)|\mathrm{S}_0|\log T$ w.p.a.1 by Assumption 5, it follows from Lemma 4 above and Theorem 1 and Corollary 1 in Elenberg et al. (2018) that

$$\ell(\widehat{\psi}) - \ell(\mathbf{0}_p) \ge (1 - T^{-1})(\ell(\psi[\mathrm{S}_0]) - \ell(\mathbf{0}_p)) \tag{14}$$

w.p.a.1.

Now, combining (12), (13), and (14) above with Theorem 6 in Elenberg et al. (2018), we obtain

$$\|\widehat{\psi}-\psi[\mathrm{S}_0]\|_2^2 \lesssim_P (|\mathrm{S}_0|+|\widehat{\mathrm{S}}|)\frac{\log(Np)}{T}+\frac{1}{T} \lesssim_P \frac{|\mathrm{S}_0|\log^2(NTp)}{T}, \tag{15}$$

where the second inequality follows from Assumption 5. Thus,

$$\begin{aligned}\|\widehat{\psi}-\psi\|_2^2 &\lesssim \|\widehat{\psi}-\psi[\mathrm{S}_0]\|_2^2+\|\psi[\mathrm{S}_0]-\psi\|_2^2\\ &\le \|\widehat{\psi}-\psi[\mathrm{S}_0]\|_2^2+\|\psi[\mathrm{S}_0]-\psi\|_1^2 \lesssim_P \frac{|\mathrm{S}_0|\log^2(NTp)}{T}\end{aligned}$$

by the triangle inequality and Assumption 2, yielding the first asserted claim. Also,

$$\begin{aligned}\|\widehat{\psi}-\psi\|_1^2 &\lesssim \|\widehat{\psi}-\psi[\mathrm{S}_0]\|_1^2+\|\psi[\mathrm{S}_0]-\psi\|_1^2\\ &\le (|\mathrm{S}_0|+|\widehat{\mathrm{S}}|)\|\widehat{\psi}-\psi[\mathrm{S}_0]\|_2^2+\|\psi[\mathrm{S}_0]-\psi\|_1^2 \lesssim_P \frac{|\mathrm{S}_0|^2\log^3(NTp)}{T}\end{aligned}$$

by the Cauchy-Schwarz inequality and Assumption 2, yielding the second asserted claim. *Q.E.D.*

APPENDIX C: PROOF OF THEOREM 2

In this Appendix, we use the same additional notation as that introduced in the previous Appendix. Also, for all $j\in[p]$, let $\ell_j\colon \mathbb{R}^{p-1}\to\mathbb{R}$ be the function defined by $\ell_j(x)=\|\widehat{\mathrm{C}}_{\{j\}}-\widehat{\mathrm{C}}_{\{j\}^c}x\|_2^2/N$ for all $x\in\mathbb{R}^{p-1}$. In addition, for all $j\in[p]$, let $\widetilde{\mathrm{S}}_j=\widehat{\mathrm{S}}_{D,j}\setminus\{j\}$ and let $\widehat{\varphi}_j=(\widehat{\varphi}_{j,1},\dots,\widehat{\varphi}_{j,p})^\top\in\mathbb{R}^p$ be the vector defined by $\widehat{\varphi}_{j,j}=0$ and

$$\widehat{\varphi}_{j,\{j\}^c}=\left(\widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}_j}^\top\widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}_j}\right)^{-1}\widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}_j}^\top\widehat{\mathrm{C}}_{\{j\}}. \tag{16}$$

To prove Theorem 2, we first state and prove seven auxiliary lemmas.

LEMMA 5: *Suppose that Assumptions 4, 8, and 14 are satisfied. Then for all $j\in[p]$, we have $\|\varphi_j\|_2\lesssim 1$.*

PROOF: Fix $j\in[p]$. It follows from the equality $R_j=\mathrm{C}_{\{j\}}-\mathrm{C}_{\mathrm{S}_{0,j}}\varphi_{j,\mathrm{S}_{0,j}}$ that

$$\varphi_{j,\mathrm{S}_{0,j}}=(\mathrm{C}_{\mathrm{S}_{0,j}}^\top\mathrm{C}_{\mathrm{S}_{0,j}})^{-1}(\mathrm{C}_{\mathrm{S}_{0,j}}^\top\mathrm{C}_{\{j\}}-\mathrm{C}_{\mathrm{S}_{0,j}}^\top R_j),$$

where the matrix $C_{S_{0,j}}^\top C_{S_{0,j}}$ is non-singular by Assumptions 4 and 8. Thus,

$$\|\varphi_{j,S_{0,j}}\|_2 \leq \frac{\|C_{S_{0,j}}^\top C_{\{j\}}\|_2 + \|C_{S_{0,j}}^\top R_j\|_2}{\lambda_{\min}(C_{S_{0,j}}^\top C_{S_{0,j}})} \lesssim 1 + \sqrt{\frac{|S_{0,j}|\log(Np)}{T}} \lesssim 1$$

by Assumptions 4, 8, and 14. Since $\varphi_{j,l} = 0$ for all $l \in S_{0,j}^c$, the asserted claim follows. *Q.E.D.*

LEMMA 6: *Suppose that Assumptions 4, 8, and 14 are satisfied. Then for all $j \in [p]$, we have $N \lesssim \|R_j\|_2^2 \lesssim N$.*

PROOF: Fix $j \in [p]$. Let $v = (v_1, \ldots, v_p)^\top \in \mathbb{R}^p$ be the vector defined by $v_j = 1$, $v_l = -\varphi_{j,l}$ for all $l \in S_{0,j}$, and $v_l = 0$ for all $l \in (S_{0,j} \cup \{j\})^c$. Then $R_j = Cv$ and $\|v\|_0 \leq \bar{s}_T + 1$ by Assumption 8. Therefore,

$$\|R_j\|_2^2 = R_j^\top R_j = (Cv)^\top (Cv) = v^\top (C^\top C) v \geq Nk\|v\|_2^2 \geq kN,$$

yielding the lower bound in the asserted claim. Further,

$$\begin{aligned} \|R_j\|_2^2 &= R_j^\top (C_{\{j\}} - C\varphi_j) \leq \|R_j\|_2 \times \|C_{\{j\}}\|_2 + \|C_{\{j\}^c}^\top R_j\|_\infty \times \|\varphi_j\|_1 \\ &\lesssim \sqrt{N}\|R_j\|_2 + N\sqrt{\log(Np)/T}\sqrt{|S_{0,j}|}\|\varphi_j\|_2 \lesssim \sqrt{N}\|R_j\|_2 + N, \end{aligned} \tag{17}$$

where the second inequality follows from Assumption 4 and 8 and the third from Assumptions 8 and 14 and Lemma 5. In turn, (17) implies the upper bound in the asserted claim. *Q.E.D.*

