3D-PCA: Factor Models with Restrictions

Martin Lettau[1]

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

Abstract

This paper proposes latent factor models for multidimensional panels called 3D-PCA. Factor weights are constructed from a small set of dimension-specific building blocks, which give rise to proportionality restrictions of factor weights. While the set of feasible factors is restricted, factors with long/short structures often found in pricing factors are admissible. I estimate the model using a 3-dimensional data set of double-sorted portfolios of 11 characteristics. Factors estimated by 3D-PCA have higher Sharpe ratios and smaller cross-sectional pricing errors than models with PCA or Fama-French factors. Since factor weights are subject to restrictions, the number of free parameters is small. Consequently, the model produces robust results in short time series and performs well in recursive out-of-sample estimations.

Keywords: Factor models, factor zoo, PCA, SVD, Tucker decomposition, tensors

JEL classification: G0, G12, C38

This draft: December 10, 2024

[1]I thank Ben Hebert, Ali Kakhbod, Valentina Raponi, and the seminar participants at Berkeley-Haas, Columbia, IESE, Johns Hopkins, Pompeu Fabra, Toulouse, and the CEMFI Big Data, INFORMS, and SITE conferences for their helpful comments.

1. Introduction

Factor models have been a central framework for modeling stock returns. Many asset pricing theories, such as the CAPM and Intertemporal-CAPM, imply factor structures for returns. The arbitrage pricing theory (APT) of Ross (1976) implies that risk premia are linked to common factors via no-arbitrage conditions, leading to statistical representations of approximate factor models (Huberman (1982), Chamberlain and Rothschild (1983)). Following these theoretical insights, Connor and Korajczyk (1986, 1988) proposed to estimate latent factors by Principal Component Analysis (PCA).

Recently, latent factor models have been used to address the "factor zoo" conundrum in asset pricing (Subrahmanyam (2010), Cochrane (2011)). This literature aims to condense the large number of characteristics associated with return spreads into a small number of pricing factors. While some recent papers use machine learning tools, e.g., elastic nets (Kozak et al. (2020)), regression trees (Bryzgalova et al. (2023)), and neural nets (Chen et al. (2024)), the majority of estimation methods for latent factor models are based on PCA. However, PCA estimations have several drawbacks. First, PCA factors often have little economic interpretation since they comprise many original test assets. Second, PCA factors have limited success in capturing the cross-section of expected returns (see, e.g., Lettau and Pelger (2020a,b)). Third, PCA estimations require a relatively long time series to obtain reliable results. Fourth, results based on recursive out-of-sample estimations in short samples tend to have poor properties.

This paper proposes an estimation method for multidimensional panels called 3D-PCA that addresses these shortcomings. Standard PCA is based on eigenvectors and eigenvalues of the data's covariance (or second-moment) matrix, which "flattens" the data and removes any multidimensionality of the data set. In contrast, 3D-PCA is specifically designed to exploit the panel structure of the data. In the empirical part of the paper, I estimate 3D-PCA using portfolios constructed from size/characteristic quintile double-sorts for a set of characteristics. Hence, the panel of portfolios has three dimensions: the characteristic, the size quintile, and the quintile of the characteristic sort. 3D-PCA specifies a small set of vectors for each of the three dimensions that serve as "building blocks" for portfolio weights that determine a factor. Different combinations of these building blocks yield different factor weights and, therefore, distinct factors. Since factors are based on the same building blocks, their factor weights are related and subject to restrictions. The restrictions can be expressed as proportionality conditions along the dimensions of the panel data. Factors that share a subset of building blocks share features implied by the common building blocks. The restrictions imply that the number of free parameters in 3D-PCA is several orders of magnitudes smaller than that for comparable PCA estimations. While these restrictions limit the set of possible factor weights, I show that factors with familiar long-short structures are permissible. In particular, the Fama-French factors SMB, HML, ... satisfy the restrictions and are, therefore, special cases of 3D-PCA factors.

I find that factors estimated by 3D-PCA have appealing features and dominate models with PCA and Fama-French factors along several dimensions. First, 3D-PCA factors have straightforward economic intuitions given by the structures of the estimated building block. These building block

vectors have level, slope, and curvature patterns found in many PCA estimations. Factor weights are given by level, slope, and curvature combinations along the different data dimensions. Some resulting factors have familiar patterns along the size and characteristic-quintile dimensions related to long-only averages, small-minus-big, and high-minus-low. In contrast to the SMB and HML factors, 3D-PCA factors combine these patterns with multiple characteristics. For example, some factors are based on (approximately) equal weights across all characteristics, while others are based on only a few characteristics. However, all factors can be interpreted as interactions of the estimated building block vectors.

