High-Dimensional Factor Models with an Application to Mutual Fund Characteristics

Martin Lettau[1]

Haas School of Business, University of California at Berkeley, Berkeley, CA 94720,
lettau@berkeley.edu, NBER, CEPR

INCOMPLETE DRAFT - DO NOT CIRCULATE

Abstract

This paper considers extensions of two-dimensional factor models to higher-dimensional data represented as tensors. I describe decompositions of tensors that generalize the standard matrix singular value decomposition and principal component analysis to higher dimensions. I estimate the model using a three-dimensional data set consisting of 25 characteristics of 1,342 mutual funds observed over 34 quarters. The tensor factor models reduce the data dimensionality by 97% while capturing 94% of the variation of the data. I relate higher-dimensional tensor models to standard two-dimensional models and show that the components of the model have clear economic interpretations. Pricing factors implied by the Tucker model outperform Fama-French factors in pricing the cross-section of mutual funds returns.

Keywords: Tucker decomposition, tensors, PCA, SVD, factor models, mutual funds, characteristics

This draft: December 28, 2022

[1]I thank Ian Dew-Becker, Christian Julliard (discussant), Dmitry Livdan, Peter Maxted, Stefan Nagel, Paulo Zaffaroni, Guofu Zhou, and seminar participants at Berkeley-Haas, Frankfurt, the London School of Economics, and the 2022 BI-SHoF Conference for helpful comments.

1. Introduction

Statistical factor models have a long tradition in finance and economics. They can be used to identify latent factors and test asset pricing implications of the arbitrage pricing theory (APT) developed by Ross (1976) and Chamberlain and Rothschild (1983) or to reduce the dimensions of large data sets to a few factors that efficiently summarize the information in the data. Tests of the APT go back to Roll and Ross (1980), Schipper and Thompson (1981), Connor and Korajczyk (1986), and Connor and Korajczyk (1988), followed by a more recent literature that has developed extensions to the standard model (e.g. Kelly, Pruitt and Su (2019), Pelger (2019), Lettau and Pelger (2020a), Lettau and Pelger (2020b), Giglio and Xiu (2021)). Some applications in macroeconomics include business-cycle forecasting (Stock, Wright and Yogo (2002), Stock and Watson (2006)), large macroeconomic modeling (Gagliardini and Gourieroux (2014), Favero, Marcellino and Neglia (2005), Forni, Hallin, Lippi and Reichlin (2000)), and monetary policy (Boivin and Ng (2006)).

Traditional factor models can be estimated in two-dimensional panel data sets, *i.e.*, the data has the form of a matrix. In this paper, I consider models that apply to higher-dimensional data. The sample used in the empirical section is three-dimensional, where the unit of observation x_{tmc} is the characteristic c of mutual fund m at time t. Since standard factor models cannot be applied to three-dimensional data, their estimation requires an ad hoc method to eliminate one of the dimensions. Consider, for example, Balasubramaniam, Campbell, Ramadorai and Ranish (2021), who analyze stock ownership in India. While their sample, consisting of stock holdings of individual investors over ten years, is three dimensional, they estimate a cross-sectional two-dimensional factor model for a single period. Hence, any time-series information is lost and not used in their cross-sectional factor model. In contrast, the techniques considered in this paper can be directly applied to higher-dimensional data without the need to collapse the data into two dimensions.

Formally, higher-order data form *tensors*, which extend the notions of vectors and matrices to higher dimensions. For example, a data set with observations x_{tmc} forms a three-dimensional tensor. Tensors were introduced by Ricci-Curbastro and Levi-Civita (1900) and have many applications in physics and engineering. Intuitively, tensor factor models are generalizations of singular-value matrix decompositions (SVD) and principal component analysis (PCA). The SVD decomposes a matrix $\mathbf{X}$ into the product of three matrices, which are formed by the eigenvectors of $\mathbf{X}^\intercal\mathbf{X}$ and $\mathbf{X}\mathbf{X}^\intercal$, $\mathbf{U}_1$ and $\mathbf{U}_2$, respectively, and the diagonal matrix $\mathbf{H}$ of associated eigenvalues: $\mathbf{X} = \mathbf{U}_1\mathbf{H}\mathbf{U}_2^\intercal$. A K-factor PCA model is equivalent to a *truncated* SVD with eigenvectors that are associated with the K largest eigenvalues.

The methods considered in this paper are based on tensor decompositions that share some, but not all, properties of the matrix SVD. The details are in Section 2, which also includes a summary of tensor algebra. The form of the tensor decomposition is similar to that of SVDs and PCA. A tensor with n dimensions can be expressed as the (tensor) product of n matrices and a "small" n-dimensional *core* tensor. Each of the n matrices corresponds to a dimension of the tensor and is similar to the $\mathbf{U}_1$ and $\mathbf{U}_2$ eigenvector matrices in the SVD and "loadings" matrices in two-dimensional factor models. The n-dimensional core tensor plays a similar role as the diagonal matrix of eigenvalues $\mathbf{H}$ in the SVD. However, since there is no equivalent notion of eigenvalues and eigenvectors

for higher-order tensors, the components of the tensor decomposition are computed differently. The decomposition of a three-dimensional $(T \times M \times C)$ tensor implies a collection of $T+M+C$ two-dimensional factor models. The factor models are interconnected by a set of linear restrictions that are functions of the elements of the core tensor and loading matrices.

Since the interpretation of factors and factor loadings is often difficult, even in two-dimensional data, I pay particular attention to the economic meaning of the components of the tensor decomposition. First, I show that the decomposition of an n-dimensional tensor implies n two-dimensional factor models, one for each of the n dimensions. For example, the three-dimensional decomposition of the date/fund/characteristic data set x_{tmc} implies a two-dimensional factor model for dates t, another two-dimensional factor model for funds m, and a third two-dimensional model for characteristics c. Each of these models has the same form and interpretation as a two-dimensional factor model. Second, I show that the elements of the core tensor in the high-dimensional decomposition can be interpreted as observations of "summary" date/fund/characteristic objects. In other words, the core tensor compresses the data set with $(T \times M \times C)$ dimensions into a low-dimensional tensor with K_1 "summary" time, K_2 "summary" fund, and K_3 "summary" characteristic observations. This intuition allows a straightforward interpretation of the elements of the tensor decomposition even though the data is high dimensional.

I implement the tensor decompositions using a data set consisting of 25 characteristics of 1,342 mutual funds observed over a sample of 34 quarters totaling 1,140,700 observations. Traditional two-dimensional factor models can only be applied to panels with a single characteristic, for example, time-series of returns across assets. Tensor methods allow the estimation of factor models of many characteristics observed for many assets over time while exploiting possible interactions in all three dimensions. Characteristics are correlated across assets; for example, the book-to-market ratios of stocks and mutual funds 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. three-dimensional tensor models can capture all three two-dimensional correlation patterns, including their interactions, in an efficient and internally-consistent representation.

I compare the fit of decompositions with a wide variety of factors and settle on a benchmark model with $(K_1, K_2, K_3) = (10, 25, 9)$ factors. This model captures 94% of the variation in the data and reduces the dimensionality of the data by 97%, which is equivalent to a two-dimensional model with three factors for a panel of size (200, 100). In other words, the dimension of the data tensor is reduced from (34, 1697, 25) to a model with a (10, 25, 9) dimensional core that can be interpreted as 25 summary mutual funds with nine summary characteristics observed over ten summary quarters.

I show that the model yields a good fit for all data points with the exception of some outliers. The model is stable over the sample, across the vast majority of mutual funds, as well as across characteristics. I find that many aspects of the tensor model are similar to patterns found in many two-dimensional factor models. The first factors along the three dimensions are "level" factors with positive "long-only" loadings, while higher-order components are "long-short" factors. Similarly, lower-order "summary" time/fund/characteristics elements in the core tensor are related to means,

and higher-order elements represent deviations from means.

Finally, I construct pricing factors that are based on the Tucker model and find that the Tucker factors outperform Fama-French factors in pricing the cross-section of mutual fund returns. Typically, PCA estimations of latent factor models are based on panels of a cross-section of asset returns observed over time. In contrast, the pricing factors implied by the Tucker model use *no* information about returns and are constructed only from the 3-dimensional data set of asset characteristics.

This paper is related to Bryzgalova, Lettau, Lerner and Pelger (2022) (BLLP), who propose an estimation methodology for two-dimensional cross-sectional panels that are observed over time. Their procedure combines two-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 two-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 is designed to impute missing data.

Since the increased availability of "big" data sets, there has been much research about efficient algorithms to reduce data dimensionality. Because of their flexibility and efficiency, methods based on tensor decompositions have become popular for modeling high-dimensional data in many areas of science. Examples include brain imaging (Möcks (1988)), fMRI processing (Andersen and Rayens (2004)), facial recognition (Vasilescu and Terzopoulos (2002)), signal processing (De Lathauwer, De Moor and Vandewalle (2000)), machine learning (Bacciu and Mandic (2020)), and MPEG watermarking (Abdallah, Ben Hamza and Bhattacharya (2007)), among many others (see Kolda and Bader (2009) and Sidiropoulos, De Lathauwer, Fu, Huang, Papalexakis and Faloutsos (2017) for recent overviews.).

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 three-dimensional. The estimation of dynamic corporate finance models often involves data sets with three or more dimensions, see Strebulaev and Whited (2012) for a survey. The investor-level data used in the household finance literature that studies portfolio holdings have more than two dimensions, see for example Odean (1998), Campbell (2006), Calvet, Campbell and Sodini (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. Based on the results in this paper, tensor decompositions are promising additions to the toolbox of economists for modeling higher-dimensional data.

The rest of the paper is organized as follows. Section 2 introduces tensors and summarizes tensor operations that are used in the paper, followed by a description of tensor decompositions. The empirical implementation is described in Section 3 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 4 shows that pricing factors derived from the Tucker model capture the cross-section of mutual fund returns. Section 5 concludes.

2. High-dimensional data

Traditional factor models used in finance and economics are based on two-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 two-dimensional data. In this section, I consider generalizations of factor models to situations when the data set has more than two dimensions. There are many potential applications of higher-dimensional models. The data set used in the empirical section below is three-dimensional and composed of observations of characteristic c of mutual fund m in period t, and I use this example to illustrate the theoretical results in this section.

Higher-dimensional data form *tensors*, which were first defined by Ricci-Curbastro and Levi-Civita (1900), 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 two-dimensional factor models, PCA, and SVD to facilitate a better understanding of the extensions to higher dimensions.

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

2.1. Two-dimensional data: SVD, PCA, and factor models

This section summarizes the SVD, factor models, and PCA for two-dimensional data represented in matrix form. Let $\mathbf{X}$ be a $(T\times N)$ data matrix with TN observations x_{tn}.[2] The SVD of $\mathbf{X}$ is given by

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

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

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

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

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*

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

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*. Factor models (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 $(N\times T)$ 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 the TN elements of $\mathbf{X}$ by a matrix $\hat{\mathbf{X}}_K$ that can be written in terms of lower-dimensional matrices such that

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

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

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

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

where $\|\mathbf{E}\| = \sqrt{\sum_{t,n} e_{tn}^2}$ is the Frobenius matrix norm. Eckart and Young (1936) showed that the solution is given by the *truncated SVD*, *i.e.* setting $\mathbf{H}_K$ to the first K rows and columns of $\mathbf{H}$ and $\mathbf{U}_K^{(1)}, \mathbf{U}_K^{(2)}$ to the first K columns of $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}$. The truncated SVD (4) is equivalent to the K-factor model

$$\mathbf{X} = \mathbf{F}_K\mathbf{B}_K^\top + \mathbf{E}_K, \tag{6}$$

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 in the remainder of the paper.

The truncated SVD (5) has an alternative representation that will be useful in understanding the tensor decompositions described below. $\mathbf{U}_K^{(1)}\mathbf{H}_K\mathbf{U}_K^{(2)\top}$ is 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 (5) as

$$\hat{\mathbf{X}}_K = \mathbf{F}_K\mathbf{B}_K^\top = \sum_{t=1}^{K}\sum_{n=1}^{K} h_{tn} \underbrace{\mathbf{u}_t^{(1)}\mathbf{u}_n^{(2)\top}}_{T\times N} = \sum_{k=1}^{K} h_{kk}\,\mathbf{u}_k^{(1)}\mathbf{u}_k^{(2)\top}. \tag{7}$$

The last equation follows from the fact that $\mathbf{H}$ 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 represents a factor in the K-factor representation $\mathbf{F}_K\mathbf{B}_K^\top$. The advantage of representation (7) 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.

2.2. From matrices to tensors

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

Throughout the paper, I use the following notation:

$$\text{scalar: } \mathsf{x} \in \mathbb{R}$$
$$\text{vector: } \mathbf{x} \in \mathbb{R}^I$$
$$\text{matrix: } \mathbf{X} \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2}$$
$$n\text{-th order tensor: } \boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \ldots \times \mathbb{R}^{I_n}$$

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 n dimensions of a tensor is called a *mode*. A tensor $\boldsymbol{\mathcal{X}}$ is *diagonal* if $\mathsf{x}_{i_1,\ldots,i_n} \neq 0$ only if $i_1 = \ldots = i_n$ and 0 otherwise.

The data set that will be used later has three dimensions: the characteristic c of mutual fund m at date t, x_{tmc}. To simplify the notation, I will therefore focus on tensors of order 3 but all results can be easily generalized to higher dimensions. Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^T \times \mathbb{R}^M \times \mathbb{R}^C$ be a three-dimensional $(T \times M \times C)$ tensor $\boldsymbol{\mathcal{X}}$ with *elements* x_{tmc}.[3]

A three-dimensional tensor can be represented by one-dimensional *fibers* and two-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 $\mathbf{x}_{t(mc)}, \mathbf{x}_{m(tc)}$, and $\mathbf{x}_{c(tm)}$, respectively.[4] Slices are created by fixing all but two indices and are written as $\mathbf{X}_{mc(t)}, \mathbf{X}_{tc(m)}, \mathbf{X}_{tm(c)}$.[5] Tensors can be written as matrices by *unfolding*, or *stacking*, two-dimensional slices along a mode n. The resulting matrix $\mathbf{X}_{(n)}$ is defined so that the number of rows equals the mode-n order of $\boldsymbol{\mathcal{X}}$. The number of columns of $\mathbf{X}_{(n)}$ is equal to the product of the dimensions along all other modes. [6]

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

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

[3] Figures A-1 to A-3 show a third-order tensor with dimensions $T = 5, M = 4, C = 3$ and illustrate the tensor operations described in this section.

[4] See Panels B, C, and D of Figure A-1.

[5] See Panels E, F, and G of Figure A-1.

[6] Figure A-2 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.

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 $\mathbf{b} \in \mathbb{R}^M$ is defined as[7]

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

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

$$\boldsymbol{\mathcal{X}} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \in \mathbb{R}^T \times \mathbb{R}^M \times \mathbb{R}^C. \tag{8}$$

An n-th order tensor is *rank*-1 if it can be written as an outer product of n vectors. More generally, a tensor is *rank-r* if it can be written as a sum of r rank-1 tensors. See Kolda and Bader (2009) for a more comprehensive discussion of tensor ranks.

Finally, 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 specified. This is not necessary for matrix multiplications since matrices have only two dimensions. 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 $(T \times M \times C)$ tensor $\boldsymbol{\mathcal{X}}$ and the $(S \times T)$ matrix $\mathbf{A}_1$ is equal to a $(S \times M \times C)$ 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 of a third-order tensor and a vector yields a matrix. For example, if $\mathbf{a}_1$ is a $(T \times 1)$ vector, then

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

is a $(M \times C)$ matrix. More generally, the n-mode product of a p-th order tensor and a vector is a tensor of order $p-1$ in which the n-th mode is removed.

The 1-mode product of a $(2 \times 2 \times 3)$ tensor with a (5×2) matrix is illustrated in Panel A of Figure A-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.[8]

Vector and 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 and $\mathbf{X}\mathbf{A}_2^\top = \mathbf{X} \times_2 \mathbf{A}_2$ is a $(T \times U)$ matrix. If $\mathbf{a}_1$ and $\mathbf{a}_2$ are $(T \times 1)$ and $(N \times 1)$ vectors, respectively, then $\mathbf{a}_1^\top \mathbf{X} = \mathbf{X} \times_1 \mathbf{a}_1^\top$ is a

[7] Panel A of Figure A-3 shows an example for $T = 5, M = 4, C = 3$.

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

$(1\times N)$ vector and $\mathbf{X}\mathbf{a}_2 = \mathbf{X} \times_2 \mathbf{a}_2^\top$ is a $(T\times 1)$ vector.

It is instructive to express the K-factor SVD-PCA model (6-7) in tensor notation. Using n-mode multiplication, we can write

$$\mathbf{F}_K \mathbf{B}_K^\top = \mathbf{F}_K \times_2 \mathbf{B}_K \tag{9}$$

$$= \mathbf{H}_K \times_1 \mathbf{U}_K^{(1)} \times_2 \mathbf{U}_K^{(2)} \tag{10}$$

$$= \sum_{k=1}^{K} h_{kk}\, \mathbf{u}_k^{(1)} \circ \mathbf{u}_k^{(2)}. \tag{11}$$

2.3. The Tucker decomposition of a tensor

This section introduces tensor representations used in the empirical estimation below. The goal is to reduce the dimensionality of the $T\times M\times C$ data tensor, $\boldsymbol{\mathcal{X}}$, while capturing most of its variation. Hence the objective is similar to the dimension reduction of the K-factor model for two-dimensional data. Let $\hat{\boldsymbol{\mathcal{X}}}$ be an approximation of $\boldsymbol{\mathcal{X}}$ that can be expressed in terms of low dimensional matrices and tensors to be specified below. The approximation error and corresponding mean-squared error (MSE) are $\boldsymbol{\mathcal{E}}$ and MSE($\hat{\boldsymbol{\mathcal{X}}}$) as follows

$$\boldsymbol{\mathcal{X}} = \hat{\boldsymbol{\mathcal{X}}} + \boldsymbol{\mathcal{E}}, \tag{12}$$

$$\text{MSE}(\hat{\boldsymbol{\mathcal{X}}}) = \frac{1}{TMC} \|\boldsymbol{\mathcal{E}}\|^2. \tag{13}$$

Most dimension reduction models for tensors are based on the *Tucker decomposition* (Tucker (1966)). For the three-dimensional $(T\times M\times C)$ tensor $\boldsymbol{\mathcal{X}}$, the Tucker decomposition with (K_1,K_2,K_3) components, denoted Tucker(K_1,K_2,K_3), is given by

$$\hat{\boldsymbol{\mathcal{X}}}(K_1,K_2,K_3) = \boldsymbol{\mathcal{G}} \times_1 \mathbf{V}^{(1)} \times_2 \mathbf{V}^{(2)} \times_3 \mathbf{V}^{(3)} \tag{14}$$

$$= \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \sum_{k_3=1}^{K_3} g_{k_1k_2k_3}\, \mathbf{v}_{k_1}^{(1)} \circ \mathbf{v}_{k_2}^{(2)} \circ \mathbf{v}_{k_3}^{(3)}, \tag{15}$$

where $\boldsymbol{\mathcal{G}}$ is a $(K_1\times K_2\times K_3)$ tensor with elements $g_{k_1k_2k_3}$ and $\mathbf{V}^{(1)},\mathbf{V}^{(2)},\mathbf{V}^{(3)}$ are $(T\times K_1),(M\times K_2),(C\times K_3)$ matrices, respectively. $\boldsymbol{\mathcal{G}}$ is called the *core* tensor and can thought of as a "compressed" version of $\boldsymbol{\mathcal{X}}$. The *loading* matrices $\mathbf{V}^{(i)}$ with columns $\mathbf{v}_j^{(i)}$ expand the core tensor to the dimension of $\boldsymbol{\mathcal{X}}$. The optimal Tucker model minimizes the MSE (13) among all $\hat{\boldsymbol{\mathcal{X}}}(K_1,K_2,K_3)$ and can be written as (14).

Note, that the Tucker decomposition is not unique and can be rotated. Let $\mathbf{S}_i$ be nonsingular $(K_i\times K_i)$ matrices for $i=1,2,3$. Then (14) also be written as

$$\hat{\boldsymbol{\mathcal{X}}}(K_1,K_2,K_3) = (\boldsymbol{\mathcal{G}} \times_1 \mathbf{S}_1 \times_2 \mathbf{S}_2 \times_3 \mathbf{S}_3) \times_1 (\mathbf{V}^{(1)}\mathbf{S}_1^{-1}) \times_2 (\mathbf{V}^{(2)}\mathbf{S}_2^{-1})(\mathbf{V}^{(3)}\mathbf{S}_3^{-1}). \tag{16}$$

Typically, (14) is normalized so that the $\mathbf{V}^{(i)}$ matrices are orthonormal, similar to the eigenvector matrices in the SVD-PCA model. The mechanism of the Tucker decomposition (14) is illustrated in Figure 3, which shows the decomposition of a $(6\times 5\times 4)$ tensor $\boldsymbol{\mathcal{X}}$ by a Tucker model with $(K_1,K_2,K_3) = (3,2,2)$ components. The core tensor $\boldsymbol{\mathcal{G}}$ compresses $\boldsymbol{\mathcal{X}}$ to a lower-dimension of $(3\times 2\times 2)$. The Tucker component matrices $\mathbf{V}^{(1)},\mathbf{V}^{(2)}$, and $\mathbf{V}^{(3)}$ expand the core tensor to the full dimension of $\boldsymbol{\mathcal{X}}$ and have the matching 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

$$(3\times2\times2)\times_1(6\times3)\times_2(5\times2)\times_3(4\times2)=(6\times5\times4).$$

2.4. The Tucker model implies interconnected two-dimensional factor models

The mechanism of the Tucker tensor decomposition can be further understood by deriving the implications of a three-dimensional Tucker tensor for fibers (vectors) and slices (matrices) that make up the tensor.[9] This section shows that the Tucker decomposition of a tensor implies a collection of connected two-dimensional factor models. Each slice of the tensor represents a two-dimensional factor model. In other words, the Tucker decomposition (14) embeds T factor models of horizontal $(M\times C)$ slices, M factor models of lateral $(T\times C)$ slices, and C factor models of $(T\times M)$ frontal slices for a total of $T+M+C$ two-dimensional factor models.

To see this, consider first a Tucker model with $(K_1,K_2,K_3)=(1,1,2)$ components so that the loading matrices $\mathbf{V}^{(1)}$ and $\mathbf{V}^{(2)}$ are $(T\times1)$ and $(M\times1)$ vectors, respectively, and $\mathbf{V}^{(3)}$ is a $(C\times2)$ matrix. Each mode-t slice $\mathbf{X}_{mc(t)}$ is a $(M\times C)$ matrix that, by rewriting (15), can be expressed as

$$\mathbf{X}_{mc(t)}=\mathrm{v}^{(1)}_{1,t}\,\mathbf{v}^{(2)}_1\circ\left(g_{111}\mathbf{v}^{(3)}_1+g_{112}\mathbf{v}^{(3)}_2\right) \tag{17}$$

$$=\mathrm{v}^{(1)}_{1,t}\,\left(\mathbf{v}^{(2)}_1\circ\tilde{\mathbf{v}}^{(3)}_1\right), \tag{18}$$

where $\tilde{\mathbf{v}}^{(3)}_1=g_{111}\mathbf{v}^{(3)}_1+g_{112}\mathbf{v}^{(3)}_2$. The term $\mathbf{v}^{(2)}_1\circ\tilde{\mathbf{v}}^{(3)}_1$ is a $(M\times C)$ matrix and represents a one-factor model in the (M,C) modes. In particular, the t-slices $\mathbf{X}_{mc(t)}$ are formed by the *same* $(M\times C)$ matrix multiplied by the scalar $\mathrm{v}^{(1)}_{1,t}$ and are thus proportional to each other.

Each t-mode fiber $\mathbf{x}_{t(mc)}$ is a $(T\times1)$ vector and can be written as

$$\mathbf{x}_{t(mc)}=\left(g_{111}\mathrm{v}^{(2)}_{1,m}\mathrm{v}^{(3)}_{1,c}+g_{112}\mathrm{v}^{(2)}_{1,m}\mathrm{v}^{(3)}_{2,c}\right)\mathbf{v}^{(1)}_1. \tag{19}$$

The term in parentheses is a scalar, so that all fibers $\mathbf{x}_{t(mc)}$ are proportional to the $(T\times1)$ vector $\mathbf{v}^{(1)}_1$. Therefore, all-time-series fibers are perfectly correlated. Panel A of Figure A-6 shows such time-series fibers $\mathbf{x}_{t(mc)}$ in a simulated $(25\times10\times5)$ Tucker model with $(K_1,K_2,K_3)=(1,1,2)$. The plot shows six (m,c) combinations (out of $10\cdot5=50$), corresponding to six time-series fibers. All time-series are proportional and perfectly correlated.

A similar calculation shows that mode-t slices and fibers of a model with two factors in the T-dimension, $(K_1,K_2,K_3)=(2,1,2)$, are given by

$$\mathbf{X}_{mc(t)}=\mathrm{v}^{(1)}_{1,t}\,\mathbf{v}^{(2)}_1\circ\left(g_{111}\mathbf{v}^{(3)}_1+g_{112}\circ\mathbf{v}^{(3)}_2\right)+\mathrm{v}^{(1)}_{2,t}\,\mathbf{v}^{(2)}_1\circ\left(g_{211}\mathbf{v}^{(3)}_1+g_{212}\circ\mathbf{v}^{(3)}_2\right) \tag{20}$$

$$=\mathbf{v}^{(2)}_1\circ\left(\mathrm{v}^{(1)}_{1,t}\,\tilde{\mathbf{v}}^{(3)}_1+\mathrm{v}^{(1)}_{2,t}\,\tilde{\mathbf{v}}^{(3)}_2\right) \tag{21}$$

$$\mathbf{x}_{t(mc)}=\left(g_{111}\mathrm{v}^{(2)}_{1,m}\mathrm{v}^{(3)}_{1,c}+g_{112}\mathrm{v}^{(2)}_{1,m}\mathrm{v}^{(3)}_{2,c}\right)\mathbf{v}^{(1)}_1+\left(g_{211}\mathrm{v}^{(2)}_{1,m}\mathrm{v}^{(3)}_{1,c}+g_{212}\mathrm{v}^{(2)}_{1,m}\mathrm{v}^{(3)}_{2,c}\right)\mathbf{v}^{(1)}_2. \tag{22}$$

The right-hand side of (21) is the outer product of an $(M\times1)$ vector $\mathbf{v}^{(2)}_1$ and a $(C\times1)$ vector $\mathrm{v}^{(1)}_{1,t}\,\tilde{\mathbf{v}}^{(3)}_1+\mathrm{v}^{(1)}_{2,t}\,\tilde{\mathbf{v}}^{(3)}_2$ and yields a $(M\times C)$ matrix. Therefore, as in the case with a single mode-t component above, each slice $\mathbf{X}_{mc(t)}$ is formed by a $(M\times C)$ one-factor model. However, (21), unlike (18), implies that the

[9] See Panels B–F of Figure A-1.

one-factor models differ across t since the term in parentheses is a linear combination of two vectors and the weights $v_{1,t}^{(1)}$ and $v_{2,t}^{(1)}$ change across t. Equation (22) shows that t-fibers $\mathbf{x}_{t(mc)}$ are linear combinations of two $(T\times 1)$ vectors, $\mathbf{v}_1^{(1)}$ and $\mathbf{v}_2^{(1)}$, and are, therefore, not perfectly correlated. Hence, the model with two t-mode factors results in richer and more complex patterns in the t-dimension. Panel B of Figure A-6 shows the same six mode-t fibers as in Panel A but for $K_1 = 2$. Although the t-fibers are highly correlated, they are not proportional. The average of the absolute correlations of all 50 (m,c)-combinations is 0.85. Panel C shows the same plot for $K_1 = 5$ components and demonstrates that higher values of K_1 result in richer patterns across t-fibers. Note, that the t-fibers are correlated (the average correlation for $K_1 = 5$ is 0.66), which is not surprising since they stem from connected two-dimensional factor models.

