

HIGH-DIMENSIONAL FACTOR MODELS AND THE  
FACTOR ZOO

Martin Lettau

WORKING PAPER 31719

NBER WORKING PAPER SERIES

HIGH-DIMENSIONAL FACTOR MODELS AND THE FACTOR ZOO

Martin Lettau

Working Paper 31719

<http://www.nber.org/papers/w31719>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

September 2023

I have nothing to disclose. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2023 by Martin Lettau. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

High-Dimensional Factor Models and the Factor Zoo  
Martin Lettau  
NBER Working Paper No. 31719  
September 2023  
JEL No. C38,G0,G12

### **ABSTRACT**

This paper proposes a new approach to the “factor zoo” conundrum. Instead of applying dimension-reduction methods to a large set of portfolios obtained from sorts on characteristics, I construct factors that summarize the information in characteristics across assets and then sort assets into portfolios according to these “characteristic factors”. I estimate the model on a data set of mutual fund characteristics. Since the data set is 3-dimensional (characteristics of funds over time), characteristic factors are based on a tensor factor model (TFM) that is a generalization of 2-dimensional PCA. I find that parsimonious TFM captures over 90% of the variation in the data set. Pricing factors derived from the TFM have high Sharpe ratios and capture the cross-section of fund returns better than standard benchmark models.

Martin Lettau  
Haas School of Business  
University of California, Berkeley  
545 Student Services Bldg. #1900  
Berkeley, CA 94720-1900  
and CEPR  
and also NBER  
lettau@haas.berkeley.edu

## 1. Introduction

This paper proposes a new approach to the “factor zoo” conundrum in asset pricing. Most of the current literature tries to resolve the multidimensional challenge (Cochrane (2011)) by sorting individual assets into portfolios according to a (potentially large) set of characteristics. Factors, or the stochastic discount factor (SDF), are estimated using the panel of portfolio returns as inputs. The econometric methods used in the second step exploit the dependence structure of the cross-section of portfolio returns and “summarize” the information in a small number of factors. One class of such models is based on Principal Component Analysis (PCA), see for example Connor and Korajczyk (1986, 1988), Pelger (2019), Kelly et al. (2019), Lettau and Pelger (2020a,b), and Giglio and Xiu (2021). More recently, the literature has also considered machine learning (ML) methods, such as Lasso regressions (Feng et al. (2020)), elastic nets (Kozak et al. (2020)), regression trees (Bryzgalova et al. (2023a)), and neural nets (Chen et al. (2023)).

The method used in this paper reverses the order of the sorting and factor construction steps. First, factors are constructed in the characteristic space rather than in the return space. These factors capture the dependence structure of characteristics across time, across individual assets, and across characteristics but do not use returns of individual assets. Second, I sort individual assets into portfolios according to these “characteristic factors” (instead of the original characteristics) and obtain asset pricing factors by subtracting the returns of the highest portfolios minus the returns of the lowest portfolios. This approach has several advantages. First, it exploits dependence along all dimensions. This is important since characteristics are likely correlated across time, across assets, and across characteristics. Second, the methodology allows for complex dependence structure across many characteristics. Univariate portfolio sorts according to individual characteristics do not take information in other, potentially correlated, characteristics into account. Sorting on multiple characteristics is often infeasible since the number of individual assets in sorted portfolios decreases geometrically in the number of sorting characteristics. Third, the methodology can accommodate samples with large numbers of individual assets and characteristics even when the number of time series observations is small.

The intuition of the methodology is as follows. Principal Component Analysis is based on an eigenvalue/eigenvector of the covariance matrix of a 2-dimensional panel data set, or, equivalently, on the Singular Value Decomposition (SVD) of the 2-dimensional data matrix.<sup>1</sup> However, the econometrician observes characteristics of individual assets over time thus forming a 3-dimensional data set. Hence, PCA-based methods are not directly applicable and instead require ad hoc methods to eliminate one of the dimensions.<sup>2</sup> The method used in this paper can be understood as a generalization of PCA factor models to higher-dimensional data sets. Formally, data sets with more than two dimensions form *tensors*, which extend the notions of vectors and matrices to higher dimensions.<sup>3</sup> Correspondingly,

---

<sup>1</sup>Alternatively, the correlation matrix can be used. The RP-PCA method of Lettau and Pelger (2020a,b) is based on different second-moment matrices as PCA inputs.

<sup>2</sup>Consider, for example, Balasubramaniam et al. (2023) who study stock ownership in India. While their sample, consisting of stock holdings of individual investors over ten years, is 3-dimensional, they estimate a cross-sectional 2-dimensional factor model for a single period. Therefore, their factor model does not exploit potentially useful time series information.

<sup>3</sup>Tensors were introduced by Ricci-Curbastro and Levi-Civita (1900) and have many applications in physics and engineering.

the high-dimensional factor models used in this paper are based on generalizations of the SVD matrix decomposition and 2-dimensional principal component analysis to tensors and can be applied to any data set with more than two dimensions.<sup>4</sup> A distinct advantage of these models is that they capture dependence structures along all dimensions simultaneously.

Tensor factor models (TFM) are related to 2-dimensional factor models in the following way.<sup>5</sup> Suppose the data set is 3-dimensional with  $T$ ,  $N$ , and  $C$  observations in the three dimensions with a total of  $TNC$  observations. In the empirical implementation, the data set consists of  $C$  characteristics of  $N$  assets observed over  $T$  time periods. An econometrician could estimate separate 2-dimensional PCA models for each time period  $t = 1, \dots, T$ . Each PCA model is based on a matrix of characteristics of individual assets at time  $t$ , hence the PCA factors capture cross-sectional correlations of characteristics across assets in a given period. Alternatively, she could estimate separate PCA models for each asset  $n = 1, \dots, N$ . Each factor model captures the time series correlations across characteristics for a given asset. Finally, to capture time series correlations of a given characteristic  $c = 1, \dots, C$ , she could estimate separate PCA models for each characteristic  $c = 1, \dots, C$ . This approach yields  $T+N+C$  individual PCA models that are estimated separately. Since each PCA only uses two dimensions, potentially useful information is lost and the procedure is therefore inefficient. In contrast, TFM can be understood as the *simultaneous* estimation of the  $T+N+C$  individual PCA models while allowing for dependence across all dimensions. In other words, the tensor factor model yields *interconnected* 2-dimensional factor models that are mutually consistent and exploit correlations in all dimensions. The tensor structure also imposes restrictions on the 2-dimensional factor models. In practice, TFM estimations are based on a single 3-dimensional representation rather than the direct estimation of  $T+N+C$  2-dimensional factor models. The 3-dimensional model in turn implies a system of interconnected  $T+N+C$  2-dimensional factor models.

Similar to PCA factor models, tensor factor models can be interpreted as dimension reduction methods. The information in a large data set is summarized by a small number of factors. In contrast to 2-dimensional PCA, the tensor factor model used in this paper allows for different numbers of factors for each dimension. In other words, the  $T$  observations of the first dimension are summarized by  $K_T$  “time” factors, the  $N$  observations in the second dimension are summarized by  $K_N$  “asset” factors, and the  $C$  observations in the third dimension are summarized by  $K_C$  “characteristic” factors. The form of TFM models is similar to that of the matrix SVD. The matrix of eigenvalues is replaced by a  $(K_T \times K_N \times K_C)$ -dimensional “core” tensor, which is multiplied by three matrices that replace the two eigenvector matrices of the SVD. The TFM can also be written as a 3-dimensional factor model so that the three matrices also have the interpretation of “loading” matrices. However, the “core” tensor and “loading” matrices in the TFM are not linked to eigenvalues and eigenvectors since these notions do not exist for tensors.

Since the objective is to construct factors that summarize the information in characteristics, I show

---

<sup>4</sup>This paper focuses on 3-dimensional applications but all results extend to higher dimensions.

<sup>5</sup>The tensor factor model used in this paper is due to Tucker (1966). There are other high-dimensional factor models that are special cases of the Tucker model, *e.g.*, the CP model (Carroll and Chang (1970), Harshman (1970)). I discuss the differences in Section 3.

that the 3-dimensional tensor factor model implies a 2-dimensional factor representation for each asset. Intuitively, the  $C$  characteristics of each asset are summarized by  $K_C$  “characteristic” factors. Since they are derived from a TFM, these “characteristic” factors exploit dependencies across characteristics, across assets, and across time. The TFM factors can be used to construct asset pricing factors. The only difference is that portfolios are constructed by sorting assets according to the  $K_C$  TFM characteristic factors instead of the  $C$  individual characteristics.

The methodology has three steps. First, I estimate the TFM and obtain the core tensor and loading matrices. Second, I compute  $K_C$  characteristic factors for each asset from the fitted TFM. Third, in each period  $t$  I sort assets according to the  $K_C$  characteristic into decile portfolios and compute the portfolio returns in  $t + 1$ . Finally, I compute  $K_C$  asset pricing factors as the differences between decile-10 and decile-1 portfolios. These pricing factors derived from the TFM can be used in asset pricing tests and can be compared to other factors, such as Fama-French and PCA-based factors. Similar to factors derived from PCA models, the TFM factors are subject to a look-ahead bias since the TFM is estimated using the entire data set. I, therefore, also construct out-of-sample factors that are based on the estimation of TFM using expanding samples, so that TFM factors and portfolios use only past information.

I implement the methodology using a data set consisting of 25 characteristics of 934 mutual funds observed over a sample of 34 quarters totaling 793,900 observations. Characteristics are correlated across assets; for example, the book-to-market ratios of stocks and mutual funds tend to move together. In addition, some characteristics of a single asset might be correlated, *e.g.*, the book-to-market and earnings-to-price ratios. Finally, the cross-sectional correlation of assets and characteristics can vary over time. Hence, applying 2-dimensional models by eliminating one dimension inevitably causes a loss of information.

I evaluate the fit of a range of tensor factor models with different numbers of factors and find that parsimonious TFM capture most of the variation in the data. I compare specifications by their fit as measured by the mean-square error or, equivalently, the  $R^2$ , as well as the parsimony of the model measured by the “compression ratio,” which is defined as one minus the ratio of the number of free parameters of a model and the total number of observations. For example, a TFM with  $(K_T, K_N, K_C) = (8, 12, 12)$  factors captures over 92.9% of the variation in the data with a compression ratio of 98.4%. Dimension reduction is particularly effective in the mutual fund dimension since this specification summarizes the information in all 980 mutual funds in only 12 “mutual fund” factors. Note that a compression ratio is 98.4% corresponds to a standard PCA model with one or two factors for a 2-dimensional data set with 100 variables. The  $R^2$  of a more parsimonious model with  $(K_T, K_N, K_C) = (3, 10, 10)$  is slightly lower, 90.4%, but the compression ratio increases to 98.8%. I find that the models yield good fits for all data points except for some outliers. In addition, the model is stable over the sample, across mutual funds, as well as across characteristics. Many aspects of the tensor model are similar to patterns found in 2-dimensional factor models. For example, the first factors along the three dimensions are “level” factors with positive “long-only” loadings, while higher-order components are “long-short” factors.

Next, I construct  $K_C$  characteristic factors that are implied by the estimated TFM. I consider two

specifications. First, I estimate the TFM model using the entire sample and then construct factors for each fund in each quarter. Since the TFM uses future information, the factors are subject to a look-ahead bias. I, therefore, also estimate TFM recursively on expanding windows and construct factors using only past information. The results for the out-of-sample factors are slightly weaker but follow the same patterns as the in-sample factors. Most factors have economic interpretations. For example, the first factor is related to means of characteristics, while the second factor is related to “value” and “growth” characteristics. To assess whether these TFM factors are related to mutual fund returns, I regress fund returns on lagged characteristic factors and find that most factors are statistically significant and capture 50% of the cross-sectional variation in fund returns. Regressing fund returns on lagged characteristics instead of TFM characteristic factors yields a worse overall fit and few characteristics are statistically significant. Even though no return information is used in the estimation of the tensor factor model, the characteristic factors derived from the TFM are related to returns.

Given the  $K_C$  characteristic factors, I sort mutual funds into decile portfolios in each quarter and obtain  $K_C$  asset pricing factors by forming returns of “long/short” decile-10 minus decile-1 portfolios. Following the same procedure, I also construct such “long/short” portfolios for each of the  $C$  original characteristics as benchmarks. I find that mean returns and Sharpe ratios of many TFM factors are substantially higher than those of characteristic portfolios. For example, the highest annualized Sharpe ratio of all  $C$  characteristic portfolios is 0.46, while the highest Sharpe ratio of in-sample TFM factors is almost twice as high, 0.89, and 0.66 for out-of-sample factors. In addition, several alphas of TFM with respect to the CAPM and 3-factor Fama-French factors are statistically significant while alphas are insignificant for characteristic portfolios. Note that there is no mechanical link between TFM factors and their returns as no return information was used in the construction of the portfolios that underlie the factors.

Finally, I use TFM pricing factors in cross-sectional asset pricing tests and compare the results to those for Fama-French models. I also construct factors using the standard PCA approach, *i.e.*, I construct decile portfolios for each of the  $C$  original factors and then compute a PCA model for the panel of  $10C$  portfolio returns. I find that the TFM factors yield smaller pricing errors than Fama-French and PCA factors. This is not only true for the in-sample TFM factors but also the out-of-sample factors. Moreover, adding TFM factors to Fama-French and PCA factor models improves the fit, while adding Fama-French or PCA factors does not improve the fit of TFM factors.

There are many potential applications of tensor-based methods to model high-dimensional data in finance and economics. For example, databases, such as CRSP and COMPUSTAT, include variables observed for individual stocks and across time and are thus inherently 3-dimensional. The estimation of dynamic corporate finance models often involves data sets with three or more dimensions, see Strebulaev and Whited (2012) for a survey. The investor-level data used in the household finance literature that studies portfolio holdings have more than two dimensions, see for example Odean (1998), Campbell (2006), Calvet et al. (2009). In asset pricing, tensor-based methods are used to study the joint behavior of asset-level characteristics and returns or asset prices across countries. A distinct advantage of high-dimensional factor models compared to 2-dimensional models is that they can be

estimated on data sets with short time series since information in all dimensions is used. This feature makes factor modeling feasible in situations in which 2-dimensional factor models are not applicable. Based on the results in this paper, tensor decompositions are promising additions to the toolbox of economists for modeling higher-dimensional data.

This paper is related to several strands of the literature. Although the term “factor zoo” was coined by Cochrane (2011), concerns about an increasing abundance of cross-sectional factors go back further, see, for example, Subrahmanyam (2010), and have generated a large and diverse literature that tries to address the multidimensionality of risk factors. The May 2020 issue of the *Review of Financial Studies* is dedicated to “new methods for the cross-section of returns,” see Karolyi and Van Nieuwerburgh (2020) for an overview of the included articles. Some approaches follow “classical” econometric methods while others use machine learning methods. Some examples of the former are Harvey et al. (2016), who suggest using a higher hurdle for  $t$ -statistics for any new factors, and Harvey and Liu (2021), who propose a bootstrap model selection approach. McLean and Pontiff (2016) use the time period after publication as an out-of-sample test. I already referenced some applications of PCA and machine learning methods that are based on portfolio sorts above. Other papers apply such methods to data sets with individual stocks rather than portfolios. For example, Pelger (2019) and Pelger and Xiong (2022) develop a PCA estimator for high-frequency observations of individual stocks to identify continuous and jump factors. Chincio et al. (2019), Freyberger et al. (2020), and Martin and Nagel (2022) use regularization methods, such as Lasso and Ridge, while Moritz and Zimmermann (2016) use random forests.

This paper is also related to the large literature on mutual fund performance that goes back to Jensen (1968). Some recent contributions are Gruber (1996), Carhart (1997), Berk and Green (2004), Berk and van Binsbergen (2015), Mamaysky et al. (2007), Fama and French (2010), Harvey and Liu (2018), and Jones and Mo (2021). There is an emerging literature that uses machine learning methods to identify funds that outperform their benchmarks. Li and Rossi (2020) uses regression trees to select funds based on the characteristics of the stocks they are holding. Kaniel et al. (2023) find that fund momentum and fund flows are the most important predictors of risk-adjusted returns based on a neural network estimation. DeMiguel et al. (2021) compare estimations using elastic nets, gradient boosting, and random forest to identify mutual funds with positive alphas.

Finally, there are a few recent papers that model high-dimensional data. Bryzgalova et al. (2023b) (BLLP) an estimation methodology for 2-dimensional cross-sectional panels that are observed over time. Their procedure combines 2-dimensional factor models that are estimated for each period with time series models of the latent factors. BLLP apply their method to infer missing values in a time series panel of stock characteristics. There are several differences between BLLP’s estimator and the methods used in this paper. First, BLLP study 2-dimensional panel data observed over time, while I focus on generic high-dimensional data that may or may not include a time dimension. Second, in its current form, the estimation method in this paper requires a balanced panel without any missing values, while BLLP’s estimation is designed to impute missing data. Chen et al. (2021) and Chen et al. (2022) develop factor models for high-dimensional time series and Babii et al. (2022) propose an estimation method for high-dimensional data set that is based on a different tensor factor model than the one used in

this paper.

The rest of the paper is organized as follows. Section 2 introduces the data set used in the paper. Section 3 introduces the tensor factor model and its estimation. The empirical implementation is described in Section 4 and includes a comparison of tensor models of different orders, a detailed analysis of the fit of the benchmark specification, and develops an economic interpretation of the components of the decomposition. Section 5 studies the pricing factors that are derived from the tensor factor model and shows that they capture the cross-section of mutual fund returns. Section 6 concludes.

## 2. Mutual fund characteristics and returns

The data set includes characteristics of active mutual funds over time and is taken from Lettau et al. (2021). I refer to their paper for a detailed description of the data. Lettau et al. (2021) construct 25 characteristics of mutual funds and exchange-traded funds (ETFs) based on portfolio holdings. Characteristics on the mutual fund level are computed as weighted averages of the characteristics of the stocks in their portfolios and are scaled from 1 (low) to 5 (high). The data set includes seven price-ratios, five growth rates of fundamentals, three “value”/“growth” Morningstar indices, momentum, reversal, size, operating profitability, investment, quality<sup>6</sup>, and four liquidity measures, see Table 1. To obtain a balanced panel with no missing data, I select all mutual funds and ETFs that are in the sample between 2010Q3 and 2018Q4.<sup>7</sup> The final sample consists of  $T = 34$  quarters,  $M = 934$  active mutual funds and ETFs, and  $C = 25$  characteristics for a total of 793,900 observations.

Table 2 shows some properties of the mutual funds in the sample. I first take time series means by funds and then compute descriptive statistics of the distribution of fund means. The median fund has a total net asset value (TNA) of \$677 mil. with an inter-quartile range of \$254 mil. to \$1.79 bil. The mean TNA of \$1.91 bil. is larger than the 75%-th percentile indicating that the TNA distribution is heavily right-skewed. The median fund holds 81 stocks with an interquartile range of 56 to 121 and a mean of 120. The lower panel breaks the sample by fund category. The sample includes 346 “growth” funds (G), 213 “cap-based” funds (C), and 202 “value” funds (V). The remaining 20% of the sample are “sector”(S, 116) and “balanced” funds (B, 45). 12 funds do not fit in this classification and are labeled as “other” (O, 12).

As is well known in the literature, mutual funds underperform broad stock market indices, and alphas of the majority of mutual funds and ETFs are negative. The annualized mean fund excess returns over the sample is 11.67%, well below the mean CRSP-VW return of 13.20%. 233 of the 934 funds in the sample have a higher return than the CRSP-VW index. The average standard deviation of 14.40% is higher than the standard deviation of CRSP-VW returns (13.05%) and excess returns of 722 funds are more volatile than the CRSP-VW index. The interquartile ranges of mean returns and mean standard deviations are (10.56%,11.84%) and (13.20%,15.65%), respectively. The market beta of most funds is close to one. The mean beta is 1.00 and the beta of three-quarters of all funds is between

---

<sup>6</sup>The quality index combines the return-to-equity, debt-to-equity, and earnings variability.

<sup>7</sup>Choosing an earlier starting date drastically reduces the number of funds.

0.95 and 1.10. Consistent with the literature on fund performance, the majority of funds in the sample underperform. The mean and median CAPM alphas are negative (-1.59% and -1.71, respectively), and only a quarter of the funds generate positive alphas. Only 10 alphas are statistically significantly positive at the 5% level. The lower panel shows the distribution of returns by fund category. Since “growth” stocks outperform “value” stocks over the sample period, the returns of “growth” funds are higher than those of “value” funds. Sector funds performed poorly on average but there is a large dispersion of mean returns across funds.

PCA/factor models exploit the dependence structure of the data. In 2-dimensional data sets, this is straightforward since the covariance matrix captures all relevant information and determines the principal components and implied factors. The dependence structure in higher dimensional data is more complex. The mutual fund data set has three dimensions: time, funds, and characteristics, and its dependence structure is in general multi-dimensional. For example, a characteristic of a given fund may be correlated with other characteristics of the fund and may also be correlated with characteristics of other funds. In addition, characteristics are correlated across time.

Figure 1 shows heatmaps of time series (upper triangle) and cross-sectional (lower triangle) correlations of characteristics. To obtain time series correlations, I first compute (time series) correlations of characteristics for each mutual fund and then take means across all funds. Cross-sectional correlations are obtained by computing (cross-sectional) correlations of characteristics for each quarter and averaging across all quarters. Comparing the two correlation measures shows that the overall patterns are similar, but cross-sectional correlations in the lower-left triangle are on average larger (in absolute value) than time series correlations in the upper-right triangle. Not surprisingly, price-ratio characteristics are positively correlated, as are characteristics related to the growth of fundamentals, but the two blocks are negatively correlated. Since the Morningstar variables MS, MULT, and GR are based on price ratios and growth rates, their correlation pattern is to a large degree mechanical. Investment, momentum, and reversal are negatively related to price ratios but positively related to growth rates, and size is positively correlated with higher liquidity.

Recall that Figure 1 is based on the means of the correlation distribution and therefore cannot capture more complex relationships in the 3-dimensional data set. For example, time series correlations across characteristics differ by mutual fund and cross-sectional correlations may vary over time. Figure E.1 shows time series and cross-sectional correlations of four characteristic pairs across funds and time, respectively. While cross-sectional correlations are relatively stable over time (with the exceptions of correlations involving momentum and reversal), there are large variations of time series correlations across funds. For example, the book-to-market and earnings-to-price ratios are positively correlated on average (the mean and median are 0.38 and 0.46, respectively). However, there is considerable variation in the (BM, EP)-correlation across funds. The interquartile range is (0.14, 0.66) and BM and EP are *negatively* correlated for 16% of all funds. Such a large dispersion of pairwise time series correlations is typical for characteristic pairs.

The discussion above described the correlation patterns across characteristics. The same analysis can be repeated for the other two dimensions of time and mutual funds. The goal of high-dimensional factor models is to capture the complex dependence structure in the data in a parsimonious model

without having to collapse the data to two dimensions and loss of information.

### 3. High-dimensional factor models

Traditional factor models used in finance and economics are based on 2-dimensional data sets, *i.e.*, the data can be represented by a matrix. A canonical example in asset pricing is the factor analysis of a panel of returns of  $N$  assets observed over  $T$  periods. Latent factors can be constructed by PCA, which is based on the eigenvalue/eigenvector decomposition of a second-moment matrix of returns, or, equivalently, by the SVD of the data matrix. The vast literature on factor models has suggested many extensions to the standard model but has been limited to 2-dimensional data. In this section, I consider generalizations of factor models to situations when the data set has more than two dimensions. The data set used in the empirical section below is 3-dimensional and composed of observations of characteristic  $c$  of asset  $n$  in period  $t$ , and I use this example to illustrate the theoretical results in this section.

Higher-dimensional data form *tensors*, which were first defined by Ricci-Curbastro and Levi-Civita (1900). Tensors generalize the notions of vectors and matrices to more than two dimensions. Many tensor operations are straightforward extensions of matrix algebra, but there are some important differences and the notation is necessarily more complex. This section defines tensors and summarizes tensor operations used in the rest of the paper. I start with a brief summary of 2-dimensional factor models, PCA, and SVD to facilitate a better understanding of the extensions to higher dimensions.

The tensor models used in this paper can be interpreted as extensions of the SVD of a matrix. Similar to SVD and PCA, the goal is to summarize the variation in the data efficiently by expressing the data tensor in terms of lower-dimensional tensors and matrices. In this sense, SVD/PCA and tensor decompositions can be thought of as dimension reduction methods. I also show that the decomposition of a 3-dimensional tensor implies a collection of 2-dimensional factor models that are linearly connected across all three dimensions and internally consistent. As with any latent factor method, it is important to pay attention to the economic meaning of the model. It turns out that the different components of tensor decompositions have clear economic interpretations, as I explain below.

#### 3.1. From matrices to tensors

As mentioned above, tensors extend the notions of vectors and matrices into higher dimensions. This section presents a brief introduction to tensor algebra and is limited to operations used in the rest of the paper. See Kolda and Bader (2009) for a concise summary and Kroonenberg (2007) for a more comprehensive treatment of tensor algebra and decompositions.

Throughout the paper, I use the following notation:

$$\begin{aligned} \text{scalar: } & \mathbf{x} \in \mathbb{R} \\ \text{vector: } & \mathbf{x} \in \mathbb{R}^I \\ \text{matrix: } & \mathbf{X} \in \mathbb{R}^{I_1 \times \mathbb{R}^{I_2}} \\ j\text{-th order tensor: } & \mathcal{X} \in \mathbb{R}^{I_1 \times \mathbb{R}^{I_2} \times \dots \times \mathbb{R}^{I_j}. \end{aligned}$$

Hence, a zero-order tensor is a scalar, a first-order tensor is a vector, a second-order tensor is a matrix, and a third-order tensor is a cuboid. Each of the  $j$  dimensions of a tensor is called a *mode*. A tensor  $\mathcal{X}$  is *diagonal* if  $x_{i_1, \dots, i_j} \neq 0$  only if  $i_1 = \dots = i_j$  and 0 otherwise.

The data set that will be used later has three dimensions: the characteristic  $c$  of asset  $n$  at date  $t$ ,  $\mathbf{x}_{tnc}$ . To simplify the notation, I will therefore focus on tensors of order  $n = 3$  but all results can be easily generalized to higher dimensions. Let  $\mathcal{X} \in \mathbb{R}^T \times \mathbb{R}^N \times \mathbb{R}^C$  be a 3-dimensional ( $T \times N \times C$ ) tensor  $\mathcal{X}$  with *elements*  $x_{tnc}$ .<sup>8</sup> Thus, the dimension of the tensor that represents the asset data set is  $34 \times 934 \times 25$ .

A 3-dimensional tensor can be expressed as collections of one-dimensional *fibers* or 2-dimensional *slices*. Fibers are vectors and correspond to rows and columns of a matrix, while slices are matrices. Fibers are defined by fixing every index but one so that  $\mathcal{X}$  has fibers along each mode, denoted by  $\mathbf{x}_{(nc)t}$ ,  $\mathbf{x}_{(tc)n}$ , and  $\mathbf{x}_{(tn)c}$ , respectively.<sup>9</sup> Slices are created by fixing all but two indices and are written as  $\mathbf{X}_{(t)nc}$ ,  $\mathbf{X}_{(n)tc}$ ,  $\mathbf{X}_{(c)tn}$ .<sup>10</sup> Tensors can be written as matrices by *unfolding*, or *stacking*, 2-dimensional slices along a mode  $n$ . The resulting matrix  $\mathbf{X}_{(i)}$  is defined so that the number of rows equals the mode- $n$  order of  $\mathcal{X}$ . The number of columns of  $\mathbf{X}_{(i)}$  is equal to the product of the dimensions along all other modes.<sup>11</sup>

### 3.2. Tensor factor models

Before introducing tensor models, consider first the familiar case of 2-dimensional factor models. Further detail are in Appendix B. Let  $\mathbf{X}$  be a ( $T \times N$ ) data matrix with  $TN$  observations  $\mathbf{x}_{tn}$ . A  $K$ -factor model is defined as

$$\mathbf{X} = \mathbf{F}_K \mathbf{B}_K^\top + \mathbf{E}_K, \quad (1)$$

where  $\mathbf{F}_K$  and  $\mathbf{B}_K$  are ( $T \times K$ ) and ( $N \times K$ ) factor and loading matrices, that can be estimated using Principal Component Analysis (PCA). Let  $\hat{\mathbf{X}}_K = \mathbf{F}_K \mathbf{B}_K^\top$  be the matrix of “fitted” values. PCA factors and loadings can be obtained from the truncated SVD of  $\mathbf{X}$ :

$$\hat{\mathbf{X}}_K = \mathbf{U}_K^{(1)} \mathbf{H}_K \mathbf{U}_K^{(2)\top} \quad (2)$$

$$= \sum_{k=1}^K h_{kk} \mathbf{u}_k^{(1)} \mathbf{u}_k^{(2)\top} \quad (3)$$

$$= \sum_{k=1}^K h_{kk} \mathbf{u}_k^{(1)} \circ \mathbf{u}_k^{(2)} \equiv \mathbf{F}_{TN}(K), \quad (4)$$

where  $\mathbf{H}_K$  is the ( $K \times K$ ) matrix with the square roots of the  $K$  largest eigenvalues,  $h_k$ , of  $\mathbf{X}^\top \mathbf{X}$  on the diagonal.<sup>12</sup>  $\mathbf{U}_K^{(1)}$  and  $\mathbf{U}_K^{(2)}$  are ( $T \times K$ ) and ( $N \times K$ ) matrices whose columns are the eigenvectors  $\mathbf{u}_k^{(1)}$  and  $\mathbf{u}_k^{(2)}$  of  $\mathbf{X}\mathbf{X}^\top$  and  $\mathbf{X}^\top \mathbf{X}$  that are associated with the  $K$  largest eigenvalues. Factors and loadings are

<sup>8</sup>Figures E.2 to E.4 show a third-order tensor with dimensions  $T = 5, N = 4, C = 3$  and illustrate the tensor operations described in this section.

<sup>9</sup>See Panels B, C, and D of Figure E.2.

<sup>10</sup>See Panels E, F, and G of Figure E.2.

<sup>11</sup>Figure E.3 shows the unfolding of a ( $5 \times 4 \times 3$ ) tensor  $\mathcal{X}$  along each mode. The resulting matrix of unfolding  $\mathcal{X}$  along mode-1,  $\mathbf{X}_{(1)}$ , has five rows and  $4 \cdot 3 = 12$  columns. Unfolding along modes two and three yields matrices  $\mathbf{X}_{(2)}$  and  $\mathbf{X}_{(3)}$  with dimensions ( $4 \times 15$ ) and ( $3 \times 20$ ), respectively.

<sup>12</sup>The *outer product*  $\circ$  of two vectors  $\mathbf{a} \in \mathbb{R}^T$  and  $\mathbf{b} \in \mathbb{R}^N$  is defined as  $\mathbf{a} \circ \mathbf{b} = \mathbf{a}\mathbf{b}^\top \in \mathbb{R}^T \times \mathbb{R}^N$ .

given by  $\mathbf{F}_K = \mathbf{U}^{(1)}\mathbf{H} = \mathbf{X}\mathbf{U}^{(2)}$  and  $\mathbf{B}_K = \mathbf{U}^{(2)}$ . I will refer to the  $K$ -factor model for the  $(T \times N)$  matrix  $\mathbf{X}$  as  $\mathbf{F}_{TN}(K)$ .<sup>13</sup>

The representations (3) and (4) show that  $\hat{\mathbf{X}}_k$  is the weighted sum of  $K$  matrices with dimensions  $(T \times N)$ , which are the outer vector product of the eigenvectors  $\mathbf{u}_k^{(1)}$  and  $\mathbf{u}_k^{(2)\top}$ . Each  $k$  in the summation represents a factor in the  $K$ -factor representation (1). The advantage of representation (3) is that it shows the contribution of each of the  $K$  factors in the fit of the model. Since the eigenvectors are normalized, the  $K$  outer vector products  $\mathbf{u}_k^{(1)}\mathbf{u}_k^{(2)\top}$  are of the same magnitude, so the weight of the contribution of each factor  $k$  is approximately equal to the  $k$ -th eigenvalue.

Next, consider a 3-dimensional tensor  $\mathcal{X}$  with dimensions  $(T \times N \times C)$ . One possible way to investigate the factor structure of  $\mathcal{X}$  is to estimate 2-dimensional factor models after collapsing one of the three dimensions. For example, fix  $t$  and estimate a 2-dimensional  $K_T$ -factor model  $\mathbf{F}_{NC}(K_T)$  for the  $t$ -th slice  $\mathbf{X}_{(t)nc}$  of  $\mathcal{X}$ .  $\mathbf{X}_{(t)nc}$  is a  $(N \times C)$  matrix, so the model represents a cross-sectional factor model for  $C$  characteristics of  $N$  assets. Alternatively, for a given asset  $n$ , one can estimate a  $K_N$ -factor model  $\mathbf{F}_{TC}(K_N)$  for the  $n$ -th slice  $\mathbf{X}_{(n)}$ , which forms a  $(T \times C)$  matrix. This model captures the time series correlations across  $C$  characteristics of asset  $n$ . Finally, fix characteristic  $c$  and consider the  $K_C$ -factor model  $\mathbf{F}_{TN}(K_C)$  for the  $c$ -th slice of  $\mathcal{X}$ . This factor model captures time series correlations across  $N$  assets for characteristic  $c$ . In principle, one could estimate 2-dimensional factor models for each  $t = 1, \dots, T$ ,  $n = 1, \dots, N$ , and  $c = 1, \dots, C$  and obtain  $T + N + C$  models. Note that estimating the models separately implies that information is lost. For example, estimating factor models for each  $t$  does not exploit potentially useful time series information.

In contrast, tensor factor models (TFM) are estimated in a single joint step that exploits all three (or more) dimensions simultaneously. The Tucker factor model (Tucker (1966)) extends the matrix SVD to higher-dimensional tensors. For the 3-dimensional case of asset characteristics, the 3-dimensional representation of the Tucker model implies 2-dimensional factor models for each dimension that have the same structure as the factor models described in the previous paragraph. In other words, the Tucker model implies a 2-dimensional factor model for each  $t = 1, \dots, T$ , each  $c = 1, \dots, C$ , and each  $n = 1, \dots, N$ . Instead of having to estimate  $T + N + C$  separate 2-dimensional factor models, it is possible to estimate a *single* 3-dimensional factor model and derive the  $T + N + C$  implied 2-dimensional factor models. Since all 2-dimensional factor models stem from the same 3-dimensional representation, they are mutually consistent. In addition, the 3-dimensional Tucker model exploits dependence in all dimensions without having to collapse any one dimension and the corresponding loss of information. Finally, the Tucker model has an order of magnitude fewer free parameters than  $T + N + C$  2-dimensional factor models.