Second, Sharpe ratios of 3D-PCA factors are substantially higher than those of PCA and Fama-French factors. The highest Sharpe ratio of estimated 3D-PCA factors is 1.08, and six 3D-PC factors have annualized Sharpe ratios above 0.5. In contrast, only a single PCA factor exceeds this benchmark, consistent with Lettau and Pelger (2020a,b) who also find low Sharpe ratios of PCA factors. Third, 3D-PCA factors capture the cross-section of portfolio returns substantially better than PCA and Fama-French factors. Cross-sectional models with three to five 3D-PCA factors estimated in-sample or out-of-sample capture between 65% and 76% of the variation in mean returns. Comparable models with Fama-French or PCA factors have substantially higher pricing errors.

While 3D-PCA pricing errors are smaller for all characteristics, the fit for momentum, reversals, and variance-related portfolios is substantially better than that of PCA and Fama-French factor models. Many factor models do not capture returns of small/low portfolios (e.g., the small/low-BM portfolio). While these portfolios are also the most mispriced in 3D-PCA models, their pricing errors are substantially smaller than those for PCA and Fama-French models.

Fourth, while PCA estimations typically require long time series, 3D-PCA can be estimated in short samples. In the data set used in this paper, estimations of 3D-PCA with as little as 12 time series observations yield robust factor estimates. 3D-PCA estimates are substantially more stable for a given sample length than those from PCA models. The reason is that the construction of 3D-PCA factors imposes restrictions on factor weights, as mentioned above. Hence, fewer free parameters must be estimated. In contrast, PCA can be interpreted as an unrestricted estimation of factor weights. The difference in degrees-of-freedom is stark for all specifications I consider. For example, in the benchmark case, the weight matrix underlying 3D-PCA factors has only 48 free parameters, while the PCA weight matrix requires estimating 7,372 parameters.

Finally, 3D-PCA performs substantially better in recursive out-of-sample estimations than PCA. When estimated in rolling 10-year windows, out-of-sample 3D-PCA factors are highly correlated with their in-sample counterparts. In contrast, out-sample PCA factors are only weakly related to in-sample PCA factors. Indeed, out-of-sample 3D-PCA factors have a slightly better cross-sectional fit than in-sample factors, suggesting that the recursive estimation picks up some time variation in the correlation structure of the test portfolios. In contrast, models with out-of-sample PCA factors have substantially higher pricing errors.

The 3D-PCA model can be estimated using tensor methods, which generalize vector and matrix algebra to higher dimensions. Formally, an n-dimensional panel data set forms an n-dimensional

tensor. For example, a 3-dimensional tensor is a cuboid. The standard PCA of a 2-dimensional matrix can be written in terms of its singular value decomposition (SVD). Tucker (1966) provides a tensor decomposition that generalizes the matrix SVD to n-dimensional tensors, see Kolda and Bader (2009) for a survey of tensor methods and Lettau (2023) for an application to mutual funds. The 3D-PCA model is an implication of the Tucker decomposition and can be estimated using standard iterative numerical methods. Finally, I extend the model to allow for arbitrary dimensions and factorizations.

This paper contributes to the literature on extensions of PCA for latent factor estimation. Connor et al. (2012), Fan et al. (2016), and Kim et al. (2021) allow factor betas to be nonparametric functions of observed characteristics. Kelly et al. (2019) propose a related Instrumented-PCA (I-PCA) model in which factor betas are linear in characteristics. Gu et al. (2021) extend I-PCA to nonlinear specifications. Lettau and Pelger (2020a,b) show that standard PCA may not be able to identify weak factors. Their Risk Premium PCA (RP-PCA) estimator can overcome this limitation when weak factors have high Sharpe ratios. 3D-PCA can be combined with methods that are based on PCA estimations. For example, 3D-PCA factors can be used in place of PCA factors in Kozak et al. (2020)'s shrinkage estimator of the stochastic discount factor and the three-pass model of Giglio and Xiu (2021). Babii et al. (2022) study an alternative tensor decomposition.

The rest of the paper is organized as follows. Section 2 presents an example that illustrates the intuition of high-dimension factor models. Section 3 introduces the data set used in the paper. Section 4 introduces general 3D-PCA model and its estimation. The empirical results are described in Section 5. Section 6 extends 3D-PCA to arbitrary dimensions, and section 7 concludes.