LEMMA 7: *Suppose that Assumptions 1, 9, and 14 are satisfied. Then for all $j \in [p]$, we have*

$$\|(\widehat{C} - C)\varphi_j\|_2 \lesssim_P \sqrt{\frac{N\log(Np)}{T}}.$$

PROOF: The claim follows from the same argument as that in the proof of Lemma 2, where we now use Assumption 9 and replace Assumption 7 by Assumption 14. *Q.E.D.*

LEMMA 8: *Suppose that Assumptions 1, 4, and 14 are satisfied. Then for any $c > 0$,*

$$-\frac{k}{\sqrt{1+c}}\|y - x\|_2^2 \geq \ell_j(y) - \ell_j(x) - \nabla \ell_j(x)^\top (y - x) \geq -\sqrt{1+c}\,K\|y - x\|_2^2$$

for all $x, y \in \mathbb{R}^{p-1}$ *satisfying* $\|y-x\|_0 \le 3\bar{s}_T + 1$ *w.p.a.1.*

PROOF: The claim follows from the same argument as that in the proof of Lemma 4, where we replace Assumption 7 by Assumption 14. *Q.E.D.*

LEMMA 9: *Suppose that Assumptions 1, 4, 8, 9, 10, and 14 are satisfied. Then for all* $j \in [p]$,

$$\|\widehat{\varphi}_j - \varphi_j\|_2 \lesssim_P \sqrt{\frac{\bar{s}_T \log(Np)}{T}}.$$

PROOF: Fix $j \in [p]$. As in the proof of Theorem 1, we start with three preliminary bounds. First,

$$\begin{aligned}\nabla \ell_j(\varphi_{j,\{j\}^c}) &= \frac{2}{N}\widehat{C}_{\{j\}^c}^\top(\widehat{C}_{\{j\}} - \widehat{C}_{S_{0,j}}\varphi_{S_{0,j}}) \\ &= \frac{2}{N}(\widehat{C}_{\{j\}^c} - C_{\{j\}^c})^\top(\widehat{C}_{\{j\}} - \widehat{C}_{S_{0,j}}\varphi_{S_{0,j}}) + \frac{2}{N}C_{\{j\}^c}^\top(\widehat{C}_{\{j\}} - \widehat{C}_{S_{0,j}}\varphi_{S_{0,j}}).\end{aligned}$$

Here,

$$\begin{aligned}\|\widehat{C}_{\{j\}} - \widehat{C}_{S_{0,j}}\varphi_{S_{0,j}}\|_2 &\le \|\widehat{C}_{\{j\}} - C_{\{j\}}\|_2 \\ &+ \|C_{\{j\}} - C_{S_{0,j}}\varphi_{S_{0,j}}\|_2 + \|(C_{S_{0,j}} - \widehat{C}_{S_{0,j}})\varphi_{S_{0,j}}\|_2 \lesssim_P \sqrt{N}\end{aligned} \tag{18}$$

by Lemmas 1, 6, and 7. Thus,

$$\left\|(\widehat{C}_{\{j\}^c} - C_{\{j\}^c})^\top(\widehat{C}_{\{j\}} - \widehat{C}_{S_{0,j}}\varphi_{S_{0,j}})\right\|_\infty \lesssim_P N\sqrt{\frac{\log(Np)}{T}}$$

by the Cauchy-Schwarz inequality and Lemma 1. Also,

$$\begin{aligned}\|C_{\{j\}^c}^\top(\widehat{C}_{\{j\}} - \widehat{C}_{S_{0,j}}\varphi_{S_{0,j}})\|_\infty &\le \|C_{\{j\}^c}^\top(\widehat{C}_{\{j\}} - C_{\{j\}})\|_\infty \\ &+ \|C_{\{j\}^c}^\top(C_{\{j\}} - C_{S_{0,j}}\varphi_{S_{0,j}})\|_\infty + \|C_{\{j\}^c}^\top(C_{S_{0,j}} - \widehat{C}_{S_{0,j}})\varphi_{S_{0,j}}\|_\infty \lesssim_P N\sqrt{\frac{\log(Np)}{T}}\end{aligned}$$

by Assumptions 4 and 8 and Lemmas 1 and 7. Thus,

$$\|\nabla \ell_j(\varphi_{j,\{j\}^c})\|_\infty \lesssim_P \sqrt{\frac{\log(Np)}{T}}, \tag{19}$$

which is the first bound we need.

Second,

$$|\ell_j(\varphi_{j,\{j\}^c})-\ell_j(\mathbf{0}_{p-1})|\leq\frac{1}{N}\|\widehat{\mathrm{C}}_{\{j\}}-\widehat{\mathrm{C}}_{\mathrm{S}_{0,j}}\varphi_{\mathrm{S}_{0,j}}\|_2^2+\frac{1}{N}\|\widehat{\mathrm{C}}_{\{j\}}\|_2^2\lesssim_P 1 \tag{20}$$

by the bound in (18) and Lemma 3, which is applicable because Assumption 14 implies Assumption 7.

Third, given that $|\mathrm{S}_{0,j}|+|\widehat{\mathrm{S}}_j|\leq 2\bar{s}_T$ w.p.a.1 by Assumptions 8 and 10 and $|\widehat{\mathrm{S}}_j|\geq(1+c)(K/k)|\mathrm{S}_{0,j}|\log T$ w.p.a.1 by Assumption 10, it follows from Lemma 8 above and Theorem 1 and Corollary 1 in Elenberg et al. (2018) that

$$\ell_j(\widetilde{\varphi}_j)-\ell_j(\mathbf{0}_p)\geq(1-T^{-1})(\ell_j(\varphi_{j,\{j\}^c})-\ell(\mathbf{0}_{p-1})) \tag{21}$$

w.p.a.1 , where we denoted $\widetilde{\varphi}_j=(\widehat{\mathrm{C}}_{\widehat{\mathrm{S}}_j}^\top\widehat{\mathrm{C}}_{\widehat{\mathrm{S}}_j})^{-1}\widehat{\mathrm{C}}_{\widehat{\mathrm{S}}_j}^\top\widehat{\mathrm{C}}_{\{j\}}$; note the difference between $\widetilde{\varphi}_j$ and $\widehat{\varphi}_{j,\{j\}^c}$ in (16).

