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, Revised May 2024

I thank Ian Dew-Becker, Christian Julliard, Ali Kahkbod, Eben Lazarus, Dmitry Livdan, Peter Maxted, Stefan Nagel, Alberto Rossi, Paulo Zaffaroni, Guofu Zhou, and seminar participants at Berkeley-Haas, CDAR, Frankfurt, the London School of Economics, the San Francisco Federal Reserve, and the 2022 BI-SHoF and 2023 AFA conferences for helpful comments. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

High-Dimensional Factor Models and the Factor Zoo
Martin Lettau
NBER Working Paper No. 31719
September 2023, Revised May 2024
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 portfolio returns 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. (2024)).

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.[1] 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.[2] 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. Correspondingly, the high-dimensional

[1] 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.

[2] 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.

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.[3] 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.[4] 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 $t = 1,...,T$. 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 $n = 1,...,N$. 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 across assets, she could estimate separate PCA models for each characteristic $c = 1,...,C$. 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

[3]This paper focuses on 3-dimensional applications but all results extend to higher dimensions.

[4]The tensor factor model used in this paper is due to Tucker (1963, 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)).

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 with 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 $(K_T, K_N, K_C) = (5, 15, 6)$ factors captures over 92% of the variation in the data with a compression ratio of 98%. 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% 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 $(K_T, K_N, K_C) = (2, 6, 5)$ is slightly lower, 87%, but the compression ratio increases to 99%. These models yield good fits for all data points except for some outliers.

Next, I construct K_C characteristic factors based on 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. To assess whether the 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 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. In contrast, the highest Sharpe ratio of in-sample TFM factors is almost twice as high, 0.84, and 0.69 for out-of-sample factors. Finally, I use TFM pricing factors in cross-sectional asset pricing tests and compare the results to those for traditional PCA and Fama-French factors. The TFM factors yield smaller pricing errors than Fama-French and PCA factors, both for in-sample and out-of-sample estimations. 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.

A distinct advantage of tensor factor models is that the number of parameters is an order of magnitude smaller than in comparable SVD/PCA models. For example, suppose there is one factor per 10 data points. In that case, the 3-dimensional TFM model has approximately 1,000 data points per estimated parameter, while there are only approximately six data points per parameter in SVD/PCA. This parsimony implies that the Tucker model can be computed in small samples. I obtain robust estimates for the data set used in this paper when the time series dimension is as low as four. It is, therefore, feasible to estimate the Tucker model in short rolling samples and assess whether the factor structure changes over time.

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. Estimating 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 (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 referenced some applications of PCA and machine learning methods 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. Chinco et al. (2019), Freyberger et al. (2020), and Martin and Nagel (2022) use regulation 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) use 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[5], 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.[6] 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). Twelve 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 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,

[5]The quality index combines the return-to-equity, debt-to-equity, and earnings variability.

[6]Choosing an earlier starting date drastically reduces the number of funds.

but there is a large dispersion of mean returns across funds.

It is well-known that time series properties differ across characteristics. Figure 1 plots the time series means, standard deviations, and first-order autocorrelations averaged across mutual funds by characteristic. ME and VOL have the highest mean scores, indicating that most mutual funds hold large and high-volume stocks. Mean scores of price-ratios tend to be below three since fund portfolios are dominated by growth stocks rather than value stocks. In addition, funds tend to hold stocks with low bid-ask spreads. MOM and REV are by far the most volatile characteristics, while ME and VOL are the least volatile. Most characteristics are persistent, with autocorrelations between 0.7 and 0.8. The exceptions are REV, BIDASK, and PSLIQ.

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 2 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. The patterns of time series and cross-sectional correlations are generally similar, but the degree of this correlation varies by characteristic. For DP, MS, SP, and BM, the patterns of time series and cross-sectional correlations are nearly identical. In contrast, the correlation of the two measures is less than 50% for QUAL, MOM, TURN, and REV.

Recall that Figure 2 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 varying over time. Figure 3 shows time series and cross-sectional correlations of four characteristic pairs, (BM,ME), (ME,EP), (BM,MOM), (BM,REV), across funds and time, respectively. Panel plots the correlations of characteristics across mutual funds over time. The (BM,ME) and (ME,EP) correlations are stable across time; however, the (BM,MOM) and (BM,REV) correlations vary substantially. Both correlation pairs are negative on average but are positive in 2017 and

2018. Panel B shows time series correlations of characteristics across mutual funds. The figure shows box plots of the distribution of pairwise correlation of BM with the other 24 characteristics.[7] For example, the median time series correlation of BM and EP across all funds is 0.46, but the box plot shows that the correlation varies substantially. The interquartile range is (0.14, 0.66), and 152 have a negative (BM,EP) correlation. The correlations of BM with the other characteristics exhibit similar variations across funds. 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.

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. I use this example to illustrate the theoretical results in this section.

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 higher-dimensional models.

The tensor models used in this paper can be interpreted as extensions of the SVD of a matrix. Like SVD/PCA, tensor decompositions summarize the variation in the data by expressing the data tensor in terms of lower-dimensional tensors and matrices. I show that the decomposition of a 3-dimensional tensor implies a collection of 2-dimensional factor models that are connected across all three dimensions and internally consistent. As with any latent factor method, it is essential 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.

[7] Figure D.1 shows the results for ME and MOM.

Throughout the paper, I use the following notation:

$$\begin{aligned} \text{scalar: } & \mathsf{x} \in \mathbb{R} \\ \text{vector: } & \mathbf{x} \in \mathbb{R}^{I} \\ \text{2-dim. matrix: } & \mathbf{X} \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \\ j\text{-th order tensor: } & \boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \ldots \times \mathbb{R}^{I_j}. \end{aligned}$$

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*.

The data set used in the empirical sections of the paper has three dimensions: the characteristic c of asset n at date t, x_{tnc}. To simplify the notation, I will therefore focus on tensors of order $j=3$ but all results can be easily generalized to higher dimensions. Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{T} \times \mathbb{R}^{N} \times \mathbb{R}^{C}$ be a 3-dimensional $(T \times N \times C)$ tensor $\boldsymbol{\mathcal{X}}$ with *elements* x_{tnc}.[8] 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 $\boldsymbol{\mathcal{X}}$ has fibers along each mode, denoted by $\mathbf{x}t(nc), \mathbf{x}_{(tc)n}$, and $\mathbf{x}_{(tn)c}$, respectively.[9] Slices are created by fixing all but two indices and are written as $\mathbf{X}_{(t)nc}, \mathbf{X}_{(n)tc}, \mathbf{X}_{(c)tn}$.[10] Tensors can be written as matrices by *unfolding*, or *stacking*, 1-dimensional fibers of a mode j as columns of a matrix. The resulting matrix $\mathbf{X}_{(j)}$ is defined so that the number of rows equals the mode-j order of $\boldsymbol{\mathcal{X}}$. The number of columns of $\mathbf{X}_{(j)}$ is equal to the product of the dimensions along all other modes.[11]

3.2. Tensor factor models

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

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

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 $\widehat{\mathbf{X}}_K = \mathbf{F}_K \mathbf{B}_K^{\top}$ be the matrix of "fitted" values. PCA factors and loadings can be

[8]Figures D.2 to D.4 show a third-order tensor with dimensions $T=5, N=4, C=3$ and illustrate the tensor operations described in this section.

[9]See Panels B, C, and D of Figure D.2.

[10]See Panels E, F, and G of Figure D.2.

[11]Figure D.3 shows the unfolding of a $(5 \times 4 \times 3)$ tensor $\boldsymbol{\mathcal{X}}$ along each mode. The resulting matrix of unfolding $\boldsymbol{\mathcal{X}}$ 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 (4×15) and (3×20), respectively.

obtained from the "truncated" SVD of $\mathbf{X}$:

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

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

$$= \sum_{k=1}^{K} h_{kk}\, \mathbf{u}_k^1 \circ \mathbf{u}_k^2 \equiv \mathsf{G}^{TN}(K), \tag{4}$$

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.[12] $\mathbf{U}_K^1$ 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. $\mathbf{u}_k^1 \circ \mathbf{u}_k^2 = \mathbf{u}_k^1 \mathbf{u}_k^{2^\top}$ denotes the outer product. Factors and loadings are given by $\mathbf{F}_K = \mathbf{U}_K^1 \mathbf{H} = \mathbf{X}\mathbf{U}_K^2$ and $\mathbf{B}_K = \mathbf{U}_K^2$. I will refer to the K-factor model for a $(T\times N)$ matrix $\mathbf{X}$ as $\mathsf{G}^{TN}(K)$.[13] The "full" SVD sets $K = \min(T,N)$ so that $\hat{\mathbf{X}}_K = \mathbf{X}$.

The representations (3) and (4) show that $\hat{\mathbf{X}}_K$ is the weighted sum of K matrices, 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 $\mathbf{u}_k^1 \circ \mathbf{u}_k^{(2)}$ 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 $\boldsymbol{\mathcal{X}}$ with dimensions $(T\times N\times C)$. One possible way to investigate the factor structure of $\boldsymbol{\mathcal{X}}$ 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 $\mathsf{G}^{NC}(K_T)$ for the t-th slice $\mathbf{X}_{(t)nc}$ of $\boldsymbol{\mathcal{X}}$. $\mathbf{X}_{(t)nc}$ 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 $\mathsf{G}^{TC}(K_N)$ 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 $\mathsf{G}^{TN}(K_C)$ for the c-th slice of $\boldsymbol{\mathcal{X}}$. This factor model captures time series correlations across N assets for characteristic c. In principle, one could estimate 2-dimensional factor models for each $t = 1,...,T, n = 1,...,N$, and $c = 1,...,C$ 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. The Tucker factor model (Tucker (1963, 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

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

[13]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.

factor model for each $t = 1,...,T$, each $c = 1,...C$, and each $n = 1,...,N$. Instead of specifying $T+N+C$ separate 2-dimensional factor models, the *single* 3-dimensional factor model can be rewritten as $T+N+C$ 2-dimensional factor models. Since all 2-dimensional factor models stem from the 3-dimensional representation, they are mutually consistent. In addition, the 3-dimensional Tucker model exploits dependence in all dimensions without collapsing 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 generalizes 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 $\boldsymbol{\mathcal{X}}$ and a matrix $\mathbf{A}_n$ is the multiplication of each n-mode fiber of $\boldsymbol{\mathcal{X}}$ by the row vectors of $\mathbf{A}_n$. For example, the mode-1 product of a $(S \times N \times C)$ tensor $\boldsymbol{\mathcal{X}}$ and a $(T \times S)$ matrix $\mathbf{A}_1$ is equal to a $(T \times N \times C)$ tensor $\boldsymbol{\mathcal{Y}}$ given by $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \times_1 \mathbf{A}_1$ (see Figure D.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 $\boldsymbol{\mathcal{X}}$ be a data tensor with dimensions $(T \times N \times C)$. The Tucker approximation $\widehat{\boldsymbol{\mathcal{X}}}$ of order (K_T, K_N, K_C) is given by

$$\widehat{\boldsymbol{\mathcal{X}}}(K_T, K_N, K_C) = \boldsymbol{\mathcal{G}} \times_1 \mathbf{V}^T_{K_T} \times_2 \mathbf{V}^N_{K_N} \times_3 \mathbf{V}^C_{K_C}, \tag{5}$$

where $\boldsymbol{\mathcal{G}}$ is a $(K_T \times K_N \times K_C)$ tensor with elements $g_{k_T k_N k_C}$ and $\mathbf{V}^T_{K_T}, \mathbf{V}^N_{K_N}, \mathbf{V}^C_{K_C}$ are $(T \times K_T), (N \times K_N)$, $(C \times K_C)$ matrices, respectively. Given the definition of the n-mode tensor product, $\widehat{\boldsymbol{\mathcal{X}}}$ is a $(T \times N \times C)$-dimensional tensor and thus has the same dimensionality as the data tensor $\boldsymbol{\mathcal{X}}$. $\boldsymbol{\mathcal{G}}$ is called the *core* tensor and can be thought of as a "compressed" version of $\boldsymbol{\mathcal{X}}$. As we will see below, the matrices $\mathbf{V}^i_{K_i}$ 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 $\boldsymbol{\mathcal{E}} = \boldsymbol{\mathcal{X}} - \widehat{\boldsymbol{\mathcal{X}}}$. The optimal Tucker model minimizes the mean-squared error (MSE) of $\boldsymbol{\mathcal{E}}$ among all $\widehat{\boldsymbol{\mathcal{X}}}(K_T, K_N, K_C)$ of the form (5).

The mechanism of the Tucker decomposition (5) is illustrated in Figure 4, which shows the decomposition of a $(6 \times 5 \times 4)$ tensor $\boldsymbol{\mathcal{X}}$ by a Tucker model of order $(K_T, K_N, K_C) = (3, 2, 2)$. The core tensor $\boldsymbol{\mathcal{G}}$ compresses $\boldsymbol{\mathcal{X}}$ to a lower-dimension of $(3 \times 2 \times 2)$. The loading matrices $\mathbf{V}^T, \mathbf{V}^N$, and $\mathbf{V}^C$ expand the core tensor to the full dimension of $\boldsymbol{\mathcal{X}}$ and have conforming dimensions of $(6 \times 3), (5 \times 2)$, and (4×2). 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 $\widehat{\boldsymbol{\mathcal{X}}}$ is a matrix and $K_i = K$. This can be seen by rewriting (2) in tensor notation:

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

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

$$\widehat{\boldsymbol{\mathcal{X}}}(K_T, K_N, K_C) = \sum_{k_T=1}^{K_T} \sum_{k_N=1}^{K_N} \sum_{k_C=1}^{K_C} g_{k_T k_N k_C} \mathbf{v}_{k_T}^T \circ \mathbf{v}_{k_N}^N \circ \mathbf{v}_{k_C}^C, \tag{7}$$

where $g_{k_T k_N k_C}$ are the elements of the core tensor $\boldsymbol{\mathcal{G}}$ and $\mathbf{v}_{k_i}^i$ 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, $\widehat{\boldsymbol{\mathcal{X}}}(K_T, K_N, K_C)$ 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 $\boldsymbol{\mathcal{G}}$ and loading matrices $\mathbf{V}_{(i)}$ are not tied to eigenvalues and eigenvectors. In general, $\boldsymbol{\mathcal{G}}$ is not diagonal, so the number of factors may differ across dimensions.[14] 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 3.7.

Neither the SVD nor the Tucker model are unique and can be rotated.[15] Let $\mathbf{S}_i, i = T, N, C$, be nonsingular $(K_i \times K_i)$ matrices. Then (5) can be written as

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

(5) is normalized so that the $\mathbf{V}^i$ matrices are orthonormal, similar to the eigenvector matrices in the SVD/PCA. One important property of the Tucker decomposition is that it 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.[16] 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.

The Tucker representation (5) can be written in matrix notation. The loading matrices $\mathbf{V}^T, \mathbf{V}^N$, and $\mathbf{V}^C$ can

[14]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 contrast to the Tucker decomposition, the CP decomposition might not exist. However, if it exists, the CP decomposition is unique under mild conditions. See Kolda and Bader (2009) for more details and Babii et al. (2022) for a recent application.

[15]The SVD can be rotated as follows: $\hat{\mathbf{X}}_K = \mathbf{U}_K^1 \mathbf{H}_K \mathbf{U}^{(2)\top} = (\mathbf{U}_K^1 \mathbf{S}_1^{-1})(\mathbf{S}_1 \mathbf{H}_K \mathbf{S}_2)(\mathbf{S}_2^{-1} \mathbf{U}^{(2)\top}) = \tilde{\mathbf{U}}_K^1 \tilde{\mathbf{H}}_K \tilde{\mathbf{U}}^{(2)\top}$, where $\mathbf{S}_1$ and $\mathbf{S}_2$ are conforming matrices. $\tilde{\mathbf{H}}_K$ is no longer diagonal and $\tilde{\mathbf{U}}_K^1, \tilde{\mathbf{U}}_K^2$ are not orthonormal.

[16]However, in most applications, the differences are small.

be combined to a single loadings matrix using the Kronecker product:

$$\mathbf{W} = \mathbf{V}^T \otimes \mathbf{V}^N \otimes \mathbf{V}^C, \tag{9}$$

where $\mathbf{W}$ is a $(TNC \times K_T K_N K_C)$-dimensional matrix. Each column of $\mathbf{W}$ corresponds to a factor, while each row is a loading for a single element of $\boldsymbol{\mathcal{X}}$. Let $\mathrm{vec}(\boldsymbol{\mathcal{X}})$ be the vectorized form of $\boldsymbol{\mathcal{X}}$. The Tucker model (5) can be written as

$$\mathrm{vec}(\widehat{\boldsymbol{\mathcal{X}}}) = \mathbf{W}\,\mathrm{vec}(\boldsymbol{\mathcal{G}}). \tag{10}$$

3.3. Factor representation

The Tucker model has a factor representation similar to (1) along each dimension.[17] Consider first the representation for the characteristic dimension implied by (5):

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{F}}^C \times_3 \mathbf{V}^C + \boldsymbol{\mathcal{E}} \tag{11}$$

$$\text{where } \boldsymbol{\mathcal{F}}^C = \boldsymbol{\mathcal{G}} \times_1 \mathbf{V}^T \times_2 \mathbf{V}^N. \tag{12}$$

$\boldsymbol{\mathcal{F}}^C$ 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 $\boldsymbol{\mathcal{F}}^C$ is similar to that of factor matrices in 2-dimensional SVD/PCA models. For a given asset n, the nth slice of $\boldsymbol{\mathcal{F}}^C, \mathbf{F}_n^C$, is a $(T \times K_C)$ matrix whose columns are the time series of K_C characteristic factors of asset n. Recall that the nth slice of $\boldsymbol{\mathcal{X}}, \mathbf{X}_{(n)tc}$, 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 $\mathbf{F}_n^C$ summarize the information in all C characteristics of asset n. The implied factor model for asset n is given by

$$\mathbf{X}_{(n)tc} = \mathbf{F}_n^C \mathbf{V}^{C^\top} + \mathbf{E}_{(n)tc}. \tag{13}$$

Thus, (11) can be understood as a collection of N 2-dimensional K_C-factor models for characteristics of individual assets $n = 1, \ldots, N$. 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 $\boldsymbol{\mathcal{F}}^C$ 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:

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{F}}^N \times_1 \mathbf{V}^N + \boldsymbol{\mathcal{E}} \tag{14}$$

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{F}}^T \times_2 \mathbf{V}^T + \boldsymbol{\mathcal{E}}. \tag{15}$$

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

[17] 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 $\mathbf{U}_K^2$ forms loadings in the $\mathbf{X}^\top$ representation.

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 $t = 1,...,T$, each $n = 1,...,N$, and each $c = 1,...,C$. 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 $\boldsymbol{\mathcal{G}}$. Recall that a 2-dimensional K-factor model $\mathsf{G}(K)$ 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:

$$\begin{aligned}\widehat{\boldsymbol{\mathcal{X}}}(K_T,K_N,K_C) &= \sum_{k_T=1}^{K_T} \mathbf{v}_{k_T}^T \circ \left[\sum_{k_N=1}^{K_N} \mathbf{v}_{k_N}^N \circ \left(\sum_{k_C=1}^{K_C} g_{k_T k_N k_C} \mathbf{v}_{k_C}^C\right)\right] && (16)\\
&= \sum_{k_T=1}^{K_T} \mathbf{v}_{k_T}^T \circ \left[\sum_{k_N=1}^{K_N} \mathbf{v}_{k_N}^N \circ \tilde{\mathbf{v}}_{k_T,k_N}^C\right] && (17)\\
&= \sum_{k_T=1}^{K_T} \mathbf{v}_{k_T}^T \circ \mathsf{G}_{k_T}^{NC}(K_N) && (18)\end{aligned}$$

where $\tilde{\mathbf{v}}_{k_T,k_N}^C = \sum_{k_C=1}^{K_C} g_{k_T k_N k_C} \mathbf{v}_{k_C}^C$. $\mathbf{v}_{k_N}^N$ 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 (16) and (17) represent a 2-dimensional K_N-factor model $\mathsf{G}_{k_T}^{NC}(K_N)$ for the N and C dimensions of $\boldsymbol{\mathcal{X}}$. 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). 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 $\mathsf{G}_{k_T}^{NC}(K_N)$ 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 $\widehat{\boldsymbol{\mathcal{X}}}$.