2. Example: 2D-PCA

Consider the nine portfolios formed by a 2-dimensional portfolio sort as the intersection of three size (ME) and three book-to-market (BM) portfolios that are used to form SMB and HML.[1] Let $\mathbf{R}_t = [R_{pq,t}]$ be the 3-by-3 matrix of excess returns of (p,q)-portfolios where p and q are size and BM terciles, respectively. Factors are constructed as linear combinations of the test assets. Let $\mathbf{W} = [w_{pq}]$ a 3-by-3 matrix of weights so that the associated factor is given by

$$F_t = \sum_{p,q} w_{pq} R_{pq,t} = \text{vec}(\mathbf{W})^\top \text{vec}(\mathbf{R}_t). \tag{1}$$

The literature has considered various methods to determine factor weights $\mathbf{W}$. For example, Fama and French (1993) use fixed weights to construct their SMB and HML factors:

$$\mathbf{W}_{\text{SMB}} = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \qquad \mathbf{W}_{\text{HML}} = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}. \tag{2}$$

Alternatively, factors can be estimated by PCA, so that factor weights are given by the eigenvectors of the second-moment matrix of returns, $\text{E}_t[\text{vec}(\mathbf{R}_t)\text{vec}(\mathbf{R}_t)^\top]$.

[1] SMB and HML are constructed using a 2-by-3 sort, but I use a 3-by-3 structure for consistency with the rest of the paper.

Instead, I consider factor weights that are constructed from separate vectors for the ME and BM dimensions. Let

$$\mathbf{v}^P = \begin{pmatrix} v_1^P \\ v_2^P \\ v_3^P \end{pmatrix}, \qquad \mathbf{v}^Q = \begin{pmatrix} v_1^Q \\ v_2^Q \\ v_3^Q \end{pmatrix}, \tag{3}$$

$$\mathbf{W}_{pq} = \mathbf{v}^P \mathbf{v}^{Q^\top}, \tag{4}$$

so that $\mathbf{W}_{pq}$ is a 3-by-3 matrix of weights given by the outer product of $\mathbf{v}^P$ and $\mathbf{v}^Q$. The factor weight of portfolio (p,q) is therefore $w_{pq} = v_p^P v_q^Q$. Suppose there are two such vectors for the size dimension, as well as for the BM dimension, and collect them in the (3×2) matrices

$$\mathbf{V}^P = \left[\mathbf{v}_1^P, \mathbf{v}_2^P\right], \qquad \mathbf{V}^Q = \left[\mathbf{v}_1^Q, \mathbf{v}_2^Q\right]. \tag{5}$$

Factors are formed by combining the columns of $\mathbf{V}^P$ and $\mathbf{V}^P$. Since both matrices have two columns, there are four possible combinations, so four different factors can be created. In other words, the columns of $\mathbf{V}^P$ and $\mathbf{V}^P$ form the building blocks of factor weights.

To develop further intuition, consider the following specific example

$$\mathbf{V}^P = \mathbf{V}^Q = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}. \tag{6}$$

Combining the columns of $\mathbf{V}^P$ and $\mathbf{V}^Q$ yields weights of four factors:

$$\mathbf{W}_{11} = \mathbf{v}_1^P \mathbf{v}_1^{Q^\top} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \qquad \mathbf{W}_{12} = \mathbf{v}_1^P \mathbf{v}_2^{Q^\top} = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix} \tag{7}$$

$$\mathbf{W}_{21} = \mathbf{v}_2^P \mathbf{v}_1^{Q^\top} = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \qquad \mathbf{W}_{22} = \mathbf{v}_2^P \mathbf{v}_2^{Q^\top} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \tag{8}$$

which can be collected in the (9×4)-matrix of factor weights $\mathbf{W}^{2D} = [\text{vec}(\mathbf{W}_{11}), \text{vec}(\mathbf{W}_{12}), \text{vec}(\mathbf{W}_{21}), \text{vec}(\mathbf{W}_{22})]$. Note that $\mathbf{W}^{2D}$ can also be expressed using the Kronecker product $\otimes$: $\mathbf{W}^{2D} = \mathbf{V}^P \otimes \mathbf{V}^Q$. The vector of four factors is given by $\mathbf{F}_t^{2D} = \mathbf{W}^{2D^\top} \text{vec}(\mathbf{R}_t)$.

Comparing the weight matrices in (7) and (8) to (2) shows that $\mathbf{W}_{21}$ is identical to SMB weights and that $\mathbf{W}_{12}$ is equal to HML weights. Hence, the Fama-French factors are special cases of 2D-PCA. Furthermore, the rows and columns of each weight matrices are proportional. In addition, the rows of $\mathbf{W}_{11}$ and $\mathbf{W}_{21}$ are proportional as are the rows of $\mathbf{W}_{12}$ and $\mathbf{W}_{22}$. The columns have similar proportionality properties. Of course, these patterns arise because the weight matrices are constructed from the same building blocks. i.e., the columns of $\mathbf{V}^P$ and $\mathbf{V}^Q$.

The patterns in the weight matrices can also be understood as restrictions imposed by the construction of the weight matrices in (3) and (4). The (9×4)-matrix of factor weights $\mathbf{W}^{2D}$ has 36 elements while the $\mathbf{V}^P$ and $\mathbf{V}^Q$ matrices have a total of 12 elements. Hence, the elements of $\mathbf{W}^{2D}$ are subject to 24 restrictions. In the general model introduced below, $\mathbf{V}^P$ and $\mathbf{V}^Q$ are assumed to be orthonormal, which imposes six additional restrictions, so that $\mathbf{W}^{2D}$ has six degrees of freedom.