Now, we proceed as in the proof of Theorem 6 in Elenberg et al. (2018). (We can not directly apply their theorem as it would be a result about $\widetilde{\varphi}_j$, and we need a result about $\widehat{\varphi}_j$.) We have

$$\ell_j(\widehat{\varphi}_{j,\{j\}^c})-\ell_j(\varphi_{j,\{j\}^c})\geq\ell_j(\widetilde{\varphi}_j)-\ell_j(\varphi_{j,\{j\}^c})\geq T^{-1}(\ell_j(\mathbf{0}_{p-1})-\ell_j(\varphi_{j,\{j\}^c})) \tag{22}$$

w.p.a.1, where the first inequality follows from the fact that $\widehat{\mathrm{S}}_j\subset\widetilde{\mathrm{S}}_j$ and the second from rearranging the terms in (21). Also, denoting $\Delta=\widehat{\varphi}_{j,\{j\}^c}-\varphi_{j,\{j\}^c}$, we have

$$\ell_j(\widehat{\varphi}_{j,\{j\}^c})-\ell_j(\varphi_{j,\{j\}^c})-\nabla\ell_j(\varphi_{j,\{j\}^c})^\top\Delta\leq -k\|\Delta\|_2^2/2 \tag{23}$$

w.p.a.1 by Lemma 8 since $\|\widehat{\varphi}_{j,\{j\}^c}-\varphi_{j,\{j\}^c}\|_0\leq 3\bar{s}_T$ w.p.a.1 by Assumptions 5, 8, and 10. Combining (22) and (23), we have

$$\begin{aligned}k\|\Delta\|_2^2/2 &\lesssim_P \nabla\ell_j(\varphi_{j,\{j\}^c})^\top\Delta+T^{-1}(\ell_j(\varphi_{j,\{j\}^c})-\ell_j(\mathbf{0}_{p-1}))\\ &\lesssim_P \|\nabla\ell_j(\varphi_{j,\{j\}^c})\|_\infty\|\Delta\|_1+\frac{1}{T}\lesssim_P\|\Delta\|_2\sqrt{\frac{\bar{s}_T\log(Np)}{T}}+\frac{1}{T},\end{aligned}$$

where the second inequality follows from (20) and the third from (19). The last bound in turn implies the asserted claim. *Q.E.D.*

LEMMA 10: *Suppose that Assumption 12 is satisfied. Then for all* $j \in [p]$,

$$\phi_{j,\{j\}^c} = \frac{1}{N} \sum_{i \in [N]} (\beta_{i,j} \eta_{j,\{j\}^c} + \beta_{i,\{j\}^c}) R_{j,i}.$$

PROOF: The asserted claim follows from substituting (2) and $v_{t,j} = \eta_{j,\{j\}^c} v_{t,\{j\}^c} + z_{t,j}$ into (9) and cancelling the matrices, which is feasible under Assumption 12. *Q.E.D.*

LEMMA 11: *Suppose that Assumptions 8 and 12 are satisfied. Then for all* $j \in [p]$, *we have* $\|\phi_j\|_\infty \lesssim \sqrt{\log(Np)/T}$.

PROOF: Fix $j \in [p]$ and recall that for all $i \in [N]$ and $l \in [p]$, we have

$$\mathrm{C}_{i,l} = \beta_{i,j} \mathrm{E}[v_{t,j} v_{t,l}] + \beta_{i,\{j\}^c}^\top \mathrm{E}[v_{t,\{j\}^c} v_{t,l}].$$

Combining this equality with (8), it follows that

$$(\mathrm{C}_{i,l})_{l \in \{j\}^c}^\top = \mathrm{E}[v_{t,\{j\}^c} v_{t,\{j\}^c}^\top](\beta_{i,j} \eta_{j,\{j\}^c} + \beta_{i,\{j\}^c}).$$

Therefore,

$$\left\| \frac{1}{N} \sum_{i \in [N]} \mathrm{E}[v_{t,\{j\}^c} v_{t,\{j\}^c}^\top](\beta_{i,j} \eta_{j,\{j\}^c} + \beta_{i,\{j\}^c}) R_{j,i} \right\|_\infty \lesssim \sqrt{\frac{\log(Np)}{T}}$$

by Assumption 8. Hence,

$$\|\phi_{j,\{j\}^c}\|_\infty = \left\| \frac{1}{N} \sum_{i \in [N]} (\beta_{i,j} \eta_{j,\{j\}^c} + \beta_{i,\{j\}^c}) R_{j,i} \right\|_\infty \lesssim \sqrt{\frac{\log(Np)}{T}},$$

where the equality follows from Lemma 10 and the inequality from Lemma 13 and Assumption 12. Combining this bound with the fact that $\phi_{j,j} = 0$, yields the asserted claim. *Q.E.D.*

LEMMA 12: *Suppose that Assumptions 8, 12, and 14 are satisfied. Then for all* $j \in [p]$,

$$R_j^\top R_j = \sum_{i \in [N]} R_{j,i} \beta_{i,j} \sigma_{z,j}^2 + o_P(N).$$

PROOF: Fix $j \in [p]$. Note that for all $i \in [N]$,

$$\begin{aligned}\mathrm{C}_{i,j} &= \beta_i^\top \mathrm{E}[v_t v_{t,j}] = \beta_{i,j}\mathrm{E}[v_{t,j}^2] + \beta_{i,\{j\}^c}^\top \mathrm{E}[v_{t,\{j\}^c} v_{t,j}] \\ &= \beta_{i,j}\sigma_{z,j}^2 + \beta_{i,j}\eta_{j,\{j\}^c}^\top \mathrm{E}[v_{t,\{j\}^c} v_{t,\{j\}^c}^\top]\eta_{t,\{j\}^c} + \beta_{i,\{j\}^c}^\top \mathrm{E}[v_{t,\{j\}^c} v_{t,\{j\}^c}^\top]\eta_{t,\{j\}^c},\end{aligned}$$

where the last equality follows from $v_{t,j} = \eta_{t,\{j\}^c}^\top v_{t,\{j\}^c} + z_{t,j}$ and (8). Also, for all $i \in [N]$ and $l \in \{j\}^c$,

$$\begin{aligned}\mathrm{C}_{i,l} &= \beta_i^\top \mathrm{E}[v_t v_{t,l}] = \beta_{i,j}\mathrm{E}[v_{t,j} v_{t,l}] + \beta_{i,\{j\}^c}^\top \mathrm{E}[v_{t,\{j\}^c} v_{t,l}] \\ &= \beta_{i,j}\eta_{j,\{j\}^c}^\top \mathrm{E}[v_{t,\{j\}^c} v_{t,l}] + \beta_{i,\{j\}^c}^\top \mathrm{E}[v_{t,\{j\}^c} v_{t,l}]\end{aligned}$$

by the same argument. Thus, $\mathrm{C}_{i,j} = \beta_{i,j}\sigma_{z,j}^2 + \sum_{l\in\{j\}^c}\mathrm{C}_{i,l}\eta_{j,l}$, and so

$$R_j^\top R_j = \sum_{i\in[N]} R_{j,i}\left(\mathrm{C}_{i,j} - \sum_{l\in\{j\}^c}\mathrm{C}_{i,l}\varphi_{j,l}\right) = \sum_{i\in[N]} R_{j,i}\left(\beta_{i,j}\sigma_{z,j}^2 + \sum_{l\in\{j\}^c}\mathrm{C}_{i,l}(\eta_{j,l} - \varphi_{j,l})\right).$$

