NBER WORKING PAPER SERIES

HIGH-DIMENSIONAL FACTOR MODELS WITH AN APPLICATION TO MUTUAL FUND CHARACTERISTICS

Martin Lettau

Working Paper 29833
http://www.nber.org/papers/w29833

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2022, Revised July 2022

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, and the 2022 BI-SHoF Conference for helpful comments. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2022 by Martin Lettau. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

High-Dimensional Factor Models with an Application to Mutual Fund Characteristics
Martin Lettau
NBER Working Paper No. 29833
March 2022, Revised July 2022
JEL No. C38,G12

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 93% 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.

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

1. Introduction

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 $\mathbf{X}^\top\mathbf{X}$ and $\mathbf{X}\mathbf{X}^\top$, $\mathbf{U}_1$ and $\mathbf{U}_2$, respectively, and the diagonal matrix $\mathbf{H}$ of associated eigenvalues: $\mathbf{X} = \mathbf{U}_1\mathbf{H}\mathbf{U}_2^\top$. 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 "representative" 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 "representative" time, K_2 "representative" fund, and K_3 "representative" 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 exam-

ple, 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 across 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 93% 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 representative mutual funds with nine representative characteristics observed over ten representative 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 "representative" time/fund/characteristics elements in the core tensor are related to means, and higher-order elements represent deviations from means.

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 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

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

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

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

where $\mathbf{U}^{(1)}$ is a $(T\times T)$ matrix of eigenvectors $\mathbf{u}_t^{(1)}$ of $\mathbf{X}\mathbf{X}^\top$ as columns, $\mathbf{U}^{(2)}$ is a $(N\times N)$ matrix of eigenvectors $\mathbf{u}_t^{(2)}$ of $\mathbf{X}^\top\mathbf{X}$ as columns, and $\mathbf{H}$ is a diagonal $(T\times N)$ matrix with diagonal elements h_i that are the square 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 $\mathbf{x}_{(\bullet mc)}$,$\mathbf{x}_{(t\bullet c)}$, and $\mathbf{x}_{(tm\bullet)}$, respectively.[4] Slices are created by fixing all but two indices and are written as $\mathbf{X}_{(t\bullet\bullet)}$,$\mathbf{X}_{(\bullet m\bullet)}$,$\mathbf{X}_{(\bullet\bullet c)}$.[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:

$$\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \rangle = \sum_{t,m,c} x_{tmc} y_{tmc}$$

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

[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.

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

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 $(1\times N)$ 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

[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.

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 eigenvec-

tor matrices in the SVD-PCA model. The mechanism of the Tucker decomposition (14) is illustrated in Figure 1, which shows the decomposition of a $(6\times5\times4)$ 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 $(3\times2\times2)$. 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 (6×3),(5×2), 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 $\mathbf{X}_{(t\bullet\bullet)}$ is a $(M\times C)$ matrix that, by rewriting (15), can be expressed as

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

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

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

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

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

The term in parentheses is a scalarso that all fibers $\mathbf{x}_{(\bullet mc)}$ are proportional to the $(T\times1)$ vector $\mathbf{v}_1^{(1)}$. Therefore, all-time-series fibers are perfectly correlated. Panel A of Figure 2 shows such time-series

[9] See Panels B–F of Figure A-1.

fibers $\mathbf{x}_{(\bullet mc)}$ 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

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

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

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

The right-hand side of (21) is the outer product of an $(M\times1)$ vector $\mathbf{v}^{(2)}_1$ and a $(C\times1)$ vector $\mathsf{v}^{(1)}_{1,t}\,\tilde{\mathbf{v}}^{(3)}_1 + \mathsf{v}^{(1)}_{2,t}\,\tilde{\mathbf{v}}^{(3)}_2$ and yields a $(M\times C)$ matrix. Therefore, as in the case with a single mode-t component above, each slice $\mathbf{X}_{(t\bullet\bullet)}$ is formed by a $(M\times C)$ one-factor model. However, (21), unlike (18), implies that the one-factor models differ across t since the term in parentheses is a linear combination of two vectors and the weights $\mathsf{v}^{(1)}_{1,t}$ and $\mathsf{v}^{(1)}_{2,t}$ change across t. Equation (22) shows that t-fibers $\mathbf{x}_{(\bullet mc)}$ are linear combinations of two $(T\times1)$ vectors, $\mathbf{v}^{(1)}_1$ and $\mathbf{v}^{(1)}_2$, 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 2 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 $K_2\le K_3$; then (15) implies that t-mode slices

$\mathbf{X}_{(t\bullet\bullet)}$ and fibers $\mathbf{x}_{(\bullet mc)}$ can be written as

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

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

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

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 $\mathbf{X}_{(t\bullet\bullet)}$ is the sum of K_1 $(M\times C)$ matrices. If there is a single K_1 factor, then all t-slices $\mathbf{X}_{(t\bullet\bullet)}$ are formed by the same $(M\times C)$ matrix multiplied by the scalar $\mathsf{v}^{(1)}_{1,t}$ and are thus proportional. Note, that this property holds for any K_2 and K_3. As K_1 increases, $\mathbf{X}_{(\bullet mc)}$ 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 $\mathbf{x}_{(\bullet mc)}$ are linear combinations of K_1 columns of the mode-1 $(T\times K_1)$ loading matrix $\mathbf{V}^{(1)} = [\mathbf{v}^{(1)}_{k_1}]$. 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 $\mathbf{V}^{(2)} = [\mathsf{v}^{(2)}_{k_2,m}]$ and $\mathbf{V}^{(3)} = [\mathsf{v}^{(3)}_{k_3,c}]$. If $K_1 = 1$, all mode-t fibers $\mathbf{x}_{(\bullet mc)}$ are perfectly correlated, no matter the values of K_2 and K_3. As for t-slices $\mathbf{X}_{(t\bullet\bullet)}$, higher values of K_1 create more complex patterns of correlations of $\mathbf{x}_{(\bullet mc)}$ 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 $\mathsf{v}^{(2)}_{k_2}$, yields a $(M\times C)$ matrix. Therefore, each $\mathbf{X}_{(t\bullet\bullet)}$ 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 $\mathsf{v}^{(1)}_{k_1,t}$ and $\mathsf{v}^{(3)}_{k_3}$ vectors but depend on the weights $g_{k_1k_2k_3}$. Note, that the vectors in the M-dimension, $\mathsf{v}^{(2)}_1,\ldots,\mathsf{v}^{(2)}_{K_2}$ 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 $\mathbf{X}_{(t\bullet\bullet)}$ in (24) is K_3, instead of K_2. In general, $\mathbf{X}_{(t\bullet\bullet)}$ slices are formed by a $(M \times C)$ factor model of order $\min(K_1, K_2)$.

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 $\mathbf{X}_{(t\bullet\bullet)}$ across t. If $K_i = 1$, then all mode-i fibers are perfectly correlated and all slices $\mathbf{X}_{(t\bullet\bullet)}$ 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 $\mathbf{X}_{(t\bullet\bullet)}$ are $(M \times C)$ factor models with $\min(K_2, K_3)$ factors for all $t = 1,\ldots,T$.
- m-slices $\mathbf{X}_{(\bullet m\bullet)}$ are $(T \times C)$ factor models with $\min(K_1, K_3)$ factors for all $m = 1,\ldots,M$.
- c-slices $\mathbf{X}_{(\bullet\bullet c)}$ are $(T \times M)$ factor models with $\min(K_1, K_2)$ factors for all $c = 1,\ldots,C$.

[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 $\mathbf{X}^{\intercal}$ are equivalent.

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 $\mathbf{S}^{\top}_{(jk)}$ are equivalent to 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 $\mathbf{X}^{\top}_{(3)}$ 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 "representative" version of the data matrix. This is 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 "representative" 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 "representative" core mutual fund k_2 and "representative" 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 fiber of this tensor, $\mathbf{s}_{(tk_2c)}$ is a time-series of characteristic c of the "representative" core mutual fund k_2.[12]

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

[12]Figure A-6 illustrates the construction of the "representative" mutual funds and "representative" characteristics $\mathbf{s}_{(tk_2k_3)}$ in Panel A and the representative 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 representative mutual funds and characteristics unchanged, see Panel A. Therefore, the resulting tensor $\boldsymbol{\mathcal{S}}^{(1)}$ consists of six periods, two representative funds (in mode-2), and two representative 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 representative fund and characteristic. The figure shows the 6 time-series observations of the first ($k_2 = 1$) representative fund and first ($k_3 = 1$) representative characteristics, $\mathbf{s}_{(t11)}$, and the second ($k_2 = 2$) representative fund and first ($k_3 = 1$) representative 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 representative mutual funds in the second mode. The darker shades indicate the time-series observations of the first and second representative funds for characteristics c, $\mathbf{s}_{(t1c)}$ and $\mathbf{s}_{(t2c)}$.

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]

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

[13]Figure A-4 illustrates the steps of the derivation for the example in Figure 1. 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 $(6\times5\times2)$. Unfolding $\boldsymbol{\mathcal{S}}_{(12)}$ along mode-3 yields $\mathbf{S}_{(1,2)}$, which is (2×30). Unfolding the original $(6\times5\times4)$ 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.

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. The CP decomposition

An alternative but related tensor decomposition was developed independently of the Tucker model. It was first proposed by Hitchcock (1927) but is known as CANDECOMP ("canonical decomposition", Carroll and Chang (1970)) or PARAFAC ("parallel factors", Harshman (1970)), usually abbreviated as *CP decomposition*. It can be written as a special case of the Tucker model (14) when the numbers of components in each mode are equal, that is $K_1=K_2=K_3=K$, and the core tensor $\boldsymbol{\mathcal{G}}$ is diagonal. The CP model with K components, CP(K), is defined as

$$\boldsymbol{\mathcal{X}}(K) = \sum_{k=1}^{K} g_{kkk}\,\mathbf{w}_k^{(1)} \circ \mathbf{w}_k^{(2)} \circ \mathbf{w}_k^{(3)} \tag{29}$$

$$= \boldsymbol{\mathcal{G}}_{\mathrm{CP}} \times_1 \mathbf{W}^{(1)} \times_2 \mathbf{W}^{(2)} \times_3 \mathbf{W}^{(3)}, \tag{30}$$

where $\boldsymbol{\mathcal{G}}_{\mathrm{CP}}$ is a diagonal $(K\times K\times K)$ tensor with diagonal elements g_{kkk} and $\mathbf{W}^{(1)}$,$\mathbf{W}^{(2)}$, and $\mathbf{W}^{(3)}$ are $(T\times K)$, $(M\times K)$, and $(C\times K)$ matrices with the normalized vectors $\mathbf{w}_k^{(1)}$,$\mathbf{w}_k^{(2)}$, and $\mathbf{w}_k^{(3)}$ as columns.[14] Recall that an n-dimensional tensor that can be written as an outer product of n vectors is a rank-one tensor, see (8). The CP decomposition is, therefore, an approximation of $\boldsymbol{\mathcal{X}}$ by the sum of K rank-one tensors.

2.8. A comparison of SVD-PCA, Tucker, and CP decompositions

It is instructive to compare the two-dimensional SVD-PCA model (10, 11), and the higher-dimensional Tucker (29, 30), and CP (14, 15) decompositions. All three models minimize a quadratic loss func-

[14] The CP decomposition is illustrated in Figure A-5.

tion subject to linear constraints. The CP model is a special case of the Tucker model with the constraint that the core tensor $\mathcal{G}_{\text{CP}}$ is diagonal while the Tucker core tensor $\mathcal{G}$ may have non-zero off-diagonal elements. In turn, the SVD decomposition is a special case of the CP decomposition for two-dimensional matrices.

The minimization problem for two dimensions (*i.e.*, matrices) has a closed-form solution given by eigenvalues and eigenvectors, while numerical methods are required for higher dimensions (*i.e.*, tensors). The SVD and Tucker decompositions always exist but are not unique and can be rotated. In contrast, the CP decomposition is unique under mild conditions but is not guaranteed to exist for all K; see Kolda and Bader (2009).

2.9. Computation

In contrast to the SVD-PCA matrix representation, there are no closed-form solutions for the Tucker and CP tensor decompositions (14) and (30) 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 for, the core tensor $\mathcal{G}$ can be constructed. A similar procedure can be used for computing the CP model. The literature has developed many refinements to this algorithm to improve the rate and speed of convergence even if the data tensor is "large".

One important property of the Tucker and CP decompositions is that they cannot be computed sequentially. Consider two CP models with K and $K' > K$ components. The first K components of the CP(K') are different from the K components of CP(K) so that CP(K) and CP(K') have to be estimated separately. This is also true for Tucker decompositions with different numbers of components (K_1, K_2, K_3). 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.

Tucker models are generally easier to estimate than CP models. One reason is that, in contrast to Tucker decompositions, a CP(K) representation might not exist for some K, leading to possible numerical difficulties. Tucker models are less restrictive and more flexible. This flexibility comes at

a cost, however, since Tucker models are not unique, while CP models, if they exist, are unique. An advantage of Tucker models is that they allow for different numbers of components for each mode, while the CP representation restricts the number of components to be the same across models. If the number of dimensions varies substantially across modes, Tucker models might be a better choice. The data set used in the next section has 34 quarterly (mode-1) and 25 characteristics (mode-3) observations, but the second-mode consists of 1,697 individual mutual funds. I will discuss the trade-offs of the estimation of Tucker and CP models in this data set in more detail below.

3. Mutual fund characteristics over time

Next, I estimate Tucker and CP 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[15], and four liquidity measures, see Table 1. To obtain a balanced panel with no missing data, I select all 1,342 mutual funds and ETFs that are in the sample for all quarters between 2010Q3 and 2018Q4.[16] 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 \$575 mil. with an inter-quartile range of just under \$200 mil. to \$1.70 bil. The mean TNA of \$2.34 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 84 stocks with an interquartile range of 54 to 165 with a mean of 190. The market beta of most mutual funds is between 0.92 and 1.06. As is well-known in the literature, mutual funds underperform broad stock market indices, and alphas of the majority are negative.