The Tucker factor model is based on a generalization of the SVD decomposition and PCA for matrices to tensors. The Tucker model is usually written in tensor notation, see Appendix A for more details. Briefly, the  $n$ -mode product of a tensor  $\mathcal{X}$  and a matrix  $\mathbf{A}_n$  is the multiplication of each  $n$ -mode

---

<sup>13</sup>Note that we could define the truncated SVD using different numbers of factors,  $K_1$  and  $K_2$ , for the two dimensions. However, since  $\mathbf{H}$  is diagonal, this SVD reduces to a  $K$ -factor SVD where  $K = \min(K_1, K_2)$ . In contrast, the equivalent object to  $\mathbf{H}$  in the tensor factor model considered below is *not* diagonal so the number of factors can differ by dimension. See Appendix B for more details.

fiber of  $\mathbf{x}$  by the row vectors of  $\mathbf{A}_n$ . For example, the mode-1 product of a  $(S \times N \times C)$  tensor  $\mathbf{x}$  and a  $(T \times S)$  matrix  $\mathbf{A}_1$  is equal to a  $(T \times N \times C)$  tensor  $\mathbf{y}$  given by  $\mathbf{y} = \mathbf{x} \times_1 \mathbf{A}_1$  (see Figure E.4). Note that the standard matrix product can be written in tensor notation:  $\mathbf{A}_1 \mathbf{x} \mathbf{A}_2^\top = \mathbf{x} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2$ .

Let  $\mathbf{x}$  be a data tensor with dimensions  $(T \times N \times C)$ . The Tucker approximation  $\hat{\mathbf{x}}$  of order  $(K_T, K_N, K_C)$  is given by

$$\hat{\mathbf{x}}(K_T, K_N, K_C) = \mathcal{G} \times_1 \mathbf{V}_{K_T}^{(T)} \times_2 \mathbf{V}_{K_N}^{(N)} \times_3 \mathbf{V}_{K_C}^{(C)}, \quad (5)$$

where  $\mathcal{G}$  is a  $(K_T \times K_N \times K_C)$  tensor with elements  $g_{k_T k_N k_C}$  and  $\mathbf{V}_{K_T}^{(T)}, \mathbf{V}_{K_N}^{(N)}, \mathbf{V}_{K_C}^{(C)}$  are  $(T \times K_T), (N \times K_N), (C \times K_C)$  matrices, respectively. Given the definition of the  $n$ -mode tensor product,  $\hat{\mathbf{x}}$  is a  $(T \times N \times C)$ -dimensional tensor and thus has the same dimensionality as the data tensor  $\mathbf{x}$ .  $\mathcal{G}$  is called the *core* tensor and can be thought of as a ‘‘compressed’’ version of  $\mathbf{x}$ . As we will see below, the matrices  $\mathbf{V}_{k_i}^{(i)}$  correspond to the loadings matrix  $\mathbf{B}_k$  in the 2-dimensional factor model (1) and I will refer to them as loadings matrices of the Tucker model. To simplify the notation, I omit the subscripts of the loading matrices henceforth. The approximation error is  $\mathcal{E} = \mathbf{x} - \hat{\mathbf{x}}$ . The optimal Tucker model minimizes the mean-squared error (MSE) of  $\mathcal{E}$  among all  $\hat{\mathbf{x}}(K_T, K_N, K_C)$  of the form (5).

The mechanism of the Tucker decomposition (5) is illustrated in Figure 3, which shows the decomposition of a  $(6 \times 5 \times 4)$  tensor  $\mathbf{x}$  by a Tucker model of order  $(K_T, K_N, K_C) = (3, 2, 2)$ . The core tensor  $\mathcal{G}$  compresses  $\mathbf{x}$  to a lower-dimension of  $(3 \times 2 \times 2)$ . The Tucker loading matrices  $\mathbf{V}^{(T)}, \mathbf{V}^{(N)}$ , and  $\mathbf{V}^{(C)}$  expand the core tensor to the full dimension of  $\mathbf{x}$  and have the matching dimensions of  $(6 \times 3), (5 \times 2)$ , and  $(4 \times 2)$ . With slight abuse of notation, the dimensions of the tensors and matrices can be expressed as  $(3 \times 2 \times 2) \times_1 (6 \times 3) \times_2 (5 \times 2) \times_3 (4 \times 2) = (6 \times 5 \times 4)$ .

The 2-dimensional  $K$ -factor SVD-PCA model (2) is a special case of the Tucker model when  $\hat{\mathbf{x}}$  is a matrix and  $K_i = K$ . This can be seen by rewriting (2) in tensor notation:

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{U}_K^{(1)} \mathbf{H}_K \mathbf{U}_K^{(2)\top} \\ &= \mathbf{H}_K \times_1 \mathbf{U}_K^{(1)} \times_2 \mathbf{U}_K^{(2)}. \end{aligned} \quad (6)$$

Thus the core tensor  $\mathcal{G}$  in (5) corresponds to  $\mathbf{H}_K$  and the matrices  $\mathbf{V}^{(i)}$  correspond to  $\mathbf{U}_K^{(j)}$ .

Recall that the  $K$ -factor SVD-PCA model can be written as the sum of  $K$  outer products of the column vectors of  $\mathbf{U}_K^{(1)}$  and  $\mathbf{U}_K^{(2)}$ , see (4). The Tucker model (5) has an analogous representation in terms of outer products of the column vectors of the loading matrices  $\mathbf{V}^{(T)}, \mathbf{V}^{(N)}, \mathbf{V}^{(C)}$ :

$$\hat{\mathbf{x}}(K_T, K_N, K_C) = \sum_{k_T=1}^{K_T} \sum_{k_N=1}^{K_N} \sum_{k_C=1}^{K_C} g_{k_T k_N k_C} \mathbf{v}_{k_T}^{(T)} \circ \mathbf{v}_{k_N}^{(N)} \circ \mathbf{v}_{k_C}^{(C)}, \quad (7)$$

where  $\mathbf{v}_{k_i}^{(i)}$  are columns of  $\mathbf{V}^{(i)}$ . The intuition of the Tucker model is similar to that of the SVD decomposition. Each outer product  $\mathbf{v}_{k_T}^{(T)} \circ \mathbf{v}_{k_N}^{(N)} \circ \mathbf{v}_{k_C}^{(C)}$  forms a  $(T \times N \times C)$  tensor that has the interpretation of a tensor factor. Hence,  $\hat{\mathbf{x}}(K_T, K_N, K_C)$  is the sum of  $K_T K_N K_C$  tensor factors. Each factor is weighted by the corresponding element of the core tensor  $g_{k_T k_N k_C}$ . Factors with larger  $g_{k_T k_N k_C}$  have higher weights in the Tucker model than factors with low  $g_{k_T k_N k_C}$ .

As we will see below, the Tucker model shares some properties with the SVD-PCA model for matrices. However, there are some important differences. First, the Tucker model allows for a different

number of factors in each dimension while the SVD-PCA specifies a single number of factors (i.e.  $\mathbf{H}_K$  is a  $K \times K$  matrix). Note that, in principle, the decomposition  $\mathbf{U}_K^{(1)} \mathbf{H}_K \mathbf{U}_K^{(2)\top}$  could be specified so that  $\mathbf{U}_K^{(1)}$  and  $\mathbf{U}_K^{(2)}$  have different numbers of columns (with appropriate dimensions for  $\mathbf{H}_K$ ). However, in the 2-dimensional SVD, the off-diagonal elements of  $\mathbf{H}_K$  are zero so that in effect the number of columns of  $\mathbf{U}_K^{(1)}$  and  $\mathbf{U}_K^{(2)}$  are the same. In contrast, the Tucker core tensor  $\mathcal{G}$  and loading matrices  $\mathbf{V}_{(i)}$  are not tied to eigenvalues and eigenvectors. In general,  $\mathcal{G}$  is not diagonal, so the number of factors may differ across dimensions.<sup>14</sup> Second, there is no closed-form solution for the optimal Tucker model, so the model has to be solved numerically. I discuss numerical solutions in section 3.6.

Neither the SVD nor the Tucker model are unique and can be rotated. Let  $\mathbf{S}_i, i = T, N, C$ , be nonsingular ( $K_i \times K_i$ ) matrices. Then (5) can be written as

$$\hat{\mathbf{x}}(K_T, K_N, K_C) = (\mathcal{G} \times_1 \mathbf{S}_T \times_2 \mathbf{S}_M \times_3 \mathbf{S}_C) \times_1 (\mathbf{V}^{(T)} \mathbf{S}_T^{-1}) \times_2 (\mathbf{V}^{(N)} \mathbf{S}_M^{-1}) (\mathbf{V}^{(C)} \mathbf{S}_C^{-1}). \quad (8)$$

Typically, (5) is normalized so that the  $\mathbf{V}^{(i)}$  matrices are orthonormal, similar to the eigenvector matrices in the SVD-PCA model.

One important property of Tucker decompositions is that they cannot be computed sequentially. Consider two Tucker decompositions with  $(K_T, K_N, K_C)$  and  $(K'_T, K'_N, K'_C), K'_j < K_j$ , respectively. The first  $K'_T, K'_N, K'_C$  components of the Tucker  $(K_T, K_N, K_C)$  model are in general different from the Tucker  $(K'_T, K'_N, K'_C)$  model.<sup>15</sup> In contrast, the first  $K$  factors of a 2-dimensional SVD-PCA factor model with  $K' > K$  factors are the same as those of a  $K$ -factor model since they are based on eigenvalues and eigenvectors.

### 3.3. Factor representation

The Tucker model can be written in factor form similar to (1) along each dimension.<sup>16</sup> Consider first the representation for the characteristic dimension implied by (5):

$$\mathbf{x} = \mathcal{F}_{tn}^{(C)} \times_3 \mathbf{V}^{(C)} + \boldsymbol{\varepsilon} \quad (9)$$

$$\text{where } \mathcal{F}_{tn}^{(C)} = \mathcal{G} \times_1 \mathbf{V}^{(T)} \times_2 \mathbf{V}^{(N)}. \quad (10)$$

$\mathcal{F}_{tn}^{(C)}$  is a  $(T \times N \times K_C)$ -dimensional ‘‘factor’’ tensor and  $\mathbf{V}^{(C)}$  is the  $(C \times K_C)$ -dimensional loadings matrix. The interpretation of  $\mathcal{F}_{tn}^{(C)}$  is similar to that of factor matrices in 2-dimensional SVD-PCA models. For a given asset  $n$ , the  $n$ th slice of  $\mathcal{F}_{tn}^{(C)}, \mathbf{F}_{t(n)}^{(C)}$ , is a  $(T \times K_C)$  matrix whose columns are the time series of  $K_C$  characteristic factors of asset  $n$ . Recall that the  $n$ th slice of  $\mathbf{X}, \mathbf{X}_{(n)tc}$ , is a  $(T \times C)$  matrix whose columns are the time series of all  $C$  characteristics. In other words, the  $K_C$  factors given by  $\mathbf{F}_{t(n)}^{(C)}$  summarize the

<sup>14</sup>The CP tensor decomposition (Carroll and Chang (1970)), Harshman (1970)) is a special case of the Tucker decomposition where the core tensor is restricted to be diagonal. The CP decomposition restricts the number of factors to be identical for all dimensions, which is a drawback when the dimensions of the data tensor differ substantially. In addition, in contrast to the Tucker decomposition, the CP decomposition might not exist. However, if it exists, the CP decomposition is unique. See Kolda and Bader (2009) for more details and Babii et al. (2022) for a recent application.

<sup>15</sup>In practice, the differences are small.

<sup>16</sup>Note that there are two isomorphic factor representations in the 2-dimensional case, one for  $\mathbf{X}$  and one for  $\mathbf{X}^\top$  in which the roles of  $\mathbf{U}_K^{(1)}$  and  $\mathbf{U}_K^{(2)}$  are reversed. The factors and loadings in the representation for  $\mathbf{X}$  are given by  $\mathbf{U}_K^{(1)}$  and  $\mathbf{U}_K^{(2)}$ , respectively, while  $\mathbf{U}_K^{(2)}$  forms factors and  $\mathbf{U}_K^{(1)}$  forms loadings in the  $\mathbf{X}^\top$  representation.

information in all  $C$  characteristics of asset  $n$ . The implied factor model for asset  $n$  is given by

$$\mathbf{X}_{(n)tc} = \mathbf{F}_{t(n)}^{(C)} \times_3 \mathbf{V}^{(C)} + \mathbf{E}_{(n)tc} = \mathbf{F}_{t(n)}^{(C)} \mathbf{V}^{(C)\top} + \mathbf{E}_{(n)tc}. \quad (11)$$

Thus, (9) can be understood as a collection of  $N$  2-dimensional  $K_C$ -factor models for characteristics of individual assets  $n = 1, \dots, N$ . The estimation of the 3-dimensional Tucker model implicitly obtains the  $N$  factor models jointly and exploits information across assets. The  $K_C$  characteristic factors encoded in  $\mathcal{F}_{tn}^{(C)}$  form the basis for the pricing factors that will be the focus of section 5.

The factor representations for the other dimensions are defined accordingly:

$$\mathbf{x} = \mathcal{F}_{tc}^{(N)} \times_3 \mathbf{V}^{(N)} + \boldsymbol{\varepsilon} \quad (12)$$

$$\mathbf{x} = \mathcal{F}_{nc}^{(T)} \times_3 \mathbf{V}^{(T)} + \boldsymbol{\varepsilon}. \quad (13)$$

The  $(T \times K_N \times C)$ -dimensional factor tensor  $\mathcal{F}_{tc}^{(N)}$  summarize the information in all  $N$  assets into  $K_N$  factors and the  $(K_T \times N \times C)$ -dimensional  $\mathcal{F}_{nc}^{(T)}$  is composed of  $K_T$  factors that summarize the information in  $T$  time periods.

#### 3.4. The Tucker model implies interconnected 2-dimensional factor models

As discussed in section 3.2, the econometrician could estimate separate 2-dimensional factor models for each  $t = 1, \dots, T$ , each  $n = 1, \dots, N$ , and each  $c = 1, \dots, C$ . Next, I show that the Tucker model (5) implies 2-dimensional factor models that are *interconnected* and subject to restrictions that are imposed by the 3-dimensional representation of the Tucker model. These restrictions would be violated if the 2-dimensional factor models were estimated separately. The factor models are formed by the columns of the loading matrices  $\mathbf{V}^{(i)}$  and weighted by the elements of the core tensor  $\mathcal{G}$ . Recall that a 2-dimensional  $K$ -factor model  $\mathbf{F}(K)$  can be written as the weighted sum of  $K$  outer products of vectors, see (4). The corresponding representation (7) expresses the Tucker model as the sum of outer products of the column vectors of the loadings matrices and can be rewritten in terms of  $K_T$  2-dimensional factor models as follows:

$$\hat{\boldsymbol{\chi}}(K_T, K_N, K_C) = \sum_{k_T=1}^{K_T} \mathbf{v}_{k_T}^{(T)} \circ \left[ \sum_{k_N=1}^{K_N} \mathbf{v}_{k_N}^{(N)} \circ \left( \sum_{k_C=1}^{K_C} \mathcal{G}_{k_T k_N k_C} \mathbf{v}_{k_C}^{(C)} \right) \right] \quad (14)$$

$$= \sum_{k_T=1}^{K_T} \mathbf{v}_{k_T}^{(T)} \circ \left[ \sum_{k_N=1}^{K_N} \mathbf{v}_{k_N}^{(N)} \circ \tilde{\mathbf{v}}_{k_T}^{(C)} \right] \quad (15)$$

$$= \sum_{k_T=1}^{K_T} \mathbf{v}_{k_T}^{(T)} \circ \mathbf{F}_{NC}^{k_T}(K_N) \quad (16)$$

where  $\tilde{\mathbf{v}}_{k_T}^{(C)} = \sum_{k_C=1}^{K_C} \mathcal{G}_{k_T k_N k_C} \mathbf{v}_{k_C}^{(C)} \cdot \mathbf{v}_{k_N}^{(N)}$  and  $\tilde{\mathbf{v}}_{k_C}^{(C)}$  are  $(K_N \times 1)$  and  $(K_C \times 1)$  vectors, so their outer product is a  $(N \times C)$  matrix. Thus, the terms in square brackets in (14) and (15) represent a 2-dimensional  $K_N$ -factor model  $\mathbf{F}_{NC}^{k_T}(K_N)$  for the  $N$  and  $C$  dimensions of  $\boldsymbol{\chi}$ . The factor model is formed by the columns of loading matrices,  $\mathbf{v}_{k_N}^{(N)}$  and  $\mathbf{v}_{k_C}^{(C)}$ , of the Tucker model (5). Note that for a given  $k_T$ , the vector  $\tilde{\mathbf{v}}_{k_T}^{(C)}$  in square brackets of (15) is the same for all  $k_N$ , so that  $\mathbf{F}_{NC}^{k_T}(K_N)$  is a *restricted* factor model. The last equality shows that the Tucker model can be written in terms of  $K_T$  2-dimensional (restricted)  $(N \times C)$  factor models.  $\mathbf{v}_{k_T}^{(T)}$  is a  $(T \times 1)$  vector and  $\mathbf{F}_{NC}^{k_T}(K_N)$  is a  $(N \times C)$  matrix so that their outer product is a

$(T \times N \times C)$ -dimensional tensor and has the same dimensions as  $\hat{\mathbf{x}}$ .

The order of the summation in the term in square brackets in (14) can be reversed so that the sum in square brackets in (15) is over  $k_C$  instead of  $k_N$ . The resulting factor model is equivalent to (16) but has  $K_C$  factors. I will choose the representation that has the fewest factors and write the resulting factor model as  $\mathbf{F}_{NC}^{k_T}(\tilde{K}_{NC})$ , where  $\tilde{K}_{NC} = \min(K_N, K_C)$ .

Next, consider estimating a 2-dimensional cross-sectional factor model for period  $t$  by computing the SVD of the  $(N \times C)$  matrix given by the  $t$ -th slice  $\mathbf{X}_{(t)nc}$ . The corresponding factor model implied by the Tucker model is given by multiplying  $\mathbf{F}_{NC}^1(K_N), \dots, \mathbf{F}_{NC}^{K_T}(K_N)$  by the  $t$ -th elements of the vectors  $\mathbf{v}_{k_T}^{(T)}$  in (16):

$$\hat{\mathbf{X}}_{(t)nc} = \sum_{k_T=1}^{K_T} \mathbf{v}_{k_T,t}^{(T)} \mathbf{F}_{NC}^{k_T}(K_N) \equiv \mathbf{F}_{NC}^{(t)}(K_N), \quad (17)$$

where  $\mathbf{v}_{k_T,t}^{(T)}$  is the  $t$ -th element of the  $k_T$ -th column of the loading matrix  $\mathbf{V}^{(T)}$ . (17) shows that the implied 2-dimensional factor model for period  $t$  is given by a weighted sum of  $K_T$  factor models  $\mathbf{F}_{NC}^{k_T}(\tilde{K}_{NC}), k_T = 1, \dots, K_T$  and is therefore also a  $\tilde{K}_{NC}$ -factor model, defined as  $\mathbf{F}_{NC}^t(\tilde{K}_{NC})$ . The weights are given by the rows of  $\mathbf{V}^{(T)}$ . Since the weights vary across periods  $t$ , the factor models  $\mathbf{F}_{NC}^t(\tilde{K}_{NC})$  change over time. However, the time- $t$  factors are based on the same  $\mathbf{F}_{NC}^{k_T}(\tilde{K}_{NC})$  and differ only by time-varying weights.

An interesting special case is when there is only a single  $T$ -factor, i.e.  $K_T = 1$ . In this case,

$$\mathbf{F}_{NC}^{(t)}(\tilde{K}_{NC}) = \mathbf{v}_{1,t}^{(T)} \mathbf{F}_{NC}^1(\tilde{K}_{NC}), \quad (18)$$

which implies that all  $t$ -slices are proportional to the same factor model  $\mathbf{F}_{NC}^1(\tilde{K}_{NC})$ , or, equivalently, proportional to the same  $(N \times C)$  matrix given by the term in square brackets in (15) that forms  $\mathbf{F}_{NC}^1(\tilde{K}_{NC})$ . The proportionality of  $t$ -slices in turn implies that all  $t$ -fibers are perfectly correlated. Recall that the  $t$ -fibers in the asset application are given by time series of mutual/asset characteristic pairs. In the special case of  $K_T = 1$ , the Tucker model implies that all  $NC$  time series are perfectly correlated. Therefore, the behavior of time series in Tucker models with a single  $T$ -factor is severely restricted. The Tucker model shares this property with SVD-PCA models since the columns of the matrix that is given by a 1-factor SVD-PCA model are also proportional to each other. Adding  $T$ -factors enriches the dynamics across time series fibers. For example,  $K_T = 2$  implies

$$\mathbf{F}_{NC}^{(t)}(K_N) = \mathbf{v}_{1,t}^{(T)} \mathbf{F}_{NC}^1(\tilde{K}_{NC}) + \mathbf{v}_{2,t}^{(T)} \mathbf{F}_{NC}^2(K_N). \quad (19)$$

Time slices are given by weighted sums of two  $(N \times C)$  factor models and are thus not proportional. However, unless the weights  $\mathbf{v}_{1,t}^{(T)}$  and  $\mathbf{v}_{2,t}^{(T)}$  differ significantly,  $t$ -slices and time series fibers will be correlated. Adding further  $T$ -factors creates scope for more complex dependence structures across  $t$ -slices and time series fibers.

Although the derivations above focus on the properties along the  $T$ -dimension of  $\mathbf{x}$ , the results

apply to the other dimensions as well and can be summarized as follows:

$$\hat{\boldsymbol{\mathcal{X}}}(K_T, K_N, K_C) = \sum_{k_T=1}^{K_T} \mathbf{v}_{k_T}^{(T)} \circ \mathbf{F}_{NC}^{k_T}(\tilde{\mathbf{K}}_{NC}) \quad (20)$$

$$= \sum_{k_N=1}^{K_N} \mathbf{v}_{k_N}^{(N)} \circ \mathbf{F}_{TC}^{k_N}(\tilde{\mathbf{K}}_{TC}) \quad (21)$$

$$= \sum_{k_C=1}^{K_C} \mathbf{v}_{k_C}^{(C)} \circ \mathbf{F}_{TN}^{k_C}(\tilde{\mathbf{K}}_{TN}) \quad (22)$$

$$\hat{\boldsymbol{\mathcal{X}}}_{(t)nc} = \mathbf{F}_{NC}^{(t)}(\tilde{\mathbf{K}}_{NC}) \quad (23)$$

$$\hat{\boldsymbol{\mathcal{X}}}_{(n)tc} = \mathbf{F}_{TC}^{(n)}(\tilde{\mathbf{K}}_{TC}) \quad (24)$$

$$\hat{\boldsymbol{\mathcal{X}}}_{(c)tn} = \mathbf{F}_{TN}^{(c)}(\tilde{\mathbf{K}}_{TN}). \quad (25)$$

The first three equations show that the Tucker model can be written as a set of 2-dimensional factor models for each dimension. The factor models are based on the columns of the loadings matrices  $\mathbf{V}^{(i)}$  and weighted by the elements of the core tensor  $\boldsymbol{\mathcal{G}}$ , and are thus connected. The Tucker model imposes the restriction that the vectors underlying the factor models  $\mathbf{F}$  are identical in one of the two dimensions. The last three equations show that slices of  $\hat{\boldsymbol{\mathcal{X}}}$  form 2-dimensional factor models. The factor models in a dimension  $n$  are connected since they are given by weighted sums of  $K_i$  factor models. If  $K_i = 1$ , the slices and fibers in dimension  $n$  are perfectly correlated.

Note that “time” does not play a special role in factor modeling. If one of the dimensions of the data set is “time”, it is not treated differently from the other dimensions. In the setting of this paper, it is important to keep this in mind when interpreting “time” factors. As we have seen above, “time” factors do not capture serial correlation but instead capture cross-sectional dependence of the time series that make up the data. If the time series in  $\boldsymbol{\mathcal{X}}$  are relatively highly correlated, relatively few “time” factors are sufficient to capture the correlations across time series. If correlations across time series are low or vary substantially, more “time” factors are needed to capture the overall dynamics of the data. Of course, in an application, the logic is reversed. If correlations across times series are high (low), the econometrician will find that few (many)  $t$ -factors are required, so that  $K_T$  is small (high).

The same intuition holds for factors in the other dimensions. If there is a single asset factor,  $K_N = 1$ , the Tucker model implies that all  $n$ -slices  $\hat{\boldsymbol{\mathcal{X}}}_{(n)tc}$ , which are matrices of dimension  $(T \times C)$  are proportional and all  $n$ -fibers  $\hat{\boldsymbol{\mathcal{X}}}_{(tc)n}$  ( $(N \times 1)$  vectors) are perfectly correlated. In other words, all characteristics are proportional across assets. If  $K_C = 1$ ,  $c$ -slices  $\hat{\boldsymbol{\mathcal{X}}}_{(c)tn}$  ( $(T \times N)$  matrices) are proportional and  $c$ -fibers  $\hat{\boldsymbol{\mathcal{X}}}_{(tn)c}$  ( $(C \times 1)$  vectors) are perfectly correlated, so all assets observations are proportional across characteristics.

The logic for 3-dimensional Tucker models extends to higher dimensional tensors. The Tucker model (5) for an  $n$ -dimensional tensor is the tensor product of an  $n$ -dimensional core tensor and  $n$  loading matrices, or, equivalently as the  $n$  sums of outer products of  $n$  vectors similar to (7). Slices and fibers can be written as combinations of factor models of lower rank that are subject to restrictions.

### 3.5. Data compression

The core tensor  $\mathcal{G}$  has  $K_T K_N K_C$  elements and the component matrices  $\mathbf{V}^{(T)}, \mathbf{V}^{(N)}$ , and  $\mathbf{V}^{(C)}$  have  $TK_T, NK_N$ , and  $CK_C$  elements, respectively. The orthonormal normalizations and unit-norm normalizations of the loading matrices  $\mathbf{V}^{(i)}$  add  $2(K_T + K_N + K_C)$  restrictions. Thus, the Tucker decomposition (5) has  $K_T K_N K_C + TK_T + NK_N + CK_C - 2(K_T + K_N + K_C)$  parameters. Define  $\kappa_i$  as the number of model parameters divided by the number of data points of a  $i$ -dimensional data set. The data-compression ratio is defined as  $1 - \kappa_i$ . For the 3-dimensional Tucker decomposition,  $\kappa_3 = (K_T K_N K_C + TK_T + NK_N + CK_C - 2(K_T + K_N + K_C)) / (TNC)$ . For the special case when the number of factors in all three dimensions is  $K = K_T = K_N = K_C$ ,  $\kappa_3$  simplifies to  $K(K^2 + T + N + C - 6) / (TNC)$ . For comparison, the 2-dimensional  $K$ -factor SVD of a  $(T \times N)$  matrix has  $K + K(T + N) - 4K$  parameters so that  $\kappa_2 = K(T + N - 3) / (TN)$ .<sup>17</sup>

Further insights about how data compression is related to the dimensionality of the data tensor can be gained by considering the limiting case when the size of the tensor approaches infinity. For simplicity, I assume that the data tensor has  $M$  observations in each dimension and that the number of factors of the TFM is  $K_j = K$  for all  $j = 1, \dots, i$ .<sup>18</sup> I assume that  $K, M \rightarrow \infty$  at the same rate so that  $Q = K/M$  is a constant. Then  $\kappa_3 = (K/M)^3 + 3MK/M^3 - 6K/M^3 = Q^3 + Q/M - 6Q/M^2 \rightarrow Q^3$  as  $K, M \rightarrow \infty$ . Hence, the 3-dimensional Tucker model compresses the data's total size by a ratio of order  $\mathcal{O}((K/M)^3)$ . If  $Q = K/M$  is 10% and  $M$  is large, *i.e.*, there is one Tucker component for every ten data dimensions,  $\kappa_3 = 0.001$  so that the Tucker model compresses the data by approximately 99.9%. A similar calculation for the 2-dimensional SVD of an  $(M \times M)$  matrix shows that  $\kappa_2 = 2Q - 3Q/M \rightarrow 2Q$ , which is an order of magnitude higher than  $\kappa_3$  of the 3-dimensional Tucker model.

This logic can be applied to Tucker models of higher dimensions. In the special case considered here, the core tensor of the Tucker decomposition of an  $i$ -dimensional tensor ( $i > 2$ ) has  $K^i$  elements and each of the  $i$  normalized loadings matrices  $\mathbf{V}^{(i)}$  is of dimension  $(M \times K)$ . Hence there are  $K^i + iMK - 2iK$  parameters, so that  $\kappa_i = Q^i + iQ/M^{i-2} - 2iQ/M^{i-1} \rightarrow Q^i$ . Hence, the order of compression ratio increases geometrically in the dimensionality of the data tensor. The intuition is that for large  $i$  the dimensionality of the core tensor,  $K^i$ , relative to the size of the data tensor,  $M^i$ , is the dominant term. Going from  $i$  to  $i+1$  dimensions multiplies the number of elements in the data tensor by  $M$ , however, the size of the core is only multiplied by  $K$ .

### 3.6. Estimation

In contrast to the SVD-PCA matrix representation, there is no closed-form solution for the Tucker decomposition (5) that minimizes the MSE of the error tensor  $\mathcal{E}$ . Kolda and Bader (2009) and Kroonenberg (2007), chapter 10, discuss several numerical solutions methods. I use the most popular algorithm, *Higher-Order Orthogonal Iteration* (HOOI) throughout the paper. As shown in Kolda and Bader (2009) and Kroonenberg (2007), it is possible to solve for the loading matrix  $\mathbf{V}^{(i)}$  when all other  $\mathbf{V}^{(j)}, j \neq i$  are known. Therefore, the Tucker model can be solved recursively by choosing some starting

<sup>17</sup>Since  $\mathbf{H}_K$  in the SVD is diagonal, it has  $K$  free parameters. In contrast, the core tensor  $\mathcal{G}$  of the Tucker decomposition is not diagonal and has  $K_T K_N K_C$  free elements.

<sup>18</sup>It is easy to show that the results extend to the general case.

values for  $\mathbf{V}^{(T)}$  and  $\mathbf{V}^{(N)}$ , solving for  $\mathbf{V}^{(T)}$ , and iteratively solving for  $\mathbf{V}^{(i)}$  until convergence. Once the  $\mathbf{V}^{(i)}$  are solved, the core tensor  $\mathcal{G}$  can be constructed. Details are in Appendix C.

To assess the precision of the estimation, I perform a Monte Carlo simulation for various combinations of tensor sizes  $(T, N, C)$  and orders  $\mathbf{K} = (K_T, K_N, K_C)$  of Tucker models. For each combination, I simulate 1,000 samples of Tucker factor models,  $\mathbf{x}_i$  and estimate the Tucker model  $\hat{\mathbf{x}}_i$  for the true model plus noise,  $\mathbf{x}_i^e = \mathbf{x}_i + \sigma_e \boldsymbol{\varepsilon}_i$ . The elements of the noise tensor are drawn from standard normal distributions. Table 3 reports the mean RMSE of  $\mathbf{x}_i - \hat{\mathbf{x}}_i$  across the 1,000 samples. The columns correspond to different values of the standard deviation of the noise tensor relative to the standard deviation of the true factor tensor.

I consider combinations of  $(T, N, C)$ , so that the tensors  $\mathbf{x}_i$  have 1,000,000 data points, which is comparable to the size of the asset sample. I choose five combinations of  $(T, N, C)$  to mimic different data patterns. In the first case, the data dimensions are equal ( $T = N = C = 100$ ). In the empirical application, the second dimension is substantially larger than the first and third dimensions. Hence, I consider three additional cases with unbalanced data dimensions:  $(100, 500, 20)$ ,  $(40, 1000, 25)$ , and  $(25, 2000, 20)$ . The case with  $(T, N, C) = (40, 1000, 25)$  closely resembles the size of the sample used in the next section. The ratio of the standard deviation of the noise term to the standard deviations of  $\mathbf{x}_i$  has five possible values: 0, 0.1, 0.25, 0.5, 1. If  $\sigma_e / \sigma_x = 0$ , the Tucker models are estimated assuming that the true factor model is observed without error, while for  $\sigma_e / \sigma_x = 1$ , the noise term is as volatile as the data in the true factor model. To make the cases comparable, I scale the tensors so that the volatility of the observed tensor  $\mathbf{x}_i^e$  is equal to one.<sup>19</sup> The starting values of each mode- $n$  Tucker loading matrix are set to the 2-dimensional SVD decompositions computed from the unfolded tensor along mode- $n$ . Typically, the HOOI algorithm converges after 20 to 40 iterations. The procedure is robust to other starting values, albeit at the cost of slower convergence.

The first and second columns of Table 3 show the order  $(K_T, K_N, K_C)$  of the Tucker models and dimensions of the simulated data tensor. The estimation errors of the Tucker models are small even when the observed tensors contain a significant amount of noise. The largest error is 3.78% for the case when the data tensor is most unbalanced,  $(T, N, C) = (25, 2000, 20)$ , the order is large,  $(K_T, K_N, K_C) = (20, 60, 20)$ , and noisy ( $\sigma_e / \sigma_x = 1$ ). The estimation errors are under 1% for all combinations of  $(T, N, C)$  and  $(K_T, K_N, K_C)$  when  $\sigma_e / \sigma_x < 0.25$ . For fixed  $(K_T, K_N, K_C)$ , the estimation error is larger for unbalanced data tensors than for balanced tensors. For fixed  $(T, N, C)$ , the error for larger underlying factor structures is higher than for small  $(K_T, K_N, K_C)$ . The estimation error for the case that resembles the size of the sample used in this paper,  $(T, N, C) = (40, 1000, 25)$ , is below 0.5% for all combinations of  $(K_T, K_N, K_C)$  that are used in the empirical analysis below.