It is straightforward to derive the corresponding expressions for general (K_1,K_2,K_3)-Tucker models, albeit the equations are more complicated. Suppose $K_2 \leq K_3$; then (15) implies that t-mode slices $\mathbf{X}_{mc(t)}$ and fibers $\mathbf{x}_{t(mc)}$ can be written as

$$\mathbf{X}_{mc(t)} = v_{1,t}^{(1)} \underbrace{\left[\sum_{k_2=1}^{K_2} \underbrace{\mathbf{v}_{k_2}^{(2)}}_{M\times 1} \circ \underbrace{\left(\sum_{k_3=1}^{K_3} g_{1k_2k_3} \underbrace{\mathbf{v}_{k_3}^{(3)}}_{C\times 1} \right)}_{C\times 1} \right]}_{M\times C} + \ldots + v_{K_1,t}^{(1)} \underbrace{\left[\sum_{k_2=1}^{K_2} \underbrace{\mathbf{v}_{k_2}^{(2)}}_{M\times 1} \circ \underbrace{\left(\sum_{k_3=1}^{K_3} g_{K_1k_2k_3} \underbrace{\mathbf{v}_{k_3}^{(3)}}_{C\times 1} \right)}_{C\times 1} \right]}_{M\times C}, \quad (23)$$

$$= \underbrace{\mathbf{v}_1^{(2)}}_{M\times 1} \circ \underbrace{\left[\sum_{k_1=1}^{K_1} v_{k_1,t}^{(1)} \left(\sum_{k_3=1}^{K_3} g_{k_1 1k_3} \underbrace{\mathbf{v}_{k_3}^{(3)}}_{C\times 1} \right) \right]}_{C\times 1} + \ldots + \underbrace{\mathbf{v}_{K_2}^{(2)}}_{M\times 1} \circ \underbrace{\left[\sum_{k_1=1}^{K_1} v_{k_1,t}^{(1)} \left(\sum_{k_3=1}^{K_3} g_{k_1K_2k_3} \underbrace{\mathbf{v}_{k_3}^{(3)}}_{C\times 1} \right) \right]}_{C\times 1} \quad (24)$$

$$\mathbf{x}_{t(mc)} = \underbrace{\mathbf{v}_1^{(1)}}_{T\times 1} \left(\sum_{k_2=1}^{K_2} \sum_{k_3=1}^{K_3} g_{1k_2k_3} v_{k_2,m}^{(2)} v_{k_3,c}^{(3)} \right) + \ldots + \underbrace{\mathbf{v}_{K_1}^{(1)}}_{T\times 1} \left(\sum_{k_2=1}^{K_2} \sum_{k_3=1}^{K_3} g_{K_1k_2k_3} v_{k_2,m}^{(2)} v_{k_3,c}^{(3)} \right). \quad (25)$$

The two expressions (23, 24) highlight different aspects of the Tucker factor model. Each term in square brackets in (23) is a $(M\times C)$ matrixso that a t-slice $\mathbf{X}_{mc(t)}$ is the sum of K_1 $(M\times C)$ matrices. If there is a single K_1 factor, then all t-slices $\mathbf{X}_{mc(t)}$ are formed by the same $(M\times C)$ matrix multiplied by the scalar $v_{1,t}^{(1)}$ and are thus proportional. Note, that this property holds for any K_2 and K_3. As K_1 increases, $\mathbf{X}_{t(mc)}$ is a weighted sum of more $(M\times C)$ matricesso that each additional K_1 component adds flexibility to the relationships of slices across $t = 1,\ldots,T$.

Equation (25) shows that fibers $\mathbf{x}_{t(mc)}$ are linear combinations of K_1 columns of the mode-1 $(T\times K_1)$ loading matrix $\mathbf{V}^{(1)} = [\mathbf{v}_{k_1}^{(1)}]$. The weights depend on elements of the core tensor, $\boldsymbol{\mathcal{G}} = [g_{k_1k_2k_3}]$, and the mode-2 and mode-3 loading matrices $\mathbf{V}^{(2)} = [v_{k_2,m}^{(2)}]$ and $\mathbf{V}^{(3)} = [v_{k_3,c}^{(3)}]$. If $K_1 = 1$, all mode-t fibers $\mathbf{x}_{t(mc)}$ are perfectly correlated, no matter the values of K_2 and K_3. As for t-slices $\mathbf{X}_{mc(t)}$, higher values of K_1 create more complex patterns of correlations of $\mathbf{x}_{t(mc)}$ across $m = 1,\ldots,M, c = 1,\ldots,C$.

In turn, (24) highlights the factor structure of t-slices. Each expression in square brackets is a

$(C\times 1)$ vector, which, when pre-multiplied by an $(M\times 1)$ vector $\mathsf{v}_{k_2}^{(2)}$, yields a $(M\times C)$ matrix. Therefore, each $\mathbf{X}_{mc(t)}$ represents a two-dimensional K_2-factor model given by the weighted sum of K_2 $(M\times C)$ matrices, see (11). The K_2 $(C\times 1)$ vectors in square brackets are composed of outer products of the same $\mathsf{v}_{k_1,t}^{(1)}$ and $\mathsf{v}_{k_3}^{(3)}$ vectors but depend on the weights $g_{k_1k_2k_3}$. Note, that the vectors in the M-dimension, $\mathsf{v}_1^{(2)},\ldots,\mathsf{v}_{K_2}^{(2)}$ are identical for all t, while the vectors in the C-dimension (in square brackets) differ across t.

Equations (23) and (24) were derived under the assumption that $K_2 \leq K_3$, but it is straightforward to derive the corresponding expressions for $K_3 \leq K_2$ by changing the sequence of taking sums over k_2 and k_3. Hence, all results hold in both cases. The only difference is that the order of the $(M\times C)$ factor model for t-slices $\mathbf{X}_{mc(t)}$ in (24) is K_3, instead of K_2. In general, $\mathbf{X}_{mc(t)}$ slices are formed by a $(M\times C)$ factor model of order $\mathsf{min}(K_1,K_2)$.

The expressions (18)–(24) apply to slices in the mode-1 t-dimension. If a dimension of a data matrix $\mathbf{X}$ or a data tensor $\boldsymbol{\mathcal{X}}$ is a "time" index, then corresponding rows, columns, or fibers form a set of time-series, usually indexed by t. However, in factor models, there is nothing special about a "time" dimension and a mode with a time index is treated as any other index or unit, *i.e.* all dimensions are interchangeable.[10] The derivation above would be identical if the term "time dimension" was replaced by "first dimension". Therefore, the results above apply, without loss of generality, to slices and fibers in other dimensions as well, *i.e.* slices in the second dimension, M, and the third dimension, C.[11]

The results in this section show that a three-dimensional Tucker model implies a rich set of $T+M+C$ interconnected two-dimensional factor models. The number of factors in dimension i, K_i, affects the complexity of the correlation structure of mode-i fibers (vectors) and the relationships of mode-i slices $\mathbf{X}_{mc(t)}$ across t. If $K_i = 1$, then all mode-i fibers are perfectly correlated and all slices $\mathbf{X}_{mc(t)}$ are proportional. Additional components, $K_i > 1$, yield different correlation patterns. The number of components in the other dimensions determines the order of the two-dimensional factor models that create all mode-i slices. These results are summarized as follows:

- $K_i = 1$: mode-i fibers are perfectly correlated, mode-i slices are proportional.
- $K_i > 1$: mode-i fibers not perfectly correlated, mode-i slices are connected but not proportional.
- t-slices $\mathbf{X}_{mc(t)}$ are $(M\times C)$ factor models with $\mathsf{min}(K_2,K_3)$ factors for all $t = 1,\ldots,T$.
- m-slices $\mathbf{X}_{tc(m)}$ are $(T\times C)$ factor models with $\mathsf{min}(K_1,K_3)$ factors for all $m = 1,\ldots,M$.
- c-slices $\mathbf{X}_{tm(c)}$ are $(T\times M)$ factor models with $\mathsf{min}(K_1,K_2)$ factors for all $c = 1,\ldots,C$.

2.5. Interpretation of the Tucker decomposition

Equation (26) suggests that the loading matrices $\mathbf{V}^{(i)}$ can be interpreted as factor loading matrices of the implied two-dimensional factor models. In turn, the columns of the matrix $\mathbf{S}_{(jk)}^{\mathsf{T}}$ are equivalent to

[10] Dynamic factor models are an exception since they specify a VAR in the time dimension, *e.g.* Forni, Hallin, Lippi and Reichlin (2000) and Stock and Watson (2006).

[11] The interchangeability of indices applies to two-dimensional SVD-PCA models as well. Suppose $\mathbf{X}$ is a $(T\times N)$ matrix where rows represent a time index. As noted in section 2.1, SVD-PCA factor models of $\mathbf{X}$ and the $(N\times T)$ matrix $\mathbf{X}^{\mathsf{T}}$ are equivalent.

factors. It is important to stress that (26) is derived from the Tucker representation (14). One could also directly estimate a two-dimensional SVD-PCA factor model for $\mathbf{X}_{(3)}^{\top}$ with K_3 factors. However, the separately estimated models are not consistent with each other.

The core tensor can also be interpreted as a compressed, or "summary" version of the data matrix. This is the easiest to see if one of the dimensions is a time index. For the data set used in the next section, the three modes are time, mutual fund, and characteristic, which I will use again as an example. The modes of the core tensor have the same modes as the data tensor, so each element g_{tmc} of $\boldsymbol{\mathcal{G}}$ can be interpreted as an observation of a "summary" element with time, mutual fund, and characteristic dimensions. The elements of $\boldsymbol{\mathcal{G}}$ are constructed from the data tensor to summarize the information in $\boldsymbol{\mathcal{X}}$ in a more compact form with dimensions (K_1,K_2,K_3) instead of (T,M,C). Multiplying $\boldsymbol{\mathcal{G}}$ with $\mathbf{V}^{(1)}$ along the first mode yields a tensor $\boldsymbol{\mathcal{S}}_{(1)}$ with dimensions (T,K_2,K_3). Each of the $K_1\times K_2$ vertical mode-1 fibers of $\boldsymbol{\mathcal{S}}_{(1)}$ is a $(T\times 1)$ vector can be written as $\mathbf{s}_{(tk_2k_3)}$. $\mathbf{s}_{(tk_2k_3)}$ represents a time-series of the "summary" core mutual fund k_2 and "summary" core characteristic k_3. This interpretation of the Tucker representation can be pushed further. If $\boldsymbol{\mathcal{S}}_{(1)}$ is multiplied by $\mathbf{V}^{(3)}$ along mode 3, the resulting tensor $\boldsymbol{\mathcal{S}}_{(13)}$ has dimension $(T\times K_2\times C)$. Each t-fiber of this tensor, $\mathbf{s}_{(tk_2c)}$ is a time-series of characteristic c of the "summary" core mutual fund k_2.[12] By the same token, $\boldsymbol{\mathcal{S}}_{(12)} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{V}^{(1)} \times_2 \mathbf{V}^{(2)}$ is a $(T\times M\times K_3)$ tensor with t-fibers $\mathbf{s}_{(tmk_1)}$ representing a time-series of "summary" characteristic k_1 of mutual fund m.

A key feature of the Tucker decomposition is that it can be rewritten in different ways to highlight certain aspects of the model. For example, the three-dimensional tensor decomposition implies one "stacked" two-dimensional factor model for each of the three modes. Suppose we are interested in the factor structure of the third C-mode, which corresponds to characteristics in the empirical implementation below. The key is to write the tensors of the Tucker representation as matrices in a specific way.[13]

[12] Figure A-5 illustrates the construction of the "summary" mutual funds and "summary" characteristics $\mathbf{s}_{(tk_2k_3)}$ in Panel A and the summary mutual fund $\mathbf{s}_{(tk_2c)}$ in Panel B. The data tensor $\boldsymbol{\mathcal{X}}$ consists of $T=6$ periods, $M=5$ mutual funds, and $C=3$ characteristics. The core tensor $\boldsymbol{\mathcal{G}}$ is compressed to $K_1=3$ periods, $K_2=2$ funds, and $K_3=2$ characteristics. Multiplying $\boldsymbol{\mathcal{G}}$ by the (6×3) loading matrix $\mathbf{V}^{(1)}$ expands the core from four to six periods while keeping the number of summary mutual funds and characteristics unchanged, see Panel A. Therefore, the resulting tensor $\boldsymbol{\mathcal{S}}_{(1)}$ consists of six periods, two summary funds (in mode-2), and two summary characteristics (in mode-3) so that each vertical fiber of $\boldsymbol{\mathcal{S}}_{(1)}$ represents a time-series of the $2\times 2=4$ combinations of the summary fund and characteristic. The figure shows the 6 time-series observations of the first ($k_2=1$) summary fund and first ($k_3=1$) summary characteristics, $\mathbf{s}_{(t11)}$, and the second ($k_2=2$) summary fund and first ($k_3=1$) summary characteristics, $\mathbf{s}_{(t21)}$, in darker shades. Panel B shows the expansion of the core tensor along the mode-1 (time) and mode-3 (characteristics) dimensions by multiplying the core tensor by $\mathbf{V}^{(1)}$ and $\mathbf{V}^{(3)}$. The resulting tensor $\boldsymbol{\mathcal{S}}_{(13)}$ consists of six periods and all four characteristics in modes one and three with the two summary mutual funds in the second mode. The darker shades indicate the time-series observations of the first and second summary funds for characteristics c, $\mathbf{s}_{(t1c)}$ and $\mathbf{s}_{(t2c)}$.

[13] Figure A-4 illustrates the steps of the derivation for the example in Figure 3. The dimensions of the data matrix $\boldsymbol{\mathcal{X}}$ are $T=6, M=5$, and $C=4$ and those of the core tensor $\boldsymbol{\mathcal{G}}$ are $(K_1,K_2,K_3)=(3,2,2)$. Since $\mathbf{V}^{(1)}$ and $\mathbf{V}^{(2)}$ are of dimensions (6×3) and (5×2), respectively, $\boldsymbol{\mathcal{S}}_{12}\times_1 \mathbf{V}^{(1)} \times_2 \mathbf{V}^{(2)}$ has dimensions $(6\times 5\times 2)$. Unfolding $\boldsymbol{\mathcal{S}}_{(12)}$ along mode-3 yields $\mathbf{S}_{(1,2)}$, which is (2×30). Unfolding the original $(6\times 5\times 4)$ data tensor $\boldsymbol{\mathcal{X}}$ along mode-3 yields a (4×30) matrix $\mathbf{X}_{(3)}$. Since $\mathbf{V}^{(1)}$ is (4×2), we obtain the factor model (26) with 30 rows and two factors by transposing $\mathbf{X}_{(3)}$,$\mathbf{S}_{(12)}$, and $\mathbf{V}^{(1)}$. This process can be repeated for modes 1 and 2. The factor model for mode-1 (T) has 20 rows and three factors and that of mode-2 (M) has 24 rows and two factors.

1. Unfold the $(T\times M\times C)$ tensor $\hat{\boldsymbol{\mathcal{X}}}$ along the third mode $\rightarrow$ $(C\times TM)$ matrix $\mathbf{X}_{(3)}$.
2. Multiply $\boldsymbol{\mathcal{G}}$ by $\mathbf{V}^{(1)},\mathbf{V}^{(2)}$ along modes 1 and 2 $\rightarrow$ $(T\times M\times K_3)$ tensor $\boldsymbol{\mathcal{S}}_{(12)}=\boldsymbol{\mathcal{G}}\times_1\mathbf{V}^{(1)}\times_2\mathbf{V}^{(2)}$
3. Unfold $\boldsymbol{\mathcal{S}}_{(12)}$ along mode-3 $\rightarrow$ $(K_3\times TM)$ matrix $\mathbf{S}_{(12)}$
4. Write the implied two-dimensional $(TM\times C)$-factor model with K_3 factors as follows

$$\underbrace{\mathbf{X}_{(3)}^{\top}}_{TM\times C}=\underbrace{\mathbf{S}_{(12)}^{\top}}_{TM\times K_3}\underbrace{\mathbf{V}^{(3)\top}}_{K_3\times C}+\underbrace{\mathbf{E}_3}_{TM\times C}=\hat{\mathbf{X}}_{(3)}^{\top}+\mathbf{E}_3. \tag{26}$$

The implied stacked two-dimensional factor models for the T- and M-modes are given by

$$\underbrace{\mathbf{X}_{(1)}^{\top}}_{MC\times T}=\underbrace{\mathbf{S}_{(23)}^{\top}}_{MC\times K_1}\underbrace{\mathbf{V}^{(1)\top}}_{K_1\times T}+\underbrace{\mathbf{E}_1}_{MC\times T}=\hat{\mathbf{X}}_{(1)}^{\top}+\mathbf{E}_1, \tag{27}$$

$$\underbrace{\mathbf{X}_{(1)}^{\top}}_{TC\times M}=\underbrace{\mathbf{S}_{(13)}^{\top}}_{TC\times K_2}\underbrace{\mathbf{V}^{(2)\top}}_{K_2\times M}+\underbrace{\mathbf{E}_2}_{TC\times M}=\hat{\mathbf{X}}_{(2)}^{\top}+\mathbf{E}_2. \tag{28}$$

The two-factor models (26, 28, 27) have the familiar properties of two-dimensional factor models and can be interpreted as such. Since they all stem from the same three-dimensional Tucker model, they are internally consistent with each other.

2.6. Data compression

The core tensor $\boldsymbol{\mathcal{G}}$ has $K_1K_2K_3$ elements and the component matrices $\mathbf{V}^{(1)},\mathbf{V}^{(2)}$, and $\mathbf{V}^{(3)}$ have TK_1,MK_2, and CK_3 elements, respectively. The orthonormal normalization adds $K_1^2+K_2^2+K_3^2$ restrictions. Thus, the Tucker decomposition (14) has $TK_1+MK_2+CK_3+K_1K_2K_3-(K_1^2+K_2^2+K_3^2)$ free parameters. Consider the special case where $T=M=C=N,K_1=K_2=K_3=K$. Furthermore, assume that the ratio of number of Tucker components K to N is a constant, $Q=K/N$. I will compare the number of elements of the data tensor, N^3, to the $3NK+K^3-3K^2$ parameters of the Tucker representation. The data-compression ratio κ is defined as the number of free parameters divided by the numbers of data points. For the three-dimensional Tucker model κ_{3d} is equal to $3K/N^2+(K/N)^3-3K^2/N^3=Q^3+3Q(1-Q)/N$ and converges to Q^3 as $N\longrightarrow\infty$. Hence, the three-dimensional Tucker model compresses the data's total size by a ratio of order $\mathcal{O}((K/N)^3)$. If $Q=K/N$ is 10%, *i.e.*, there is one Tucker component for every ten data dimensions, $\kappa_{3d}=0.001$ so that the Tucker model compresses the data by approximately 99.9%. The two-dimensional K-factor SVD-PCA model has $2NK-K^2$ free parameters so that the compression ratio is $\kappa_{2d}=2Q(1-Q)$. For $Q=0.1$, $\kappa_{2d}=0.18$, which is an order of magnitude higher than κ_{3d} of the Tucker model. Hence higher-order tensor decompositions achieve more efficient dimensionality reduction than two-dimensional factor models.

2.7. Computation

In contrast to the SVD-PCA matrix representation, there are no closed-form solutions for the Tucker decomposition (14) that minimize the MSE (13). Kolda and Bader (2009) and Kroonenberg (2007), chapter 10, discuss several numerical solutions methods. I will use the most popular algorithm of *alternating least squares* (ALS). 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 $\mathbf{V}^{(j)}, j > 1$, solving for $\mathbf{V}^{(1)}$, and iteratively solving for $\mathbf{V}^{(i)}$ until convergence. Once the $\mathbf{V}^{(i)}$ are solved, the core tensor $\boldsymbol{\mathcal{G}}$ can be constructed.

One important property of the Tucker decomposition is that they cannot be computed sequentially. Consider two Tucker models with Tucker decompositions with (K_1, K_2, K_3) and $(K_1', K_2', K_3'), K_j' < K_j$, respectively. The first K_1', K_2', K_3' components of the Tucker (K_1, K_2, K_3) model are in general different from the Tucker (K_1', K_2', K_3') model. In contrast, the first K factors of a two-dimensional SVD-PCA factor model with $K' > K$ factors are the same as those of a K-factor model since they are based on eigenvalues and eigenvectors.

3. Mutual fund characteristics over time

Next, I estimate Tucker models using a data set on mutual fund characteristics. The data is taken from Lettau, Ludvigson and Manoel (2021) and I refer to their paper for a detailed description and construction of the data. Lettau, Ludvigson and Manoel (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[14], and four liquidity measures, see Table 1. To obtain a balanced panel with no missing data, I select all 980 mutual funds and ETFs that are in the sample for all quarters between 2010Q3 and 2018Q4.[15] The final sample consists of $T = 34$ quarters, $M = 1{,}342$ mutual funds and ETFs, and $C = 25$ characteristics for a total of 1,140,700 observations.

Table 2 shows some properties of the mutual funds in the sample. I first take means across all funds and then compute descriptive statistics of the distribution of fund means. The median fund has a total net asset value (TNA) of \$590 mil. with an inter-quartile range of just under \$195 mil. to \$1.72 bil. The mean TNA of \$2.29 bil. is larger than the 75%-th percentile indicating that the TNA distribution is heavily right-skewed, as is the distribution of the number of stocks in mutual fund portfolios. The median fund holds 89 stocks with an interquartile range of 58 to 170 with a mean of 196. The market beta of most mutual funds is between 0.52 and 1.06. As is well known in the literature, mutual funds underperform broad stock market indices, and alphas of the majority are negative.

The data set is arranged in a three-dimensional tensor $\boldsymbol{\mathcal{X}}$. The first mode represents the time index t, the second mode represents mutual funds m, and the third mode represents characteristics c so that the data tensor $\boldsymbol{\mathcal{X}}$ has dimension $(T \times M \times C) = (34 \times 880 \times 25)$.

Any factor model depends on the correlation structure of the data. In two-dimensional data, the correlation matrix is also two-dimensional and can be easily understood. Computing correlations in higher-dimensional data is more complicated. Suppose we are particularly interested in correlations

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

[15] Choosing an earlier starting date drastically reduces the number of funds without missing values.

across characteristics (mode-3). One possibility is to unfold the three-dimensional data tensor along the characteristic dimension into a (15×45628) matrix and compute the correlation matrix of the transpose. However, there might be important interactions across time and/or mutual funds, which would not be captured by this correlation matrix. Alternatively, one can compute correlation matrices holding either the time or fund index fixed. In other words, for each date t, we can calculate the cross-sectional correlation of characteristics across funds, and for each fund m, we can compute the time-series correlations of characteristics.

Figure 1 shows a heatmap of the means of the two correlation measures. The upper right triangle plots the mean of time-series correlations across mutual funds, and the lower triangle shows the mean cross-sectional correlations across dates. Comparing the two correlation measures shows that the overall patterns are similar, but cross-sectional correlations in the lower-left triangle are on average larger in absolute value than time-series correlations in the upper-right triangle. Not surprisingly, price-ratio characteristics are positively correlated, as are characteristics related to the growth of fundamentals, but the two blocks are negatively correlated. Since the Morningstar variables MS, MULT, and GR are based on price-ratios and growth rates, their correlation pattern is to a large degree mechanical. Investment, momentum, and reversal are negatively related to price ratios but positively related to growth rates, and size is positively correlated with higher liquidity.

Recall that Figure 1 is based on the means of the correlation distribution and therefore cannot capture more complex relationships in the three-dimensional data set. The goal of higher-dimensional factor models is to capture important patterns in the joint distribution of the data that go beyond linear correlations. However, some features of factor models are directly related to the correlations shown in Figure 1.

3.1. Estimation of Tucker models

I start with the estimation of Tucker models for a wide range of combinations of (K_1,K_2,K_3). Each model is estimated using the ALS method described in Section 2.7. The starting values of each mode-i Tucker loading matrix are set to the two-dimensional SVD decompositions computed from the unfolded tensor along mode i. Typically, the ALS algorithm converges after 10 to 50 iterations depending on the number of components. The procedure is robust to other starting values, albeit at the cost of slower convergence.

For each combination of (K_1,K_2,K_3), I compute the Tucker decomposition (6) and the associated $\text{MSE}(K_1,K_2,K_3)$. The MSE can be compared to the total variance of the data tensor $\text{Var}(\boldsymbol{\mathcal{X}}) = 0.54$. Since the mean of the errors of the Tucker approximation $\boldsymbol{\mathcal{E}}$ is close to zero, $1-\text{MSE}(K_1,K_2,K_3)/\text{Var}(\boldsymbol{\mathcal{X}}) \approx 1-\text{Var}(\boldsymbol{\mathcal{E}}(K_1,K_2,K_3))/\text{Var}(\boldsymbol{\mathcal{X}})$ can be interpreted as a pseudo-R^2 of the model.

The results are shown in the first three panels of Figure 4. Each plot shows the MSE as a function of the number of components along one mode while keeping the numbers of the other two components fixed. Panel A plots the $\text{MSE}(K_1,K_2,K_3)$ as a function of K_1 for four different combinations of (K_2,K_3): $\text{MSE}(K_1,1,1)$ in red, $\text{MSE}(K_1,10,5)$ in blue, $\text{MSE}(K_1,20,5)$ in orange, and $\text{MSE}(K_1,40,15)$ in black. The "minimal" Tucker model with a single component in each mode collapses the 34 quarters, 980 mutual funds, and 25 characteristics into a single "summary" mutual fund with a single

"summary" characteristic observed at one "summary" quarter. The MSE of the minimal Tucker(1, 1, 1) model represented by the left-most point on the red line in Panel A is 0.32, corresponding to an R^2 of 40%. Note, that increasing K_1 while keeping $K_2 = K_3 = 1$ does not reduce the MSE further.

However, the MSE is reduced significantly when K_2 and K_3 are larger than one. The blue line sets $K_1 = 10$ and $K_3 = 5$. The MSE is 0.08 for a single K_1 component, which is equivalent to an R^2 of 84% and thus twice as high as the R^2 of the minimal Tucker(1, 1, 1) model. Changing K_1 from one to five decreases the MSE to 0.07 but increasing K_1 further has a negligible effect on the MSE. The Tucker(1,10,5) model has 13,503 degrees of freedom, thus reducing the data dimensionality of 1,140,700 by 99%. The dimension reduction is equivalent to approximating a two-dimensional panel of 100 variables and 200 time-series observations with a factor model with two principal components.

Panel A also shows the MSE for $(K_1,20,10)$ and $(K_1,40,20)$. The MSE is close to 0.07 when $K_1 = 1$ for both cases and decreases for K_1 values up to 10 before leveling off around 0.04 for (10,20,10) and 0.03 for (10,40,15). The corresponding R^2's are 92% and 96%, respectively. These two specifications have 28,830 and 56,470 degrees of freedom, respectively, compressing the data dimensions by 98% and 95%, respectively. These values compare to the dimension reduction of a 3-factor SVD-PCA model for a panel of size (200,100).

Panel B has the same format but shows the MSE as a function of the number of mode-2 components, K_2. Recall that the second mode of the data tensor $\boldsymbol{\mathcal{X}}$ corresponds to the 980 mutual funds in the sample. Hence, I consider a broader range of values of K_2 from 1 to 100. Based on the results in Panel A, I fix the number of mode-1 components at 10 and plot the MSE for three values of K_3, 5, 10, 15. The MSE for all three specification models with $K_2 = 1$ is 0.32 and only slightly below the MSE of the minimal (1,1,1) model. Increasing K_2 drastically lowers the MSE but the effect flattens out for $K_2 \gtrsim 20$. The MSE for the (10,40,10) and (10,40,15) models are 0.03 and 0.02, respectively, corresponding to R^2's of 94% and 95%. The dimension reduction ratio is close 95% for both specifications.

Finally, Panel C plots the MSE as a function of K_3 for three combinations of K_1 and K_2: $(10,10,K_3)$, $(10,40,K_3)$, and $(20,40,K_3)$. The MSE declines steeply for $K_3 \leq 3$ and declines at a lower rate for larger K_3. The MSE of the $(10,40,K_3)$ and $(20,40,K_3)$ models are almost identical and lower than the MSE of the $(10,10,K_3)$ model. Note, that models with a single mode-1 component yield a reasonably good fit (see Panel A, Tucker(1, 10, 5), Tucker(1, 20, 10), Tucker(1, 40, 15)), while the fit of models with a single mode-2 or mode-3 components have a poor fit. This difference is due to the particular structure of the estimated core tensor of the Tucker model, as I will show in the next section.