Therefore,

$$\begin{aligned}\left|R_j^\top R_j - \sum_{i\in[N]} R_{j,i}\beta_{i,j}\sigma_{z,j}^2\right| &\le \left|R_j^\top \mathrm{C}_{\{j\}^c}(\eta_{j,\{j\}^c} - \varphi_{j,\{j\}^c})\right| \\ &\le \Big(\|\eta_{j,\{j\}^c}\|_1 + \|\varphi_{j,\{j\}^c}\|_1\Big)\|\mathrm{C}_{\{j\}^c}^\top R_j\|_\infty \\ &\lesssim (1+\sqrt{\bar{s}_T})N\sqrt{\log(Np)/T},\end{aligned}$$

where the last inequality follows from Assumptions 8 and 12 and Lemmas 5 and 14. Combining this bound with Assumption 14 yields the asserted claim. *Q.E.D.*

PROOF OF THEOREM 2: Fix $j \in [p]$ and, for brevity, write $\widetilde{\mathrm{S}}$, $\widehat{\varphi}$, and φ instead of $\widetilde{\mathrm{S}}_j = \widehat{\mathrm{S}}_{D,j} \setminus \{j\}$, $\widehat{\varphi}_j$, and φ_j, respectively. Then

$$\begin{aligned}\sqrt{T}(\widehat{\psi}_{D,j} - \psi_j) &= \frac{\sqrt{T}(\widehat{\mathrm{C}}_{\{j\}} - \widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}}\widehat{\varphi}_{\widetilde{\mathrm{S}}})^\top(\widehat{\mathrm{Y}} - (\widehat{\mathrm{C}}_{\{j\}} - \widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}}\widehat{\varphi}_{\widetilde{\mathrm{S}}})\psi_j)}{(\widehat{\mathrm{C}}_{\{j\}} - \widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}}\widehat{\varphi}_{\widetilde{\mathrm{S}}})^\top(\widehat{\mathrm{C}}_{\{j\}} - \widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}}\widehat{\varphi}_{\widetilde{\mathrm{S}}})} \\ &= \frac{\sqrt{T}(\widehat{\mathrm{C}}_{\{j\}} - \widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}}\widehat{\varphi}_{\widetilde{\mathrm{S}}})^\top(\widehat{\mathrm{Y}} - \widehat{\mathrm{C}}_{\{j\}}\psi_j - \widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}}\psi_{\widetilde{\mathrm{S}}})}{(\widehat{\mathrm{C}}_{\{j\}} - \widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}}\widehat{\varphi}_{\widetilde{\mathrm{S}}})^\top(\widehat{\mathrm{C}}_{\{j\}} - \widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}}\widehat{\varphi}_{\widetilde{\mathrm{S}}})} \qquad (24)\end{aligned}$$

where the first line follows from the Frisch-Waugh-Lowell theorem, and the second from observing that $(\widehat{\mathrm{C}}_{\{j\}}-\widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}}\widehat{\varphi}_{\widetilde{\mathrm{S}}})^\top\widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}}=\mathbf{0}_{|\widetilde{\mathrm{S}}|}$. To derive the asymptotic distribution of the quantity on the right-hand side of (24), we start with a few preliminary bounds. First,

$$\|\widehat{\mathrm{C}}\widehat{\varphi}-\widehat{\mathrm{C}}\varphi\|_2^2 \lesssim_P N\|\widehat{\varphi}-\varphi\|_2^2 \lesssim_P N\bar{s}_T\log(Np)/T, \tag{25}$$

where the first inequality follows from Lemma 3 since $\|\widehat{\varphi}-\varphi\|_0\le|\widehat{\mathrm{S}}|+|\widehat{\mathrm{S}}_j|+|\mathrm{S}_{0,j}|\le 3\bar{s}_T$ w.p.a.1 by Assumptions 5, 8, and 10 and the second from Lemma 9. Second,

$$\begin{aligned}\|\psi[\widehat{\mathrm{S}}_{D,j}]-\psi[\mathrm{S}_0]\|_2^2 &\lesssim \|\psi[\widehat{\mathrm{S}}_{D,j}]-\psi\|_2^2+\|\psi-\psi[\mathrm{S}_0]\|_2^2 \lesssim \|\psi[\widehat{\mathrm{S}}]-\psi\|_2^2+\|\psi-\psi[\mathrm{S}_0]\|_2^2\\ &\lesssim \|\widehat{\psi}-\psi\|_2^2+\|\psi-\psi[\mathrm{S}_0]\|_2^2 \lesssim_P \|\widehat{\psi}-\psi[\mathrm{S}_0]\|_2^2+\|\psi-\psi[\mathrm{S}_0]\|_2^2 \lesssim_P \bar{s}_T\log(Np)/T,\end{aligned} \tag{26}$$

where the first inequality follows from the triangle inequality, the second from the fact that $\widehat{\mathrm{S}}\subset\widehat{\mathrm{S}}_{D,j}$, the third from the fact that $\operatorname{supp}(\widehat{\psi})=\widehat{\mathrm{S}}$, the fourth from the triangle inequality, and the fifth from the intermediate bound in (15) in the proof of Theorem 1 and Assumption 2. Third,

$$\|\widehat{\mathrm{C}}_{\widehat{\mathrm{S}}_{D,j}}\psi_{\widehat{\mathrm{S}}_{D,j}}-\widehat{\mathrm{C}}_{\mathrm{S}_0}\psi_{\mathrm{S}_0}\|_2^2 \lesssim_P N\|\psi[\widehat{\mathrm{S}}_{D,j}]-\psi[\mathrm{S}_0]\|_2^2 \lesssim_P N\bar{s}_T\log(Np)/T, \tag{27}$$

where the first inequality follows from Lemma 3 since $\|\psi[\widehat{\mathrm{S}}_{D,j}]-\psi[\mathrm{S}_0]\|_0\le|\widehat{\mathrm{S}}|+|\widehat{\mathrm{S}}_j|+1+|\mathrm{S}_0|\le 3\bar{s}_T+1$ w.p.a.1 by Assumptions 2, 5, and 10, and the second from (26). Fourth,

$$\begin{aligned}\|\widehat{\mathrm{C}}_{\mathrm{S}_0}\psi_{\mathrm{S}_0}-\widehat{\mathrm{C}}\psi\|_2 &= \Big\|\sum_{j\in\mathrm{S}_0^c}\widehat{\mathrm{C}}_{\{j\}}\psi_j\Big\|_2 \le \sum_{j\in\mathrm{S}_0^c}\|\widehat{\mathrm{C}}_{\{j\}}\psi_j\|_2\\ &\le \max_{j\in\mathrm{S}_0^c}\|\widehat{\mathrm{C}}_{\{j\}}\|_2\sum_{j\in\mathrm{S}_0^c}|\psi_j| \lesssim_P \sqrt{N\bar{s}_T\log(Np)/T},\end{aligned} \tag{28}$$

where the first inequality follows from the triangle inequality and the third from Lemma 3 and Assumption 2.