## 4. Empirical results

### 4.1. Fit of Tucker models

I start with the estimation of Tucker models for a wide range of combinations of  $\mathbf{K} = (K_T, K_N, K_C)$ . For each combination of  $\mathbf{K} = (K_T, K_N, K_C)$ , I compute the Tucker decomposition (1) and the associated mean

<sup>19</sup>The scaling of  $\mathbf{x}_i^e$  has an immaterial impact on the results.

square error  $\text{MSE}(\mathbf{K})$ . Since the mean of the errors of the Tucker approximation,  $\boldsymbol{\mathcal{E}}$ , is generally close to zero,  $1 - \text{MSE}(\mathbf{K}) / \text{Var}(\boldsymbol{\mathcal{X}}) \approx 1 - \text{Var}(\boldsymbol{\mathcal{E}}(\mathbf{K})) / \text{Var}(\boldsymbol{\mathcal{X}})$  can be interpreted as the  $R^2$  of the model. Since there is no formal test for the optimal choice of  $(K_T, K_N, K_C)$  in Tucker models, I follow the literature and choose “reasonable” parameters based on fit and parsimony.<sup>20</sup>

The results are shown in Figure 4. Each plot shows the MSE as a function of the number of components along one mode while keeping the numbers of the other two components fixed. Panel A plots the MSE of Tucker( $K_T, K_N, K_C$ ) models as a function of  $K_T$  for four different combinations of  $(K_N, K_C)$ :  $(K_T, 1, 1)$ ,  $(K_T, 4, 4)$ ,  $(K_T, 15, 12)$ , and  $(K_T, 25, 15)$ . The “minimal” Tucker model with a single component in each mode collapses the 34 quarters, 934 mutual funds, and 25 characteristics into a single “summary” mutual fund with a single “summary” characteristic observed at one “summary” quarter. The MSE of the minimal Tucker(1,1,1) model represented by the left-most point on the blue line in Panel A is 0.32, corresponding to an  $R^2$  of 40%. The model has a compression ratio of 99.88%. Note that increasing  $K_T$  while keeping  $K_N = K_C = 1$  does not reduce the MSE further. However, the MSE is reduced significantly when  $K_N$  and  $K_C$  are larger than one. For  $K_T = 4$  and  $K_C = 4$ , the MSE is 0.08 for a single  $K_T$  component, which is equivalent to an  $R^2$  of 84% and thus twice as high as the  $R^2$  of the minimal Tucker(1,1,1) model. Increasing  $K_T$  has a negligible effect on the MSE. The fit of Tucker models is improved by choosing higher values of  $K_N$  and  $K_C$ . The MSE for  $(K_T, 12, 12)$  range from 0.07 for  $K_T = 1$  to about 0.04 for  $K_T \gtrsim 5$ . Increasing  $K_N$  and  $K_C$  to 25 and 15, respectively, has only a negligible effect on the model fit.

Panel B has the same format but shows the MSE as a function of the number of mode-2 components,  $K_N$ . Recall that the second mode of the data tensor  $\boldsymbol{\mathcal{X}}$  corresponds to the 934 mutual funds in the sample. Hence, I consider a broader range of values of  $K_N$  from 1 to 40 for combinations  $(1, K_N, 4)$ ,  $(3, K_N, 10)$ ,  $(8, K_N, 12)$ , and  $(12, K_N, 15)$ . In all three cases, setting  $K_N$  below three yields poor fits. Furthermore, the MSE of  $(1, K_N, 4)$  models are considerably higher than those of  $(8, K_N, 12)$  models, however,  $(12, K_N, 15)$  models do not significantly improve the fit further. Finally, Panel C plots the MSE as a function of  $K_C$  for three combinations of  $K_T$  and  $K_N$ :  $(1, 4, K_C)$ ,  $(3, 10, K_C)$ ,  $(8, 12, K_C)$ , and  $(12, 25, K_C)$ . The MSE declines steeply for  $K_C \lesssim 4$  and at a lower rate for larger  $K_C$ . The MSE of the  $(8, 12, K_C)$  and  $(10, 25, K_C)$  models are almost identical.

Table 4 and Figure 5 compare the fit of Tucker models with  $(1, 4, 4)$ ,  $(3, 10, 10)$ ,  $(8, 12, 12)$ , and  $(10, 20, 15)$  components in more detail. For comparison, the overall MSE of the four models are 0.084, 0.053, 0.039, and 0.028, respectively, which correspond to  $R^2$  of 85.0%, 90.4%, 92.9%, and 94.9%, while their respective compression ratios are 99.5%, 98.8%, 98.4%, and 97.3%. Table 4 shows descriptive statistics of the distributions of the errors for the four models. The mean and median errors are close to zero so the MSE are close to the variance. Skewness is close to zero but errors are mildly leptokurtic for all specifications. The percentiles in Panel B show the improvement of the fit of models with higher  $(K_T, K_N, K_C)$ . The interquartile range of the  $(1, 4, 4)$  model is almost twice as large as that of the  $(10, 20, 15)$  and, while the minimum and maximum errors are similar, there are fewer data points with large errors.

---

<sup>20</sup>The results in the paper are robust and do not rely on a particular choice of  $(K_T, K_N, K_C)$ .

Figure 5 shows the MSE for each dimension separately rather than the overall MSE. Consider first the MSE by quarter in Panel A. The mean square error for (1,4,4) ranges from 0.06 in the middle of the sample to 0.12 at the start of the sample period. Adding further components improves the fit across the sample considerably. The MSE of the (3,10,10) model is around 0.05 for most of the sample but increases to 0.08 in 2018. The two “largest” models of orders (8,12,12), and (10,20,15) further reduce the MSE across the whole sample. Panel B shows the fit of all 934 mutual funds. Funds are sorted by MSE from smallest to largest. For all four models, the errors are small for the majority of funds but there are some funds with large errors. For example, there are 56 funds with MSE over 0.2 for the (1,4,4) model. The tail of the error distribution shrinks for larger models and there are only 25, 12, and two funds with MSE larger than 0.2 for (3,10,10), (8,12,12), and (10,20,15) models, respectively.

The MSE by characteristics is plotted in Panel C. The errors of the (1,4,4) model are considerably higher than for the larger models for all characteristics. For most characteristics, the improvements in fit for the largest (10,20,15) relative to the (3,10,10) model are relatively small. The exceptions are momentum (MOM) and reversal (REV). Note that these are the two characteristics with the highest errors for all four Tucker models. The reason is that MOM and REV are volatile and exhibit low persistence compared to the other characteristics. It turns out there is a link between the time series behavior of characteristics and the number of time components,  $K_T$ . Tucker models with low  $K_T$  can capture the variation in high-persistence characteristics but perform worse for less persistent characteristics. For example, the MSE of MOM and REV is more than twice as high as the MSE of the other (more persistent) characteristics for the Tucker(3,10,10) model. Increasing  $K_T$  to 8 and 10 improves the fit for MOM and REV but has a relatively small impact on the other characteristics.

Finally, I plot the time series of actual and fitted BM, MOM, and REV for individual mutual funds for the Tucker(10,20,15) model in Figure 6.<sup>21</sup> For each characteristic, I plot the 75th and 90th MSE percentiles funds, as well as the “worst-case” fund with the highest MSE in the sample.<sup>22</sup> Each panel plots the observed characteristic in black and the fitted values of the Tucker model in orange. The legends include the wficn identifier of the fund that is plotted as well as the MSE of the Tucker model.

Consider first the BM ratio in Panel A. The observed BM of the 75th-percentile fund is stable over the sample and varies between 1.4 and 1.9. The fitted BM matches the level but is smoother than the observed data. The pattern is similar for the 90th-percentile fund in the middle panel. The book-to-market ratio of the fund with the highest MSE in the data set increases from 1.3 to 3.8 over the sample. The fitted BM also increases but more slowly resulting in large errors at the beginning and end of the sample.

Momentum, shown in the middle row, is more volatile and less persistent than the book-to-market ratio. The Tucker model matches the level and variation of the observed data of the 75th-percentile fund well, resulting in a good fit. While the average fitted BM of the 90th-percentile fund is close to the observed average, the model does not match the time series variation as well, resulting in an MSE of 0.37 compared to 0.25 of the 75th-percentile fund. MOM of the “worst-case” fund declines between 2010 and 2012 before increasing sharply to 4.5 in 2013. After 2013, MOM slowly declines to 2 after

<sup>21</sup>The corresponding plots for the (1,4,4), (3,10,10), and (8,12,12) are in Figures E.8, E.9, and E.10 in the Appendix.

<sup>22</sup>The histograms of the MSE distributions are plotted in Figure E.7 in the Appendix.

2017. The fitted MOM matches the decline after 2013 but exhibits large errors before 2013. The fits for reversals are shown in the bottom row. For all three cases, the Tucker model captures the time series variation fairly well even though REV often changes substantially quarter-by-quarter, especially for the 90th percentile and “worst case” funds. However, the variation in the fitted REV is often more muted than the observed data.

I conclude that parsimonious Tucker factor models yield good approximations of the mutual fund sample with  $(34,934,25)$  observations. The MSE of the model with  $(K_T, K_N, K_C) = (10, 20, 15)$  components captures 94.9% of the variation in the data while compressing the data dimensions by 97.3%. Even a “small” model with  $(1, 4, 4)$  components yields an  $R^2$  of 85%. In the next subsection, I use the Tucker model with  $(K_T, K_N, K_C) = (3, 10, 10)$  to illustrate the properties of the Tucker tensor factor model.<sup>23</sup>

#### 4.2. Properties of the Tucker model

Recall that the Tucker decomposition (5) of a 3-dimensional tensor consists of a core tensor  $\mathcal{G}$  and loading matrices  $\mathbf{V}^{(T)}, \mathbf{V}^{(N)}, \mathbf{V}^{(C)}$ . For the Tucker(3,10,10) model estimated for a data tensor with dimensions  $(34 \times 934 \times 25)$ ,  $\mathcal{G}$  is a  $(3 \times 10 \times 10)$ -dimensional tensor,  $\mathbf{V}^{(T)}, \mathbf{V}^{(N)}$ , and  $\mathbf{V}^{(C)}$  are  $(3 \times 34), (10 \times 934)$ , and  $(10 \times 25)$ -dimensional matrices, respectively. This section focuses on the properties of these variables.

Panel A of Figure 7 plots the 20 largest elements (by absolute value) of the core tensor  $\mathcal{G}$  on a log-scale. The first core element with index  $(1, 1, 1)$  is the largest element, with a value of 9.15. The next two largest values are 1.19 and 0.86 for the elements with indices  $(1, 2, 2)$  and  $(1, 3, 3)$ , respectively, followed by six elements with values between 0.1 and 0.03. Recall the Tucker decomposition is a generalization of the SVD decomposition for matrices. In many economic and finance-related applications, the first eigenvalue is often significantly larger than the other eigenvalues. Even though the core tensor is not related to eigenvalues, the spectrum of the core elements is similar to typical eigenvalue spectrums. Furthermore, the five largest and seven of the nine largest core elements have a mode-1 index of one. In other words, the core tensor is dominated by elements from the first (time) index, suggesting that the first (time) dimension plays a particularly important role. This explains the earlier result that Tucker models with a single mode-1 time component can have surprisingly good fits (recall Figure 4).

Next, I analyze the structures of the loading matrices  $\mathbf{V}^{(T)}, \mathbf{V}^{(N)}, \mathbf{V}^{(C)}$ . Recall that the loading matrices of the Tucker decomposition are similar to the loading matrices of eigenvectors in the SVD matrix decomposition and can be interpreted accordingly. I have also shown in (20)-(22) that the 3-dimensional Tucker decomposition implies three 2-dimensional factor representations in which  $\mathbf{V}^{(T)}, \mathbf{V}^{(N)}, \mathbf{V}^{(C)}$  are loading matrices for appropriately defined factors.

Panel A of Figure 8 shows the heatmap of the  $(3 \times 34)$ -dimensional matrix  $\mathbf{V}^{(T)}$ . Rows correspond to the 34 time series observations, and columns correspond to the three mode-1 components of the

---

<sup>23</sup>The properties of the other specifications considered in this section are similar. For example, the first three  $T$ -factors of the  $(8, 12, 12)$  and  $(10, 25, 15)$  models are close to the three  $T$ -factors of the  $(3, 10, 10)$  model. The same is true for the first 10  $N$  and  $C$ -factors.

Tucker model. The first row represents the first quarter in the sample, 2010Q3, and the last quarter, 2018Q4, is in the bottom row. All elements of the first column of the heatmap are between 0.51 and 0.52, suggesting that the first component has the interpretation of a mean, or “level” factor, similar to many PCA applications. All other columns have positive and negative elements and have the same interpretation as higher-order eigenvectors as “long/short” factors. For example, the values of the second component are negative over the first part of the sample and positive in the latter part and thus can be interpreted as a “slope” component. The loadings of the third component are positive at the beginning and end of the sample and negative in the middle and are akin to a “curvature” factor.”

The loading matrix for the second mode,  $\mathbf{V}^{(N)}$ , in Panel B has 934 rows and 12 rows and is more difficult to visualize. To make the heatmap readable, I plot only the first 10 columns and sort each column from high to low. Hence, each of the 934 rows plots different mutual funds. The first component has only positive values and represents a “level” factor. All higher-order components are “long/short” factors with positive and negative values.

Finally, the first five columns of (25×10) loading matrix of the characteristic mode,  $\mathbf{V}^{(C)}$ , are displayed in Panel C. As for  $\mathbf{V}^{(T)}$  and  $\mathbf{V}^{(N)}$ , the elements of the first column of  $\mathbf{V}^{(C)}$  are positive so that the first component has the interpretation of a “level” factor. The second and third components are related to cross-characteristic correlation patterns, which were shown in Figure 1. The most obvious correlation blocks are price multiples and growth rates of fundamentals (plus INV) that are positively correlated within but negatively correlated with each other. The elements of the second column of  $\mathbf{V}^{(C)}$  bear this pattern out. The nine largest elements of the second component for the characteristics are related to price multiples, while the six smallest (and negative) elements are related to growth-related characteristics (plus INV). Therefore, the second component can be interpreted as a “value/growth” factor. The two characteristics with the largest elements of the third column, ME and VOL, are highly correlated, and both are strongly negatively correlated with BIDASK, which has the lowest weight in the third column. The fourth and fifth factors also have pairs of either positively or negatively correlated characteristics. For example, (DP, TURN) and (BM, OP) are pairwise negatively correlated, while (ME, VOL) and (MOM, REV) are positively correlated.

#### 4.3. Implications for 2-dimensional Factor Representations

In Section 3.4 I showed that the 3-dimensional Tucker model implies interconnected 2-dimensional factor representations, see (20)-(22). For simplicity, I explain the properties of these representations in a “small” estimated Tucker model with  $(K_T, K_N, K_C) = (2, 3, 3)$  model. The results for larger models with a better fit are almost identical. Consider first the representation in (21). Since  $K_N = 3$ , the Tucker model can be written as the sum of three outer products of the three columns of the loading matrix  $\mathbf{V}^{(N)}$  and 2-dimensional factor models  $\mathbf{F}_{TC}^{(k_N)}$  that capture the time series correlations across characteristics. Since  $\tilde{K}_{TC} = \min(K_T, K_C) = 2$ , the three  $\mathbf{F}_{TC}^{(k_N)}$  ( $\tilde{K}_{TC}$ ) models have two factors. Each  $\mathbf{F}_{TC}^{(k_N)}$  is a  $(T \times C) = (34 \times 25)$ -dimensional matrix that is constructed from the  $(2 \times 3 \times 3)$ -dimensional core tensor  $\mathcal{G}$  and the time and characteristic loading matrices  $\mathbf{V}^{(T)}$  and  $\mathbf{V}^{(C)}$  that are of dimensions  $(T \times K_T) = (34 \times 2)$  and  $(C \times K_C) = (25 \times 3)$ . Rows and columns of  $\mathbf{F}_{TC}^{(k_N)}$  correspond to quarters and characteristics, respectively.

Figure 9 compares the first factor matrices  $\mathbf{F}_{TC}^{(1)}$ ,  $\mathbf{F}_{NC}^{(1)}$  and  $\mathbf{F}_{TN}^{(1)}$  to the corresponding means of the

data. The dashed line in the left figure of Panel A plots the means of characteristics across all mutual funds in 2018Q4. For example, the means of BM and ME across funds in 2018Q4 are 2.38 and 4.13, respectively. The solid line shows the 2018Q4 row of the first factor matrix,  $F_{TC}^{(k_N)}$ ,  $k_N = 1$ . The first factor is close to characteristic means suggesting that it can be interpreted as a mean, or “level”, factor. Although the figure shows only one quarter, this pattern is true across the sample, as we will see below. The right panel has the same format but shows two columns of  $F_{TC}^{(1)}$ , BM in blue and MOM in orange. The dashed lines are the BM and MOM means in each quarter. The mean BM ratio across mutual funds is close to 2.5 throughout the sample while the mean MOM varies between 3.0 and 3.6. As in the left panel, the corresponding columns of  $F_{TC}^{(1)}$  are close to the characteristic means. Hence the rows and columns of the first factor matrix  $F_{TC}^{(1)}$  are closely related to time series and characteristic means across mutual funds of the data tensor.

Panels B and C show that this result is also true for the implied factor models in the other two dimensions,  $F_{NC}^{(1)}$  and  $F_{TN}^{(1)}$ . The rows and columns of first ( $N \times C$ ) factor matrix  $F_{NC}^{(1)}$  are related to means across quarters of funds (rows) and characteristics (columns), see Panel B. Finally, Panel C shows the first ( $T \times N$ ) factor matrix  $F_{TN}^{(1)}$  and means across characteristics. Hence, the first components of the Tucker factor model have the familiar property of 2-dimensional PCA models that the first factor can be interpreted as a mean, or “level”, factor.

While Figure 9 focused on the first components, Figure 10 shows the first three components of the factor matrices  $F_{TC}^{(k_N)}$ ,  $k_N = 1, 2, 3$ . Panel A has the same format as Panel A of Figure 9, however, it shows all  $T = 34$  rows of  $F_{TC}^{(1)}$  instead of only the row corresponding to 2018Q4. Note that the lines are close to each other implying that the factor representation is stable over time. The rows of the second factor matrix  $F_{TC}^{(2)}$  are plotted in Panel B. While the elements of the first factor matrix are positive, the second factor matrix has positive and negative elements and is thus a “long/short” factor. Note that the positive (long) elements are related to price multiples while the negative (short) elements are related to growth rates suggesting that the second factor is a “value/growth” factor. The third component matrix, shown in Panel C, is also a “long/short” factor that is long in ME and VOL and short in BIDASK.

As shown in (21), the Tucker factor model is the sum of the outer product of the  $F_{TC}^{(k_N)}$ ,  $k_N = 1, 2, 3$  matrices and the three column vectors of the ( $N \times K_N$ ) = (934  $\times$  3) loading matrix  $\mathbf{V}^{(N)} = [\mathbf{v}_1^{(N)}, \mathbf{v}_2^{(N)}, \mathbf{v}_3^{(N)}]$ . Each  $F_{TC}^{(k_N)}$  is ( $T \times C$ ) = (34  $\times$  25)-dimensional so that the outer products  $\mathbf{v}_{k_N}^{(N)} \circ F_{TC}^{(k_N)}$  yield tensors that have the same dimensions as the data tensor, ( $T \times N \times C$ ) = (34  $\times$  934  $\times$  25).

The three column vectors  $\mathbf{v}_{k_N}^{(N)}$  are shown in Panel D of Figure 10. Each vector is sorted from small to large. The elements of the first vector  $\mathbf{v}_1^{(N)}$  are close to one so that the first factor for all mutual funds is almost identical and equal to  $F_{TC}^{(k_N)}$ , i.e.  $\mathbf{v}_{1,m}^{(N)} \circ F_{TC}^{(1)} \approx F_{TC}^{(1)}, \forall m$ . Since  $F_{TC}^{(1)}$  has the interpretation as a “level” factor, the first  $k_N = 1$  factor is a “level” factor for all mutual funds. The first columns of the other two loading matrices,  $\mathbf{v}_1^{(T)}$  and  $\mathbf{v}_1^{(C)}$ , are also close to one, so that the first factors in all three 2-dimensional representations of the 3-dimensional Tucker model, (20)-(22), represent “level” factors. In contrast to the first column, the second and third columns of  $\mathbf{V}^{(N)}$  have positive and negative elements. Hence, loadings on  $F_{TC}^{(2)}$  and  $F_{TC}^{(3)}$  are positive for some mutual funds and negative for others.

#### 4.4. Subsample stability

Next, I investigate the stability of the Tucker decomposition over time. First, I split the sample into two subsamples of equal length: 2010Q3-2014Q4 and 2015Q1 to 2018Q4. Second, I estimate the model of recursive 10-year windows. I compare the fit across subsamples, the estimated core tensors  $\mathcal{G}$ , and the loading matrices  $\mathbf{V}^{(i)}$ . Since the variance of the subsample data varies, I report the  $R^2$  instead of the  $\text{MSE}(\mathbf{x})$ . Since the subsamples have shorter time spans, I reduce the number of mode-1 components from 8 to three so that  $(K_T, K_N, K_C) = (3, 12, 12)$ . In addition to splitting the sample, I also estimate the Tucker decomposition in recursive 10-year subsamples. The results are similar to those in the split sample and are reported in the Appendix.

The overall fits in the split samples are similar to the fit in the complete data set. The  $R^2$  for the two subsamples are similar, 0.94 compared to 0.93. Panel A of Figure 11 shows the largest 25 elements of the core tensors  $\mathcal{G}$  for the whole sample (black), the first half (orange), and the second half (blue) on a log-10 scale. The pattern of core values is remarkably stable across the samples. In all three cases, the  $(1,1,1)$  elements of  $\mathcal{G}$  are by far the largest core values followed by the  $(1,2,2)$  and  $(1,3,3)$  elements while the remaining core values are considerably smaller.

Figure 12 plots the first three columns of the three loading matrices in each subsample. The first row shows the mode-1 (time) loading matrices. Note, that the time spans in the split samples do not overlap, while the mutual fund and characteristic indices are identical. Despite the lack of overlap, the estimated loading matrices  $\mathbf{V}^{(1)}$  in the two subsamples are similar. The first column vectors of  $\mathbf{V}^{(1)}$  are long-only “mean” factors and are almost identical. The second and third factors differ more significantly across subsamples but have the familiar “slope” and “curvature” patterns. The modest instability of higher-order “time” components has only a second-order effect on the fit of the model since the elements of the core tensor associated with higher-order factors are an order of magnitude smaller than the values of the first component. The  $\mathbf{V}^{(2)}$  and  $\mathbf{V}^{(3)}$  loadings matrices of the second and third modes in rows 2 and 3 are almost identical in both subsamples. I also consider the fit in rolling subsamples of 10 quarters. The results are in Figures E.11 and E.12 in the Appendix and confirm the stability of the results even in short subsamples.

### 5. Mutual fund returns

In this section, I construct asset pricing factors from the estimated Tucker model. The pricing factors can be used in the estimation of linear factor models for the cross-section of mutual fund returns. Note that there is a fundamental difference between factors derived from the 3-dimensional Tucker model and factors constructed from 2-dimensional PCA factor models as in Kelly et al. (2019), Pelger (2019), Lettau and Pelger (2020a), Lettau and Pelger (2020b), Giglio and Xiu (2021)), among others. Such models are based on panels of time series of *returns* of portfolios that are constructed from sort on a large set of characteristics. Hence PCA factors capture the dependence across portfolio returns. In contrast, the Tucker factors summarize the information of a large set of *characteristics* of many assets and thus depend on the dependence of characteristics rather than returns. The dimension reduction is performed in the characteristics space rather than the return space. Furthermore, the

construction of the Tucker factors uses only information in characteristics and is independent of returns.

A distinct advantage of the 3-dimensional Tucker decomposition is that it exploits dependence in all three dimensions. In particular, factors derived from the Tucker model capture dependencies among all  $C$  characteristics. In contrast, portfolios based on sorts on univariate or bivariate characteristic sorts can only capture the correlations among the characteristics that are chosen by the econometrician. Higher-order sorts are infeasible because the number of firms in each portfolio decreases geometrically in the sort dimension. Tucker models do suffer from this curse of dimensionality and can capture dependencies across a large number of characteristics.

The intuition of the construction of Tucker factors is as follows. Suppose we want to summarize the information in all  $C$  characteristics of each mutual fund  $n$  in a small number of factors  $K_C$ . One possibility is to estimate a 2-dimensional SVD-PCA model with  $K_C$  factors for each mutual fund  $n = 1, \dots, N$ . Each factor model is estimated using the  $(T \times C)$ -dimensional matrix of  $C$  characteristics of fund  $n$  observed over  $T$  periods. This approach has several drawbacks. First, each fund factor model uses only information from one mutual fund, which is inefficient since characteristics are correlated across funds. Second, each factor model is subject to estimation error, especially if  $T$  is small and  $C$  is large. Third, the factor models are potentially inconsistent across funds. For example, the order of factors might differ. Suppose there is one “value”/“growth” factor and one profitability factor. For one fund, the “value”/“growth” factor might be the second factor and the profitability factor the second, while this order could be reversed for another fund.

The Tucker factor model offers an alternative method to construct characteristic factors for each mutual fund that do not suffer from these issues. In section 3.3, I showed that the Tucker model implies 2-dimensional factor models for each dimension. The representation for the characteristic dimension is given in (9)-(11). (11) shows that the Tucker model implies a 2-dimensional factor model for the characteristics of each mutual fund.  $\mathbf{F}_{t(n)}^{(C)}$  is the  $(T \times K_C)$ -dimensional matrix whose columns are the time series of the  $K_C$  characteristic factors of fund  $n$ . The advantage of the Tucker model is that  $\mathbf{F}_{t(n)}^{(C)}$  of all mutual funds can be computed from its 3-dimensional representation, see (9) and (10), rather than having to estimate the factors separately for each fund. Moreover, since all  $\mathbf{F}_{t(n)}^{(C)}, n = 1, \dots, N$  factor matrices stem from the same model, they are mutually consistent.

The standard approach is to form portfolios based on sorts of the  $C$  original characteristics. Instead, I compute portfolios using sorts of the  $K_C$  characteristic factors  $\mathbf{F}_{t(n)}^{(C)}$ . For each of the  $K_C$  factors, I sort mutual funds into 10 deciles according to the characteristic factor in quarter  $t$ . Hence, the 10% of funds with the lowest (highest) characteristic factors in period  $t$  are in the first (tenth) deciles. Next, I compute equally-weighted returns in the next quarter  $t + 1$  of the mutual funds in each portfolio.<sup>24</sup> I repeat this procedure for  $t = 1, \dots, T - 1$  and each of the  $K_C$  Tucker components to obtain  $10K_C$  time series of portfolio returns. Given the decile portfolios, I form “long/short” pricing factors, denoted  $\mathbf{f}_{\text{Tck},t}$ , by subtracting the returns of the decile-1 portfolios from the returns of the decile-10 portfolios

---

<sup>24</sup>I use equal-weighted portfolios throughout the paper since value-weighted returns would be dominated by a small number of very large mutual funds. However, the main results of the paper are robust to the weighting scheme in the construction of portfolios.

yielding  $K_C$  Tucker characteristic factors.<sup>25</sup>

Since the Tucker model is estimated using the entire sample, there is a look-ahead bias in the construction of the portfolios. I, therefore, also consider a recursive specification that uses only past data in the construction of portfolio returns. I estimate the Tucker model for expanding subsamples using data from  $t = 1, \dots, T'$  for  $T' = K_T, \dots, T - 1$ . For the subsample ending in  $T'$ , I form decile portfolios in period  $T'$  and compute the returns of decile portfolios in  $T' + 1$ . This procedure yields portfolio returns for periods  $K_T + 1, \dots, T$  that are based only on past information and thus not subject to a look-ahead bias. As for in-sample portfolios, I form  $K_C$  “high-minus-low” pricing factors. It turns out that the correlations of out-of-sample portfolios with the corresponding in-sample portfolios are above 0.95 for all specifications studied below indicating that the Tucker model is stable over time.

### 5.1. Properties of Tucker characteristic factors

Next, I study the properties of the characteristic factors implied by the Tucker model. The results in the remainder of this section are for the Tucker model of order (3,10,10). Since the recursive model uses data with shorter time series, I choose the Tucker(2,10,10) model as a benchmark for the out-of-sample specification. Results for higher-order models are similar. First, I compute the average characteristics of the mutual funds that are in a particular portfolio. Figure 13 plots the average (net) characteristics of the first five factors. Since characteristics are measured in an [1,5] interval and factors are the difference between two portfolios, the possible range of long/short portfolios is [-4,4]. ME and VOL characteristics of the first factor are close to minus two implying that ME and VOL of the decile-1 portfolio are substantially higher than ME and VOL of the decile-10 portfolio. Hence the first characteristic factor is long (short) in mutual funds that hold small (large) stocks with low (high) volume. By the same token, it is long (short) in funds with low (high) price multiples and high (low) bid-ask spreads. The second factor, shown in Panel B, is related to price multiples (negative) and growth rates (positive) and has therefore the interpretation of a “value/growth” factor. The third factor is positively related to ME and VOL while the fourth factor is negatively related to TURN, VOL, and ME and positively related to OP and QUAL.

Note that the characteristics of the Tucker factors are consistent with the composition of the characteristics loading matrix in Panel C of Figure 8 and the factors of the 2-dimensional representation in Figure 10. For example, the loadings of growth rates and price multiples of the second factor are positive and negative, respectively, which is also true for the second factor,  $F_{TC}^2$ , in Panel B of Figure 10 and the characteristic factor in Figure 13. The interpretation of the third factor is also consistent across representations.<sup>26</sup>

The factors also vary in terms of the types of funds that are in their underlying portfolios. For example, 60% and 79% of the funds in the decile-10 portfolio of the second and third factors are “growth” funds while 65% and 48% of the funds in their decile-1 portfolios are “value” funds. On the

---

<sup>25</sup>To distinguish characteristic factors derived from the Tucker model and pricing factors, I use lower case  $f_t$  for the latter. Since the signs of Tucker components are not identified, I normalize the pricing factors so that their means are positive.

<sup>26</sup>The interpretation of the first factor is slightly different since it is related to means so that its loadings in Figures 8 and 10 are positive, unlike those of the higher order factors. Since the first Tucker factor is nevertheless a long/short factor, its characteristics are not directly related to its loadings.

other hand, less than 7% of the funds in the tenth deciles are “value” funds and less than 2% of the funds in the first decile portfolio are “growth” funds. The differences between the compositions of the decile-1 and decile-10 portfolios of the other factors are less pronounced. For the first factor, 45% (15%) of funds in the tenth (first) deciles are cap-based while 30% (6%) of the first (tenth) decile of the fourth factor are “sector”funds.

Table 5 shows descriptive statistics of excess returns of in-sample (Panel A) and out-of-sample (Panel B) Tucker factors of the five factors with the highest mean returns. For comparison, I compute decile portfolios based on univariate sorts on each of the  $C = 25$  individual characteristics. The results for the five characteristics with the largest mean returns are in Panel C. The fifth in-sample Tucker factor has the highest mean excess return of 5.52% (annualized) and a Sharpe ratio of 0.89. The alphas from time-series regressions on the market (CAPM) and the 3-factor Fama-French model are 4.88% and 5.23%, respectively, and statistically significant at the 5% level.<sup>27</sup> The mean returns of the next four factors range from 2.56% to 3.66% with Sharpe ratios between 0.42 and 0.72. All CAPM alphas are statistically significant at least at the 10% level and three of five FF3 alphas are statistically significant. The pattern for out-of-sample factors in Panel B is similar but somewhat weaker. Mean returns and Sharpe ratios are slightly lower and three of five CAPM and FF3 alphas are statistically significant.

Compared to the returns of Tucker factors, the returns of portfolios based on traditional characteristic sorts are substantially lower. The long/short portfolio constructed from sorts on expected long-term growth (ELTG) yields the highest mean return, followed by portfolios based on reversal (REV) and the growth index (GR). Except for the quality (QUAL) portfolio, Sharpe ratios are lower than those of Tucker factors. Among all 25 characteristic portfolios, only two (QUAL and OP) have significant CAPM alphas while only the QUAL FF3-alpha is significant.

## 5.2. Tucker factors and the cross-section of mutual fund returns

Next, I study whether the characteristic factors of the tensor model are relevant for the cross-section of mutual fund returns. First, I investigate whether Tucker characteristic factors are directly linked to mutual fund returns. Second, I use the returns of long/short factors that are derived from the Tucker model as factors in linear cross-sectional asset pricing models.

To assess whether characteristics are related to returns, the standard approach is to regress excess returns on lagged characteristics. Instead, I use lagged Tucker characteristic factors  $\mathbf{F}_{t,(n)}^{(C)}$  as independent variables. There are several methods to estimate such regressions. The most popular approach is due to Fama and MacBeth (Fama and MacBeth (1973)), which estimates cross-sectional regressions in each time period. Coefficient estimates are obtained by sample means of the  $T$  regression coefficients. This method allows for time variation in betas but requires a large time series dimension for a reliable second-stage estimation. The data set used in this paper has a relatively short time series dimension  $T = 34$  but a large cross-section of  $N = 934$  mutual funds. I, therefore, use a panel regression

<sup>27</sup>Note that the time series is relatively short since  $T = 34$ . Since the  $t$ -statistic of a sample mean of an i.i.d. time series of returns is  $t\text{-stat} = \bar{R}_t / \sigma(\bar{R}_t) = \sqrt{T} \bar{R}_t / \sigma(R_t) = \sqrt{T} \text{SR}$ , statistical significance at 1%, 5%, and 10% levels require SR of at least 0.29, 0.35, and 0.45, respectively.

to estimate the model and include time fixed effects to capture variation across quarters:

$$R_{m,t+1}^e = \alpha + \beta^\top \mathbf{F}_{t,(n)}^{(C)} + \gamma_t + e_{m,t+1}, \quad (26)$$

where  $R_{m,t+1}^e$  is the excess return of fund  $n$  in quarter  $t+1$  and  $\mathbf{F}_{t,(n)}^{(C)}$ ,  $\gamma_t$  are time fixed effects. The  $\mathbf{F}_{t,(n)}^{(C)}$  factors are normalized to a unit standard deviation to make the regression coefficients comparable.

The results are reported in Table 6 for in-sample (Panel A) and out-of-sample (Panel B) Tucker characteristic factors. The table shows the regression coefficients of the  $K_C = 10$  Tucker factors and three  $t$ -statistics in parentheses based on heteroskedasticity-corrected HAC, time-clustered, and entity-clustered standard errors. The between- $R^2$  in the last column measures the fit across funds after all time effects are removed.

Consider first the results for in-sample factors. Recall that factors are scaled such that the mean returns of the associated long/short factors are positive. Although there is no mechanical link between long/short factors and regression coefficients in (26), nine out of 10 regression coefficients are positive. The HAC  $t$ -statistics and  $t$ -statistics based on entity-clustered standard errors are similar for most factors. Time-clustered  $t$ -statistics are significantly smaller suggesting that the identification stems from the mutual fund dimension rather than the time dimension, which is not surprising given the short sample span. Eight out of 10 coefficients are statistically significant using HAC and entity-clustered  $t$ -statistics and three are significant using time-clustered standard errors. The between- $R^2$  is 50%, which implies that Tucker factors capture half the variation in returns across mutual funds. The results for out-sample factors in Panel B are comparable to those for in-sample factors. Seven out of 10 coefficients are statistically significant using HAC and entity-clustered  $t$ -statistics. The between- $R^2$  is 37%. For comparison, I run the panel regression (26) using original characteristics as independent variables. The coefficients and  $t$ -statistics for the 10 characteristics with the largest coefficients are reported in Panel C. The regression coefficients have similar magnitudes as those for Tucker factors but, except for REV, the  $t$ -statistics are smaller. Note that the between- $R^2$  of the regression with 25 characteristics is lower than that of the regression with 10 Tucker factors.