These results suggest that Tucker models with $K_1 \approx 10, K_2 \approx 15$, and $K_3 \approx 10$ components offer the best fit (R^2 over 90%) *vs.* parsimony (over 94% dimension reduction) trade-off. I choose the model with $(K_1,K_2,K_3) = (8,15,8)$ components as the benchmark specification for the rest of the paper. The MSE of this model is 0.03, and the R^2 is 94%, while compressing the 1,140,700 observations by 97% to a model with 35,559 degrees of freedom. The dimension of the data tensor is reduced from (34,1324,25) to a Tucker model with a (8,15,8) dimensional core, which can be interpreted as 25 summary mutual funds with nine summary characteristics observed over 10 summary periods.

3.2. The fit of the Tucker benchmark model

In this section, I will analyze the fit of the Tucker(8, 15, 8) benchmark model in more detail. Table 3 shows descriptive statistics of the distributions of the errors. The mean and median error of the Tucker(8, 15, 8) models are close to zero and the standard deviation of the Tucker model is 0.18. The error distribution exhibits significant skewness, but has fat tails, as indicated by excess kurtosis of 3.2. The percentiles reported in Panel B confirm the presence of some extreme outliers. The interquartile range of the errors is (-0.098, 0.097) and 90% and 99% of the errors are in the intervals (-0.292, 0.293) and (-0.590, 0.609), respectively. The high kurtosis of the error distribution is primarily driven by the presence of outliers. The minimum and maximum errors are -1.782 and 1.888, respectively. The outliers can be seen in Panel A of Figure 5, which shows the errors sorted from most negative to most positive. The figure shows that errors are small for the majority of observations, suggesting that the Tucker model yields an overall good fit. However, the plot also shows the outliers at both ends of the distribution.

Since the data is three-dimensional, it is difficult to visualize the patterns of model fit. It is helpful to "collapse" one or two dimensions and analyze the resulting two or three dimensions. I start by collapsing two dimensions and studying the model errors along the remaining dimension. The remaining three panels of Figure 5 show MSE when errors are aggregated along each of the modes, time, mutual funds, and characteristics. Panel B shows the MSE across funds and characteristics for each quarter in the sample. The plot shows that there is little variation across quarters, indicating that the model yields a good fit over the whole sample. Panel C shows the MSE aggregated across the 980 mutual funds, sorted from the fund with the smallest MSE to the fund with the largest MSE. The MSE of the Tucker model is below 0.05 (0.1) for 83% (95%) of all funds. However, there are a few outlier mutual funds with much larger MSEs, for example, thirteen funds have MSEs higher than 0.15. I conclude that the Tucker(8, 15, 8) yields a good fit for all but very few mutual funds.

Finally, MSEs aggregated on the characteristic level are plotted in Panel C. There are several interesting patterns. 23 out of 25 have an MSE below 0.05, while the MSEs of momentum and reversals are by far the highest (0.57 and 0.69, respectively). The MSEs of size (me) and volume (vol) are the lowest (0.013 and 0.014, respectively). Why does the fit vary across characteristics? It turns out there is a link between fit and the time series behavior of characteristics. Momentum and reversal are the two most volatile characteristics and have the lowest serial correlations (along with PS liquidity). Moreover, the volatilities and serial of the fitted values are significantly different from those in the data. For example, the mean autocorrelations of momentum and reversal are 0.65 and 0.41, respectively, but the mean autocorrelations of the fitted model are markedly higher (0.73 and 0.72, respectively). More generally, the fit is better for more persistent and less volatile characteristics than for less persistent and more volatile characteristics.[16] Of course, the model fit depends on other factors, in particular, the factor structure across characteristics, as well.

Finally, I plot the time-series of individual mutual fund/characteristics pairs to illustrate the fit of the models in more detail. I consider characteristics related to the non-market factors of the

[16] Figure A-7 in the Appendix shows standard deviations and autocorrelations by characteristic in the data and fitted values.

5-factor Fama-French model, the book-to-market ratio (BM), profitability (OP), investment (INV), and momentum (MOM). For each characteristic, I show the fitted time-series of the mutual funds with the lowest and highest MSEs in the sample, as well as a fund with a "fair" fit.[17] Figure 8 shows a row for each of the four characteristics with three plots for individual funds. Each panel plots the data in black and the fitted values of the Tucker model in orange. The "best" and "worst"-fit funds are in the first and third columns, respectively. The legends include the wficn of the fund that is plotted as well as the MSE of the Tucker model.

The respective characteristics of the funds with the best fits in the left column are stable over the sample and the fitted time-series match the level and most of the small variations data closely. The MSEs are below 0.005. The funds in the middle column exhibit more time-series variation. For example, the book-to-market ratio in Panel A decreases over the sample from about 3 to 1.5 (orange line) and the fitted BM ratio tracks the decline reasonably well, albeit more slowly. The patterns of OP and INV of the funds in the middle column in Panels B and C are similar and the fitted values yield similar fits. Momentum is typically less persistent than the other characteristics and the fund in Panel C exhibit significant high-frequency variation. The fitted time-series of the Tucker model matches observed MOM fairly well but its MSE is about twice as high as for the other characteristics in the middle column.

The right column of Figure 8 shows the funds with the highest MSEs in the sample. The book-to-market ratio of the fund in panel A fluctuates between 3.9 and 4.6. The fitted time-series mirrors the movement of the observed BM but is too low throughout the sample resulting in a poor fit and high MSE of 0.22. The operating profitability of the fund in Panel B profitability equals the maximum value of five throughout the sample. Since the observed data is at a boundary, this is a challenging scenario for factor models. The fitted values for both models are around four, leaving a level gap over the entire sample. The MSE of the Tucker model is 0.74 and is the fourth-highest among all fund/characteristic combinations, so this plot shows one of the worst-case fits of the Tucker model in the entire sample. Finally, momentum of the fund shown in Panel D changes substantially over the sample and ranges from 1.8 to 4.9. The fitted time-series captures the pattern of the observed data to a degree but does not match the swings fully.

3.3. Properties of the core tensor and loading matrices of the Tucker model

Recall that the Tucker decomposition (14) of a three-dimensional tensor consists of a core tensor $\boldsymbol{\mathcal{G}}$ and loading matrices $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$. For the Tucker(8, 15, 8) model estimated for a data tensor with dimensions $(34 \times 880 \times 25)$, $\boldsymbol{\mathcal{G}}$ is a $(8 \times 15 \times 8)$-dimensional tensor, $\mathbf{V}_1$ is (8×34), $\mathbf{V}_2$ is (15×880), and $\mathbf{V}_3$ is (8×25). This section focuses on the properties of these variables. Panel A of Figure 9 plots the 25 largest elements by the absolute value of the core tensor $\boldsymbol{\mathcal{G}}$ on a log-scale, and Table 4 lists the largest 15 elements with their indices. The first core element with index (1, 1, 1) is the largest element, with a value of 1.46. The next two largest values are 0.18 and 0.14 for the elements with indices (1, 2, 2) and (1, 3, 3), respectively, followed by five elements with values between 0.02 and

[17] The MSEs of the chosen funds are in the top 10% of all funds and understate the "typical" fit of the model.

0.04. Recall the Tucker decomposition is a generalization of the SVD decomposition for matrices. In many economic and finance-related applications, the first eigenvalue is often significantly larger than the other eigenvalues. Even though the core tensor is not related to eigenvalues, the spectrum of the core elements is similar to typical eigenvalue spectrums. Furthermore, the five largest and 12 of the 13 largest core elements have a mode-1 index of one. In other words, the core tensor is dominated by elements from the first (time) index, suggesting that the first (time) dimension plays a particularly important role. This explains the earlier result that Tucker models with a single mode-1 time component have a surprisingly good fit (recall Figure 4).

Next, I analyze the structures of the loading matrices $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$. Recall that the loading matrices in the higher-dimensional Tucker decomposition are similar to the matrices of eigenvectors in the SVD matrix decomposition (1) and can be interpreted accordingly. The $\mathbf{V}_i$ matrices are also related to the factor and loading matrices of the two-dimensional factor model (3). I have also shown in (26) that the three-dimensional Tucker decomposition implies three two-dimensional factor representations in which $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$ are loading matrices for appropriately defined factors.

Panel A of Figure 10 shows the heatmap of the (8×34)-dimensional matrix $\mathbf{V}_1$. Rows correspond to the 34 time-series observations, and columns correspond to the 10 mode-1 components of the Tucker model. The first row represents the first quarter in the sample, 2010Q3, and the last quarter, 2018Q4, is in the bottom row. All elements of the first column of the heatmap are between 1.32 and 2.41, suggesting that the first component has the interpretation of a mean, or "level" factor. This is similar to the first eigenvector with only positive (or all negative) elements, as is often the case in finance applications. All other columns have positive and negative elements and have the same interpretation as higher-order eigenvectors as "long"/"short" factors. For example, the values of the second component are negative over the first part of the sample and positive in the latter part and thus a "slope" component. In contrast, the third component represents "curvature."

The loading matrix for the second mode, $\mathbf{V}_2$, in Panel B has 980 rows and 25 rows and is more difficult to visualize. To make the heatmap readable, I plot only the first 10 columns and sort each column from high to low. Hence, each of the 980 rows plots different mutual funds. The first component has again only positive values and represents a "level" factor. All higher-order components are "long/short" factors.

Finally, the (25×8) loading matrix of the third mode representing characteristics is displayed in Panel C. As before, the first component has the interpretation of a "level" factor. The values of the second component for the characteristics related to price-multiples are positive, while values of growth-related characteristics, and investment are negative. The second component is related to average correlations across characteristics, which were discussed above and are shown in Figure 1. Blue blocks have positive values in the second component and blocks in red have negative values. Since the heatmap in Figure 1 is based on average correlations across funds or dates, the second characteristic component can be interpreted as picking up the mean correlations. The interpretations of higher-order components in Panel C is less obvious. For example, the third component is composed of relatively few characteristics with large weights (BM, SP, ADJBM, BIDASK, and TURN have positive weights, and DP, ME, OP, and VOL have negative weights). Some characteristics are

significant for many components (BM, DP, TURN), while others are represented in few components (PSLIQ, MS).

3.4. Summary funds/characteristics

As shown in Section 2.5, the elements of the Tucker core tensor can be interpreted as (8,15,8) "summary" time/fund/characteristic observations. I follow the example given in that section and compute the time-series of the (15×8) summary fund/characteristics by multiplying the core tensor $\boldsymbol{\mathcal{G}}$ by the mode-1 loading matrix $\mathbf{V}_1$ to obtain $\boldsymbol{\mathcal{S}}_{(1)}$. Each fiber (k_2,k_3) of $\boldsymbol{\mathcal{S}}_{(1)}$ represents a time-series of the k_2-th summary mutual fund with summary characteristic k_3.

Figure 11 plots the time-series of the first four summary fund/characteristic combinations. The first column shows the first two "diagonal" combinations, and the second column shows the first two "off-diagonal" combinations. The plots show that the diagonal and off-diagonal combinations have different properties. The diagonal elements are "level" factors with positive values over the entire sample, while the off-diagonal combinations flip signs and are thus "long/short" factors. Furthermore, the magnitudes of the summary fund/characteristics decrease by row. The first summary fund/characteristic ranges from 2.47 to 2.54, while the second diagonal combination ranges from 0.26 to 0.35. Hence the first diagonal summary fund/characteristic combination is the dominant level-factor followed by the second, third, and so on.

Following the exposition in Section 2.5, I multiply $\boldsymbol{\mathcal{S}}_{(1)}$ by $\mathbf{V}_3$ along mode-3 to obtain $\boldsymbol{\mathcal{S}}_{(13)}$ with dimension $(34,25,25)$. Each fiber $\mathbf{s}_{(tk_2c)}, k_2 = 1,...,25$ of $\boldsymbol{\mathcal{S}}_{(1)}$ corresponds to a time-series of characteristic c of the k_2-th summary mutual fund. Figure 12 shows time-series plots of $\mathbf{s}_{(tk_2c)}$ for c = BM, OP, INV, MOM. The panels in the left column plot the first three summary mutual funds, *i.e.*, $k_2 = 1,2,3$, while the right column plots the first summary fund ($k_2 = 1$) along with the mean of the data tensor across all mutual funds in each t and given characteristic c. To highlight their comovement, I standardize the series shown in the right column.

Panel A shows the plot for the book-to-market ratio. Since the level of the first fund differs from the levels of the other summary funds, I use a different y-scale for the first fund. The book-to-market ratio of the first fund ranges from 3.5 to 3.75, while BM of the second and third fund are much low and ranges between 0.55 and 0.75. Moreover, the first summary mutual fund is closely linked to the cross-sectional mean of the book-to-market ratio of all funds in the sample, as shown in Panel B. The lower frequency variations of the first summary mutual fund and the time-series of BM-means across funds are almost identical. In other words, the book-to-market ratio of the first summary fund can be interpreted as a mean book-to-market factor. The higher-order summary mutual funds capture different aspects of the joint time-series and cross-sectional BM distribution that are unrelated to the mean but important for the overall fit of the model.

The corresponding plots for profitability, investment, and momentum confirm this interpretation of the summary mutual funds. The levels of the first funds are larger than those of the other funds, and the first summary funds are linked to the cross-sectional means. This is also true for the other 21 characteristics that are not plotted.

By the token, the mode-3 fibers $\mathbf{s}_{(tmk_3)}$ of the tensor $\boldsymbol{\mathcal{S}}_{(12)} = \boldsymbol{\mathcal{G}}\times_1\mathbf{V}_1\times_2\mathbf{V}_2$ are time-series of the

$k_3 = 1,...,9$ components of mutual fund m. I will use the estimated $\mathbf{s}_{(tmk_3)}$ to construct asset pricing factors in Section 4.

The contributions of summary quarters/funds/characteristics to the overall fit of the model can also be illustrated by considering an individual mutual fund for a given characteristic. Figure 13 shows the fit for momentum of the fund shown in the left plot of Panel D in Figure 8.[18] Panel A shows the cumulative contributions of the mode-1 components in the overall fit of the momentum time-series of this fund. I compute the cumulative contributions as follows. Recall that the Tucker model can be written in terms of the weighted sum of the outer products of the columns of $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$, see (15). To compute the cumulative fit of the mode-1 contributions, I calculate the sum over all 25 mode-2 and nine mode-3 components but truncate the mode-1 sum at $\overline{K}_1 = 1,2,...,K_1$ as follows

$$\hat{\boldsymbol{\mathcal{X}}}(\overline{K}_1,25,9) = \sum_{t=1}^{\overline{K}_1} \sum_{m=1}^{25} \sum_{c=1}^{9} g_{tmc} \mathbf{v}_{1t} \circ \mathbf{v}_{2m} \circ \mathbf{v}_{3c}. \tag{29}$$

Panel A shows the fitted time-series for $\overline{K}_1 = 1,5,10$ as well as the observed momentum time-series (dashed orange). The time-series with only a single mode-1 component (displayed in light blue) is essentially constant over the sample and captures the average momentum of the fund of 2.9. The model with five mode-1 components (in dark blue) features a modest variation over the sample that mimics the low-frequency movement of observed momentum. Including all 10 mode-1 components yields a fitted time-series that matches the low and high frequencies of the data closely.

The time-series of the mode-2 and mode-3 components, $\hat{\boldsymbol{\mathcal{X}}}(10,\overline{K}_2,9)$ and $\hat{\boldsymbol{\mathcal{X}}}(10,25,\overline{K}_3)$, respectively, shown in panels B and C are similar. The fitted time-series with single mode-2 and mode-3 components display little variation over the sample and yield a poor fit. The specifications with five components capture the low- and high-frequency movements of the data significantly better. Adding additional components improves the fit further, especially in the peaks and troughs.

3.5. Subsample stability

Next, I investigate the stability of the Tucker decomposition over time. First, I split the sample into two subsamples of equal length: 2010Q3-2014Q4 and 2015Q1 to 2018Q4. Next, I estimate the model of recursive 10-year windows. I compare the estimated core tensors $\boldsymbol{\mathcal{G}}$, loading matrices $\mathbf{V}^{(i)}$ and the fit across subsamples. Since the variance of the subsample data varies, I report the pseudo-R^2, $1-\text{MSE}(\hat{\boldsymbol{\mathcal{X}}})/\text{Var}(\boldsymbol{\mathcal{X}})$ instead of the $\text{MSE}(\boldsymbol{\mathcal{X}})$. Since the subsamples have shorter time spans, I reduce the number of mode-1 components from 10 to five so that $(K_1,K_2,K_3) = (5,15,8)$. In addition to splitting the sample, I also estimate the Tucker decomposition in recursive 10-year subsamples. The results are similar to those in the split sample and are reported in the Appendix.

The overall fits in the split samples are similar to the fit in the complete data set. The pseudo-R^2 $1-\text{MSE}(\hat{\boldsymbol{\mathcal{X}}})/\text{Var}(\boldsymbol{\mathcal{X}})$ for the first subsample is 0.95 and 0.94 for the second subsample and thus almost identical to the fit of the full sample fit (0.94). Panel A of Figure 14 shows the largest 25 elements of the core tensors $\boldsymbol{\mathcal{G}}$ for the whole sample (black), the first half (orange), and the second

[18] The fund is the "Energy Select Sector SPDR" (wficn 500490, ticker XLE) and is chosen for illustration only. However, the results shown in Figure 13 are representative of most funds in the sample.

half (blue) on a log-10 scale. The pattern of core values is remarkably stable across the samples. In all three cases, the (1,1,1) elements of $\boldsymbol{\mathcal{G}}$ are by far the largest core values followed by the (1,2,2) and (1,3,3) elements while the remaining core values are considerably smaller.

Figure 15 plots the first three columns of the three loading matrices in each subsample. The first row shows the mode-1 (time) loading matrices. Note, that the time spans in the split samples do not overlap, while the mutual fund and characteristic indices are identical. Despite the lack of overlap, the estimated loading matrices $\mathbf{V}^{(1)}$ in the two subsamples are similar. The first column vectors of $\mathbf{V}^{(1)}$ are long-only "mean" factors and are almost identical. The second and third factors differ more significantly across subsamples but have the familiar "slope" and "curvature" patterns. The modest instability of higher-order "time" components has only a second-order effect on the fit of the model since the elements of the core tensor associated with higher-order factors are an order of magnitude smaller than the values of the first component. The $\mathbf{V}^{(2)}$ and $\mathbf{V}^{(3)}$ loadings matrices of the second and third modes in rows 2 and 3 are almost identical in both subsamples. This evidence suggests that the Tucker estimation is robust across subsamples.

I conclude that the Tucker model with (8, 15, 8) components yields a good approximation to the mutual fund data. The model captures 94% of the variation of the data while compressing the size of the sample by 97%. The model errors are small for all observations except for a small number of outliers with large errors. The model is stable across the sample and fits well for most mutual funds and characteristics. In the next section, I show that the components of the Tucker model are related to mutual funds returns.

4. Mutual fund returns

As shown above, the components of the Tucker model can be used to construct a small number of factors that summarize the information of the C individual characteristics. In this section, I show how to construct *pricing factors* from the estimated Tucker model that can be used in cross-sectional asset pricing tests. I find that these "Tucker characteristic factors" yield smaller cross-sectional pricing errors than standard benchmark models.

Recall that the core tensor $\boldsymbol{\mathcal{G}}$ in the benchmark Tucker model has dimensions $(K_T \times K_M \times K_C)$ and the loadings matrices associated with the time and mutual fund dimensions, $\mathbf{V}^{(1)}$ and $\mathbf{V}^{(2)}$, have dimensions $(T \times K_T)$ and $(M \times K_M)$, respectively. Multiplying the core tensor with the loading matrices yields a tensor with dimensions $(T \times M \times K_C)$: $\boldsymbol{\mathcal{S}}_{(12)} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{V}^{(1)} \times_2 \mathbf{V}^{(2)}$. For a given mutual fund m, $\mathbf{S}_{tc(m)}$ is a $(T \times K_C)$ matrix (*i.e.*, the m-th slice of $\boldsymbol{\mathcal{S}}_{(12)}$). The $k_3 = 1,...,K_C$ columns of $\mathbf{S}_{tc(m)}$, $\mathbf{s}_{t(mk_3)}$ are time-series of "characteristic factors" that have the same interpretations as factors in 2-dimensional PCA applications. To see this, consider the estimation of a K_C-factor PCA model for the time-series of characteristics of a single mutual fund m given by the $(T \times C)$ matrix $\mathbf{X}_{tc(m)}$. The K_C PCA factors summarize the information across C factors of fund m and form a $(T \times C)$ matrix, which corresponds to the characteristic factors constructed from the Tucker model, $\mathbf{S}_{tc(m)}$. The difference is that PCA factors have to be estimated for each of the M funds separately while the Tucker factors in $\mathbf{S}_{tc(m)}$ are based on a *single* estimation that incorporates information of all M funds simultaneously.

Next, I study how the Tucker characteristic factors are related to the returns of mutual funds. I follow the standard procedure and construct characteristic portfolios. Instead of using individual characteristics, portfolios are based on the K_C Tucker characteristic factors that summarize the information of the C characteristics. 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. I repeat this procedure for $t = 1,...,33$ and each of the K_C Tucker components and obtain $10 \cdot K_C$ time series of portfolio returns. Given the decile portfolios, I form "high-minus-low" pricing factors by subtracting the returns of the decile-1 portfolios from the returns of the decile-10 portfolios yielding K_C Tucker characteristic factors.[19]

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 $t = 1,...,T'$ for $T' = K_T,...,T-1$. For the subsample ending in T', I form decile portfolios in period T' and compute the returns of equally-weighted decile portfolios in $T'+1$. This procedure yields portfolio returns for periods $K_T,...,T$ that are based only on past information and thus not subject to a look-ahead bias. As for in-sample portfolios, I form K_C "high-minus-low" pricing factors. It turns out that the correlations of out-of-sample portfolios with the corresponding in-sample portfolios are above 0.95 for all specifications studied below indicating that the Tucker model is stable over time.

4.1. 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. The K_C decile-10 minus decile-1 portfolios are similar to the standard long-short portfolios typically used in asset pricing tests but are based on characteristic factors implied by the Tucker model instead of characteristics themselves. I compare the performance of the Tucker characteristic factors to several benchmarks, including Fama-French and SVD-PCA factors.

Since all pricing factors are excess returns, I run time-series regressions of excess returns of mutual funds on pricing factors $\mathbf{F}_t$:

$$R^e_{mt} = \alpha_m + \boldsymbol{\beta}_m \, \mathbf{F}_t + e_{mt}, \quad m = 1,...,M. \tag{30}$$

The "market" excess return of the CRSP-VW index is included as the first element of $\mathbf{F}_t$ in all specifications. Let L be the number of factors in $\mathbf{F}_t$. The pricing error of fund m is α_m. I evaluate models by their root-mean-squared pricing error (RMSPE) and the mean-absolute pricing errors (MAPE) across all mutual funds:

$$\text{RMSPE} = \sqrt{\frac{1}{M}\sum_{m=1}^{M} \alpha_m^2}, \quad \text{MAPE} = \frac{1}{M}\sum_{m=1}^{M} |\alpha_m|. \tag{31}$$

[19] Since the signs of Tucker components are not identified, I normalize the pricing factors so that their mean is positive.

I also compute the mean pricing error (MPE) as a measure of the average over or under-performance of mutual funds.[20]

I start by comparing the pricing errors of various Tucker(K_T,K_M,K_C) models. For each model, I compute pricing errors for models with $L = 2,3,4,5$ factors. Recall that the first factor in $\mathbf{F}_t$ is the market excess return. Rather than considering all possible combinations of all K_C Tucker factors, I add factors sequentially, i.e., I start by comparing all models with one Tucker factor in addition to the market excess return and pick the factor with the lowest RMSPE. I repeat this procedure up to a maximum of $L = 5$ factors. The results for the in-sample and out-of-sample factors are reported in Table 5 and 7, respectively. The RMSPEs are in the top panels and the selected factors are in the bottom panels.

The results for the benchmark Tucker model with (8, 15, 8) components show that the pricing error of 2.77% for $L = 3$ factors is significantly lower than for $L = 2$ (3.23%) but adding further factors does not improve the fit. The model for $L = 3$ includes the third and fourth Tucker characteristic factors. Tucker models with additional components do not lower pricing errors but more parsimonious models improve the fit. This is possible since the first (K_T,K_M,K_C) components of a (K'_T,K'_M,K'_C) are not necessarily equal to the components of a (K_T,K_M,K_C) Tucker model. The next three rows show results for Tucker specifications with $K_T = 8,2,1$ and $K_M = K_C = 8$. More parsimonious models yield smaller pricing errors but note that the third and fourth factors are included in all cases, which suggests that these two factors play a particularly important role. Since higher-order factors appear to be less relevant, I next consider models with $K_C = 4$. Comparing several specifications, I find that a three-factor model with the market and the third and fourth factors of the (1,4,4) Tucker model captures most of the information that is relevant for returns. Lowering K_C further yields significantly higher pricing errors. The pricing errors for models with out-of-sample factors in Table 7 are somewhat higher than for in-sample factors but the patterns are similar to the results for in-sample factors. The (2,4,4) Tucker model slightly outperforms the (1,4,4) model. As before, the third and fourth factors are included in the model.

Note that the two preferred models have fewer components that the benchmark model in the previous section with (8,15,8) components. That specification was selected to yield the best fit for characteristics, not returns. The additional components improve the fit for characteristics but are not relevant for returns. This pattern is also typical in SVD-PCA estimations with return data where some factors are relevant for the cross-section of return while other factors are not (see Lettau and Pelger (2020a) and Lettau and Pelger (2020b)).

Next, I compare the Tucker models to other benchmark factor models. The results of the in-sample and out-of-sample portfolios are in Tables 6 and 8, which show which factors are included in the pricing model, the number of factors L, the mean pricing error, MPE, and the RMSPE for each model. Panel A of Table 6 reports results for characteristics factors derived from the Tucker(1,4,4) model with $L = 2,3,4,5$. The RMSPEs of this specification were already reported in Table 5. The mean-absolute pricing error (MAPE) has the same pattern as the RMSE. Note that the MAPE is lower than

[20] All models considered in this paper are formally rejected by the GRS test that tests the null hypothesis that all α_m are jointly zero. I, therefore, do not report the results of this test.

the RMSE, which suggests that some outliers with very large (absolute) pricing errors. The MPE for all cases is negative, which indicates that mutual funds underperform on average after accounting for risk exposure to the Tucker factors.

Panel B of Table 6 reports results for combinations of the market excess return and Fama-French factors SMB, HML, RMW, CMA, and MOM.[21] The RMSPEs range from 3.37% for the CAPM to 2.42% for the 4-factor model including SMB, HML, and MOM. Adding RMW and CMA increases the pricing error. The mean pricing error is negative for all specifications. In comparison to the Tucker factors, however, the Fama-French models perform poorly as even the 2-factor Tucker model has lower errors than the 4-factor Fama-French model.

Of course, the Fama-French factors are constructed from the returns of individual stocks while the test assets are mutual fund returns. I, therefore, construct factors that use the same characteristics but are formed from mutual funds returns. Since the vast majority of funds invest in large stocks (see Lettau et al. (2021)), I do not use size double-sorts and construct univariate (decile) characteristic sorts based on ME, BM, OP, INV, and MOM. I form equal-weighted decile portfolios and obtain "high-minus-low" factors by subtracting the returns of the first decile portfolios from the tenth decile portfolio. The results for these "mutual-fund" Fama-French factors are reported in Panel C. The pricing errors are comparable to those of the original Fama-French factors and larger than the RMSPE of the 2-factor Tucker model.

Next, I consider SVD-PCA models that use information in all 25 characteristics as in Kelly, Pruitt and Su (2019), Pelger (2019), Lettau and Pelger (2020a), Lettau and Pelger (2020b), Giglio and Xiu (2021), among many others. For each characteristic, I form decile portfolios and construct equal-weighted returns. I then estimate SVD-PCA factor models using the $10 \cdot 25 = 250$ time-series of returns as input. In addition, I consider SVD-PCA models that only use 25 "high-minus-low" portfolios which are the difference between the returns of the decile-10 and decile-1 portfolios.

It is important to highlight a fundamental difference between the Tucker vs. SVD-PCA factors. SVD-PCA models are estimated using time-series of mutual fund *returns* and thus exploit the co-movement in the return space. In contrast, the Tucker model is estimated using *only characteristics* of mutual funds and exploits 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.