[15] The quality index combines the return-to-equity, debt-to-equity, and earnings variability.
[16] Choosing an earlier starting date drastically reduces the number of funds without missing values.

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 $(T \times M \times C) = (34 \times 1342 \times 25)$.

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 across characteristics (mode-3). One possibility is to unfold the three-dimensional data tensor along the characteristic dimension into a (25×45628) 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 3 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 3 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 3.

3.1. Estimation of Tucker and CP 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 2.9. 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, 1,342 mutual funds, and 25 characteristics into a single "representative" mutual fund with a single "representative" characteristic observed at one "representative" 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 1,342 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 $K_1 \approx 10, K_2 \approx 30$, and $K_3 \approx 10$ components offer the best fit (R^2 over 95%) *vs.* parsimony (over 94% dimension reduction) trade-off. I choose the model with $(K_1,K_2,K_3) = (10,25,9)$ components as the benchmark specification for the rest of the paper. The MSE of this model is 0.03, and the R^2 is 93%, 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 (10,25,9) dimensional core, which can be interpreted as 25 representative mutual funds with nine representative characteristics observed over 10 representative periods.

Before analyzing the fit and properties of this model in more detail, I estimate CP(K) models

of the form (29). Recall that in CP(K) models, the number of components is the same along each mode, in contrast to the Tucker models, which allows for different numbers of components for each mode. This restriction may not be an issue if the dimensions of the data along each mode are similar; however, this might not be true if the mode dimensions vary substantially, as is the case in the mutual fund sample. The time and characteristic dimensions of the sample are relatively small, 34 and 25, respectively, while the mutual fund dimension is 1,342. Hence, it would not be surprising if Tucker models with more components along the mode-2 mutual fund dimension than along the first time and third characteristic dimensions would be preferable to CP models with equal mode-dimensions. Recall that the benchmark Tucker model has 25 mutual fund components but only 10 and 9 date and characteristic dimensions, respectively.

With this caveat in mind, I will proceed with the estimation of CP models and compare their fit to the fit of Tucker models. Panel D of Figure 4 plots the MSE for $K = 1,...,20$. The iterative ALS estimation becomes less robust for $K > 15$ and sometimes does not converge. The MSE shown in the figure is the mean across iterations that did converge. However, I will not consider $K > 15$ in the rest of the paper. The pattern of the MSE is similar to that of the Tucker $(10,10,K_3)$ model shown in Panel C (blue line). The MSE drops from 0.33 for $K = 1$ to 0.04 for $K = 15$ and to 0.03 for $K \gtrsim 18$. I choose the CP(12) model as the benchmark specification because it combines a good fit with a parsimonious number of mode-1 and mode-3 components. This specification has 16,812 degrees of freedom, thus compressing the data by 99%.

3.2. The fit of Tucker and CP benchmark models

In this section, I will analyze the fit of the two benchmark models, Tucker(10, 25, 9) and CP(12), in more detail. Table 3 shows descriptive statistics of the distributions of the errors. The mean and median error of the Tucker(10, 25, 9) and CP(12) models are both close to zero. The standard deviation of the Tucker model is 0.19 and lower than the standard deviation of the CP model of 0.22. Neither error distribution exhibits significant skewness, but both have fat tails, as indicated by excess kurtosis of over 3. The percentiles reported in Panel B confirm the presence of some extreme outliers. The inter-quartile range of the errors of the Tucker(10, 25, 9) is (-0.101, 0.099) and 90% and 99% of the errors are in the intervals (-0.302, 0.303) and (-0.612, 0.629), respectively. The high kurtosis of the error distribution is primarily driven by the presence of outliers. The minimum and maximum errors are -2.299 and 2.102, respectively. Errors of the CP(12) model are more spread

out than those of the Tucker model but the overall shape of the error distribution is similar. The outliers can be seen in Panel A of Figure 5, which shows the errors sorted from most negative to most positive. The errors of the Tucker model are in black, and the CP errors are in orange. The figure shows that errors are small for the majority of observations, suggesting that the Tucker and CP models yield 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. I will also compute some results when only one of the dimensions is collapsed. 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 for each quarter in the sample. The blue line is for the Tucker(10, 25, 9) model, and the orange line is for the CP(12) model. The plot shows that there is little MSE variation across quarters, especially for the Tucker model, indicating that the model yields a good fit over the whole sample. The MSE of the CP model is somewhat larger at the start and end of the sample period than in the middle. Panel C shows the MSE aggregated across the 1,342 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 80% (95%) of all funds. The corresponding numbers are slightly lower for the CP model, 71% and 90%, respectively. However, there are a few outlier mutual funds with much larger MSE. Three funds have a Tucker MSE higher than 0.2, and the CP model has 13 such mutual funds. I conclude that the Tucker(10, 25, 9) in particular yields a good fit for all but very few mutual funds.

Finally, MSE aggregated on the characteristic level are plotted in Panel C. There are several interesting patterns. For all but two characteristics (the adjusted book-to-market ratio, *adjbm*, and the Pastor-Stambaugh liquidity measure, *psliq*), the MSE of the Tucker model is lower than those of the CP model, but the patterns across characteristics are similar. The characteristics with the lowest MSE are size (*me*) and volume (*dvol*) for both models. The Tucker errors of 17 out of 25 characteristics are between 0.01 and 0.04, while the MSE of 15 characteristics is between 0.02 and 0.05 in the CP model. In both models, two characteristics have significantly higher MSE than the rest, momentum (*mom*) and reversals (*rev*). I will return to the fit across characteristics below after discussing the

properties of the components of the Tucker and CP models.

Next, I analyze model errors when they are collapsed across funds, that is I compute the MSE across all mutual funds for each quarter/characteristics pair. The resulting two-dimensional 34-by-25 matrices for the Tucker (Panel A) and CP (Panel B) models are shown as heatmaps in Figure 6. Lighter (darker) shaded of blue indicates low (high) MSE. The heatmaps reveal several interesting patterns. First, they facilitate a better comparison of the fit of Tucker and CP models. The overall lighter shades in Panel A compared to those in Panel B confirms that the Tucker model has a better overall fit than the CP model, particularly in the first three quarters of the sample. This difference is especially pronounced for characteristics related to growth rates and value/growth. Second, the heatmaps show that momentum and reversals are associated with the highest MSE across the entire sample. Third, even though Panel B in Figure 5 showed little evidence of time variation in the model fit when errors were collapsed by fund and characteristic, the two-dimensional heatmaps show that there is significant time variation when errors are aggregated only across mutual funds. For example, the MSE of momentum and reversals varies considerably across quarters. The momentum MSE in 2013Q1 is 0.27 and almost twice as high as in 2015Q2 (0.14). The range of MSE is even larger for reversals, 0.15 in 2014Q1 to 0.32 in 2010Q4, and in particular for PS liquidity, 0.11 in 2012Q4 to 0.38 in 2017Q1. On the other hand, the MSE is stable across the sample for most characteristics, including size, the book-to-market ratio, and the dividend-price ratio.

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 5-factor Fama-French model, the book-to-market ratio (*bm*), profitability (*op*), investment (*inv*), and momentum (*mom*). For each of these characteristics, I choose three representative mutual funds that illustrate the fit of the Tucker and CP models. I always include the "worst-case" mutual funds with the worst fit. Figure 7 shows a row for each of the four characteristics with three plots for individual funds. Each panel plots the data in orange, the fitted values of the Tucker model in black, and the CP model in blue. The "worst-case" funds are in the third column. The legends include the wficn of the fund that is plotted as well as the MSE of the Tucker and CP models.

The first row shows plots of the book-to-market ratio. The left panel shows an example in which both models yield a good fit. The BM ratio of the fund increases over the sample from about 2.4 to 3.4 (orange line). BM ratios of individual stocks are persistent, so it is not surprising the BM ratio of

fund portfolios are persistent as well. The MSE of CP is slightly lower than that of the Tucker model for this fund, but both capture the overall pattern of the observed data but the fitted time-series are smoother than the time-series of the observed data. The middle panel shows an example in which the fit of the CP model is poor, but the Tucker model yields a good fit. The BM ratio of this fund is around 3 from 2011 to the end of 2013 but drops to around 1.75 for the rest of the sample. The Tucker model captures this sudden change reasonably well but smoothes the drop considerably. The fitted BM ratio of the CP model does not capture the drop in the data resulting in a large MSE. The right panel shows the "worst-case" funds with the highest BM MSE. The fit of the Tucker model is reasonably good over the first half of the sample but cannot capture the sudden rise in 2014 and under-fits the BM ratio for the rest of the sample. In contrast, the fitted values of the CP model approximate the data poorly.

The second row shows three examples of profitability. In the left panel, the Tucker model has a better fit, while the center panel shows an example where the CP yields a significantly better fit than the Tucker model. The worst-case mutual fund with the poorest fit is shown in the right panel. For this fund, 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 0.91 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.

The third and fourth rows show plots for investment and momentum. The left panels show examples when both models fit the observed data well, while the center panels show cases when the Tucker model yields a better fit than the CP model. The Tucker model also produces reasonable fits for the worst-case funds displayed in the panels on the right but the CP performs particularly poorly. The CP fit for momentum in the bottom-right panel is the fifth-worst in the entire sample.

In summary, the Tucker model with (10, 25, 9) components yields a good approximation to the observed mutual fund data. It captures 93% of the variation in the data while reducing the full dimension of the data by 97%. The model is stable across the sample and fits well for most mutual funds but there are some outlier funds with poor fits. It also produces good approximation across characteristics with momentum and reversals having the largest MSE. The CP model with 12 components is more parsimonious than the Tucker(10, 25, 9) model but comes at the cost of an overall somewhat

worse fit. Next, I will analyze the properties of the estimated models and develop an intuition of the various components of each model. I will focus on the Tucker model since it fits better than the CP and its components have more direct economic interpretations.

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(10, 25, 9) model estimated for a data tensor with dimensions $(34 \times 1342 \times 25)$, $\boldsymbol{\mathcal{G}}$ is a $(10 \times 25 \times 9)$-dimensional tensor, $\mathbf{V}_1$ is (10×34), $\mathbf{V}_2$ is (25×1342), and $\mathbf{V}_3$ is (9×25). This section focuses on the properties of these variables. Panel A of Figure 8 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 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 13 of the 15 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, see Figure 4.

The remaining panels of Figure 8 show the largest 10 values of the two-dimensional cores that are implied by the three-dimensional Tucker model. The plots correspond to the eigenvalue spectrums of standard two-dimensional SVD-PCA models. The plot of the core values for the two-dimensional factor model for dates is shown in Panel B, while Panels C and D show the core values of the implied models for mutual funds and characteristics, respectively. The core values have the familiar patterns of eigenvalue spectra in two-dimensional factor models. The largest core value is dominant, especially so for the spectrum of the implied model for mutual funds in Panel B, followed by some smaller but significant values, before trailing off toward zero.

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 9 shows the heatmap of the (10×34)-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 1,342 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 1,342 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 (25×9) 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 3. Blue blocks have positive values in the second component and blocks in red have negative values. Since the heatmap in Figure 3 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 (in absolute value), *bm, sp, adjbm, bidask, turn* have positive weights, and *dp, me, op, dvol* have negative weights. Some characteristics are signifi-

cant for many components, *e.g. bm, dp, turn*, while others are represented in few components, *e.g. psliq, ms*.

3.4. Representative funds/characteristics

As shown in Section 2.5, the elements of the Tucker core tensor can be interpreted as (10,25,9) "representative" time/fund/characteristic observations. I follow the example given in that section and compute the time-series of the (25×9) representative 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 representative mutual fund with representative characteristic k_3.

Figure 10 plots the time-series of eight representative fund/characteristic combinations. The first column shows the first four "diagonal" combinations, and the second column shows the first four "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 representative fund/characteristics decrease by row. The first representative fund/characteristic ranges from 2.47 to 2.54, while the second, third, and fourth diagonal combinations range from 0.26 to 0.35, 0.22 to 0.26, and 0.04 to 0.09, respectively. Hence the first diagonal representative 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 representative mutual fund. Figure 11 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 representative mutual funds, *i.e.*, $k_2 = 1,2,3$, while the right column plots the first representative 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 representative 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 representative 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 representative mutual fund and the time-series of *bm*-means across funds are almost identical. In other words, the book-to-market ratio of first representative funds can be interpreted as a mean book-to-market factor. The higher-order representative 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 representative mutual funds. The levels of the first funds are larger than those of the other funds, and the first representative funds are linked to the cross-sectional means. This is also true for the other 21 characteristics that are not plotted.

The contributions of representative 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 12 shows the fit for momentum of the fund shown in the left plot of Panel D in Figure 7.[17] 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

$$\hat{\boldsymbol{\mathcal{X}}}(\overline{K}_1,25,9) = \sum_{t=1}^{\overline{K}_1}\sum_{m=1}^{25}\sum_{c=1}^{9} g_{tmc}\mathbf{v}_{1t}\circ\mathbf{v}_{2m}\circ\mathbf{v}_{3c}. \tag{31}$$

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

[17] The fund is the "Energy Select Sector SPDR" (wficn 500490, ticker XLE) and is chosen for illustration only. However, the results shown in Figure 12 are representative of most funds in the sample.

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 $\mathcal{G}$, loading matrices $\mathbf{V}^{(i)}$ and the fit across subsamples. Since the variance of the subsample data varies, I report the pseudo-R^2, $1-\text{MSE}(\hat{\mathcal{X}})/\text{Var}(\mathcal{X})$ instead of the $\text{MSE}(\mathcal{X})$. Since the subsamples have shorter time spans, I reduce the number of mode-1 components from 10 to sevenso that $(K_1,K_2,K_3) = (7,25,9)$.