The order of the summation in the term in square brackets in (16) can be reversed so that the sum in square brackets in (17) is over k_C instead of k_N. The resulting factor model is equivalent to (18) but has K_C factors. I will choose the representation that has the fewest factors and write the resulting factor model as $\mathsf{G}_{k_T}^{NC}(\tilde{K}_{NC})$, 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 $\mathsf{G}_1^{NC}(K_N),...,\mathsf{G}_{K_T}^{NC}(K_N)$ by the t-th elements of the vectors $\mathbf{v}_{k_T}^T$ in (18):

$$\hat{\mathbf{X}}_{(t)nc} = \sum_{k_T=1}^{K_T} \mathsf{v}_{k_T,t}^T \; \mathsf{G}_{k_T}^{NC}(K_N) \equiv \mathsf{G}_{(t)}^{NC}(K_N), \tag{19}$$

where $\mathsf{v}_{k_T,t}^T$ is the t-th element of the k_T-th column of the loading matrix $\mathbf{V}^T$. (19) shows that the implied 2-dimensional factor model for period t is given by a weighted sum of K_T factor models $\mathsf{G}_{k_T}^{NC}(\tilde{K}_{NC}), k_T = 1,...,K_T$ and is therefore also a $\tilde{K}_{NC}$-factor model, defined as $\mathsf{G}_{(t)}^{NC}(\tilde{K}_{NC})$. The weights are given by the rows of $\mathbf{V}^T$. Since the weights vary across periods t, the factor models $\mathsf{G}_t^{NC}(\tilde{K}_{NC})$ change over time. However, the time-t factors are based on the same $\mathsf{G}_{k_T}^{NC}(\tilde{K}_{NC})$ 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,

$$\mathsf{G}^{NC}_{(t)}(\tilde{K}_{NC}) = \mathsf{v}^T_{1,t}\, \mathsf{G}^{NC}_{(t)}(\tilde{K}_{NC}), \tag{20}$$

which implies that all t-slices are proportional to the same factor model $\mathsf{G}^{NC}_{(1)}(\tilde{K}_{NC})$, or, equivalently, proportional to the same $(N\times C)$ matrix given by the term in square brackets in (17) that forms $\mathsf{G}^{NC}_{(1)}(\tilde{K}_{NC})$. 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

$$\mathsf{G}^{NC}_{(t)}(K_N) = \mathsf{v}^T_{1,t}\, \mathsf{G}^{NC}_1(\tilde{K}_{NC}) + \mathsf{v}^T_{2,t}\, \mathsf{G}^{NC}_2(K_{NC}). \tag{21}$$

Time slices are given by weighted sums of two $(N\times C)$ factor models and are thus not proportional. However, unless the weights $\mathsf{v}^T_{1,t}$ and $\mathsf{v}^T_{2,t}$ 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 $\boldsymbol{\mathcal{X}}$, the results apply to the other dimensions as well and can be summarized as follows:

$$\begin{align}
\widehat{\boldsymbol{\mathcal{X}}}(K_T,K_N,K_C) &= \sum_{k_T=1}^{K_T} \mathbf{v}^T_{k_T} \circ \mathsf{G}^{NC}_{k_T}(\tilde{K}_{NC}) \tag{22}\\
&= \sum_{k_N=1}^{K_N} \mathbf{v}^N_{k_N} \circ \mathsf{G}^{TC}_{k_N}(\tilde{K}_{TC}) \tag{23}\\
&= \sum_{k_C=1}^{K_C} \mathbf{v}^C_{k_C} \circ \mathsf{G}^{TN}_{k_C}(\tilde{K}_{TN}), \tag{24}
\end{align}$$

and

$$\begin{align}
\widehat{\mathbf{X}}_{(t)nc} &= \mathsf{G}^{NC}_{(t)}(\tilde{K}_{NC}) \tag{25}\\
\widehat{\mathbf{X}}_{(n)tc} &= \mathsf{G}^{TC}_{(n)}(\tilde{K}_{TC}) \tag{26}\\
\widehat{\mathbf{X}}_{(c)tn} &= \mathsf{G}^{TN}_{(c)}(\tilde{K}_{TN}). \tag{27}
\end{align}$$

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 G are identical in one of the two dimensions. The last three equations show that slices of $\widehat{\boldsymbol{\mathcal{X}}}$ 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, 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 $\hat{\mathbf{X}}_{(n)tc}$, which are matrices of dimension $(T\times C)$ are proportional and all n-fibers $\hat{\mathbf{x}}_{(tc)n}$ ($(N\times 1)$ vectors) are perfectly correlated. In other words, all characteristics are proportional across assets. If $K_C = 1$, c-slices $\hat{\mathbf{X}}_{(c)tn}$ ($(T\times N)$ matrices) are proportional and c-fibers $\hat{\mathbf{x}}_{(tn)c}$ ($(C\times 1)$ 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. Number of parameters

The core tensor $\boldsymbol{\mathcal{G}}$ 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 Tucker model has $K_T K_N K_C + TK_T + NK_N + CK_C - K_T^2 - K_N^2 - K_C^2$ free parameters where the last three terms are due to the rotational indeterminacy, see (8). Define κ_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, $\kappa_3 = (K_T K_N K_C + TK_T + NK_N + CK_C - K_T^2 - K_N^2 - K_C^2)/(TNC)$. For comparison, the 2-dimensional K-factor SVD of a $(T\times N)$ matrix has $K(T+N+1) - K(K+1) = K(T+N) - K^2$ free parameters so that $\kappa_2 = (K(T+N)-K^2)/(TN)$.

Further insights about how data compression is related to the dimensionality of the data tensor can 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.[18] I assume that $K, M \to \infty$ at the same rate so that $Q = K/M$ is a constant. Then $\kappa_3 = (K/M)^3 + 3K/M^2 - 3K^2/M^3 = Q^3 + 3Q/M - 3Q/M^2 \to Q^3$ as $K, M \to \infty$. 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%, i.e., there is one Tucker component for every ten data dimensions, and M is large, $\kappa_3 = 0.001$ so that the Tucker model compresses the data by approximately 99.9%. A similar calculation for the 2-dimensional SVD of an $(M\times M)$

[18] It is easy to show that the results extend to the general case.

matrix shows that $\kappa_2 \to 2Q(1-Q)$, which is an order of magnitude higher than κ_3 of the 3-dimensional Tucker model. For $Q = 0.1$, $\kappa_2 = 0.18$ and the compression ratio is only 82%.

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 $K^i + iMK - iK^2$ parameters, so that $\kappa_i = Q^i + iQ/M^{i-2} - iQ^2/M^{i-2} \to Q^i$. Hence, the order of the 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. *Determining the number of factors*

While there are formal selection criteria to determine the number of factors for 2-dimensional factor models, e.g., Bai and Ng (2002), to my knowledge, no formal methods exist for Tucker tensor models. In 2-dimensional SVD/PCA, the contribution of a factor to the overall variation in the data is given by the associated eigenvalue of the eigenvector that is used to construct the factor. The appropriate number of factors is based on the spectrum of eigenvalues. Since the Tucker decomposition allows for a different number of factors in each dimension, we require measures of the contributions of factors in each of the dimensions. Consider the full Tucker decomposition with $K_T = T, K_N = N, K_C = C$, so that $\widehat{\boldsymbol{\mathcal{X}}}(K_T, K_N, K_C) = \boldsymbol{\mathcal{X}}$. Since the $\mathbf{V}^i$ matrices are orthonormal, the overall variation of the data tensor $\boldsymbol{\mathcal{X}}$ is equal to the squared norm of the core tensor $\boldsymbol{\mathcal{G}}$:

$$\|\boldsymbol{\mathcal{X}}\|^2 = \sum_{t,n,c} x_{tnc}^2 = \sum_{t,n,c} g_{tnc}^2 = \|\boldsymbol{\mathcal{G}}\|^2. \tag{28}$$

This result is reminiscent of the fact that in matrix SVD/PCA, the sum of eigenvalues is equal to the overall data variation. (28) implies that a squared element of the core tensor $\boldsymbol{\mathcal{G}}$ can be interpreted as the contribution of the corresponding factor to the overall variation of the data. Next, consider a model with a subset of factors. Let $\mathcal{K}$ be the set of included factor combinations (k_T, k_N, k_C), and $\mathcal{K}'$ be the corresponding set of excluded factor combinations. It is straightforward to show that the sum of squared fitted values, the sum of squared approximation errors, and the implied R-squared are given by

$$\|\widehat{\boldsymbol{\mathcal{X}}}\|^2 = \sum_{(k_T,k_N,k_C)\in\mathcal{K}} g_{k_T k_N k_C}^2, \tag{29}$$

$$\|\boldsymbol{\mathcal{E}}\|^2 = \sum_{(k_T,k_N,k_C)\in\mathcal{K}'} g_{k_T k_N k_C}^2, \tag{30}$$

$$R_{\mathcal{K}}^2 = \frac{\|\widehat{\boldsymbol{\mathcal{X}}}\|^2}{\|\boldsymbol{\mathcal{X}}\|^2} \tag{31}$$

respectively. For example, the maximal R^2 of any model with K factor combinations is obtained by setting $\mathcal{K}$ to the K largest elements of $\boldsymbol{\mathcal{G}}$. The contribution of a single factor is given by the sum of the squared associated elements of $\mathcal{G}$:

$$\Lambda_{k_T} = \sum_{n,c} g^2_{k_T nc} \tag{32}$$

$$\Lambda_{k_N} = \sum_{t,c} g^2_{t k_N c} \tag{33}$$

$$\Lambda_{k_C} = \sum_{t,n} g^2_{t n k_C}, \tag{34}$$

where k_T,k_N, and k_C are single T, N, and C-factors, respectively. Plotting Λ_{k_T},Λ_{k_N}, and Λ_{k_C} corresponds to a scree plot of eigenvalues in SVD/PCA. The difference is that there is a separate scree plot for each of the three dimensions.

Another metric for determining the number of factors is the R^2. In a 2-dimensional SVD/PCA model with K factors, the R^2 equals the sum of the K largest eigenvalues divided by the sum of all eigenvalues. In the Tucker model, the R-squared of a model with up to K_i factors of dimension i is given by $R^2_{K_i} = \sum_{k_i=1}^{K_i} \Lambda_{k_i} / \|\mathcal{X}\|^2$ for $i = T,N,C$. The spectra of Λ_{k_i} and $R^2_{K_i}$ can be used to determine appropriate factor orders for each dimension i.

3.7. Estimation

In contrast to the SVD/PCA matrix representation, no closed-form solution exists 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 follow the literature and use the most popular algorithm, *Higher-Order Orthogonal Iteration* (HOOI) throughout the paper.[19] 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 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, $\mathcal{X}_i$ and estimate the Tucker model $\widehat{\mathcal{X}}_i$ for the true model plus noise, $\mathcal{X}^e_i = \mathcal{X}_i + \sigma_e \mathcal{E}_i$. The elements of the noise tensor are drawn from standard normal distributions. Table 3 reports the mean RMSE of $\mathcal{X}_i - \widehat{\mathcal{X}}_i$ 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 $\mathcal{X}_i$ have 1,000,000 data points, 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 $(T,N,C) =$

[19] HOOI is also known as Alternating Least Squares (ALS).

(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 $\mathcal{X}_i$ 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 $\mathcal{X}_i^e$ is equal to one.[20] 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

This section presents the estimation results of Tucker models for mutual fund characteristics. I start with a discussion about the appropriate number of factors of each of the three dimensions of the data set and assess the fits of models with different numbers of factors. Second, I analyze the properties of the estimated models. Third, I compare 3-dimensional models to 2-dimensional SVD/PCA models.

4.1. Fit of Tucker models

In 2-dimensional SVD/PCA models, the choice of the number of factors is based on the spectrum of eigenvalues. As discussed in section 3.6, in the Tucker model, the squared elements of the core tensor can be used to determine the appropriate number of factors. The contribution of a factor in dimension i to the overall fit, Λ_{k_i} is equal to the sum of squared core elements that correspond to the factor, see (32)-(34). To obtain the Λ_{k_i}, I first estimate the full Tucker model with $(K_T,K_N,K_C) = (T,N,C)$).[21] The left column of Figure 5 shows scree plots of $\Lambda_{k_T}, \Lambda_{k_N}$, and Λ_{k_T} on log-scales. The right column shows the implied R^2 of models with up to K_i factors of dimension i, which given by $R^2_{K_i} = \sum_{k_i=1}^{K_i} \Lambda_{k_i} / \|\mathcal{X}\|^2$ for $i = T,N,C$.

Figure 5 shows that the Λ_{k_i} have similar properties as scree plots of eigenvalues in SVD/PCA estimations. Λ_{k_T}, in Panel A, are monotonically declining, which implies that the lower-order factors contribute more to

[20] The scaling of $\mathcal{X}_i^e$ has an immaterial impact on the results.

[21] Since mutual fund characteristics are measured on a somewhat arbitrary scale of [1,5], I demean the data tensor $\mathcal{X}$ in all estimations. This has a negligible effect on the results.

the model's fit than higher-order factors. The first element, $k_T = 1$ is by far the largest, and the higher-order Λ_{k_T} decline approximately at an exponential rate. Panel B plots the implied R^2 and shows that including only the first Λ_{k_T} yields an R^2 of almost 90%. Adding further factors increases the R^2 albeit at a declining rate. For example, the R^2 for $K_T = 5$ is 96%.

For the mutual fund dimension, shown in Panels B and D, the first two Λ_{k_N} are dominant, while the higher-order elements decrease exponentially. Recall that there are 934 mutual funds in the data set. Including only the first factor yields an R^2 of 57%; adding the second factor raises the R^2 to 84%, and a six-factor specification captures over 90% of the data variation. Hence, most information in almost a thousand funds can be captured by a low-dimensional Tucker model. Panel E plots Λ_{k_C} for the characteristic dimension. As for Λ_{k_N}, the first two elements stand out. The R^2 of specifications with one, two, and five factors are 57%, 84%, and 92%, respectively. The results show that the information in the large data of mutual fund characteristics can be well-captured by low-dimensional Tucker models.

Next, I discuss the fit and properties of three benchmark specifications in more detail. I set (K_T, K_N, K_C) to the smallest values that yield R^2 above 85%, 90%, and 95% in each dimension. The corresponding models have $(1,3,3)$, $(2,6,5)$, and $(5,15,8)$ factors. The most parsimonious model has $K_T K_N K_C = 9$ factors, while the other two specifications have 60 and 600. While 600 factors may appear high, the "effective" number of factors is much smaller. Recall that the contribution of a factor combination (k_T, k_N, k_C) to the overall fit is the corresponding squared element of the core tensor, $g^2_{k_T k_N k_C}$. Most core entries are small, so most factors are largely irrelevant to the fit. Panel A of Figure C.1 plots the 25 largest core elements. The horizontal lines correspond to contributions to the overall variance of 0.1%, 0.3%, and 0.5%. Only four, seven, and 18 core elements contribute more than 0.5%, 0.3%, and 0.1% to the variance of the fitted values. The corresponding R^2 are shown in Panel B. The two largest core elements account for 81% of the variance, and the eight largest elements contribute 85%.

Table 4 reports the properties of these three benchmark models. Consider first the MSE in the first row. The MSE of of the $(1,3,3)$ model is 0.09, which corresponds to an R^2 of 83.63%. Increasing the number of factors to $(2,6,5)$ and$(5,15,8)$ decreases the MSE to 0.07 and 0.04. The R^2 are 87.13% and 92.12%.[22] As discussed in the previous paragraph, most factors play only a minor role in the fit. The next row shows the R-square, defined as $\tilde{R}^2$, when all combinations of factors (k_T, k_N, k_C) that contribute less than 0.1% are excluded. $\tilde{K}$ is the number of remaining factors. The $(1,3,3)$ model has nine factors, but only three contribute meaningfully to its fit. The R^2 remains essentially when the other six factors are removed. The $(2,6,5)$ model has 60 factors, of which nine are relevant. $\tilde{R}^2$ is 86.49%, which is only slightly lower than the R^2 with all 60 factors. Finally, 18 of the 600 factors of the $(5,15,8)$ model contribute more than 0.1% to the fit. The $\tilde{R}^2$ is 87.18%, which is somewhat lower than the R^2 of 92.12%, but the number of effective factor combinations is reduced substantially.

The last row of Panel A shows the compression ratio, defined as one minus the number of model param-

[22]The R^2 are slightly lower than those plotted in Figure 5 because they are based on truncated specifications instead of the full specification $(K_T, K_N, K_C) = (34, 934, 25)$.

eters divided by the number of data points. The compression ratio of the largest model with $(K_T, K_N, K_C) = (5,15,8)$ factors is 98.15%, implying that there are 54 data points for each free parameter. The compression ratios of the two smaller models are 99.27% and 99.63% and thus even higher, corresponding to 137 and 274 data points per parameter. Note that the compression ratios are similar even though the total number of factors, $K_T K_N K_C$, differ significantly: the three specifications have 9, 60, and 600 factors. However, the number of parameters increases slower than the total number of factors. For comparison, consider estimating traditional factor models for a 2-dimensional slice of the data tensor. For example, fix a characteristic and estimate SVD/PCA using the resulting matrix with 34 time series observations of 934 mutual funds. The compression ratios of models with two, five, and ten factors are 94%, 85%, and 70%, respectively, corresponding to 16, 7, and 3 data points per parameter. In other words, Tucker models with more factors have fewer parameters than corresponding higher-dimensional SVD/PCA models.

Panel B of Table 4 reports moments of the errors, $\boldsymbol{\mathcal{E}}$. The mean and median error of all three specifications is close to zero. The percentiles of the error distributions show that the (5,15,8) model implies fewer large errors (in absolute value) than the (1,3,3) model. For example, in the model with (1,3,3) factors, there are 254 errors above one in absolute value compared to only 24 in the (5,15,8) model.

Next, I study mean squared errors for each dimension, i.e., I compute the MSE for each quarter, each mutual fund, and each characteristic. The results are shown in Figure 6. Consider first the MSE by quarter in Panel A. The mean square error for the model with (1,3,3) factors ranges from 0.069 in 2014Q1 to 0.123 in 2010Q4. The mean MSE across all quarters is 0.092. Adding further components improves the fit across the sample. For (5,15,8) factors the MSE ranges from 0.036 to 0.055 with an overall mean of 0.044. Panel B shows the fit of all 934 mutual funds. Funds are sorted by MSE from smallest to largest. For all three specifications, the errors are small for the majority of funds, but there are some funds with large errors. The average MSE for the three models is 0.044, 0.072, and 0.092. As for the unconditional errors, larger models reduce the frequency of outliers in the error distribution. For example, for the (1,3,3) model, 62 funds have an MSE over 0.2. For the (2,6,5) and (5,15,8) models, this is the case for only 40 and 14 funds, respectively. The highest fund MSE are 1.150, 0.982, and 0.569.

The MSE by characteristics is plotted in Panel C. Consider the specification with (1,3,3) factors first. MOM and REV have the largest pricing errors, followed by TURN, while the fit is best for ME and VOL. The mean MSE across all characteristics is 0.09. The MSE of the (2,6,5) model are lower for all characteristics. The improvement is minor for some but more substantial for others. The mean MSE is 0.07. Increasing the number of factors to (5,15,8) further lowers the MSE, and the average MSE is 0.04.

Note that MOM, REV, and TURN errors are substantially lower for the model with (5,15,8) factors than those with fewer factors. Recall that these three characteristics are also the most volatile (see Figure 1). To investigate how adding factors affects the fit of characteristics with different volatilities, I add factors in one dimension while holding the numbers of factors in the other dimensions fixed. Figure 7 plots the corresponding MSE of ME and BM, which have low volatilities, and MOM and REV.[23] The baseline model

[23]The other low-volatility characteristics behave similarly as ME and BM.

is the specification with (5,15,8) factors. Panel A varies K_T from two to twenty while holding K_N and K_C fixed.[24] The figure shows that the MSE of ME and BM remains almost constant when the K_T is increased. On the other hand, the fit of MOM and REV improved considerably. The MSE of MOM for $K_T = 2$ is 0.13 but lowered to 0.08 for $K_T = 5$, corresponding to the benchmark specification. Increasing K_T to ten decreases the MSE further to 0.06. The pattern of REV is similar. Its MSE for $K_T = 2,5,10$ are 0.15, 0.09, and 0.07, respectively. The effect of K_N, shown in Panel B, is more uniform across characteristics. The MSE of all characteristics decreases when K_N is increased. Panel C shows the effect of K_C on the MSE. The ME error is almost unaffected by increasing K_C beyond two, while K_C moderately impacts the BM fit. The MSE of MOM and REV is above 0.17 for $K_T \leq 3$ but decreases to around 0.07-0.08 when a fourth factor is added. Increasing K_T beyond four has only a minor effect on the MSE. These results suggest that few K_T and K_C components are necessary to yield good fits for characteristics with low volatilities. However, $K_T \geq 5$ and $K_C \geq 4$ are required to fit MOM and REV well.

I conclude that parsimonious Tucker factor models yield good approximations of the mutual fund sample with (34,934,25) observations. The model with $(K_T, K_N, K_C) = (5,15,8)$ components captures 92% of the variation in the data while compressing the data dimensions by 98%. More parsimonious models with (2,6,5) and (1,3,3) components yield R^2 of 87% and 84% and have compression ratios above 99%. pond to factors and columns to characteristics. Blue (red) cells represent positive (negative) loadings. 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 2. 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 factor in the second row of $\mathbf{V}^C$ bear this pattern out. The nine largest elements 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.