The results of the PCA specifications with 250 decile and 25 "high-minus-low" portfolios are reported in Panel D and E, respectively. Given the number of factors in the model, L, the pricing errors of the PCA models are lower than those for Fama-French style factors but higher than those for Tucker factors. For example, among all 3-factor models, the Tucker model has the lowest RMSPE of 1.90%, followed by the SVD-PCA model based on 10-1 factors (2.40%) and the SVD-PCA model with all 250 decile portfolios (2.57%). The pricing errors of the specifications with Fama-French factors are 2.65% and 2.84%, respectively. The pattern is similar for $L = 4,5$. Due to the relatively short sample, the table only reports results for models with up to five factors. Ignoring this issue for the moment, increasing the number of SVD-PCA decile factors reduces the RMSPE further, for example, the RMSPE

[21] The data are downloaded from Ken French's website (date: 10/5/2022).

is 2.02 when eight factors are included, which still exceeds the pricing error of the 2-factor Tucker model. In contrast, increasing the number of SVD-PCA high-minus-low factors does not lower the RMSPE further.

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 a single factor of a different type to the specifications in Table 6. Consider, for example, the Tucker model with three factors with an RMSPE of 1.90%. I then add a single factor from one of the other factor types and re-estimate this model with four factors. The added factor is the one with the lowest pricing errors in each group: SMB, ME, the third factor of SVD-PCA for decile portfolios, and the second SVD-PCA factor for 10-1 portfolios. The RMSPEs of these four specifications are reported in the RMSPE* column. Adding any one of these factors does not improve the RMSPE of the three-factor model with only Tucker factors. This is also true for the four and five-factor Tucker models.[22]

The RMSPE* column in the remaining panels reports the root-mean-squared pricing error when the fourth Tucker factor is added to the respective models. Including the Tucker factor lowers the RMSPE in *all* cases. Note that the improvement is often significant. For example, for models with three factors, $L = 3$, the root-mean-squared pricing errors are improved by 15% to 31% when the fourth Tucker factor is added.

The results reported in Tables 5 and 6 suggest that a parsimonious 3-factor model that includes the third and fourth Tucker factors in addition to the market excess return captures the cross-section of mean mutual fund returns. Its pricing errors are smaller than those for models with Fama-French-type factors are well as models with SVD-PCA factors based on the covariance matrix of returns. Adding other factors increases pricing errors. I will refer to this model as the preferred in-sample pricing model.

As mentioned above, the Tucker and PCA-based factors underlying the results in Tables 5 and 6 are constructed from full-sample estimations of the Tucker model and thus subject to a look-ahead bias. Tables 7 and 8 report results when the models are estimated using expanding windows so that the resulting pricing factors are based only on past information and thus not subject to a look-ahead bias. Table 7 compares root-mean-squared pricing errors for Tucker models of different orders. In most cases, pricing errors for out-of-sample factors are only slightly higher than those for in-sample factors reported in Table 5 confirming the previous result that the Tucker model is stable over time. Moreover, the same Tucker factors, three and four, contain the most pricing information in all specifications as in the in-sample case. While the preferred in-sample model included a single time component, the specification with (2, 4, 4) components yields slightly lower pricing errors than the (1, 4, 4) model. Note, that the lowest RMSE of 1.87 in Table 5 is given by the model with (1, 8, 8) components and $L = 6$ factors. However, I choose the more parsimonious specification since the difference is small.

Table 8 compares the results of out-of-sample factors of the Tucker and SVD-PCA models. The results for the out-of-sample Tucker model in Panel A are almost identical to the in-sample results

[22] These results are robust to adding further factors to Tucker models.

in Panel A of Table 6: adding other factors raises the RMSPE in all cases (see the RMSPE* column) and mean pricing errors are negative. In contrast, the results for the out-of-sample SVD-PCA factors are significantly higher than for their in-sample counterparts. For the SVD-PCA based on all decile portfolios, the lowest in-sample RMSPE is 2.17% for $L = 5$ but the corresponding out-of-sample RMSPE is 2.65%, an increase of 22%. The lowest RMSPE for the SVD-PCA factors based on 10-1 "high-minus-low" portfolios, the smallest RMSPE rises from 2.10% to 2.61%, a 24% increase. Hence, the SVD-PCA estimations are less stable over time than the estimation of the Tucker model

I conclude that the characteristic factors derived from the Tucker decomposition successfully price the cross-section of mutual fund returns, 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.

4.2. Properties of Tucker characteristic factors

As in all estimations of latent factor models, it is important to develop an economic intuition of the pricing factors. To this end, I study the properties of the Tucker characteristic factors next. To understand the composition of a factor portfolio, I first compute the average characteristics of the portfolios that underlie the factor. For each decile portfolio and each of the $C = 25$ characteristics in the sample, in each quarter I calculate the mean of the characteristic of all funds in the portfolio. To obtain the "net" characteristics of the "high-minus-low" pricing factors, I take the difference between the characteristics of the decile-10 and decile-1 portfolios. Recall that characteristics are scaled and range from 1 ("low") to 5 ("high"), so that the possible range of "net" characteristics is from -4 to 4.

Figure 16 plot the ME, BM, and MOM characteristics of the four in-sample pricing portfolios, respectively, and Figure 17 shows time-series means of all 25 characteristics of each factor. The results for out-of-sample factors are similar and shown in the appendix (Figures A-10 and A-11). Panel A of Figure 16 shows the results for the first in-sample Tucker factor. ME is negative over the sample implying that ME of the decile-1 portfolio of the first factor is larger than ME of the decile-10 portfolio. Since ME is the average size of stocks in mutual fund portfolios, the first Tucker pricing factor is long in funds that hold smaller stocks and short in funds that hold large stocks. Hence, it is not surprising that this factor is positively correlated with SMB (0.87). BM and MOM of this factor are close to zero and the first factor is BM and MOM-neutral. Note that ME and BM are very stable over the sample and the time-variation of MOM is modest suggesting that the composition of the factor portfolio is stable over time.

ME and BM of the second Tucker factor in Panel B are negative while momentum is positive. Even though ME and BM of the pricing factor appear similar, the characteristics of the underlying decile-1 and 10 portfolios differ significantly. ME of the decile-1 portfolio is close to the maximum value of five throughout the sample, hence the funds in this portfolio invest in very large stocks. ME of the

decile-10 portfolio is close to 3.5, so funds in this portfolio invest in mid-cap stocks. In contrast, the book-to-market ratio of the bottom decile portfolio is around 1.7, which implies that the funds in this portfolio are low-BM growth stocks. The BM ratio of the top decile portfolio is close to the midpoint of 3. Thus the second factor is long in growth funds and therefore negatively correlated with HML (-0.64). Momentum is positive and varies between 0 and 1, thus the mutual funds in the second factor have on average a slight tilt toward high momentum. In contrast to the other factors, ME of the third factor is strongly positive and it, therefore, correlates negatively with SMB. BM is negative and MOM is close to zero. ME, BM, and MOM of the fourth factor are relatively close to zero.

Figure 17 shows bar plots of the time-series means of all 25 characteristics of the four Tucker factors. All price-ratio characteristics of the first factor are negative, so the mutual funds that make up this factor tend to hold growth stocks rather than value stocks. Moreover, the funds included in the first factor are small (negative ME), low volume (negative VOL), and have high bid-ask spreads (positive BIDASK). The price-ratio characteristics of the second Tucker factor are even more negative and thus tilted more strongly towards growth funds. Instead, the characteristics related to fundamental growth rates (EG, ELTG, BG, CFG, SG, and GR) are positive. ME and VOL are negative while INV is positive. Hence, mutual funds in this portfolio tend to invest in smaller stocks with high fundamental growth rates but low price-ratios, and higher than average investment. Mutual funds in the third factor hold very large stocks with high volume (ME and VOL are large) and liquid stocks with low bid-ask spreads (BIDASK is negative). The (net) price ratios are negative but not as small as those for the third factor. Note, that some of the characteristics of the third factor have the opposite pattern of those in the first factor, hence it is not surprising that their correlation is negative (-0.42), though the average return of both factors is positive. The fourth Tucker factor is the only factor where some price ratios are negative while others are positive. For example, the BM is negative and DP is positive, so mutual funds in this factor hold stocks with low book-to-market ratios that pay higher than average dividends. In addition, their stocks are small with low turnover and volume (ME, TURN, and VOL are negative) but are more profitable (positive OP) and have a high quality index.

5. Conclusion

In this paper, I explore tensor-based methods to model high-dimensional data. The Tucker and CP tensor decompositions extend the singular value decomposition and principal components analysis for matrices to more than two dimensions. I show that the decomposition of an n-dimensional tensor implies n two-dimensional factor models and can therefore be interpreted accordingly. The core tensor of the Tucker model forms a compressed version of the data tensor with elements akin to summary observations along each of the data dimensions.

I estimate tensor decompositions using a three-dimensional data set of characteristics of mutual funds across time. I find that over 90% of the variation in the data can be captured by low-order tensor representations that compress the data by over 95%. The components of the estimated tensor models share many features that are typically found in applications of two-dimensional factor models. The first factors of the tensor representation are “long-only” and are related to means, while other factors are “long/short” and capture deviations from means.

The elements of the core tensor of the Tucker decomposition form a compressed version of the data tensor and are "summary" date/mutual fund/characteristic combinations. I show that some date/mutual fund/characteristic combinations capture means while others capture deviations from means. For example, the time-series of the first "summary" mutual fund for a given characteristic is highly correlated with the time-series of cross-sectional means of the characteristic. Higher-order "summary" funds add components beyond cross-sectional means and improve the fit of the model.

References

Abdallah, Emad E., A. Ben Hamza, and Prabir Bhattacharya (2007) "MPEG Video Watermarking Using Tensor Singular Value Decomposition," in Kamel, Mohamed and Aurélio Campilho eds., *Lecture Notes in Computer Science*, 4633, 772–783, Springer, Berlin, Heidelberg, 10.1007/978-3-540-74260-9_69.

Andersen, Anders H. and William S. Rayens (2004) "Structure-Seeking Multilinear Methods for the Analysis of fMRI Data," *NeuroImage*, 22 (2), 728–739, 10.1016/J.NEUROIMAGE.2004.02.026.

Bacciu, Davide and Danilo P. Mandic (2020) "Tensor Decompositions in Deep Learning," *ESANN 2020 - Proceedings, 28th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 441–450, https://arxiv.org/abs/2002.11835v1.

Balasubramaniam, Vimal, John Y. Campbell, Tarun Ramadorai, and Benjamin Ranish (2021) "Who Owns What? A Factor Model for Direct Stock Holding," *NBER Working Paper Series*, 29065, 10.3386/W29065.

Boivin, Jean and Serena Ng (2006) "Are More Data Always Better for Factor Analysis?," *Journal of Econometrics*, 132 (1), 169–194, 10.1016/J.JECONOM.2005.01.027.

Bryzgalova, Svetlana, Martin Lettau, Sven Lerner, and Markus Pelger (2022) "Missing Financial Data."

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, 10.1162/QJEC.2009.124.1.301.

Campbell, John Y. (2006) "Household Finance," *Journal of Finance*, 61 (4), 1553–1604, 10.1111/J.1540-6261.2006.00883.X.

Chamberlain, Gary and Michael Rothschild (1983) "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51 (5), 1281–1304, 10.2307/1912275.

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, 10.1016/0304-405X(86)90027-9.

——— (1988) "Risk and return in an equilibrium APT: Application of a new test methodology," *Journal of Financial Economics*, 21 (2), 255–289, 10.1016/0304-405X(88)90062-1.

De Lathauwer, Lieven, Bart De Moor, and Joos Vandewalle (2000) "On the best rank-1 and rank-(R_1,R_2,...,R_N) approximation of higher-order tensors," *SIAM Journal on Matrix Analysis and Applications*, 21 (4), 1324–1342, 10.1137/S0895479898346995.

Eckart, Carl and Gale Young (1936) "The Approximation of One Matrix by Another of Lower Rank," *Psychometrika*, 1 (3), 211–218, 10.1007/BF02288367.

Favero, Carlo A., Massimiliano Marcellino, and Francesca Neglia (2005) "Principal Components at Work: The Empirical Analysis of Monetary Policy With Large Data Sets," *Journal of Applied Econometrics*, 20 (5), 603–620, 10.1002/JAE.815.

Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin (2000) "The Generalized Dynamic-Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82 (4), 540–554, 10.1162/003465300559037.

Gagliardini, Patrick and Christian Gourieroux (2014) "Efficiency in Large Dynamic Panel Models With Common Factors," *Econometric Theory*, 30 (5), 961–1020, 10.1017/S0266466614000024.

Giglio, Stefano and Dacheng Xiu (2021) "Asset pricing with omitted factors," *Journal of Political Economy*, 129 (7), 1947–1990, https://doi.org/10.1086/714090.

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, 10.2139/ssrn.3032013.

Kolda, Tamara G. and Brett W. Bader (2009) "Tensor decompositions and applications," *SIAM Review*, 51 (3), 455–500, 10.1137/07070111X.

Kroonenberg, Pieter M. (2007) *Applied Multiway Data Analysis*, Wiley Blackwell: Hoboken, NJ, 10.1002/9780470238004.

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, 10.3386/w25381.

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, 10.1093/RFS/HHAA020.

——— (2020b) "Estimating latent asset-pricing factors," *Journal of Econometrics*, 218 (1), 1–31, 10.1016/J.JECONOM.2019.08.012.

Möcks, Joachim (1988) "Topographic Components Model for Event-Related Potentials and Some Biophysical Considerations," *IEEE Transactions on Biomedical Engineering*, 35 (6), 482–484, 10.1109/10.2119.

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, 10.1016/j.jeconom.2018.09.004.

Ricci-Curbastro, Gregorio and Tullio Levi-Civita (1900) "Méthodes De Calcul Différentiel Absolu Et Leurs Applications," *Mathematische Annalen*, 54 (1), 125–201, 10.1007/BF01454201.

Roll, Richard and Stephen A. Ross (1980) "An Empirical Investigation of the Arbitrage Pricing Theory," *The Journal of Finance*, 35 (5), 1073–1103, 10.1111/J.1540-6261.1980.TB02197.X.

Ross, Stephen A. (1976) "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, 13 (3), 341–360, 10.1016/0022-0531(76)90046-6.

Schipper, Katherine and Rex Thompson (1981) "Common Stocks as Hedges Against Shifts in the Consumption or Investment Opportunity Set," *Journal of Business*, 54 (2), 305–328.

Sidiropoulos, Nicholas D., Lieven De Lathauwer, Xiao Fu, Kejun Huang, Evangelos E. Papalexakis, and

Christos Faloutsos (2017) “Tensor Decomposition for Signal Processing and Machine Learning,” *IEEE Transactions on Signal Processing*, 65 (13), 3551–3582, 10.1109/TSP.2017.2690524.

Stock, James H. and Mark W. Watson (2006) “Forecasting with Many Predictors,” in Pesaran, M Hashem and Martin Weale eds., *Handbook of Economic Forecasting, Volume 1*, 1, 515–554, Elsevier: Oxford, United Kingdom.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo (2002) “A survey of weak instruments and weak identification in generalized method of moments,” *Journal of Business and Economic Statistics*, 20 (4), 518–529, 10.1198/073500102288618658.

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, 10.1561/0500000035.

Tucker, Ledyard R. (1966) “Some mathematical notes on three-mode factor analysis,” *Psychometrika*, 31 (3), 279–311, 10.1007/BF02289464.

Vasilescu, M. Alex O. and Demetri Terzopoulos (2002) “Multilinear Analysis of Image Ensembles: TensorFaces,” in Heyden, Anders, Gunnar Sparr, Mads Nielsen, and Peter Johansen eds., *Lecture Notes in Computer Science*, 2350, 447–460, Springer: Berlin, Heidelberg, 10.1007/3-540-47969-4_30.

Tables and figures

Table 1: Mutual fund characteristics

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

Notes: The table lists the mutual fund characteristics used in the paper. See Lettau, Ludvigson and Manoel (2021) for a detailed description of the data.

Table 2: Sample statistics

	Means	Std. Dev.	25% pct.	50% pct.	75% pct.
TNA ($ mil.)	1822.68	4388.98	214.25	633.60	1685.83
No. of stocks	118.10	182.38	55.74	78.88	119.42
Market β	1.01	0.15	0.95	1.03	1.10
Return (% p.a.)	11.61	2.96	10.52	11.81	13.18
CAPM α (% p.a.)	-1.68	2.84	-3.08	-1.76	-0.08
4-factor α (% p.a.)	-0.71	2.32	-1.96	-0.75	0.67

Notes: The table reports summary statistics of the sample. I report statistics of the distribution of means by mutual fund. The sample period is 2010Q3 to 2018Q4.

Table 3: Distributions of errors of the Tucker (10, 25,9) model

Panel A: Moments	
Mean	0.000
Median	0.000
Std. Dev.	0.184
Skew	0.052
Kurt.	3.243
Panel B: Percentiles	
min.	−1.782
0.005	−0.590
0.05	−0.292
0.25	−0.098
0.75	0.097
0.95	0.293
0.995	0.609
max.	1.888

Notes: The table reports summary statistics of the distributions of errors of the Tucker(8, 15, 8) model. The sample period is 2010Q3 to 2018Q4.

Table 4: Tucker core - 15 largest elements by absolute value

k_1	k_2	k_3	Core $\mathbf{c}_{k_1,k_1,k_1}$
1	1	1	1.46
1	2	2	0.18
1	3	3	0.14
1	4	4	0.04
1	6	6	0.02
2	8	2	0.02
1	5	6	0.02
1	7	8	0.02
1	4	5	0.02
1	6	7	0.01
1	11	5	0.01
1	7	7	0.01
1	10	9	0.01
1	9	7	0.01
3	1	4	0.01

Notes: The table lists the 20 largest elements of the core tensor of the Tucker(8, 15, 8) model as well as the corresponding (k_1,k_2,k_3) indices. The sample period is 2010Q3 to 2018Q4.

Table 5: Pricing Errors: Comparison of Tucker Models – In-sample

Tucker model	Number of Factors L			
	2	3	4	5
Panel A: RMSE				
(8, 15, 8)	2.77	1.99	1.97	1.97
(2, 8, 8)	2.37	1.96	1.90	1.89
(1, 8, 8)	2.29	1.90	1.87	1.84
(2, 4, 4)	2.33	1.92	1.93	1.98
(1, 4, 4)	**2.29**	**1.90**	**1.91**	**1.92**
(2, 3, 3)	2.79	2.70	2.83	N/A
(1, 3, 3)	2.79	2.70	2.59	N/A
Panel A: Included Factors				
(8, 15, 8)	(3)	(3, 4)	(3, 4, 5)	(3, 4, 5, 2)
(2, 8, 8)	(4)	(3, 4)	(3, 4, 5)	(3, 4, 5, 2)
(1, 8, 8)	(4)	(3, 4)	(3, 4, 5)	(3, 4, 5, 7)
(2, 4, 4)	(4)	(3, 4)	(3, 4, 2)	(3, 4, 2, 1)
(1, 4, 4)	**(4)**	**(3, 4)**	**(3, 4, 2)**	**(3, 4, 2, 1)**
(2, 3, 3)	(3)	(2, 3)	(2, 3, 1)	N/A
(1, 3, 3)	(3)	(2, 3)	(2, 3, 1)	N/A

Table 6: Pricing Errors: Tucker (1,4,4) – In-sample

Factors	L	MPE	MAPE	RMSPE	RMSPE*
Panel A: Tucker (1,4,4)					
Mkt-RF, 4	2	-0.92	1.85	2.29	2.23, 1.98, 2.12, 2.28
Mkt-RF, 3, 4	3	-0.75	1.49	1.90	2.13, 1.98, 1.90, 1.92
Mkt-RF, 3, 4, 2	4	-0.79	1.50	1.91	2.28, 1.99, 2.22, 1.94
Mkt-RF, 3, 4, 2, 1	5	-0.72	1.46	1.92	2.39, 2.05, 2.29, 1.97
Panel B: Fama-French Factors					
Mkt-RF, SMB, HML	3	-0.89	1.90	2.65	2.19
Mkt-RF, SMB, HML, MOM	4	-0.79	1.78	2.41	2.16
Mkt-RF, SMB, HML, RMW, CMA	5	-0.69	1.93	2.68	2.13
Mkt-RF, SMB, HML, RMW, CMA, MOM	6	-0.58	1.83	2.47	2.14
Panel C: Mutual Fund Fama-French Factors					
Mkt-RF, ME, BM	3	-1.55	2.11	2.84	1.94
Mkt-RF, ME, BM, MOM	4	-1.61	2.18	2.96	2.03
Mkt-RF, ME, BM, OP, INV	5	-1.05	1.75	2.22	2.03
Mkt-RF, ME, BM, OP, INV, MOM	6	-1.07	1.89	2.43	2.21
Panel D: SVD-PCA Deciles					
Mkt-RF, 3	2	-1.09	2.03	2.80	2.14
Mkt-RF, 3, 6	3	-0.91	1.86	2.57	2.06
Mkt-RF, 3, 6, 5	4	-0.84	1.81	2.36	2.07
Mkt-RF, 3, 6, 5, 2	5	-0.96	1.64	2.17	1.87
Panel E: SVD-PCA 10-1					
Mkt-RF, 2	2	-1.03	1.88	2.63	2.04
Mkt-RF, 2, 5	3	-0.94	1.65	2.40	2.03
Mkt-RF, 2, 5, 4	4	-0.87	1.56	2.10	2.03
Mkt-RF, 2, 5, 4, 3	5	-0.89	1.57	2.12	2.06

The sample period is 2010Q3 to 2018Q4.

Table 7: Pricing Errors: Comparison of Tucker Models – Out-of-sample

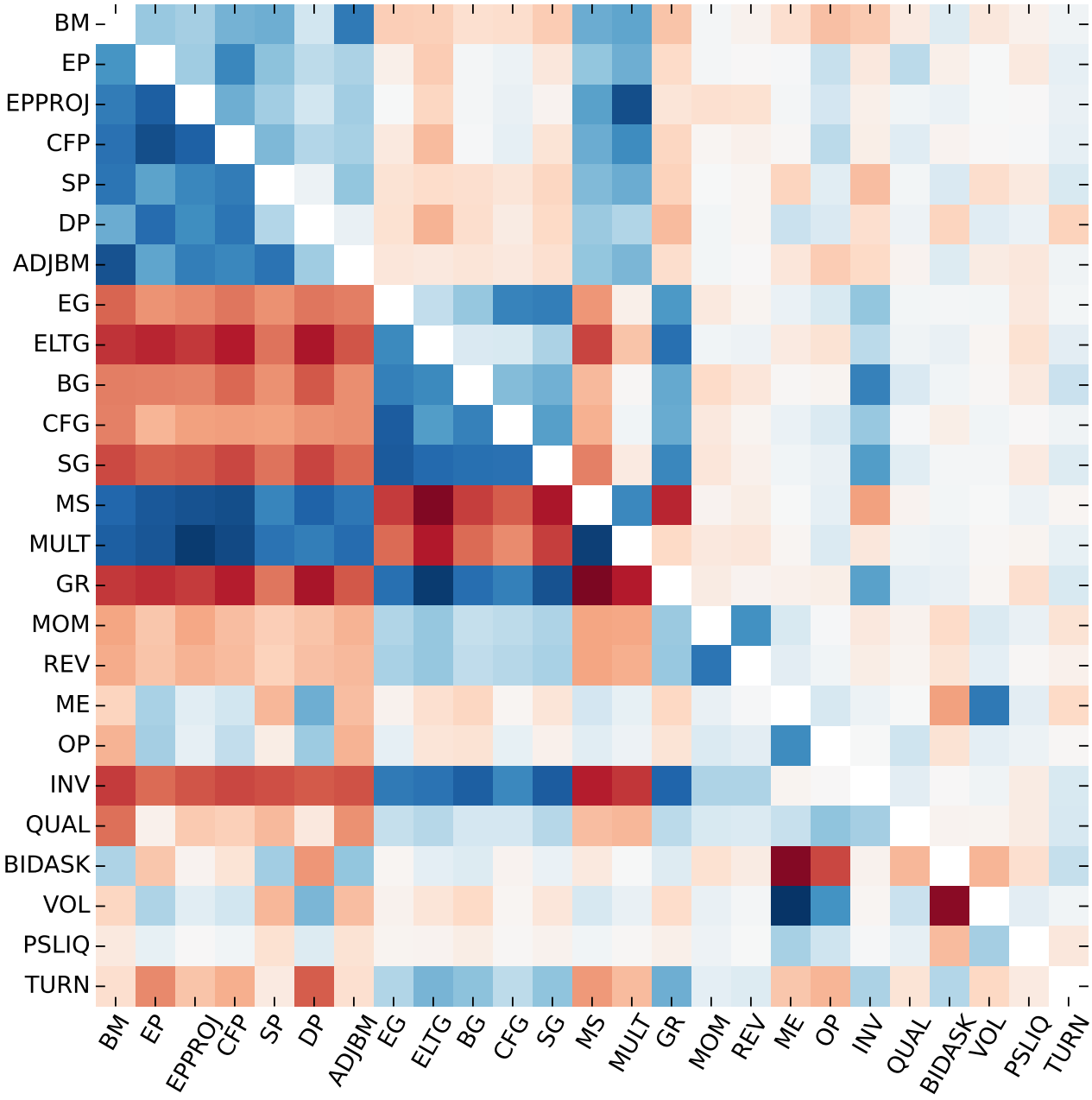
Tucker model	Number of Factors L			
	2	3	4	5
Panel A: RMSE				
(8, 15, 8)	2.81	2.11	2.11	2.11
(2, 8, 8)	2.30	1.97	1.96	1.92
(1, 8, 8)	2.31	1.90	1.89	1.87
(2, 4, 4)	**2.34**	**1.88**	**1.88**	**1.94**
(1, 4, 4)	2.31	1.90	1.90	1.90
(2, 3, 3)	2.83	2.70	2.24	N/A
(1, 3, 3)	2.82	2.70	2.25	N/A
Panel A: Included Factors				
(8, 15, 8)	(4)	(3, 4)	(3, 4, 5)	(3, 4, 5, 2)
(2, 8, 8)	(4)	(3, 4)	(3, 4, 8)	(3, 4, 8, 2)
(1, 8, 8)	(4)	(3, 4)	(3, 4, 6)	(3, 4, 6, 2)
(2, 4, 4)	(4)	(3, 4)	(3, 4, 2)	(3, 4, 2, 1)
(1, 4, 4)	**(4)**	**(3, 4)**	**(3, 4, 2)**	**(3, 4, 2, 1)**
(2, 3, 3)	(3)	(2, 3)	(2, 3, 1)	N/A
(1, 3, 3)	(3)	(2, 3)	(2, 3, 1)	N/A

Table 8: Linear Factor Models: Tucker (2,4,4) – Out-of–sample

Factors	L	MPE	MAPE	RMSPE	RMSPE*
Panel A: Tucker (2,4,4)					
Mkt-RF, 4	2	-0.94	1.88	2.34	2.28, 2.01, 2.02, 2.18
Mkt-RF, 3, 4	3	-0.76	1.47	1.88	2.15, 1.94, 2.02, 1.90
Mkt-RF, 3, 4, 2	4	-0.79	1.47	1.88	2.20, 1.97, 2.06, 1.90
Mkt-RF, 3, 4, 2, 1	5	-0.63	1.47	1.94	2.34, 2.04, 2.03, 1.98
Panel C: SVD-PCA Deciles					
Mkt-RF, 3	2	-1.19	2.23	3.00	2.58
Mkt-RF, 3, 4	3	-1.11	2.12	2.80	2.57
Mkt-RF, 3, 4, 5	4	-1.06	2.13	2.74	2.52
Mkt-RF, 3, 4, 5, 2	5	-1.37	2.09	2.65	2.39
Panel D: SVD-PCA 10-1					
Mkt-RF, 2	2	-1.23	2.12	2.83	2.46
Mkt-RF, 2, 4	3	-1.17	2.07	2.65	2.46
Mkt-RF, 2, 4, 1	4	-1.26	2.06	2.63	2.40
Mkt-RF, 2, 4, 1, 6	5	-1.23	2.03	2.61	2.40

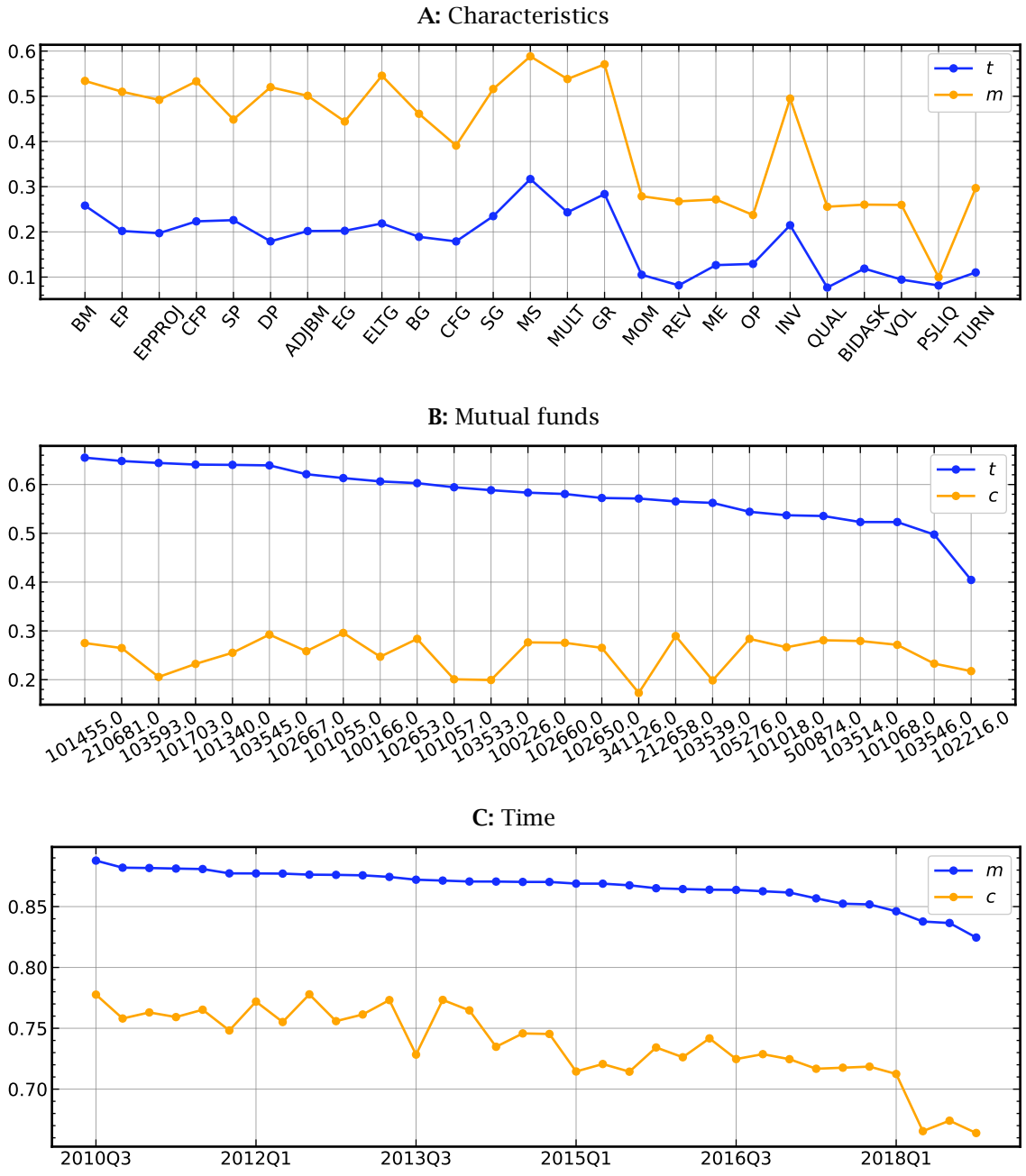
The sample period is 2010Q3 to 2018Q4.