On the one hand, the $\mathbf{W}^{2D}$ matrix is highly parametrized, and the set of feasible $\mathbf{W}^{2D}$ is limited; on the other hand, the parametrization allows for long/short patterns often found in pricing factors. In PCA, the weight matrix $\mathbf{W}$ is composed of eigenvectors and is, therefore, orthonormal but not subject to other restrictions. The PCA weight matrix has 36 elements, and orthonormality imposes seven restrictions. Hence, the weight matrix has $36 - 7 = 29$ free parameters, almost five times the number of free parameters of $\mathbf{W}^{2D}$.

3. Data

The data set consists of excess returns of double-sorted portfolios of 11 characteristics: book-to-market (BM), operating profitability (OP), investment (INV) [2], momentum (MOM), long and short-term reversal (REV, SREV), accruals (AC), beta (BETA), net share issuance (NSI), daily variance (VAR), and daily residual variance (RVAR). Each set of double-sorted portfolios is based on the intersection of five size (ME) quintile portfolios and five quintiles of characteristic $c \in$(BM, OP, ..., RVAR). The total number of portfolios is $11 \times 25 = 275$. The size and characteristic quintiles are denoted P1 to P5 and Q1 to Q5, respectively. The sorts are arranged so that the high-return portfolios are in the fifth quintile and the low-return portfolios are in the first quintile. For example, since small stocks have higher average returns than big stocks, P5 refers to the quintile with the smallest stocks, and P1 is the quintile with the biggest stocks.[3] The data set has 724 monthly observations from July 1963 to October 2023. All data are downloaded from Ken French's website.

Figure 1 shows heatmaps of annualized mean returns of the 5-by-5 quintiles for each of the 11 characteristics.[4] The bottom right panel shows the mean across all characteristics. The annualized mean excess return of the CRSP value-weighted index (CRSP-VW) over the sample period is 7.29%. Portfolios with mean returns that are higher (lower) than the mean CRSP-VW return are plotted in blue (red). As usual, the highest and lowest returns are associated with the highest and lowest quintiles, although the pattern is not always monotonic. The small-high momentum portfolios MOM-P5Q5 and MOM-P4Q5 have the highest mean returns (15.2% and 14.3%, respectively) followed by the P4Q5 RVAR and VAR portfolios (13.8% and 13.6%). The RVAR and VAR P5Q1 portfolios have a negative average excess return, while the means of SREV and MOM P5Q1 portfolios are below 1%. Note that the interaction of size and characteristics quintiles varies across characteristics. While the size effect is positive for all Q5 quintiles, it reverses in Q1 portfolios for some characteristics but not others. For example, the return of the big/low-BM portfolio (BM-P1Q1) is 6.9%, which is higher than the mean of the small/low-BM portfolio (BM-P5Q1, 3.0%). On the other hand, the returns of the AC-P5Q1 accruals portfolio are higher than those of the AC-P1Q1 portfolio.

Table 1 reports annualized means, standard deviations, and Sharpe ratios of long-short portfolios for each characteristic. SMB_c is the difference between the small-stock portfolios (P5) and the big-stock portfolios (P1) of the double sort of characteristic c averaged of all characteristic quintiles.

[2] INV is defined as asset growth.
[3] Thus, P1 (P5) correspond to high (low) OP, NSI, REV, and SREV.
[4] Sharpe ratios are plotted in Figure C.1 in the Appendix.

HML_c is constructed accordingly, but is the difference of Q5 and Q1 portfolios averaged over all size quintiles. Finally, $CROSS_c$ is the average of the P1Q1 and P5Q5 portfolios minus the average of the P1Q5 and P5Q1 portfolios. SMB_c is positive for all characteristics; hence, small stocks earn, on average, higher returns than big stocks in double-sorts. However, there is substantial variation in the magnitudes of the size premium. SMB_c of OP is 3.91%, three times the SMB_c of SREV. The average SMB_c across all characteristics is 2.75%. For most characteristics, the premium in the characteristic dimension is larger than the size premium. The average HML_c of MOM is 10.43% and is by far the largest in the sample. The second largest HML_c is SREV, 7.37%. In contrast, HML_c of BETA is only 0.7%. The average HML_c is 4.83%, substantially larger than the average SMB_c.