4.2. Estimation in rolling windows

The results presented so far are based on full-sample estimations. In this section, I estimate Tucker models with (2,6,5) factors in rolling subsamples to assess the stability of Tucker models in shorter samples. The results for the other specifications are similar. Panel A of Figure 9 plots the time series of R^2 for three different window lengths: $h = 4,8,12$. Each dot corresponds to the last quarter of a particular window. The R^2 from rolling samples are between 89% and 94% for all window sizes, which is slightly higher than the

[24] I omit the case of $K_T = 1$ to make the figures more readable.

full sample R^2 of 87%. The fit is stable over time for all values of h, suggesting that the estimations yield reliable results across the sample.

Next, I study how the estimated components of the TFM vary with the length of the windows. Panel B to D plot the squares of the ten largest elements of the core tensor for $h = 4, 8, 12$. Each line corresponds to the core tensor estimated in one of the rolling windows. For $h = 4$, the largest two core elements are between 27 and 34, and 13 and 17, across windows. The higher-order core elements are small and almost identical. Increasing the window length to $h = 8$ and $h = 12$ yields slightly more stability across windows, but the effect is small.

Figure 10 shows the estimation of the loadings of the first factors, i.e., the first columns of the loading matrices $\mathbf{V}^i$. The results for the second factor are in Figure C.2. The results for $\mathbf{V}^C$ are in the first row. Since each window is of length h, $\mathbf{V}^T$ is a (h, K_T)-dimensional matrix. Each line corresponds to the estimation in a particular subsample and plots the first column of $\mathbf{V}^T, \mathbf{v}_1^C$. The four elements of $\mathbf{v}_1^C$ are close to 0.5, and there is little variation across subsamples. The average standard deviation is 0.0072. Increasing the window length to $h = 8$ and $h = 12$, shown in Panels B and C, yields similar results. The standard deviations across the longer windows are 0.0068 and 0.0062 and thus slightly lower than for $h = 4$.

The plots in the second row show the first columns of $\mathbf{V}^N$. Since $\mathbf{V}^N$ has $N = 934$ rows, I plot results for 25 randomly selected funds, but they are representative of the distribution across all mutual funds. The 25 funds are sorted by their standard deviations of estimated loadings across subsamples and are plotted on the x-axis. The estimated loadings are on the y-axis. The plots show that the estimates are consistent across windows and display only slight variation, even for a window length of four. Increasing h from four to eight and 12 reduces the variation as the average standard deviations decrease from 0.0042 for $h = 4$ to 0.0027 for $h = 12$. The results are similar for characteristics loading plotted in the third row. The estimates are similar across subsamples for $h = 4$, and increasing h reduces the variation further.

These results show that the Tucker model yields reliable estimates even for very short samples. Estimations with only four time series observations produce consistent results across subsamples with little variation across windows. The estimation is stable because the number of estimated parameters remains low relative to the sample size in short subsamples. The compression ratio for the full sample estimation is 99%, corresponding to 137 observations per estimated parameter. For $h = 4, 8, 12$, the compression ratios are 94%, 97%, and 98%, respectively, and the ratios of the number of data points to the number of parameters are 16, 33, and 49. The parsimony of the Tucker model is a distinct advantage compared to SVD/PCA models that typically require long samples.

4.3. Tucker vs. 2-dimensional SVD/PCA estimation

As mentioned in the introduction, an alternative to the 3-dimensional Tucker model is estimating a series of 2-dimensional SVD/PCA models along each dimension. For example, one could estimate matrix SVD/PCA models for each quarter t, which capture the factor structures of characteristics across mutual funds across time. Alternatively, estimating a 2-dimensional model for each mutual fund yields factor models of characteristics over time for a given fund. Finally, estimating SVD/PCA models for each characteristic captures the factor structures across funds across time for the given characteristic. Of course, any 2-dimensional model inevitably loses information. Any information in the time series is lost in the estimation by quarter, and estimation by fund does not exploit any cross-fund information. In this section, I investigate the properties of such 2-dimensional models and compare them to the Tucker model.

Let's say we are interested in the characteristic loadings. In the Tucker model, characteristic loadings are given by the columns of the $\mathbf{V}^C$ matrix. To compare the Tucker model to 2-dimensional factor models, I first estimate SVD/PCA with K_C factors for each quarter $t = 1, \ldots, T$. Let $\mathbf{X}_t$ be the (N, C)-matrix of characteristics of mutual funds in quarter t. Rows and columns of $\mathbf{X}_t$ correspond to funds and characteristics, respectively. $\mathbf{U}_K^1$ in (2) is the $(N \times K)$-matrix of eigenvalues of $\mathbf{X}_t \mathbf{X}_t^\intercal$ and $\mathbf{U}_K^2$ is the $(C \times K)$-matrix of eigenvalues of $\mathbf{X}_t^\intercal \mathbf{X}_t$. Characteristic loadings are thus given by the columns of $\mathbf{U}_K^2$.[25] The left column of Figure 11 shows the loadings of the first four factors. Each of the $T = 34$ lines corresponds to the estimation using data from a given quarter. Characteristics are ordered by their mean loadings across all quarters. For comparison, characteristic loadings estimated from a Tucker (2, 6, 5) model are plotted in black.

Panel A shows that the estimation of loadings of the first factor across quarters is stable for some characteristics but varies substantially for others. The loadings for SP and BM range from -0.35 to -0.30, and -0.31 to -0.26, respectively. On the other hand, for example, REV and MOM loadings vary significantly across quarters and range from -0.15 to 0.24, and -0.03 to 0.23, respectively. Note that the overall pattern of loadings is similar to the Tucker characteristic loadings (shown in black). Panel A of Figure 12 plots loadings across quarters instead of characteristics. Each line corresponds to the loadings of a characteristic over time. The plot confirms that most characteristics' loadings of the first factor are relatively stable over time. The average time series standard deviation across all characteristics, shown in the upper left corner of the plot, is 0.03. The loadings of the second characteristic factor shown in Panels C of Figures 11 and 12 exhibit somewhat more variations across quarters than those of the first factor. On average, the time series standard deviation of the second factor is 0.04.

In contrast to the first two factors, the loadings of the third factor exhibit substantial instability. Panel E of Figure 11 shows that loadings of most characteristics vary significantly across estimations by quarter. For example, the EPPROJ and EP loading ranges from -0.38 to 0.41, and -0.21 to 0.43, respectively. As a result, the SVD/PCA loadings are only weakly related to the Tucker estimates. The corresponding time series plot in Panel E of Figure 12 shows significant instability across the sample, in particular, in 2011 and 2016 when the signs of more than ten loadings switch signs. The mean standard deviation is 0.14, which is an order of magnitude than the standard deviation of the first two factors. The fourth factor, shown in Panel G, shows similar instability as the third factor. Unreported results indicate that factor loadings of higher-order factors are increasingly unstable across the sample.

Next, I estimate characteristic loadings of 2-dimensional SVD/PCA factor models for each mutual fund, $n = 1, \ldots, N$. Let $\mathbf{X}_n$ be the (T, C)-matrix of time series of characteristics for mutual fund n. $\mathbf{U}_K^1$ in (2) is the

[25] Since the signs of eigenvectors can vary across estimations, I normalize the signs so that they are consistent across quarters.

$(T \times K)$-matrix of eigenvalues of $\mathbf{X}_t \mathbf{X}_t^\top$ and $\mathbf{U}_K^2$ is the $(C \times K)$-matrix of eigenvalues of $\mathbf{X}_t^\top \mathbf{X}_t$, so that characteristic loadings are given by the columns of $\mathbf{U}_K^2$. The results are reported in the right columns of Figures 11 and 12. The plots show that SVD/PCA estimations by mutual fund yield unstable estimates of characteristic loadings. Recall that each line in Panel A corresponds to loadings of the first factor from the estimation for a single mutual fund. Except for PSLIQ, estimated loadings of characteristics differ substantially from fund to fund and often have different signs. The SVD/PCA loadings of the second, third, and fourth factors vary even more across funds and are essentially unrelated to the Tucker loadings. The results in the right column of Figure 12 show plots of loadings across the 25 largest mutual funds in the sample. Even for large funds, the estimates are unstable. Clearly, the estimation of characteristic loadings by fund yields unreliable results.

Another possible method to reduce the dimensionality is to take means across a given mode. For example, instead of estimating separate SVD/PCA models for all quarters, one could estimate factor models for the time series means of characteristics. As we have seen above, estimations for individual funds are unreliable, but perhaps a single estimation of average characteristics taken across all mutual funds yields better results. I compare the results of such mean estimations to those from the Tucker model in Figure 13, which shows loadings of the first four factors. Each panel shows the Tucker estimates in blue and the SVD/PCA results for time series and fund means in black and orange, respectively. I also plot the means of the SVD/PCA estimates for individual quarters and funds in dashed lines. Panels A and B show that the SVD/PCA estimate of the first two factors using time series means and the mean of the estimations by quarter are essentially identical to those from the Tucker model. The loadings for the specification using means across funds are also reasonably close to the Tucker estimates for the first factor; however, they deviate for the second factor. For example, the estimated MOM and REV loadings are near zero for the Tucker model but around 0.5 for the SVD/PCA model using fund means.

For the third and fourth factors, the results from 2-dimensional models are inconsistent and deviate significantly from those of the Tucker model. Consider first the estimation based on means across quarters (solid black) vs. means of SVD/PCA models for individual quarters (dashed black). The two approaches yield substantially different results for most characteristics. For example, the OP loadings of the third factor are -0.41 and 0.13, respectively. The correlations between the two loadings estimates are only 0.09 and 0.27 for the third and fourth factors. The corresponding results for estimations by mutual funds (in orange) yield similar results that are inconsistent for different specifications.

Finally, note that the number of estimated parameters in the 2-dimensional SVD/PCA models is significantly larger than the number of parameters in the Tucker model. The Tucker model with $(7, 13, 13)$ factors has 13,501 parameters and 793,900 data points. The SVD/PCA model for a given quarter has 12,298 parameters and 23,350 data points, so there are less than two data points for each parameter. The total number of parameters for all $T = 34$ quarters is 307,450, 31 times more than in the Tucker model. Each SVD/PCA model for a mutual fund has 598 parameters but only 850 data points, so there are 558,532 parameters for all $N = 934$ funds.[26]

This section has shown that estimating characteristic loadings using 2-dimensional factor models is problematic. First, the results differ when loadings are estimated by quarter or mutual fund. Second, estimations across quarters and funds are unstable, especially for higher-order factors. In other words, the 2-dimensional factor models are mutually inconsistent and unreliable. In contrast, the 3-dimensional Tucker model yields a unique set of characteristic loadings that is consistent across time and mutual funds.

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,b), 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 $n = 1,\ldots,N$. 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

[26] The 2-dimensional specifications have more parameters than the Tucker model even when the number of factors is reduced. For example, the total number of parameters for $K = 4$ factors for the estimation by quarter and by fund are 129,880 and 205,480, respectively.

“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 (11)-(13). (13) shows that the Tucker model implies a 2-dimensional factor model for the characteristics of each mutual fund. $\mathbf{F}_n^C$ 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 $\mathbf{F}_n^C$ of all mutual funds can be computed from its 3-dimensional representation, see (11) and (12), rather than having to estimate the factors separately for each fund. Moreover, since all $\mathbf{F}_n^C, n = 1, \ldots, N$ 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 $\mathbf{F}_n^C$. 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.[27] I repeat this procedure for $t = 1, \ldots, T-1$ 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 yielding K_C Tucker characteristic factors.[28]

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 $t = 1, \ldots, T'$ for $T' = K_T, \ldots, T-1$. 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 $K_T + 1, \ldots, T$ 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. The correlations of out-of-sample portfolios with the corresponding in-sample portfolios are above 0.95.

I compare the Tucker factors to several benchmarks. First, I construct traditional PCA-based factors as follows. For each characteristic, I create decile-sorted portfolios in quarter t and compute their returns in quarter $t+1$. I obtain factors by estimating the SVD/PCA model for the panel of $10C = 250$ portfolio returns. I consider an in-sample estimation, as well as a recursive estimation using expanding windows using the same methodology as for the TFM. Second, I consider the Fama-French factors SML, HML, MOM, CMA, and INV.

Table 5 shows descriptive statistics of excess returns of in-sample (Panel A) and out-of-sample (Panel

[27] 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.

[28] To distinguish characteristic factors derived from the Tucker model and pricing factors, I use lowercase $\mathbf{f}_t$ for the latter. Since the signs of Tucker components are not identified, I normalize the pricing factors so that their means are positive.

B) factors derived from a Tucker model with $(2,6,5)$ factors. The fourth in-sample Tucker factor has the highest Sharpe ratio of 0.84, followed by the third with a Sharpe ratio of 0.70. The Sharpe ratios of the remaining factors are 0.55, 0.36, and 0.22. The Sharpe ratios of the out-of-sample Tucker factors range from 0.69 to 0.32.

The Sharpe ratio of the CRSP-VW index is 0.94 over the sample period. It is well known that the returns of many characteristic factors have been lower than in previous decades. The mean MOM return is 3.72%, and its Sharpe ratio is 0.42. The mean returns of SMB, HML, RMW, and CMA are close to zero or negative.[29] Panels D and E show the returns of SVD/PCA factors. The first in-sample and out-of-sample factor is long-only and highly correlated with the CRSP-VW index. The Sharpe ratios are slightly lower than that of the CRSP-VW index. The third in-sample and out-of-sample factors have Sharpe ratios of 0.67 and 0.64, respectively, but the Sharpe ratios of the remaining factors are substantially smaller.

5.1. 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 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 $\mathbf{F}^{C}_{t,(n)}$ as independent variables. There are several methods to estimate such regressions. The most popular approach is due to 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 to estimate the model and include time fixed effects to capture variation across quarters:

$$R^{e}_{n,t+1} = \alpha + \boldsymbol{\beta}^{\top} \mathbf{F}^{C}_{t} + \gamma_t + e_{n,t+1}, \tag{35}$$

where $R^{e}_{n,t+1}$ is the excess return of fund n in quarter $t+1$ and $\mathbf{F}^{C}_{t,(n)}$. γ_t are time fixed effects. The $\mathbf{F}^{C}_{t}$ 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 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 removing time effects.

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

[29] The CMA factor has a Sharpe ratio of 0.01 and is omitted from the table.

factors and regression coefficients in (35), all regression coefficients are positive. The HAC t-statistics and t-statistics based on entity-clustered standard errors are similar for most factors and indicate statistical significance of all coefficients. Time-clustered t-statistics are significantly smaller, suggesting that the identification stems from the mutual fund dimension rather than the time dimension, which is unsurprising given the short sample span. The between-R^2 is 41%, which implies that Tucker factors capture a substantial part of the variation in returns across mutual funds. The results for out-sample factors in Panel B are comparable to those for in-sample factors. For comparison, I run the panel regression (35) using original characteristics as independent variables. The coefficients and t-statistics for the five 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 only slightly higher than that of the regression with five Tucker factors.

These results suggest that the Tucker characteristic factors are linked to mutual fund returns. Next, I 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 and 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 $n = 1, \ldots, N$:

$$R^e_{nt} = \alpha_n + \boldsymbol{\beta}_n^\top \mathbf{f}_t + e_{nt}, \tag{36}$$

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 (36), including the market return but excluding the constant. The pricing error, or alpha, of fund n is α_m. I evaluate models by their root-mean-squared pricing error (RMSPE), the mean-absolute pricing errors (MAPE), and the mean pricing error (MPE) across all mutual funds:

$$\text{RMSPE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\alpha_n^2}, \quad \text{MAPE} = \frac{1}{N}\sum_{n=1}^{N}|\alpha_n|, \quad \text{MPE} = \frac{1}{N}\sum_{n=1}^{N}\alpha_n. \tag{37}$$

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.[30] 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 $L = 4$.

Consider first the CAPM (first row of Panel B). The RMSPE and MAPE of the CAPM are 3.27% and 2.45%, respectively. The average CAPM alpha is -1.66% indicating that mutual funds underperform on average relative to the CAPM. The pricing errors of 215 mutual funds are statistically significantly different from

[30]Moreover, models with more than $L = 4$ factors do not improve the results in any of the specifications.

zero, and alphas of 709 of the 934 funds in the sample are negative.[31]

Panel A reports the results of in-sample factors derived from the Tucker model with (2,6,5) components. Adding the third Tucker factor to the CRSP-VW index reduced the RMSPE and MAPE to 2.15% and 1.67%. The fit improves further when the first Tucker factor is included. The RMSPE and MAPE of this model with $L = 3$ factors are 1.85% and 1.45%. Adding a fourth factor has a minor effect on pricing errors. For the model with three factors, the pricing errors of 71 mutual funds are individually statistically significant compared to 215 for the CAPM. 607 funds have negative alphas.

Consider next the specifications with Fama-French factors in Panel B. The pricing errors are substantially higher than those for Tucker factors. For example, the RMSPE of the 4-factor model with SMB, HML, and MOM is 2.35% is higher than that of the Tucker model with only two factors. 115 individual pricing errors are significant. Note that including the profitability and investment factors CMA and INV increases the RMSE and the corresponding results are, therefore, not reported.

The results for SVD/PCA factors are in Panel D. Models with SVD/PCA factors are comparable to Fama-French models but worse than models with Tucker factors. For example, the RMSPE of the $L = 3$ model is 2.43% compared to 2.59%, and 1.85% for 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 Tucker factors are added to the model, the RMSPE is reduced to 1.85%, see the "RMSPE*-Tucker" column. When SVD/PCA factors are included instead, the RMSPE shrinks to 2.43% (RMSPE*-PCA column). Adding Fama-French or PCA factors to Tucker factors worsens the fit as the RMSPE increase. In contrast, adding Tucker factor to Fame-French or PCA factors improves the fit in all cases. These results suggest 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. Since the estimation windows in the out-of-sample estimation are short, I also consider the more parsimonious Tucker specification with (1,3,3) factors. The pricing errors, reported in

[31]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.

Panel B, are only slightly higher than those of the model with $(2,6,5)$ factors. 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 three out-of-sample factors is 2.76% compared to 2.43% in the in-sample model.

Figure 14 illustrates the fit of the models. It shows scatter plots of mean fund returns on the x-axes and fitted mean returns om the y-axes. The patterns of fitted mean returns of the in-sample and out-of-sample Tucker models in Panels A and B are similar. The pricing errors for most funds are small but there are some outliers with large errors. Panels C and D show of Figure 14 show fitted returns of the CAPM and the 3-factor Fama-French models. Their pricing errors are substantially larger that those for the Tucker models in Panels A and B. The fit is worse particularly for mutual funds with low returns. Pricing errors of the model with in-sample SVD/PCA factors are similar to those of the 3-factor Fama-French model. The pattern of fitted mean returns of out-of-sample SVD/PCA model is similar but pricing errors are magnified.

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.

5.2. Pricing errors

Next, I investigate the properties of pricing errors of some of the Tucker factor models studied above. I focus on models with $L = 3$ factors, 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). Since there are only 12 funds in the "other" category, I exclude these funds from the discussion. 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 4.19%, followed by "cap-based" funds with a mean of 2.34%. Leaving "other" funds aside, "growth" and "balanced" funds have the lowest pricing errors with means of 1.85% and 1.91%, respectively.

Consider next the fit by model. The pricing errors of the in-sample and out-of-sample Tucker models are similar. Without exception, RMSPE are the smallest for Tucker factors across all categories, showing that the superior 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, so the regression is given by

$$\alpha_n = \gamma_0 + \boldsymbol{\gamma}^\top \mathbf{z}_n + e_n,. \tag{38}$$

I standardize the independent variables to have zero means and unit standard deviations to compare the regression coefficients. 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.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 for the in-sample Tucker model with (2,6,5) factors. The three largest coefficients are 2.18 (ME), 1.90 (INV, and 1.82 (MS), and the three smallest are -2.03 (VOL), -1.32 (SG), and -0.99 (BG). Note that patterns of regression coefficients are related to the fit of the Tucker model (Panel C of Figure 6). For example, the MSE of the Tucker model are highest for MOM and REV, which are the characteristics with the largest coefficients in absolute value. The mean of the absolute values of the coefficients is 0.74, and 11 are statistically significant at the 5% level. The results for the model with out-of-sample Tucker in the second column are similar to those for in-sample factors. The mean of absolute coefficients is 0.73.