The overall fits in the split samples are similar to the fit in the complete data set. The pseudo-R^2 $1-\text{MSE}(\hat{\mathcal{X}})/\text{Var}(\mathcal{X})$ for the first subsample is 0.95 and 0.94 for the second subsample. Panel A of Figure ?? shows the largest 25 elements of the core tensors $\mathcal{G}$ for the whole sample (black), the first half (orange), and the second half (blue) on a log-10 scale. The pattern of core values is remarkably stable across the samples. Tn all three cases, the $(1,1,1)$ elements of $\mathcal{G}$ are by far the largest core values followed by the $(1,2,2)$ and $(1,3,3)$ elements while the remaining core values are considerably smaller.

Figure 14 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.

In addition to splitting the sample, I also estimate the Tucker decomposition in recursive 10-year subsamples. The pseudo-R^2 for the 23 separate windows, shown in Panel B of Figure ??, are between 95% and 96%, which confirms a stable fit across the sample period. Panel C plots the 25

largest elements of each subsample $\mathcal{G}$so that each line corresponds to one window. The pattern of the core tensors is remarkably stable without any outliers, especially since the sample size of each subsample window is relatively short. Each panel of Figure 15 shows loading matrices $\mathbf{V}^{(i)}$ for all 23 subsample windows and confirms the conclusions drawn from Figure 14. The loadings matrices are almost identical across the sample with the exceptions of the second and third "time" factors, which have only a modest effect on the fit of the model

4. 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 representative 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 "representative" 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 "representative" mutual fund for a given characteristic is highly correlated with the time-series of cross-sectional means of the characteristic. Higher-order "representative" 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.

Carroll, J. Douglas and Jih Jie Chang (1970) "Analysis of Individual Differences in Multidimensional Scaling via an *N*-Way Generalization of "Eckart-Young" Decomposition," *Psychometrika*, 35 (3), 283–319, 10.1007/BF02310791.

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.

Harshman, Richard A. (1970) "Foundations of the PARAFAC Procedure: Models and Conditions for an "Explanatory" Multimodal Factor Analysis," *UCLA Working Papers in Phonetics*, 16 (10), 1–84.

Hitchcock, Frank L. (1927) "The Expression of a Tensor or a Polyadic as a Sum of Products," *Journal of Mathematics and Physics*, 6 (1-4), 164–189, 10.1002/SAPM192761164.

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, J.H., J.H. Wright, and M. 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 (*e_g*) Long-term earnings (*e_lt_g*) Book value (*b_g*) Cash flow (*cf_g*), Sales (*s_g*)
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 (*dvol*)
Other	Market cap (*me*) Operating profitability (*op*) Investment (*inv*) Quality (*qual*)

Table 2: Sample statistics

	Mean	Std. Dev.	25% pct.	50% pct.	75% pct.
TNA ($ mil.)	2263.99	8001.44	195.00	587.40	1722.85
No. of stocks	192.20	342.92	56.29	87.05	164.36
Market β	1.01	0.23	0.95	1.00	1.06
Return (% *p.a.*)	10.19	4.10	7.95	10.12	12.24
4-factor α (% *p.a.*)	-0.18	0.89	-0.53	-0.16	0.19

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

	Tucker(10, 25, 9)	CP(12)
	Panel A: Moments	
Mean	0.000	0.000
Median	0.000	0.001
Std. Dev.	0.184	0.209
Skew	0.052	−0.023
Kurt.	3.243	3.263
	Panel B: Percentiles	
min.	−1.782	−1.907
0.005	−0.590	−0.677
0.05	−0.292	−0.332
0.25	−0.098	−0.111
0.75	0.097	0.112
0.95	0.293	0.329
0.995	0.609	0.671
max.	1.888	2.102

Notes: The table reports statistics of the distributions of errors of the Tucker(10, 25, 9) and CP(12) models. 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(10, 25, 9) model as well as the corresponding (k_1,k_2,k_3) indices. The sample period is 2010Q3 to 2018Q4.

Figure 1: Tucker decomposition $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \times_3 \mathbf{V}_3$

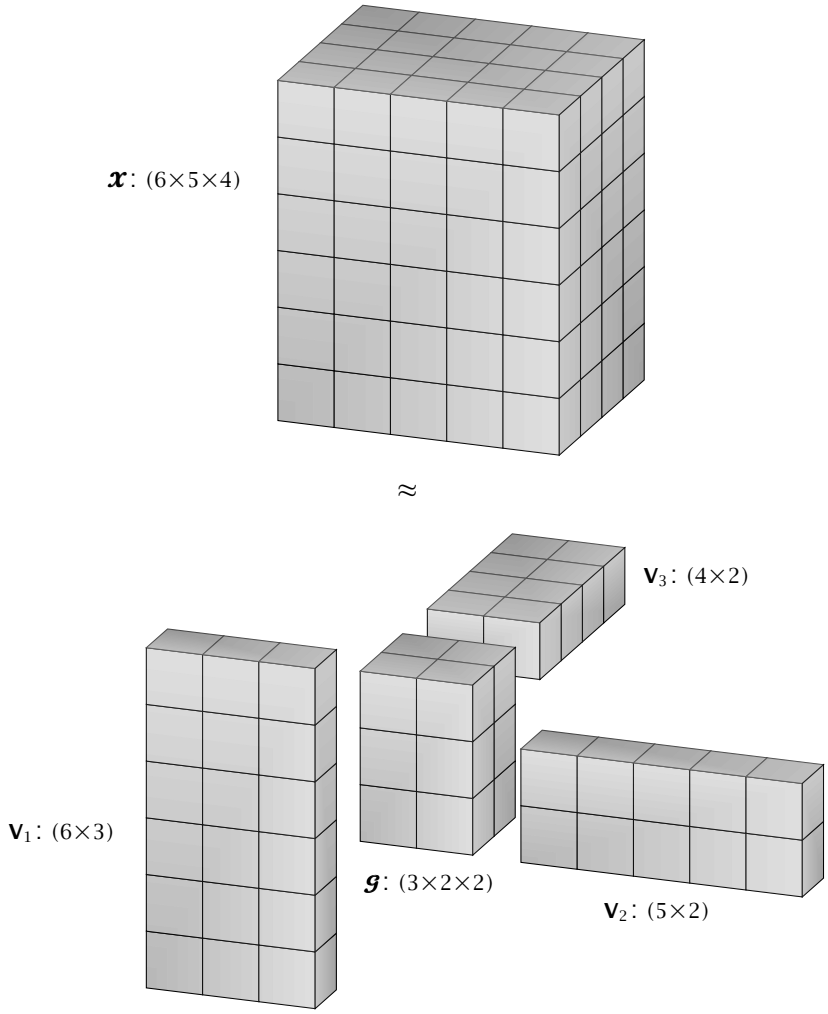
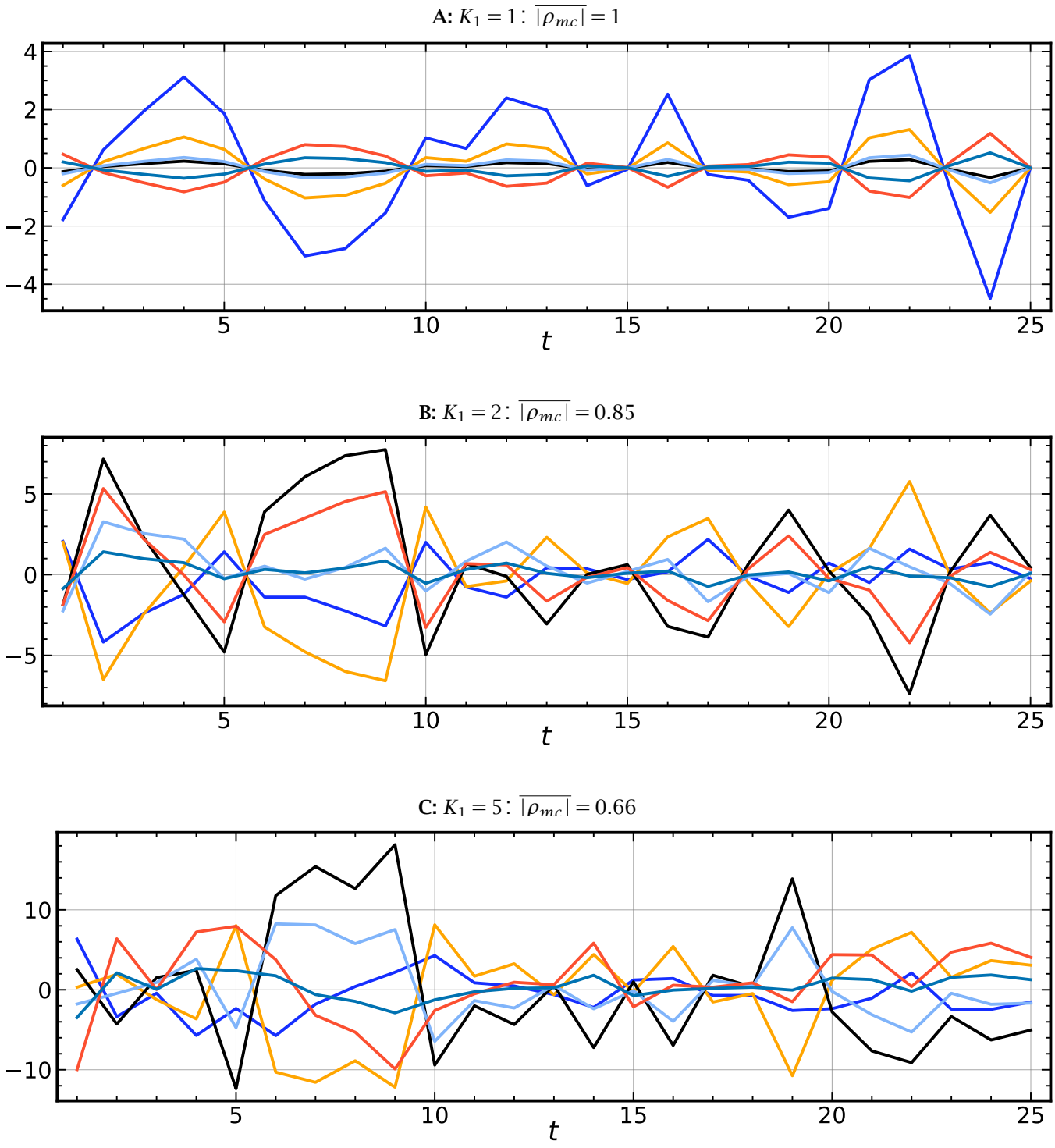
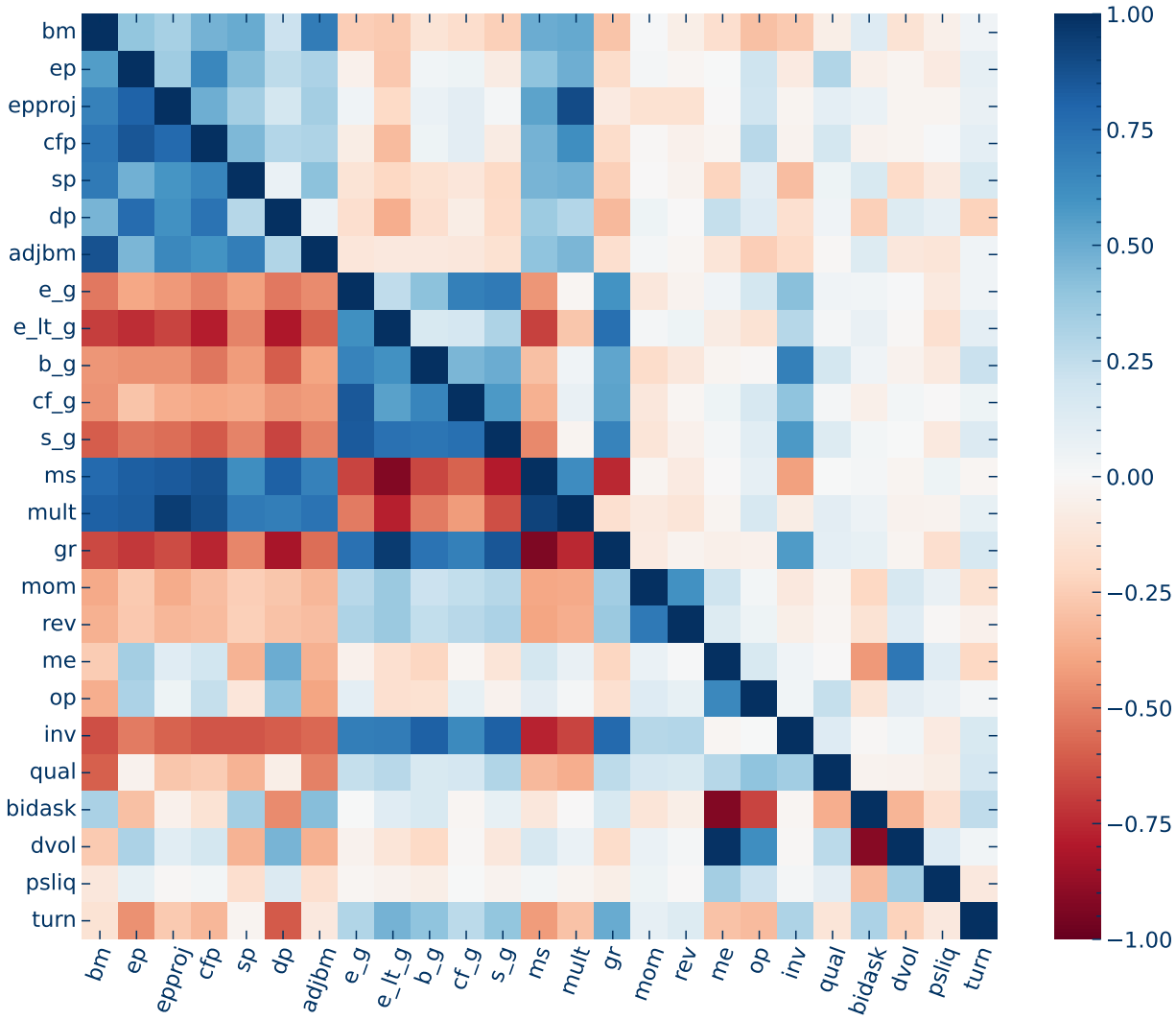