Thus, getting back to (24), we have $\widehat{\mathrm{C}}_{\{j\}}-\widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}}\widehat{\varphi}_{\widetilde{\mathrm{S}}}=\widehat{\mathrm{C}}_{\{j\}}-\mathrm{C}_{\{j\}}+\mathrm{C}\varphi+R_j-\widehat{\mathrm{C}}_{\widetilde{S}}\widehat{\varphi}_{\widetilde{\mathrm{S}}}$, where $\|\widehat{\mathrm{C}}_{\{j\}}-\mathrm{C}_{\{j\}}\|_2^2\lesssim_P N\log(Np)/T$ by Lemma 1 and

$$\|\widehat{\mathrm{C}}_{\widetilde{S}}\widehat{\varphi}_{\widetilde{\mathrm{S}}}-\mathrm{C}\varphi\|_2^2=\|\widehat{\mathrm{C}}\widehat{\varphi}-\mathrm{C}\varphi\|_2^2\lesssim\|\widehat{\mathrm{C}}\widehat{\varphi}-\widehat{\mathrm{C}}\varphi\|_2^2+\|\widehat{\mathrm{C}}\varphi-\mathrm{C}\varphi\|_2^2\lesssim_P N\bar{s}_T\log(Np)/T$$

by the triangle inequality, (25), Assumption 8, and Lemma 7. In addition, $\widehat{\mathrm{Y}}-\widehat{\mathrm{C}}_{\{j\}}\psi_j-\widehat{\mathrm{C}}_{\widetilde{\mathrm{S}}}\psi_{\widetilde{\mathrm{S}}}=\widehat{\mathrm{Y}}-\mathrm{Y}+\mathrm{C}\psi-\widehat{\mathrm{C}}_{\widehat{\mathrm{S}}_{D,j}}\psi_{\widehat{\mathrm{S}}_{D,j}}$, where $\|\widehat{\mathrm{Y}}-\mathrm{Y}\|_2^2\lesssim_P N\log(Np)T$ by Assumption 1 and

$$\begin{aligned}\|\widehat{\mathrm{C}}_{\widehat{\mathrm{S}}_{D,j}}\psi_{\widehat{\mathrm{S}}_{D,j}}-\mathrm{C}\psi\|_2^2 &\lesssim \|\widehat{\mathrm{C}}_{\widehat{\mathrm{S}}_{D,j}}\psi_{\widehat{\mathrm{S}}_{D,j}}-\widehat{\mathrm{C}}_{\mathrm{S}_0}\psi_{\mathrm{S}_0}\|_2^2+\|\widehat{\mathrm{C}}_{\mathrm{S}_0}\psi_{\mathrm{S}_0}-\widehat{\mathrm{C}}\psi\|_2^2+\|\widehat{\mathrm{C}}\psi-\mathrm{C}\psi\|_2^2\\ &\lesssim_P N\bar{s}_T\log(Np)/T\end{aligned}$$

by the triangle inequality, (27), (28), and Lemma 2. Moreover,

$$|\sqrt{T}R_j^\top\widehat{\mathrm{C}}_{\widehat{\mathrm{S}}^c_{D,j}}\psi_{\widehat{\mathrm{S}}^c_{D,j}}|\lesssim N\sqrt{\log(Np)}\|\psi_{\widehat{\mathrm{S}}^c_{D,j}}\|_1\le N\sqrt{\log(Np)}\|\widehat{\psi}-\psi\|_1=o_P(N),$$

where the first inequality follows from Assumption 8, the second from the fact that $\widehat{\mathrm{S}}\subset\widehat{\mathrm{S}}_{D,j}$, and the third form Theorem 1 and Assumption 14. Combining these bounds and using Lemma 6 and Assumption 14, it follows that

$$\sqrt{T}(\widehat{\psi}_{D,j}-\psi_j)=\frac{\sqrt{T}R_j^\top(\widehat{\mathrm{Y}}-\widehat{\mathrm{C}}\psi)+o_P(N)}{R_j^\top R_j+o_P(N)}.$$

Next, observe that $R_j^\top(\widehat{\mathrm{Y}}-\widehat{\mathrm{C}}\psi)=\mathrm{I}_1+\mathrm{I}_2+\mathrm{I}_3+\mathrm{I}_4$, where

$$\begin{aligned}\mathrm{I}_1&=\sum_{i\in[N]}R_{j,i}\beta_i^\top\left(\mathrm{E}[v_tv_t^\top]-\frac{1}{T}\sum_{t=1}^T f_t(f_t-\bar{f})^\top\right)\psi,\\ \mathrm{I}_2&=\frac{1}{T}\sum_{i\in[N]}\sum_{t\in[T]}R_{j,i}\beta_i^\top(f_t-\mathrm{E}[f_t]),\quad \mathrm{I}_3=\frac{1}{T}\sum_{i\in[N]}\sum_{t\in[T]}R_{j,i}\varepsilon_{i,t},\\ \mathrm{I}_4&=-\frac{1}{T}\sum_{i\in[N]}\sum_{t\in[T]}R_{j,i}\varepsilon_{i,t}(f_t-\bar{f})^\top\psi.\end{aligned}$$