The coefficients of the CAPM, the 3-factor Fama-French, and the PCA models in columns three to six are generally larger (in absolute value) and more statistically significant than those for the Tucker models. The means of (absolute) values of the coefficients are 1.46, 1.26, 1.09, and 1.62, respectively. 19, 16, 15, and 19 of the 25 coefficients are statistically significant at the 5% level. While the Tucker models do not capture the link between characteristics and returns completely, 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 and report the results in Table 11. 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. As a benchmark, I first regress the difference of mean returns of individual funds, $\overline{R}_m$, and the grand mean of fund returns, $\overline{\overline{R}}_m$ 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 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, implying no strong link between fund size and mean returns.

The results in this section show that factors derived from TFM are linked to mutual fund returns. Tucker factors have higher Sharpe ratios and smaller cross-sectional pricing errors than Fama-French and PCA factors. A 3-factor model with the CRSP-VW index and two Tucker factors yields smaller pricing errors than other models with more factors. Out-of-sample Tucker factors that are not subject to a look-ahead bias are highly correlated with their in-sample counterparts and capture the cross-section of fund returns as well.

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.

Bai, Jushan and Serena Ng (2002) “Determining the Number of Factors in Approximate Factor Models,” *Econometrica*, 70 (1), 191–221.

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,” *Review of Financial Economics*.

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 (2024) “Deep Learning in Asset Pricing,” *Management Science*, 20 (2), 714–750.

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, and Mao Ye (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 (r1, r2,r3) 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.

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, Avanidhar (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. (1963) "Implications of Factor Analysis of Three-Way Matrices for Measurement of Change," in Harris, Chester W. ed., *Problems in Measuring Change*, 122–137, University of Wisconsin Press.

——— (1966) "Some mathematical notes on three-mode factor analysis," *Psychometrika*, 31 (3), 279–311.

Table 1: Mutual fund characteristics

Category	Characteristics	Category	Characteristics
Multiples	Book-to-market (BM)	Morningstar	Value/growth (MS)
	Earnings-to-price (EP)		Multiples (MULT)
	Projected ep (EPPROJ)		Growth rates (GR)
	Cash flow-to-price (CFP)	Momentum/reversal	Cumulative return $t-7$ to $t-2$ (MOM)
	Sales-to-price (SP)		Cumulative return $t-12$ to $t-7$ (REV)
	Dividend-to-price (DP)	Liquidity	Bid-ask spread (BIDASK)
	Industry-adjusted book-to-market (ADJBM)		Pastor-Stambaugh (PSLIQ)
Growth rates	Earnings (EG)		Turnover (TURN)
	Long-term earnings (LTEG)		Volume (VOL)
	Book value (BG)	Other	Market cap (ME)
	Cash flow (CFP)		Operating profitability (OP)
	Sales (SG)		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.
			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 β		1.00	0.15	0.95	1.03	1.10
CAPM α (% p.a.)		-1.59	2.77	-3.01	-1.71	-0.06
			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)	σ_e/σ_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), $\boldsymbol{\mathcal{X}}_i$ and estimate the Tucker model for the true model plus noise, $\boldsymbol{\mathcal{X}}^e = \boldsymbol{\mathcal{X}} + \sigma_e \boldsymbol{\mathcal{E}}$. The table reports the mean RMSE of $\boldsymbol{\mathcal{X}} - \widehat{\boldsymbol{\mathcal{X}}}$. The columns correspond to different values of the standard deviation of the noise tensor, σ_e, relative to the standard deviation of the true factor tensor, σ_x.

Table 4: Properties of Tucker models

	(K_T, K_N, K_C)		
	(1, 3, 3)	(2, 6, 5)	(5, 15, 8)
	Panel A: Properties		
MSE	0.09	0.07	0.04
R^2	83.63%	87.13%	92.12%
$\tilde{R}^2$	83.63%	86.49%	87.18%
$\tilde{K}$	3	9	18
Comp. ratio	99.63%	99.27%	98.15%
	Panel B: Errors		
Mean	−0.01	−0.00	−0.00
Median	−0.00	−0.00	−0.00
Min	−2.57	−2.77	−2.14
1%	−1.06	−0.89	−0.67
5%	−0.49	−0.42	−0.33
25%	−0.17	−0.15	−0.12
75%	0.16	0.14	0.11
95%	0.46	0.42	0.32
99%	0.90	0.87	0.68
Max	2.97	2.99	2.95

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: Factor returns

A: In-sample Tucker factors					
	4	3	1	2	5
Mean	3.71	3.57	4.71	3.02	1.07
Std. dev.	4.39	5.10	8.55	8.42	4.97
SR	0.84	0.70	0.55	0.36	0.22
B: Out-of-sample Tucker factors					
	5	3	1	2	4
Mean	2.63	3.23	3.67	3.01	1.50
Std. dev.	3.78	5.51	8.60	8.11	4.66
SR	0.69	0.59	0.43	0.37	0.32
C: Fama-French factors					
	MKT	MOM	RMW	SMB	HML
Mean	12.20	3.72	1.24	−0.57	−1.35
Std. dev.	12.92	8.80	5.89	7.59	8.69
SR	0.94	0.42	0.21	−0.08	−0.16
D: In-sample SVD/PCA factors					
	1	3	5	4	2
Mean	9.68	8.56	0.84	0.85	2.93
Std. dev.	12.92	12.88	5.65	7.12	26.22
SR	0.75	0.67	0.15	0.12	0.11
E: Out-of-sample SVD/PCA factors					
	1	3	5	2	4
Mean	11.48	8.58	2.05	8.54	0.87
Std. dev.	12.92	13.43	5.58	26.53	6.70
SR	0.89	0.64	0.37	0.32	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), Fama-French factors (Panel C), and factors based on fund characteristics (Panel D). Tucker factors are derived from Tucker(2,6,5) models. The sample period is 2010Q3 to 2018Q4.

Table 6: Panel regression with Tucker factors

	1	2	3	4	5	R^2
			A: In-sample factors			
$\boldsymbol{\beta}$	0.25	0.11	0.19	0.10	0.26	0.41
t-HAC	(8.76)	(5.59)	(7.26)	(3.08)	(6.94)	
t-entity	(9.03)	(6.47)	(5.52)	(2.67)	(10.50)	
t-time	(0.97)	(0.59)	(1.80)	(0.98)	(1.66)	
			B: Out-of-sample factors			
$\boldsymbol{\beta}$	0.04	0.26	0.22	0.09	0.26	0.28
t-HAC	(1.67)	(11.76)	(7.19)	(2.94)	(8.36)	
t-entity	(1.37)	(10.68)	(4.69)	(3.73)	(5.74)	
t-time	(0.19)	(1.37)	(1.95)	(0.95)	(2.63)	
			C: Characteristic factors			
	MS	MULT	REV	EP	ELTG	
$\boldsymbol{\beta}$	0.61	0.47	0.44	0.33	0.28	0.45
t-HAC	(2.33)	(1.73)	(9.04)	(4.10)	(2.00)	
t-entity	(2.48)	(1.57)	(8.66)	(4.94)	(1.97)	
t-time	(1.04)	(0.64)	(2.24)	(1.37)	(0.51)	

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

$$R^e_{n,t+1} = \alpha + \boldsymbol{\beta}^\top \mathbf{F}^C_{t,(n)} + \gamma_t + e_{n,t+1},$$

where $R^e_{n,t+1}$ is the excess return of fund n in quarter $t+1$ and $\mathbf{F}^C_{t,(n)}$. γ_t are time fixed effects. The $\mathbf{F}^C_{t,(n)}$ 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 – In-sample factors

Factors	L	MPE	MAPE	RMSPE	RMSPE*		
					Tucker	FF	PCA
			A: Tucker(2,6,5) factors				
MKT, 3	2	-0.43%	1.67%	2.15%		2.45%	2.18%
MKT, 1, 3	3	-0.48%	1.45%	1.85%		2.15%	1.92%
MKT, 1, 3, 5	4	-0.54%	1.42%	1.82%		2.15%	1.85%
			B: CAPM and Fama-French factors				
MKT	1	-1.66%	2.45%	3.27%	1.85%		2.43%
MKT, SMB	2	-0.77%	2.09%	2.85%	1.94%		2.61%
MKT, SMB, HML	3	-0.83%	1.87%	2.59%	2.15%		2.44%
MKT, SMB, HML, MOM	4	-0.73%	1.75%	2.35%	2.15%		2.45%
			C: PCA factors				
MKT, 3	2	-0.75%	1.88%	2.58%	1.88%	2.64%	
MKT, 3, 5	3	-0.70%	1.83%	2.43%	1.92%	2.44%	
MKT, 3, 5, 2	4	-0.79%	1.74%	2.32%	2.07%	2.44%	

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") are added to the specification. The sample period is 2010Q3 to 2018Q4.

Table 8: Pricing errors – Out-of-sample factors

Factors	L	MPE	MAPE	RMSPE	RMSPE*		
					Tucker	FF	PCA
			A: Tucker(2,6,5) factors				
MKT, 3	2	-0.44%	1.63%	2.10%		2.12%	2.29%
MKT, 1, 3	3	-0.50%	1.41%	1.79%		1.80%	1.99%
MKT, 1, 3, 2	4	-0.55%	1.40%	1.79%		1.79%	2.14%
			B: Tucker(1,3,3) factors				
Mkt-RF, 3	2	-0.42%	1.67%	2.19%		2.21%	2.46%
Mkt-RF, 1, 3	3	-0.46%	1.44%	1.84%		1.86%	2.04%
Mkt-RF, 1, 3, 2	4	-0.52%	1.43%	1.84%		1.84%	2.19%
			C: PCA factors				
MKT, 3	2	-1.02%	2.10%	2.92%	2.19%	2.80%	
MKT, 3, 4	3	-0.97%	2.01%	2.76%	2.27%	2.73%	
MKT, 3, 4, 5	4	-0.94%	1.98%	2.74%	2.28%	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	Mean
Tucker IS	1.74%	1.63%	1.81%	1.50%	2.70%	0.92%	1.72%
Tucker OOS	1.71%	1.67%	1.66%	1.50%	2.69%	0.95%	1.69%
CAPM	3.37%	2.25%	2.05%	3.06%	5.84%	1.19%	2.96%
FF3	2.54%	2.04%	1.85%	1.92%	4.86%	1.20%	2.40%
PCA IS	2.12%	1.77%	1.81%	1.92%	3.73%	1.42%	2.13%
PCA OOS	2.55%	2.09%	1.89%	2.22%	5.31%	1.31%	2.56%
Mean	2.34%	1.91%	1.85%	2.02%	4.19%	1.17%	2.24%

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

	Tucker IS	Tucker OOS	CAPM	FF3	PCA IS	PCA OOS
const	−0.48***	−0.50***	−1.66***	−0.83***	−0.36***	−0.97***
	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)
BM	−0.16	−0.27	−1.99***	−2.03***	−0.70**	−1.60***
	(0.32)	(0.30)	(0.33)	(0.34)	(0.33)	(0.36)
EP	0.44*	0.33	0.68**	0.58**	1.19***	0.48
	(0.26)	(0.25)	(0.27)	(0.28)	(0.27)	(0.30)
EPPROJ	−0.65	−1.14**	−3.12***	−2.98***	−1.07*	−3.35***
	(0.59)	(0.56)	(0.61)	(0.63)	(0.61)	(0.67)
CFP	−0.38	−0.67**	−2.16***	−1.73***	−1.43***	−1.86***
	(0.32)	(0.31)	(0.33)	(0.35)	(0.33)	(0.37)
SP	−0.04	0.00	0.20	0.10	0.54***	0.00
	(0.19)	(0.18)	(0.20)	(0.21)	(0.20)	(0.22)
DP	−0.69***	−0.71***	−1.21***	−1.12***	−0.62***	−1.32***
	(0.23)	(0.22)	(0.24)	(0.24)	(0.23)	(0.26)
ADJBM	−0.18	−0.20	0.59***	0.37*	0.28	0.40*
	(0.19)	(0.18)	(0.19)	(0.20)	(0.19)	(0.21)
EG	0.19	0.48	0.93***	0.41	0.33	1.23***
	(0.33)	(0.31)	(0.34)	(0.36)	(0.34)	(0.38)
ELTG	−0.15	−0.09	0.75	−0.21	−0.68	1.68**
	(0.69)	(0.66)	(0.72)	(0.74)	(0.71)	(0.79)
BG	−0.99***	−0.86***	−1.11***	−1.01***	−1.05***	−0.95***
	(0.29)	(0.27)	(0.30)	(0.31)	(0.30)	(0.33)
CFG	0.13	−0.03	−0.31	−0.21	0.47*	−0.39
	(0.27)	(0.26)	(0.28)	(0.29)	(0.28)	(0.31)
SG	−1.32***	−0.85**	1.99***	1.11**	−0.45	2.09***
	(0.42)	(0.40)	(0.43)	(0.45)	(0.43)	(0.47)
MS	1.82	1.12	1.86	3.20**	2.57**	−0.66
	(1.22)	(1.16)	(1.27)	(1.31)	(1.25)	(1.39)
MULT	0.08	1.06	3.25**	2.69*	−0.55	4.83***
	(1.38)	(1.31)	(1.43)	(1.48)	(1.42)	(1.57)
GR	0.58	−0.47	−4.31***	−1.41	0.92	−6.74***
	(1.15)	(1.09)	(1.19)	(1.23)	(1.18)	(1.30)
MOM	0.95***	0.89***	0.56	0.56	0.97***	0.95**
	(0.34)	(0.32)	(0.35)	(0.36)	(0.35)	(0.38)
REV	−0.83**	−0.66*	0.12	0.04	−0.56	−0.51
	(0.38)	(0.36)	(0.39)	(0.40)	(0.39)	(0.43)
BIDASK	1.71***	1.63***	1.63***	1.36***	1.86***	1.60***
	(0.27)	(0.26)	(0.28)	(0.29)	(0.28)	(0.31)
PSLIQ	0.72***	0.71***	0.93***	1.02***	0.69***	0.81***
	(0.18)	(0.17)	(0.19)	(0.19)	(0.18)	(0.20)
TURN	0.26**	0.20**	−0.48***	−0.30**	0.33***	−0.30**
	(0.11)	(0.10)	(0.11)	(0.12)	(0.11)	(0.12)
VOL	−2.03***	−1.97***	−2.51***	−3.77***	−3.74***	−3.11***
	(0.57)	(0.54)	(0.59)	(0.61)	(0.59)	(0.65)
ME	2.18***	2.16***	3.13***	2.92***	4.28***	3.45***
	(0.55)	(0.52)	(0.57)	(0.59)	(0.56)	(0.62)
OP	0.03	0.06	0.76***	0.44**	0.40**	0.53***
	(0.17)	(0.16)	(0.17)	(0.18)	(0.17)	(0.19)
INV	1.90***	1.65***	1.35***	1.36***	1.42***	1.21***
	(0.33)	(0.32)	(0.35)	(0.36)	(0.34)	(0.38)
QUAL	−0.14	−0.06	−0.45***	−0.55***	−0.09	−0.47***
	(0.10)	(0.10)	(0.11)	(0.11)	(0.11)	(0.12)
Adj. R^2	0.21	0.22	0.65	0.51	0.44	0.51

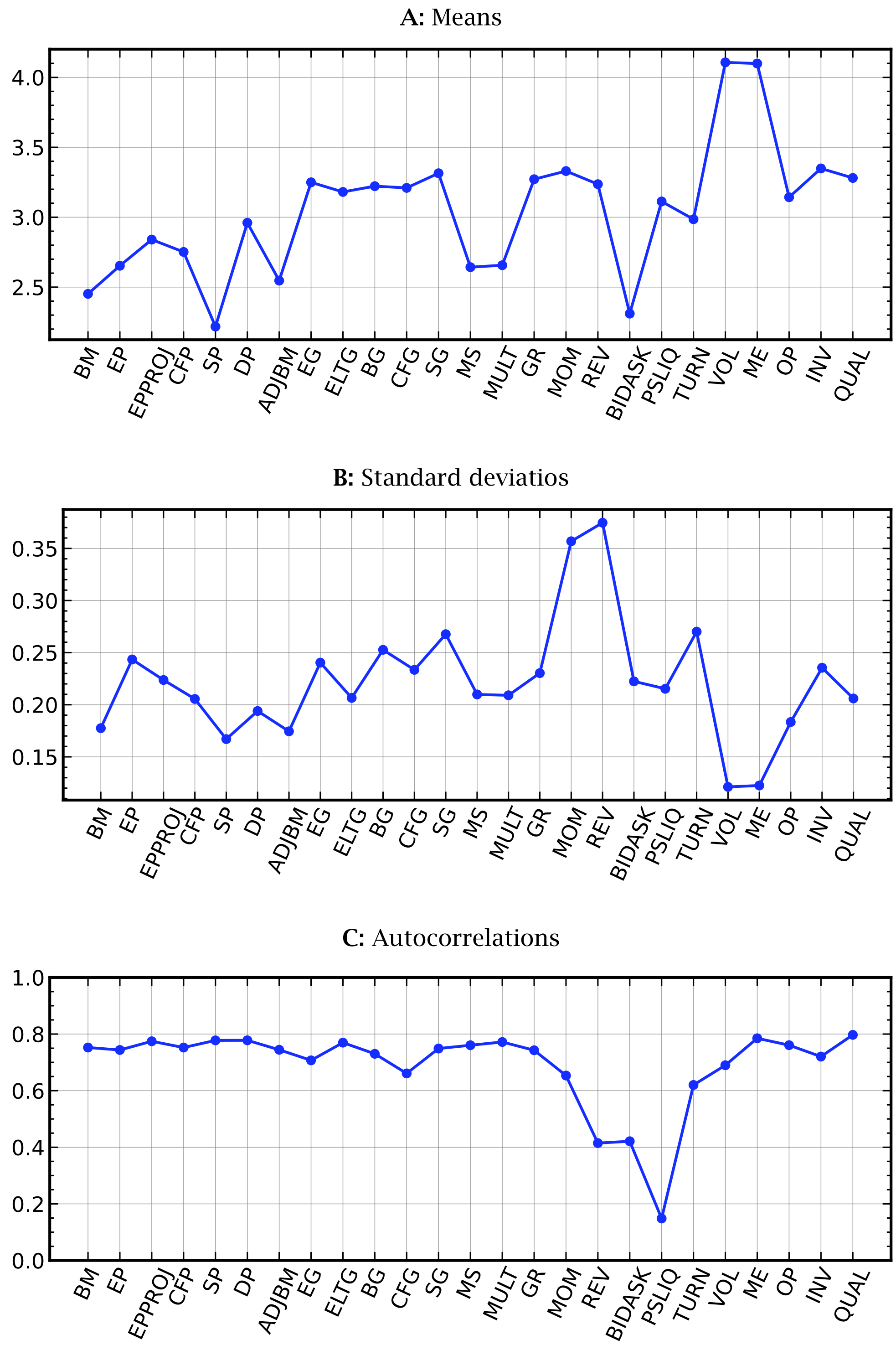
This table reports the results of regressions of pricing errors from factor models on the average expense ratio (Exp. ratio), the average number of stocks (No. stocks), and the average active share. The variables are standardized to have means of zero and unit standard deviations. The factors in the 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
const	0.10	−0.43***	−0.44***	−1.59***	−0.74***	−0.30***	−0.91***
	(0.09)	(0.06)	(0.06)	(0.09)	(0.08)	(0.07)	(0.09)
log(TNA)	0.12	0.12*	0.13**	0.20**	0.09	0.14*	0.21**
	(0.11)	(0.07)	(0.07)	(0.10)	(0.09)	(0.08)	(0.10)
log(No. stocks)	0.19*	0.21***	0.22***	−0.39***	0.44***	0.13	−0.01
	(0.11)	(0.07)	(0.07)	(0.11)	(0.10)	(0.09)	(0.10)
Turnover	0.27**	0.14**	0.16**	−0.25**	−0.31***	0.40***	−0.28***
	(0.11)	(0.07)	(0.07)	(0.10)	(0.09)	(0.08)	(0.10)
Exp. ratio	−0.34***	−0.28***	−0.26***	−0.28**	−0.23**	−0.25***	−0.22**
	(0.12)	(0.07)	(0.07)	(0.11)	(0.10)	(0.09)	(0.11)
Active share	−0.28**	0.41***	0.35***	−0.64***	0.51***	0.03	0.09
	(0.12)	(0.08)	(0.07)	(0.11)	(0.10)	(0.09)	(0.11)
Adj. R^2	0.05	0.08	0.08	0.12	0.05	0.06	0.03