These results suggest that the Tucker characteristic factors are linked to mutual fund returns. Moreover, this link appears stronger than the relationship between original characteristics and returns. Next, investigate whether the associated long/short factors  $\mathbf{f}_{\text{Tck},t}$  are also related to returns of mutual funds. Since the  $\mathbf{f}_{\text{Tck},t}$  factors are excess returns, they can be used as factors in cross-sectional asset pricing models. In the remainder of this section, I compare the cross-sectional fit of  $\mathbf{f}_{\text{Tck},t}$  factors to the fit of Fama-French portfolios. I also construct traditional PCA-based factors as follows. For each characteristic, I construct decile-sorted portfolios in quarter  $t$  and compute their returns in quarter  $t+1$  yielding time series of  $10K_C = 250$  portfolios. I then estimate the principal components for this panel of portfolio returns that represent SVD-PCA factors. As for Tucker factors, I consider an in-sample estimation as well as a recursive estimation using expanding windows. I use the first 10 in-sample SVD-PCA factors in the further analysis. Since the first  $K$  observations are lost in the recursive estimation of a SVD-PCA model with  $K$  factors, I use only the first five recursive SVD-PCA factors.

Since all pricing factors under consideration are excess returns, I run time series regressions of

excess returns of mutual funds on a set of candidate pricing factors for each fund  $n = 1, \dots, N$ :

$$R_{nt}^e = \alpha_n + \boldsymbol{\beta}_n^\top \mathbf{f}_t + e_{nt}, \quad (27)$$

where  $\mathbf{f}_t$  are (excess return) pricing factors.  $\mathbf{f}_t$  includes the excess return of the CRSP-VW index as proxy for the market in all specifications. Let  $L$  be the number of factors in (27) including the market return but excluding the constant. The pricing error of fund  $n$  is  $\alpha_n$ . I evaluate models by their root-mean-squared pricing error (RMSPE) and the mean-absolute pricing errors (MAPE) across all mutual funds:

$$\text{RMSPE} = \sqrt{\frac{1}{N} \sum_{n=1}^N \alpha_n^2}, \quad \text{MAPE} = \frac{1}{N} \sum_{n=1}^N |\alpha_n|. \quad (28)$$

I also compute the mean pricing error (MPE) as a measure of the average over or under-performance of mutual funds with respect to  $\mathbf{f}_t$ . The results for the in-sample and out-of-sample portfolios are in Tables 7 and 8, respectively. Given the relatively short sample, I consider models with no more than four factors.<sup>28</sup> For each set of factors, I add individual factors recursively to avoid searching over many combinations. In other words, I start by adding the factor that improves the RMSE the most. Then I search over the remaining factors to find the one with the lowest RMSE. I continue until the number of factors reaches  $L = 4$ .

Panel A of Table 7 shows the results for the CAPM and Fama-French models as benchmarks. The RMSPE and MAPE of the CAPM are 3.27% and 2.45%, respectively. For comparison, the average mean return of the mutual funds in the sample is 10.58%. The average CAPM alpha is -1.66% indicating that mutual funds underperform on average relative to the CAPM. The alphas of 709 of the 934 funds in the sample are negative. Adding SMB, HML, and MOM as factors lowers pricing errors substantially.<sup>29</sup> For example, the pricing errors of the 4-factor model are reduced by about one-third relative to the CAPM. Alphas of 125 mutual funds are individually statistically significant but only four funds have significantly positive alphas.<sup>30</sup>

Panel B reports the results of the in-sample factors derived from the Tucker model with (3,10,10) components. The Tucker models outperform the Fama-French models with comparable numbers of factors. For example, the RMSPE of the specification with three factors (the market and the fourth and 10th factors) is 1.90% compared to 2.59% for the 3-factor Fama-French model. The RMSPE of the respective models with four factors are 1.87% and 2.35%. The pattern for the MAPE is similar. Note that the pricing errors are small compared to the average mean fund return of 10.58% suggesting that the specifications with  $L = 3, 4$  capture the cross-section of mean mutual fund returns. For the model with  $L = 3$  factors, 46 (6) funds have individually significant (positive) alphas.

It turns out that more parsimonious Tucker models perform almost as well as the Tucker(3,10,10)

<sup>28</sup>Moreover, models with more than  $L = 4$  factors do not improve the results in any of the specifications.

<sup>29</sup>Including the CMA and INV profitability and investment factors increase the RMSE and their results are, therefore, not reported.

<sup>30</sup>Using the GRS-test, the null hypothesis that all alphas are jointly zero is rejected well below the 1% level for all models considered in this section. I, therefore, do not include the GRS statistic but report the number of individually significant pricing errors instead.

specification. Panel C shows results for a Tucker model of degree (1,4,4). The RMSPE and MAPE of the models with three or four are almost identical to those of the larger model. I conclude, therefore, that a model with the market, as well as the second and fourth factors of the parsimonious Tucker(1,4,4) model, successfully captures the cross-section of mean mutual returns.

Finally, the results for SVD-PCA factors are in Panel D. Models with SVD-PCA factors perform better than comparable Fama-French models but worse than models with Tucker factors. For example, the RMSPE of the  $L = 3$  model is 2.22% compared to 2.59%, 1.90% and 1.91% the corresponding Fama-French and Tucker models. The results for the model with four SVD-PCA factors are similar. Hence, factors derived from the 3-dimensional Tucker representations outperform SVD-PCA factors even though SVD-PCA models are estimated using time series of mutual fund *returns* and thus exploit the co-movement in the return space. In contrast, Tucker models are estimated using *only characteristics* of mutual funds and exploit the 3-dimensional co-movements of characteristics of mutual funds observed over time. In contrast to the SVD-PCA factors, the Tucker characteristic factors are constructed without any information about returns.

So far, I have only considered models that use only factors of one type. Next, I investigate the fit of models that combine different factor types. Given the relatively short time series span of the sample, I do not run horse races with a large number of factors. Instead, I add two factors of a different type to the specifications in Table 7. Consider, for example, the CAPM, which has an RMSPE of 3.27%. When the fourth and 10th Tucker factors are added to the model, the RMSPE is reduced to 1.90%, see the RMSPE\*-Tucker column. When the third and sixth SVD-PCA factors are included instead, the RMSPE shrinks to 2.22% (RMSPE\*-PCA column). By the same token, the RMSPE of the  $L = 3$  Tucker model *increase* from 1.90% to 2.24% and 1.91%, respectively, when SMB and HML, and the third and sixth SVD-PCA factors are added (RMSPE\*-FF and RMSPE\*-PCA columns). Without exception, adding Tucker factors to Fama-French and SVD-PCA models reduces the pricing errors, while adding Fama-French and SVD-PCA factors to Tucker factors does *not* lower the RMSPE, confirming that Tucker factors contain more relevant information about the cross-section of fund returns than Fama-French and SVD-PCA factors.

Table 8 reports the pricing errors for out-of-sample Tucker and SVD-PCA factors. The results for Tucker factors resemble those of the in-sample factors, confirming that the Tucker model yields consistent results even for short samples. In contrast, pricing errors for out-of-sample SVD-PCA factors are substantially higher than corresponding in-sample factors. For example, the RMSPE of the model with four out-of-sample factors is 2.74% compared to 2.06% for the in-sample model.

I conclude that the characteristic factors derived from the Tucker decomposition price the cross-section of mutual fund returns better than popular benchmark models, whether the factors are based on full-sample or recursive out-of-sample estimations. A parsimonious specification with two Tucker factors in addition to the market excess return outperforms standard benchmark models with Fama-French factors and factors based on SVD-PCA estimation of panels of mutual fund returns even when these models include more factors. Adding Fama-French-type factors or SVD-PCA factors to the 2-factor Tucker model increases the pricing errors. In contrast, the pricing errors of Fama-French and SVD-PCA specifications decrease when Tucker factors are added.

### 5.3. Pricing errors

Next, I investigate the properties of pricing errors of some of the Tucker factor models studied above. I focus on models with  $L = 3$  factor but the results for the other models are similar. Table 9 reports root-mean-square pricing errors by fund type: cap-based (C), growth (G), value (V), balanced (B), and “sector”(S), other (O). First, consider the fit by fund type. Sector funds are associated with the highest pricing errors for all models with a mean RMSPE across models of 4.20%, followed by “cap-based” funds with a mean of 2.36%. Leaving “other” funds aside, growth and “balanced” funds have the smallest pricing errors with means of 1.82% and 1.88%, respectively.

Consider next the fit by model. Without exception, RMSPE are the smallest for Tucker factors across all categories showing that the fit of these models is not due to a specific type of fund. Except for “sector” funds, the RMSPE are well below 2% for all categories. The differences in fit compared to the other models are particularly large for “sector” and “cap-based” funds and relatively small for “growth” funds. The in-sample PCA factor model yields a reasonably good fit while the pricing errors for the CAPM, the 3-factor Fama-French model, and the specification with out-of-sample PCA factors are substantially higher.

How are pricing errors related to the properties of mutual funds? To answer this question, I regress pricing errors on observable mutual fund properties. Let  $\mathbf{z}_n$  be a vector of observable properties of fund  $n$ . I consider two specifications. First, I regress pricing errors  $\alpha_n$  of a model on  $\mathbf{z}_n$ , and, second, I use absolute pricing errors as dependent variables:

$$\alpha_n = \gamma_0 + \boldsymbol{\gamma}^\top \mathbf{z}_n + e_n, \quad (29)$$

$$|\alpha_n| = \delta_0 + \boldsymbol{\delta}^\top \mathbf{z}_n + \nu_n. \quad (30)$$

To make the regression coefficients comparable, I standardize the independent variables to have zero means and unit standard deviations. I consider two sets of independent variables.

First, I use time series means of the 25 characteristics as independent variables. If an asset pricing model captures a possible link between a characteristic and returns correctly, its pricing errors should not depend on the characteristic itself. In other words, this regression is a specification test of a factor model.<sup>31</sup> The results are in Table 10. The table does not report standard errors for brevity but indicates statistical significance at the 1%, 5%, and 10% levels. The first column shows the results when mean fund returns (net of the grand mean) are used as dependent variables. For this sample period, 18 out of 25 characteristics are significantly related to mean returns at the 5% level. The largest and most negative coefficients are those of ME and MS, and VOL and BM, respectively. The mean of the absolute values of all coefficients is 1.48.

With few exceptions, the coefficients for the alphas of the in-sample Tucker models are substantially smaller. The three largest coefficients are 1.60 (INV), 1.42 (MOM), and 1.41 (BIDASK) and the three smallest are -1.55 (REV), -1.09 (EPPROJ), and -1.08 (SG). Note that patterns of regression coefficients are related to the fit of the Tucker model (See Panel C of Figure 5). For example, the MSE of the Tucker model are highest for MOM and REV, which are the characteristics with the largest coefficients

---

<sup>31</sup>The regression using absolute alphas is of less interest but is reported in Table D.2.

in absolute value. The mean of the absolute values of the coefficients is 0.68, which is substantially smaller than the mean of mean returns in the first column. The results for the model with out-of-sample Tucker in column 3 are similar to those for in-sample factors.

The coefficients of the CAPM, the 3-factor Fama-French, and the PCA models in columns 4 to 7 are generally larger (in absolute value) and more statistically significant than the coefficients for the Tucker models. The respective numbers of statistically significant (5% level) coefficients are 19, 16, 15, and 19; similar to the 16 significant coefficients for raw mean returns. The means of (absolute) values of the coefficients are 1.46, 1.26, 1.09, and 1.62, which is also close to 1.48 for raw returns. While the Tucker models do not capture the link between characteristics and returns perfectly, the pricing errors of the CAPM, the 3-factor Fama-French model, as well the PCA models inherit many of the properties of raw returns, suggesting that the models do not capture the relationships of characteristics and fund returns.

Next, I regress pricing errors on other observable fund variables. The regressors are time series means of the log of total net assets (TNA), the log of the number of stocks in the fund's portfolio, portfolio turnover, the expense ratio, and the active share. Table 11 reports the results of regressions (29) and (29) in Panels A and B, respectively. As a benchmark, I first regress the difference of mean returns of individual funds,  $\bar{R}_m$ , and the grand mean of fund returns,  $\bar{\bar{R}}_m$  on characteristics, see the first column of Table 11. The expense ratio coefficient is the largest in absolute value and is significant at the 1% level. Since it is negative, funds with higher expense ratios have on average lower mean returns (net of expenses). Similarly, the negative and significant coefficient on active share implies that funds with a higher active share tend to have lower returns. On the other hand, mutual funds with high turnover and a larger number of stocks in their portfolios have higher mean returns. The coefficient on the log of TNA is positive but statistically insignificant.

Pricing errors, or alphas, reflect the performance of a mutual fund after removing the exposure of fund returns to the factors of the asset pricing model under consideration. Hence regressions of pricing errors on characteristics reveal patterns in alphas that are not accounted for by the factors. Coefficients of some characteristics are consistent across models. For example, funds with high expense ratios have lower alphas for all models. The log(TNA) coefficients are all positive and mostly significant implying that larger funds tend to have high alphas. Turnover has a positive coefficient for the three models with the best fits (the two Tucker models and the in-sample PCA model) but negative coefficients for the other three models. Recall that firms with high active shares are associated with low mean returns. In contrast, the coefficients on Tucker models are positive and statistically significant.

The dependent variable in (30) is the absolute pricing error, hence this regression asks how the degree of mispricing is related to fund properties. First, the coefficients on the log of the average number of stocks in fund portfolios are strongly and significantly negatively linked to absolute pricing errors. Hence, funds that hold portfolios with many stocks are better priced by all models than funds with few stocks in their portfolios. In contrast, mutual funds with high (low) turnover have large (small) absolute pricing errors. Expense ratios are positively linked to pricing errors and are statistically significant for the three models with the best overall fits. The coefficients for log TNA are positive but

economically and statistically insignificant while the signs on the active share coefficients can have either sign.

## 6. Conclusion

This paper makes two contributions. First, I use tensor factor models (TFM) to summarize the information in a 3-dimensional data set of characteristics of mutual funds observed over time. TFM exploit dependencies in all three dimensions simultaneously and allow for complex patterns across characteristics. I find that parsimonious TFM capture over 90% of the variation in the data while compressing the data by over 95%. The factors of the tensor model share many of the familiar properties of 2-dimensional factor analysis. The estimation is stable over time and yields reliable results in short samples.

Second, I propose an alternative approach to resolving the “factor zoo” conundrum in asset pricing using TFM. The standard approach first sorts assets into characteristic portfolios and then addresses the “factor zoo” puzzle using panels of portfolio returns using dimension reduction methods. Instead, I first reduce the dimensionality of the data in the characteristic space using TFM and then form portfolios based on a small number of TFM characteristic factors. This methodology allows for dependencies across all characteristics and uses information across mutual funds and across time. I find that the TFM characteristic factors are related to the returns of mutual funds and capture about 50% of the cross-sectional variation in returns. Mean returns and Sharpe ratios of the TFM pricing factors are higher than those of pricing factors obtained from PCA estimations of returns of portfolios based on the original characteristics. Moreover, the TFM factors capture the cross-section of mean mutual fund returns better than Fama-French models and PCA-based pricing factors.

## References

- Andersson, Claus A. and Rasmus Bro (1998) "Improving the speed of multi-way algorithms: Part I. Tucker3," *Chemometrics and Intelligent Laboratory Systems*, 42 (1-2), 93-103.
- Babii, Andrii, Eric Ghysels, and Junsu Pan (2022) "Tensor Principal Component Analysis," SSRN Electronic Journal.
- Balasubramaniam, Vimal, John Y. Campbell, Tarun Ramadorai, and Benjamin Ranish (2023) "Who Owns What? A Factor Model for Direct Stockholding," *The Journal of Finance*, 78 (3), 1545-1591.
- Berk, Jonathan B. and Jules H. van Binsbergen (2015) "Measuring skill in the mutual fund industry," *Journal of Financial Economics*, 118 (1), 1-20.
- Berk, Jonathan B. and Richard C. Green (2004) "Mutual Fund Flows and Performance in Rational Markets," *Journal of Political Economy*, 112 (6), 1269-1295.
- Bryzgalova, Svetlana, Jiantao Huang, and Christian Julliard (2023a) "Bayesian Solutions for the Factor Zoo: We Just Ran Two Quadrillion Models," *Journal of Finance*, 78 (1), 487-557.
- Bryzgalova, Svetlana, Martin Lettau, Sven Lerner, and Markus Pelger (2023b) "Missing Financial Data," NBER.
- Calvet, Laurent E., John Y. Campbell, and Paolo Sodini (2009) "Fight or Flight? Portfolio Rebalancing by Individual Investors," *Quarterly Journal of Economics*, 124 (1), 301-348.
- Campbell, John Y. (2006) "Household Finance," *Journal of Finance*, 61 (4), 1553-1604.
- Carhart, Mark M. (1997) "On persistence in mutual fund performance," *Journal of Finance*, 52 (1), 57-82.
- Carroll, J. Douglas and Jih Jie Chang (1970) "Analysis of individual differences in multidimensional scaling via an  $n$ -way generalization of "Eckart-Young" decomposition," *Psychometrika*, 35 (3), 283-319.
- Chen, Luyang, Markus Pelger, and Jason Zhu (2023) "Deep Learning in Asset Pricing," *Management Science*.
- Chen, Rong, Han Xiao, and Dan Yang (2021) "Autoregressive models for matrix-valued time series," *Journal of Econometrics*, 222, 539-560.
- Chen, Rong, Dan Yang, and Cun Hui Zhang (2022) "Factor Models for High-Dimensional Tensor Time Series," *Journal of the American Statistical Association*, 117 (537), 94-116.
- Chinco, Alex, Adam D Clark-joseph, John Campbell et al. (2019) "Sparse Signals in the Cross-Section of Returns," *The Journal of Finance*, 74 (1), 449-492.
- Cochrane, John H. (2011) "Presidential Address: Discount Rates," *Journal of Finance*, 66 (4), 1047-1108.
- 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.
- (1988) "Risk and return in an equilibrium APT: Application of a new test methodology," *Journal of Financial Economics*, 21 (2), 255-289.
- DeMiguel, Victor, Javier Gil-Bazo, Francisco J. Nogales, and Andre A. P. Santos (2021) "Machine Learning and Fund Characteristics Help to Select Mutual Funds with Positive Alpha," *SSRN Electronic Journal*.

- Eckart, Carl and Gale Young (1936) "The Approximation of One Matrix by Another of Lower Rank," *Psychometrika*, 1 (3), 211-218.
- Elden, Lars and Berkant Savas (2009) "A Newton-Grassmann Method for Computing the Best Multilinear Rank  $(r_1, r_2, r_3)$  Approximation of a Tensor," *SIAM Journal on Matrix Analysis and Applications*, 31 (2), 248-271.
- Fama, Eugene F. and Kenneth R. French (2010) "Luck versus Skill in the cross-section of mutual fund returns," *Journal of Finance*, 65 (5), 1915-1947.
- Fama, Eugene F. and James D. MacBeth (1973) "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81 (3), 607-636.
- Feng, Guanhao, Stefano Giglio, and Dacheng Xiu (2020) "Taming the Factor Zoo: A Test of New Factors," *The Journal of Finance*, 75 (3), 1327-1370.
- Freyberger, Joachim, Andreas Neuhierl, and Michael Weber (2020) "Dissecting Characteristics Nonparametrically," *Review of Financial Studies*, 33 (5), 2326-2377.
- Giglio, Stefano and Dacheng Xiu (2021) "Asset pricing with omitted factors," *Journal of Political Economy*, 129 (7), 1947-1990.
- Gruber, Martin J. (1996) "Another puzzle: The growth in actively managed mutual funds," *Journal of Finance*, 51 (3), 783-810.
- Harshman, Richard A. (1970) "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis," *UCLA Working Papers in Phonetics*, 16 (10), 1-84.
- Harvey, Campbell R. and Yan Liu (2018) "Detecting repeatable performance," *Review of Financial Studies*, 31 (7), 2499-2552.
- (2021) "Lucky factors," *Journal of Financial Economics*, 141 (2), 413-435.
- Harvey, Campbell R., Yan Liu, and Heqing Zhu (2016) "... and the Cross-Section of Expected Returns," *Review of Financial Studies*, 29 (1), 5-68.
- Jensen, Michael C. (1968) "The performance of mutual funds in the period 1945-1964," *Journal of Finance*, 23 (2), 389-416.
- Jones, Christopher S. and Haitao Mo (2021) "Out-of-Sample Performance of Mutual Fund Predictors," *The Review of Financial Studies*, 34 (1), 149-193.
- Kaniel, Ron, Zihan Lin, Markus Pelger, and Stijn Van Nieuwerburgh (2023) "Machine-learning the skill of mutual fund managers," *Journal of Financial Economics*, 150 (1), 94-138.
- Karolyi, G. Andrew and Stijn Van Nieuwerburgh (2020) "New Methods for the Cross-Section of Returns," *The Review of Financial Studies*, 33 (5), 1879-1890.
- Kelly, Bryan T., Seth Pruitt, and Yinan Su (2019) "Characteristics are covariances: A unified model of risk and return," *Journal of Financial Economics*, 134 (3), 501-524.
- Kolda, Tamara G. and Brett W. Bader (2009) "Tensor decompositions and applications," *SIAM Review*, 51 (3), 455-500.
- Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh (2020) "Shrinking the Cross-Section," *Journal of Financial Economics*, 135 (2), 271-292.
- Kroonenberg, Pieter M. (2007) *Applied Multiway Data Analysis*, Wiley Blackwell: Hoboken, NJ.

- Lettau, Martin, Sydney C. Ludvigson, and Paulo Manoel (2021) "Characteristics of Mutual Fund Portfolios: Where Are the Value Funds?," Working Paper 25381, National Bureau of Economic Research.
- Lettau, Martin and Markus Pelger (2020a) "Factors That Fit the Time Series and Cross-Section of Stock Returns," *The Review of Financial Studies*, 33 (5), 2274-2325.
- (2020b) "Estimating latent asset-pricing factors," *Journal of Econometrics*, 218 (1), 1-31.
- Li, Bin and Alberto G. Rossi (2020) "Selecting Mutual Funds from the Stocks They Hold: A Machine Learning Approach," *SSRN Electronic Journal*.
- Mamaysky, Harry, Matthew Spiegel, and Hong Zhang (2007) "Improved forecasting of mutual fund alphas and betas," *Review of Finance*, 11 (3), 359-400.
- Martin, Ian W.R. and Stefan Nagel (2022) "Market efficiency in the age of big data," *Journal of Financial Economics*, 145 (1), 154-177.
- McLean, R. David and Jeffrey Pontiff (2016) "Does Academic Research Destroy Stock Return Predictability?," *Journal of Finance*, 71 (1), 5-32.
- Moritz, Benjamin and Tom Zimmermann (2016) "Tree-Based Conditional Portfolio Sorts: The Relation between Past and Future Stock Returns," *SSRN Electronic Journal*.
- Odean, Terrance (1998) "Are Investors Reluctant to Realize their Losses?," *Journal of Finance*, 53 (5), 1775-1798.
- Pelger, Markus (2019) "Large-dimensional factor modeling based on high-frequency observations," *Journal of Econometrics*, 208 (1), 23-42.
- Pelger, Markus and Ruoxuan Xiong (2022) "Interpretable Sparse Proximate Factors for Large Dimensions," *Journal of Business and Economic Statistics*, 40 (4), 1642-1664.
- Ricci-Curbastro, Gregorio and Tullio Levi-Civita (1900) "Méthodes De Calcul Différentiel Absolu Et Leurs Applications," *Mathematische Annalen*, 54 (1), 125-201.
- Strebulaev, Ilya A. and Toni M. Whited (2012) "Dynamic Models and Structural Estimation in Corporate Finance," *Foundations and Trends in Finance*, 6 (1-2), 1-163.
- Subrahmanyam, Avaniidhar (2010) "The Cross-Section of Expected Stock Returns: What Have We Learnt from the Past Twenty-Five Years of Research?," *European Financial Management*, 16 (1), 27-42.
- Tucker, Ledyard R. (1966) "Some mathematical notes on three-mode factor analysis," *Psychometrika*, 31 (3), 279-311.

**Table 1: Mutual Fund Characteristics**

Category	Characteristics
Multiples	Book-to-market (BM) Earnings-to-price (EP) Projected ep (EPPROJ) Cash flow-to-price (CFP) Sales-to-price (SP) Dividend-to-price (DP) Industry-adjusted book-to-market (ADJBM)
Growth rates	Earnings (EG) Long-term earnings (LTEG) Book value (BG) Cash flow (CFP) Sales (SG)
Morningstar	Value/growth (MS) Multiples (MULT) Growth rates (GR)
Momentum/reversal	Cumulative return $t-7$ to $t-2$ (MOM) Cumulative return $t-12$ to $t-7$ (REV)
Liquidity	Bid-ask spread (BIDASK) Pastor-Stambaugh (PSLIQ) Turnover (TURN) Volume (VOL)
Other	Market cap (ME) Operating profitability (OP) Investment (INV) Quality (QUAL)

Note: The table lists the mutual fund characteristics used in the paper. See Lettau et al. (2021) for a detailed description of the data.

**Table 2: Sample Statistics**

	No.	Mean	Std. Dev.	25% pct.	50% pct.	75% pct.
Panel A: Properties of Mutual Funds						
TNA (\$ mil.)		1909.70	4477.82	254.04	676.82	1789.74
No. of stocks		120.22	186.07	56.43	80.72	120.90
Mean Excess Return (% p.a.)		11.67	2.93	10.56	11.84	13.20
Std. Dev. (% p.a.)		14.40	2.17	13.20	14.36	15.65
CAPM $\beta$		1.00	0.15	0.95	1.03	1.10
CAPM $\alpha$ (% p.a.)		-1.59	2.77	-3.01	-1.71	-0.06
Panel B: Mean Returns by Fund Type						
Growth	346	12.75	1.87	11.64	12.89	14.05
Cap-based	213	11.23	3.44	10.14	11.80	12.99
Value	202	10.68	1.20	10.03	10.85	11.53
Sector	116	11.21	5.28	9.44	11.87	14.78
Balanced	45	11.14	1.64	10.84	11.65	12.17
Other	12	11.86	0.72	11.72	11.89	12.29

Note: The table reports summary statistics of the distributions of means by mutual fund. The sample period is 2010Q3 to 2018Q4.

**Table 3: Monte Carlo Simulation**

$(K_T, K_N, K_C)$	$(T, N, C)$	$\sigma_e / \sigma_x$				
		0	0.1	0.25	0.5	1
(2, 4, 4)	(100, 100, 100)	0.00%	0.00%	0.00%	0.00%	0.01%
	(100, 500, 20)	0.00%	0.00%	0.00%	0.01%	0.02%
	(50, 1000, 20)	0.00%	0.00%	0.01%	0.01%	0.02%
	(25, 2000, 20)	0.00%	0.00%	0.01%	0.01%	0.03%
(5, 10, 5)	(100, 100, 100)	0.00%	0.00%	0.01%	0.02%	0.04%
	(100, 500, 20)	0.00%	0.01%	0.02%	0.03%	0.07%
	(50, 1000, 20)	0.00%	0.01%	0.02%	0.05%	0.09%
	(25, 2000, 20)	0.00%	0.01%	0.03%	0.06%	0.13%
(10, 10, 10)	(100, 100, 100)	0.00%	0.01%	0.03%	0.06%	0.11%
	(100, 500, 20)	0.00%	0.02%	0.04%	0.08%	0.15%
	(50, 1000, 20)	0.00%	0.02%	0.05%	0.10%	0.20%
	(25, 2000, 20)	0.00%	0.03%	0.07%	0.13%	0.27%
(10, 20, 10)	(100, 100, 100)	0.00%	0.02%	0.05%	0.10%	0.19%
	(100, 500, 20)	0.00%	0.03%	0.07%	0.15%	0.29%
	(50, 1000, 20)	0.00%	0.04%	0.10%	0.19%	0.39%
	(25, 2000, 20)	0.00%	0.05%	0.13%	0.26%	0.53%
(20, 40, 20)	(100, 100, 100)	0.00%	0.11%	0.27%	0.54%	1.10%
	(100, 500, 20)	0.00%	0.14%	0.35%	0.70%	1.46%
	(50, 1000, 20)	0.00%	0.17%	0.43%	0.87%	1.83%
	(25, 2000, 20)	0.00%	0.22%	0.56%	1.14%	2.41%
(20, 60, 20)	(100, 100, 100)	0.00%	0.15%	0.39%	0.77%	1.58%
	(100, 500, 20)	0.00%	0.20%	0.51%	1.04%	2.20%
	(50, 1000, 20)	0.00%	0.25%	0.64%	1.30%	2.80%
	(25, 2000, 20)	0.00%	0.34%	0.84%	1.73%	3.73%

Note: This table reports results from a Monte Carlo simulation for estimations of Tucker models for different dimensions of the data tensor and orders of the factor models. For a given size  $(T, N, C)$ , I simulate 1,000 samples of Tucker factor models with order  $(K_T, K_N, K_C)$ ,  $\mathbf{x}_i$  and estimate the Tucker model for the true model plus noise,  $\mathbf{x}^e = \mathbf{x} + \sigma_e \mathbf{E}$ . The table reports the mean RMSE of  $\mathbf{x} - \hat{\mathbf{x}}$ . The columns correspond to different values of the standard deviation of the noise tensor,  $\sigma_e$ , relative to the standard deviation of the true factor tensor,  $\sigma_x$ .

**Table 4:** Distributions of Errors of Tucker Models

	$(K_T, K_N, K_C)$			
	(1,4,4)	(3,10,10)	(8,12,12)	(10,20,15)
Panel A: Moments				
MSE	0.08	0.05	0.04	0.03
$R^2$	0.85	0.90	0.93	0.95
Mean	-0.00	0.00	0.00	-0.00
Median	0.00	0.00	0.00	-0.00
Std. Dev.	0.29	0.23	0.20	0.17
Skew	-0.06	0.10	0.04	0.05
Kurt.	4.05	4.98	4.04	4.44
Panel B: Percentiles				
Min	-2.41	-2.40	-2.24	-2.20
0.005	-0.98	-0.74	-0.63	-0.53
0.05	-0.45	-0.36	-0.31	-0.26
0.25	-0.16	-0.13	-0.11	-0.09
0.75	0.16	0.13	0.11	0.09
0.95	0.45	0.36	0.31	0.27
0.995	0.91	0.75	0.64	0.53
Max	2.87	2.98	2.89	2.83

Note: The table reports summary statistics of the distributions of errors of Tucker models with  $(K_T, K_N, K_C)$  components. The sample period is 2010Q3 to 2018Q4.

**Table 5: Returns of Tucker Factors**

A: In-sample Tucker Factors					
	5	9	4	3	10
Mean	5.52	3.66	3.30	3.19	2.56
Std. Dev.	6.23	7.96	4.60	7.62	6.08
SR	0.89	0.46	0.72	0.42	0.42
CAPM $\alpha$	4.88**	5.38*	5.20***	5.58**	5.09**
	(2.04)	(1.80)	(3.25)	(2.01)	(2.41)
FF3 $\alpha$	5.23**	3.60	4.04**	1.44	4.11**
	(2.18)	(1.49)	(2.57)	(1.22)	(2.07)
B: Out-of-sample Tucker Factors					
	5	2	10	1	4
Mean	4.00	3.27	2.84	2.80	2.74
Std. Dev.	6.03	8.17	4.14	5.35	5.24
SR	0.66	0.40	0.69	0.52	0.52
CAPM $\alpha$	4.66**	0.84	4.18***	1.26	5.13***
	(2.02)	(0.28)	(2.79)	(0.64)	(2.89)
FF3 $\alpha$	3.67	0.89	4.02***	3.34***	3.58**
	(1.57)	(0.45)	(2.62)	(2.93)	(2.13)
C: PCA Factors					
	ELTG	REV	GR	QUAL	INV
Mean	2.92	2.39	2.04	1.98	1.72
Std. Dev.	8.72	6.36	8.56	4.32	7.79
SR	0.34	0.38	0.24	0.46	0.22
CAPM $\alpha$	-0.78	0.53	-1.64	3.67**	0.34
	(-0.26)	(0.23)	(-0.56)	(2.41)	(0.11)
FF3 $\alpha$	-0.18	0.05	-1.11	1.87*	-0.28
	(-0.08)	(0.03)	(-0.51)	(1.78)	(-0.13)

Note: This table reports annualized means, standard deviations, Sharpe ratios, CAPM alphas, and alphas of the 3-factor Fama-French model of excess returns of in-sample and out-of-sample Tucker factors (Panels A and B, respectively), factors based on fund characteristics (Panel C), and factors derived from in-sample and out-of-sample PCA estimations of the panel of decile portfolios of all 25 characteristics. In-sample Tucker factors are derived from a Tucker(3,10,10) and out-of-sample factors are based on a Tucker(2,20,10) model.  $t$ -statistics of alphas are in parentheses. The sample period is 2010Q3 to 2018Q4.

**Table 6:** Panel Regression with Tucker Factors

	1	2	3	4	5	6	7	8	9	10	$R^2$
Panel A: In-sample Factors											
$\beta$	0.36	0.10	0.34	0.26	0.30	0.29	0.35	-0.05	0.09	0.22	0.50
$t$ -HAC	(6.29)	(3.61)	(7.94)	(7.83)	(9.13)	(7.36)	(8.47)	(-1.81)	(3.04)	(5.62)	
$t$ -entity	(6.49)	(3.51)	(9.08)	(8.77)	(10.47)	(8.26)	(14.37)	(-1.92)	(3.28)	(6.87)	
$t$ -time	(2.27)	(0.42)	(1.71)	(2.27)	(1.73)	(2.19)	(1.56)	(-0.45)	(0.35)	(0.92)	
Panel B: Out-of-sample Factors											
$\beta$	0.07	0.29	0.04	0.19	0.27	0.13	0.14	0.10	-0.02	0.03	0.37
$t$ -HAC	(1.87)	(11.41)	(1.18)	(6.22)	(8.73)	(4.25)	(4.93)	(3.43)	(-0.60)	(1.28)	
$t$ -entity	(2.12)	(13.76)	(1.12)	(4.27)	(6.47)	(4.79)	(5.39)	(4.10)	(-0.57)	(1.72)	
$t$ -time	(0.40)	(1.09)	(0.19)	(1.73)	(2.32)	(1.21)	(1.13)	(1.21)	(-0.18)	(0.53)	
Panel C: Characteristics											
	MS	MULT	REV	EP	ELTG	BIDASK	ADJBM	CFG	OP	VOL	$R^2$
$\beta$	0.61	0.47	0.44	0.33	0.28	0.18	0.14	0.13	0.08	0.04	0.45
$t$ -HAC	(2.33)	(1.73)	(9.04)	(4.10)	(2.00)	(2.46)	(1.92)	(2.51)	(1.36)	(0.18)	
$t$ -entity	(2.48)	(1.57)	(8.66)	(4.94)	(1.97)	(2.17)	(2.39)	(2.14)	(1.48)	(0.18)	
$t$ -time	(1.04)	(0.64)	(2.24)	(1.37)	(0.51)	(0.71)	(0.63)	(0.81)	(0.64)	(0.08)	

Note: This table shows the results of panel regressions of mutual fund excess returns on Tucker factors:

$$R_{n,t+1}^e = \alpha + \beta^\top \mathbf{F}_{t,(n)}^{(C)} + \gamma_t + e_{n,t+1},$$

where  $R_{n,t+1}^e$  is the excess return of fund  $n$  in quarter  $t+1$  and  $\mathbf{F}_{t,(n)}^{(C)}$ .  $\gamma_t$  are time fixed effects. The  $\mathbf{F}_{t,(n)}^{(C)}$  factors are normalized to a unit standard deviation to make the regression coefficients comparable. The between  $R^2$  measures the fit across funds after all time effects are removed. Results for in-sample factors are reported in Panel A and Panel B reports results for out-of-sample factors. The table shows  $t$ -statistics based on heteroskedasticity-corrected HAC, time-clustered, and entity-clustered standard errors. The sample period is 2010Q3 to 2018Q4.