Recall that $CROSS_c$ captures the difference between the low/low and high/high, and the low/ high and high/low corner portfolios. The third column of Table 1 shows that there is large variation of mean $CROSS_c$ returns across characteristics. The reason is that the size premium inverts across the Q1 to Q5 quintiles for some characteristics but not for others. $CROSS_c$ is high for the characteristic with inverted size premia and low for those without inversion. The highest $CROSS_c$ returns of 13.70% and 13.10% are associated with VAR and RVAR, respectively, which also the highest among all long-short portfolios shown in the table. As shown in Figure 1, the size premium is positive for their Q5 portfolios (P5Q5-P1Q5 is positive) but strongly negative for the Q1 portfolios (P5Q1-P1Q1 is negative). As a result the return of the $CROSS_c$ portfolios are particularly high and outstrips their HML_c by a factor of three. $CROSS_c$ of SREV, BM, and MOM are also high. On the other hand, $CROSS_c$ of AC and BETA is negative since the size premium is positive for the Q1 and Q5 portfolios.

The standard deviations of all SMB_c and $CROSS_c$ portfolios are between 12% and 18% and thus comparable. In contrast, the volatilities of the HML_c vary across characteristics. The volatility of the ACC HML_c is only 4.91% and thus particular low. The reason is that the corresponding Q1 and Q5 AC portfolios are highly correlated; the mean correlation across the five size portfolios is 0.91. As a result, the volatility of the long-short portfolio is low. This pattern is also true for OP, INV, and NSI portfolios but to a lesser extent (their mean correlations are between 85% and 90%). However, the correlations of the MOM and VAR Q1 to Q5 portfolios are below 0.7, so their HML_c portfolios are more volatile. The differences in volatilities of HML_c portfolios will be important for the estimation of the factor models considered in the rest of the paper.

4. 3-dimensional factor models: 3D-PCA

The standard method to estimate latent factors from time series of portfolio returns is to "flatten" the multidimensional panel into a 2-dimensional matrix and apply PCA. Since PCA treats each column the same, any information related to the construction of the portfolios is lost. In the context of the data set described in the previous section, the portfolio construction is 3-dimensional: characteristic/size quintile/characteristic quintile. In this section, I describe a 3-dimensional factor model that exploits the multidimensional structure of the portfolios. Section 6 shows how the model can be extended to arbitrary dimensions.

The data is assembled in a four-dimensional panel $\boldsymbol{\mathcal{X}} = [x_{tcpq}]$, where the first dimension is time, the second dimension corresponds to characteristics, the third and fourth dimensions are the size and characteristic percentile portfolios. Hence, x_{tcpq} is the excess return of the portfolio given by the p-th size and q-th characteristic portfolio of characteristic c in period t. The data set consists of $P \times Q$ double-sorted portfolios of C characteristics observed over T periods. The total number of portfolios is $N = CPQ$. Let $\mathbf{X}_{(T)}$ be the $(T \times CPQ)$-dimensional matrix whose columns are the time series of portfolio returns.

4.1. From matrices to tensors

Multidimensional panels can be represented as *tensors*, which generalize the notions of vectors and matrices to higher dimensions. For example, a 3-dimensional tensor is a cubeoid, while the data set used in this paper forms a 4-dimensional tensor. While tensor notation is more complex and differs from the familiar matrix notation in several ways, tensor algebra is a straightforward extension of matrix algebra, see Kolda and Bader (2009) and Lettau (2023).

Throughout the paper, I use the following notation:

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

Hence, a zero-order tensor is a scalar, a first-order tensor is a vector, a second-order tensor is a matrix, and a third-order tensor is a cuboid. Each of the m dimensions of a tensor is called a *mode*.

The *slice* of a m-dimensional tensor is given by the $m-1$-dimensional tensor when one of the indices is fixed at a given value. For example, a t-slice of $\boldsymbol{\mathcal{X}}$ is a 3-dimensional tensor of size $(C \times P \times Q)$, $\boldsymbol{\mathcal{X}}_{(t)cpq}$, consisting of N portfolio returns in period t. A *fiber* of $\boldsymbol{\mathcal{X}}$ is a vector that is obtained by fixing all indices but one. For example, the mode-t fiber, $\mathbf{x}_{(cpq)t}$, is the $(T \times 1)$ vector of the time series of the p-th size quintile and q-th quintile portfolio of characteristic c. Figure C.2 shows a 3-dimensional tensor and its fibers and slices. When the context is clear, I will sometimes write $\boldsymbol{\mathcal{X}}_{(t)cpq}$ and $\mathbf{x}_{(cpq)t}$ as $\boldsymbol{\mathcal{X}}_t$ and $\mathbf{x}_t$.