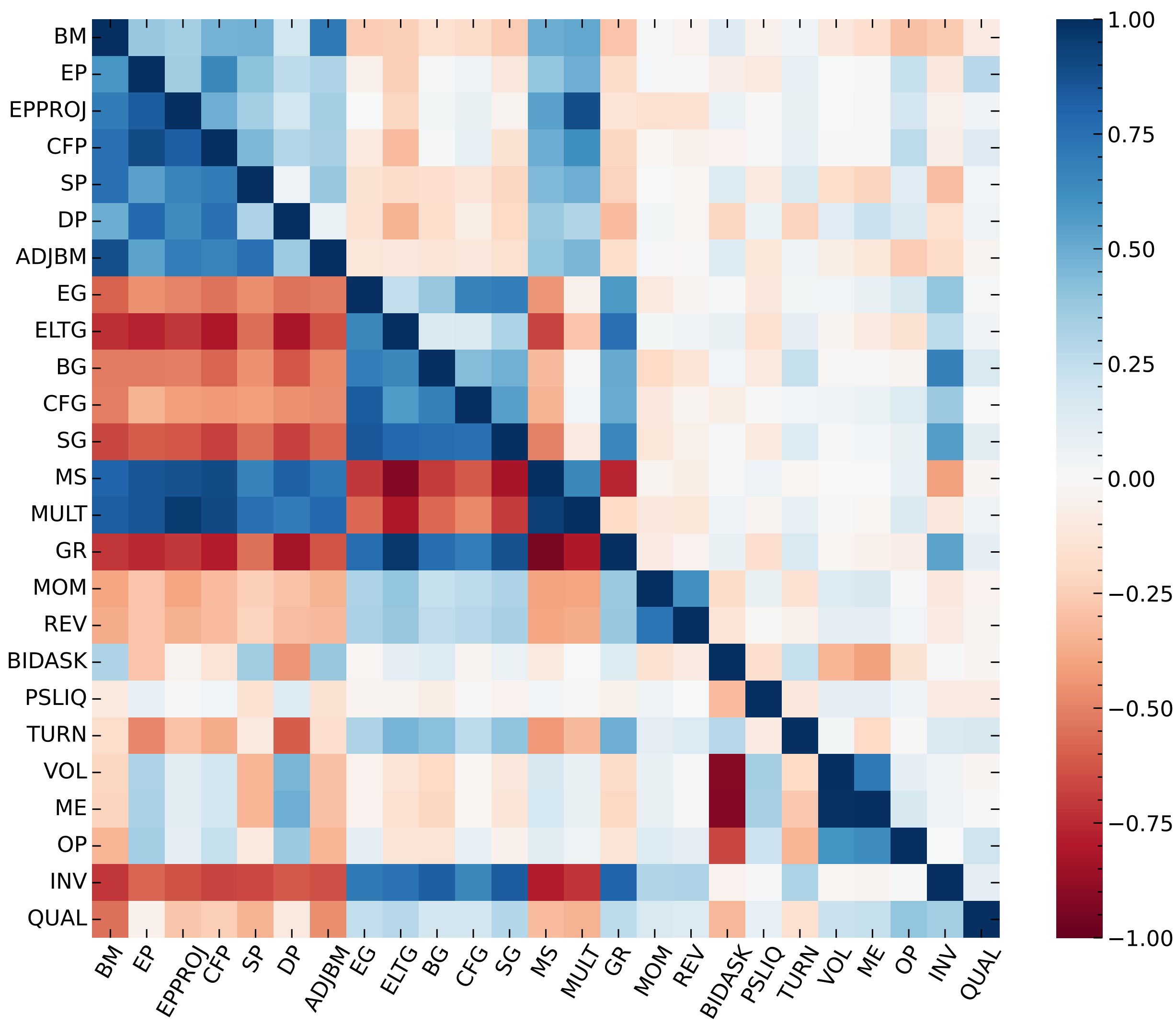
This table reports the results of regressions of pricing errors from factor models on the average log TNA, log of numbers of stocks in a fund's portfolio, turnover, expense ratio, and the average active share. The variables are standardized to have means of zero and unit standard deviations. The factors 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: Standard deviations and autocorrelations by characteristics



Notes: The figure plots the time series means, standard deviations, and first-order autocorrelations of characteristics averaged across mutual funds. The sample period is 2010Q3 to 2018Q4.

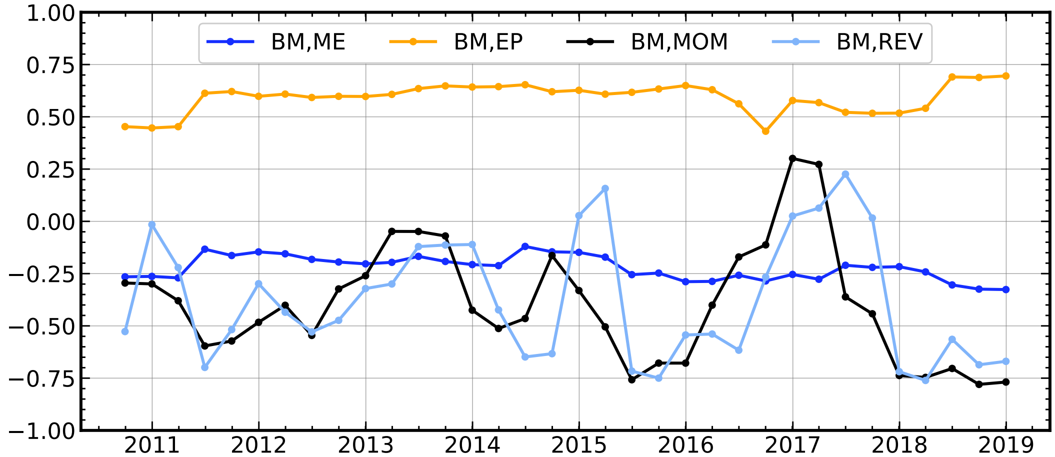
Figure 2: Cross-correlations of characteristics



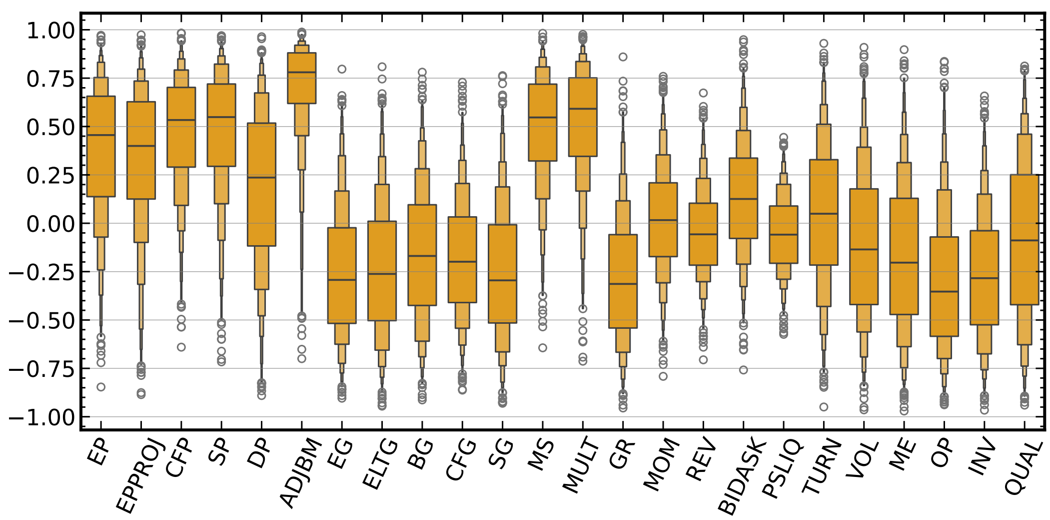
Notes: The figure shows the heatmap of pairwise correlations of mutual fund characteristics. First, I compute times-series correlations of characteristics by mutual funds and then average across funds, Second, I compute cross-sectional correlations of characteristics by quarter and then average across 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 3: Time-series and cross-sectional correlations

A: Cross-correlations over time



B: Time-series correlations across mutual funds – BM



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 box plots of the distributions of pairwise characteristic correlations of BM. The sample period is 2010Q3 to 2018Q4.

Figure 4: Tucker decomposition $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \times_3 \mathbf{V}_3$

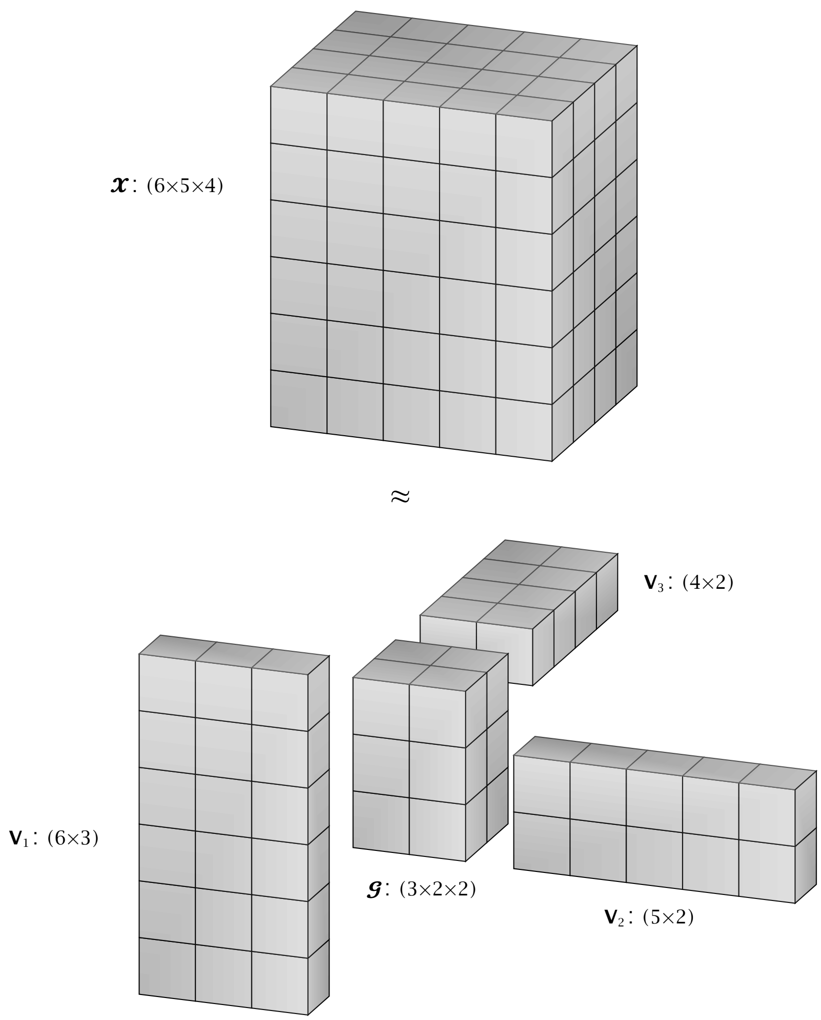
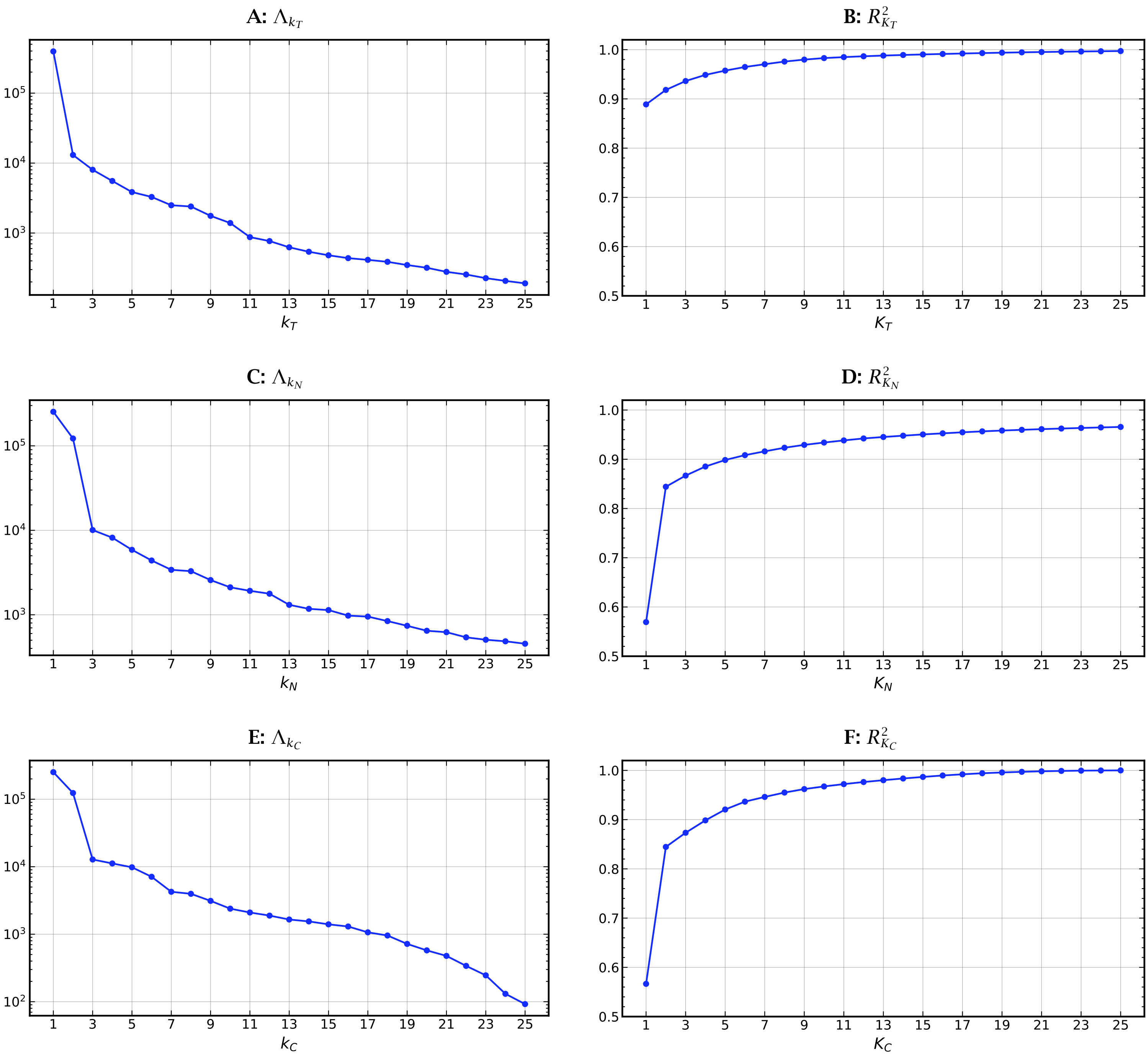


Figure 5: Core tensor – scree plots



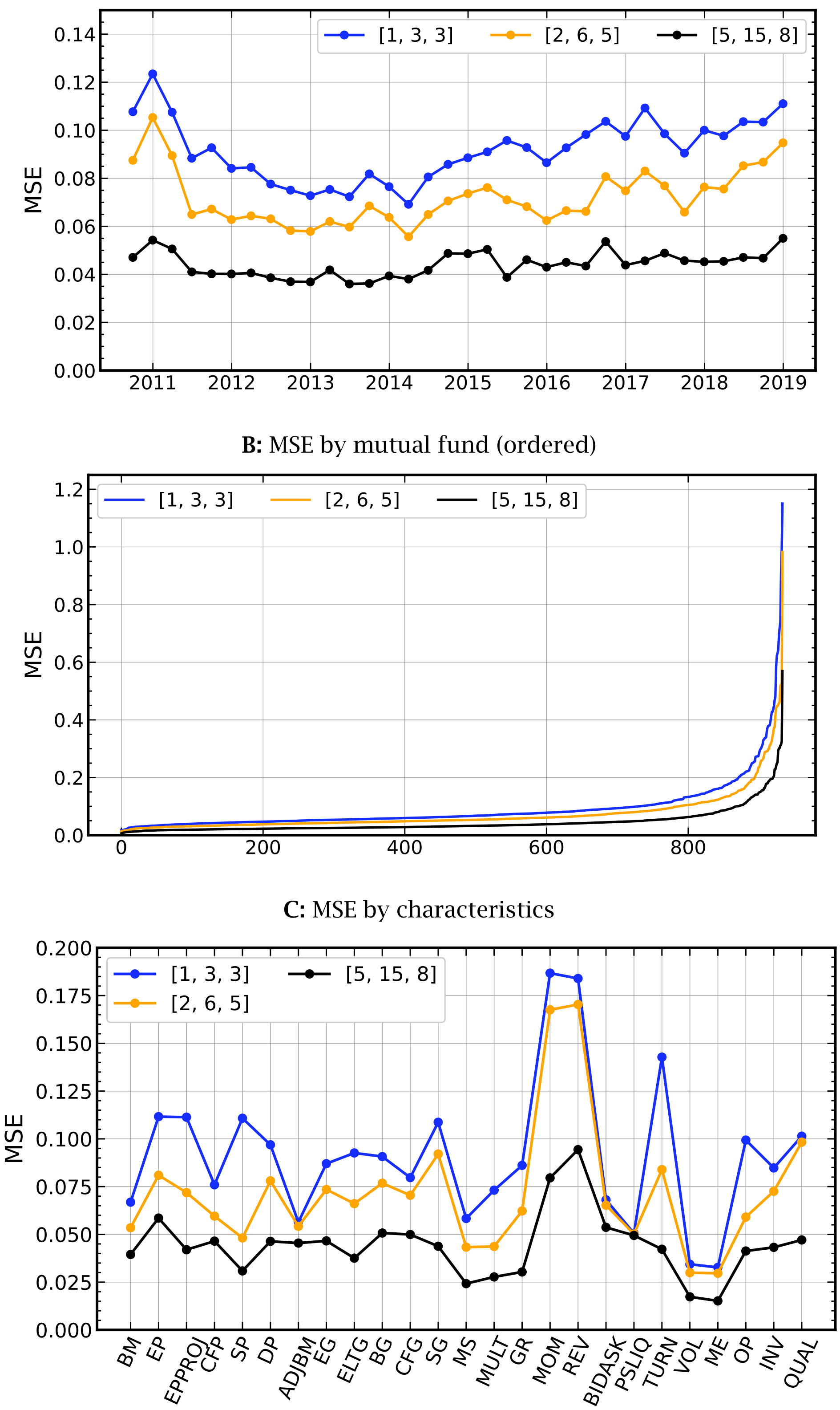
Notes: The left column shows plots the contributions of factors in the T,N, and C dimensions given by $\Lambda_{k_T} = \sum_{n,c} g^2_{k_T nc}$,$\Lambda_{k_C} = \sum_{t,c} g^2_{tk_N c}$, and $\Lambda_{k_C} = \sum_{t,n} g^2_{tnk_C}$, respectively. The right columns show the $R^2_{K_i} = \sum_{k_i=1}^{K_i} \Lambda_{k_i} / \|\boldsymbol{\mathcal{X}}\|^2$ when up to K_i factors are included. The plots are based on a Tucker(34,934,25) model. The sample period is 2010Q3 to 2018Q4.