Figure 1: Cross-correlations of characteristics



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

Figure 2: 2-dimensional correlations



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

Figure 3: Tucker decomposition $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \times_3 \mathbf{V}_3$

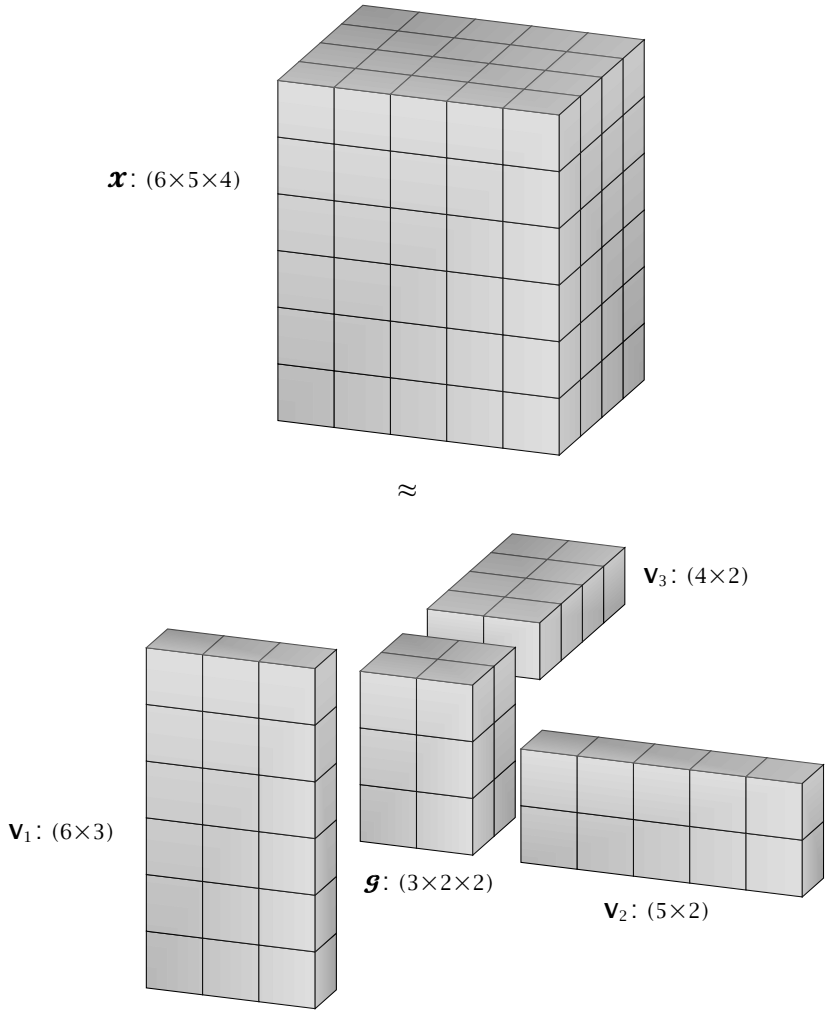
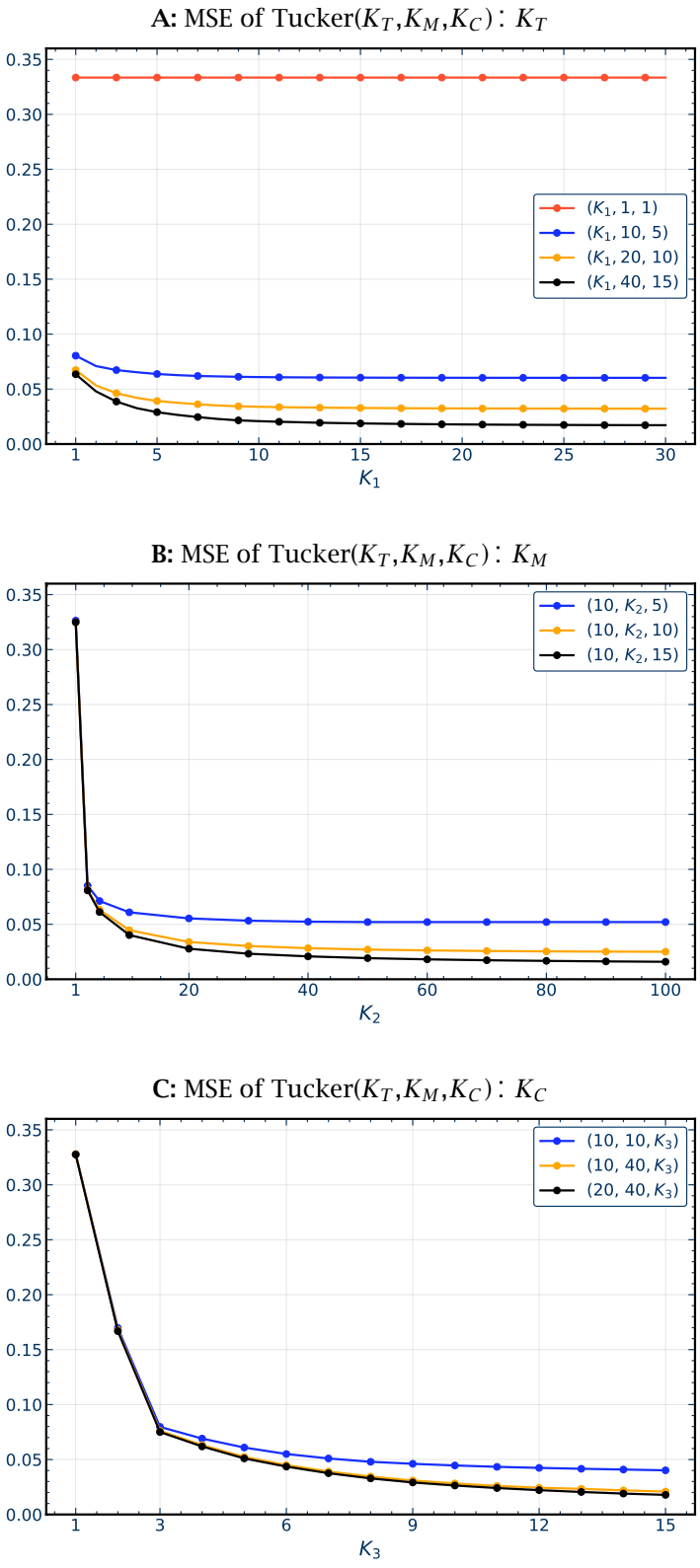
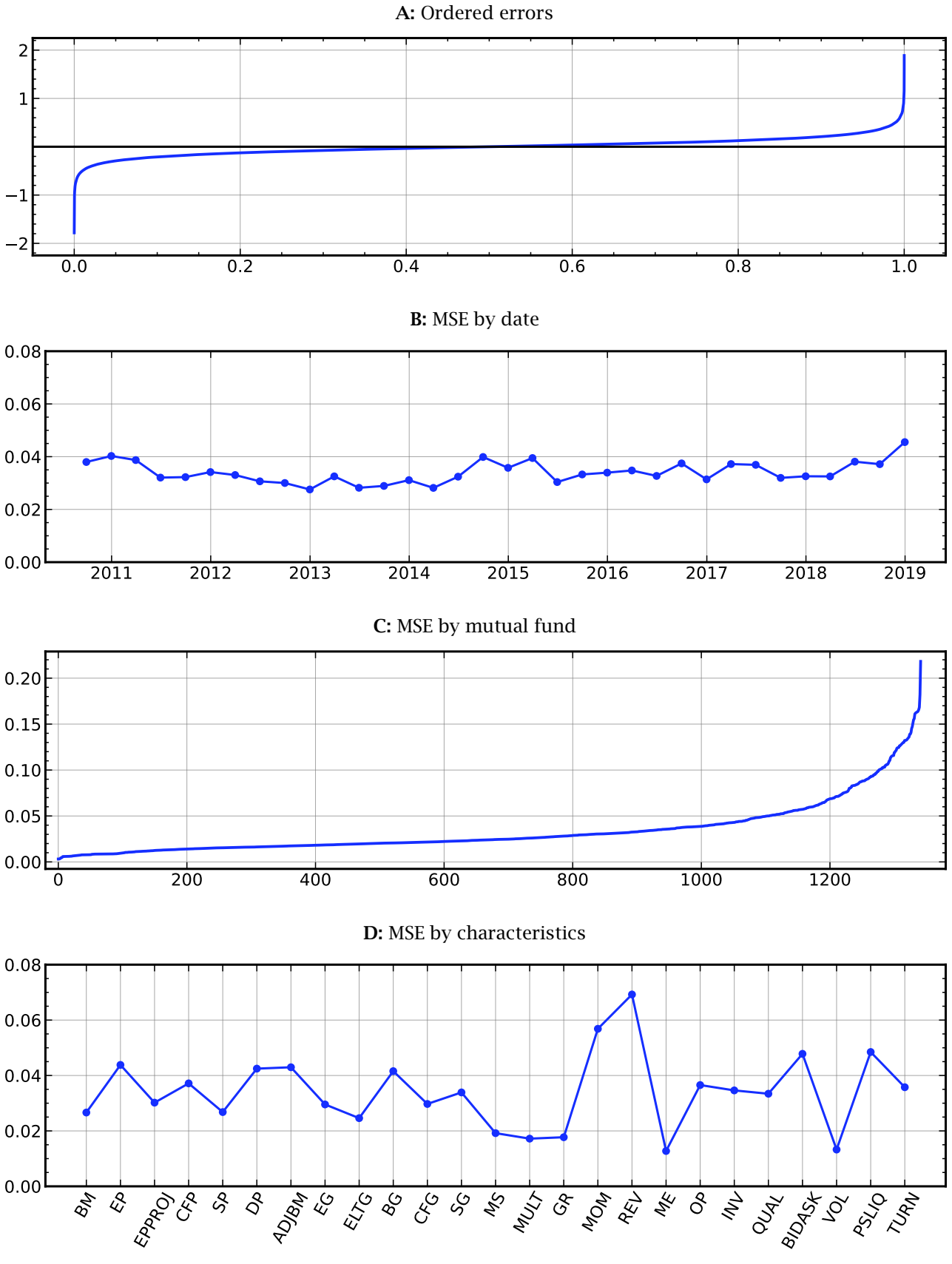


Figure 4: MSE of Tucker model



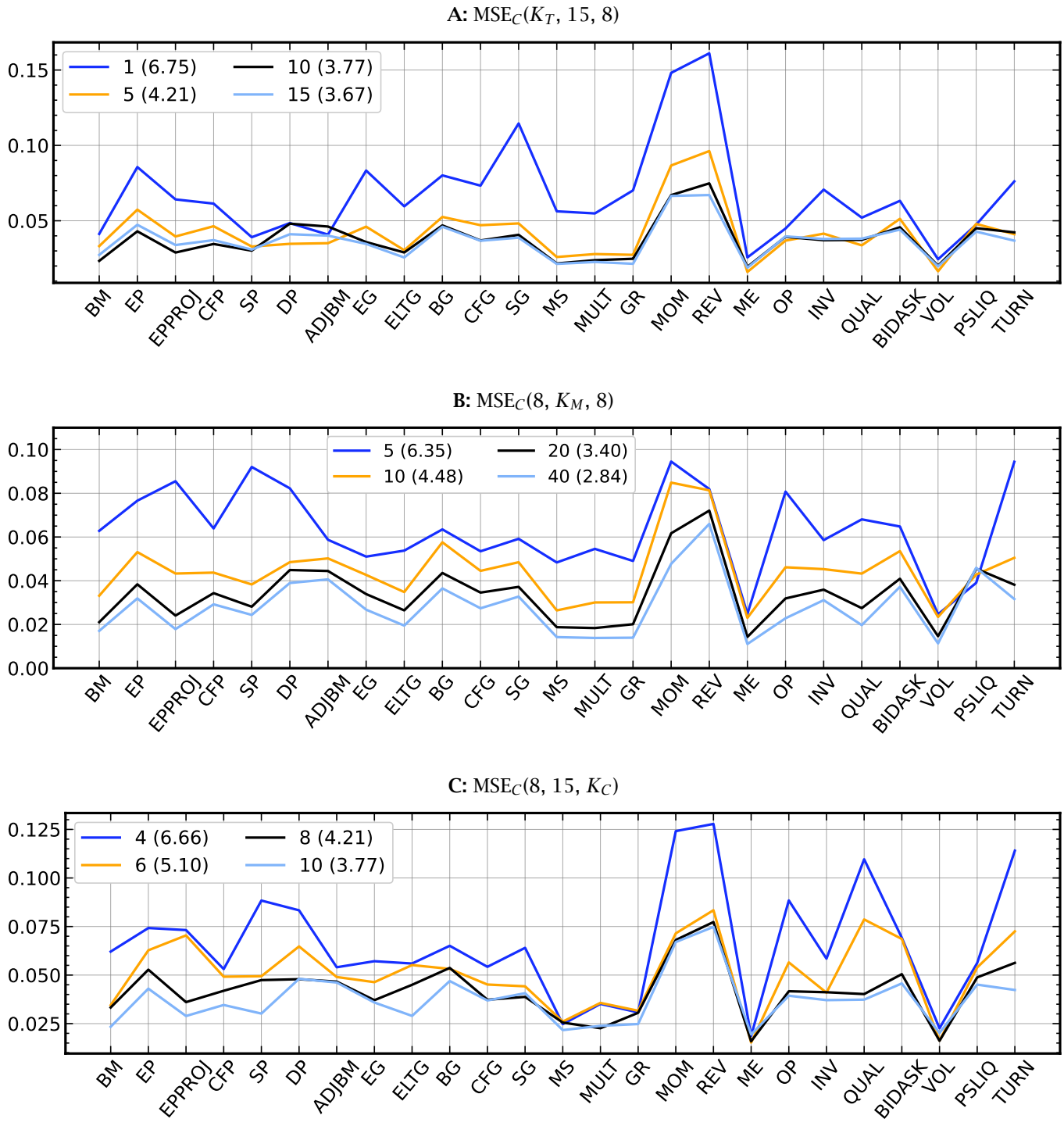
Notes: The figure plots mean-squared errors of the Tucker(K_T,K_M,K_C) model as a function of K_T (Panel A), K_M (Panel B), and K_C (Panel C). The sample period is 2010Q3 to 2018Q4.

Figure 5: Errors of Tucker(10, 25, 9) and CP(12) models



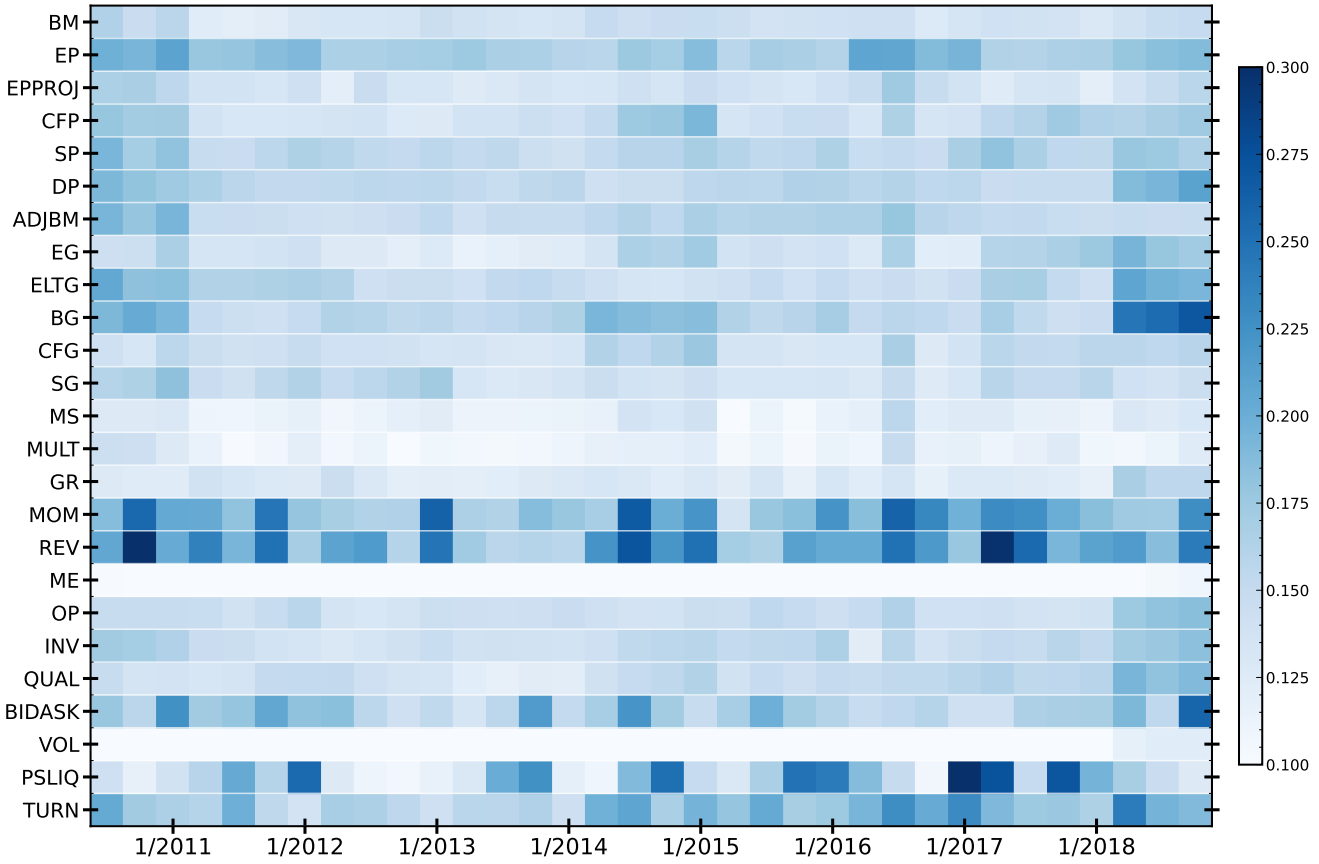
Notes: Panel A plots the sorted errors of the Tucker(10, 25, 9) and CP(12) models. Panels B, C, and D shows the means of MSE by dates, mutual funds, and characteristics, respectively. The sample period is 2010Q3 to 2018Q4.

Figure 6: MSE by characteristic



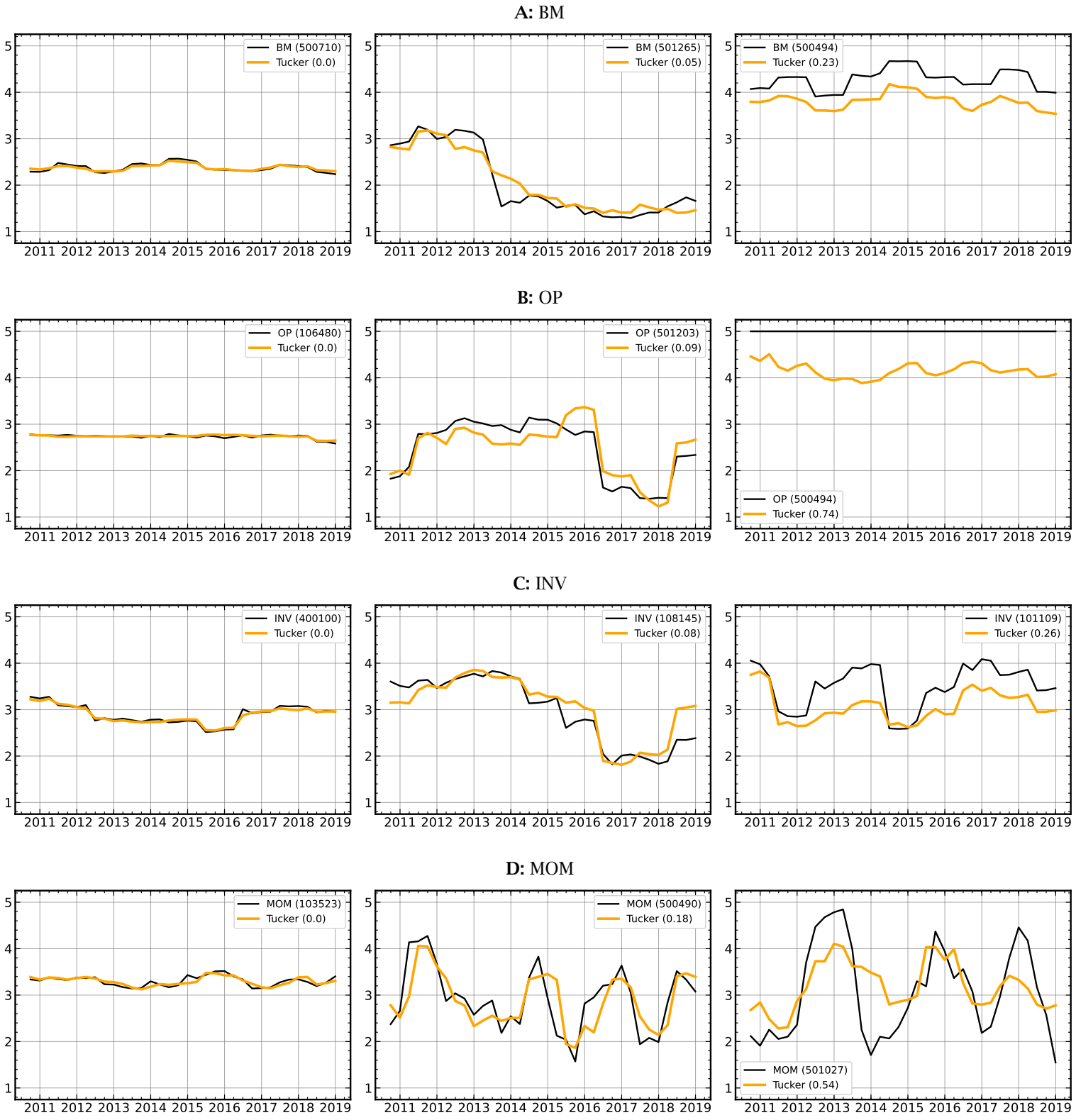
Notes: The figure plots mean-squared errors of the Tucker(K_T,K_M,K_C) model as a function of K_T (Panel A), K_M (Panel B), and K_C (Panel C). The sample period is 2010Q3 to 2018Q4.

Figure 7: Errors by data/characteristic



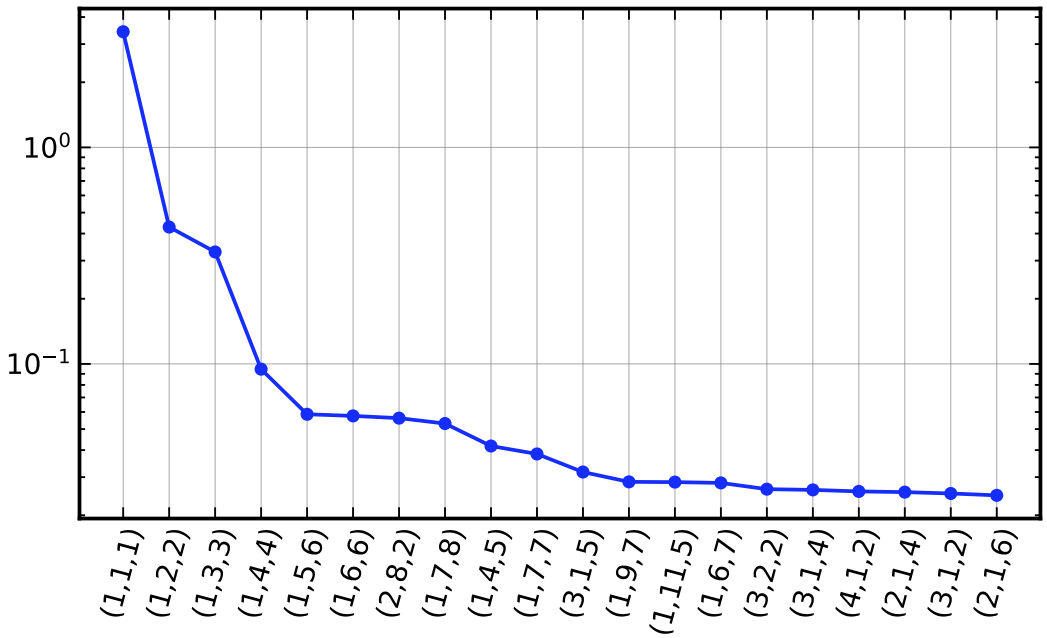
Notes: The figure shows the heatmap of means across mutual funds of the MSE for each date and characteristic combination. The sample period is 2010Q3 to 2018Q4.