Here,

$$\mathrm{I}_4=-\frac{1}{T}\sum_{i\in[N]}\sum_{t\in[T]}R_{j,i}\varepsilon_{i,t}v_t^\top\psi+o_P(N/\sqrt{T})$$

by Lemma 6 and Assumptions 1 and 14. Thus,

$$\mathrm{I}_3+\mathrm{I}_4=\frac{1}{T}\sum_{i\in[N]}\sum_{t\in[T]}R_{j,i}\varepsilon_{i,t}m_t=o_P(N/\sqrt{T})$$

by Assumption 11. Similarly,

$$\mathrm{I}_1+\mathrm{I}_2=\frac{1}{T}\sum_{i\in[N]}\sum_{t\in[T]}R_{j,i}\beta_i^\top\Big(v_t m_t-\mathrm{E}[v_t m_t]\Big)+o_P(N/\sqrt{T}),$$

again by Lemma 6 and Assumptions 1 and 14. Thus, substituting $v_{t,j}=\eta_{j,\{j\}^c}v_{t,\{j\}^c}+z_{t,j}$ and using Lemma 10 ,

$$\begin{aligned}\mathrm{I}_1+\mathrm{I}_2=&\frac{1}{T}\sum_{i\in[N]}\sum_{t\in[T]}R_{j,i}\beta_{i,j}(z_{t,j}m_t-\mathrm{E}[z_{t,j}m_t])\\&+\frac{N}{T}\sum_{t\in[T]}\phi_{j,\{j\}^c}^\top(v_{t,\{j\}^c}m_t-\mathrm{E}[v_{t,\{j\}^c}m_t])+o_P(N/\sqrt{T})\\=&\frac{1}{T}\sum_{i\in[N]}\sum_{t\in[T]}R_{j,i}\beta_{i,j}(z_{t,j}m_t-\mathrm{E}[z_{t,j}m_t])+o_P(N/\sqrt{T}),\end{aligned}$$

where the second equality follows from Lemma 11 and Assumptions 1, 13, and 14. Moreover, $R_j^\top R_j=\sum_{i\in[N]}R_{j,i}\beta_{i,j}\sigma_{z,j}^2+o_P(N)$ by Lemma 12.

Combining presented bounds and using Lemma 6, it follows that

$$\sqrt{T}(\widehat{\psi}_{D,j}-\psi_j)=\frac{1}{\sqrt{T}}\sum_{t\in[T]}\frac{z_{t,j}m_t-\mathrm{E}[z_{t,j}m_t]}{\sigma_{z,j}^2}+o_P(1),$$

which implies the asserted claim by invoking Assumption 15. *Q.E.D.*

APPENDIX D: TECHNICAL LEMMAS

LEMMA 13: *Let $A=(A_{i,j})_{i,j\in[n]}\in\mathbb{R}^{n\times n}$ be a symmetric matrix such that $\lambda_{\max}(A)\le C_1$, $\lambda_{\min}(A)\ge 1/C_1$, and $\|A^{-1}\|_{\infty,1}\le C_1$ for some constant $C_1\ge 1$. For any $k\in[n]$, let $B=(A_{i,j})_{i,j\in[n]\setminus k}$ be a sub-matrix of A. Then B is invertible and satisfies $\|B^{-1}\|_{\infty,1}\le C$, where C is a constant depending only on C_1.*

PROOF: Without loss of generality, we will assume that $k=n$, so that

$$A=\begin{pmatrix}B & b\\ b^T & a\end{pmatrix}, \tag{29}$$

where $B=(A_{i,j})_{i,j\in[n-1]}$, $b=(A_{i,n})_{i\in[n-1]}$, and $a=A_{n,n}$. Since $\lambda_{\min}(B)\ge\lambda_{\min}(A)>0$, the matrix B is invertible, yielding the first asserted claim.

To prove the second asserted claim, note that

$$\|B^{-1}\|_{\infty,1} = \sup_{y\in\mathbb{R}^{n-1}:\,\|y\|_\infty=1} \|B^{-1}y\|_\infty.$$

Further, fix any $y\in\mathbb{R}^{n-1}$ such that $\|y\|_\infty = 1$ and let $x\in\mathbb{R}^{n-1}$ and $z\in\mathbb{R}$ be such that

$$A\begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix}. \tag{30}$$

Note that such x and z exist because A is invertible. Also,

$$\left\|\begin{pmatrix} x \\ z \end{pmatrix}\right\|_\infty \le \|A^{-1}\|_{\infty,1}\left\|\begin{pmatrix} y \\ 0 \end{pmatrix}\right\|_\infty \le C_1\|y\|_\infty$$

by assumptions of the lemma, and so $\|x\|_\infty \le C_1\|y\|_\infty$ and $|z| \le C_1\|y\|_\infty$. In addition, substituting (29) into (30), it follows that $Bx + bz = y$, and so $B^{-1}y = x + B^{-1}bz$. Thus, by the triangle inequality,

$$\|B^{-1}y\|_\infty \le \|x\|_\infty + \|B^{-1}b\|_\infty \times |z| \le C_1(1+\|B^{-1}b\|_\infty)\|y\|_\infty. \tag{31}$$

It thus remains to bound $\|B^{-1}b\|_\infty$. To do so, observe that it follows from (29) that

$$A\begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix} = \begin{pmatrix} b \\ b^\top B^{-1}b \end{pmatrix}.$$

Thus,

$$\begin{aligned}\|B^{-1}b\|_\infty &\le \|A^{-1}\|_{\infty,1}\Big(\|b\|_\infty + |b^\top B^{-1}b|\Big) \\ &\le \|A^{-1}\|_{\infty,1}\left(\|b\|_2 + \frac{\|b\|_2^2}{\lambda_{\min}(A)}\right) \le \|A^{-1}\|_{\infty,1}\left(\lambda_{\max}(A) + \frac{\lambda_{\max}(A)^2}{\lambda_{\min}(A)}\right),\end{aligned}$$

where the first inequality follows from the definition of the $\ell_{\infty,1}$-norm, the second from noting that $\|b\|_\infty \le \|b\|_2$ and $\|B^{-1}\|_2 = 1/\lambda_{\min}(B) \le 1/\lambda_{\min}(A)$, and the third from noting that $\|b\|_2 \le \|A\|_2$ by the definition of the ℓ_2-norm. Combining this bound with (31) and using $\|y\|_\infty = 1$ yields the second asserted claim and completes the proof of the lemma. *Q.E.D.*

LEMMA 14: *Let* $x=(x_1,\dots,x_p)^\top \in R^p$ *be a random vector such that the matrix* $\mathrm{E}[xx^\top]$ *is non-singular and denote* $\Omega=(\mathrm{E}[xx^\top])^{-1}$. *For any* $j\in[p]$, *consider the least-squares projection*

$$x_j = x_{\{j\}^c}^\top \gamma_j + \epsilon_j, \quad \mathrm{E}[x_{\{j\}^c}\epsilon_j]=\mathbf{0}_{p-1}.$$

Then $\|\gamma_j\|_1=\mathrm{E}[\epsilon_j^2]\sum_{l\in[p]}|\Omega_{lj}|-1$.