Figure 2: Time series fibers $\mathbf{x}_{(\bullet mc)}$ in $(K_1,1,2)$ Tucker model



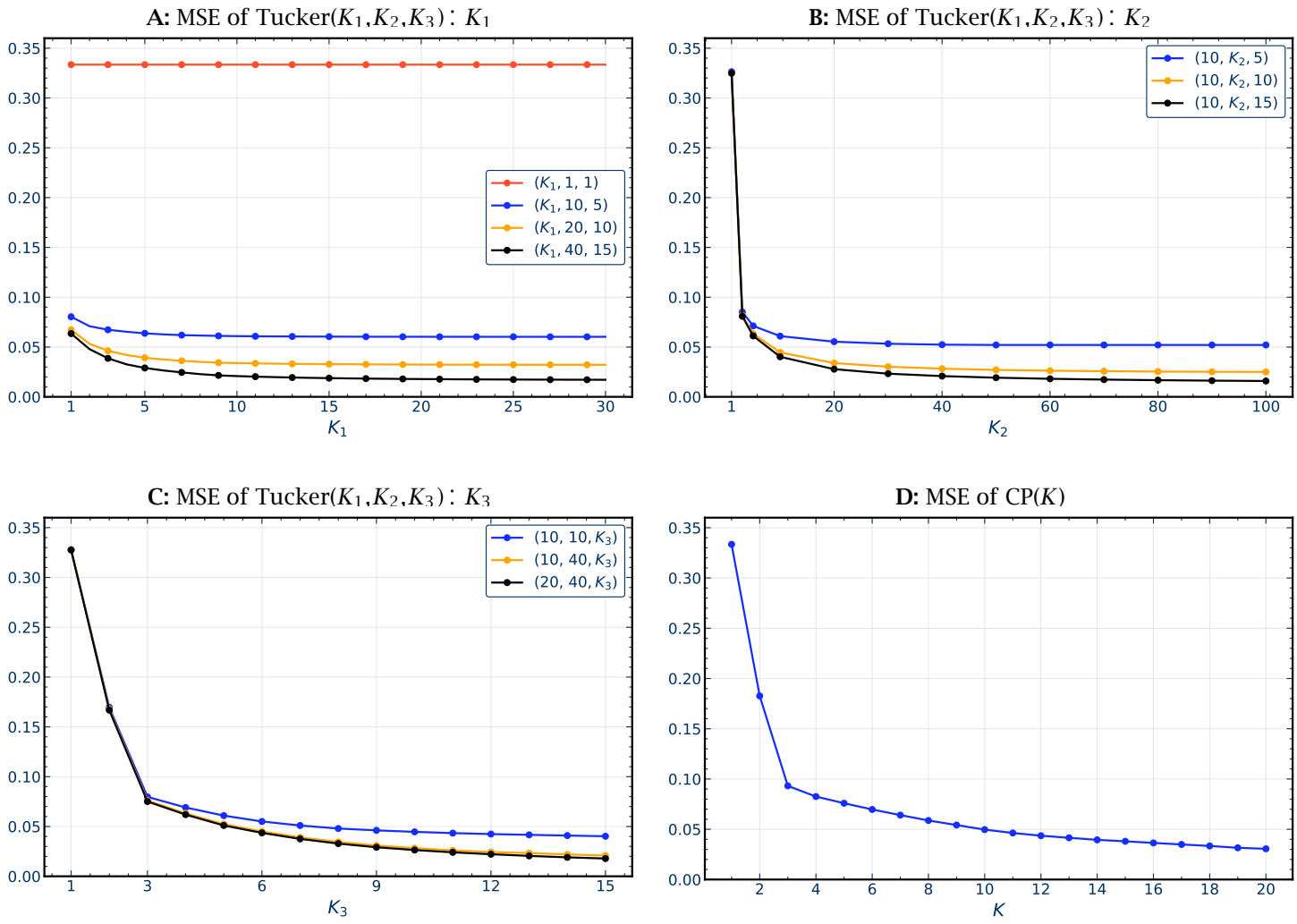
Notes: The figure plots fibers (vectors) $\mathbf{x}_{(\bullet mc)}$ in a simulated $(25\times10\times5)$ Tucker model with $(K_1,1,2)$ components for $K_1 = 1,2,5$. Each mode-t fiber is a (25×1) vector, The panel shows six lines corresponding to a subset of (m,c) combinations.

Figure 3: Cross-correlations of characteristics



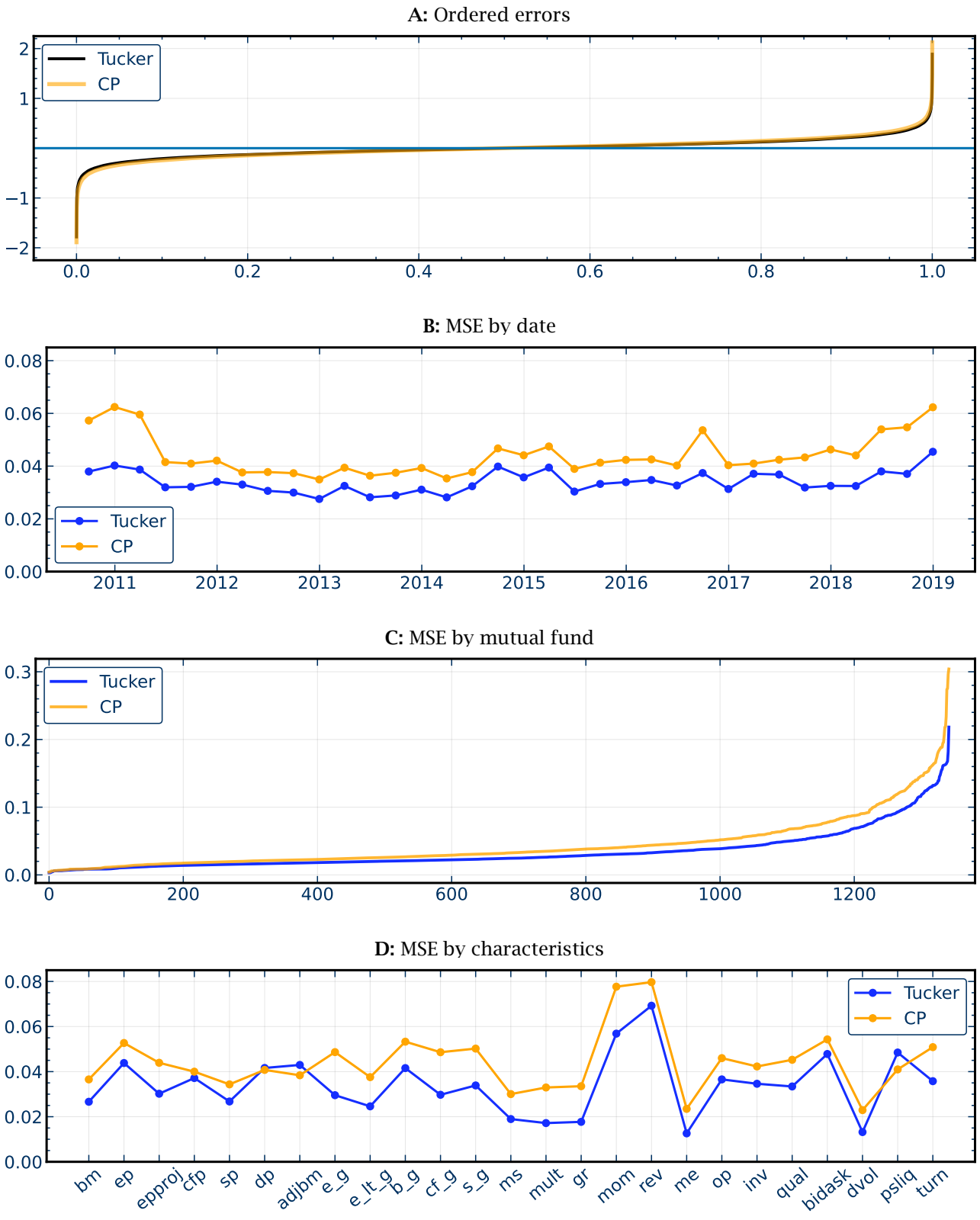
Notes: The figure shows the heatmap of pairwise correlations of mutual fund characteristics. I compute times-series characteristic correlations for all mutual fund pairs as well as cross-sectional correlations for all date pairs. The upper triangle shows the means across fund correlations and the lower triangle shows the mean across date correlations. The sample period is 2010Q3 to 2018Q4.

Figure 4: MSE of Tucker and CP models



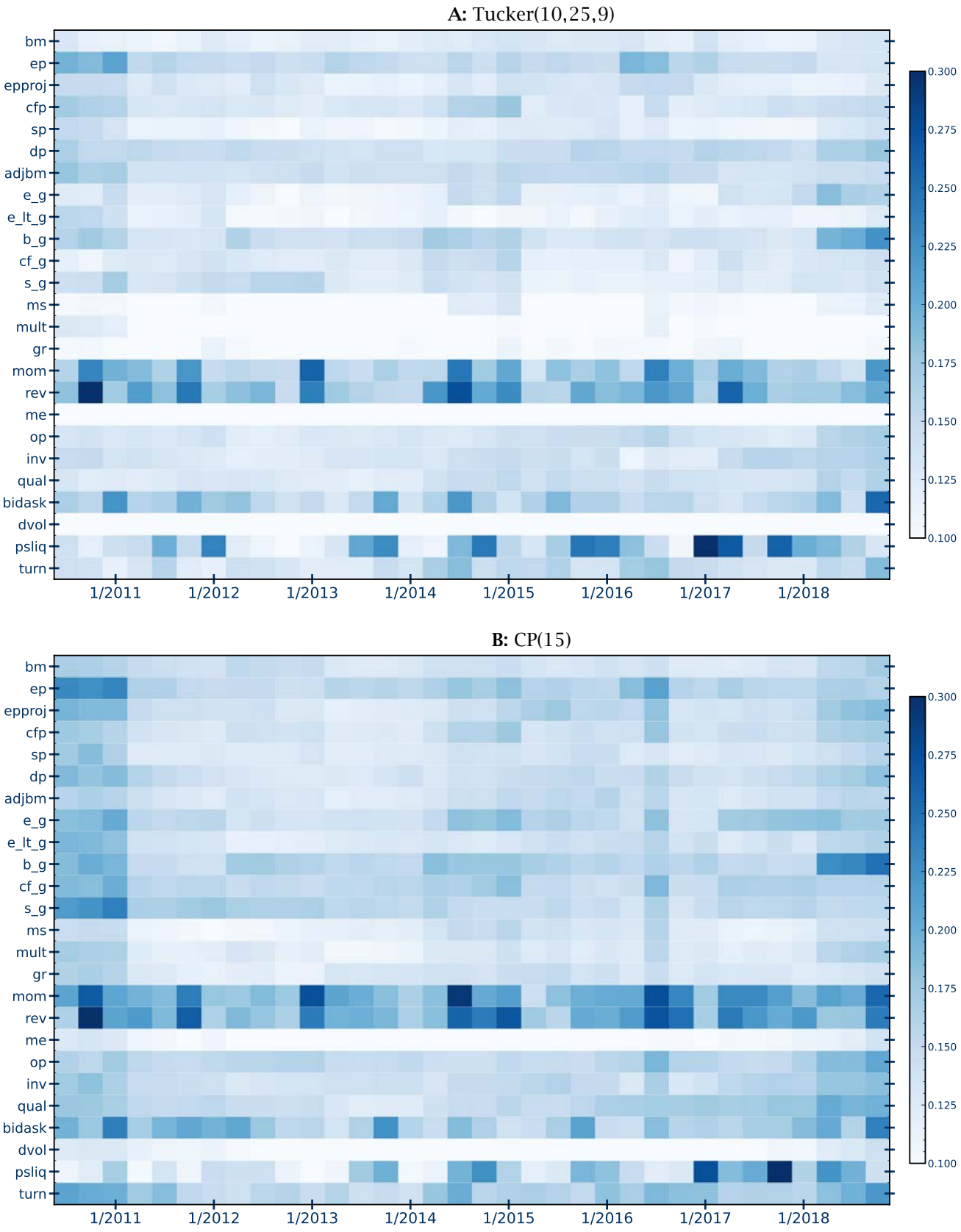
Notes: The figure plots mean-squared errors of Tucker(K_1,K_2,K_3) models in Panels A, B, and C, CP(K) models in Panel D. The sample period is 2010Q3 to 2018Q4.