**Table 7: Pricing Errors for In-sample Factors**

Factors	L	MPE	MAPE	RMSPE	RMSPE*		
					Tucker	FF	PCA
Panel A: CAPM and Fama-French Factors							
MKT	1	-1.66%	2.45%	3.27%	1.90%		2.22%
MKT, SMB	2	-0.77%	2.09%	2.85%	2.02%		2.71%
MKT, SMB, HML	3	-0.83%	1.87%	2.59%	2.24%		2.48%
MKT, SMB, HML, MOM	4	-0.73%	1.75%	2.35%	2.23%		2.39%
Panel B: Tucker(3,10,10) Factors							
MKT, 10	2	-1.34%	1.89%	2.44%		2.21%	1.94%
MKT, 4, 10	3	-0.17%	1.46%	1.90%		2.24%	1.91%
MKT, 4, 10, 1	4	-0.41%	1.43%	1.87%		2.26%	1.88%
Panel C: Tucker(1,4,4) Factors							
MKT, 4	2	-0.29%	2.18%	2.80%		2.47%	2.73%
MKT, 2, 4	3	-0.65%	1.48%	1.91%		2.46%	1.88%
MKT, 2, 3, 4	4	-0.53%	1.45%	1.87%		2.44%	1.87%
Panel D: PCA Factors							
MKT, 3	2	-0.75%	1.88%	2.58%	1.90%	2.64%	
MKT, 3, 6	3	-0.36%	1.66%	2.22%	1.91%	2.48%	
MKT, 3, 6, 2	4	-0.48%	1.53%	2.06%	1.88%	2.22%	

Note: This table shows the results of time series estimations of linear asset pricing models. The factors are Fama-French factors (Panel A), factors derived from Tucker(3,10,10) and Tucker(1,4,4) models (Panels B and C, respectively), and PCA factors (Panel D).  $L$  is the number of factors, MPE is the mean pricing error, MAPE is the mean absolute pricing error, and RMSE is the root-mean-square pricing error. RMSPE\* is the root mean square pricing error when Tucker factors (“Tucker” column), SMB and HML (column “FF”), or PCA factors (column “PCA”) is added to the specification. The sample period is 2010Q3 to 2018Q4.

**Table 8: Pricing Errors for Out-of-sample Factors**

Factors	L	MPE	MAPE	RMSPE	RMSPE*		
					Tucker	FF	PCA
Panel A: Tucker(2,20,10) Factors							
MKT, 4	2	-0.38%	1.87%	2.40%		2.42%	2.28%
MKT, 2, 4	3	-0.63%	1.39%	1.79%		1.80%	2.12%
MKT, 2, 4, 3	4	-0.52%	1.40%	1.78%		1.77%	2.13%
Panel B: Tucker(1,4,4) Factors							
MKT, 4	2	-0.44%	2.04%	2.62%		2.55%	2.47%
MKT, 2, 4	3	-0.76%	1.45%	1.87%		1.83%	2.28%
MKT, 2, 3, 4	4	-0.59%	1.42%	1.83%		1.85%	2.28%
Panel C: PCA Factors							
MKT, 3	2	-1.02%	2.10%	2.92%	2.42%		2.80%
MKT, 3, 4	3	-0.97%	2.01%	2.76%	2.53%		2.73%
MKT, 3, 4, 5	4	-0.94%	1.98%	2.74%	2.56%		2.72%

Note: See note of Table 7 but for out-of-sample factors. The sample period is 2010Q3 to 2018Q4.

**Table 9: RMSPE by Fund Type**

	C	B	G	V	S	O
Tucker IS	1.85%	1.53%	1.65%	1.82%	2.80%	0.99%
Tucker OOS	1.72%	1.63%	1.68%	1.49%	2.66%	0.86%
CAPM	3.37%	2.25%	2.05%	3.06%	5.84%	1.19%
FF3	2.54%	2.04%	1.85%	1.92%	4.86%	1.20%
PCA IS	2.12%	1.77%	1.81%	1.92%	3.73%	1.42%
PCA OOS	2.55%	2.09%	1.89%	2.22%	5.31%	1.31%

Note: This table shows average RMSPE of the factor models with  $L = 3$  factors in Tables 7 and 8 by fund types: cap-based (C), growth (G), value (V), balanced (B), and "sector"(S), other (O). The sample period is 2010Q3 to 2018Q4.

**Table 10: Pricing Errors and Fund Characteristics**

	$\bar{R}_n - \overline{\bar{R}}_n$	Tucker IS	Tucker OOS	CAPM	FF3	PCA IS	PCA OOS
const	-0.00	-0.17***	-0.63***	-1.66***	-0.83***	-0.36***	-0.97***
ADJBM	1.02***	-0.58***	-0.13	0.59***	0.37*	0.28	0.40*
BG	-1.23***	-0.66**	-0.92***	-1.11***	-1.01***	-1.05***	-0.95***
BIDASK	0.54*	1.41***	1.78***	1.63***	1.36***	1.86***	1.60***
BM	-3.02***	0.42	-0.19	-1.99***	-2.03***	-0.70**	-1.60***
CFG	0.70**	-0.25	-0.24	-0.31	-0.21	0.47*	-0.39
CFP	-2.07***	0.28	-0.46	-2.16***	-1.73***	-1.43***	-1.86***
DP	-1.45***	-0.48**	-0.66***	-1.21***	-1.12***	-0.62***	-1.32***
EG	-0.98**	0.51	0.52*	0.93***	0.41	0.33	1.23***
ELTG	0.80	0.63	-0.44	0.75	-0.21	-0.68	1.68**
EP	1.15***	-0.38	-0.01	0.68**	0.58**	1.19***	0.48
EPPROJ	-1.04	-1.09*	-1.09**	-3.12***	-2.98***	-1.07*	-3.35***
GR	-0.23	-0.61	-0.94	-4.31***	-1.41	0.92	-6.74***
INV	1.85***	1.60***	1.67***	1.35***	1.36***	1.42***	1.21***
ME	4.82***	0.97*	1.88***	3.13***	2.92***	4.28***	3.45***
MOM	-0.46	1.42***	0.91***	0.56	0.56	0.97***	0.95**
MS	4.53***	0.80	0.13	1.86	3.20**	2.57**	-0.66
MULT	0.52	0.64	1.38	3.25**	2.69*	-0.55	4.83***
OP	0.41**	0.10	0.07	0.76***	0.44**	0.40**	0.53***
PSLIQ	1.07***	0.33*	0.63***	0.93***	1.02***	0.69***	0.81***
QUAL	-0.90***	0.21**	-0.00	-0.45***	-0.55***	-0.09	-0.47***
REV	1.52***	-1.55***	-0.72**	0.12	0.04	-0.56	-0.51
SG	0.41	-1.08***	-0.54	1.99***	1.11**	-0.45	2.09***
SP	0.33	0.02	-0.01	0.20	0.10	0.54***	0.00
TURN	-0.21*	0.50***	0.22**	-0.48***	-0.30**	0.33***	-0.30**
VOL	-5.79***	-0.55	-1.14**	-2.51***	-3.77***	-3.74***	-3.11***
R <sup>2</sup>	0.62	0.30	0.23	0.66	0.52	0.46	0.52

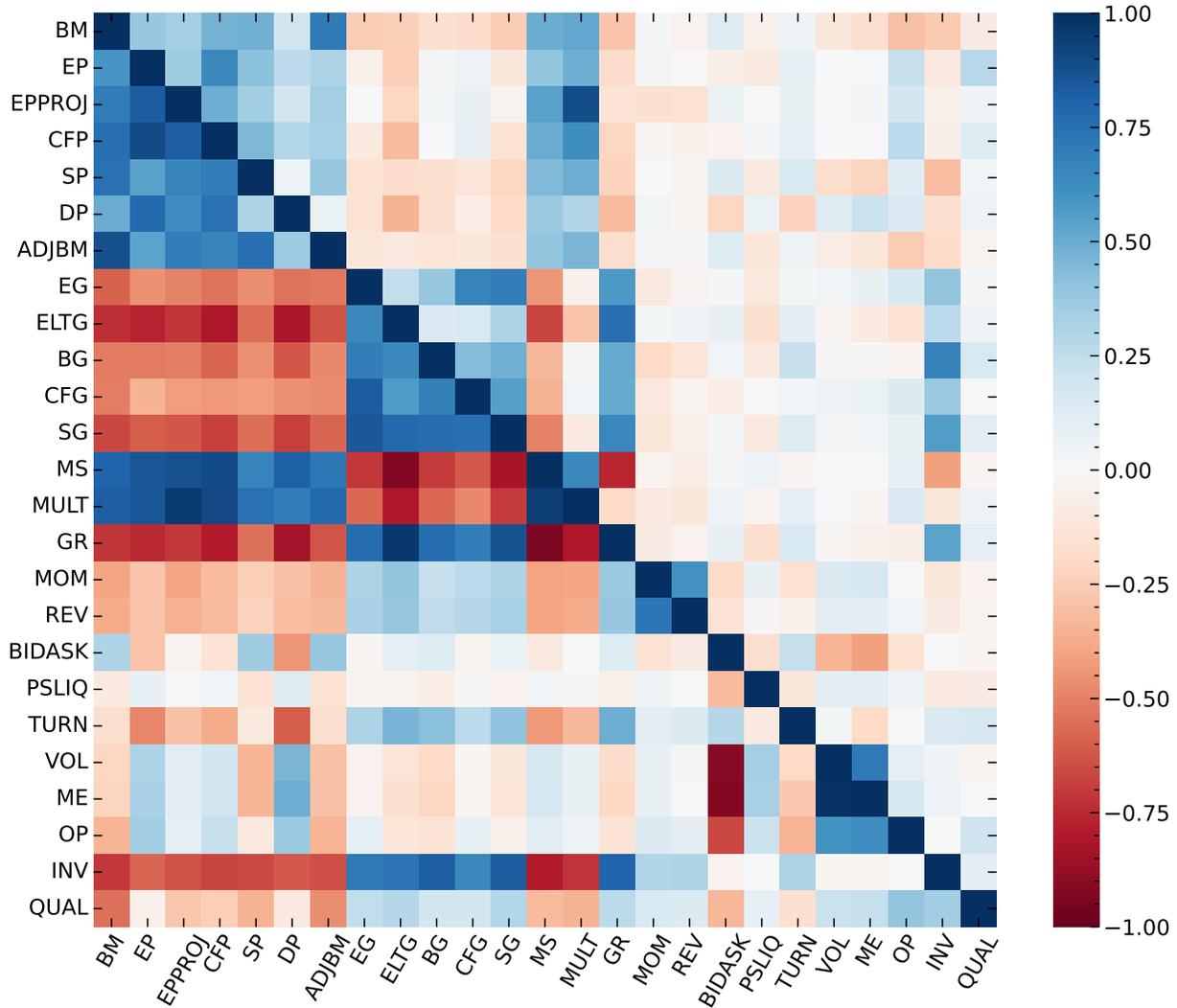
This table reports the results of regressions of pricing errors from factor models on the average expense ratio (Exp. ratio), the average number of stocks (No. stocks), and the average active share. The variables are standardized to have means of zero and unit standard deviations. The factors in the models are those listed in Tables 7 and 8 for  $L = 3$ . Statistical significance at the 1%, 5%, and 10% levels is indicated by three, two, and one star, respectively. The sample period is 2010Q3 to 2018Q4.

**Table 11: Pricing Errors and Fund Properties**

	$\overline{R}_n - \overline{\overline{R}}_n$	Tucker IS	Tucker OOS	CAPM	FF3	PCA IS	PCA OOS
Panel A: Pricing Errors							
const	0.10 (0.09)	-0.14** (0.06)	-0.60*** (0.06)	-1.59*** (0.09)	-0.74*** (0.08)	-0.30*** (0.07)	-0.91*** (0.09)
log(TNA)	0.12 (0.11)	0.23*** (0.07)	0.16** (0.07)	0.20** (0.10)	0.09 (0.09)	0.14* (0.08)	0.21** (0.10)
log(No. stocks)	0.19* (0.11)	-0.03 (0.07)	0.04 (0.07)	-0.39*** (0.11)	0.44*** (0.10)	0.13 (0.09)	-0.01 (0.10)
Turnover	0.27** (0.11)	0.64*** (0.07)	0.17** (0.07)	-0.25** (0.10)	-0.31*** (0.09)	0.40*** (0.08)	-0.28*** (0.10)
Exp. ratio	-0.34*** (0.12)	-0.24*** (0.07)	-0.28*** (0.07)	-0.28** (0.11)	-0.23** (0.10)	-0.25*** (0.09)	-0.22** (0.11)
Active share	-0.28** (0.12)	0.29*** (0.08)	0.19*** (0.07)	-0.64*** (0.11)	0.51*** (0.10)	0.03 (0.09)	0.09 (0.11)
Adjusted $R^2$	0.05	0.17	0.05	0.12	0.05	0.02	0.06
Panel B: Absolute Pricing Errors							
const	1.84*** (0.07)	1.46*** (0.04)	1.39*** (0.04)	1.84*** (0.07)	1.84*** (0.05)	1.63*** (0.05)	1.98*** (0.06)
log(TNA)	0.18** (0.08)	0.02 (0.05)	0.03 (0.04)	0.18** (0.08)	0.02 (0.06)	0.07 (0.05)	0.11 (0.07)
log(No. stocks)	-0.33*** (0.08)	-0.32*** (0.05)	-0.25*** (0.04)	-0.33*** (0.08)	-0.32*** (0.07)	-0.43*** (0.05)	-0.28*** (0.07)
Turnover	0.59*** (0.08)	0.21*** (0.05)	0.12*** (0.04)	0.59*** (0.08)	0.21*** (0.06)	0.29*** (0.05)	0.18*** (0.07)
Exp. ratio	0.06 (0.08)	0.14*** (0.05)	0.19*** (0.05)	0.06 (0.08)	0.09 (0.07)	0.10* (0.06)	0.11 (0.08)
Active share	0.12 (0.08)	-0.08* (0.05)	-0.03 (0.05)	0.12 (0.08)	0.10 (0.07)	-0.15*** (0.06)	0.26*** (0.08)
Adjusted $R^2$	0.08	0.10	0.10	0.14	0.08	0.10	0.09

This table reports the results of regressions of pricing errors from factor models on the average log TNA, log of numbers of stocks in a fund's portfolio, turnover, expense ratio, and the average active share. The variables are standardized to have means of zero and unit standard deviations. The factors in the models are those listed in Tables 7 and 8 for  $L = 3$ . Statistical significance at the 1%, 5%, and 10% levels is indicated by three, two, and one star, respectively. The sample period is 2010Q3 to 2018Q4.

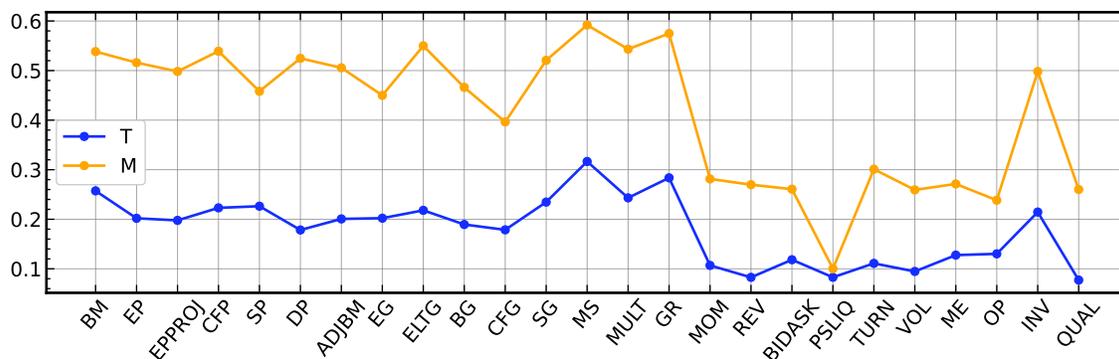
**Figure 1: Cross-correlations of Characteristics**



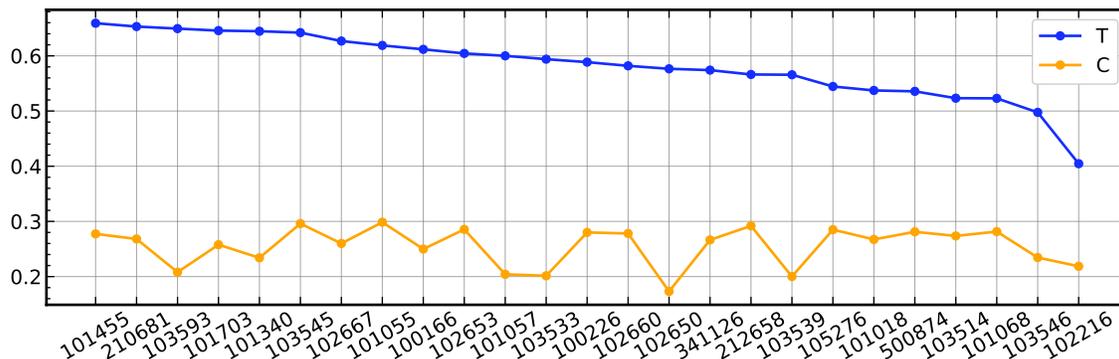
Notes: The figure shows the heatmap of pairwise correlations of mutual fund characteristics. First, I compute times-series correlations of characteristics by mutual funds and then average across funds, Second, I compute cross-sectional correlations of characteristics by quarter and then average across quarters. The lower left triangle shows cross-sectional correlations and the upper right triangle shows time series correlations. The sample period is 2010Q3 to 2018Q4.

**Figure 2: 2-dimensional Correlations**

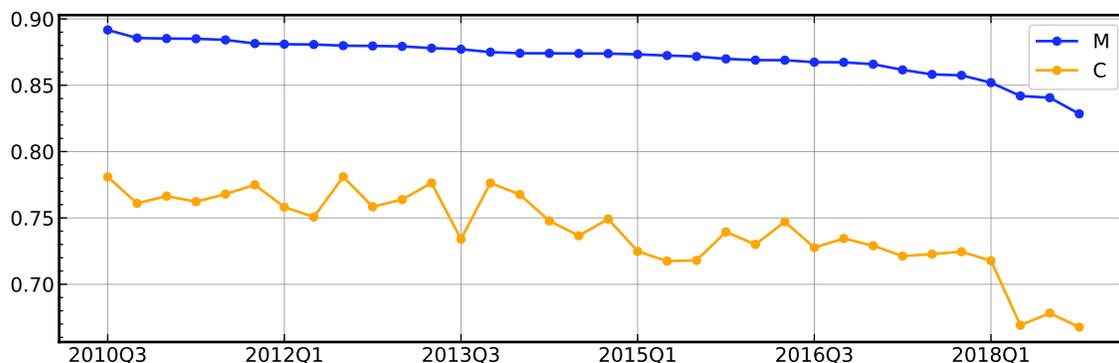
**A: Characteristics**



**B: Mutual funds**

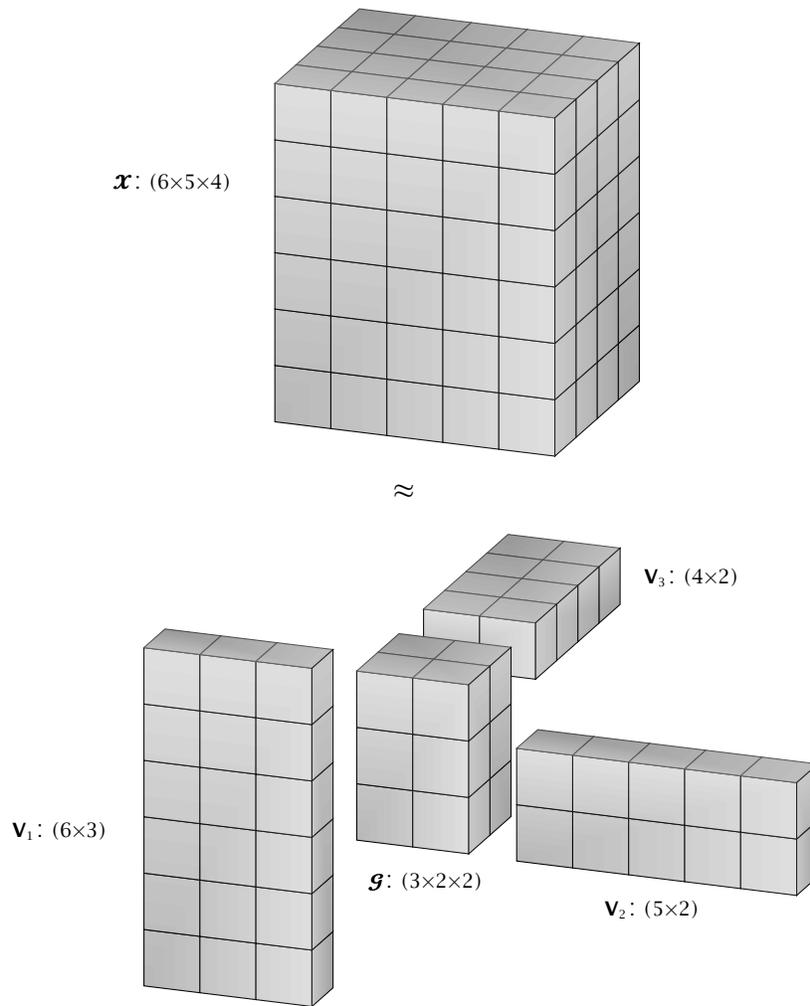


**C: Time**

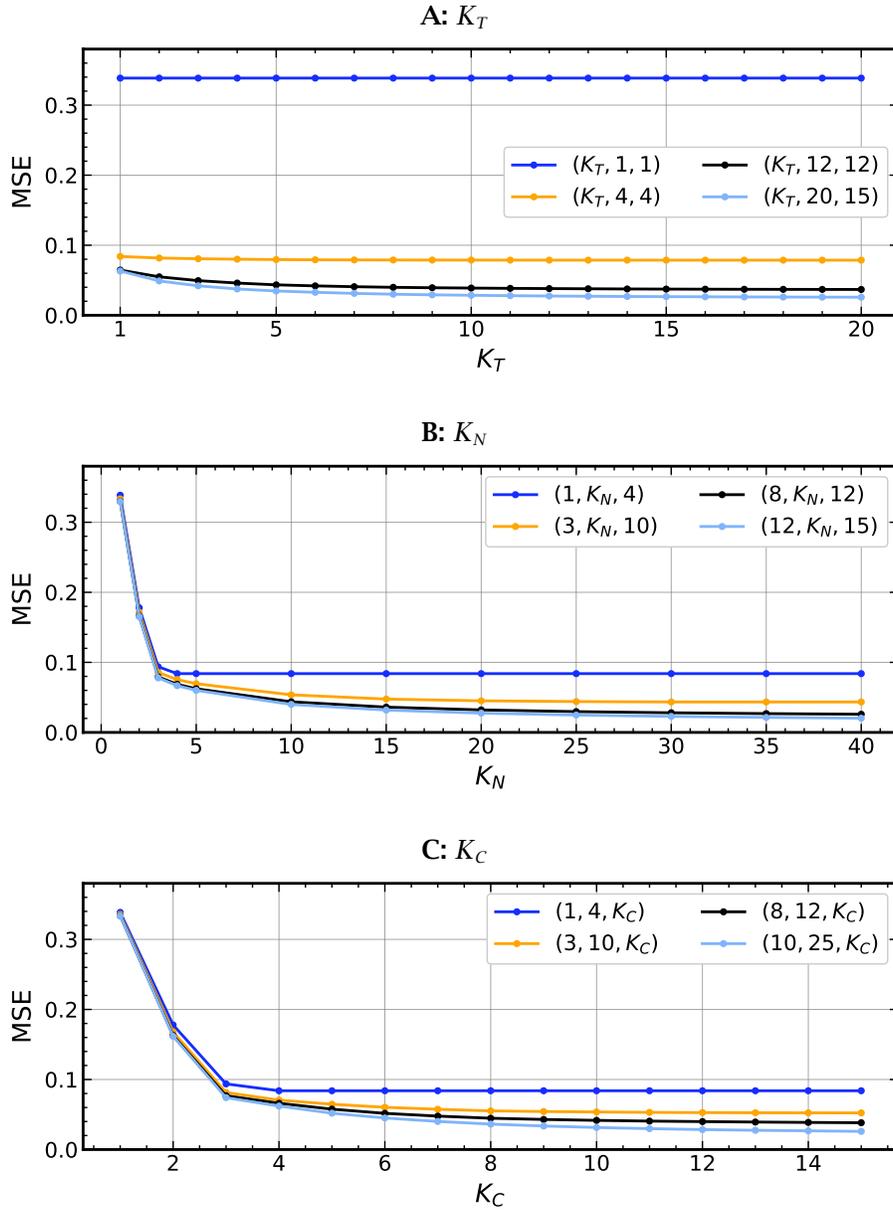


Notes: This figure shows averages of 2-dimensional correlations implied by the data tensor  $\mathcal{X}$ . For each dimension, I compute mean correlations grouped by the other two dimensions. For characteristics, I compute mean times-series correlations by mutual fund as well as mean cross-sectional correlations by quarter. For mutual funds, I compute mean times-series correlations by characteristic as well as mean cross-sectional correlations by quarter. For the time dimension, I compute mean correlations by mutual fund as well as mean correlations by characteristic. The top panel plots mean time series (blue) and cross-sectional (orange) correlations for each characteristic. The middle panel plots mean time series (blue) and cross-sectional (orange) correlations for 20 randomly selected mutual funds. The bottom panel plots mean correlations across funds (blue) and across characteristics (orange) for each quarter.

**Figure 3:** Tucker Decomposition  $\mathbf{x} = \mathcal{G} \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \times_3 \mathbf{V}_3$



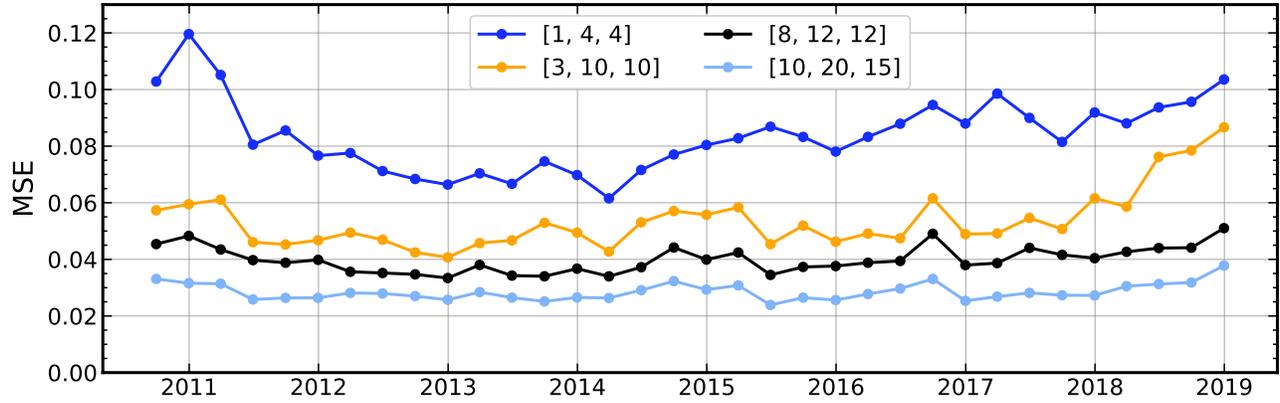
**Figure 4: MSE of Tucker( $K_T, K_N, K_C$ ) models**



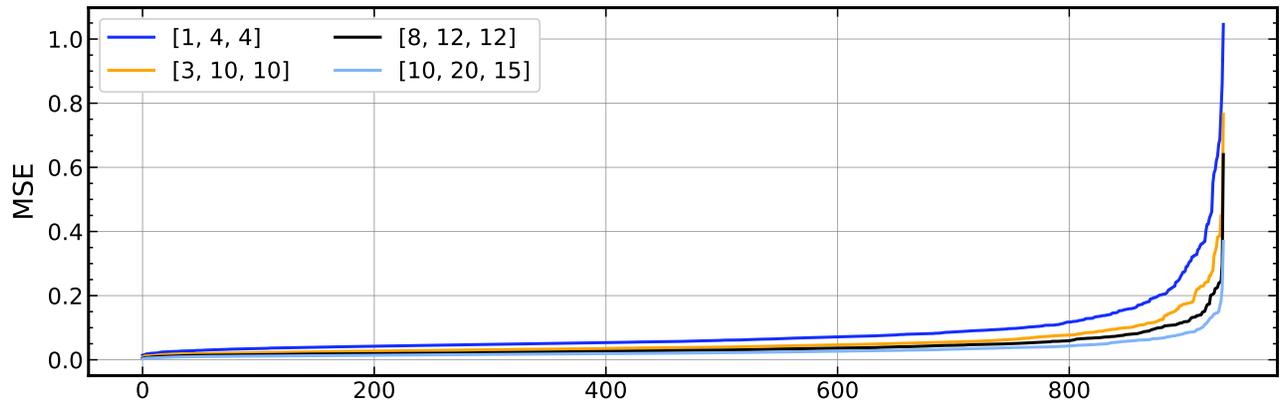
Notes: The figure plots mean-squared errors of the Tucker( $K_T, K_N, K_C$ ) model as a function of  $K_T$  (Panel A),  $K_N$  (Panel B), and  $K_C$  (Panel C). The sample period is 2010Q3 to 2018Q4.

**Figure 5: Fit of Tucker models by Dimension**

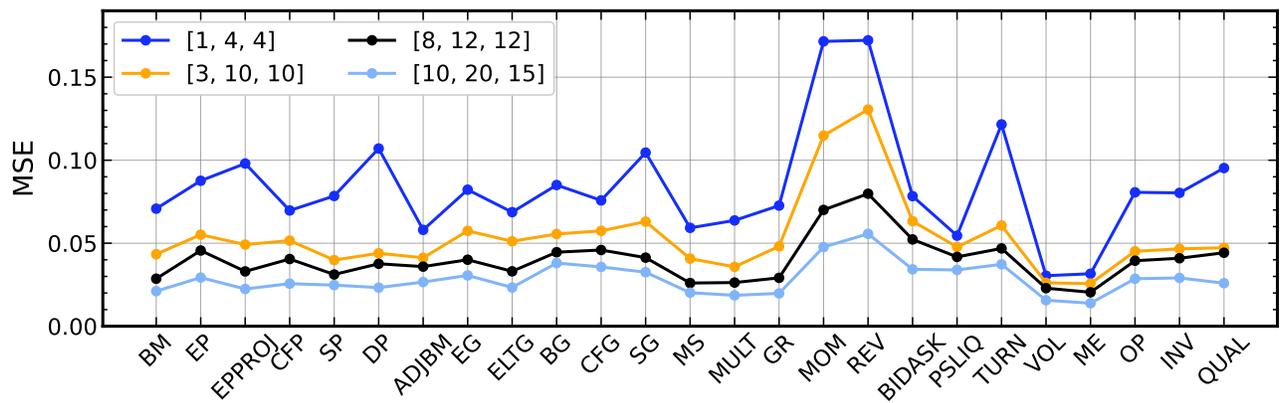
**A: MSE by quarters**



**B: MSE by mutual fund (ordered)**

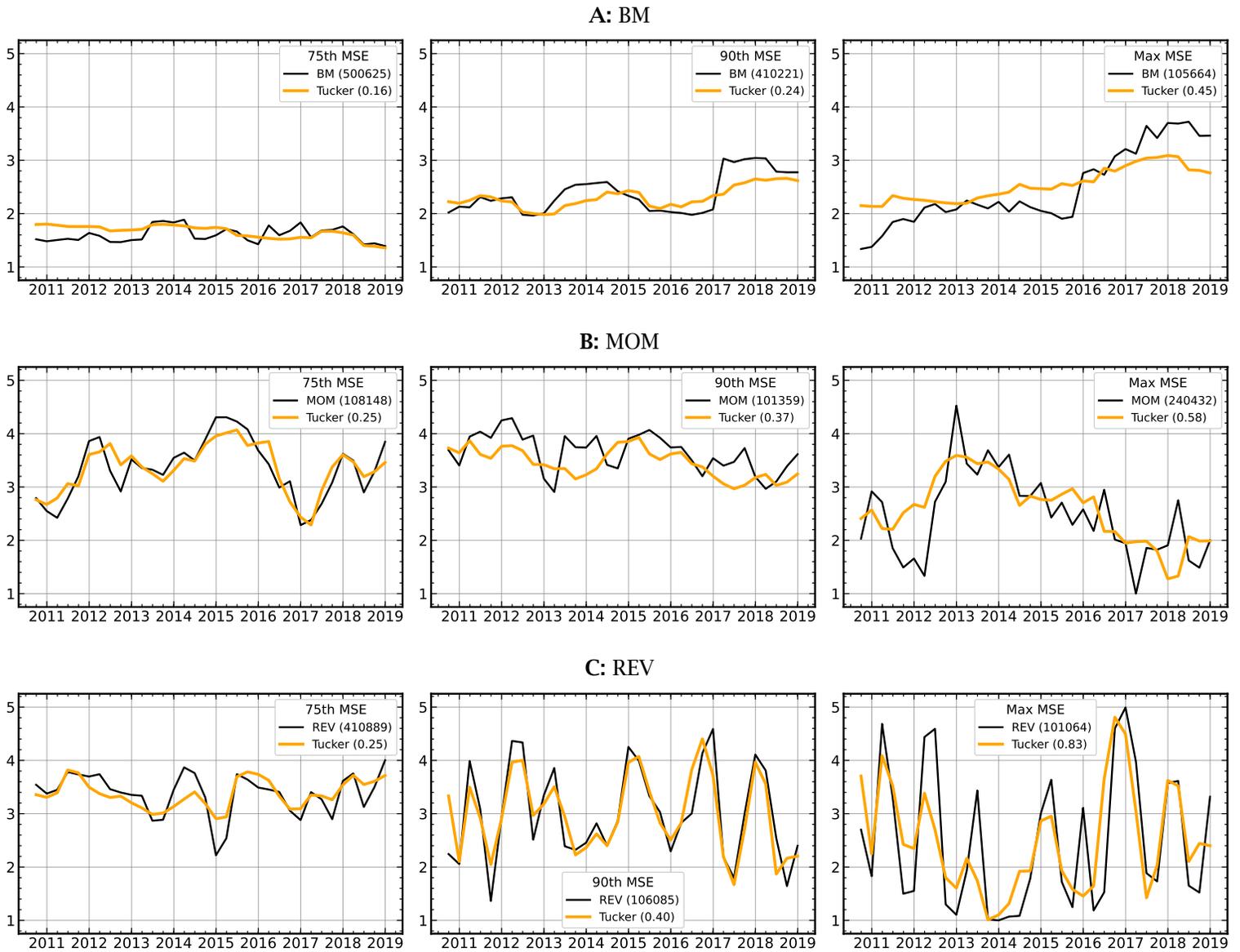


**C: MSE by characteristics**



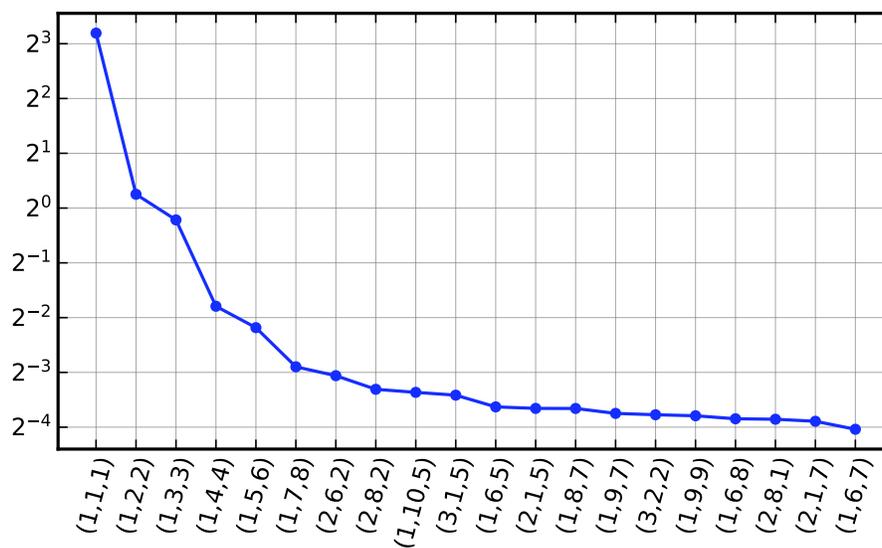
Notes: The figure plots the means of mean-squared errors (MSE) by quarters, mutual funds, and characteristics, respectively. The sample period is 2010Q3 to 2018Q4.