Figure 6: Fit of Tucker models by dimension

A: MSE by quarters

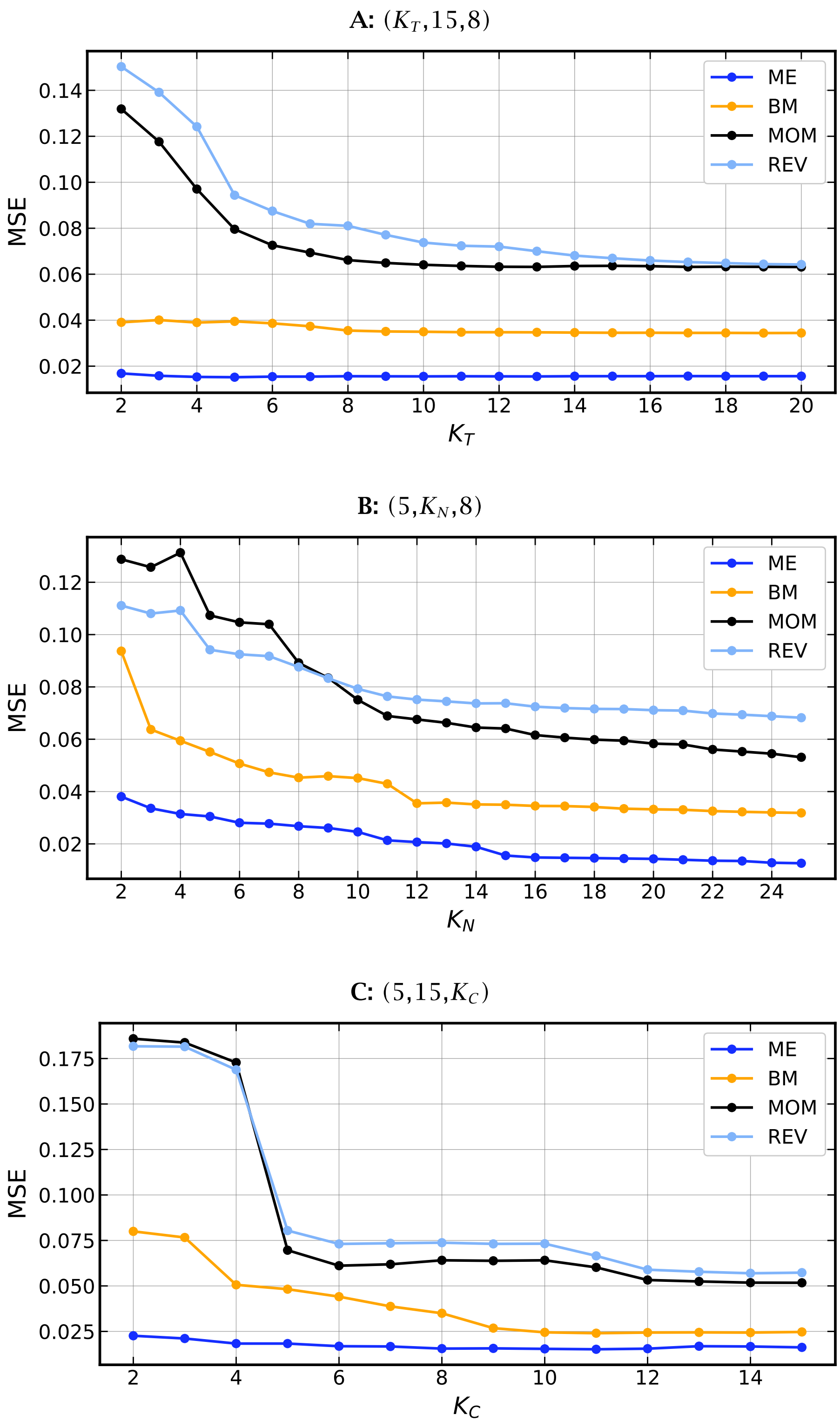
B: MSE by mutual fund (ordered)

C: MSE by characteristics



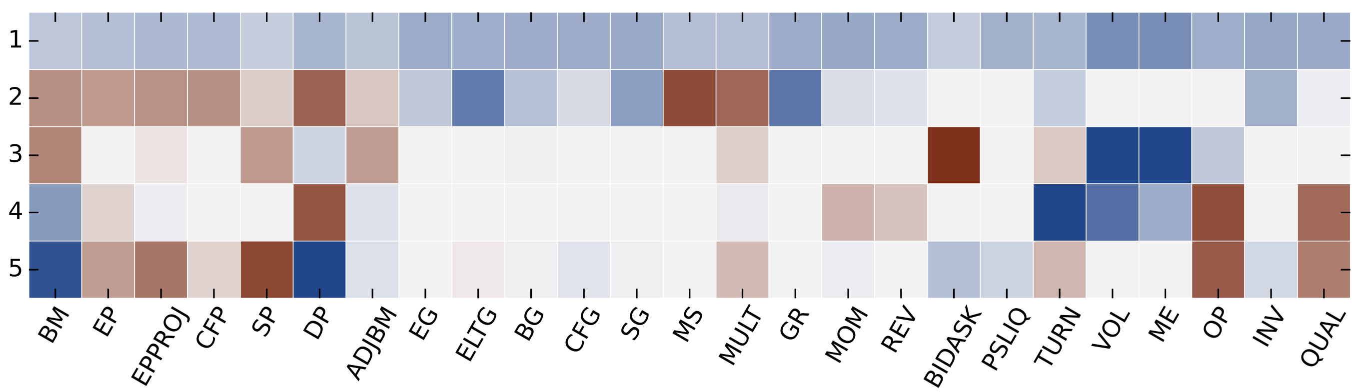
Notes: The figure plots the means of mean-squared errors (MSE) by quarters, mutual funds, and characteristics, respectively. The sample period is 2010Q3 to 2018Q4.

Figure 7: MSE of ME, BM, MON, and REV by K_i



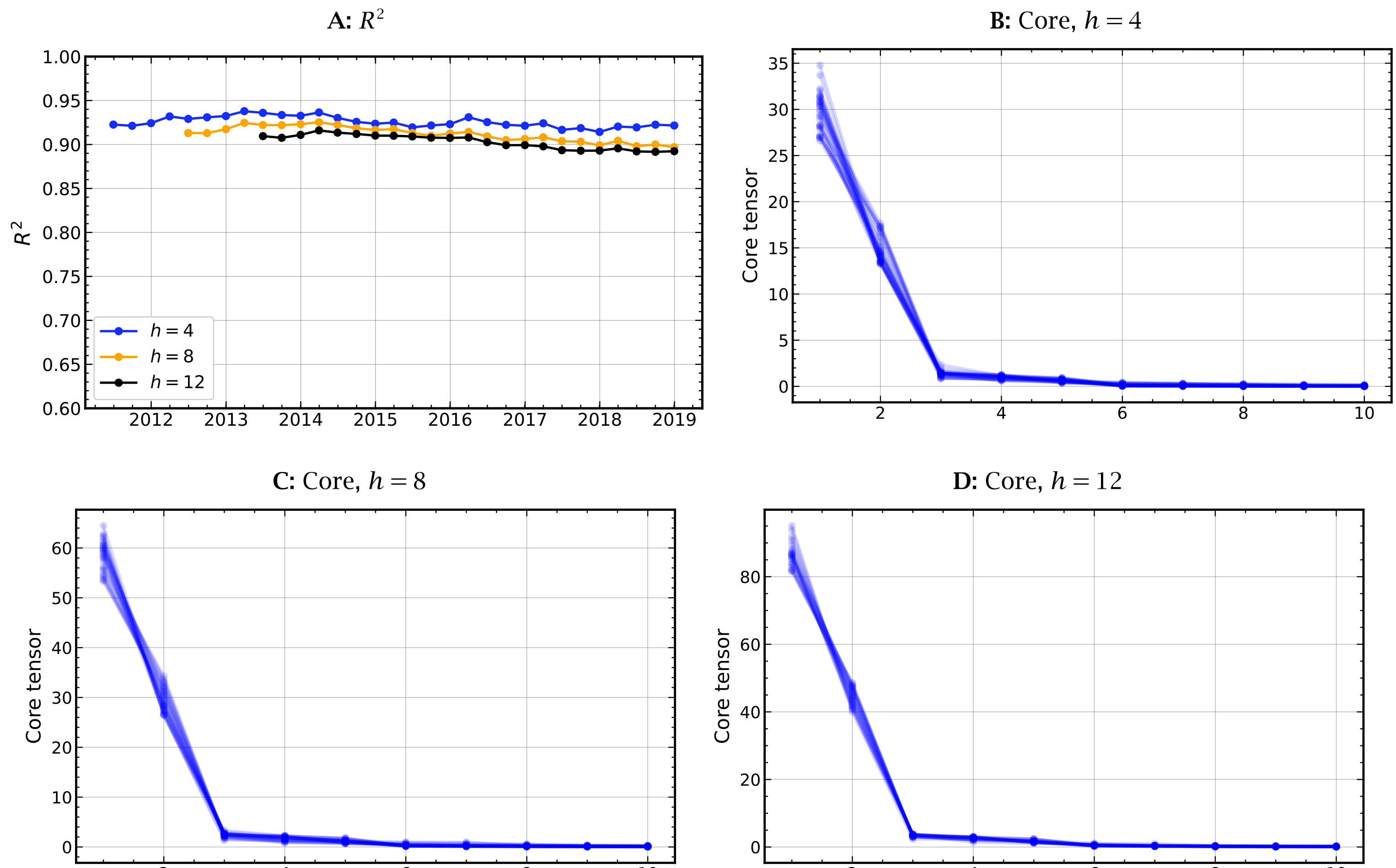
Notes: This figure shows the mean squared ME, BM, MOM, and REV errors. Panel A plots the MSE in $(K_T,15,8)$ Tucker models, where K_T ranges from 1 to 20. Panels B and C show results for $(5,K_N,8), K_N = 1,...,25$ and $(5,15,K_C), K_C = 1,...,15$, respectively. The sample period is 2010Q3 to 2018Q4.

Figure 8: Loading matrix $\mathbf{V}^C$



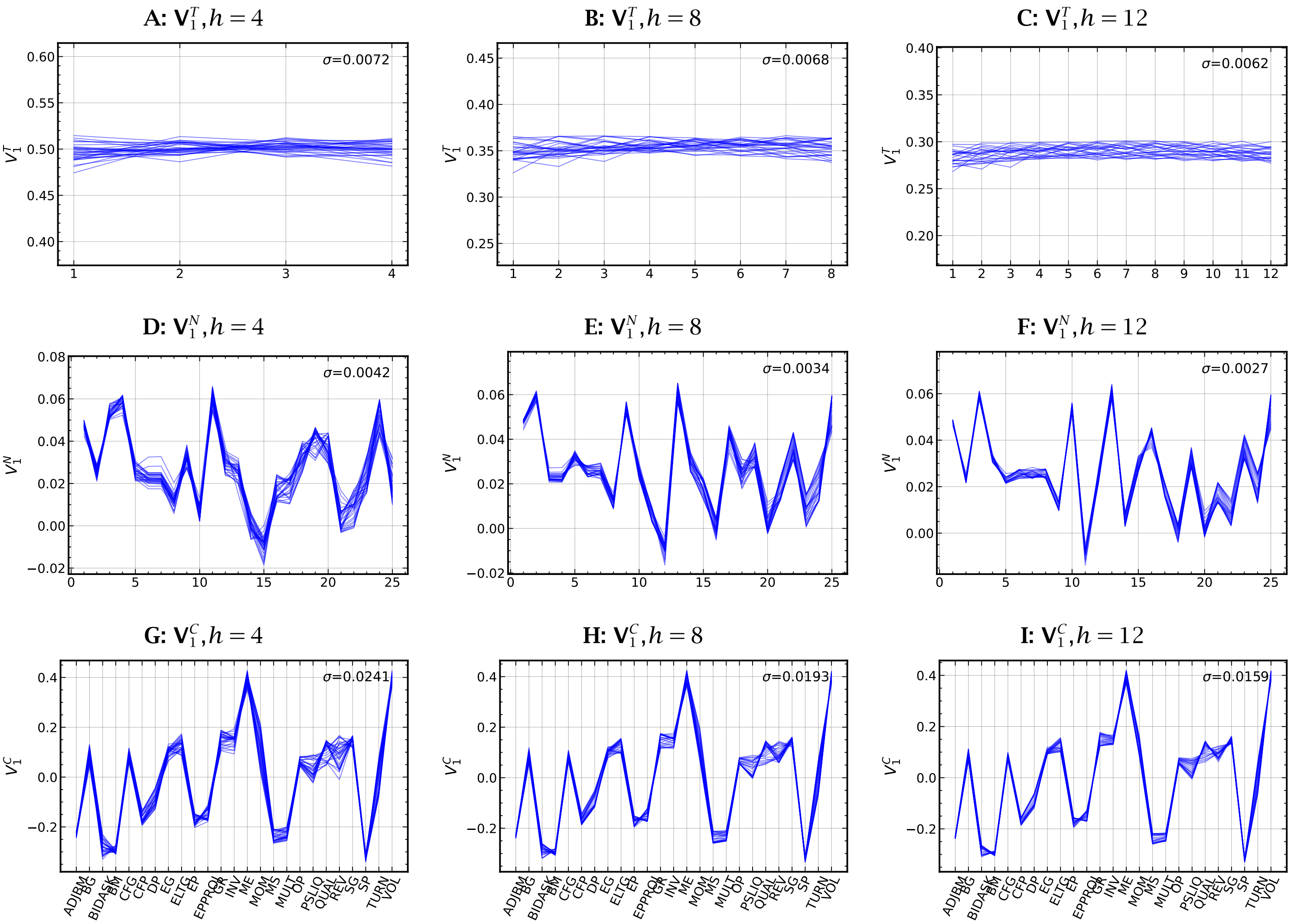
Notes: The figure shows a heatmap of the characteristic loading matrix $\mathbf{V}_C$ of the Tucker(2,6,5) model. Positive values are shown in blue, and negative values are in red. The sample period is 2010Q3 to 2018Q4.

Figure 9: Rolling subsamples



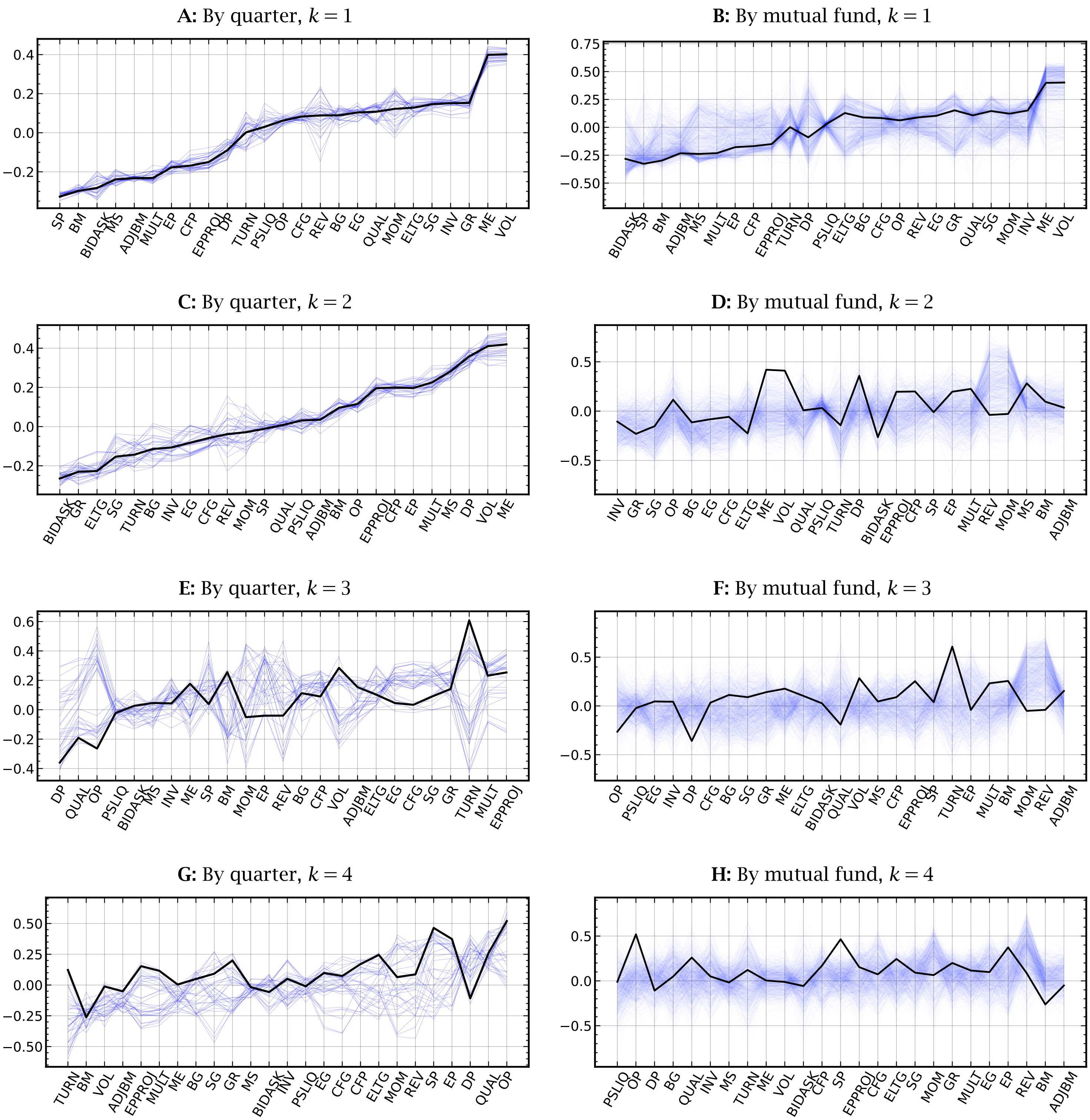
Notes: Panel A plots the R^2 of Tucker models estimated in rolling subsamples of length h. The Tucker models have $(K_T, K_N, K_C) = (2,6,5)$ components. Panel B to D plot the squares of the ten largest elements of the core tensor (divided by 10^3) for $h = 4,8,12$. The sample period is 2010Q3 to 2018Q4.

Figure 10: Rolling subsamples – $\mathbf{V}_1^i$



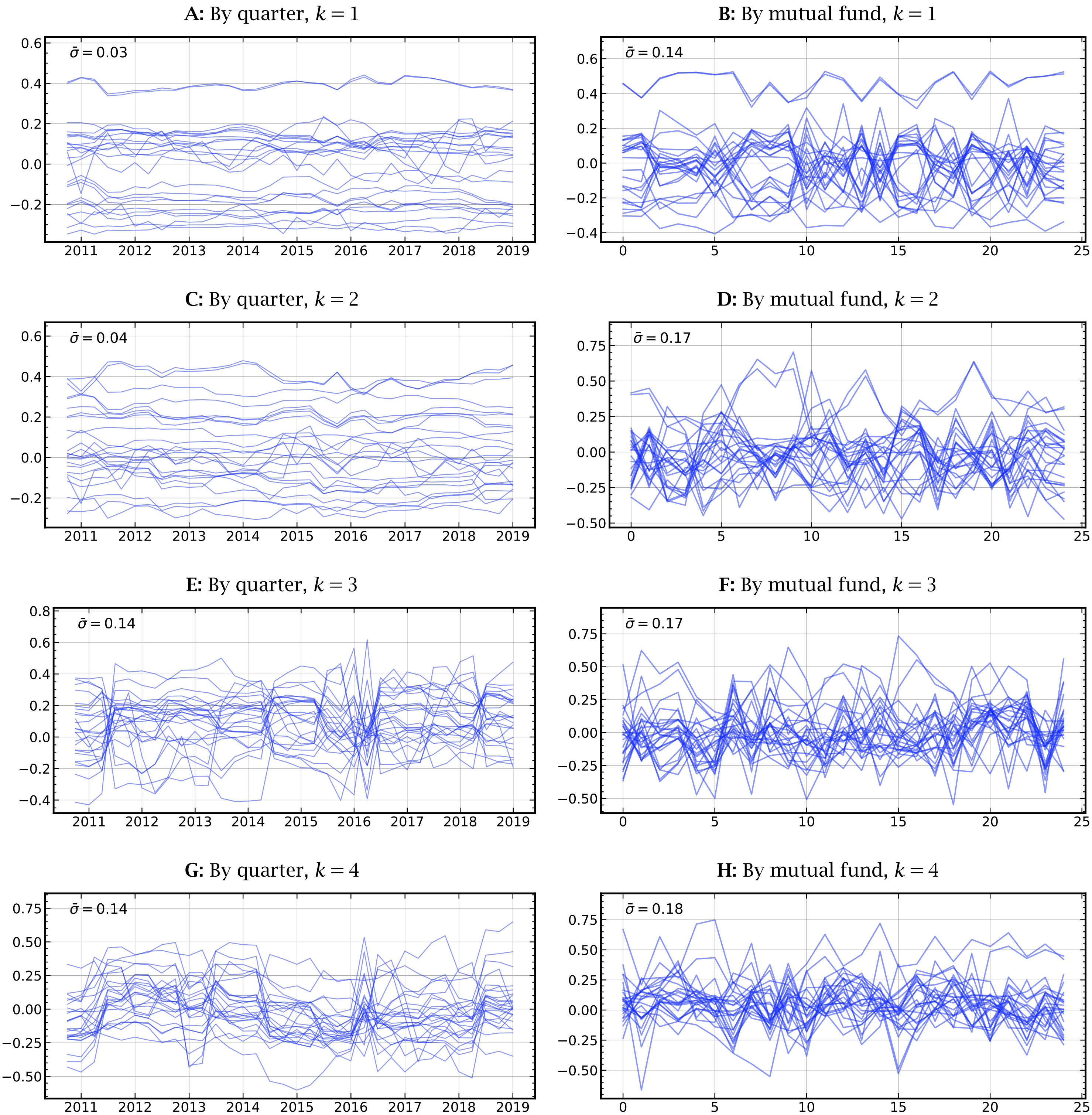
Notes: This figure plots the first loading vectors of mode $i, \mathbf{V}_1^i, i = T, N, C$, of Tucker models estimated in rolling subsamples of length $h = 4, 8, 12$. The Tucker models have $(K_T, K_N, K_C) = (2, 6, 5)$ components. Each line corresponds to a row of a loading vector estimated in a subsample. The $\mathbf{V}_i^N$ plots represent 25 randomly drawn mutual funds. Each plot shows the average time series standard deviation of the loading vectors across subsamples. The sample period is 2010Q3 to 2018Q4.