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



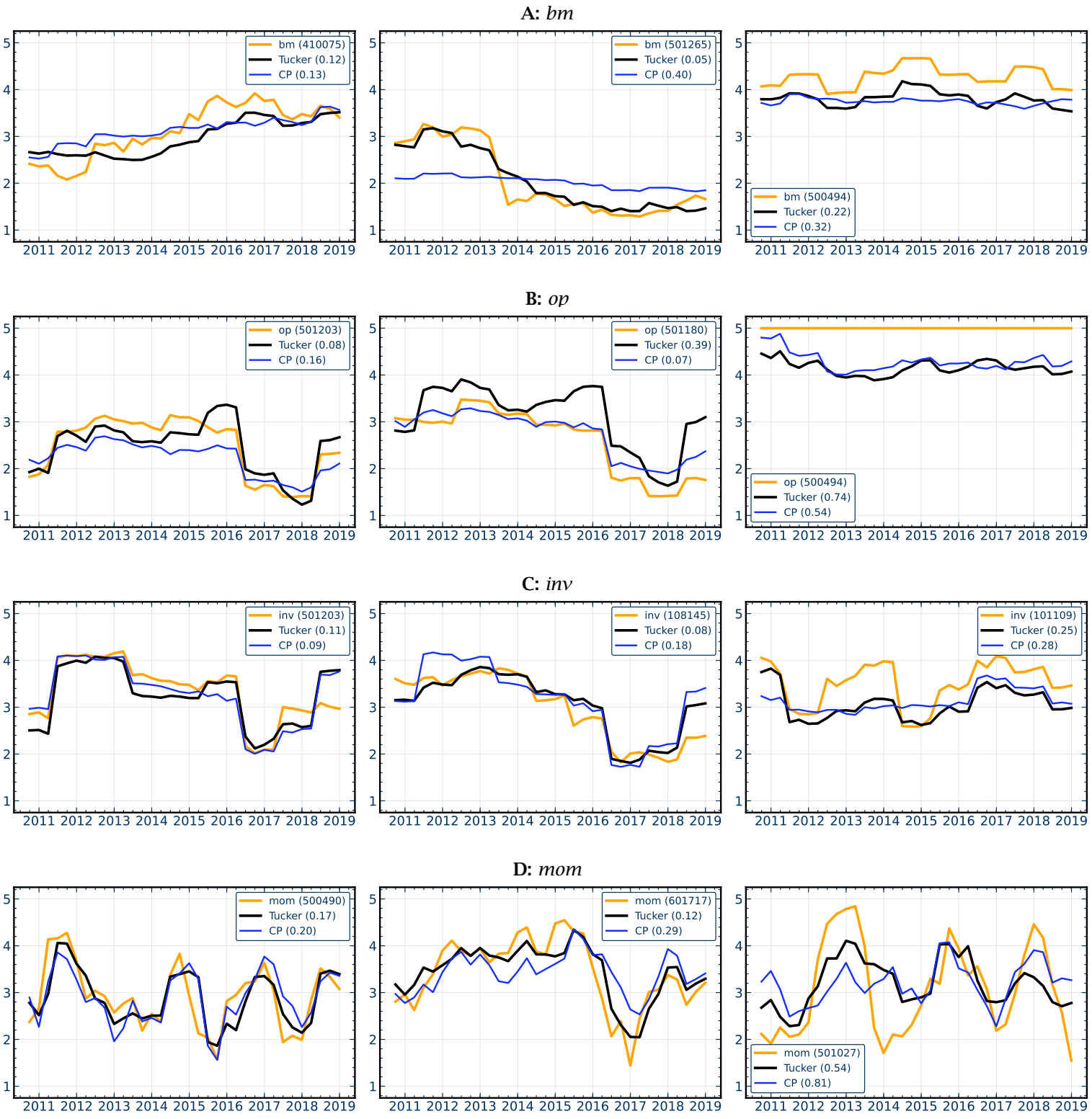
Notes: Panel A plots the sorted errors of the Tucker(10, 25, 9) and CP(12) models. Panels B, C, and D shows the means of MSE by dates, mutual funds, and characteristics, respectively. The sample period is 2010Q3 to 2018Q4.

Figure 6: Errors by data/characteristic



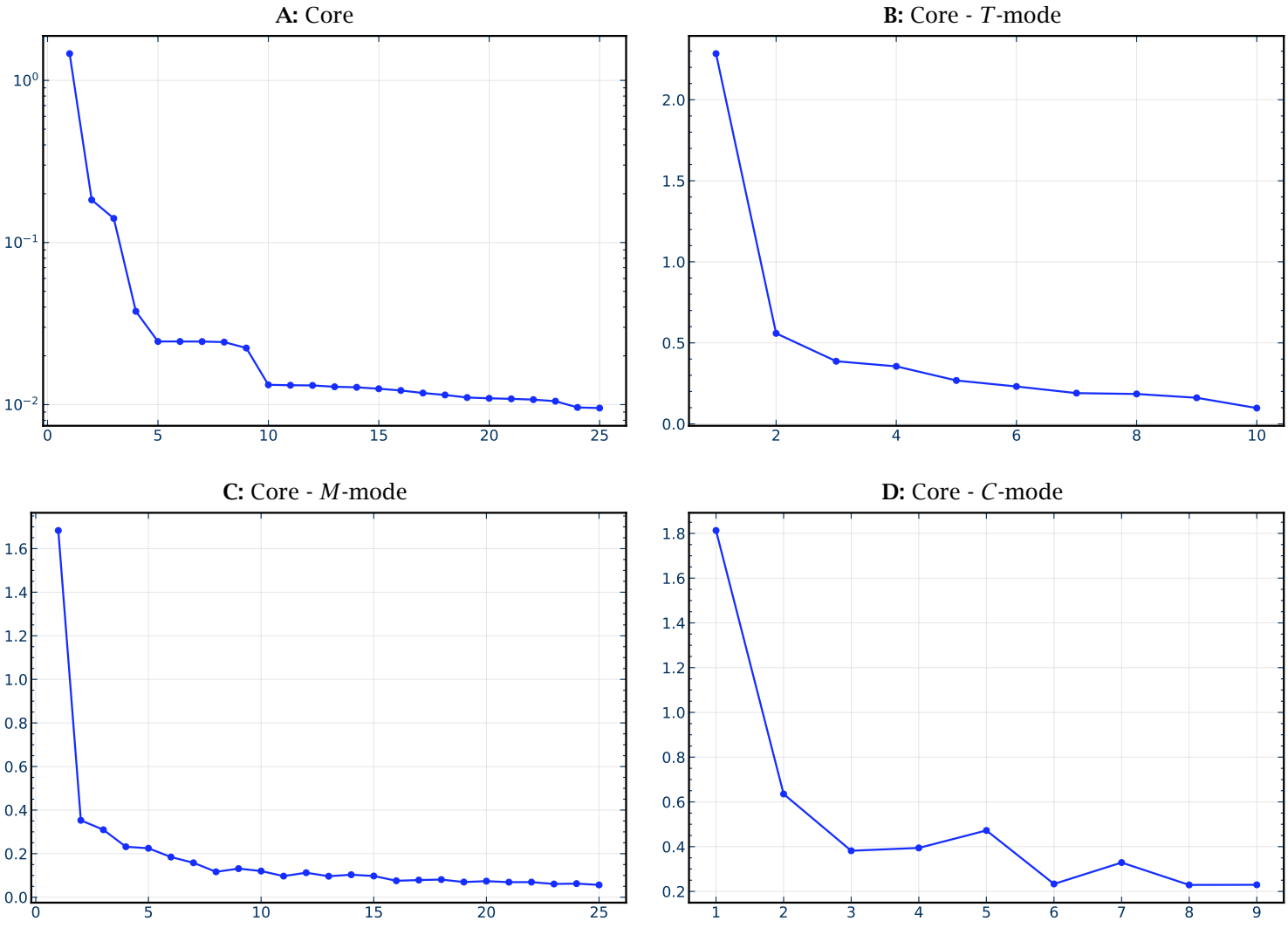
Notes: The figure shows the heatmap of means across mutual funds of the MSE for each date and characteristic combination. The sample period is 2010Q3 to 2018Q4.

Figure 7: Fit of *bm, op, inv*, and *mom* - Examples



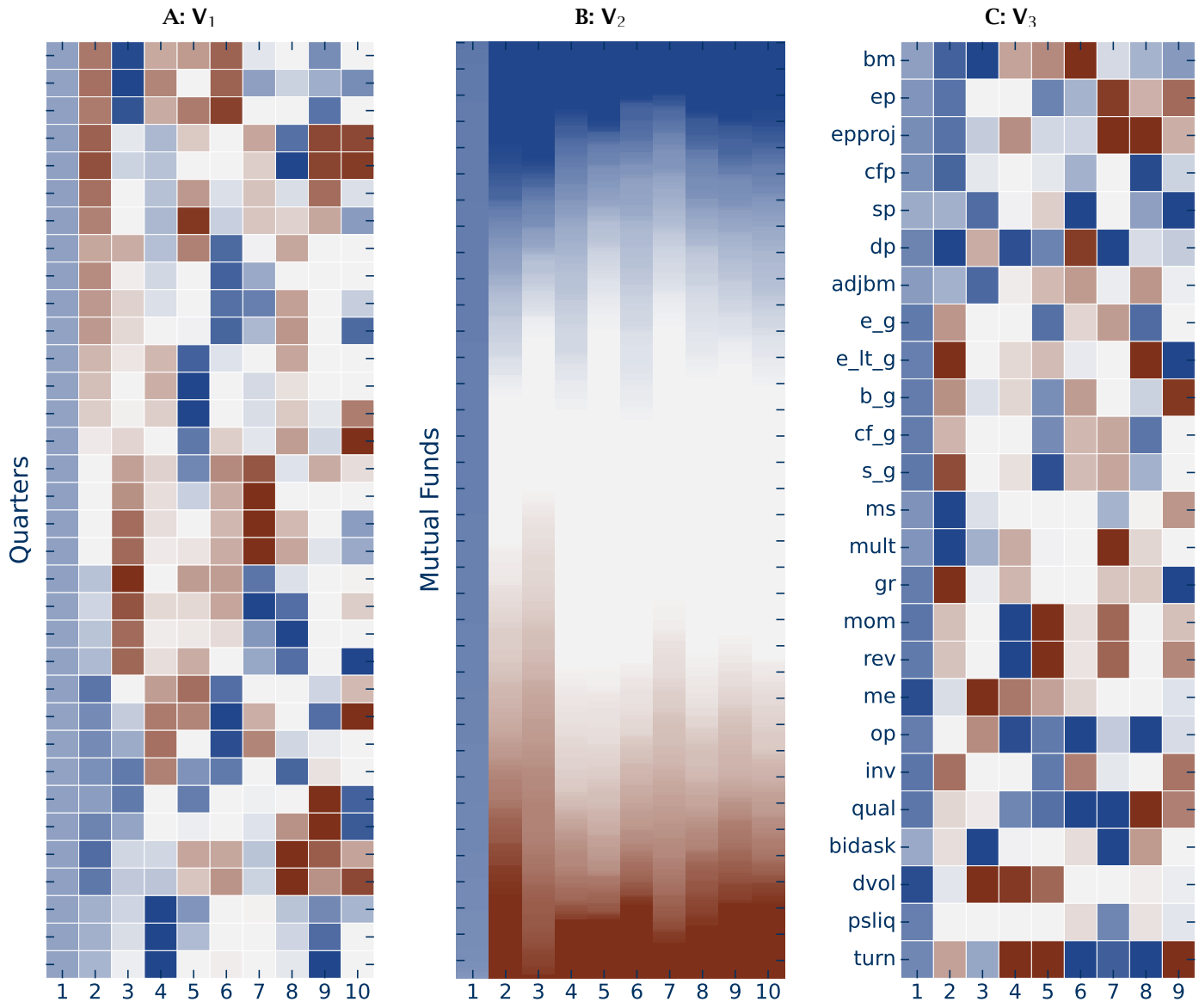
Notes: The figures shows time-series plots of the observed data and fitted values of the Tucker(10, 25, 9) and CP(12) models of the book-to-market ratio (Panel A), operating profitability (Panel B), investment (Panel C), and momentum (Panel D) of individual mutual funds. The legends include the wficn of the plotted fund and the MSE of the Tucker(9, 25, 10) and CP(12) models. The sample period is 2010Q3 to 2018Q4.

Figure 8: Tucker(10, 25, 9) core tensor



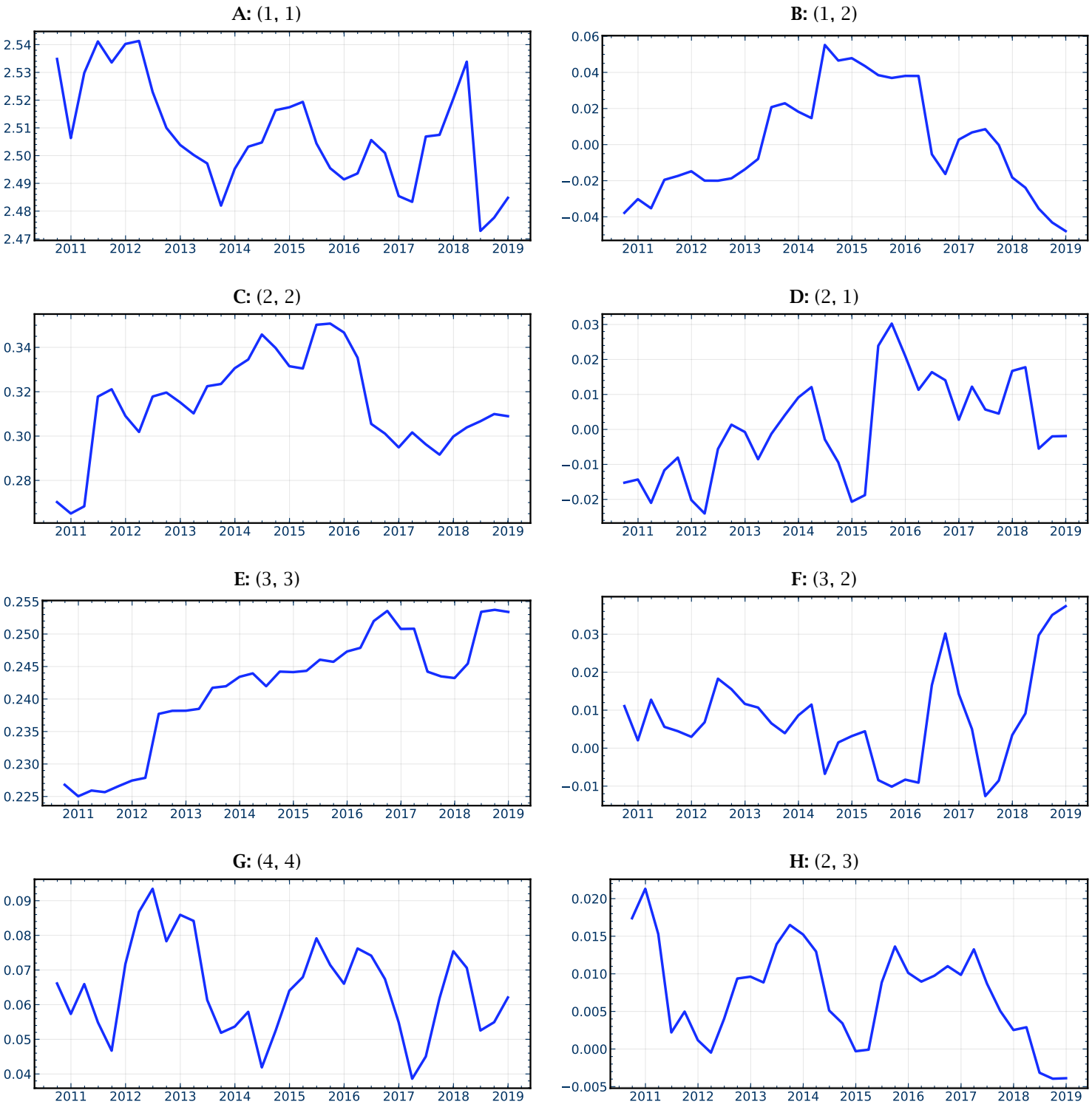
Notes: Panel A plots the 25 largest elements by the absolute value of the core tensor $\mathcal{G}$ of the Tucker(10, 25, 9) model on a log-scale. Panels B, C, and D plot the spectrums of the core values of the two-dimensional representations for dates, mutual funds, and characteristics. The sample period is 2010Q3 to 2018Q4.