Figure 8: Fit of Tucker(8, 15, 8) model - Examples



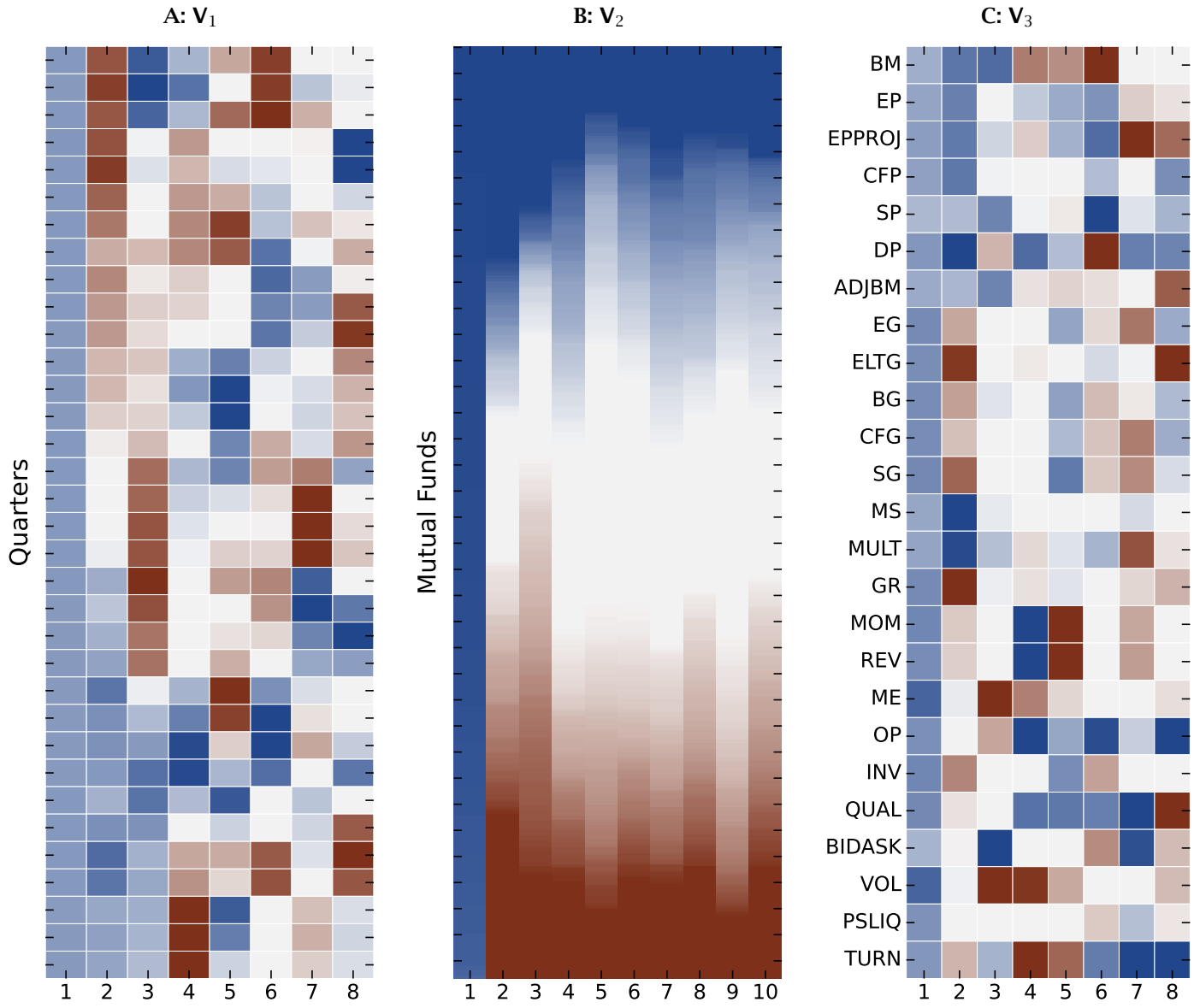
Notes: The figures show time-series plots of the observed data and fitted values of the Tucker(8, 15, 8) model of the book-to-market ratio (Panel A), operating profitability (Panel B), investment (Panel C), and momentum (Panel D) of individual mutual funds. The legends include the wficn of the plotted fund and the MSE of the Tucker(9, 25, 10) model. The sample period is 2010Q3 to 2018Q4.

Figure 9: Tucker(8, 15, 8) core tensor



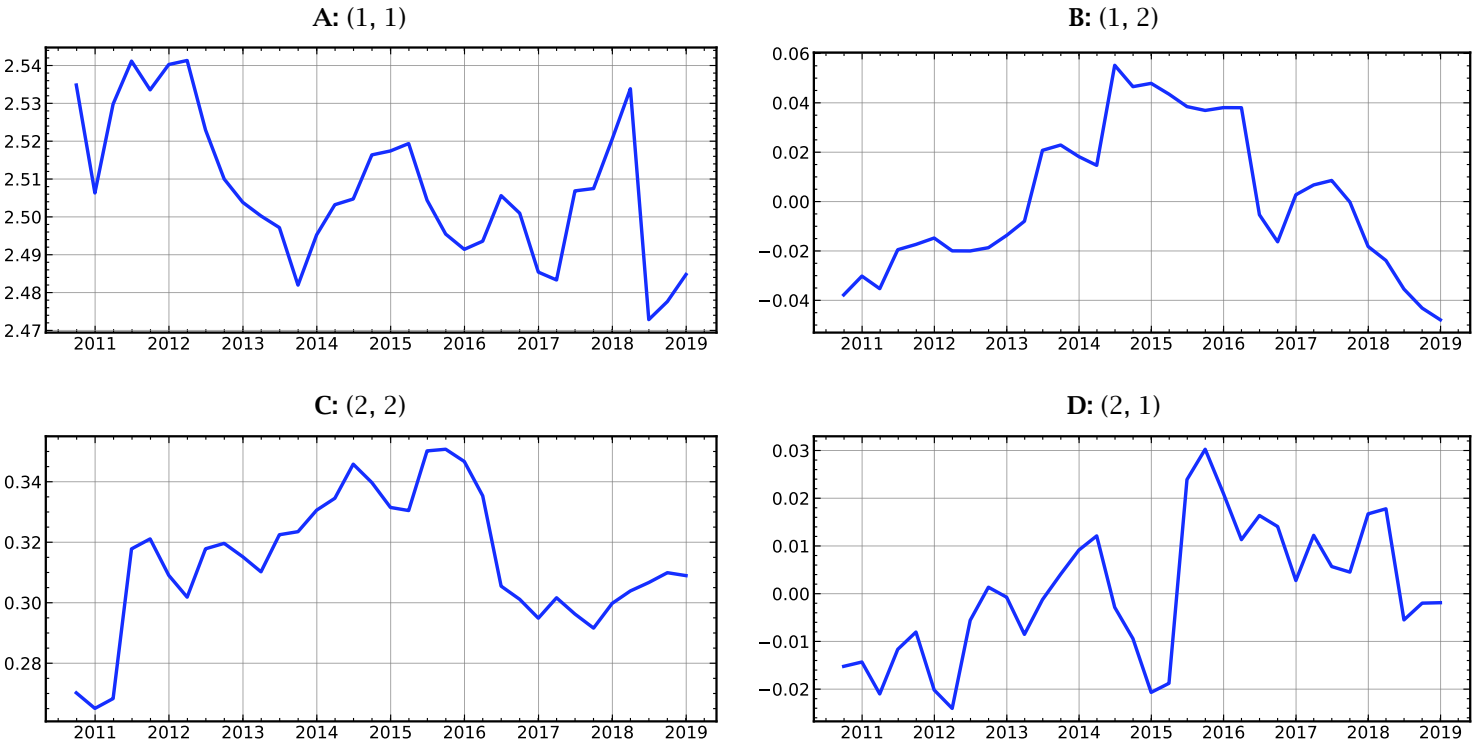
Notes: The figure plots the 20 largest elements by the absolute value of the 3-dimensional core tensor $\mathcal{G}$ of the Tucker(8, 15, 8) model on a log scale. The x-axis shows the indices of $\mathcal{G}$. The sample period is 2010Q3 to 2018Q4.

Figure 10: Component matrices of Tucker(8, 15, 8) models



Notes: The figure shows heatmaps of the Tucker component matrices $\mathbf{V}_i$ of the Tucker(8, 15, 8) model. Positive values are shown in blue and negative values in red. Panel A shows the (34×8) component matrices $\mathbf{V}_1$, Each row corresponds to a quarter starting in 201Q3 at the top to 2018Q4 at the bottom. The columns correspond to the $K_T = 8$ mode-1 components. The second component matrix $\mathbf{V}_2$ has 980 rows and 15 columns. Panel B shows the heatmap of the first 10 columns of $\mathbf{V}_2$. To visualize 980 rows, I sort each column of $\mathbf{V}_2$so that the first row of each column plots the funds with the largest values at the top and the funds with the smallest values at the bottom. Panel C shows the heatmap of the (25×8)-dimensional matrix $\mathbf{V}_3$. The sample period is 2010Q3 to 2018Q4.

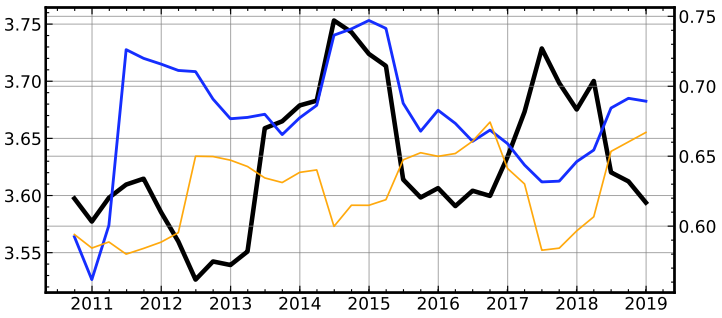
Figure 11: Summary funds/characteristics of Tucker(8, 15, 8) model



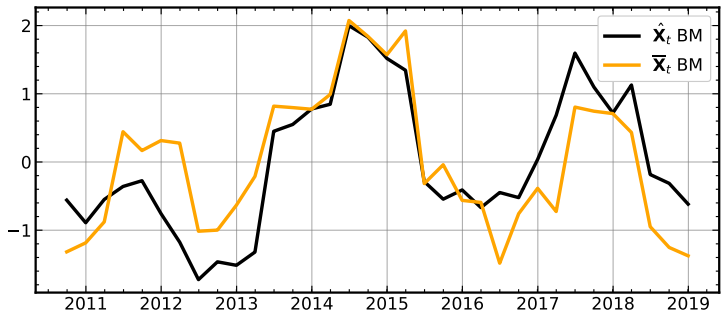
Notes: The figures shows time-series plots of the mode-1 fibers of the (34, 25, 9)-dimensional tensor $\boldsymbol{\mathcal{S}}_{(1)} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{V}_1$. The (k_2, k_3) panels plot the 34 time-series observations of the k_2-th mode-2 and k_3-th mode-3 fibers. The sample period is 2010Q3 to 2018Q4.

Figure 12: Characteristics of summary funds in Tucker(8, 15, 8) model

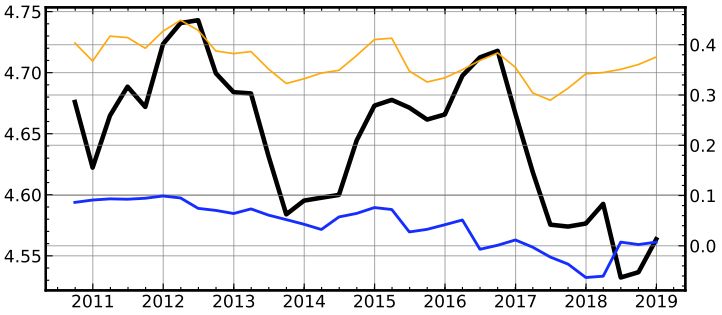
A: Repr. funds 1 to 3 - BM



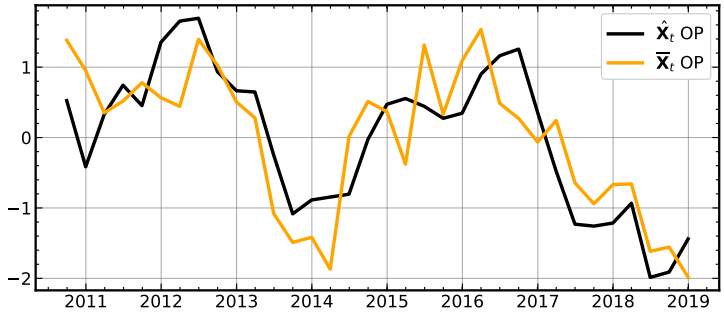
B: Repr. fund 1 vs. mean - BM



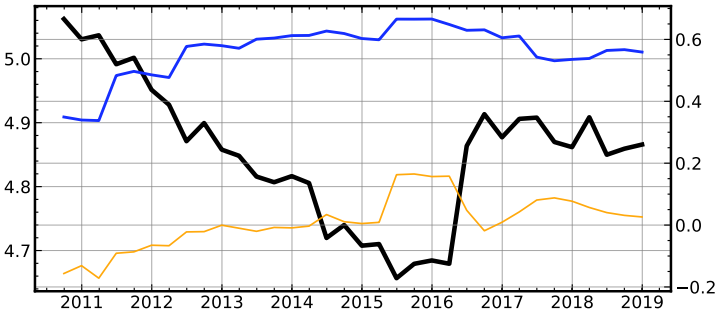
C: Repr. funds 1 to 3 - OP



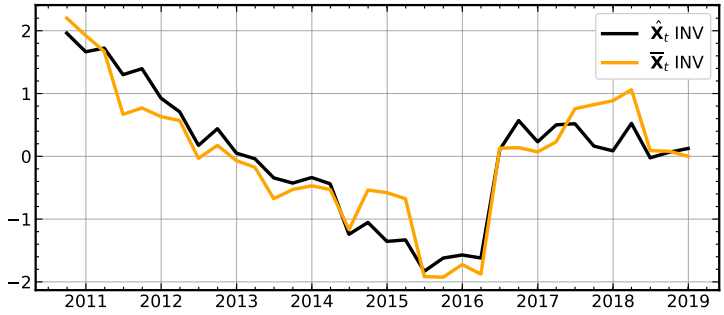
D: Repr. fund 1 vs. mean - OP



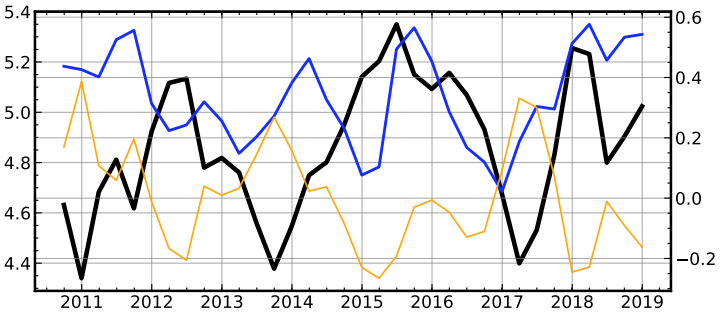
E: Repr. funds 1 to 3 - INV



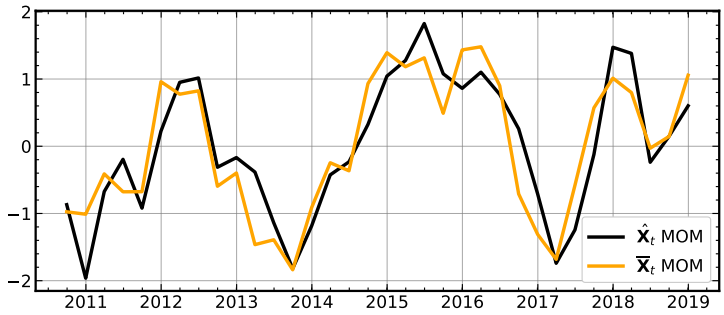
F: Repr. fund 1 vs. mean - INV



G: Repr. funds 1 to 3 - MOM

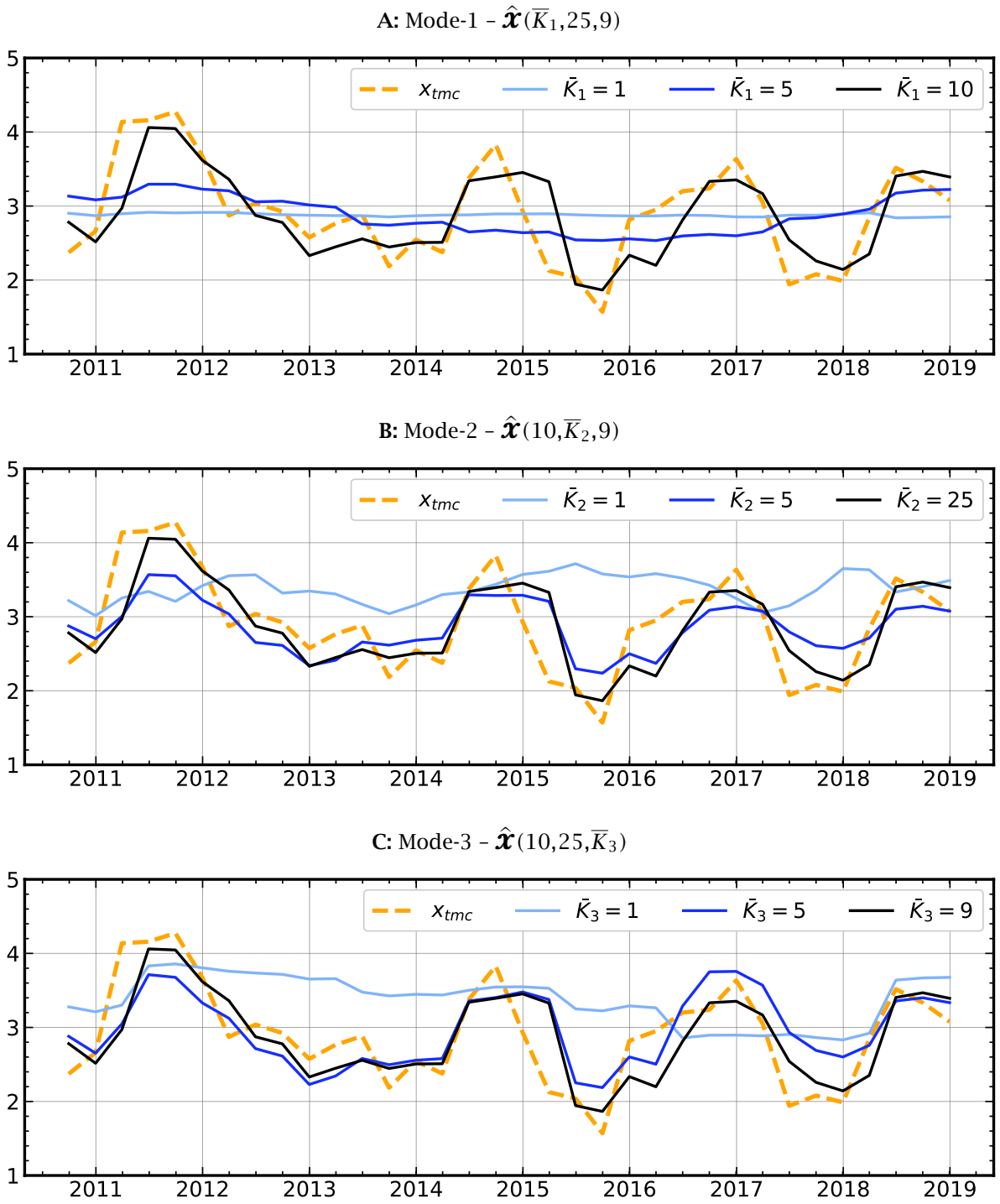


H: Repr. fund 1 vs. mean - MOM



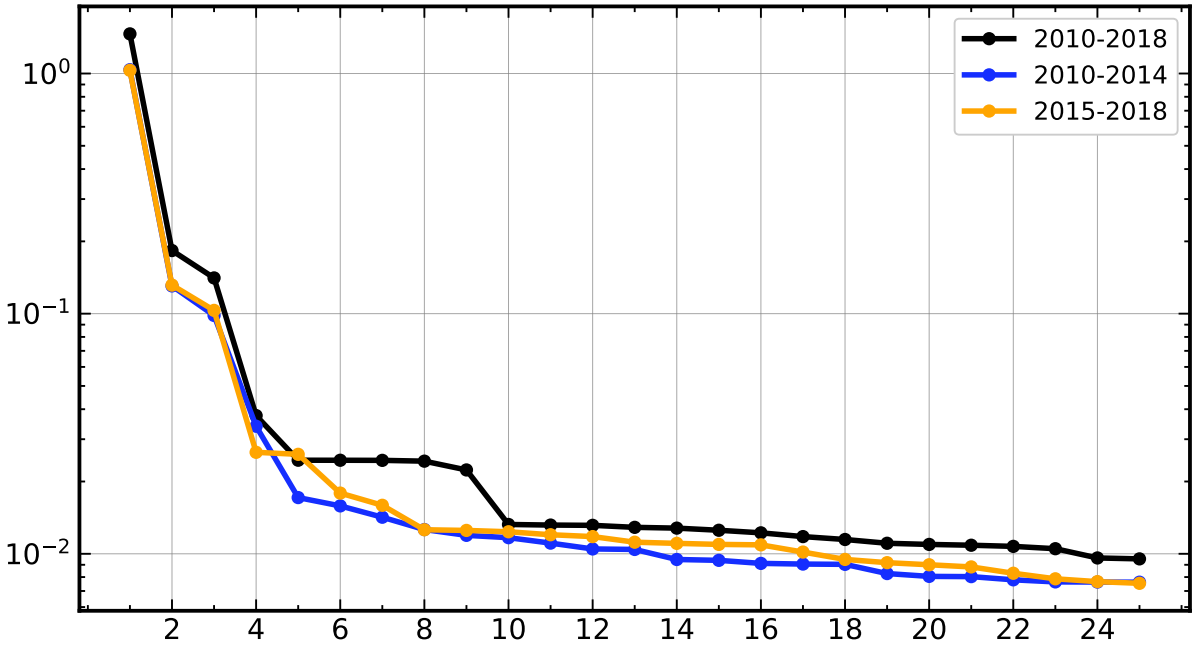
Notes: The figures shows time-series plots of the fibers of the (34, 25, 25)-dimensional tensor $\boldsymbol{\mathcal{S}}_{(13)} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{V}_1 \times_3 \mathbf{V}_3$ for book-to-market ratio (row 1), operating profitability (row 2), investment (row 3), and momentum (row 4). The left column shows the first three fibers where the first fiber is plotted on the left scale and the second and third fibers on the right scale. The right column shows the standardized plots of the first fiber and the time-series of cross-sectional means across characteristics. The sample period is 2010Q3 to 2018Q4.

Figure 13: Contributions of components in Tucker(8, 15, 8) model



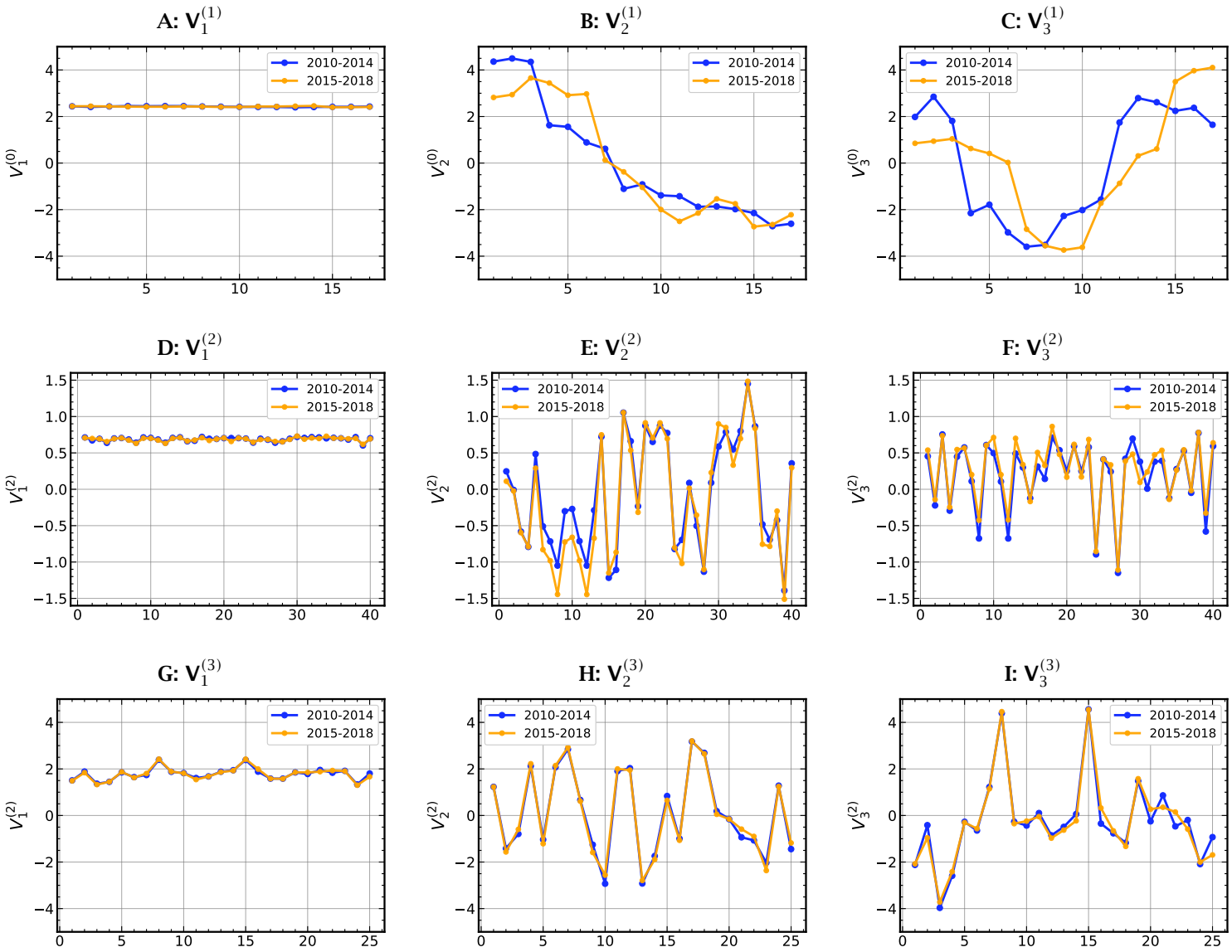
Notes: This figure shows cumulative components for momentum of a single mutual fund (wficn 500490) for the Tucker(8, 15, 8) model. The top panel plots the fitted time-series of Tucker($\overline{K}_1$,25,9) models defined in (29), *i.e.*, fixing $K_M = 25$ and $K_C = 9$ and varying the number of mode-1 components $\overline{K}_1 = 1,5,10$. The observed data is plotted in orange. Panels B and C show the plots for Tucker($10,\overline{K}_2,9$), $\overline{K}_2 = 1,5,25$ and Tucker($10,25,\overline{K}_3$), $\overline{K}_3 = 1,5,9$ models, respectively. The sample period is 2010Q3 to 2018Q4.

Figure 14: Subsamples - Core tensors



Notes: This figure plots the first three loading vectors of mode i,$\mathbf{v}_j^{(i)}$,$i = 1,2,3$,$j = 1,2,3$ of Tucker models that are estimated over two subsamples consisting of 17 quarters: 2010Q3-2014Q4 and 2015Q1-2018Q4. The Tucker models have $(K_T,K_M,K_C) = (7,25,9)$ components. The blue lines correspond to the first half of the sample and the orange lines to the second half.