**Figure 6: Fit of Tucker(10,20,15) Model of Individual Funds**



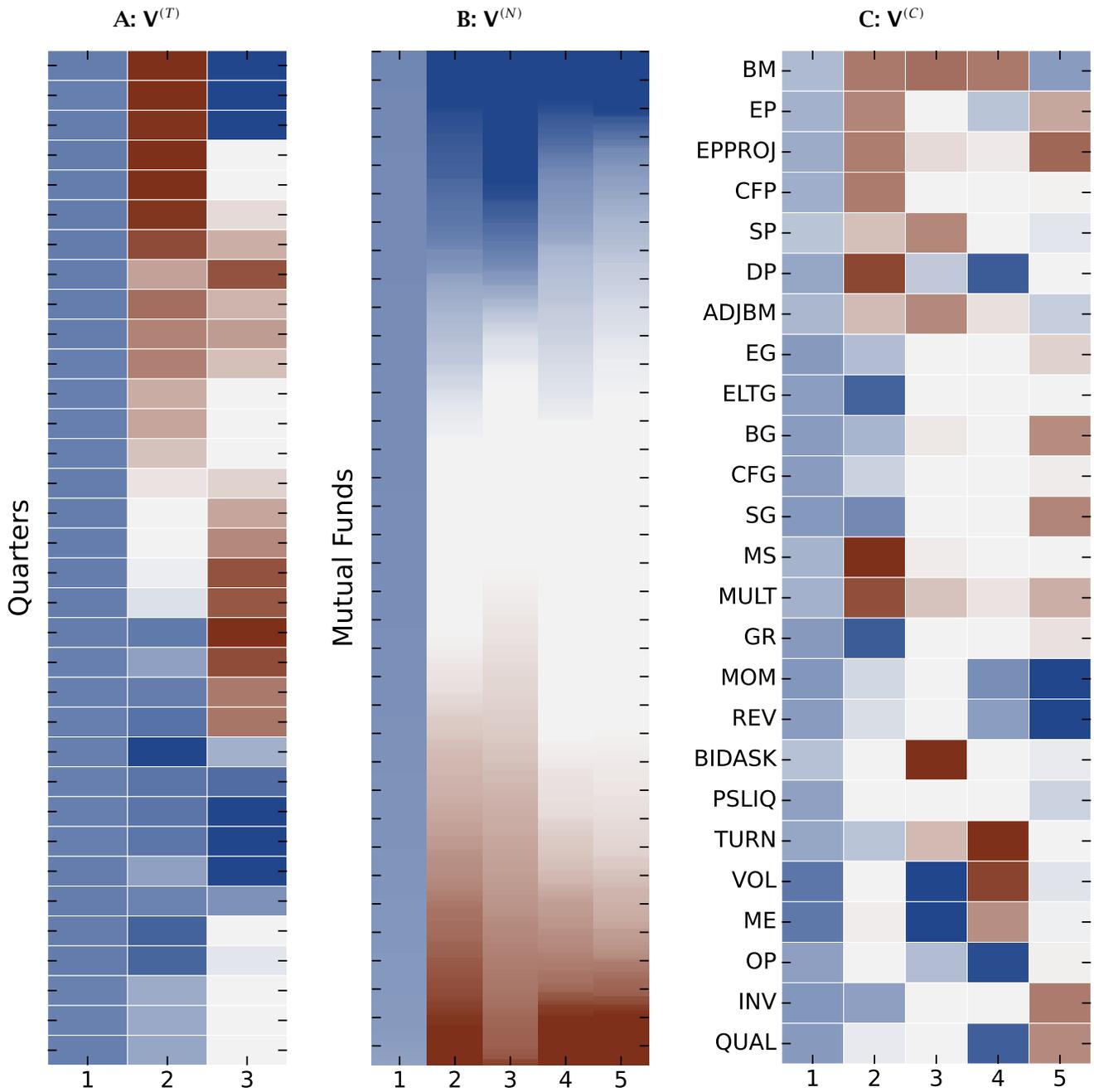
Notes: The figures show time series plots of the observed data and fitted values of the Tucker(10,20,15) model of the book-to-market ratio (Panel A), momentum (Panel B), and reversal (Panel C) of individual mutual funds. The funds in the left, middle, and right columns are the mutual funds that represent the 75th and the 90th percentiles, and the highest MSE of the MSE distribution of funds for a given characteristic. The legends include the wfcin of the plotted fund and the mean square error in parentheses. The sample period is 2010Q3 to 2018Q4.

**Figure 7: Core Tensor of Tucker(3,10,10) Model**



Notes: The figure plots the 20 largest elements by the absolute value of the 3-dimensional core tensor  $\mathcal{G}$  of the Tucker(8,12,12) model on a log scale. The  $x$ -axis shows the indices of  $\mathcal{G}$ . The sample period is 2010Q3 to 2018Q4.

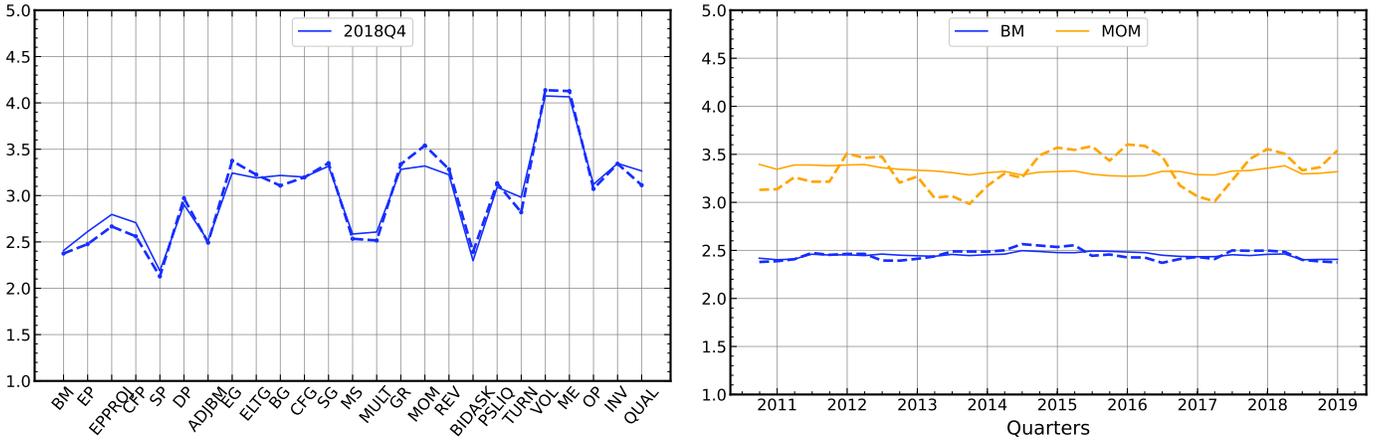
**Figure 8: Loading Matrices of Tucker(3,10,10) Model**



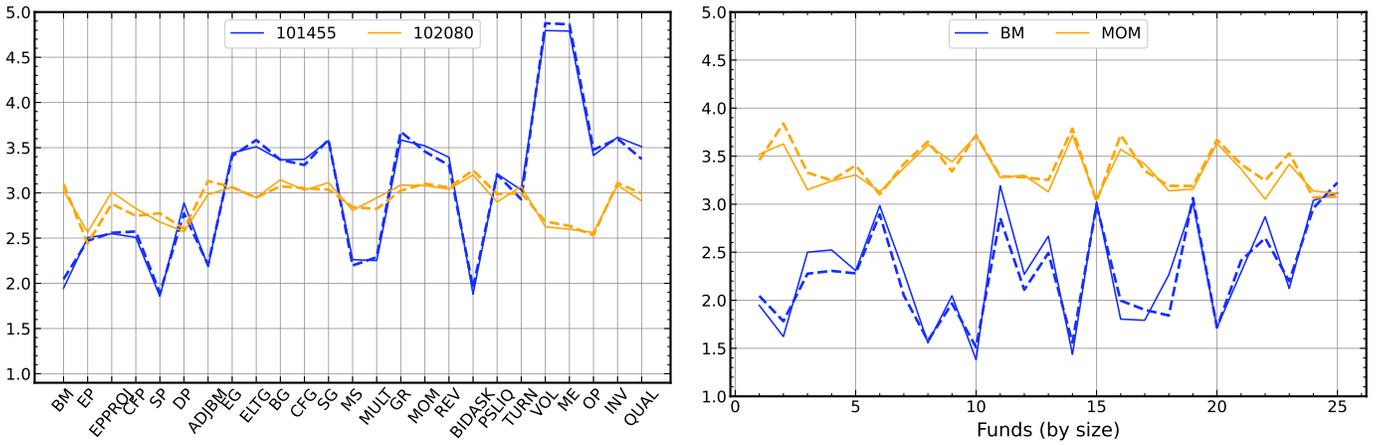
Notes: The figure shows heatmaps of the Tucker component matrices  $\mathbf{V}_i$  of the Tucker(3,10,10) model. Positive values are shown in blue and negative values in red. Panel A shows the  $(34 \times 3)$  component matrices  $\mathbf{V}^{(T)}$ . Each row corresponds to a quarter starting in 2010Q3 at the top to 2018Q4 at the bottom. The columns correspond to the  $K_T = 3$  mode-1 components. The second component matrix  $\mathbf{V}^{(N)}$  has 934 rows and 10 columns. Panel B shows the heatmap of the first five columns of  $\mathbf{V}^{(N)}$ . To visualize 934 rows, I sort each column of  $\mathbf{V}^{(N)}$  so that the first row of each column plots the funds with the largest values at the top and the funds with the smallest values at the bottom. Panel C shows the heatmap of the first five columns of the  $(25 \times 10)$ -dimensional matrix  $\mathbf{V}^{(C)}$ . The sample period is 2010Q3 to 2018Q4.

**Figure 9: 2-dimensional Factor Representations and Data Means**

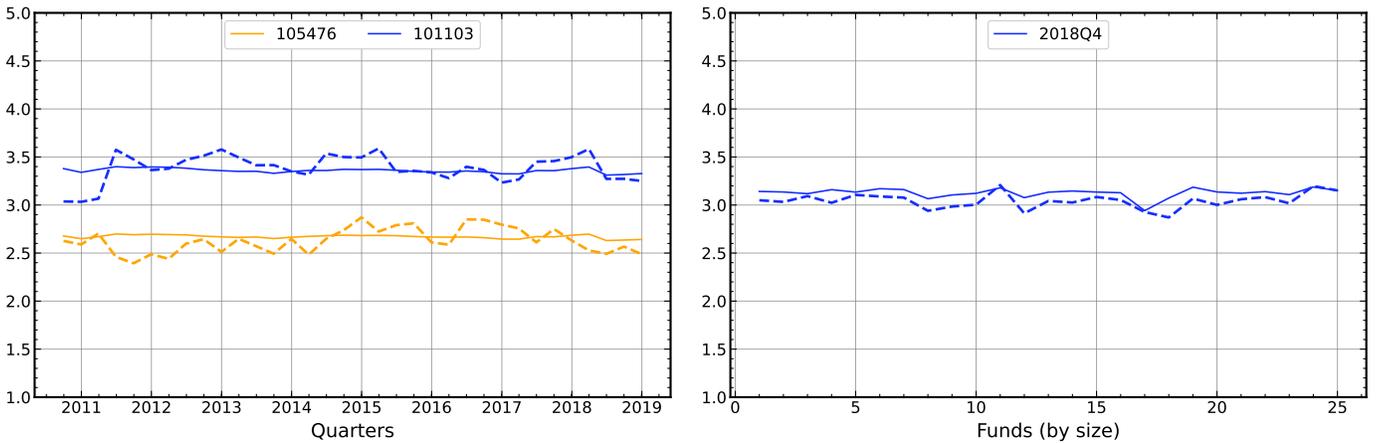
**A:  $F_{TC}^1$  and Means by Mutual Fund**



**B:  $F_{NC}^1$  and Means by Quarter**

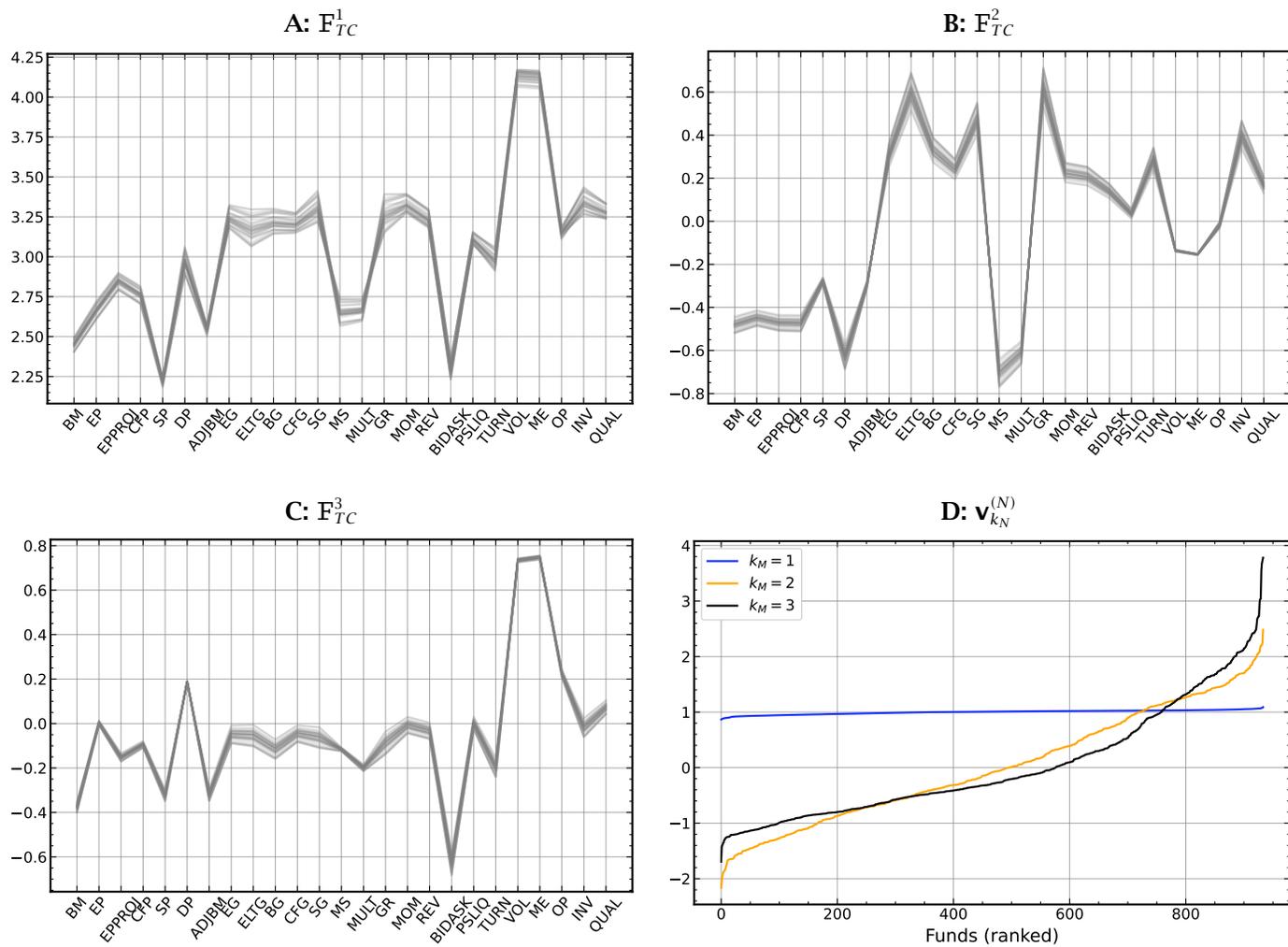


**C:  $F_{TN}^1$  and Means by Characteristic**



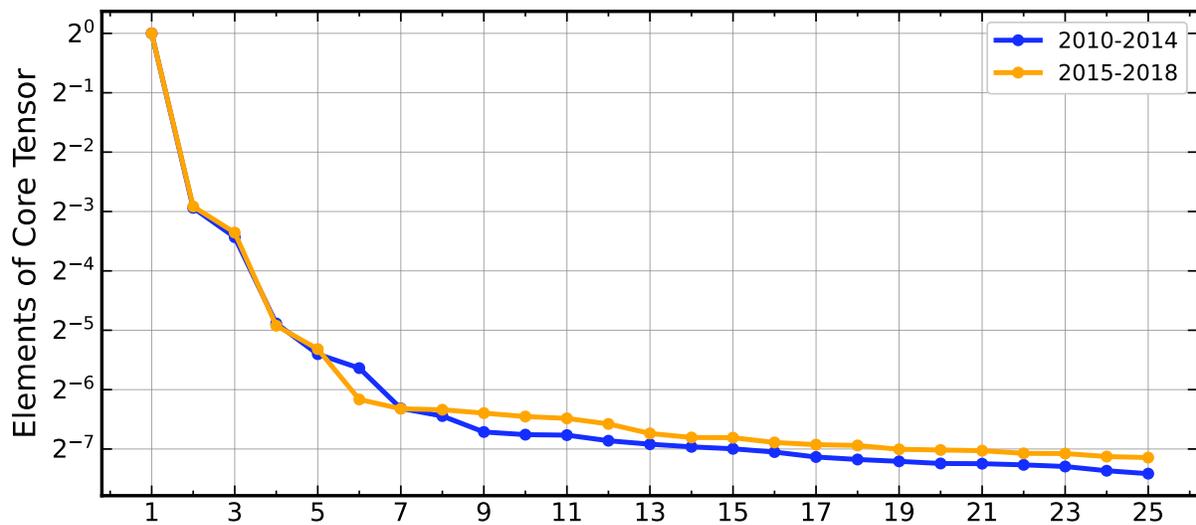
Notes: This figure compares the first factor matrices  $F_{TC}^{(1)}$ ,  $F_{NC}^{(1)}$  and  $F_{TN}^{(1)}$  of a Tucker(2,3,3) model to the corresponding means of the data. Means are plotted as dashed lines while columns and rows of factor matrices are plotted as solid lines. The sample period is 2010Q3 to 2018Q4.

**Figure 10: 2-dimensional Factor Representation  $F_{TC}^{k_N}$**



Notes: Panel A, B, and C plot the rows of the first three factor matrices  $F_{TC}^{(k_N)}$ ,  $k_N = 1, 2, 3$  of a Tucker(2,3,3) model. Panel D shows the three columns of the loading matrix  $V^{(N)}$ . Each line corresponds to a row of  $F_{TC}^{(k_N)}$ . The sample period is 2010Q3 to 2018Q4.

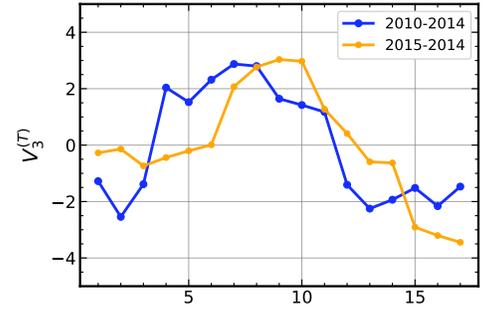
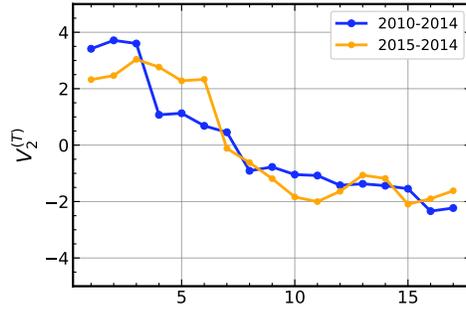
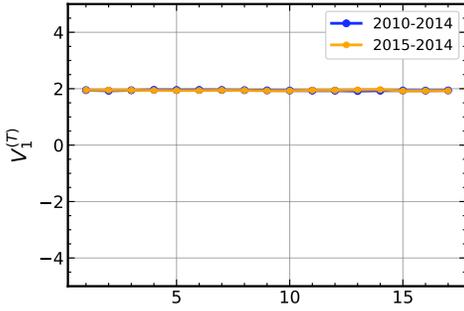
Figure 11: Subsamples - Core Tensors



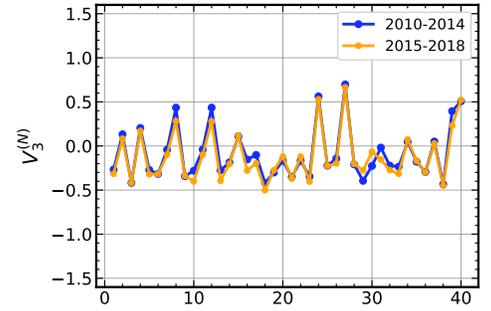
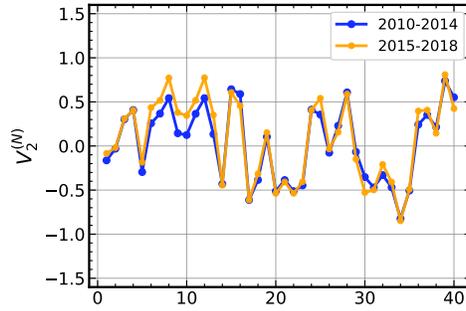
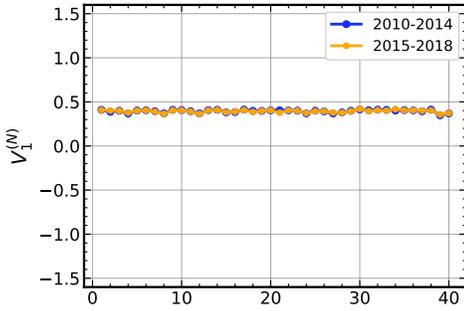
Notes: The figure plots the 20 largest elements by the absolute value of the 3-dimensional core tensors  $\mathcal{G}$  of subsample Tucker(3,12,12) models on a log scale. The Tucker models are estimated over two subsamples consisting of 17 quarters: 2010Q3-2014Q4 and 2015Q1-2018Q4. The blue lines correspond to the first half of the sample and the orange lines to the second half.

**Figure 12: Subsamples - Loading Matrices**

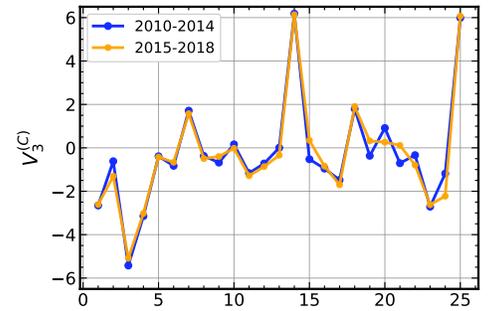
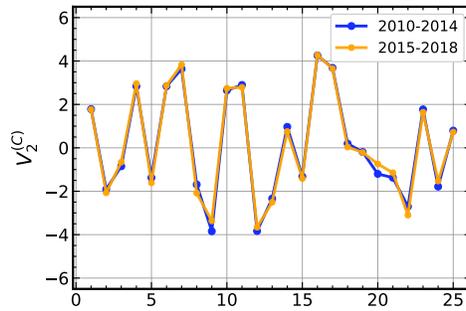
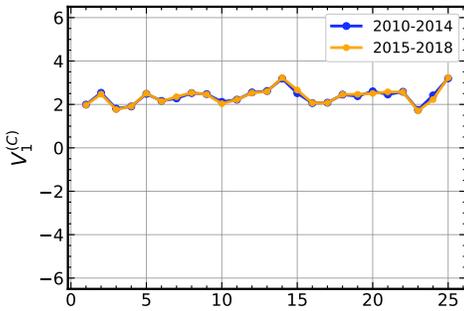
**A:  $V_i^{(T)}$**



**B:  $V_i^{(N)}$**

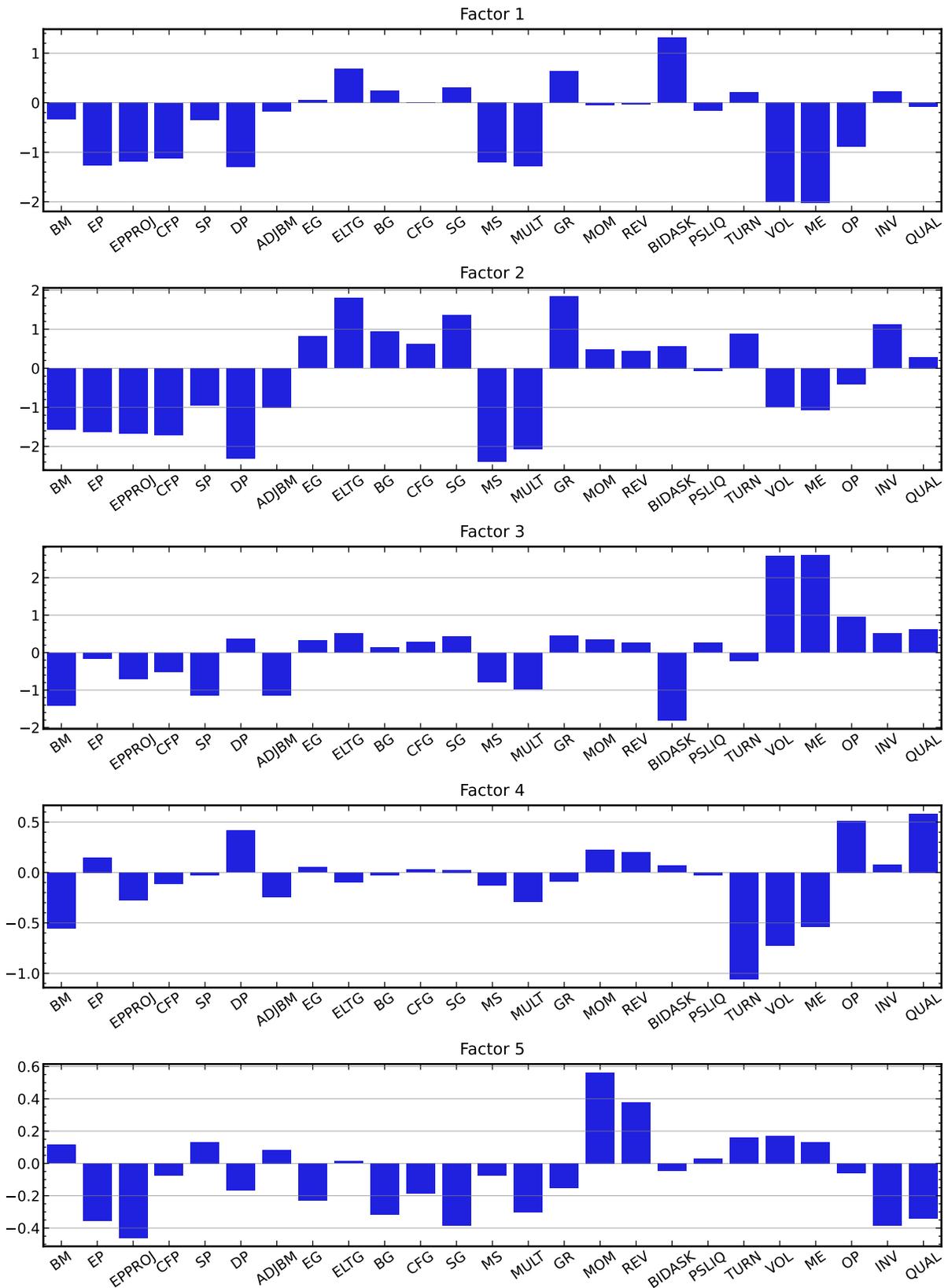


**C:  $V_i^{(C)}$**



Notes: This figure plots the first three loading vectors of mode  $i, \mathbf{V}_j^{(i)}, i = T, N, C, j = 1, 2, 3$  of Tucker models that are estimated over two subsamples consisting of 17 quarters: 2010Q3-2014Q4 and 2015Q1-2018Q4. The Tucker models have  $(K_T, K_N, K_C) = (3, 12, 12)$  components. The blue lines correspond to the first half of the sample and the orange lines to the second half.

**Figure 13: Mean Characteristics of Tucker Factors**



Notes: This figure shows the characteristics of the first five pricing factors  $f_{Tck,t}$  implied by the Tucker(3,10,10) model.

## Appendix A. Tensor operations

Let  $\mathcal{X}$  be a  $(T \times N \times C)$ -dimensional tensor. A 3-dimensional tensor can be expressed as collections of one-dimensional *fibers* and 2-dimensional *slices*. Fibers are vectors and correspond to rows and columns of a matrix, while slices are matrices. Fibers are defined by fixing every index but one so that  $\mathcal{X}$  has fibers along each mode, denoted by  $\mathbf{x}_{(nc)t}$ ,  $\mathbf{x}_{(tc)n}$ , and  $\mathbf{x}_{(tn)c}$ , respectively.<sup>32</sup> Slices are created by fixing all but two indices and are written as  $\mathbf{X}_{(t)nc}$ ,  $\mathbf{X}_{(n)tc}$ ,  $\mathbf{X}_{(c)tn}$ .<sup>33</sup>

A tensor can be written as a matrix by *unfolding* one dimension. For example, unfolding  $\mathcal{X}$  along the first dimension arranges the dimension-1 fibers as columns of the unfolded matrix  $\mathbf{X}_{(1)}$ , which is of dimension  $(T \times NC)$ . Correspondingly, unfolding  $\mathcal{X}$  along the second and third dimensions yields a  $(N \times TC)$ -matrix  $\mathbf{X}_{(2)}$  and a  $(C \times TI)$ -matrix  $\mathbf{X}_{(3)}$ , respectively.

The *inner product* of two tensors of equal dimensions is the sum of the products of the individual tensor elements as follows:

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{t,n,c} x_{tnc} y_{tnc}$$

and the *norm* of  $\mathcal{X}$  is  $\|\mathcal{X}\| = \langle \mathcal{X}, \mathcal{X} \rangle^{1/2}$ . The *outer product*  $\circ$  of two vectors  $\mathbf{a} \in \mathbb{R}^T$  and  $\mathbf{b} \in \mathbb{R}^N$  is defined as<sup>34</sup>

$$\mathbf{X} = \mathbf{a} \circ \mathbf{b} = \mathbf{a} \mathbf{b}^T \in \mathbb{R}^T \times \mathbb{R}^N,$$

so that  $\mathbf{X}$  is a  $(T \times N)$  matrix. The *outer product* of three vectors,  $\mathbf{a} \in \mathbb{R}^T$ ,  $\mathbf{b} \in \mathbb{R}^N$ ,  $\mathbf{c} \in \mathbb{R}^C$ , yields a 3-dimensional  $(T \times N \times C)$  tensor

$$\mathcal{X} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \in \mathbb{R}^T \times \mathbb{R}^N \times \mathbb{R}^C. \quad (\text{A.1})$$

Tensors can be multiplied by vectors and matrices of appropriate dimensions. Since tensors have arbitrary dimensions, the mode that is multiplied by the matrix has to be explicitly specified. The product of a tensor  $\mathcal{X}$  and a matrix  $\mathbf{A}_n$  is called *n-mode multiplication*, where  $n$  specifies the mode that is multiplied by  $\mathbf{A}_n$ . For example, the mode-1 product of the  $(T \times N \times C)$  tensor  $\mathcal{X}$  and the  $(S \times T)$  matrix  $\mathbf{A}_1$  is equal to a  $(S \times N \times C)$  tensor  $\mathcal{Y}$  given by

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{A}_1.$$

The  $n$ -mode product tensor is constructed by multiplying each mode- $n$  fiber by each row vector of  $\mathbf{A}_n$ . In general, the  $n$ -mode is written as  $\mathcal{X} \times_n \mathbf{A}_n$ . The number of columns of  $\mathbf{A}_n$  must equal the  $n$ -mode dimension of  $\mathcal{X}$  while the  $n$ -mode dimension of  $\mathcal{X} \times_n \mathbf{A}_n$  is equal to the number of rows of  $\mathbf{A}_n$ . The  $n$ -mode product can be chained:

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3$$

<sup>32</sup>See Panels B, C, and D of Figure E.2.