A tensor can be *matricized* by stacking its fibers in a matrix. For example, $\mathbf{X}_{(T)} = \text{mat}_T(\boldsymbol{\mathcal{X}})$ is the $(T \times CPQ)$ matrix whose rows are the t-fibers $\mathbf{x}_{(cpq),t}$. In other words, the columns of $\mathbf{X}_{(T)}$ are the time series of returns of the CPQ portfolios. Figure C.3 shows the matricization of a 3-dimensional tensor along each dimension. Similarly, $\mathbf{x} = \text{vec}(\boldsymbol{\mathcal{X}})$ denotes the vectorized tensor of dimension$(TCPQ \times 1)$. The n-mode product, denoted $\times_n$, is the multiplication of a tensor and a conforming matrix along the n-th dimension. For example, the mode-1 product of a $(I \times J \times K)$ tensor $\boldsymbol{\mathcal{X}}$ and a $(L \times I)$ matrix $\mathbf{A}_1$ is equal to a $(L \times J \times K)$ tensor $\boldsymbol{\mathcal{X}} \times_1 \mathbf{A}_1$ (see Figure C.4). Note that the standard matrix product can be written in tensor notation: $\mathbf{A}_1 \mathbf{X} \mathbf{A}_2^\top = \mathbf{X} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2$.

Let $\mathbf{a}_1, \ldots, \mathbf{a}_m$ be m vectors of lengths $i_1, \ldots, i_m$. The outer product $\circ$ of the m vectors is a m-dimensional tensor of size $(i_1 \times \cdots \times i_m)$: $\boldsymbol{\mathcal{X}} = \mathbf{a}_1 \circ \cdots \circ \mathbf{a}_m$. Panel C of Figure C.4 illustrates the outer

product of three vectors. Note that the outer product of two vectors $\mathbf{a}_1 \circ \mathbf{a}_2 = \mathbf{a}_1\mathbf{a}_2^\intercal = \mathbf{X}$ yields a $(i_1 \times i_2)$ matrix $\mathbf{X}$.

4.2. The 3D-PCA model

High-dimensional factor models are a natural extension of standard 2-dimensional models, which can summarized as follows. The singular value decomposition (SVD) of a matrix $\mathbf{X}$ of asset returns is given by

$$\mathbf{X} = \mathbf{U}^{(1)}\,\mathbf{H}\,\mathbf{U}^{(2)\intercal}, \tag{9}$$

where the columns of $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ are the eigenvectors of $\mathbf{X}\mathbf{X}^\intercal$ and $\mathbf{X}^\intercal\mathbf{X}$, respectively, and $\mathbf{H}$ is the diagonal matrix of the square roots of the corresponding non-zero eigenvalues. Since $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ are orthogonal, (9) can be written in factor form. Factors are linear combinations of asset returns and are given by $\mathbf{F} = \mathbf{U}^{(1)}\,\mathbf{H} = \mathbf{X}\,\mathbf{U}^{(2)}$ where the columns of $\mathbf{U}^{(2)}$ are the factor weights. The truncated K-factor model uses only eigenvectors of the K largest eigenvalues so that (9) becomes an approximation: $\mathbf{X} \approx \mathbf{U}_K^{(1)}\,\mathbf{H}_K\,\mathbf{U}_K^{(2)\intercal}$. Note that the SVD can be written in tensor notation as

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

$$\mathbf{F} = \mathbf{H} \times_1 \mathbf{U}^{(1)} = \mathbf{X} \times_2 \mathbf{U}^{(2)^\intercal}. \tag{11}$$

Tucker (1966) shows that the matrix SVD can be extended to higher dimensional tensors. The Tucker decomposition states that a m-dimensional tensor can be written in terms of a "small" m-dimensional "core" tensor and m matrices. The Tucker decomposition with (K_T, K_C, K_P, K_Q) factors is given by

$$\boldsymbol{\mathcal{X}} \approx \boldsymbol{\mathcal{G}} \times_1 \mathbf{V}^{(T)} \times_2 \mathbf{V}^{(C)} \times_3 \mathbf{V}^{(P)} \times_3 \mathbf{V}^{(Q)}, \tag{12}$$

where $\times_n$ denotes the n-mode tensor product. $\boldsymbol{\mathcal{G}}$ is a $(K_T \times K_C \times K_P \times K_Q)$ *core* tensor and $\mathbf{V}^{(T)}$, $\mathbf{V}^{(C)}$, $\mathbf{V}^{(P)}$, $\mathbf{V}^{(Q)}$ are $(T \times K_T), (C \times C), (P \times K_P), (Q \times K_Q)$ matrices, respectively. As in the SVD, each $\mathbf{V}^{(i)}$ is normalized to be orthonormal. When $K_T = T, K_C = C, K_P = P, K_Q = Q$, the inequality in (12) becomes an equality. Figure C.5 shows the Tucker decomposition of a 3-dimensional tensor. The matrix SVD (10) is a special case of the tensor Tucker decomposition (12) when $\boldsymbol{\mathcal{X}}$ is 2-dimensional. The $\mathbf{V}^{(i)}$ matrices correspond to $\mathbf{U}^{(i))}$ and $\boldsymbol{\mathcal{G}}$ corresponds to $\mathbf{H}$. Kolda and Bader (2009) provide a detailed treatment of the Tucker decomposition while Lettau (2023) develops the intuition in the context of a finance application.