Figure 11: SVD/PCA characteristic factors



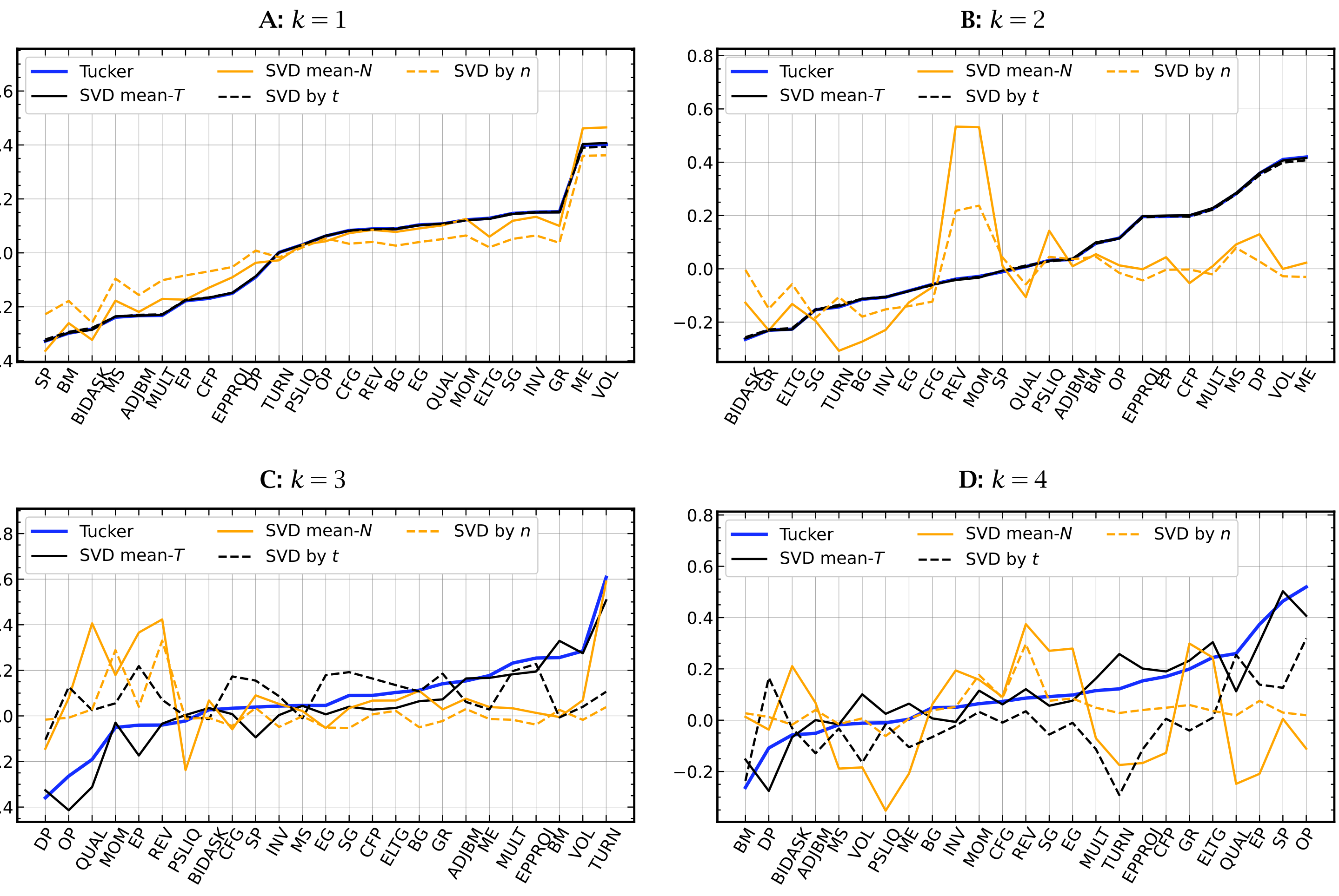
Notes: This figure shows the loadings of the first four characteristic factors estimated from 2-dimensional SVD/PCA for each quarter (left column) and each mutual fund (right column). Each line corresponds to factor loadings of a characteristic estimated for a quarter and mutual fund, respectively. The x-axes are ordered by mean loadings across all quarters and funds. The black lines show the characteristic loadings estimated from a Tucker (2, 6, 5) model. The sample period is 2010Q3 to 2018Q4.

Figure 12: SVD/PCA characteristic factors by quarter and mutual fund



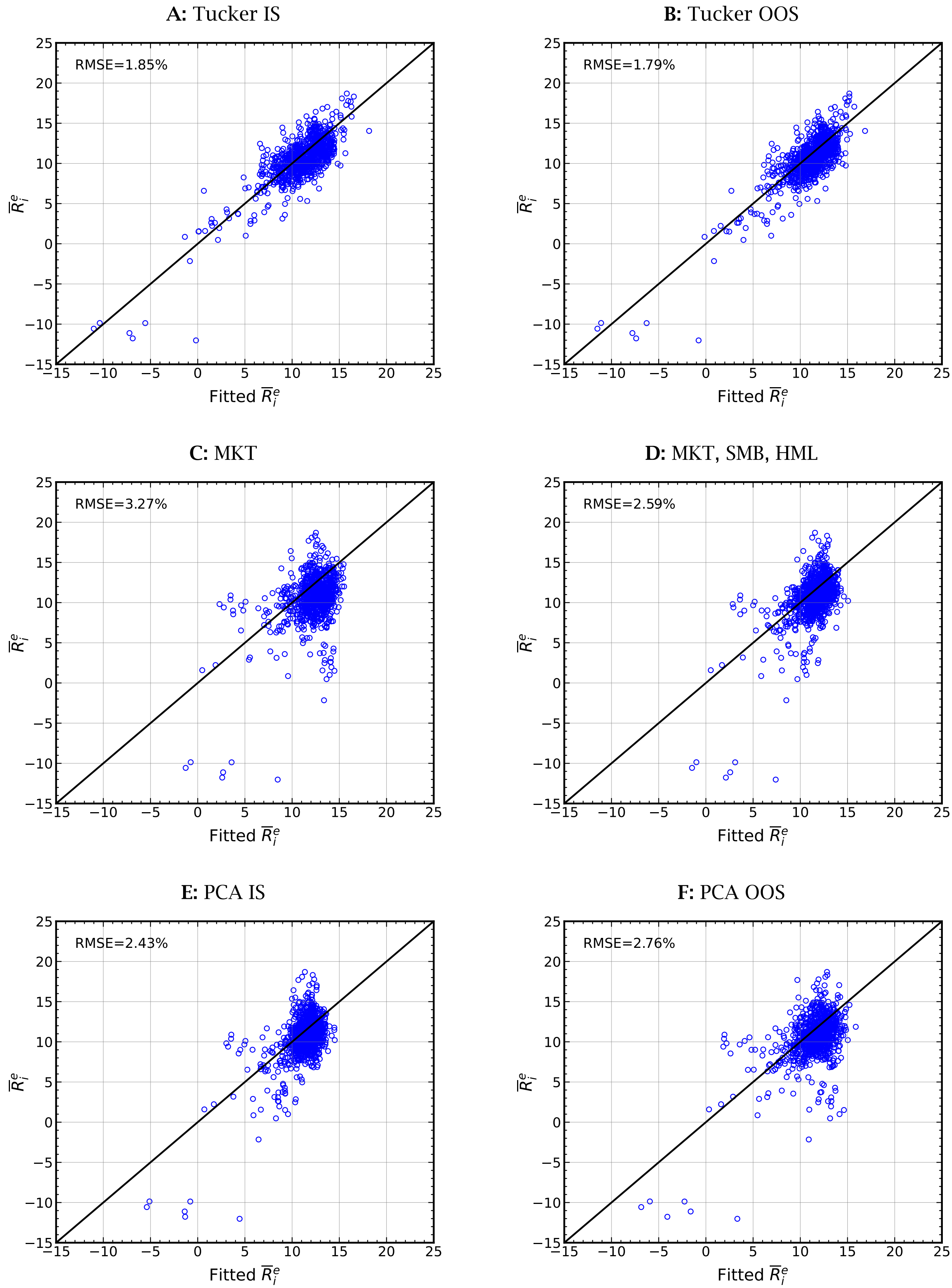
Notes: This figure shows the loadings of the first four characteristic factors estimated from 2-dimensional SVD/PCA for each quarter (left column) and each mutual fund (right column). Each line in the plots in the left column corresponds to loadings of a characteristic estimated across quarters. Each line in the plots in the right column corresponds to loadings of a characteristic estimated for the 25 largest mutual funds in the sample. The black lines show the characteristic loadings estimated from a Tucker (7, 13, 13) model. The sample period is 2010Q3 to 2018Q4.

Figure 13: SVD/PCA characteristic factors – Means



Notes: This figure compares SVD/PCA estimations of characteristic loadings of the first four factors to those obtained from a Tucker (2,6,5) model. The blue line shows the Tucker estimates. The solid black and orange lines show the results of SVD/PCA estimations for the means across quarters and mutual funds, respectively. The dashed lines plot the means of SVD/PCA estimations for each quarter and mutual fund. The sample period is 2010Q3 to 2018Q4.

Figure 14: Mean vs. fitted returns



Notes: The figure shows scatter plots of mean fund returns and fitted returns of in-sample and out-of-sample Tucker, the Fama-French, and PCA models with $L = 3$ factors. The sample period is 2010Q3 to 2018Q4.

Appendix A Tensor operations

Let $\boldsymbol{\mathcal{X}}$ be a $(T \times N \times C)$-dimensional tensor. A 3-dimensional tensor can 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 $\boldsymbol{\mathcal{X}}$ has fibers along each mode, denoted by $\mathbf{x}_{(nc)t}, \mathbf{x}_{(tc)n}$, and $\mathbf{x}_{(tn)c}$, respectively.[32] Slices are created by fixing all but two indices and are written as $\mathbf{X}_{(t)nc}, \mathbf{X}_{(n)tc}, \mathbf{X}_{(c)tn}$.[33]

A tensor can be written as a matrix by *unfolding* one dimension. For example, unfolding $\boldsymbol{\mathcal{X}}$ 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 $\boldsymbol{\mathcal{X}}$ 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:

$$\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \rangle = \sum_{t,n,c} \mathsf{x}_{tnc} \mathsf{y}_{tnc}$$

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

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

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

$$\boldsymbol{\mathcal{X}} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \in \mathbb{R}^T \times \mathbb{R}^N \times \mathbb{R}^C. \tag{A.1}$$

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 $\boldsymbol{\mathcal{X}}$ 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 $\boldsymbol{\mathcal{X}}$ and the $(S \times T)$ matrix $\mathbf{A}_1$ is equal to a $(S \times N \times C)$ tensor $\boldsymbol{\mathcal{Y}}$ given by

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

The n-mode product tensor is constructed by multiplying each mode-n fiber by each row vector of $\mathbf{A}_1$. In general, the n-mode is written as $\boldsymbol{\mathcal{X}} \times_n \mathbf{A}_n$. The number of columns of $\mathbf{A}_n$ must equal the n-mode dimension of $\boldsymbol{\mathcal{X}}$ while the n-mode dimension of $\boldsymbol{\mathcal{X}} \times_n \mathbf{A}_n$ is equal to the number of rows of $\mathbf{A}_n$. The n-mode product can be chained:

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

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 (5×2) matrix is illustrated in Panel A of Figure D.4.

---

[32] See Panels B, C, and D of Figure D.2.

[33] See Panels E, F, and G of Figure D.2.

[34] Panel A of Figure D.4 shows an example for $T = 5, N = 4, C = 3$.

Each mode-1 fiber of $\boldsymbol{\mathcal{X}}$ is a vector of length 2 and is multiplied by each of the row vectors of $\mathbf{A}_1$ so that $\boldsymbol{\mathcal{X}}$ with mode-1 dimension T is transformed into the product tensor $\boldsymbol{\mathcal{Y}}$ 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 (2×4) matrix but is displayed as a (4×2) matrix. It is standard practice to rotate tensors, matrices, and vectors in illustrations so that the mode dimensions match.[35]

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 C.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}.[36] The SVD of $\mathbf{X}$ is given by

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

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

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*

$$\mathbf{X} = \mathbf{F}_N\mathbf{B}_N^\top, \tag{B.3}$$

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 matrices such that

$$\mathbf{X} = \hat{\mathbf{X}}_K + \mathbf{E}_K, \tag{B.4}$$

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

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)

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

[35] There is no "transpose" operator for tensors, and it may be helpful to think about tensor multiplications without the notion of a matrix transpose.

[36] The row index t is generic and does not necessarily have to be a "time" index.

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 $\mathbf{U}^1, \mathbf{U}^2$:

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

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

$$\mathbf{X} = \mathbf{F}_K \mathbf{B}_K^\top + \mathbf{E}_K, \tag{B.7}$$

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

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

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

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:

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

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 $\boldsymbol{\mathcal{G}}$ in the Tucker decomposition is *not* diagonal. Consequently, the number of factors can differ by dimension.[37]

Appendix C Higher-Order Orthogonal Iteration (HOOI)

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

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

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

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

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

Since the $\mathbf{V}^i$ matrices are orthonormal, the squared norm of the approximation error $\boldsymbol{\mathcal{E}} = \boldsymbol{\mathcal{X}} - \widehat{\boldsymbol{\mathcal{X}}}$ can written as

$$\begin{aligned}
\|\boldsymbol{\mathcal{E}}\|^2 &= \|\boldsymbol{\mathcal{X}}\|^2 - 2\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{G}} \times_1 \mathbf{V}^T \times_2 \mathbf{V}^N \times_3 \mathbf{V}^C \rangle + \|\boldsymbol{\mathcal{G}} \times_1 \mathbf{V}^T \times_2 \mathbf{V}^N \times_3 \mathbf{V}^C\|^2 && \text{(C.2)}\\
&= \|\boldsymbol{\mathcal{X}}\|^2 - 2\langle \boldsymbol{\mathcal{X}} \times_1 \mathbf{V}^{T\top} \times_2 \mathbf{V}^{(M)\top} \times_3 \mathbf{V}^{C\top}, \boldsymbol{\mathcal{G}} \rangle + \|\boldsymbol{\mathcal{G}}\|^2 && \text{(C.3)}\\
&= \|\boldsymbol{\mathcal{X}}\|^2 - 2\langle \boldsymbol{\mathcal{G}}, \boldsymbol{\mathcal{G}} \rangle + \|\boldsymbol{\mathcal{G}}\|^2 && \text{(C.4)}\\
&= \|\boldsymbol{\mathcal{X}}\|^2 - \|\boldsymbol{\mathcal{G}}\|^2 && \text{(C.5)}\\
&= \|\boldsymbol{\mathcal{X}}\|^2 - \|\boldsymbol{\mathcal{X}} \times_1 \mathbf{V}^{T\top} \times_2 \mathbf{V}^{(M)\top} \times_3 \mathbf{V}^{C\top}\|^2. && \text{(C.6)}
\end{aligned}$$

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

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

This maximization problem can be rewritten in matrix form as

$$\max_{\mathbf{V}^C} \|\mathbf{V}^{C\top}\mathbf{W}_C\| \tag{C.8}$$

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

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

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 $\boldsymbol{\mathcal{X}}$ tensor. For example, unfold $\boldsymbol{\mathcal{X}}$ 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.

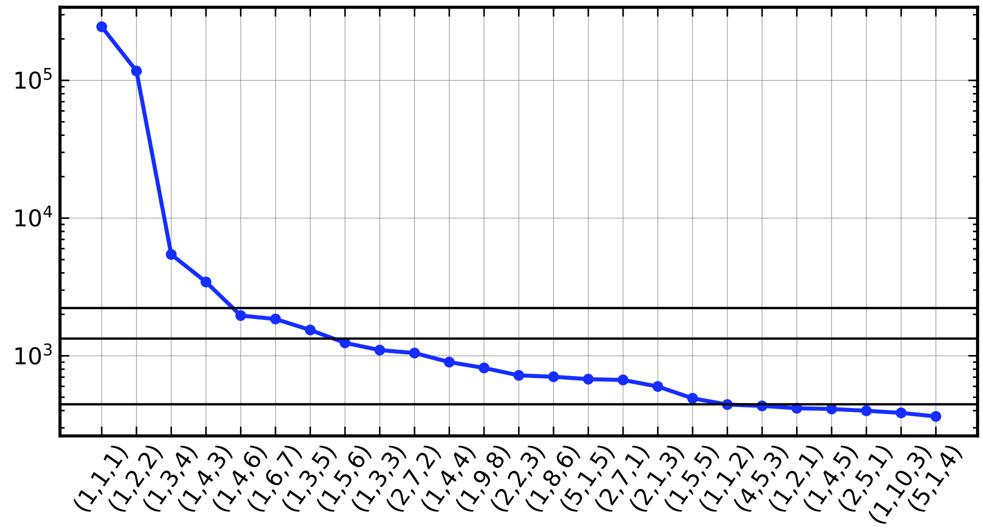
The HOOI estimator converges after 5 to 20 iterations for the data set used in this paper. In addition to setting the initial $\mathbf{V}^i$ using the method described above, I also choose initial values randomly. The numerical computations are robust and converge to the same optimum.

Table C.1: Summary of tensor notation and operations

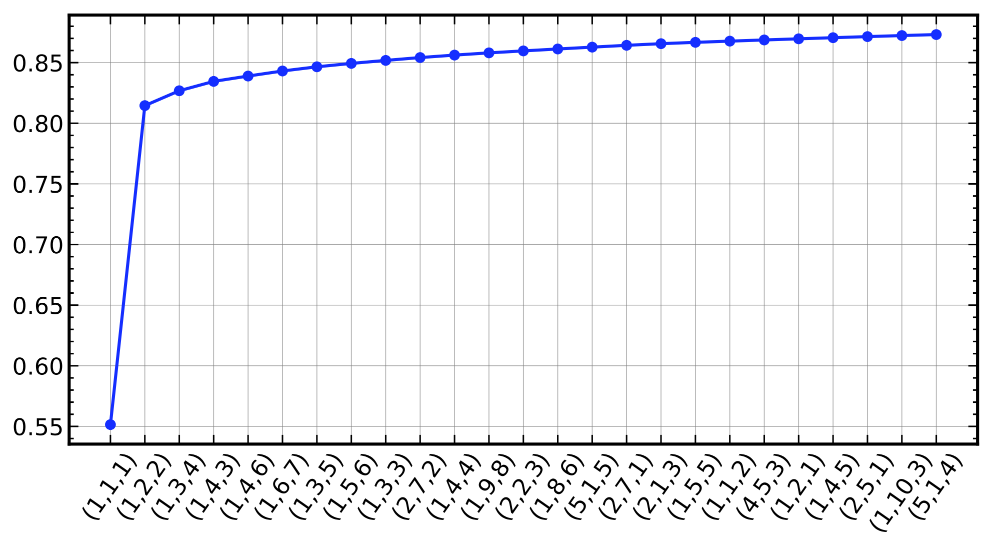
Operation	2-dimensional matrix	3-dimensional tensor	n-dimensional tensor
	$\mathbf{X} = [\mathsf{x}_{ij}]$	$\boldsymbol{\mathcal{X}} = [\mathsf{x}_{ijk}]$	$\boldsymbol{\mathcal{X}} = [\mathsf{x}_{i_1,\ldots,i_j}]$
Fibers	$\mathbf{x}_{(i)j}, \mathbf{x}_{(j)i}$	$\mathbf{x}_{(jk)i}, \mathbf{x}_{(ik)j},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