<sup>33</sup>See Panels E, F, and G of Figure E.2.

<sup>34</sup>Panel A of Figure E.4 shows an example for  $T = 5, N = 4, C = 3$ .

where  $\mathbf{A}_2$  and  $\mathbf{A}_3$  are conforming matrices. The order of the multiplications in the chain is irrelevant.

The 1-mode product of a  $(2 \times 2 \times 3)$  tensor with a  $(5 \times 2)$  matrix is illustrated in Panel A of Figure E.4. Each mode-1 fiber of  $\mathbf{X}$  is a vector of length 2 and is multiplied by each of the row vectors of  $\mathbf{A}_1$  so that  $\mathbf{X}$  with mode-1 dimension  $T$  is transformed into the product tensor  $\mathbf{Y}$  with mode-1 dimension  $S$ . All other dimensions are the same. Panel C shows an example of a mode-2 product. Note that  $\mathbf{A}_2$  is a  $(2 \times 4)$  matrix but is displayed as a  $(4 \times 2)$  matrix. It is standard practice to rotate tensors, matrices, and vectors in illustrations so that the mode dimensions match.<sup>35</sup>

The standard matrix products can be written in  $n$ -mode tensor notation. Let  $\mathbf{X}, \mathbf{A}_1$ , and  $\mathbf{A}_2$  be  $(T \times N)$ ,  $(S \times T)$ , and  $(U \times N)$  matrices, respectively. Then  $\mathbf{A}_1 \mathbf{X} = \mathbf{X} \times_1 \mathbf{A}_1$  is a  $(S \times N)$  matrix,  $\mathbf{X} \mathbf{A}_2^\top = \mathbf{X} \times_2 \mathbf{A}_2$  is a  $(T \times U)$  matrix, and  $\mathbf{A}_1 \mathbf{X} \mathbf{A}_2^\top = \mathbf{X} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2$  is a  $(S \times U)$  matrix.

The tensor operations are summarized in Table D.1.

## Appendix B. The Singular Value Decomposition (SVD) of a matrix

Let  $\mathbf{X}$  be a  $(T \times N)$  data matrix with  $TN$  observations  $x_{tn}$ .<sup>36</sup> The SVD of  $\mathbf{X}$  is given by

$$\mathbf{X} = \mathbf{U}^{(1)} \mathbf{H} \mathbf{U}^{(2)\top} \quad (\text{B.1})$$

$$= \sum_{i=1}^{\min(M,N)} h_i \mathbf{u}_i^{(1)} \mathbf{u}_i^{(2)\top}, \quad (\text{B.2})$$

where  $\mathbf{U}^{(1)}$  is a  $(T \times T)$  matrix of eigenvectors  $\mathbf{u}_t^{(1)}$  of  $\mathbf{X}\mathbf{X}^\top$  as columns,  $\mathbf{U}^{(2)}$  is a  $(N \times N)$  matrix of eigenvectors  $\mathbf{u}_t^{(2)}$  of  $\mathbf{X}^\top \mathbf{X}$  as columns, and  $\mathbf{H}$  is a diagonal  $(T \times N)$  matrix with diagonal elements  $h_i$  that are the squares roots of non-zero eigenvalues of  $\mathbf{X}\mathbf{X}^\top$ . The eigenvalues are in descending order and the eigenvectors in  $\mathbf{U}^{(1)}$  and  $\mathbf{U}^{(2)}$  are ordered accordingly.

The SVD of  $\mathbf{X}$  implies a *factor representation*

$$\mathbf{X} = \mathbf{F}_N \mathbf{B}_N^\top, \quad (\text{B.3})$$

where  $\mathbf{F}_N = \mathbf{X}\mathbf{U}^{(2)} = \mathbf{U}^{(1)} \mathbf{H}$  and  $\mathbf{B}_N = \mathbf{U}^{(2)}$  are of dimensions  $(T \times N)$  and  $(N \times N)$ , respectively. The columns of  $\mathbf{F}_N$  are *factors*, and the columns of  $\mathbf{B}_N$  are *factor loadings*. The isomorphic factor representation for  $\mathbf{X}^\top$  is given by  $\mathbf{X}^\top = \mathbf{F}_T \mathbf{B}_T^\top$ , where  $\mathbf{F}_T = \mathbf{X}^\top \mathbf{U}^{(1)} = \mathbf{U}^{(2)} \mathbf{H}^\top$  and  $\mathbf{B}_T = \mathbf{U}^{(1)}$ . Hence, the interpretations of factor and loading matrices are reversed in the two representations.

Factor models (B.3) are not unique and can be rotated by any nonsingular  $(N \times N)$  matrix  $\mathbf{S}$ :  $\mathbf{X} = \mathbf{F}_N \mathbf{S} \mathbf{S}^{-1} \mathbf{B}_N^\top$ . Note that it is also possible to compute the SVD of the  $(N \times N)$  matrix  $\mathbf{X}^\top$  instead of  $\mathbf{X}$ . The representations are equivalent, but the roles of  $\mathbf{U}^{(1)}$  and  $\mathbf{U}^{(2)}$  are reversed so that factors of the SVD of  $\mathbf{X}$  become factor loadings in the SVD of  $\mathbf{X}^\top$ , and *vice versa*.

Suppose we want to approximate  $\mathbf{X}$  by a matrix  $\hat{\mathbf{X}}_k$  that can be written in terms of lower-dimensional

<sup>35</sup>There is no “transpose” operator for tensors, and it may be helpful to think about tensor multiplications without the notion of a matrix transpose.

<sup>36</sup>The row index  $t$  is generic and does not necessarily have to be a “time” index.

matrices such that

$$\mathbf{X} = \hat{\mathbf{X}}_K + \mathbf{E}_K, \quad (\text{B.4})$$

$$\text{where } \hat{\mathbf{X}}_K = \tilde{\mathbf{U}}_K^{(1)} \tilde{\mathbf{H}}_K \tilde{\mathbf{U}}_K^{(2)\top}, \quad (\text{B.5})$$

and  $\tilde{\mathbf{H}}_K, \tilde{\mathbf{U}}_K^{(1)}, \tilde{\mathbf{U}}_K^{(2)}$  are  $(K \times K), (T \times K), (N \times K)$  matrices. The optimal  $\hat{\mathbf{X}}_K$  minimizes the mean-squared-error (MSE)

$$\text{MSE}(\hat{\mathbf{X}}_K) = \frac{1}{MN} \|\mathbf{E}_K\|^2,$$

where  $\|\mathbf{E}\| = \sqrt{\sum_{t,n} e_{tn}^2}$  is the Frobenius matrix norm. Eckart and Young (1936) showed that the solution is given by the *truncated SVD*, i.e., setting  $\tilde{\mathbf{H}}_K$  to the first  $K$  rows and columns of  $\mathbf{H}$  and  $\tilde{\mathbf{U}}_K^{(1)}, \tilde{\mathbf{U}}_K^{(2)}$  to the first  $K$  columns of  $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}$ :

$$\hat{\mathbf{X}}_K = \mathbf{U}_K^{(1)} \mathbf{H}_K \mathbf{U}_K^{(2)\top}. \quad (\text{B.6})$$

The truncated SVD (B.6) is equivalent to the  $K$ -factor model

$$\mathbf{X} = \mathbf{F}_K \mathbf{B}_K^\top + \mathbf{E}_K, \quad (\text{B.7})$$

where  $\mathbf{F}_K = \mathbf{U}_K^{(1)} \mathbf{H}_K$  and  $\mathbf{B}_K = \mathbf{U}_K^{(2)}$  are  $(T \times K)$  and  $(N \times K)$  matrices, respectively. Thus, the truncated SVD is equal to the first  $K$  principal components of  $\mathbf{X}^\top \mathbf{X}$ . I will refer to this model as SVD-PCA throughout the paper.

The truncated SVD has an alternative representation that is useful for understanding tensor decompositions.  $\mathbf{U}_K^{(1)} \mathbf{H}_K \mathbf{U}_K^{(2)\top}$  is equivalent to the weighted sum of the outer products of the column vectors of  $\mathbf{U}_K^{(1)}$  and the row vectors of  $\mathbf{U}_K^{(2)\top}$ . This can be seen by writing (B.6) as

$$\hat{\mathbf{X}}_K = \sum_{t=1}^K \sum_{n=1}^K h_{tn} \underbrace{\mathbf{u}_t^{(1)} \mathbf{u}_n^{(2)\top}}_{T \times N} \quad (\text{B.8})$$

$$= \sum_{k=1}^K h_{kk} \mathbf{u}_k^{(1)} \mathbf{u}_k^{(2)\top}. \quad (\text{B.9})$$

The second equality follows from the fact that  $\mathbf{H}_K$  is a diagonal matrix.  $\hat{\mathbf{X}}_K$  is the weighted sum of  $K$  matrices with dimensions  $(T \times N)$ , which are the outer vector product of the eigenvectors  $\mathbf{u}_k^{(1)}$  and  $\mathbf{u}_k^{(2)\top}$  of  $\mathbf{X}\mathbf{X}^\top$  and  $\mathbf{X}^\top \mathbf{X}$ , respectively. Each  $k$  in the summation corresponds to a factor in the  $K$ -factor representation (B.7). The advantage of representation (B.8) is that it shows the contribution of each of the  $K$  factors in the fit of the model. Since the eigenvectors are normalized, the  $K$  outer vector products  $\mathbf{u}_k^{(1)} \mathbf{u}_k^{(2)\top}$  are of the same magnitude, so the weight of the contribution of each factor  $k$  is approximately equal to the  $k$ -th eigenvalue.

In the truncated SVD (B.4)-(B.6) the number of factors is  $K$ . Note that we could define an asymmetric SVD that has different numbers of factors for the two dimensions:

$$\hat{\mathbf{X}}_{(K_1, K_2)} = \mathbf{U}_{K_1}^{(1)} \mathbf{H}_{K_1, K_2} \mathbf{U}_{K_1}^{(2)\top}, \quad (\text{B.10})$$

where  $\mathbf{H}_{K_1, K_2}$ ,  $\mathbf{U}_{K_1}^{(1)}$ ,  $\mathbf{U}_{K_2}^{(2)}$  are  $(K_1 \times K_2)$ ,  $(N \times K_1)$ ,  $(N \times K_2)$  matrices. However, since  $\mathbf{H}_{K_1, K_2}$  is diagonal, the asymmetric SVD reduces to a  $K$ -factor SVD where  $K = \min(K_1, K_2)$ . In contrast to the 2-dimensional matrix SVD, the core tensor  $\mathcal{G}$  in the Tucker decomposition is *not* diagonal. Consequently, the number of factors can differ by dimension.<sup>37</sup>

### Appendix C. Higher-Order Orthogonal Iteration (HOOI)

The objective is to find  $\mathcal{G}$  and orthonormal  $\mathbf{V}^{(T)}, \mathbf{V}^{(N)}, \mathbf{V}^{(C)}$  such that

$$\|\boldsymbol{\varepsilon}\| = \|\boldsymbol{\mathcal{X}} - \mathcal{G} \times_1 \mathbf{V}^{(T)} \times_2 \mathbf{V}^{(N)} \times_3 \mathbf{V}^{(C)}\|$$

is minimized. Given the loading matrices  $\mathbf{V}^{(i)}$ , the optimal core tensor  $\mathcal{G}$  satisfies

$$\mathcal{G} = \boldsymbol{\mathcal{X}} \times_1 \mathbf{V}^{(T)\top} \times_2 \mathbf{V}^{(N)\top} \times_3 \mathbf{V}^{(C)\top}. \quad (\text{C.1})$$

Since the  $\mathbf{V}^{(i)}$  matrices are orthonormal, the squared norm of the approximation error  $\boldsymbol{\varepsilon} = \boldsymbol{\mathcal{X}} - \hat{\boldsymbol{\mathcal{X}}}$  can be written as

$$\|\boldsymbol{\varepsilon}\|^2 = \|\boldsymbol{\mathcal{X}}\|^2 - 2\langle \boldsymbol{\mathcal{X}}, \mathcal{G} \times_1 \mathbf{V}^{(T)} \times_2 \mathbf{V}^{(N)} \times_3 \mathbf{V}^{(C)} \rangle + \|\mathcal{G} \times_1 \mathbf{V}^{(T)} \times_2 \mathbf{V}^{(N)} \times_3 \mathbf{V}^{(C)}\|^2 \quad (\text{C.2})$$

$$= \|\boldsymbol{\mathcal{X}}\|^2 - 2\langle \boldsymbol{\mathcal{X}} \times_1 \mathbf{V}^{(T)\top} \times_2 \mathbf{V}^{(N)\top} \times_3 \mathbf{V}^{(C)\top}, \mathcal{G} \rangle + \|\mathcal{G}\|^2 \quad (\text{C.3})$$

$$= \|\boldsymbol{\mathcal{X}}\|^2 - 2\langle \mathcal{G}, \mathcal{G} \rangle + \|\mathcal{G}\|^2 \quad (\text{C.4})$$

$$= \|\boldsymbol{\mathcal{X}}\|^2 - \|\mathcal{G}\|^2 \quad (\text{C.5})$$

$$= \|\boldsymbol{\mathcal{X}}\|^2 - \|\boldsymbol{\mathcal{X}} \times_1 \mathbf{V}^{(T)\top} \times_2 \mathbf{V}^{(N)\top} \times_3 \mathbf{V}^{(C)\top}\|^2. \quad (\text{C.6})$$

Suppose we know  $\mathbf{V}^{(T)}$  and  $\mathbf{V}^{(N)}$ . Then  $\mathbf{V}^{(C)}$  can be obtained as

$$\max_{\mathbf{V}^{(C)}} \|\boldsymbol{\mathcal{X}} \times_1 \mathbf{V}^{(T)\top} \times_2 \mathbf{V}^{(N)\top} \times_3 \mathbf{V}^{(C)\top}\|. \quad (\text{C.7})$$

This maximization problem can be rewritten in matrix form as

$$\max_{\mathbf{V}^{(C)}} \|\mathbf{V}^{(C)\top} \mathbf{W}_C\| \quad (\text{C.8})$$

$$\text{where } \mathbf{W}_C = \mathbf{X}_{(C)} (\mathbf{V}^{(T)} \otimes \mathbf{V}^{(N)}), \quad (\text{C.9})$$

The matrix  $\mathbf{X}_{(C)}$  is the unfolded tensor  $\boldsymbol{\mathcal{X}}$  in the  $C$ -dimension, and  $\otimes$  is the Kronecker matrix product. The optimal  $\mathbf{V}^{(C)}$  is given by the first  $K_C$  eigenvectors of  $\mathbf{W}_C \mathbf{W}_C^\top$ , or, equivalently, by the first  $K_C$  left singular vectors of  $\mathbf{W}_C$ .

Since one  $\mathbf{V}^{(i)}$  can be computed if the other two are known, we can use the following recursive algorithm known as Higher-Order Orthogonal Iteration (HOOI):

1. Pick initial values for  $\mathbf{V}^{(T)}, \mathbf{V}^{(N)}$ .

---

<sup>37</sup>The CP tensor decomposition is a special case of the Tucker decomposition and imposes the restriction that the core tensor  $\mathcal{G}$  is diagonal, which implies that the number of factors is the same,  $K_i = K$ .

2. Compute  $\mathbf{V}^{(C)}$  as the first  $K_C$  left singular vectors of  $\mathbf{X}_{(C)}(\mathbf{V}^{(T)} \otimes \mathbf{V}^{(N)})$ .
3. Compute  $\mathbf{V}^{(T)}$  as the first  $K_T$  left singular vectors of  $\mathbf{X}_{(T)}(\mathbf{V}^{(N)} \otimes \mathbf{V}^{(C)})$ .
4. Compute  $\mathbf{V}^{(N)}$  as the first  $K_N$  left singular vectors of  $\mathbf{X}_{(N)}(\mathbf{V}^{(T)} \otimes \mathbf{V}^{(C)})$ .
5. Repeat Steps 2. to 4. recursively until a convergence criterion is satisfied.
6. Compute  $\mathcal{G} = \mathcal{X} \times_1 \mathbf{V}^{(T)\top} \times_2 \mathbf{V}^{(M)\top} \times_3 \mathbf{V}^{(C)\top}$ .

The literature has developed numerous numerical improvements of the HOOI estimator, see, for example, Andersson and Bro (1998). Several other algorithms exist, including nonlinear Newton-Grassmann optimization (Elden and Savas (2009)). Starting values of the  $\mathbf{V}^{(i)}$  matrices are determined by applying 2-dimensional SVD to unfolded matrices of the  $\mathcal{X}$  tensor. For example, unfold  $\mathcal{X}$  along the first dimension, which yields a  $(T \times NC)$ -dimensional matrix  $\mathbf{X}_{(T)}$ . The initial  $\mathbf{V}^{(T)}$  can be chosen as the first  $K_T$  left singular vectors of  $\mathbf{X}_{(T)}$ . Initial  $\mathbf{V}^{(N)}$  and  $\mathbf{V}^{(C)}$  can be set accordingly.

For the data set used in this paper, the HOOI estimator converges after 20 to 40 iterations. In addition to setting the initial  $\mathbf{V}^{(i)}$  using the method described above, I also choose initial values randomly. The numerical computations are robust and converge to the same optimum.

**Table D.1:** Summary of tensor notation and operations

Operation	2-dimensional matrix	3-dimensional tensor	$n$ -dimensional tensor
	$\mathbf{X} = [\mathbf{x}_{ij}]$	$\boldsymbol{\mathcal{X}} = [\mathbf{x}_{ijk}]$	$\boldsymbol{\mathcal{X}} = [\mathbf{x}_{i_1, \dots, i_j}]$
Fibers	$\mathbf{x}_{(i)j}, \mathbf{x}_{(j)i}$	$\boldsymbol{\mathcal{X}}_{(jk)i}, \boldsymbol{\mathcal{X}}_{(ik)j}, \boldsymbol{\mathcal{X}}_{(ij)k}$	$\boldsymbol{\mathcal{X}}_{(j \neq i)i}$
Slices		$\mathbf{X}_{(i)jk}, \mathbf{X}_{(j)ik}, \mathbf{X}_{(k)ij}$	$\mathbf{X}_{(i)i \neq j}$
Matricization		$\mathbf{X}_{(1)}: I \times J \times K \rightarrow I \times JK$ $\mathbf{X}_{(2)}: I \times J \times K \rightarrow J \times IK$ $\mathbf{X}_{(3)}: I \times J \times K \rightarrow K \times IJ$	$\boldsymbol{\mathcal{X}}_{(p)}: (I_1 \times \dots \times I_j) \rightarrow I_p \times (\prod_{i \neq p} I_i)$
Inner product	$\mathbf{x}^\top \mathbf{y} = \langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{ij} \mathbf{x}_{ij} \mathbf{y}_{ij}$	$\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \rangle = \sum_{i,j,k} \mathbf{x}_{ijk} \mathbf{y}_{ijk}$	$\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \rangle = \sum_{i_1, \dots, i_j} \mathbf{x}_{i_1, \dots, i_j} \mathbf{y}_{i_1, \dots, i_j}$
Outer product	$\mathbf{xy}^\top = \mathbf{x} \circ \mathbf{y}$	$\mathbf{x} \circ \mathbf{y} \circ \mathbf{z}$	$\mathbf{x}_1 \circ \dots \circ \mathbf{x}_j$
Norm	$\ \mathbf{X}\  = \langle \mathbf{X}, \mathbf{X} \rangle = \sqrt{\sum_{ij} \mathbf{x}_{ij}^2}$	$\ \boldsymbol{\mathcal{X}}\  = \langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}} \rangle = \sqrt{\sum_{ijk} \mathbf{x}_{ijk}^2}$	$\ \boldsymbol{\mathcal{X}}\  = \langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}} \rangle = \sqrt{\sum_{i_1, \dots, i_N} \mathbf{x}_{i_1, \dots, i_j}^2}$
$n$ -mode multiplication	$\mathbf{A}_1 \mathbf{X} \mathbf{A}_2^\top = \boldsymbol{\mathcal{X}} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2$	$\boldsymbol{\mathcal{X}} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3$	$\boldsymbol{\mathcal{X}} \times_1 \mathbf{A}_1 \times_2 \dots \times_n \mathbf{A}_n$
Decompositions	$\mathbf{U}_{1K} \mathbf{H}_K \mathbf{U}_{2K}^\top = \mathbf{H}_K \times_1 \mathbf{U}_{1K} \times_2 \mathbf{U}_{2K}$ $= \sum_{k=1}^K h_k \mathbf{u}_{1k} \mathbf{u}_{2k}^\top$	$\boldsymbol{\mathcal{G}} \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \times_3 \mathbf{V}_3$ $= \sum_{k=1}^K g_k \mathbf{w}_{1k} \circ \mathbf{w}_{2k} \circ \mathbf{w}_{3k}$	$\boldsymbol{\mathcal{G}} \times_1 \mathbf{V}_1 \times_2 \dots \times_n \mathbf{V}_n$ $= \sum_{k=1}^K g_k \mathbf{w}_{1k} \circ \dots \circ \mathbf{w}_{nk}$

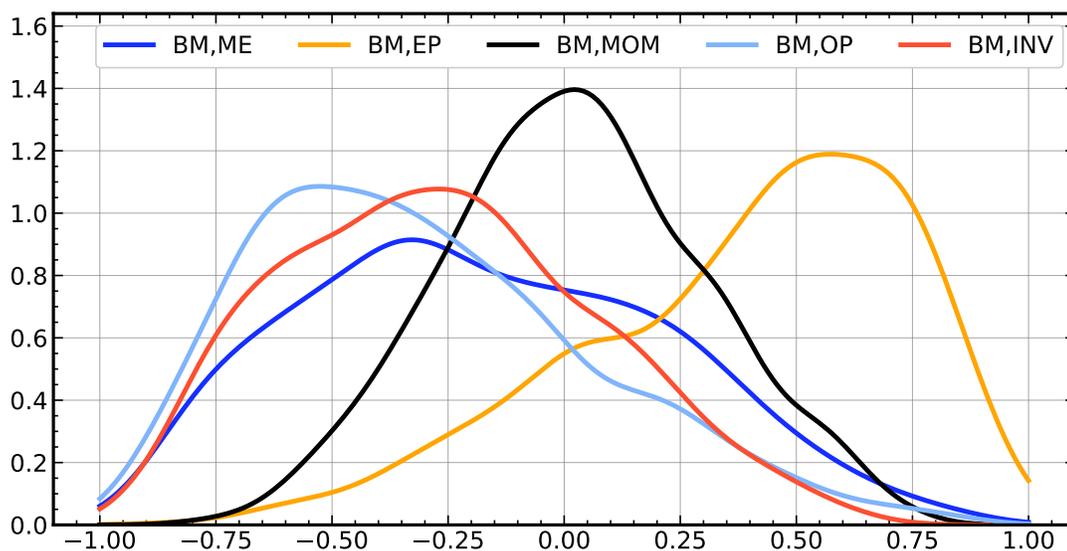
**Table D.2:** Absolute Pricing Errors and Fund Characteristics

	$\bar{R}_m - \bar{\bar{R}}_m$	Tucker IS	Tucker OOS	CAPM	FF3	PCA IS	PCA OOS
const	1.89***	1.46***	1.39***	10.73***	1.87***	1.66***	2.01***
ADJBM	-1.03**	0.06	0.33***	-0.46**	-0.35**	-0.26*	0.25
BG	0.45	-0.34*	0.33*	-0.62**	0.81***	0.35	0.56**
BIDASK	0.99***	0.39**	0.20	1.51***	0.09	0.64***	-0.26
BM	-0.69**	0.32	0.14	-0.33	0.88***	0.42*	0.69**
CFG	-1.79***	0.36*	0.01	-0.69**	-0.84***	-0.31	-0.41
CFP	0.85***	-0.02	0.12	-2.46***	0.81***	0.60**	1.31***
DP	-0.34	0.04	0.06	-1.33***	-0.10	0.47***	0.15
EG	0.95***	0.03	0.02	0.71**	-0.31	-0.16	-1.11***
ELTG	-2.97***	0.86*	0.38	1.11	-2.14***	-0.85	-3.69***
EP	-0.82***	-0.13	-0.43**	0.32	-1.46***	-0.44**	-1.94***
EPPROJ	-2.40***	-0.72*	-0.55	-1.06*	-1.36***	-0.24	-0.14
GR	1.03	-1.95**	-1.39*	-3.56***	0.56	1.29	4.75***
INV	-0.91***	0.72***	0.34	0.77**	-0.08	-0.01	0.61*
ME	4.76***	-0.94**	-0.99***	0.60	-2.81***	-2.11***	-1.57***
MOM	1.06***	0.08	-0.81***	0.32	-0.08	0.16	-0.32
MS	-6.23***	-2.23**	-2.10**	0.32	-5.86***	-1.14	-2.51**
MULT	6.08***	2.01**	1.82*	2.88*	5.32***	0.80	1.97
OP	-0.56***	0.05	0.33***	0.54***	0.22	-0.16	0.15
PSLIQ	-0.57***	0.02	-0.09	0.13	-0.39**	-0.28*	-0.27
QUAL	-0.81***	0.04	0.07	-0.02	0.40***	0.30***	0.37***
REV	-1.39***	-0.24	0.73***	0.61	-0.03	-0.44	0.19
SG	-0.07	-0.32	-0.42	1.27***	-0.87**	-0.19	-1.69***
SP	-0.70***	-0.02	-0.21	-0.07	-0.72***	-0.13	-0.47***
TURN	-0.10	0.12	0.04	-0.21*	-0.01	0.03	0.04
VOL	6.86***	1.41***	1.43***	0.42	3.96***	3.34***	2.32***
$R^2$	0.49	0.14	0.10	0.49	0.41	0.28	0.39

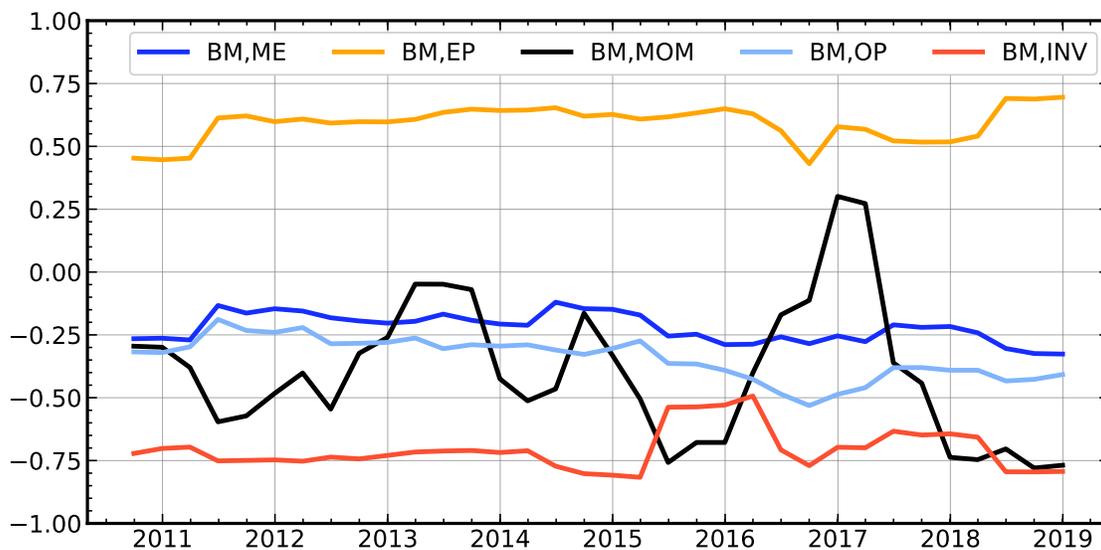
This table reports the results of regressions of pricing errors from factor models on the average expense ratio (Exp. ratio), the average number of stocks (No. stocks), and the average active share. The variables are standardized to have means of zero and unit standard deviations. The factors in the models are those listed in Tables 7 and 8 for  $L = 3$ . Statistical significance at the 10%, 5%, and 1% levels is indicated by one, two, and three stars, respectively. The sample period is 2010Q3 to 2018Q4.

**Figure E.1: Time-series and cross-sectional correlations**

**A: Time-series correlations across mutual funds**



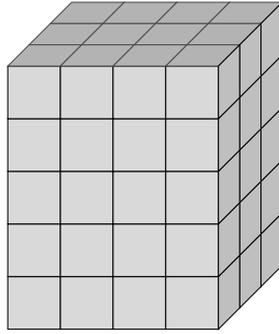
**B: Cross-correlations over time**



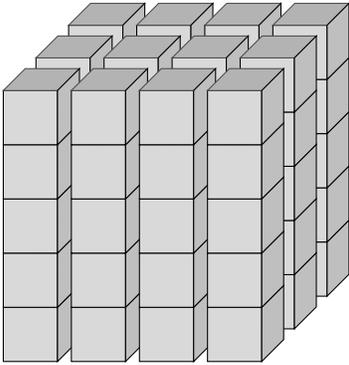
Notes: Panel A shows the distributions of pairwise time series correlations of (BM, ME), (BM, EP), (BM, MOM), (BM, OP), and (BM, INV) across mutual funds. Panel B shows the corresponding pairwise cross-sectional correlations over time. The sample period is 2010Q3 to 2018Q4.

**Figure E.2: Tensor fibers and slices**

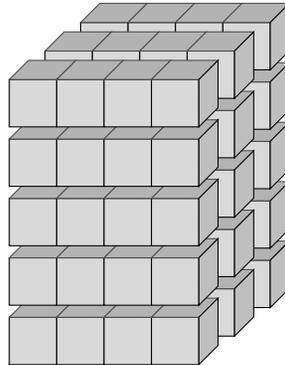
**A: Tensor  $\mathcal{X}$ : (5×4×3)**



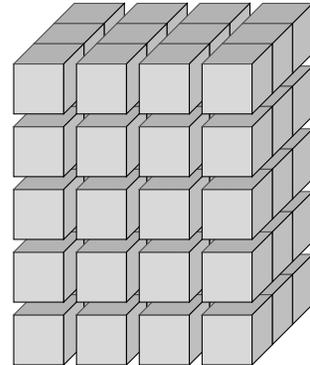
**B: Mode-1 fibers  $\mathbf{x}_{(nc)t}$**



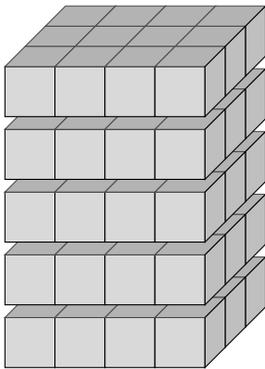
**C: Mode-2 fibers  $\mathbf{x}_{(tc)n}$**



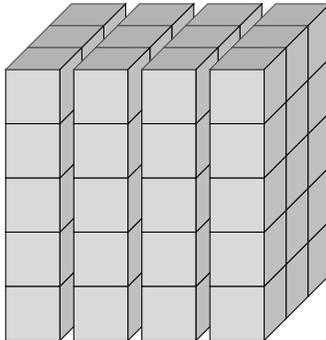
**D: Mode-3 fibers  $\mathbf{x}_{(tn)c}$**



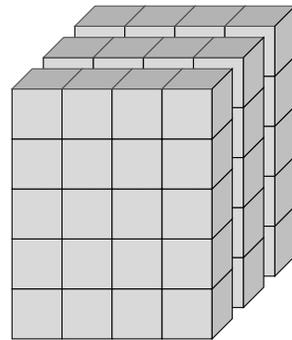
**E: Horizontal slices  $\mathbf{X}_{(t)nc}$**



**F: Lateral slices  $\mathbf{X}_{(n)tc}$**

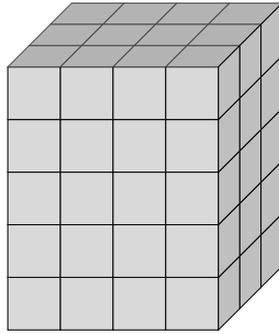


**G: Frontal slices  $\mathbf{X}_{(c)tn}$**

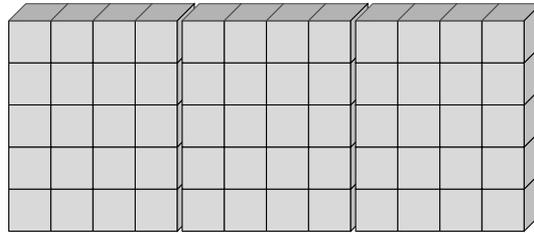


**Figure E.3: Tensor as matrices**

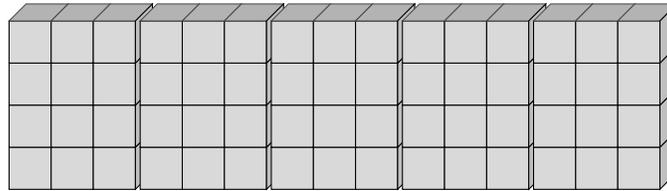
**A: Tensor  $\mathcal{X}$ : (5×4×3)**



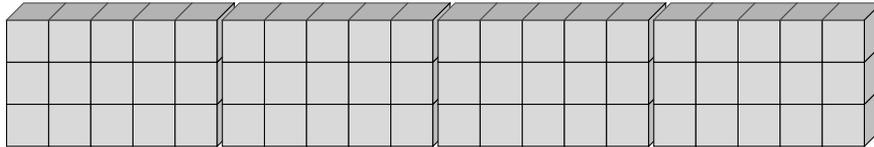
**B:  $\mathbf{X}_{(1)}$ : (5×12)**



**C:  $\mathbf{X}_{(N)}$ : (4×15)**

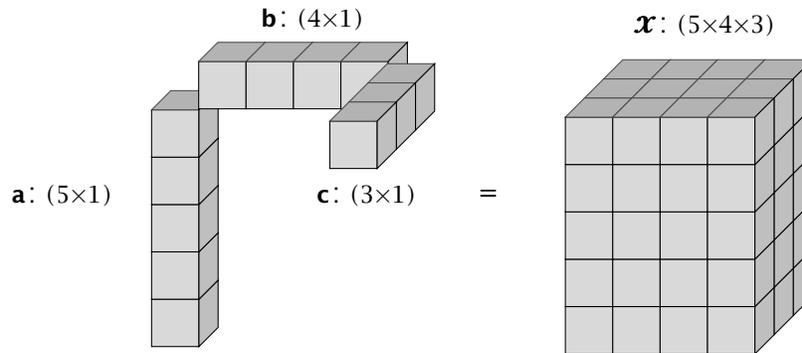


**D:  $\mathbf{X}_{(C)}$ : (3×20)**

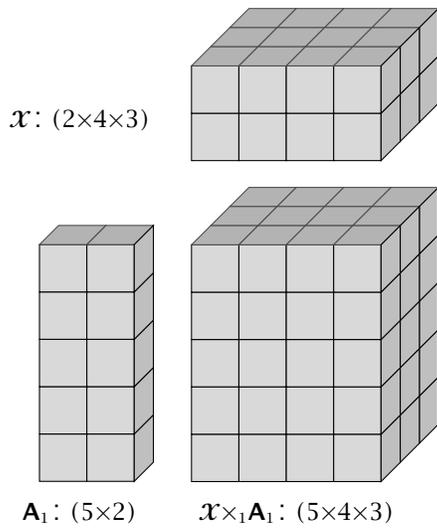


**Figure E.4: Tensor multiplication**

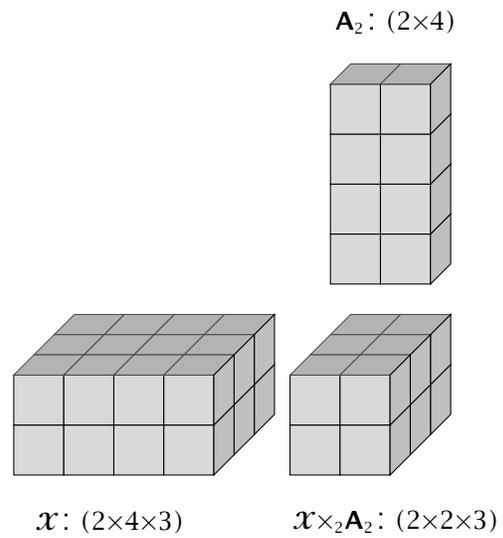
**A: Outer product  $\mathcal{X} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$**