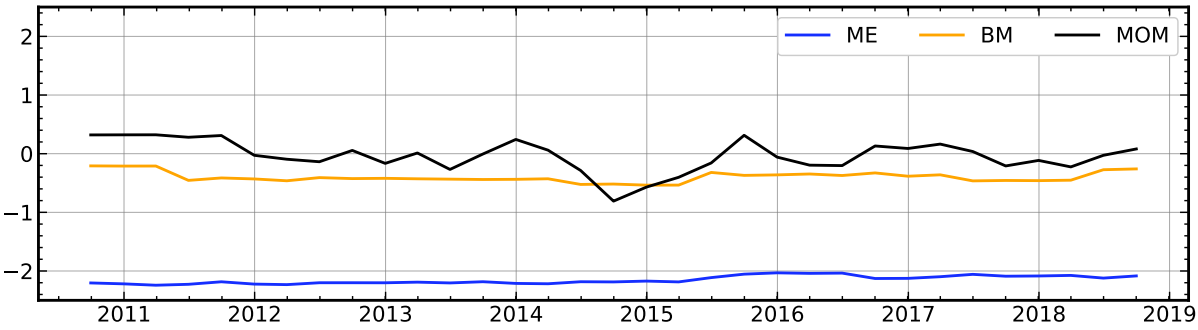
Figure 15: Subsamples - Loading matrices



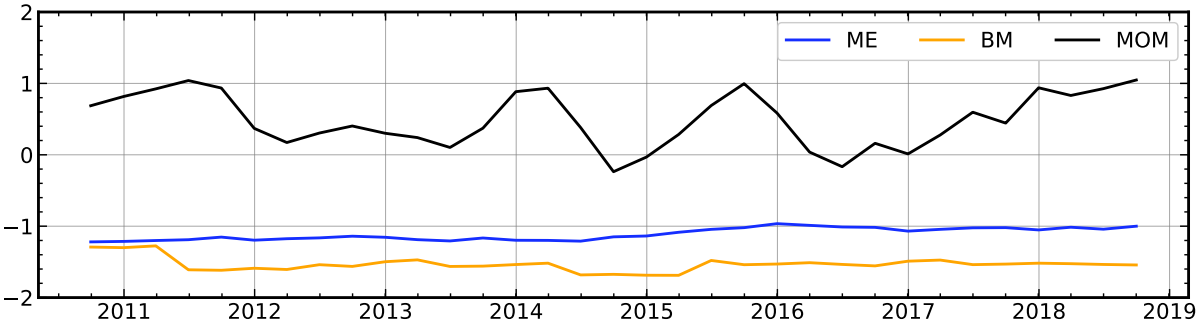
Notes: This figure plots the first three loading vectors of mode i,$\mathbf{V}_j^{(i)}$,$i = 1,2,3$,$j = 1,2,3$ of Tucker models that are estimated over two subsamples consisting of 17 quarters: 2010Q3-2014Q4 and 2015Q1-2018Q4. The Tucker models have $(K_T,K_M,K_C) = (7,25,9)$ components. The blue lines correspond to the first half of the sample and the orange lines to the second half.

Figure 16: Factor Characteristics over Time - In-sample Factors

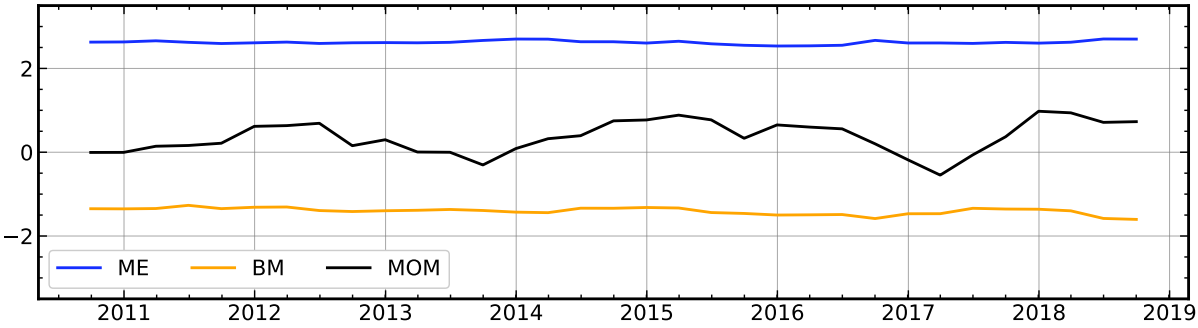
A: Factor 1



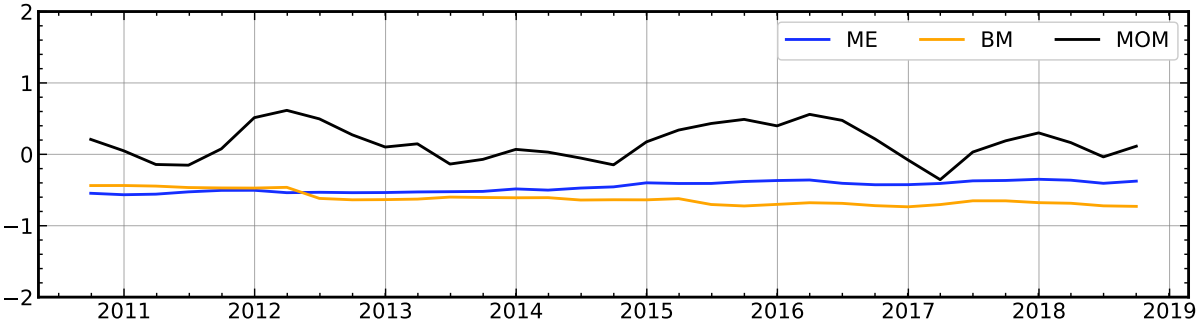
B: Factor 2



C: Factor 3

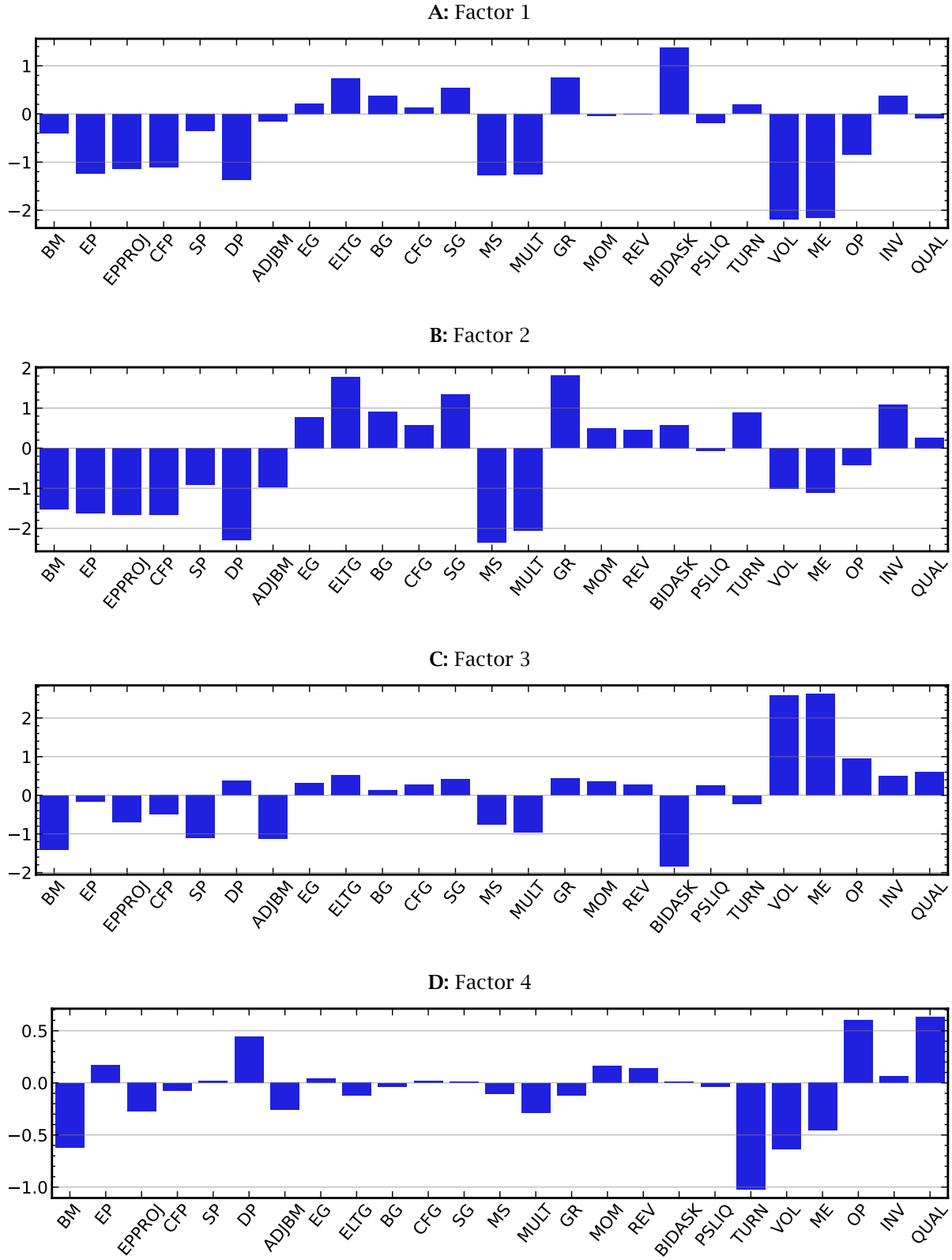


D: Factor 4



Notes: This figure shows the characteristics of the Tucker characteristic factors.

Figure 17: Mean Characteristics of Tucker Factors



Notes: This figure shows the characteristics of the Tucker characteristic factors.

Table A-1: Summary of tensor notation and operations

Operation	two-dimensional matrix	three-dimensional tensor	n-dimensional tensor
	$\mathbf{X} = [\mathsf{x}_{ij}]$	$\boldsymbol{\mathcal{X}} = [\mathsf{x}_{ijk}]$	$\boldsymbol{\mathcal{X}} = [\mathsf{x}_{i_1,\ldots,i_n}]$
norm	$\|\mathbf{X}\| = \sqrt{\sum_{ij} \mathsf{x}_{ij}^2}$	$\|\boldsymbol{\mathcal{X}}\| = \sqrt{\sum_{ijk} \mathsf{x}_{ijk}^2}$	$\|\boldsymbol{\mathcal{X}}\| = \sqrt{\sum_{i_1,\ldots,i_N} \mathsf{x}_{i_1,\ldots,i_n}^2}$
matricization		$\mathbf{X}_{(1)}$: $T \times M \times C \to T \times MC$ $\mathbf{X}_{(2)}$: $T \times M \times C \to M \times TC$ $\mathbf{X}_{(3)}$: $T \times M \times C \to C \times TM$	$\mathbf{X}_{(p)}$: $(I_1 \times \cdots \times I_n) \to I_p \times (\prod_{i \neq p} I_i)$
inner product	$\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{ij} \mathsf{x}_{ij} \mathsf{y}_{ij}$	$\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \rangle = \sum_{i,j,k} \mathsf{x}_{ijk} \mathsf{y}_{ijk}$	$\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \rangle = \sum_{i_1,\ldots,i_n} \mathsf{x}_{i_1,\ldots,i_n} \mathsf{y}_{i_1,\ldots,i_n}$
outer product	$\mathsf{x}\mathsf{y}^\top$	$\mathsf{x} \circ \mathsf{y} \circ \mathsf{z}$	$\mathsf{x}_1 \circ \cdots \circ \mathsf{x}_n$
matrix multiplication	$\mathbf{A}_1 \mathbf{X}$ $\mathbf{X}\mathbf{A}_2^\top$	$\boldsymbol{\mathcal{X}} \times_1 \mathbf{A}_1$ $\boldsymbol{\mathcal{X}} \times_2 \mathbf{A}_2$	$\boldsymbol{\mathcal{X}} \times_1 \mathbf{A}_1 \times_2 \cdots \times_n \mathbf{A}_n$
vector multiplication	$\mathbf{a}_1^\top \mathbf{X}$ $\mathbf{X}\mathbf{a}_2$	$\mathbf{X} \times_1 \mathbf{a}_1^\top$ $\mathbf{X} \times_2 \mathbf{a}_2^\top$	$\boldsymbol{\mathcal{X}} \times_1 \mathbf{a}_1^\top \times_2 \cdots \times_n \mathbf{a}_n^\top$
decompositions	$\mathbf{U}_{1K} \mathbf{H}_K \mathbf{U}_{2K}^\top$ $\sum_{k=1}^K h_k \, \mathbf{u}_{1k} \mathbf{u}_{2k}^\top$	$\boldsymbol{\mathcal{G}} \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \times_3 \mathbf{V}_3$ $\sum_{k=1}^K g_k \mathbf{w}_{1k} \circ \mathbf{w}_{2k} \circ \mathbf{w}_{3k}$	$\boldsymbol{\mathcal{G}} \times_1 \mathbf{V}_1 \times_2 \cdots \times_n \mathbf{V}_n$ $\sum_{k=1}^K g_k \mathbf{w}_{1k} \circ \cdots \circ \mathbf{w}_{nk}$

Table A-2: Mutual Fund Returns: Fit of Linear Factor Models

K	1	2	3	4	5
			Panel A: Full Sample		
2	2.62%	3.57%	3.94%	3.94%	4.07%
3	2.53%	2.55%	2.58%	2.62%	2.63%
4	2.33%	2.44%	2.49%	2.52%	2.54%
5	2.27%	2.28%	2.30%	2.34%	2.37%
6	2.19%	2.25%	2.26%	2.28%	2.29%
			Panel B: Recursive Sample		
2	3.00%	3.51%	3.74%	3.94%	4.04%
3	2.33%	2.59%	2.69%	2.91%	3.01%
4	2.34%	2.47%	2.48%	2.52%	2.54%
5	2.45%	2.49%	2.51%	2.53%	2.54%
6	2.52%	2.55%	2.56%	2.56%	2.56%

Notes: This table shows the cross-sectional fit of linear asset pricing models for mean returns of 1,378 mutual funds. The cross-sectional regression does not include a constant, so residuals represent pricing errors. ME is the mean pricing error, and MAE and RMSE are the mean-absolute and root-mean-square pricing errors. MSE and MSR are the mean-squared errors and excess mutual funds returns, so 1-MSE/MSR is a summary measure of model fit similar to the R^2. K is the number of factors of the linear pricing model. The first four rows report results for models with combinations of Fama-French factors. The models in the remaining rows use factors derived from the Tucker decomposition of mutual fund characteristics. The sample period is 2010Q3 to 2018Q4.

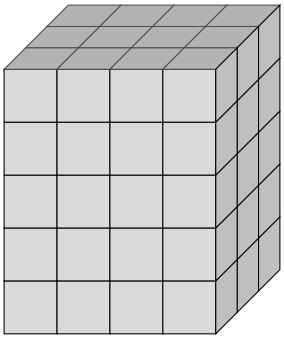
Table A-3: Mutual Fund Returns – Fit of Linear Factor Models

	K	MPE	RMSPE	RMSPE*
Panel A: Fama-French Factors				
Mkt-RF	1	-1.98%	4.14%	2.91%
Mkt-RF, SMB, HML	3	-1.07%	3.05%	3.00%
Mkt-RF, SMB, HML, MOM	4	-1.04%	2.99%	3.01%
Mkt-RF, SMB, HML, RMW, CMA	5	-0.92%	3.10%	2.95%
Mkt-RF, SMB, HML, RMW, CMA, MOM	6	-0.88%	3.06%	2.97%
Panel B: Tucker Factors				
Mkt-RF, 4	2	-1.43%	2.84%	2.77%
Mkt-RF, 3, 4	3	-0.89%	2.49%	2.79%
Mkt-RF, 1, 3, 4	4	-0.89%	2.32%	2.54%
Mkt-RF, 1, 2, 3, 4	5	-0.59%	2.29%	2.56%
Mkt-RF, 1, 2, 3, 4, 5	6	-0.60%	2.32%	2.67%
Panel C: Fama-French Factors, recursive				
Mkt-RF	1	-1.90%	4.11%	2.95%
Mkt-RF, SMB, HML	3	-1.12%	3.46%	2.83%
Mkt-RF, SMB, HML, MOM	4	-0.98%	3.39%	2.88%
Mkt-RF, SMB, HML, RMW, CMA	5	-1.08%	3.45%	2.82%
Mkt-RF, SMB, HML, RMW, CMA, MOM	6	-0.94%	3.39%	2.87%
Panel D: Tucker Factors, recursive				
Mkt-RF, 3	2	-0.84%	3.44%	3.36%
Mkt-RF, 1, 3	3	-0.47%	2.95%	2.83%
Mkt-RF, 1, 2, 3	4	-0.51%	2.90%	2.82%
Mkt-RF, 1, 2, 3, 4	5	-0.26%	2.91%	2.98%

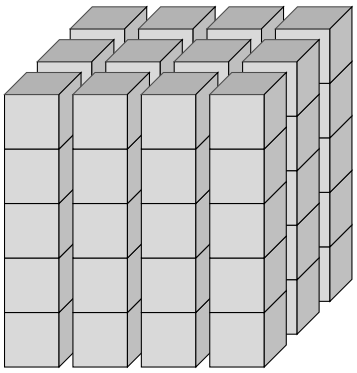
Notes: This table shows results of time-series estimations of linear asset pricing models. Panel A, C and B, D report results for models with combinations of Fama-French factors and Tucker pricing factors, respectively. The result in Panels A and B are based on the full sample estimation of the Tucker(5,10,5) model, while results in Panels C and D are based on a recursive estimation of Tucker(5,25,9) models with expanding windows. The first window has five time-series observations. K is the number of factors, MPE is the mean pricing error, and RMSE is the root-mean-square pricing error. RMSPE* is the pricing error when the third and fourth (Panel B) and first and third (Panel D) Tucker factors are added to the Fama-French models in Panel A, and SMB and HML are added to the Tucker models in Panel B.
The sample period is 2010Q3 to 2018Q4.

Figure A-1: Tensor fibers and slices

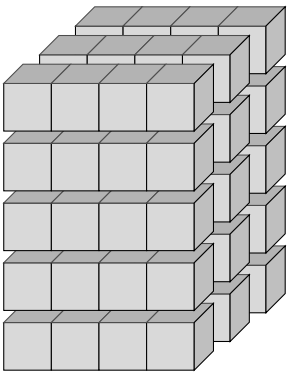
A: Tensor $\boldsymbol{\mathcal{X}}$: $(5 \times 4 \times 3)$



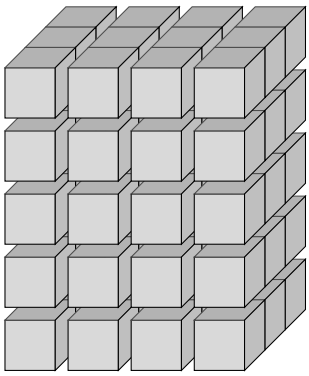
B: Mode-1 fibers $\mathbf{x}_{t(mc)}$



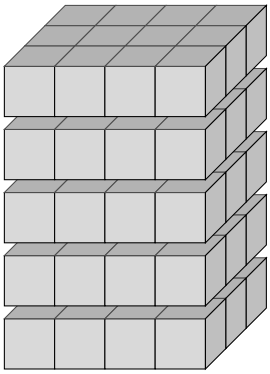
C: Mode-2 fibers $\mathbf{x}_{m(tc)}$



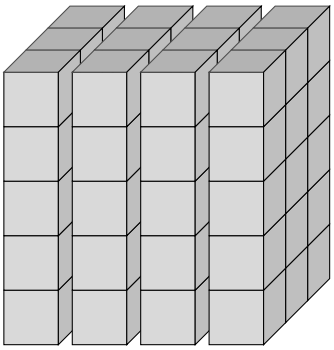
D: Mode-3 fibers $\mathbf{x}_{c(tm)}$



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



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



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

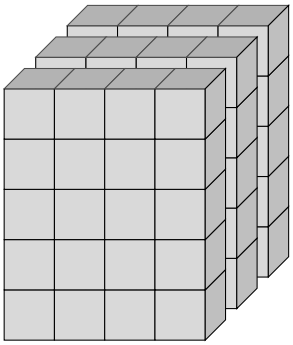
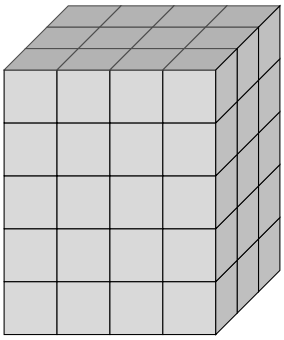
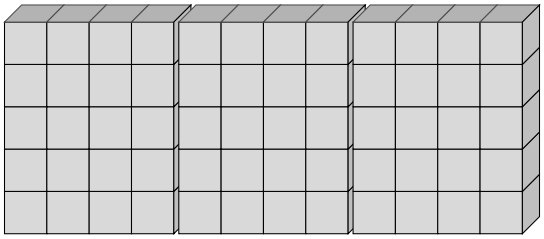


Figure A-2: Tensor as matrices

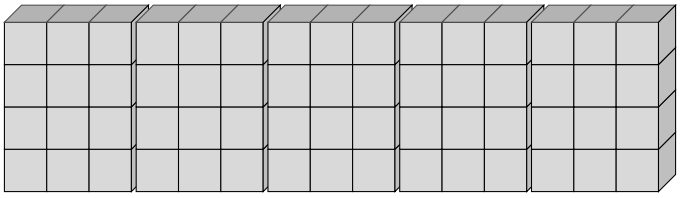
A: Tensor $\boldsymbol{\mathcal{X}}$: $(5 \times 4 \times 3)$



B: $\mathbf{X}_{(1)}$: (5×12)



C: $\mathbf{X}_{(2)}$: (4×15)



D: $\mathbf{X}_{(3)}$: (3×20)

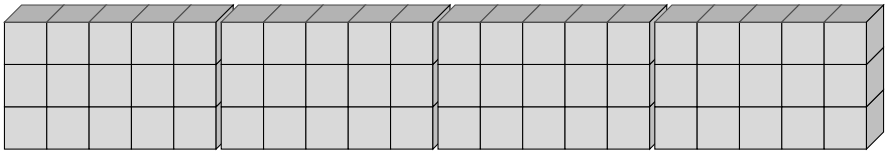


Figure A-3: Tensor multiplication

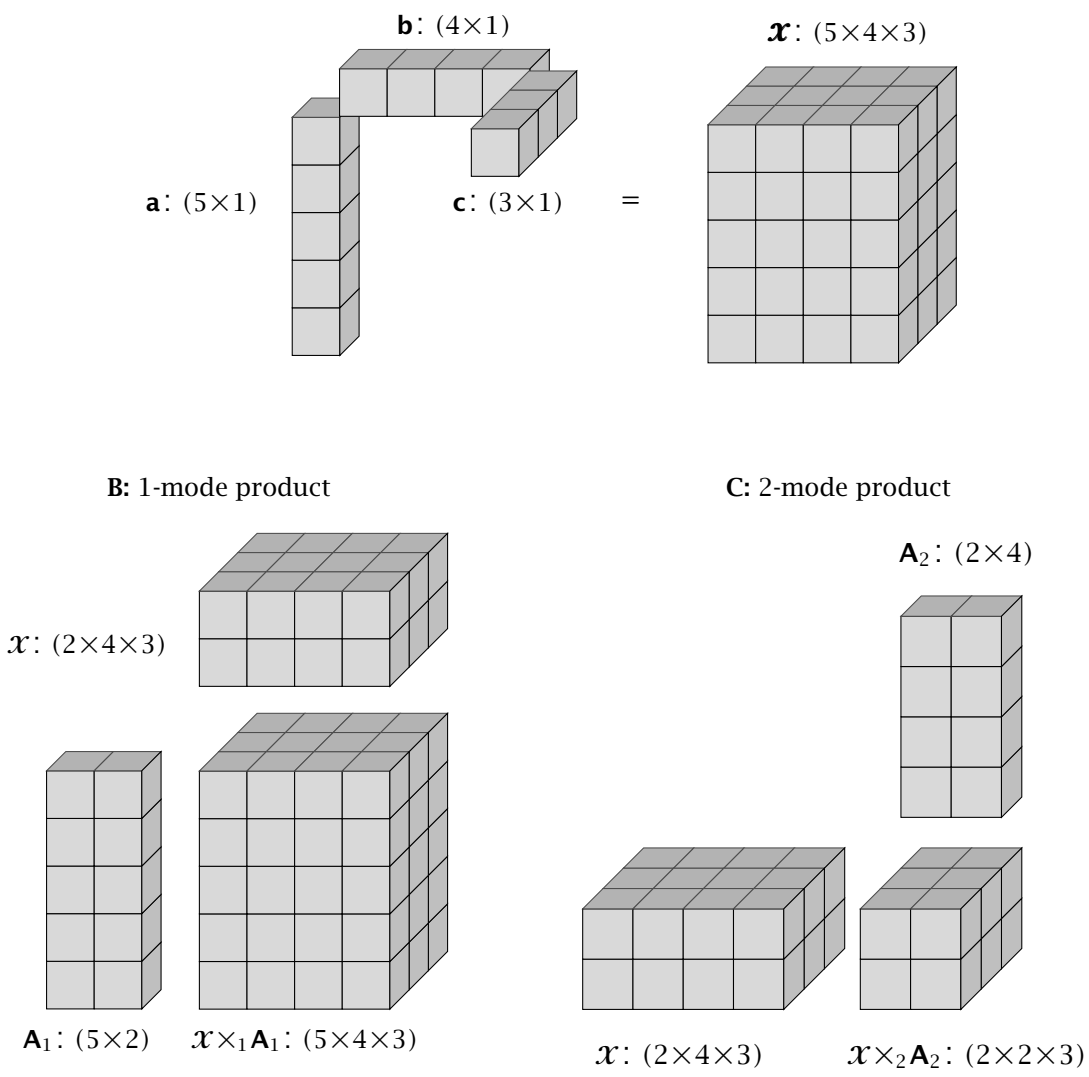
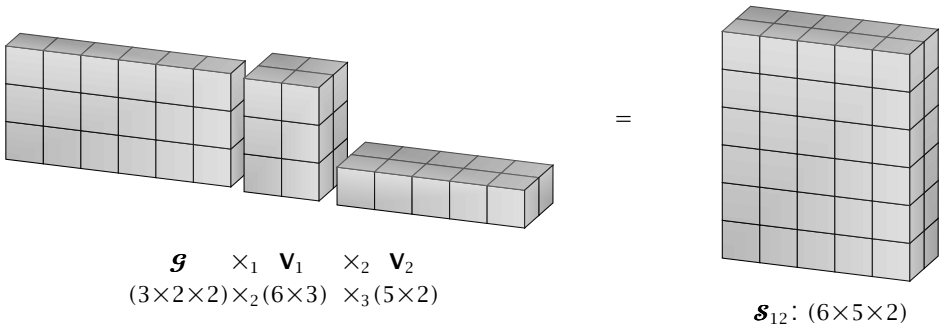


Figure A-4: Tucker model as two-dimensional factor model



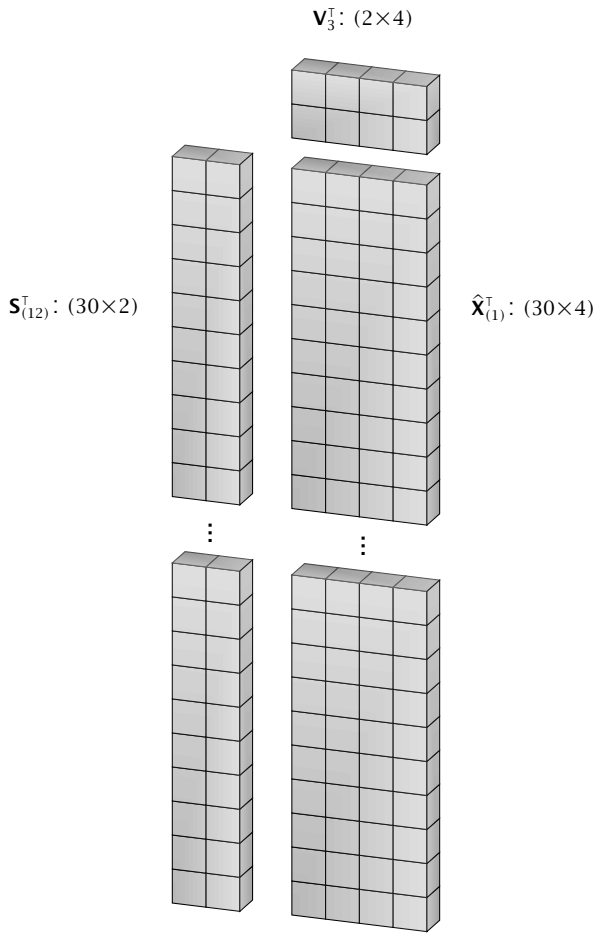
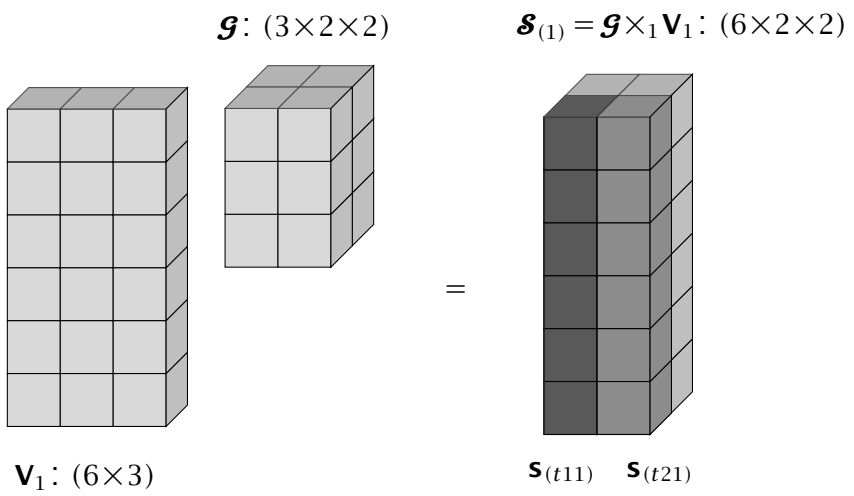