The 3D-PCA model is based on a simplified version of the Tucker decomposition (12) and is defined as follows. Since we are interested in the factor structure of the portfolios, the time dimension does not have to be factored, which suggests a *partial* Tucker decomposition with K_C, K_P, and K_Q factors for the C, P, and Q dimensions:[5]

$$\boldsymbol{\mathcal{X}} \approx \boldsymbol{\mathcal{F}} \times_2 \mathbf{V}^{(C)} \times_3 \mathbf{V}^{(P)} \times_4 \mathbf{V}^{(Q)}. \tag{13}$$

$\boldsymbol{\mathcal{F}}$ is a $(T \times K_C \times K_P \times K_Q)$-dimensional tensor and $\mathbf{V}^{(C)}, \mathbf{V}^{(P)}$, and $\mathbf{V}^{(Q)}$ are matrices of dimensions $(K_C \times C)$,

[5]All results reported in the paper are similar when the full Tucker model is used.

$(K_P \times P)$, and $(K_Q \times P)$, respectively.[6] $\boldsymbol{\mathcal{F}}$ has the interpretation of factors similar to the factors in the SVD decomposition (11). To see this, multiply both sides of (13) by $\mathbf{V}^{(C)^\top}, \mathbf{V}^{(P)^\top}$, and $\mathbf{V}^{(C)^\top}$:

$$\boldsymbol{\mathcal{F}} \approx \boldsymbol{\mathcal{X}} \times_2 \mathbf{V}^{(C)^\top} \times_3 \mathbf{V}^{(P)^\top} \times_4 \mathbf{V}^{(Q)^\top}. \tag{14}$$

Note that (14) has a similar form as the PCA factors in (11). The difference is that PCA factors are the product of a matrix with a weight matrix, while 3D-PCA factors are given by multiplying a tensor by three weight matrices.

Using the properties of the n-mode product, (13) can be written in terms of matrices. Define $\mathbf{W}^{3D}$ as the Kronecker product of $\mathbf{V}^{(C)}, \mathbf{V}^{(P)}$, and $\mathbf{V}^{(Q)}$:

$$\mathbf{W}^{3D} = \mathbf{V}^{(C)} \otimes \mathbf{V}^{(P)} \otimes \mathbf{V}^{(Q)}, \tag{15}$$

which is a $(CPQ \times K_C K_P K_Q)$-dimensional matrix. Note that $\mathbf{W}^{3D}$ is orthonormal since the $\mathbf{V}^{(i)}$ matrices are orthonormal. Recall that $\mathbf{X}_{(T)}$ is the $(T \times CPQ)$ matrix whose columns are the time series of portfolio returns. Multiplying $\mathbf{X}_{(T)}$ by $\mathbf{W}^{3D}$ yields a matrix of factors:

$$\mathbf{F}_{(T)} = \mathbf{X}_{(T)} \mathbf{W}^{3D}. \tag{16}$$

$\mathbf{F}_{(T)}$ is a $(T \times K_C K_P K_Q)$-dimensional matrix with columns F_{cpq}. Note that $\mathbf{F}_{(T)}$ is the matricized factor tensor, $\mathbf{F}_{(T)} = \text{mat}_T(\boldsymbol{\mathcal{F}})$ and $\mathbf{W}$ is the matrix of factor weights. Each of its $K_C K_P K_Q$ columns is a $(CPQ \times 1)$ vector of weights that yields a factor when multiplied by $\mathbf{X}_{(T)}$.

A key property of the 3D-PCA model is that the weight matrix $\mathbf{W}^{3D}$ is determined by the Tucker matrices $\mathbf{V}^{(C)}, \mathbf{V}^{(P)}$, and $\mathbf{V}^{(Q)}$. The intuition of constructing the columns of $\mathbf{W}$ is as follows. Let $\mathbf{v}_c^{(C)}, \mathbf{v}_p^{(P)}$, and $\mathbf{v}_q^{(Q)}$ be the c-th, p-th, and q-th columns of $\mathbf{V}^{(C)}, \mathbf{V}^{(P)}$, and $\mathbf{V}^{(Q)}$, respectively. Their outer product $\boldsymbol{\mathcal{W}}_{cpq} = \mathbf{v}_c^{(C)} \circ \mathbf{v}_p^{(P)} \circ \mathbf{v}_q^{(Q)}$ is a three-dimensional tensor of dimension $(C \times P \times Q)$ with elements