**B: 1-mode product**

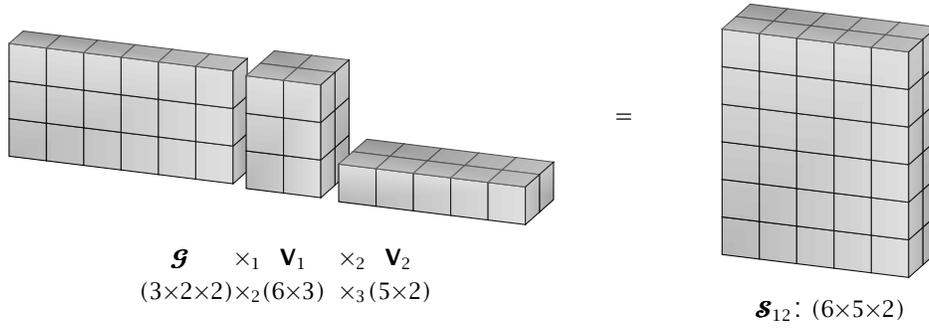


**C: 2-mode product**

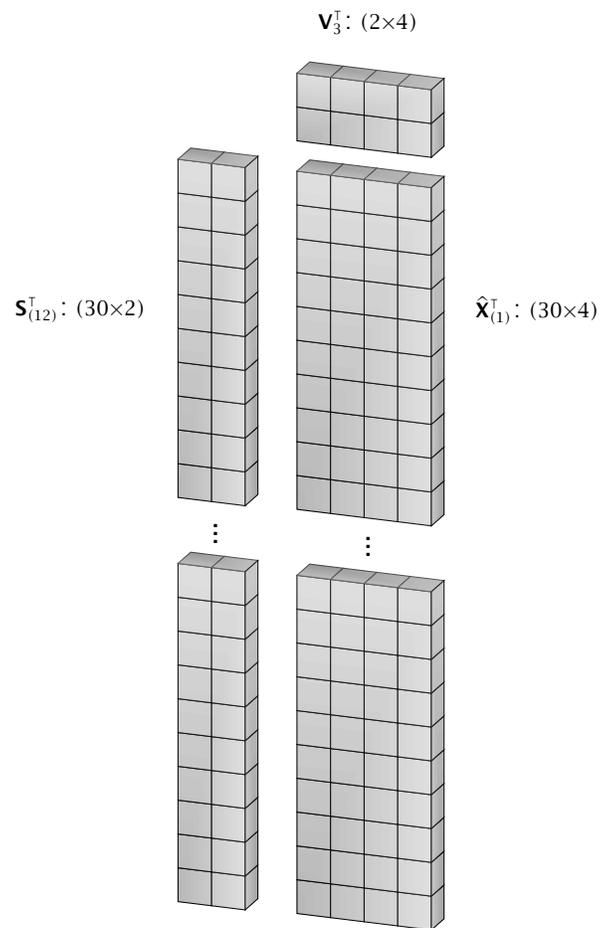


**Figure E.5:** Tucker model as 2-dimensional factor model

$$\mathbf{A}: \mathcal{G} \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \rightarrow \mathcal{S}_{12}$$

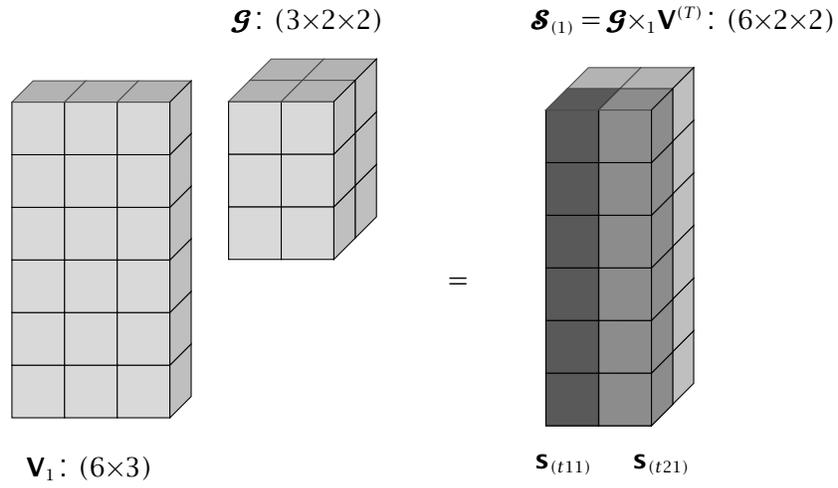


$$\mathbf{B}: \hat{\mathbf{X}}_{(C)}^T = \mathbf{S}_{(12)}^T \mathbf{V}_3^T$$

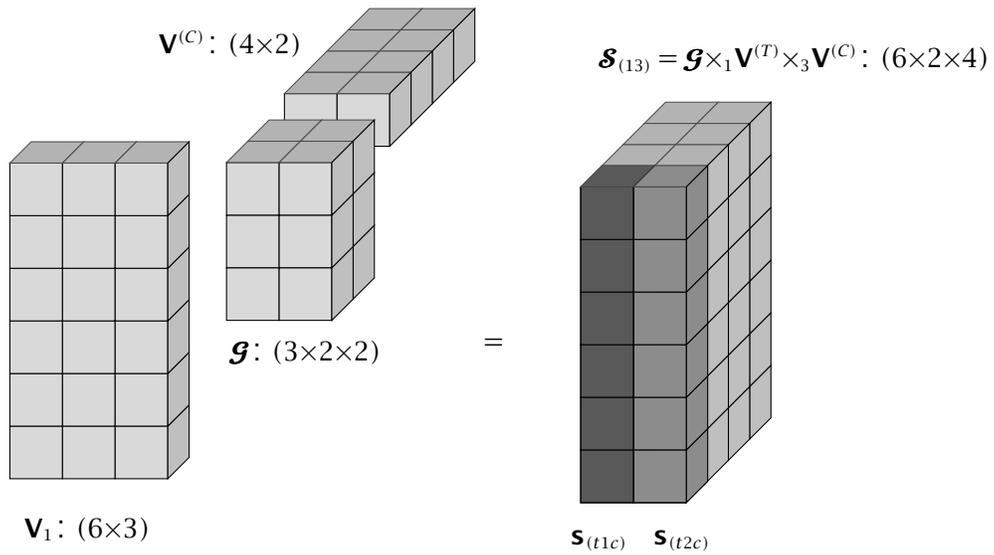


**Figure E.6:** Tucker model: Intuition

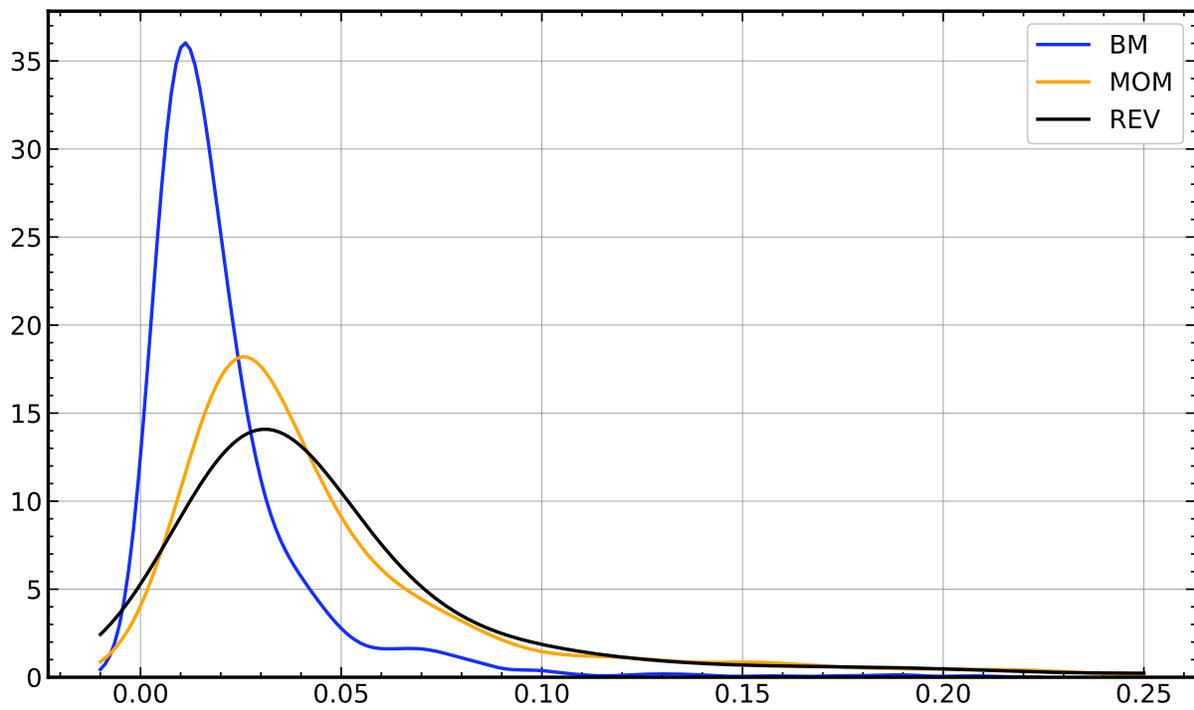
**A:** “summary” modes-(2,3) ( $M, C$ )



**B:** “summary” mode-2 ( $M$ )



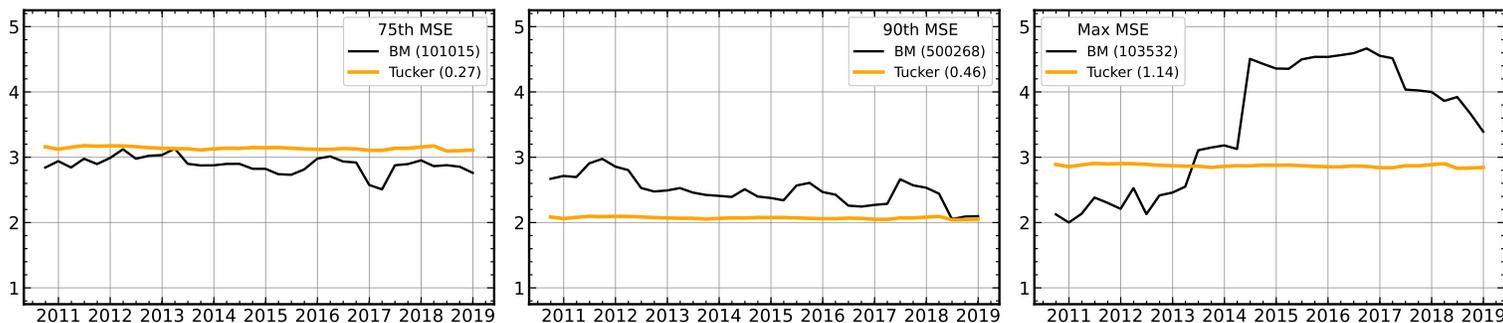
**Figure E.7: MSE Histograms of Tucker(10,25,15) Model**



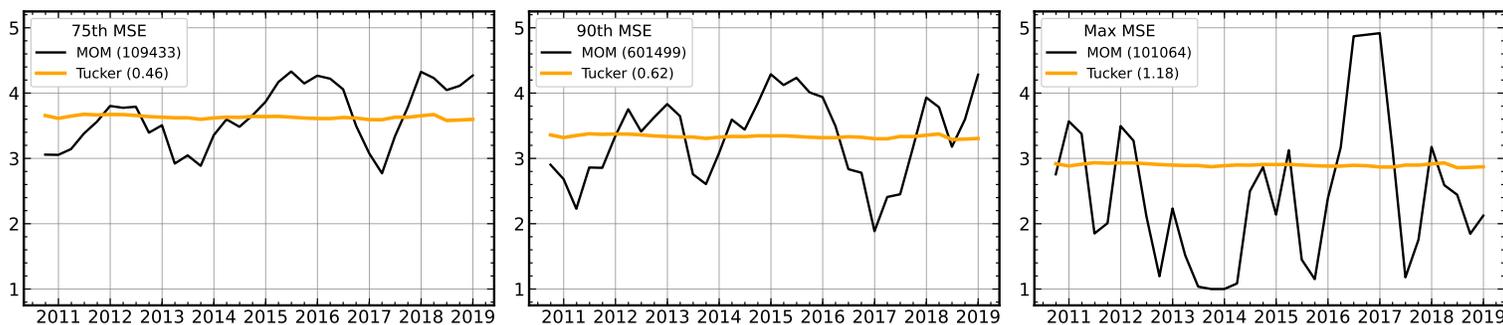
Notes: The figure shows histograms of errors of the Tucker(10,20,15) model for the book-to-market ratio, momentum, and reversals. The sample period is 2010Q3 to 2018Q4.

**Figure E.8: Fit of Tucker(1,4,4) Model of Individual Funds**

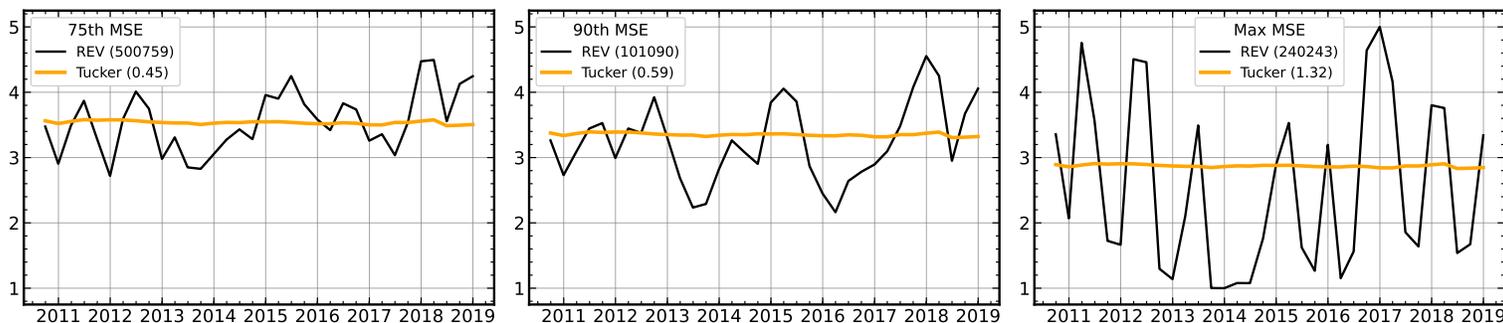
**A: BM**



**B: MOM**



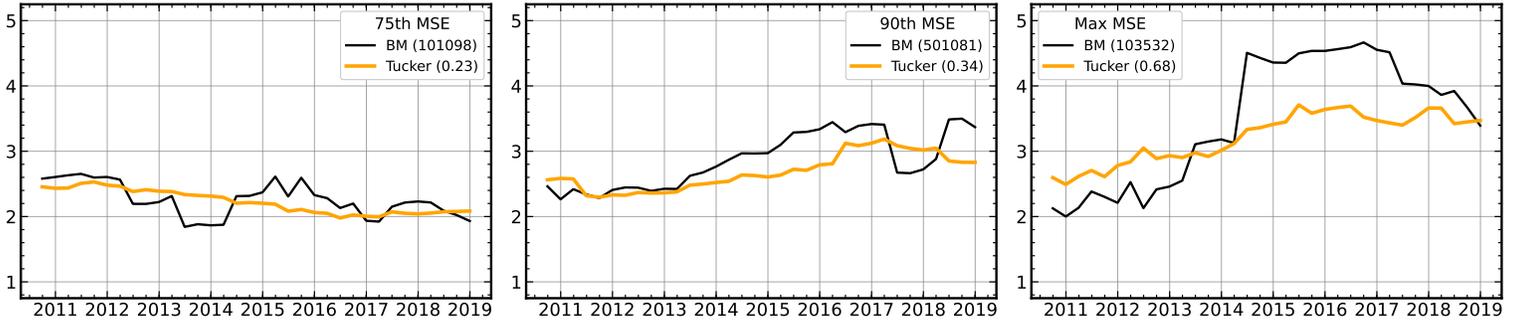
**C: REV**



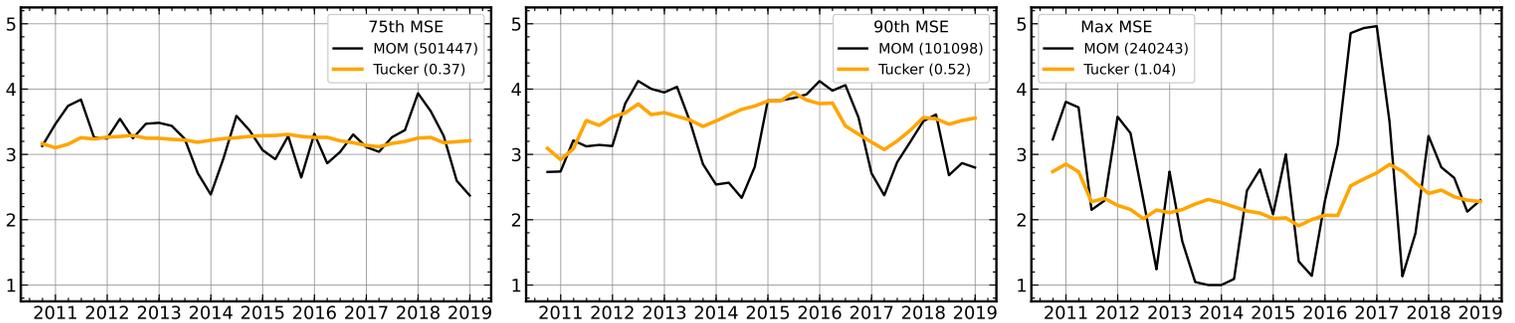
Notes: The figures show time series plots of the observed data and fitted values of the Tucker(1,4,4) model of the book-to-market ratio (Panel A), momentum (Panel B), and reversal (Panel C) of individual mutual funds. The funds in the left, middle, and right columns are the mutual funds that represent the 75th and the 90th percentiles, and the highest MSE of the MSE distribution of funds for a given characteristic. The legends include the wficn of the plotted fund and the mean square error in parentheses. The sample period is 2010Q3 to 2018Q4.

**Figure E.9: Fit of Tucker(3,10,10) Model of Individual Funds**

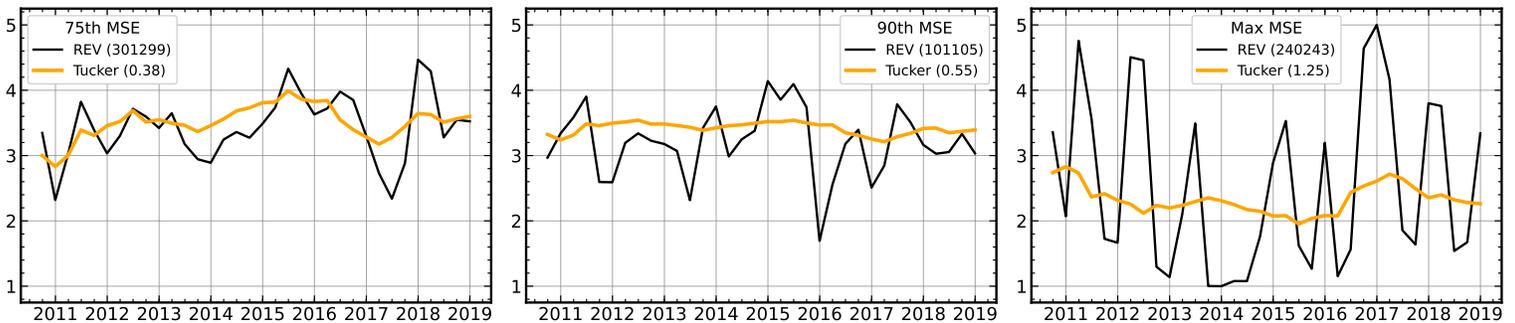
**A: BM**



**B: MOM**

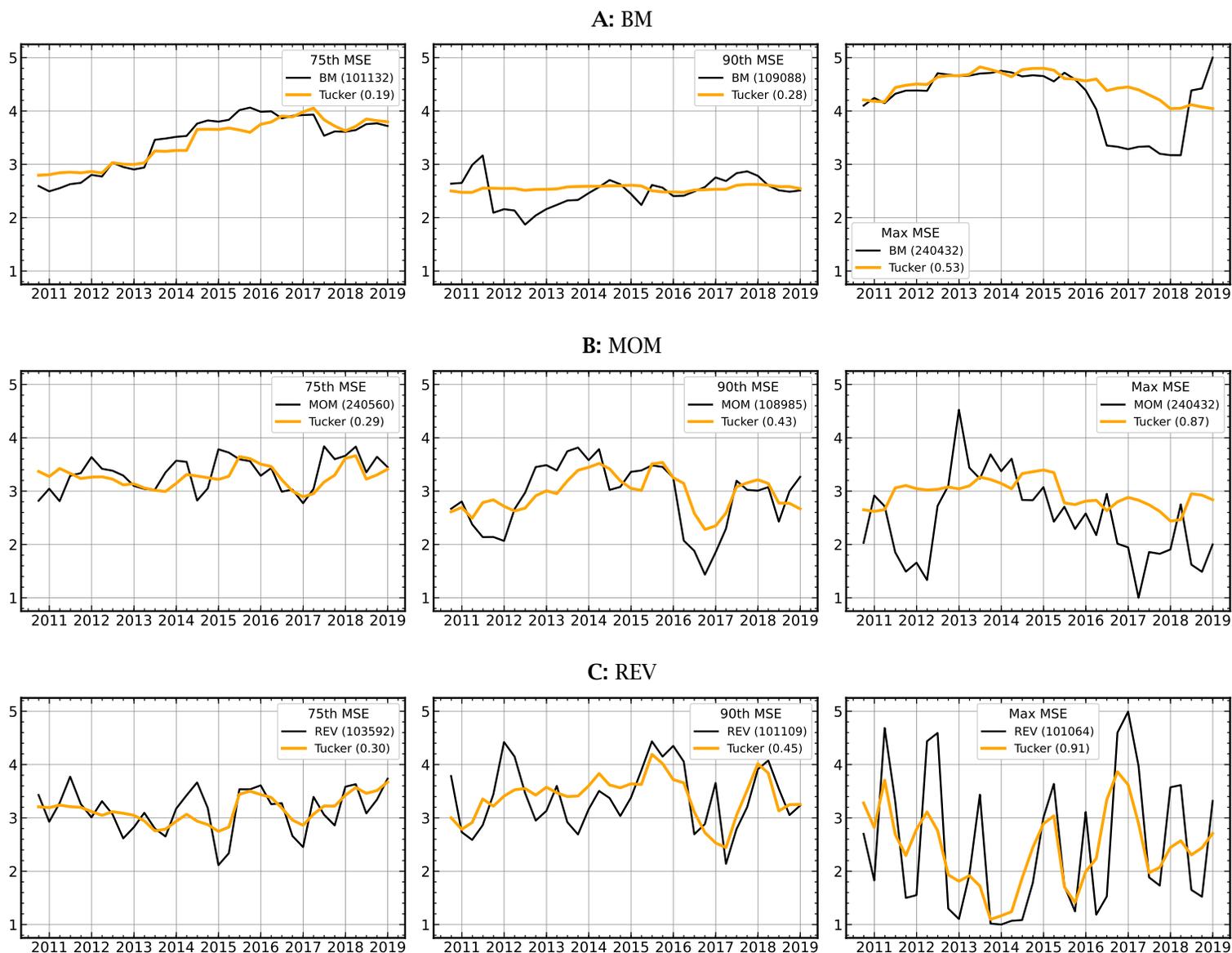


**C: REV**



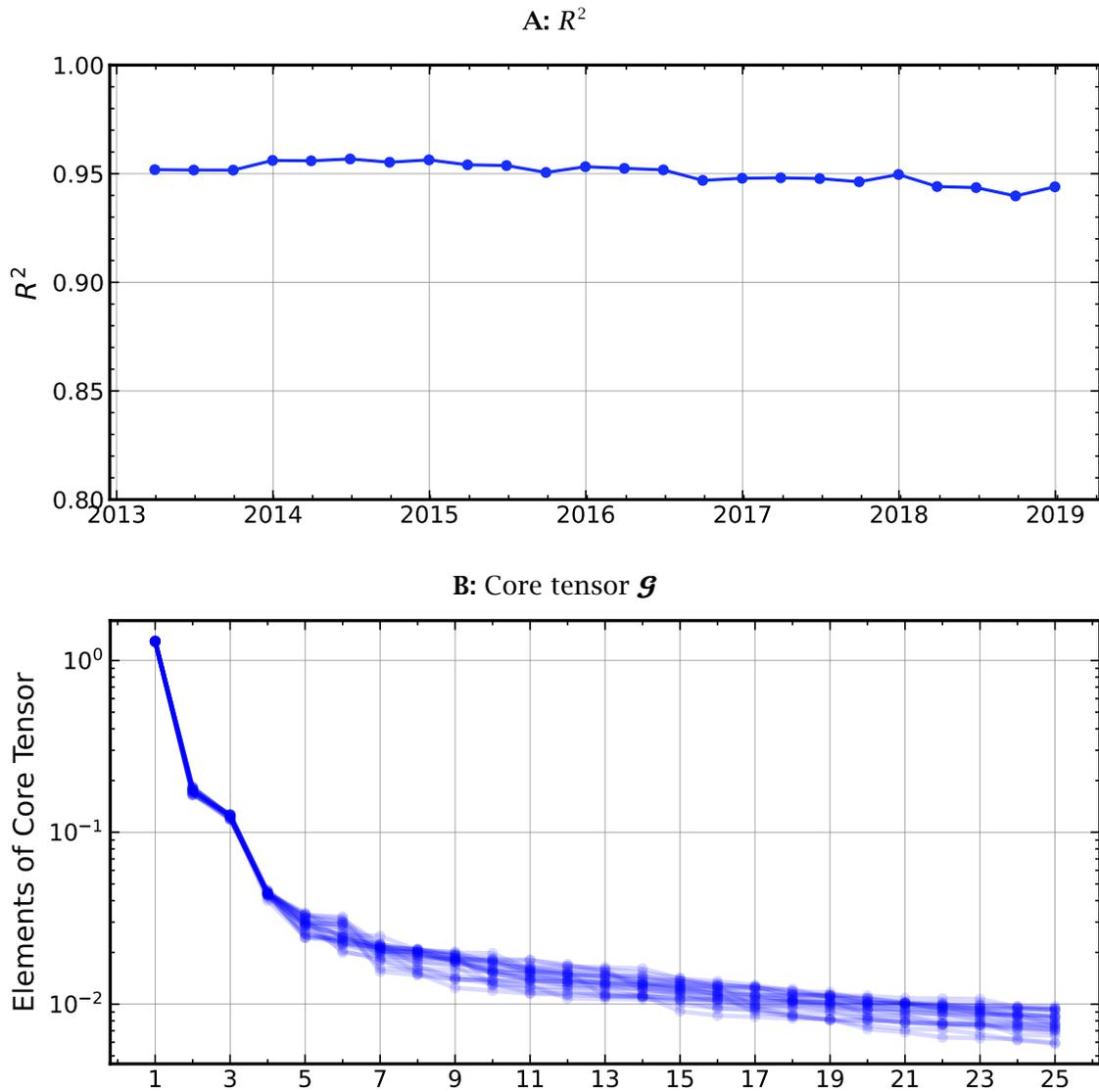
Notes: The figures show time series plots of the observed data and fitted values of the Tucker(3,10,10) model of the book-to-market ratio (Panel A), momentum (Panel B), and reversal (Panel C) of individual mutual funds. The funds in the left, middle, and right columns are the mutual funds that represent the 75th and the 90th percentiles, and the highest MSE of the MSE distribution of funds for a given characteristic. The legends include the wficn of the plotted fund and the mean square error in parentheses. The sample period is 2010Q3 to 2018Q4.

**Figure E.10: Fit of Tucker(8,12,12) Model of Individual Funds**



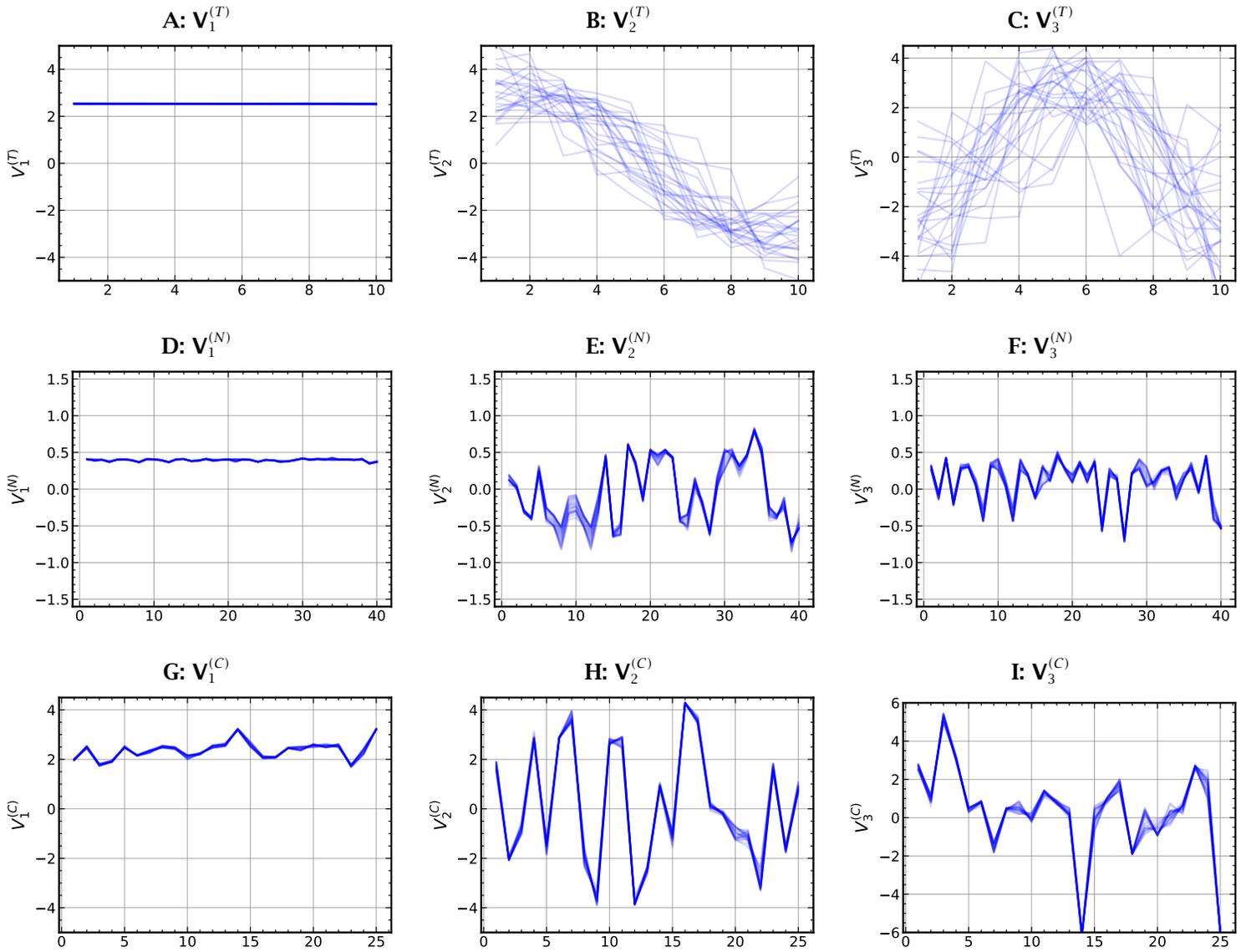
Notes: The figures show time series plots of the observed data and fitted values of the Tucker(8,12,12) model of the book-to-market ratio (Panel A), operating profitability (Panel B), investment (Panel C), and momentum (Panel D) of individual mutual funds. The funds in the left, middle, and right columns are the mutual funds that represent the 75th and the 90th percentiles, and the highest MSE of the MSE distribution of funds for a given characteristic. The legends include the wficm of the plotted fund and the mean square error in parentheses. The sample period is 2010Q3 to 2018Q4.

**Figure E.11: Recursive 10-Year Subsamples**



Notes: This figure plots the first three loading vectors of mode  $i, \mathbf{V}_j^{(i)}, i = T, N, C, j = 1, 2, 3$  of Tucker models that are estimated over sliding windows. Each window consists of 10 quarters so there are 23 subsamples. The Tucker models have  $(K_T, K_N, K_C) = (3, 12, 12)$  components. Panel B plots 23 lines corresponding to a subsample core tensor  $\mathcal{G}$ .

**Figure E.12: Recursive 10-Year Samples -  $\mathbf{V}_j^{(i)}$**



Notes: This figure plots the first three loading vectors of mode  $i$ ,  $\mathbf{V}_j^{(i)}$ ,  $i = T, N, C, j = 1, 2, 3$  of Tucker models that are estimated over sliding windows. Each window consists of 10 quarters so there are 23 subsamples. The Tucker models have  $(K_T, K_N, K_C) = (3, 12, 12)$  components. Each panel plots 23 lines corresponding to a subsample loading vector  $\mathbf{V}_j^{(i)}$ . The sample period is 2010Q3 to 2018Q4.