Figure C.1: Squared core elements

A: Largest core elements

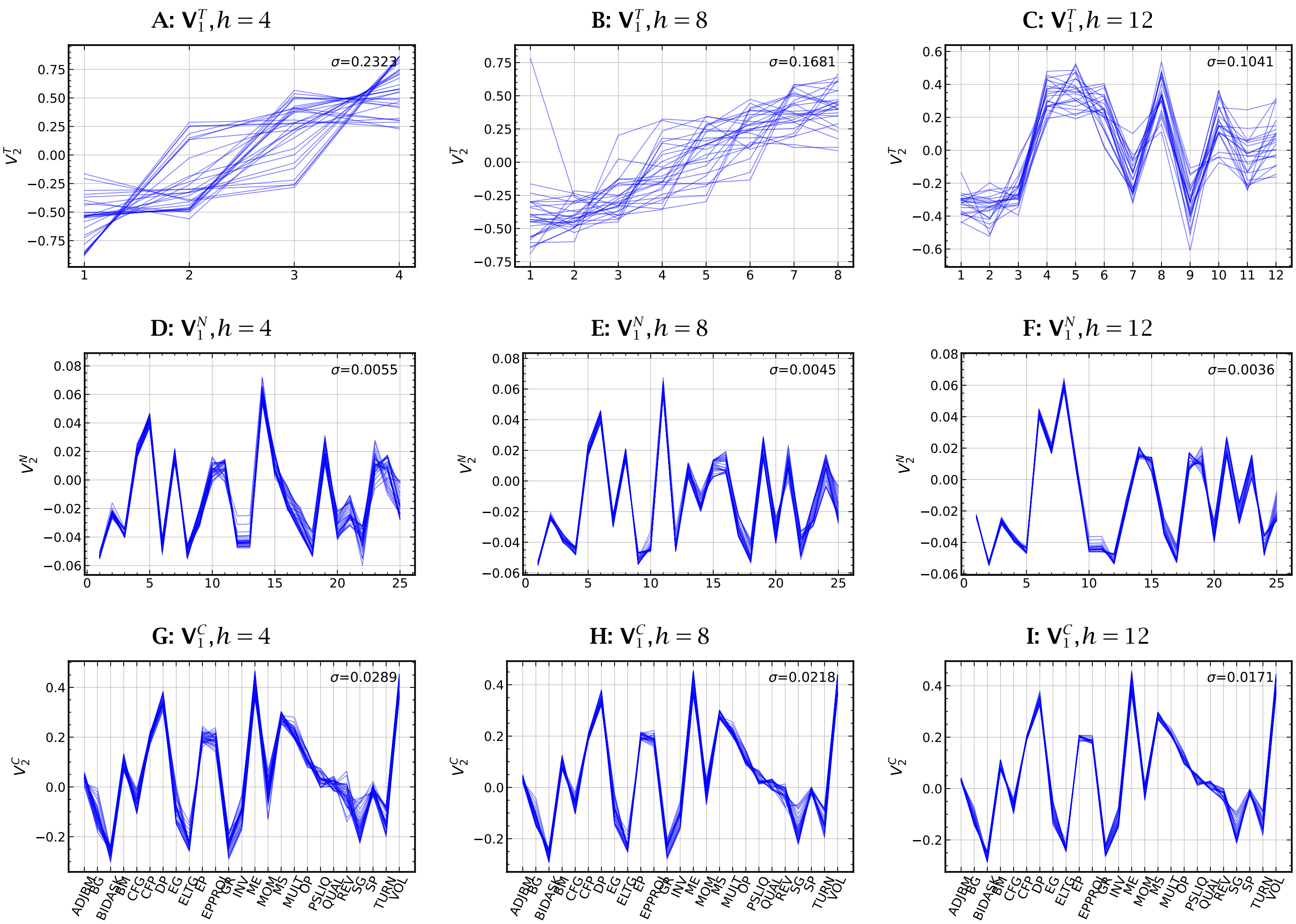


B: R^2



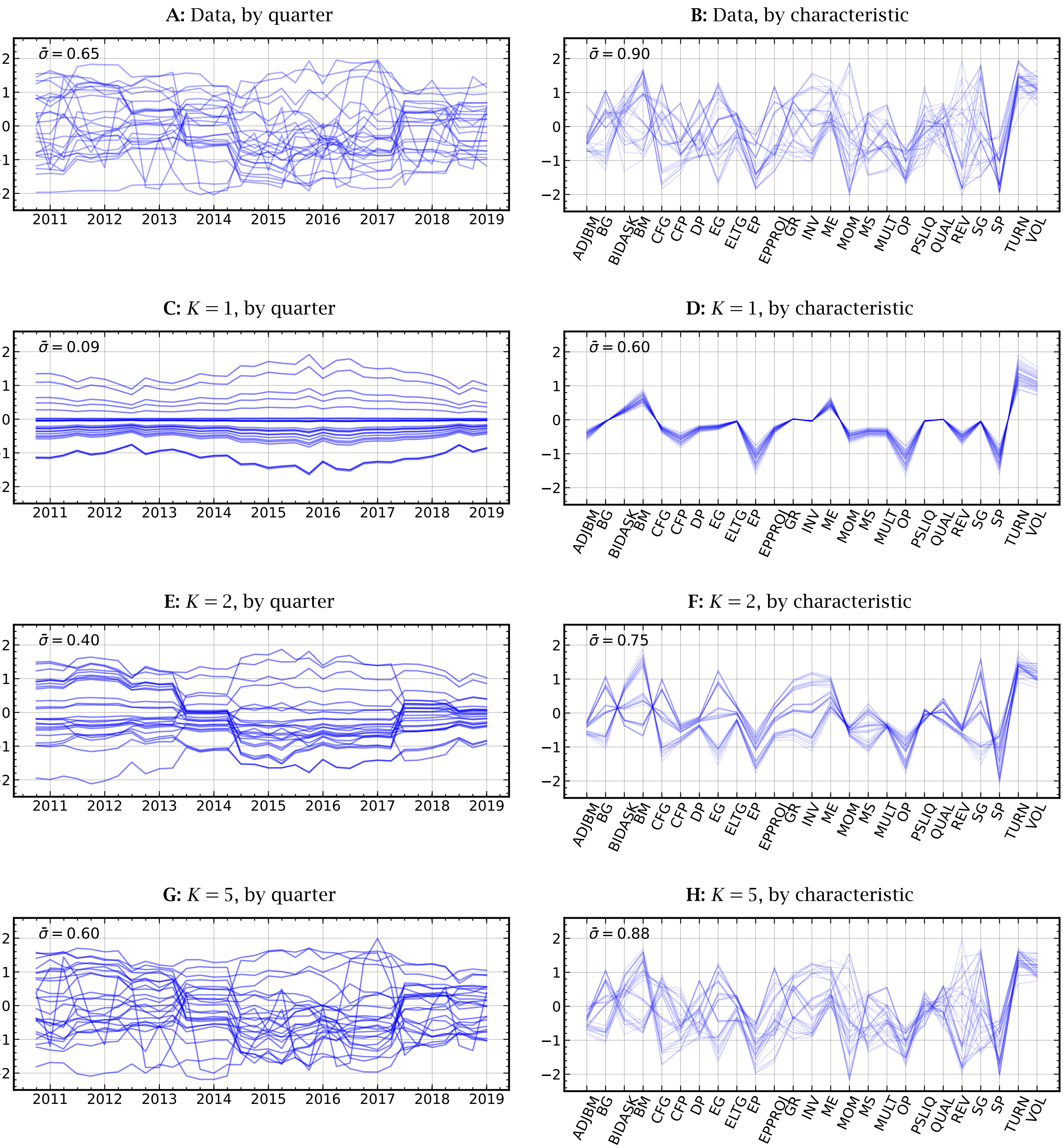
Notes: Panel A plots the squares of the largest 25 elements of the core tensor $\mathcal{G}$ of the full Tucker(34,934,25) model. The horizontal lines correspond to the contributions to the overall variance of 0.1%, 0.3% and 0.5%. Panel B plots the corresponding R^2. The sample period is 2010Q3 to 2018Q4.

Figure C.2: Rolling subsamples – $\mathbf{V}_2^i$



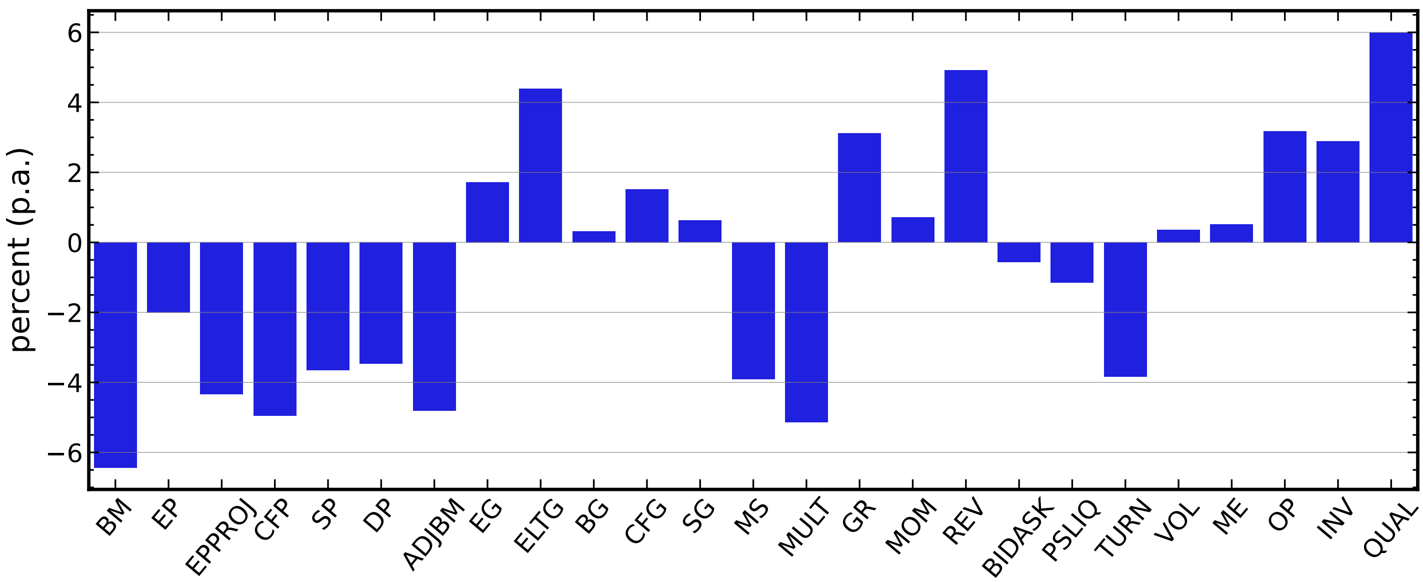
Notes: This figure plots the second loading vectors of mode i,$\mathbf{V}_2^i, i = T,N,C$, of Tucker models estimated in rolling subsamples of length $h =$ 5. The Tucker models have $(K_T,K_N,K_C) = (2,6,5)$ components. Each line corresponds to a row of a loading vector estimated in a subsample. The $\mathbf{V}_2^N$ plots represent 25 randomly drawn mutual funds. Each plot shows the average time series standard deviation of the loading vectors across subsamples. The sample period is 2010Q3 to 2018Q4.

Figure C.3: SVD/PCA estimation for a single mutual fund



Notes: This figure presents results from the SVD/PCA estimation of characteristics of a single mutual fund (wficn=102893). Panels A and B plots the observed by time and characteristics, respectively. The other panels plot the fitted values of SVD/PCA estimations with $K = 1,2,5$ factors. The left panels show the time series, while the right panels plot the fitted values by characteristics. The sample period is 2010Q3 to 2018Q4.

Figure C.4: Means of characteristic factors



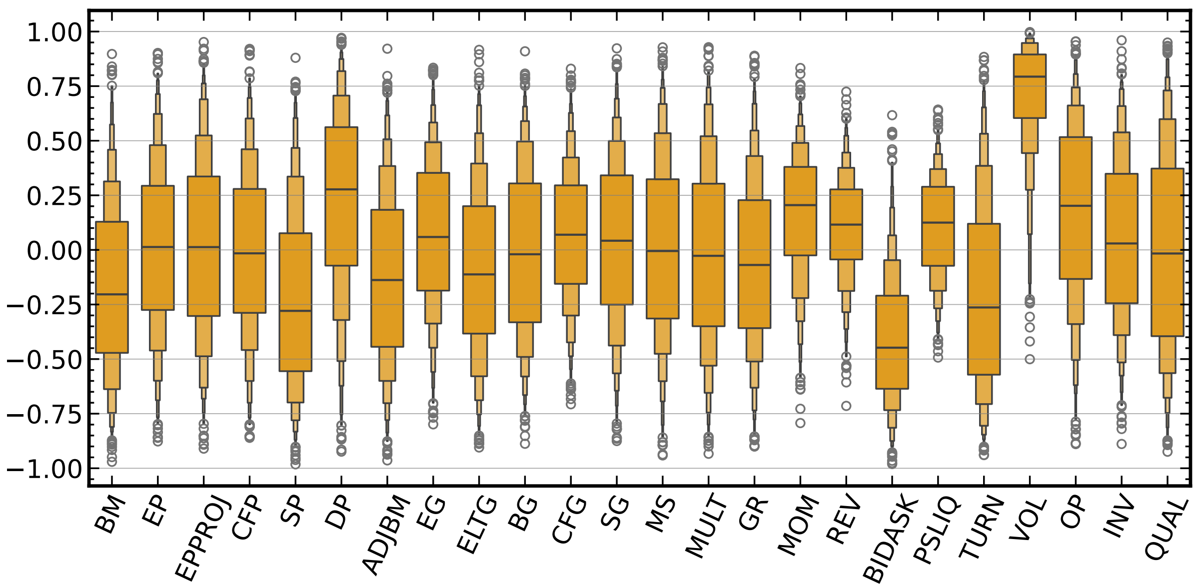
Notes: The figure plots the annualized means of characteristic factors. The sample period is 2010Q3 to 2018Q4.

Internet Appendix

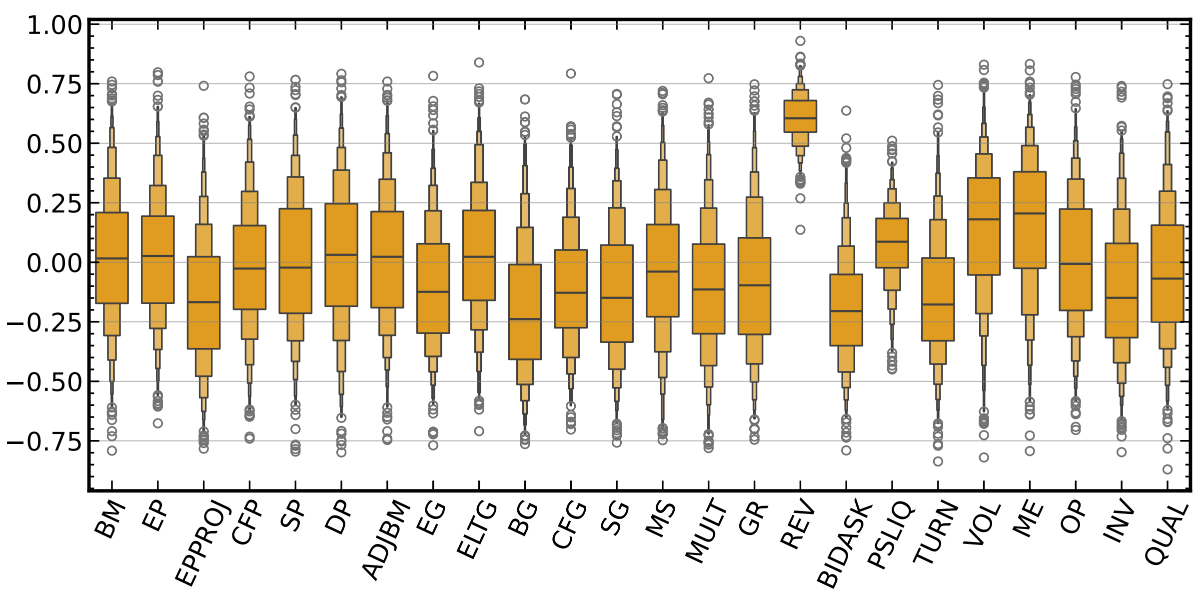
“High-Dimensional Factor Models and the Factor Zoo”

Figure D.1: Cross-sectional correlations

A: Time-series correlations across mutual funds – ME



B: Time-series correlations across mutual funds – MOM



Notes: The figure shows box plots of the distributions of pairwise characteristic correlations of ME (Panel A) and MOM (Panel B). The sample period is 2010Q3 to 2018Q4.

Figure D.2: Tensor fibers and slices

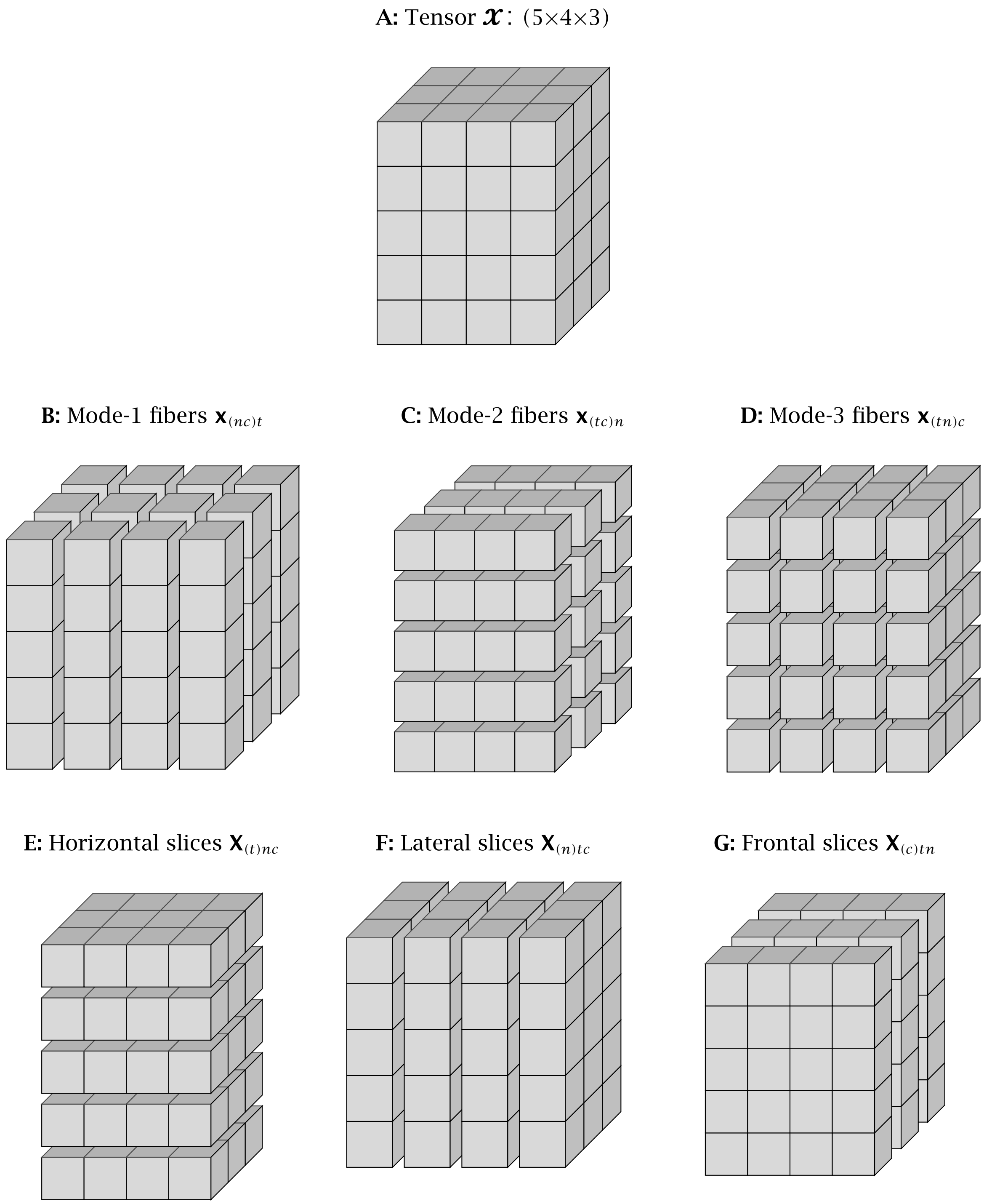
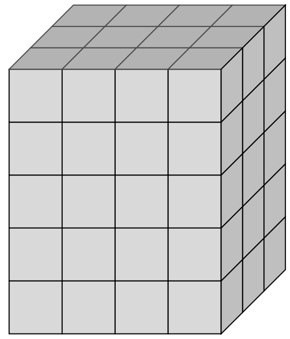
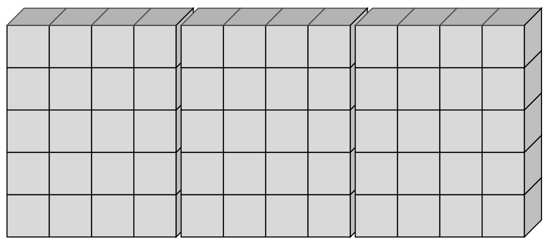


Figure D.3: Tensor as matrices

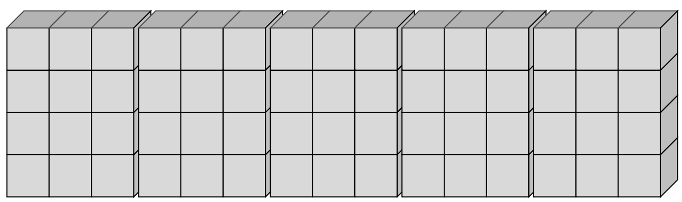
A: Tensor $\boldsymbol{\mathcal{X}}$: $(5\times4\times3)$



B: $\mathbf{X}_{(1)}$: (5×12)



C: $\mathbf{X}_{(N)}$: (4×15)



D: $\mathbf{X}_{(C)}$: (3×20)

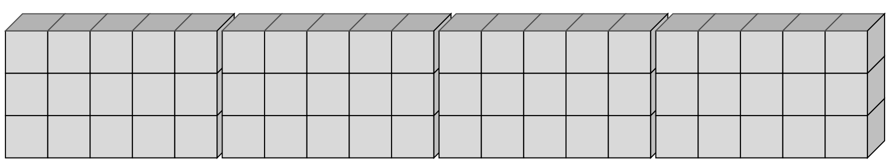


Figure D.4: Tensor multiplication