PROOF: Without loss of generality, we will assume that $j=p$. Then

$$\mathrm{E}[xx^\top]=\mathrm{E}\left[\begin{pmatrix} x_{\{p\}^c} \\ x_p \end{pmatrix}\begin{pmatrix} x_{\{p\}^c} \\ x_p \end{pmatrix}^\top\right]=\begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where $A=\mathrm{E}[x_{\{p\}^c}x_{\{p\}^c}^\top]$, $B=\mathrm{E}[x_{\{p\}^c}x_p]$, $C=\mathrm{E}[x_p x_{\{p\}^c}^\top]$, and $D=\mathrm{E}[x_p^2]$. Then

$$\Omega=\begin{pmatrix} A^{-1}+A^{-1}BCA^{-1}/E & -A^{-1}B/E \\ -CA^{-1}/E & 1/E \end{pmatrix},$$

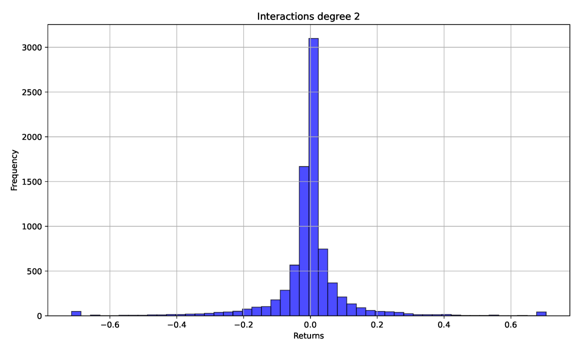
where $E=D-CA^{-1}B$. Therefore, $\sum_{l\in[p]}|\Omega_{lp}|=(\|A^{-1}B\|_1+1)/E$. However, $A^{-1}B=\gamma_p$ and $E=\mathrm{E}[\epsilon_p^2]$. The asserted claim follows. *Q.E.D.*

APPENDIX E: ADDITIONAL FIGURES & TABLES

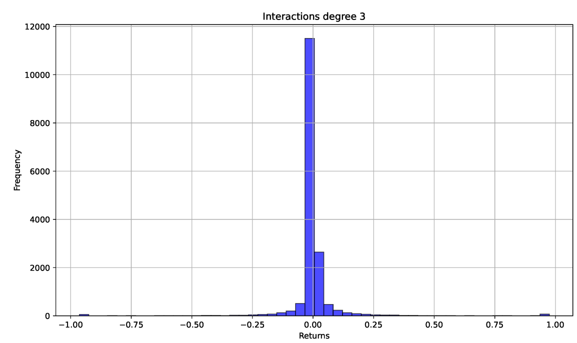
FIGURE E.1.—**Distribution of Higher-Order Factors**

This figure plots the distribution of the high-order factors grouped in pairwise interactions of degree 2 (Panel A), pairwise interactions of degree 2 (Panel B), powers of degree 2 (Panel C) and powers of degree 2 (Panel D). The distribution is clipped at 4 standard deviations from the mean, so that any value below the lower bound is replaced by the lower bound and any value above the upper bound is replaced by the upper bound.

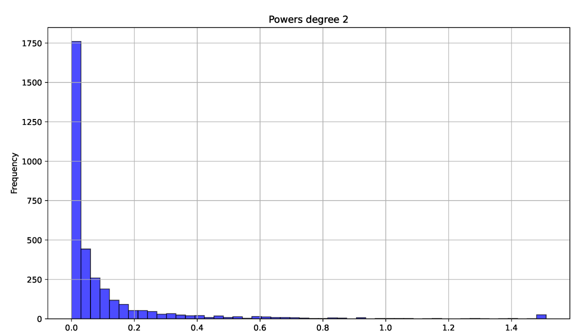
(a)



(b)



(c)



(d)

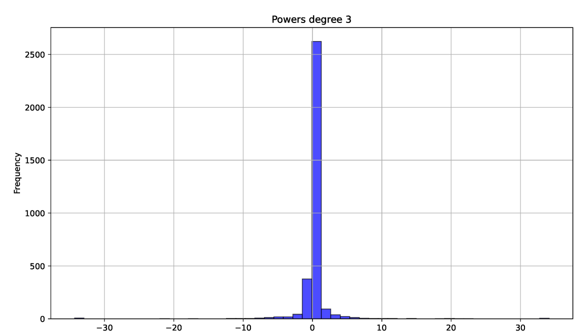


FIGURE E.2.—R^2 of Factor Mimicking Regressions

This table reports, for each of the high-order selected factors with the FS-FMB procedure, the adjusted R-squared of factor mimicking regressions. The factor mimicking regressions are time-series OLS regressions of high-order factor returns on the zoo factor from Jensen et al. (2023).

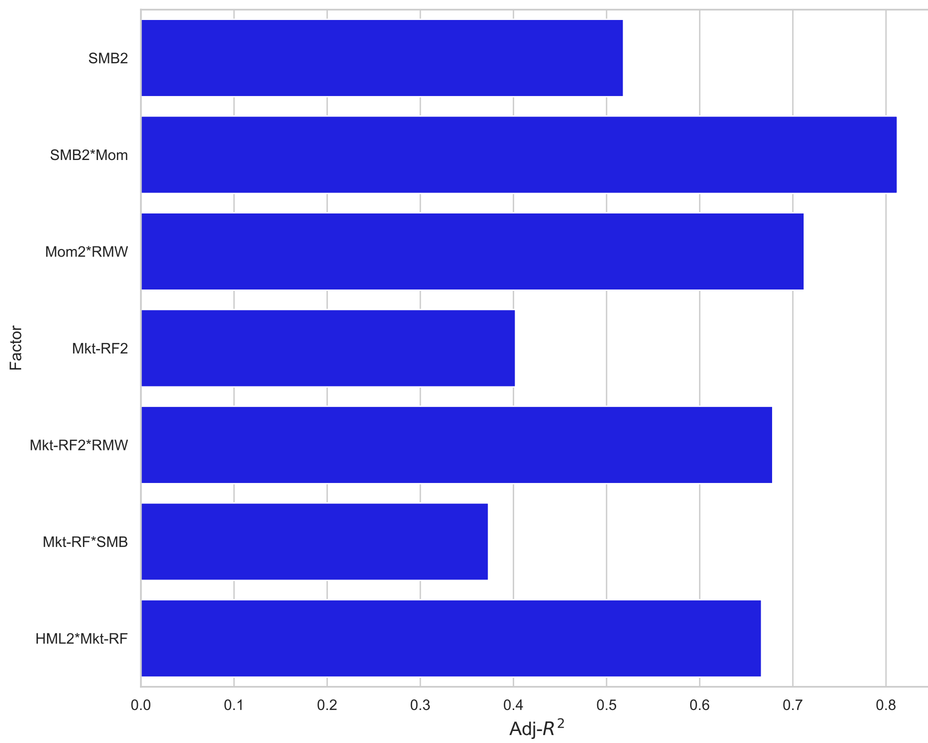


TABLE E.I

CROSS-SECTIONAL PERFORMANCE RESTRICTED MODELS

This table reports the results from the second step of the FMB procedure imposing the restriction that the risk prices of the tradable factors must equal to their sample means. The tradable factors are the FF5M factors. The high-order factors are those selected by the forward selection FMB procedure (see Table II). The table reports the adjusted cross-sectional R-squared, the estimate for the intercept (α) and associated t-statistic. The t-statistics is based on Newey-West corrected standard errors.