Figure 9: Component matrices of Tucker(10, 25, 9) models



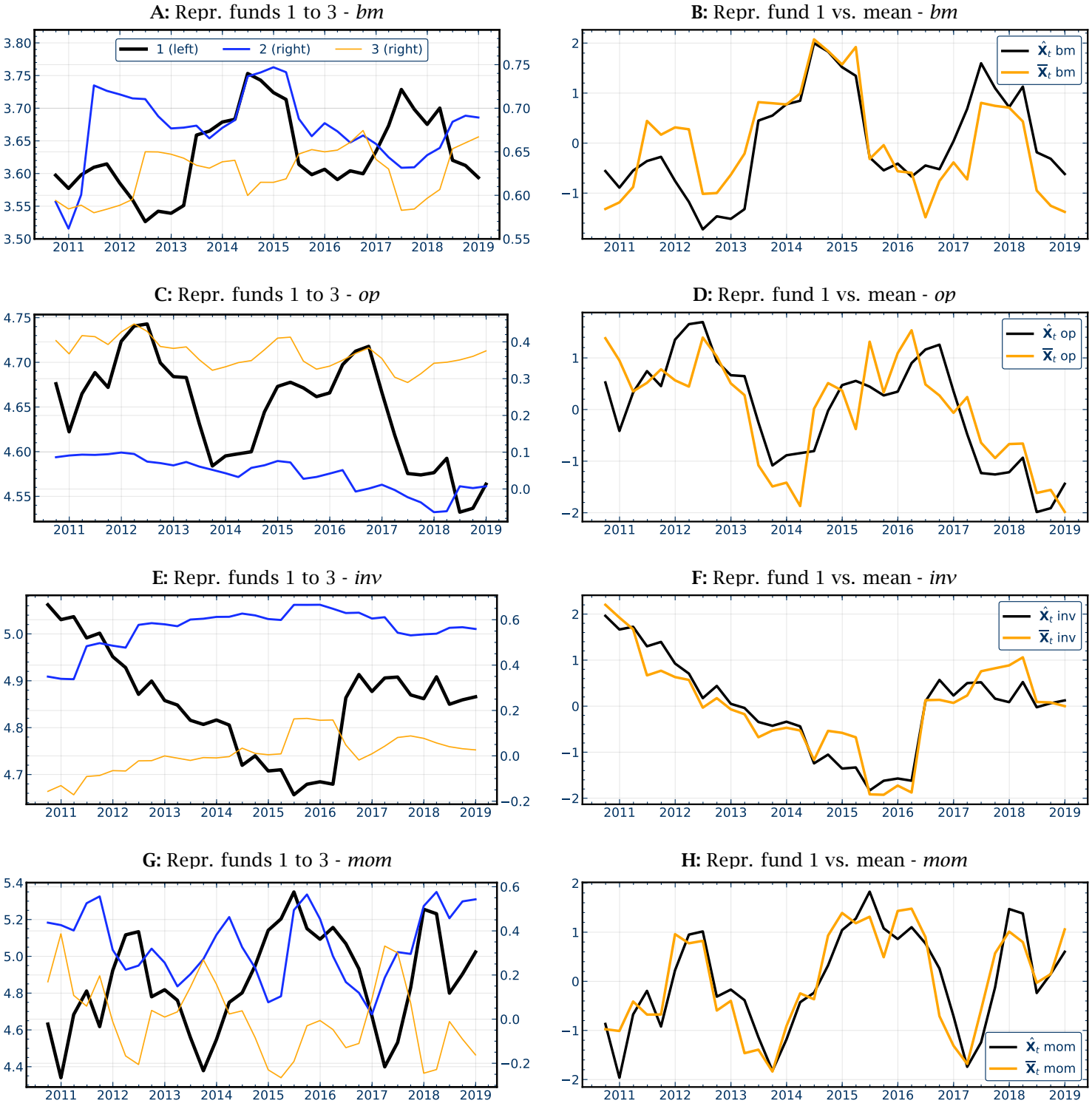
Notes: The figure shows heatmaps of the Tucker component matrices $\mathbf{V}_i$ of the Tucker(10, 25, 9) model. Positive values are shown in blue and negative values in red. Panel A shows the (34×10) component matrices $\mathbf{V}_1$, Each row corresponds to a quarter starting in 201Q3 at the top to 2018Q4 at the bottom. The columns correspond to the $K_1 = 10$ mode-1 components. The second component matrix $\mathbf{V}_2$ has 1,342 rows and 25 columns. Panel B shows the heatmap of the first 10 columns of $\mathbf{V}_2$. In order to visualize 1,342 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 (25×9)-dimensional matrix $\mathbf{V}_3$. The sample period is 2010Q3 to 2018Q4.

Figure 10: Representative funds/characteristics of Tucker(10, 25, 9) model



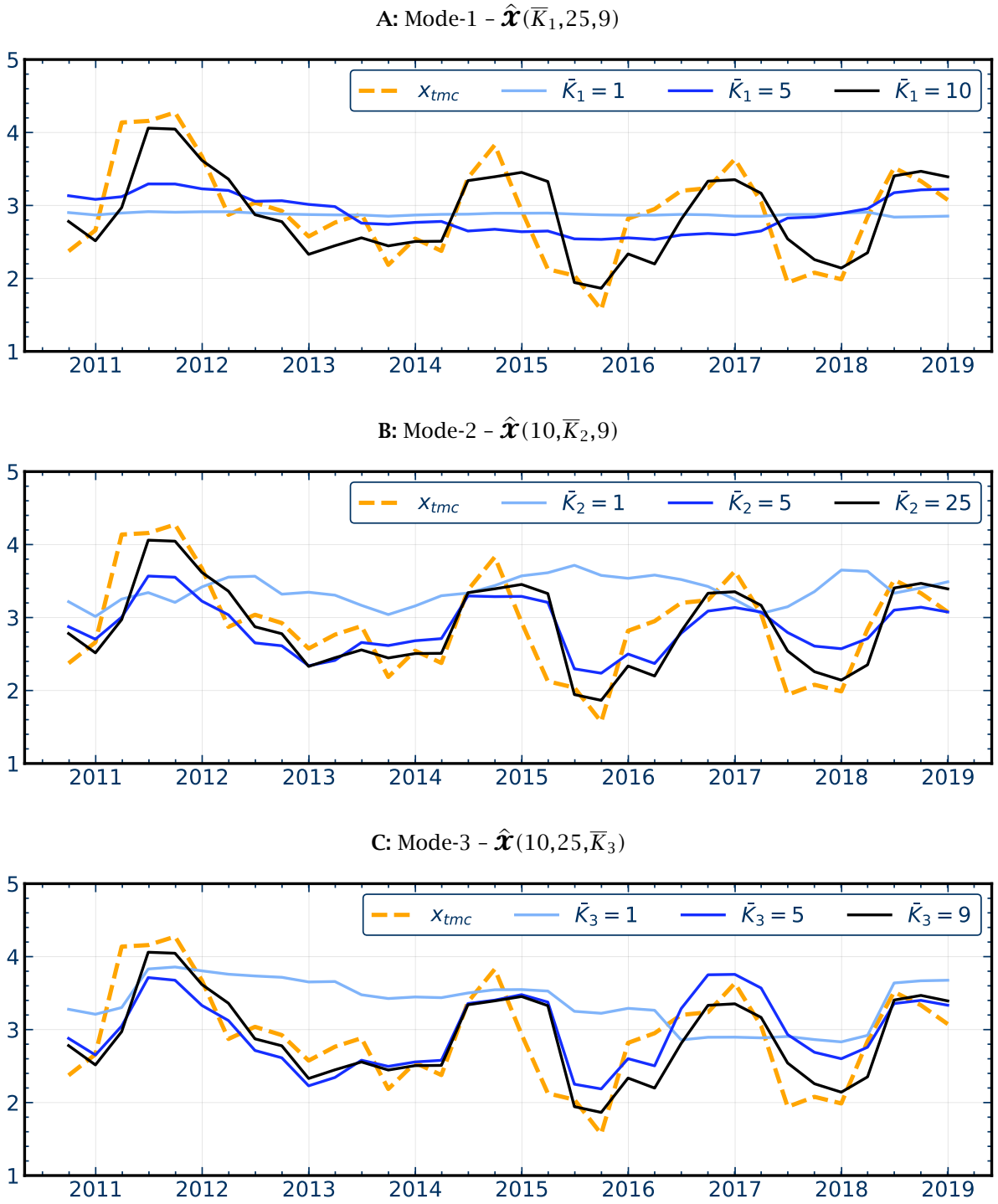
Notes: The figures shows time-series plots of the mode-1 fibers of the (34, 25, 9)-dimensional tensor $\boldsymbol{\mathcal{S}}_{(1)} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{V}_1$. The (k_2, k_3) panels plot the 34 time-series observations of the k_2-th mode-2 and k_3-th mode-3 fibers. The sample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

Figure 11: Characteristics of representative funds in Tucker(10, 25, 9) model



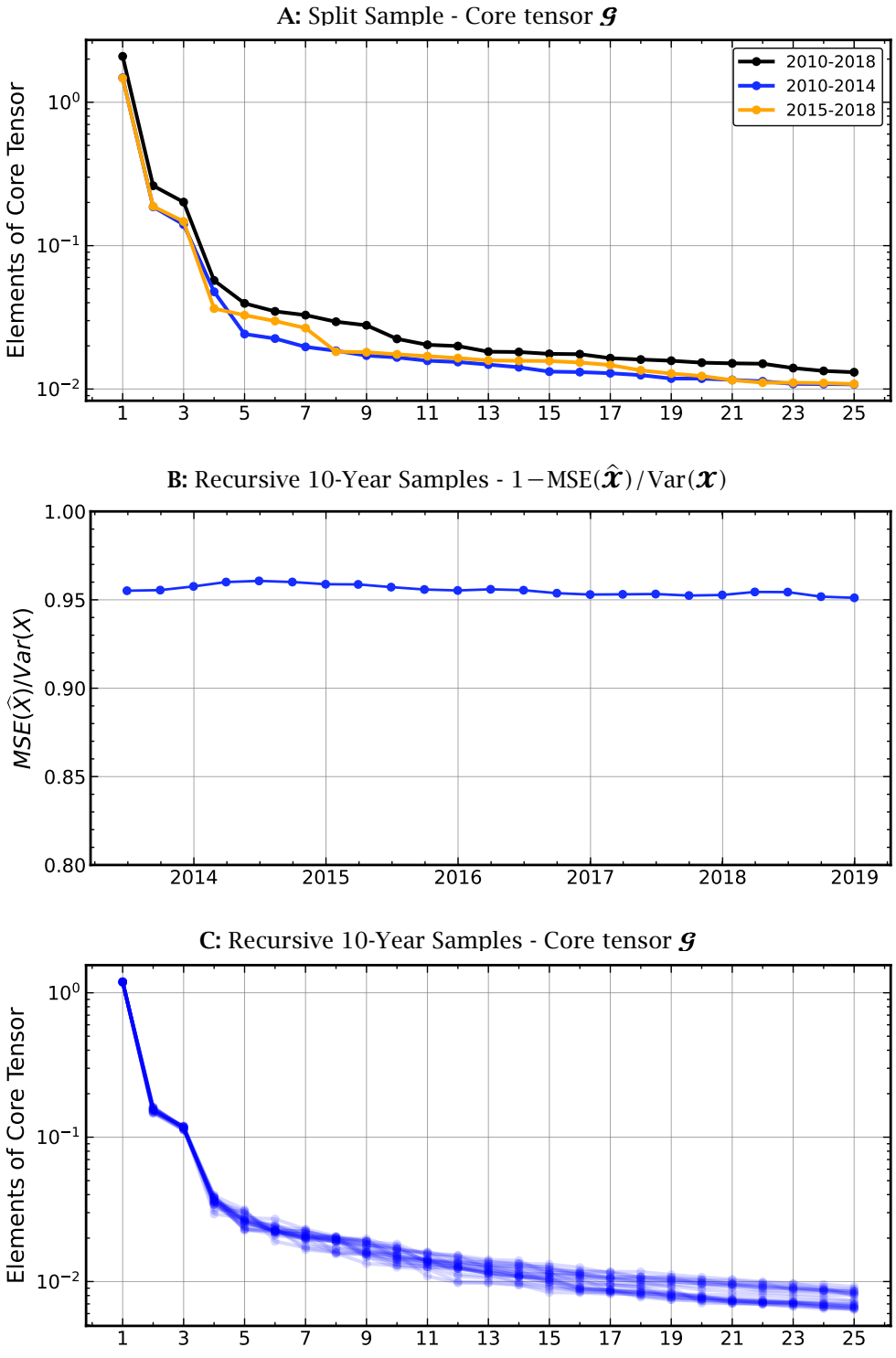
Notes: The figures shows time-series plots of the fibers of the (34, 25, 25)-dimensional tensor $\boldsymbol{\mathcal{S}}_{(13)} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{V}_1 \times_3 \mathbf{V}_3$ for book-to-market ratio (row 1), operating profitability (row 2), investment (row 3), and momentum (row 4). The left columns 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 columns 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 12: Contributions of components in Tucker(10, 25, 9) model