Figure A-5: Tucker model: Intuition

A: "summary" modes-(2,3) (M,C)



B: "summary" mode-2 (M)

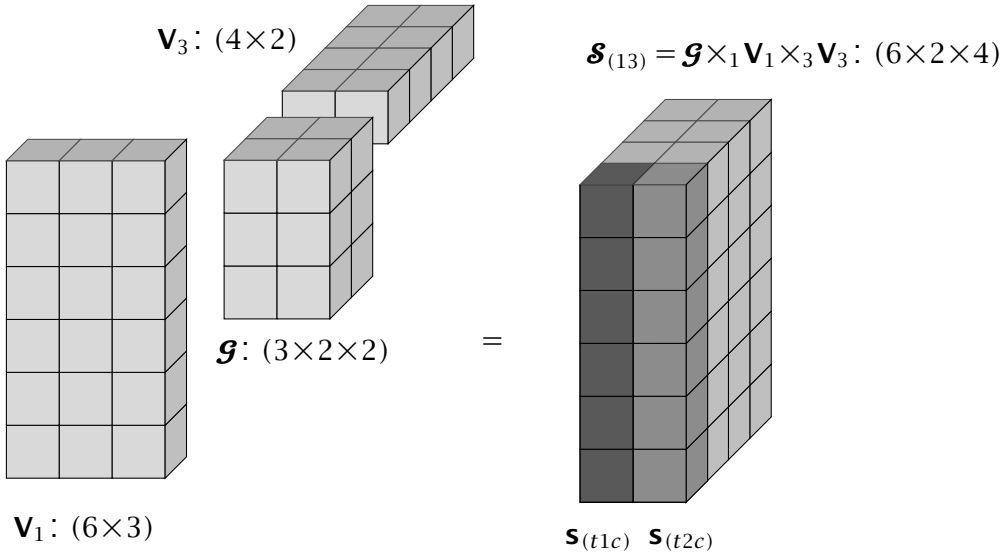
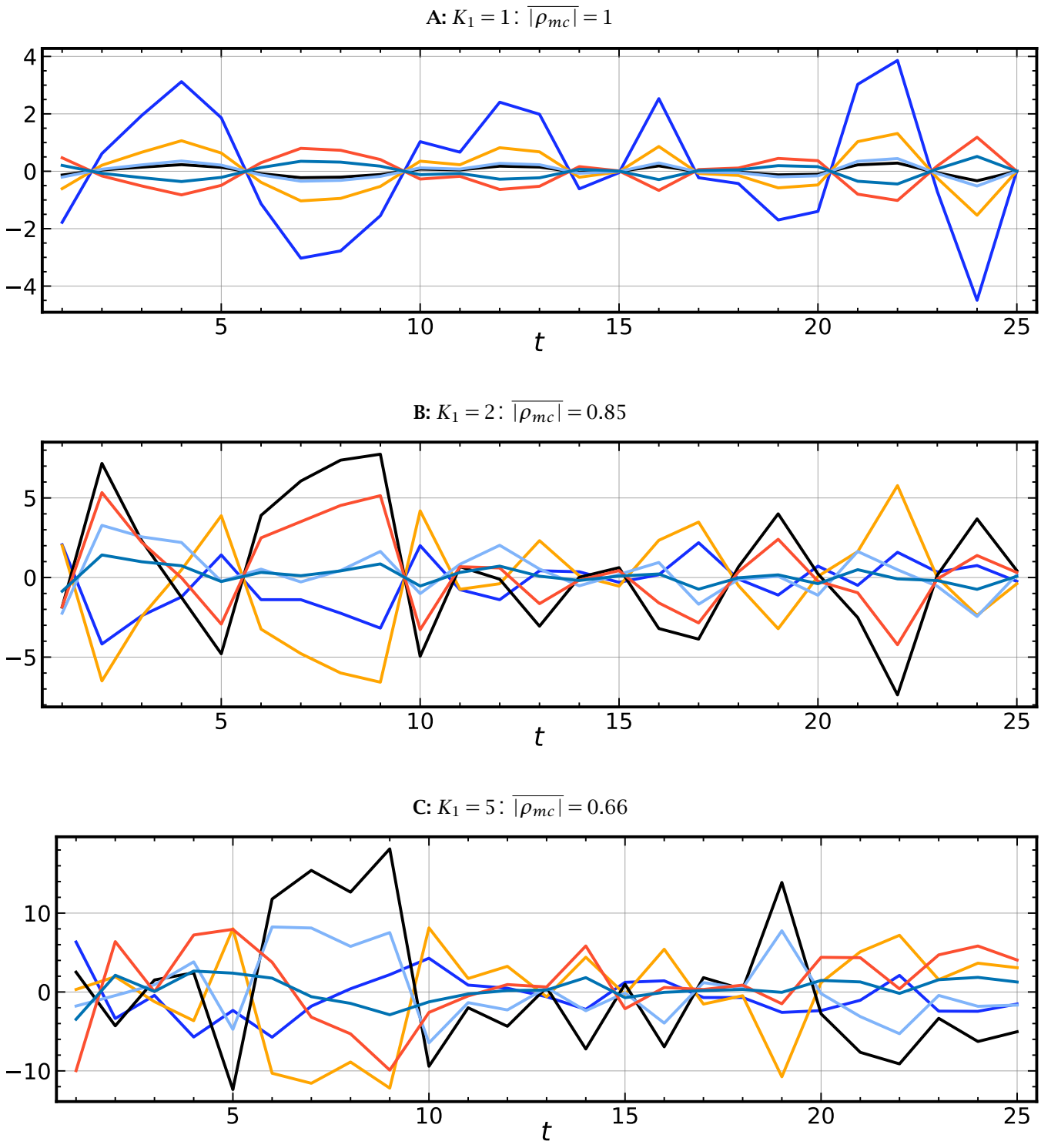


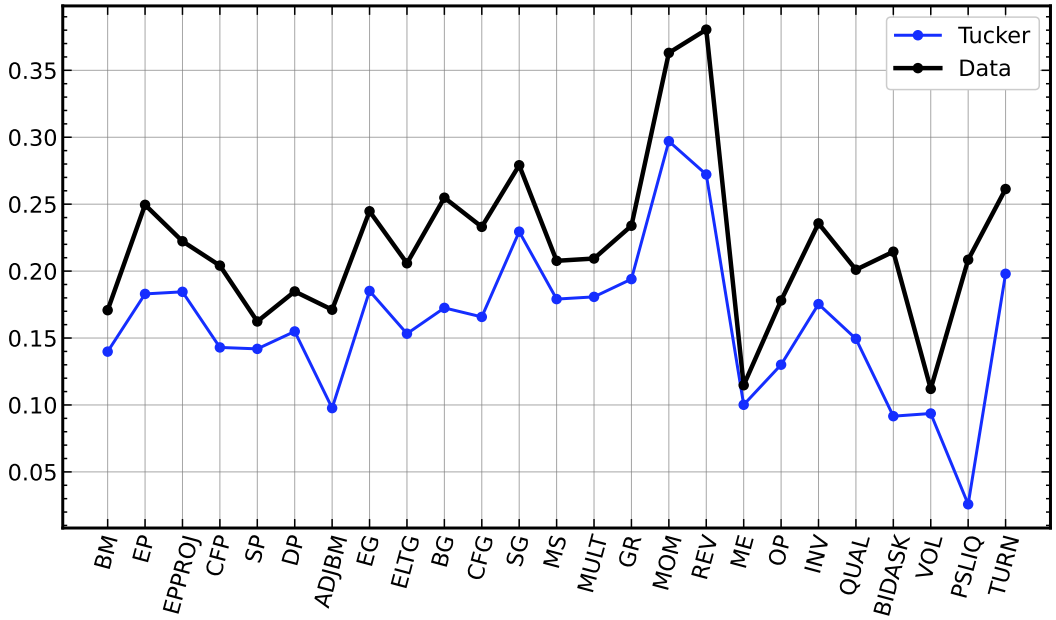
Figure A-6: Time series fibers $\mathbf{x}_{t(mc)}$ in $(K_1,1,2)$ Tucker model



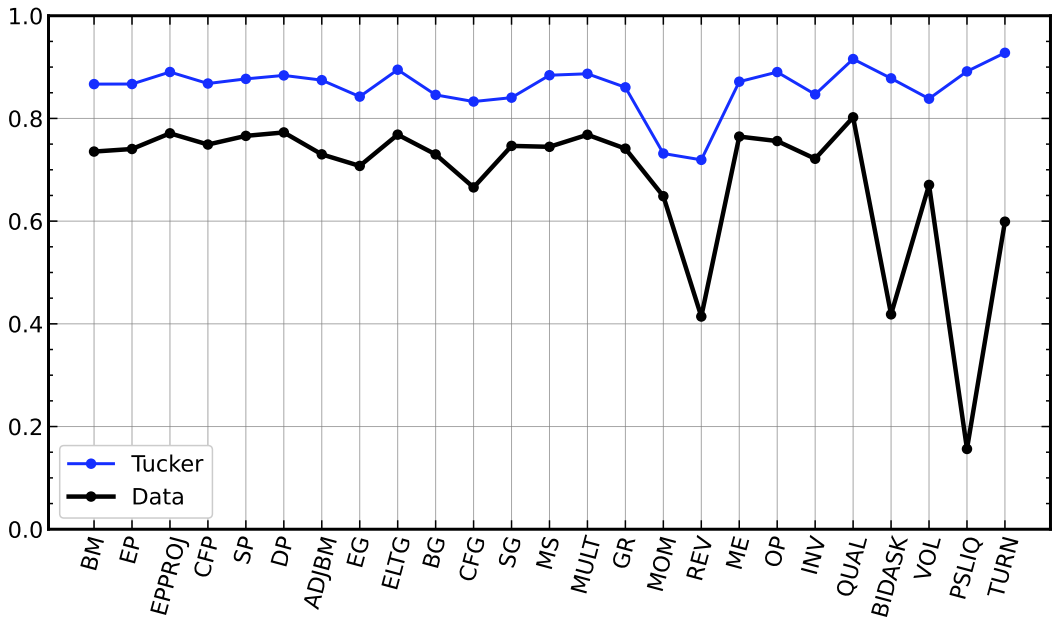
Notes: The figure plots fibers (vectors) $\mathbf{x}_{t(mc)}$ in a simulated $(25\times10\times5)$ Tucker model with $(K_1,1,2)$ components for $K_1 = 1,2,5$. Each mode-t fiber is a (25×1) vector, The panel shows six lines corresponding to a subset of (m,c) combinations.

Figure A-7: Cross-correlations of characteristics

A: Std. Dev. of Observed and Fitted Characteristics

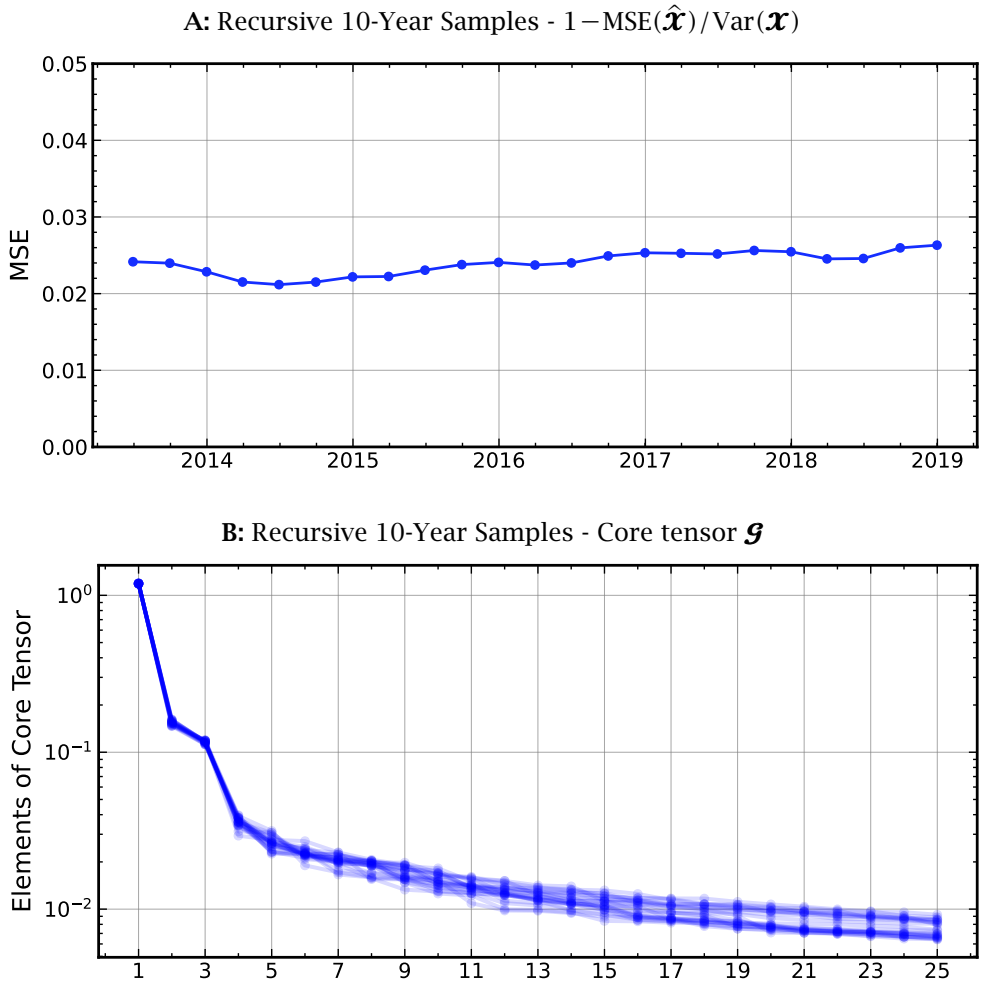


B: Autocorrelation of Observed and Fitted Characteristics



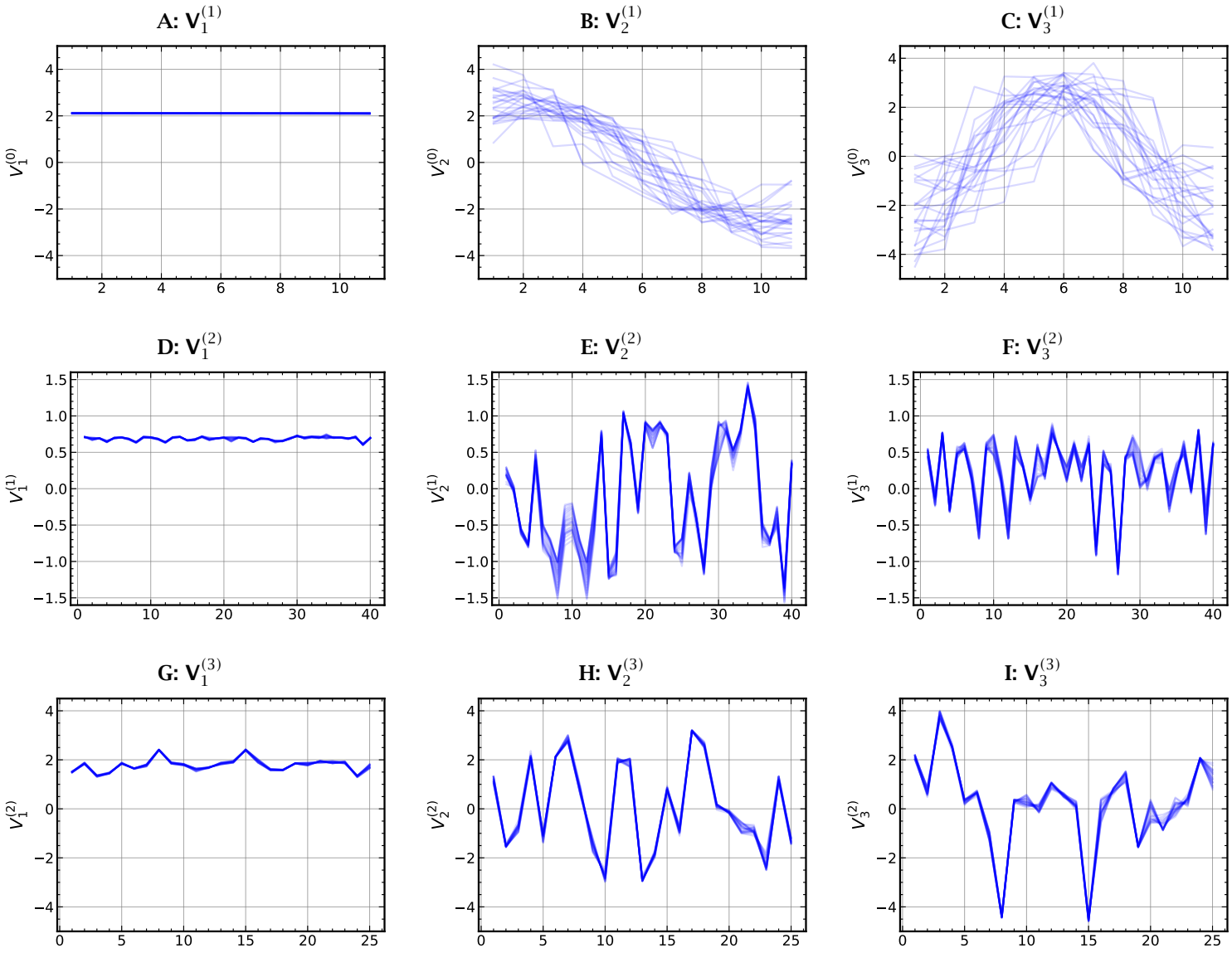
Notes: Panel A shows the time-series mean standard deviations across stocks by characteristics in the data (black) and for the fitted values of the Tucker (10, 25, 9) model. Panel B shows the first-order autocorrelations. The sample period is 2010Q3 to 2018Q4.

Figure A-8: Subsamples



Notes: This figure plots the first three loading vectors of mode $i, \mathbf{V}_j^{(i)}, i = 1,2,3, j = 1,2,3$ of Tucker models that are estimated over two subsamples consisting of 17 quarters: 2010Q3-2014Q4 and 2015Q1-2018Q4. The Tucker models have $(K_1, K_2, K_3) = (7,25,9)$ components. The blue lines correspond to the first half of the sample and the orange lines to the second half.

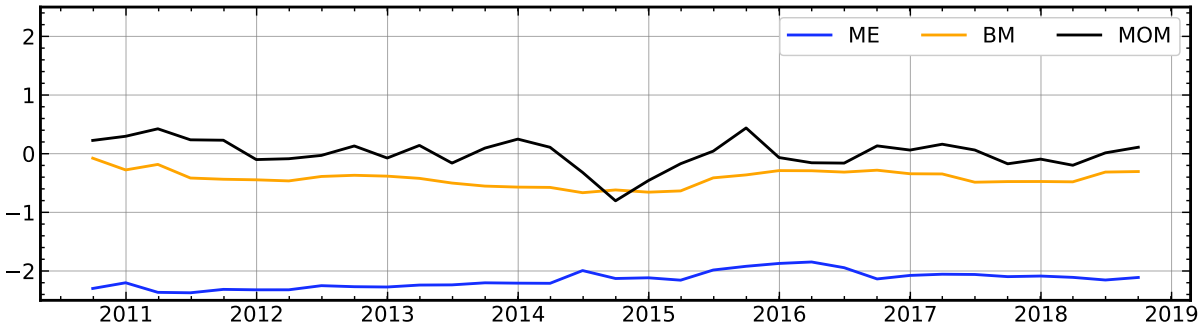
Figure A-9: Recursive 10-Year Samples - $V_j^{(i)}$



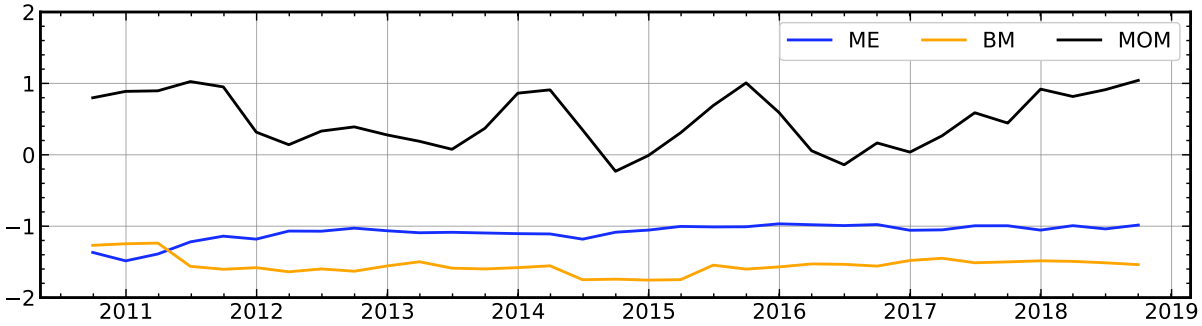
Notes: This figure plots the first three loading vectors of mode i,$\mathbf{V}_j^{(i)}$,$i = 2,3$,$j = 1,2,3$ of Tucker models that are estimated over sliding windows. Each window consists of 10 quarters so there are 23 subsamples. The Tucker models have $(K_1,K_2,K_3) = (7,25,9)$ components. Each panel plots 23 lines corresponding to a subsample loading vector $\mathbf{V}_j^{(i)}$. The sample period is 2010Q3 to 2018Q4.

Figure A-10: Factor Characteristics over Time - Out-of-sample Factors

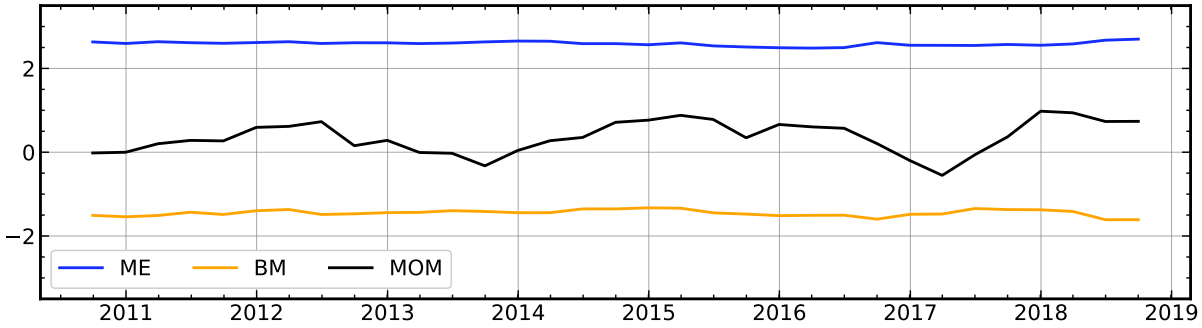
A: Factor 1



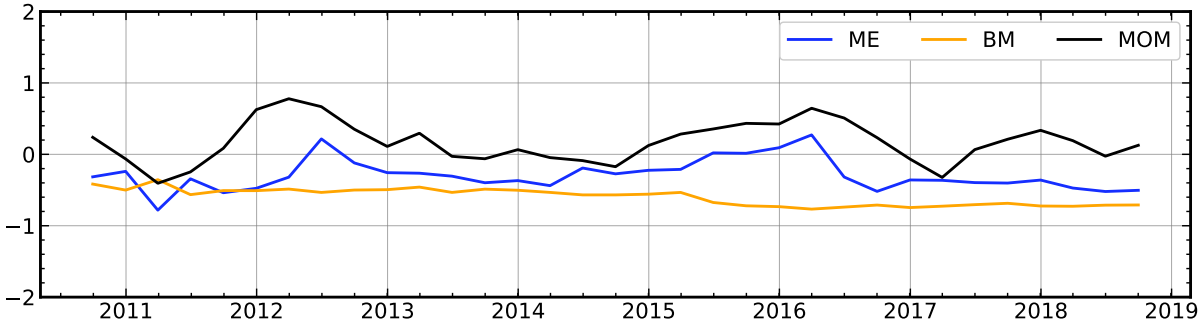
B: Factor 2



C: Factor 3

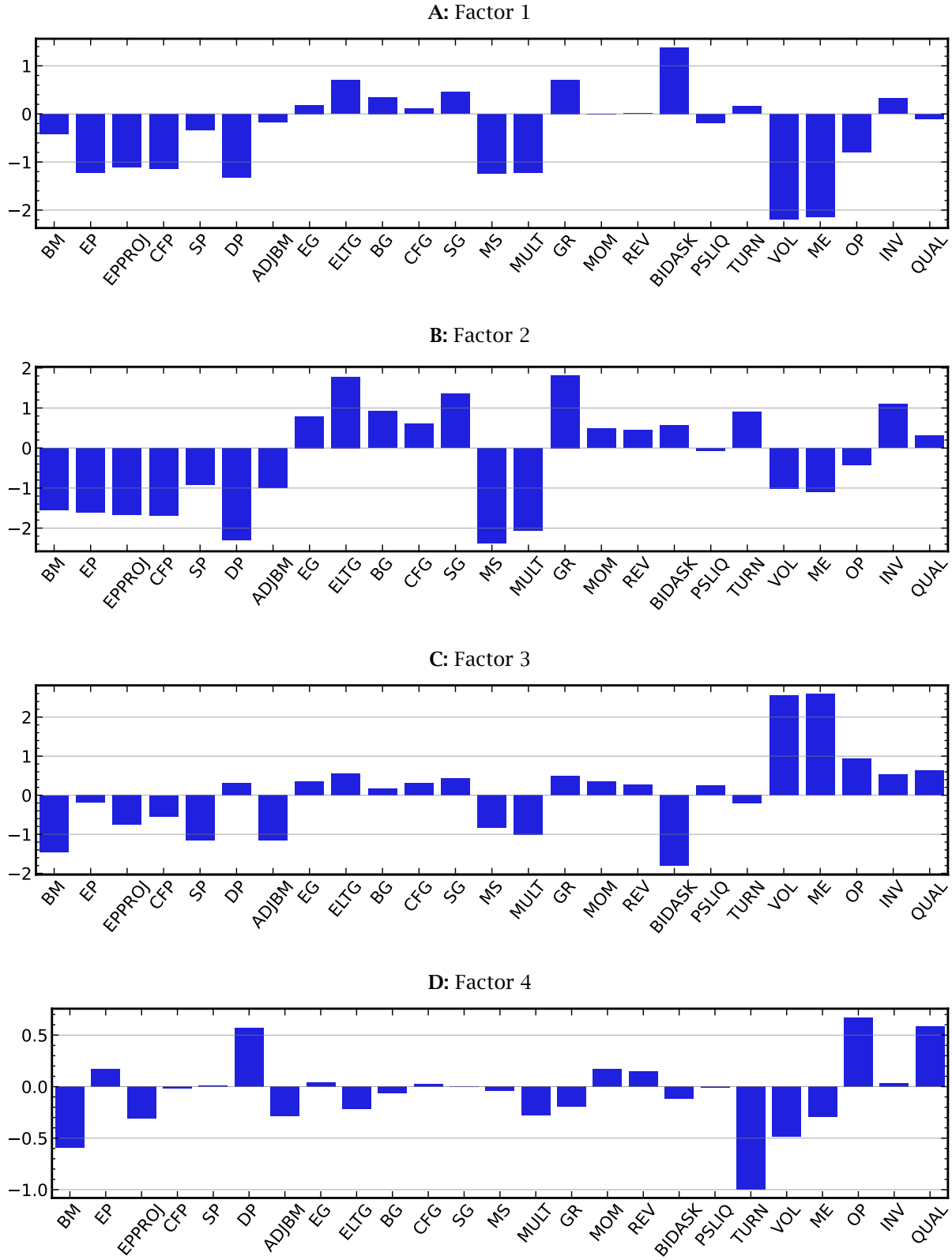


D: Factor 4



Notes: This figure shows the characteristics of the Tucker characteristic factors.

Figure A-11: Mean Characteristics of Tucker Factors - Out-of-sample



Notes: This figure shows the characteristics of the Tucker characteristic factors.