#	Model	Adj. R-squared	α	t-stat (α)
1	CAPM	-0.439	0.001	0.28
2	FF3	-0.371	0.001	0.334
3	FF5	-0.111	0.004	2.263
4	FF5M	0.022	0.006	2.839
5	Higher-Order	0.405	0.004	2.663

TABLE E.II

COMPARISON OF R-SQUAREDS WITH RANDOM TRAIN AND TEST SAMPLE

This table reports the in-sample and out-of-sample R-squareds in the cross-sectional regression step of the Fama-MacBeth method for different models. We randomly generate training and test samples of equal length by randomly drawing monthly returns without replacement, and estimate for each asset the covariance with respect to the factors and the SDF loadings on the training sample, and compare actual mean returns with model predictions in the test sample. For the test sample, we report R-squareds aftering re-centering the predicted values from the model. For the FF5M + High-Order model, we consider the high-order factors selected by the forward selection FMB procedure (see Table II).

	(1) CAPM	(2) FF3	(3) FF5	(4) FF5M	(5) Higher-Order
R^2_{train}	0.06	0.167	0.274	0.316	0.498
R^2_{test}	0.063	0.04	0.076	0.085	0.14

x

TABLE E.III

LOADINGS OF ALTERNATIVE ZOO FACTORS

This table reports the fraction of zoo factors with loading significantly different from zero at the 5% confidence level on each of the FF5M factors and high-order factors selected by the FS-FMB procedure. Significance is evaluated by t-statistics based on Newey-West corrected standard errors. The zoo factors are 159 factors constructed in Chen and Zimmermann (2022).

Factor	Frac Sig 5%	Frac Sig 5%
Mkt-RF	0.478	0.509
SMB	0.572	0.579
HML	0.459	0.421
RMW	0.572	0.553
CMA	0.428	0.491
Mom	0.535	0.421
SMB2		0.365
SMB2*Mom		0.478
Mom2*RMW		0.358
Mkt-RF2		0.302
Mkt-RF2*RMW		0.208
Mkt-RF*SMB		0.302
HML2*Mkt-RF		0.302
# zoo factors	159	159

TABLE E.IV

FORWARD SELECTION FMB PROCEDURE WITH ALTERNATIVE HIGHER-ORDER FACTORS

This table reports the results of the estimation of the FS-FMB procedure for alternative sets of higher-order factors. Panel A considers the FF5M factors and all the high-order factors of degree 2 (i.e., $f_i \times f_j$ and f_i^2). Panel B considers the FF5M factors and all the high-order factors up to degree 4 (i.e., $f_i \times f_j$, $f_i^2 \times f_j$, $f_i^2 \times f_j^2$, $f_i^3 \times f_j$, f_i^2, f_i^3, and f_i^4). Panel C considers the FF5M factors and all the pairwise interactions of degree 2 and 3 (i.e., $f_i \times f_j$ and $f_i^2 \times f_j$). Panel D considers the FF5M factors all the powers of degree 2 and 3 (i.e., f_i^2 and f_i^3). For each step, we report the selected factor (h_j) adjusted cross-sectional R-squared, the estimate for the intercept (α) and associated t-statistic. The t-statistic is based on Newey-West corrected standard errors.

Step	h_j	Adj. R-squared	α	t-stat (α)
Panel A: high-order degree 2				
1	SMB2	0.41	0.003	1.821
2	HML*Mom	0.464	0.004	2.033
3	Mkt-RF2	0.503	0.005	2.738
4	SMB*Mom	0.517	0.005	2.692
Panel B: high-order degree up to 4				
1	SMB2*CMA2	0.423	0.005	2.581
2	Mkt-RF2*SMB2	0.464	0.004	2.329
3	Mom3*CMA	0.485	0.004	2.062
4	Mkt-RF2*RMW	0.498	0.004	1.931
5	HML2	0.528	0.003	1.589
6	Mkt-RF2	0.547	0.004	2.271
7	CMA2*Mom	0.57	0.004	2.13
8	CMA2*SMB	0.587	0.004	2.074
9	Mkt-RF2*Mom	0.6	0.004	2.132
Panel C: interactions degree 2 and 3				
1	SMB2*Mkt-RF	0.394	0.003	1.456
2	CMA2*SMB	0.437	0.003	1.429
3	Mkt-RF2*SMB	0.466	0.004	2.023
4	CMA2*RMW	0.487	0.004	2.084
5	HML*CMA	0.509	0.004	2.321
6	CMA2*Mkt-RF	0.527	0.004	2.248
7	CMA2*Mom	0.542	0.004	2.044
Panel D: powers degree 2 and 3				
1	SMB2	0.41	0.003	1.821
2	Mkt-RF2	0.444	0.005	2.478
3	RMW3	0.467	0.005	2.807

TABLE E.V

FORWARD SELECTION FMB PROCEDURE STARTING FROM THE CAPM

This table reports the results from the second step of the FMB method with selection by the FS-FMB procedure starting from the CAPM. At each step, we augment the CAPM by one additional factor selected from the remaining FF5M factors and the higher-order factor (h_j) which maximizes the adjusted cross-sectional R-squared. The method stops when the improvement in the R-squared in a given step is smaller than 1 pp. The table reports the adjusted cross-sectional R-squared, the selected factor and its associated t-statistic, the estimate for the intercept (α) and associated t-statistic. The t-statistics is based on Newey-West corrected standard errors.

Step	h_j	Adj. R-squared	α	t-stat (α)
1	HML2	0.285	0.007	2.916
2	Mom2*Mkt-RF	0.323	0.005	2.153
3	SMB2*Mkt-RF	0.377	0.006	2.48
4	Mkt-RF2	0.406	0.005	2.405
5	HML*Mom	0.429	0.004	2.337
6	Mom	0.485	0.005	2.682
7	CMA2*Mom	0.499	0.006	3.087
8	SMB2*HML	0.509	0.005	2.777
9	Mom2*SMB	0.537	0.005	2.976
10	Mkt-RF3	0.552	0.005	2.722