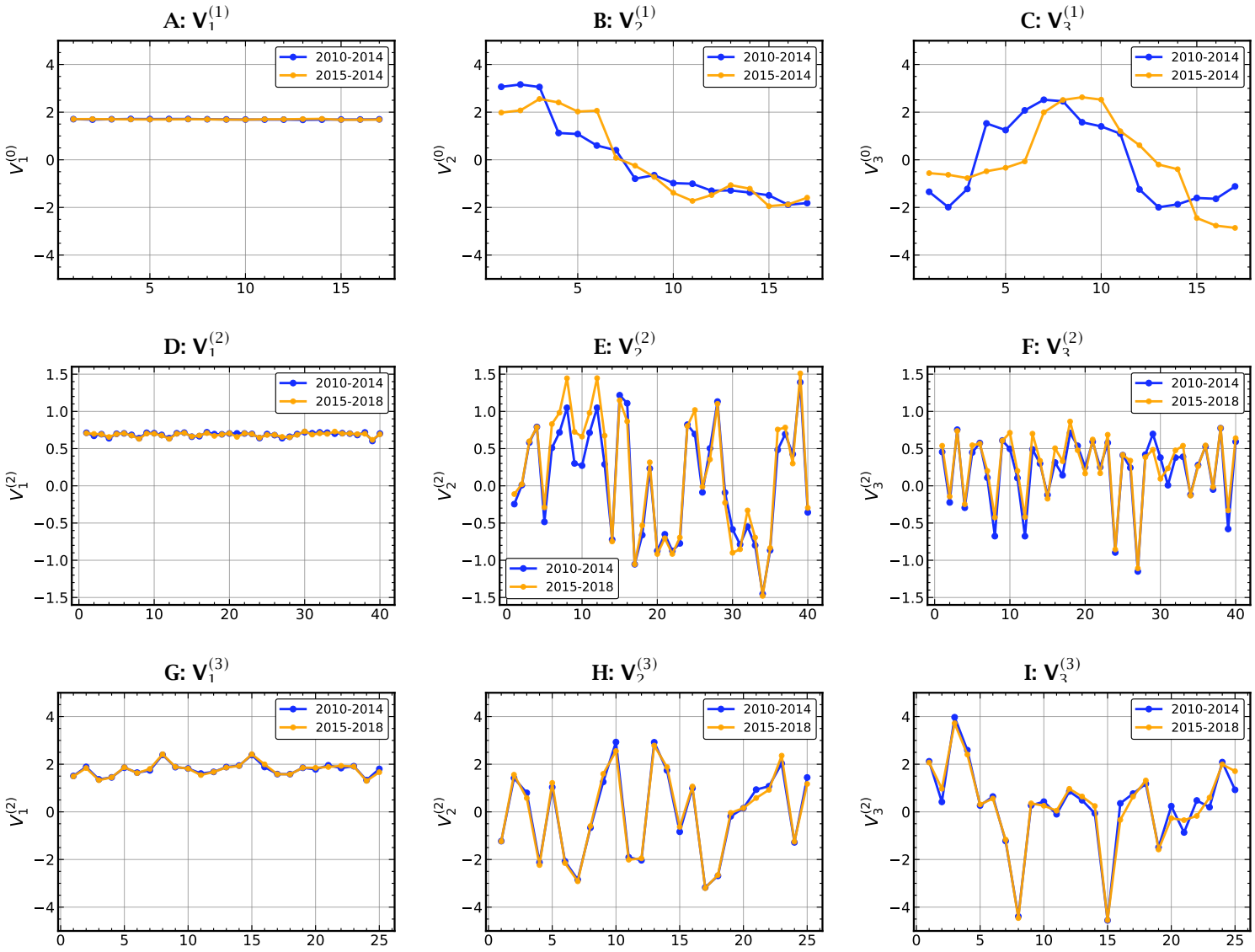
Notes: This figure shows cumulative components for momentum of a single mutual fund (wficn 500490) for the Tucker(10, 25, 9) model. The top panel plots the fitted time-series of Tucker($\overline{K}_1$,25,9) models defined in (31), *i.e.*, fixing $K_2 = 25$ and $K_3 = 9$ and varying the number of mode-1 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 13: Subsamples



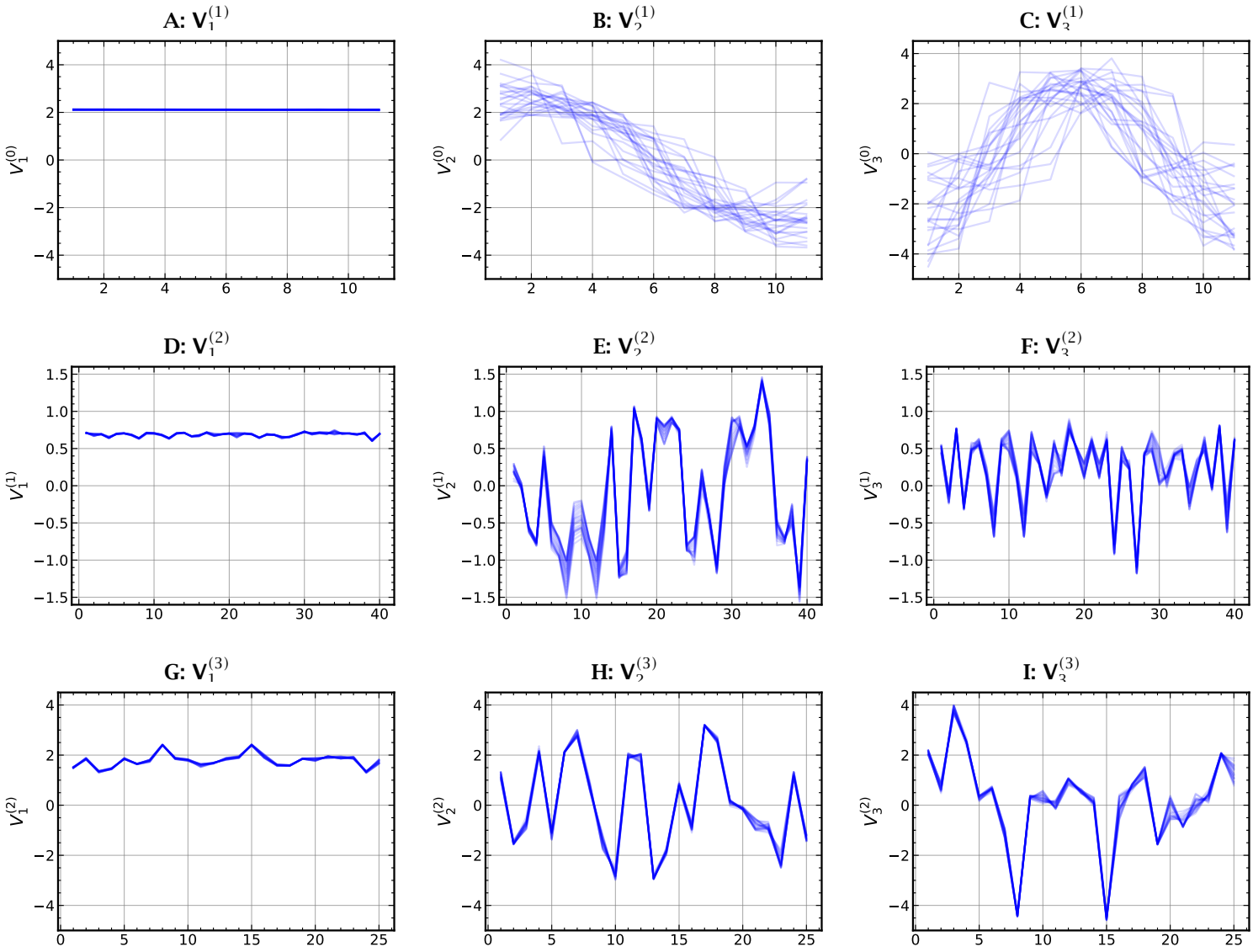
Notes: This figure plots the first three loading vectors of mode i,$\mathbf{V}_j^{(i)}$,$i = 1,2,3$,$j = 1,2,3$ of Tucker models that are estimated over two subsamples consisting of 17 quarters: 2010Q3-2014Q4 and 2015Q1-2018Q4. The Tucker models have $(K_1,K_2,K_3) = (7,25,9)$ components. The blue lines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

Figure 14: Split Samples - $V_j^{(i)}$



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 15: Recursive 10-Year Samples - $V_j^{(i)}$



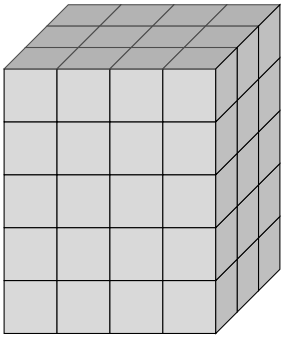
Notes: This figure plots the first three loading vectors of mode i,$\mathbf{V}_j^{(i)}$,$i = 2,3$,$j = 1,2,3$ of Tucker models that are estimated over sliding windows. Each window consists of 10 quarters so there are 23 subsamples. The Tucker models have $(K_1,K_2,K_3) = (7,25,9)$ components. Each panel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

Table A-1: Summary of tensor notation and operations

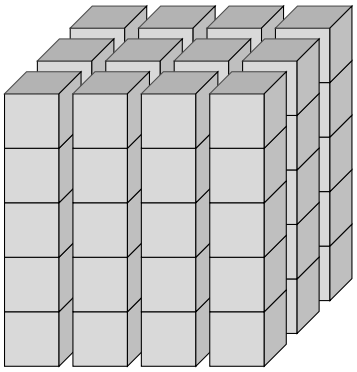
Operation	two-dimensional matrix	three-dimensional tensor	n-dimensional tensor
	$\mathbf{X} = [\mathsf{x}_{ij}]$	$\boldsymbol{\mathcal{X}} = [\mathsf{x}_{ijk}]$	$\boldsymbol{\mathcal{X}} = [\mathsf{x}_{i_1,\ldots,i_n}]$
norm	$\|\mathbf{X}\| = \sqrt{\sum_{ij} \mathsf{x}_{ij}^2}$	$\|\boldsymbol{\mathcal{X}}\| = \sqrt{\sum_{ijk} \mathsf{x}_{ijk}^2}$	$\|\boldsymbol{\mathcal{X}}\| = \sqrt{\sum_{i_1,\ldots,i_N} \mathsf{x}_{i_1,\ldots,i_n}^2}$
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	$\mathbf{U}_{1K} \mathbf{H}_K \mathbf{U}_{2K}^\top$ $\sum_{k=1}^K h_k \, \mathbf{u}_{1k} \mathbf{u}_{2k}^\top$	$\boldsymbol{\mathcal{G}} \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \times_3 \mathbf{V}_3$ $\sum_{k=1}^K g_k \mathbf{w}_{1k} \circ \mathbf{w}_{2k} \circ \mathbf{w}_{3k}$	$\boldsymbol{\mathcal{G}} \times_1 \mathbf{V}_1 \times_2 \cdots \times_n \mathbf{V}_n$ $\sum_{k=1}^K g_k \mathbf{w}_{1k} \circ \cdots \circ \mathbf{w}_{nk}$

Figure A-1: Tensor fibers and slices

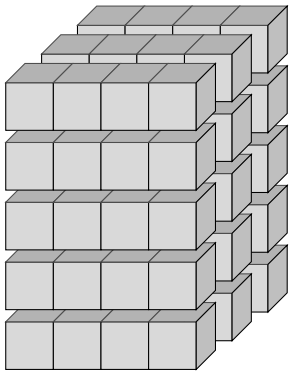
A: Tensor $\boldsymbol{\mathcal{X}}$: $(5 \times 4 \times 3)$



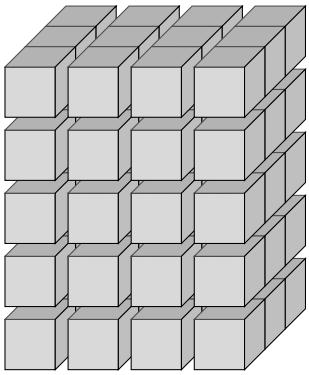
B: Mode-1 fibers $\mathbf{x}_{(\bullet mc)}$



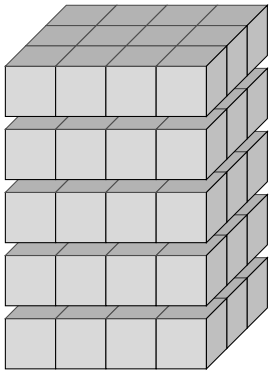
C: Mode-2 fibers $\mathbf{x}_{(t \bullet c)}$



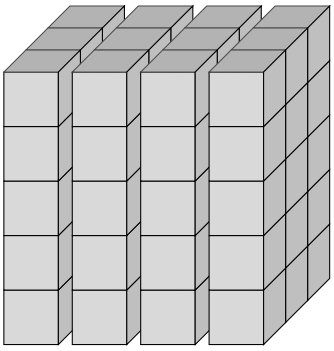
D: Mode-3 fibers $\mathbf{x}_{(tm \bullet)}$



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



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



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

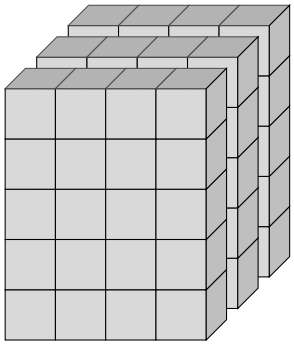
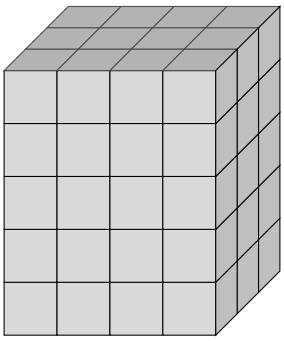
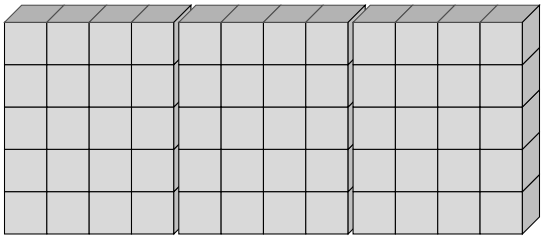


Figure A-2: Tensor as matrices

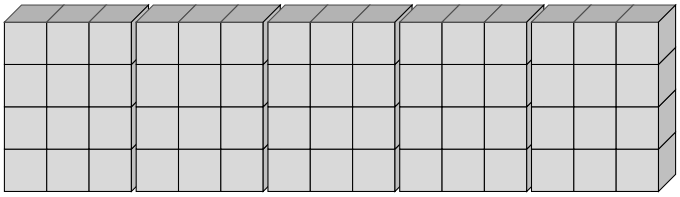
A: Tensor $\boldsymbol{\mathcal{X}}$: $(5 \times 4 \times 3)$



B: $\mathbf{X}_{(1)}$: (5×12)



C: $\mathbf{X}_{(2)}$: (4×15)



D: $\mathbf{X}_{(3)}$: (3×20)

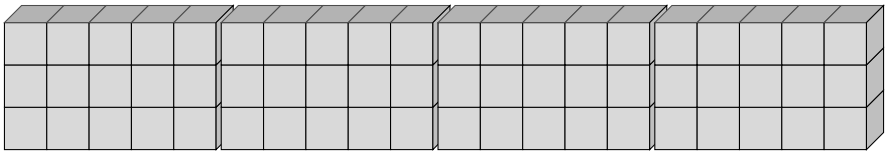


Figure A-3: Tensor multiplication

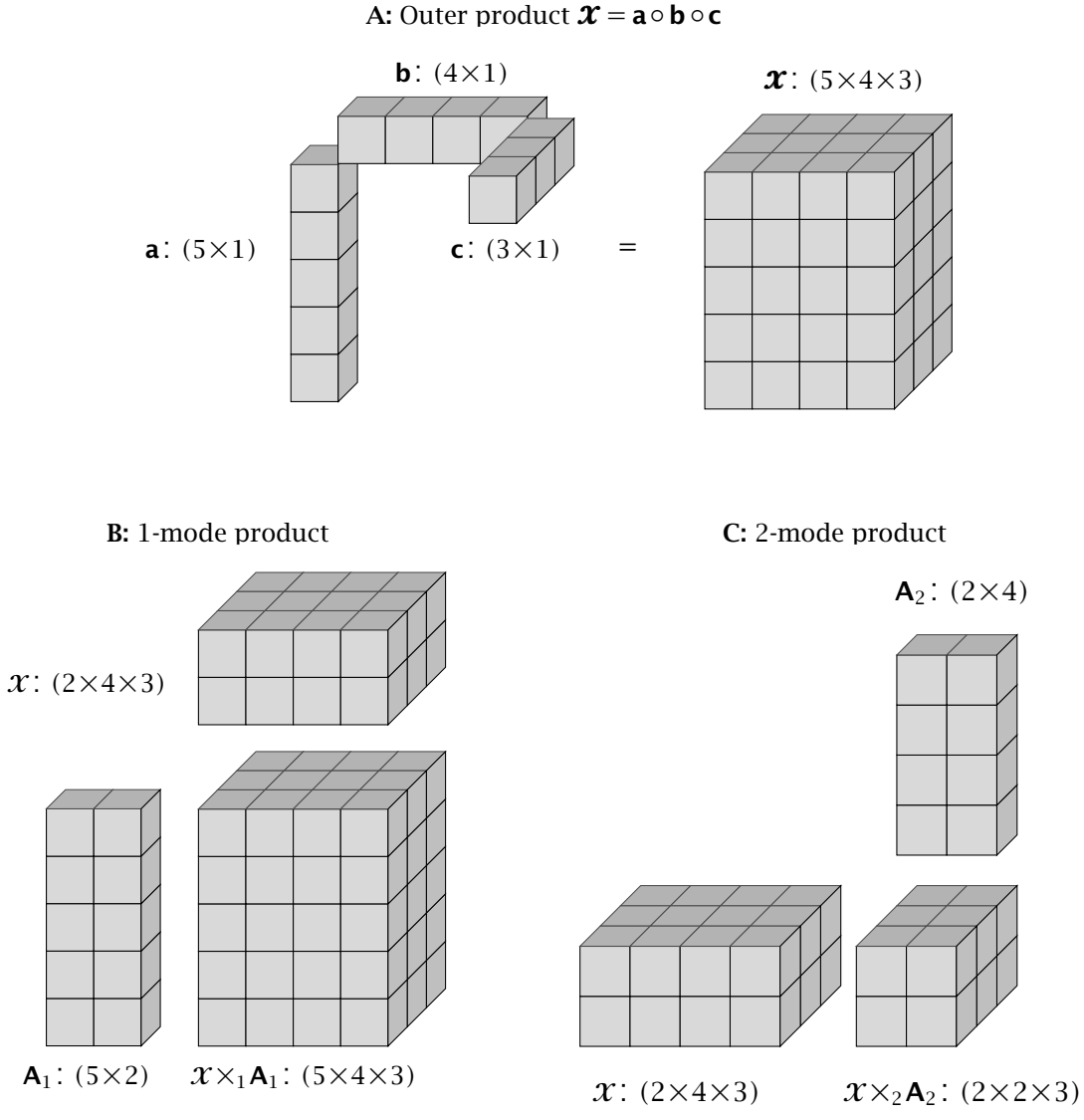


Figure A-4: Tucker model as two-dimensional factor model

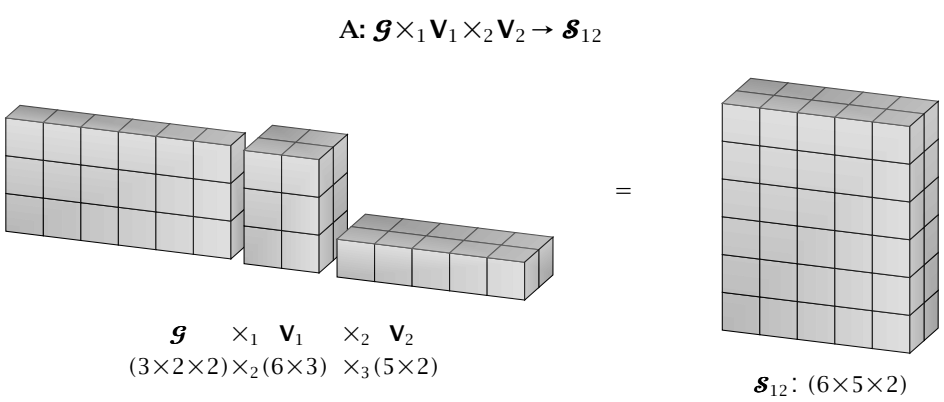


Figure A-5: CP(K) decomposition

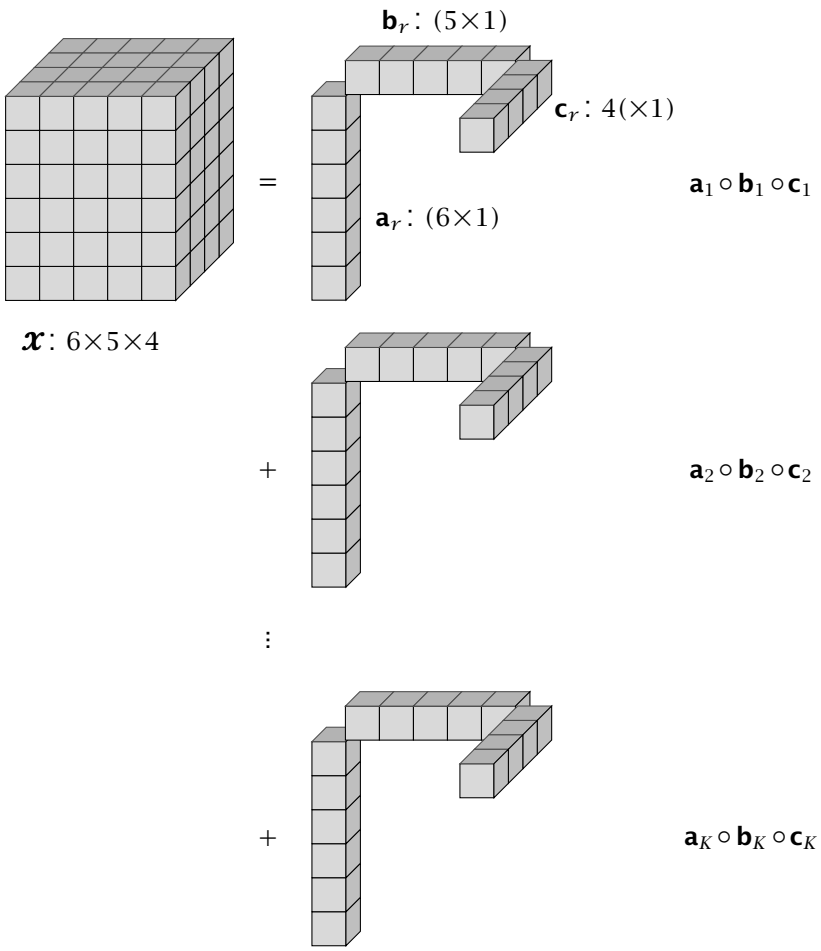
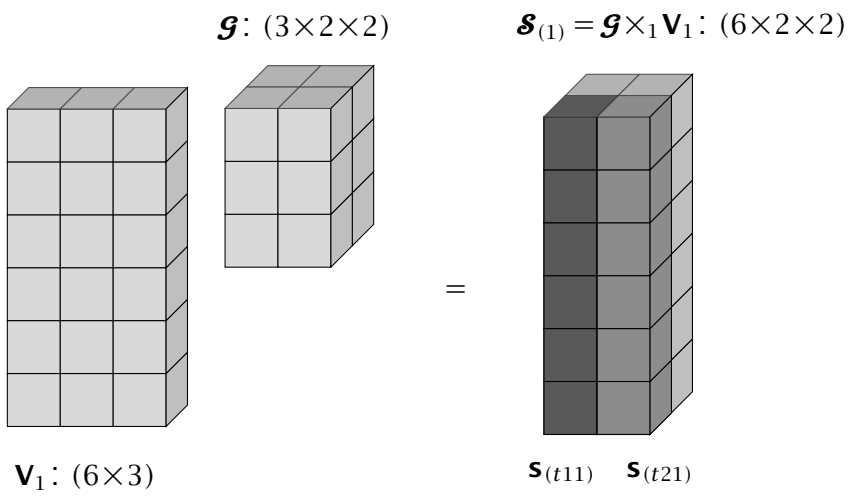
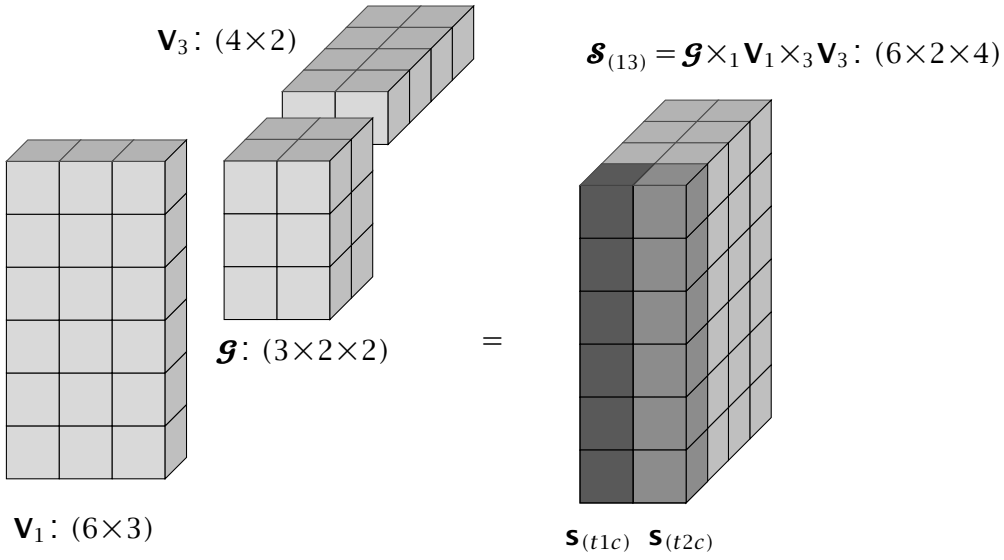


Figure A-6: Tucker model: Intuition

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



B: "Representative" mode-2 (M)