$$w_{cpq,ijk} = v_{ci}^{(C)} \, v_{pj}^{(P)} \, v_{qk}^{(Q)} \quad \text{for} \quad i = 1, \ldots, C, j = 1, \ldots, P, k = 1, \ldots, Q, \tag{17}$$

where $v_{ci}^{(C)}, v_{pj}^{(P)}$ and $v_{qk}^{(Q)}$ are the i-th, j-th, and k-the elements of $\mathbf{v}_c^{(C)}, \mathbf{v}_p^{(P)}$, and $\mathbf{v}_q^{(Q)}$ respectively. $\boldsymbol{\mathcal{W}}_{cpq}$ is comprised of weights that can be used to construct factors. The (c, p, q)-factor is given by

$$F_{cpq,t} = \sum_i \sum_j \sum_k w_{cpq,ijk} \, x_{tijk} \tag{18}$$

$$= \text{vec}(\boldsymbol{\mathcal{W}}_{cpq})^\top \mathbf{x}_t. \tag{19}$$

Repeating this procedure for all combinations of (c, p, q) yields $K_C K_P K_Q$ factors that are collected in $\boldsymbol{\mathcal{F}}$. The columns of the $\mathbf{V}^{(C)}, \mathbf{V}^{(P)}$, and $\mathbf{V}^{(Q)}$ matrices can be understood as building blocks that create the set of factor weights. The total number of vectors is $K_C + K_P + K_Q$, which creates $K_C K_P K_Q$ different factors.

Since the factor weights of 3D-PCA are given by outer products of a small number of vectors, they are subject to restriction. For example, all 2-dimensional slices of the $\boldsymbol{\mathcal{W}}_{cpq}$ factor tensors are proportional. To see this, fix c, p and q and consider the slices of $\boldsymbol{\mathcal{W}}_{cpq}$ corresponding to characteristics. Recall that $\boldsymbol{\mathcal{W}}_{cpq} = \mathbf{v}_c^{(C)} \circ \mathbf{v}_p^{(P)} \circ \mathbf{v}_q^{(Q)}$. Hence, the slice corresponding to characteristic i is given

[6] Note that $\boldsymbol{\mathcal{F}}$ has T elements in the first (time) dimension while $\boldsymbol{\mathcal{G}}$ has K_T elements. In the full decomposition (12), the implied $\boldsymbol{\mathcal{F}}$ is given by $\boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{V}^{(T)}$.

by $v_{ci}^{(C)}\left(\mathbf{v}_p^{(P)} \circ \mathbf{v}_q^{(Q)}\right)$.[7] Hence all c-slices are proportional to the $(P \times Q)$-dimensional matrix $\mathbf{v}_p^{(P)} \circ \mathbf{v}_q^{(Q)}$. Next, consider the weight tensor $\boldsymbol{\mathcal{W}}_{c'pq}$ that is created by the outer product of the same $v_p^{(P)}$ and $v_q^{(Q)}$ vectors but a different vector $v_{c'}^{(C)}$. Since $\boldsymbol{\mathcal{W}}_{c'pq} = \mathbf{v}_{c'}^{(C)} \circ \mathbf{v}_p^{(P)} \circ \mathbf{v}_q^{(Q)}$, all c-slices $v_{c'i}^{(C)}\left(\mathbf{v}_p^{(P)} \circ \mathbf{v}_q^{(Q)}\right)$ are also proportional to $\mathbf{v}_p^{(P)} \circ \mathbf{v}_q^{(Q)}$. Therefore, all factor weights that stem from the same columns of $\mathbf{V}^{(P)^\top}$ and $\mathbf{V}^{(Q)^\top}$ are proportional. The same argument holds for combinations of columns of $\mathbf{V}^{(C)^\top}$ and $\mathbf{V}^{(P)^\top}$, and columns of $\mathbf{V}^{(C)^\top}$, and $\mathbf{V}^{(Q)^\top}$. Therefore, the set of factor weights of the 3D-PCA model is subject to within-factor and across-factor restrictions.

It is helpful to compare the factors based on the Tucker decomposition (13) to factors estimated by PCA. To compute PCA, $\boldsymbol{\mathcal{X}}$ has to be written as a matrix. Unfolding along the time dimensions yields the $(T \times CPQ)$ matrix $\mathbf{X}_{(T)}$. Note that in contrast to 3D-PCA, estimating PCA using $\mathbf{X}_{(T)}$ does not take the multidimensional structure of the triple-sorted portfolios into account. Factor weights are given by the eigenvectors of $\mathbf{X}_{(T)}^\top \mathbf{X}_{(T)}$. The only restriction imposed by PCA is that the columns of the weight matrix are orthonormal. In this sense, PCA yields unrestricted factor weights, while 3D-PCA factor weights are subject to restrictions.

It is instructive to compare the number of free parameters of comparable PCA and 3D-PCA models. Let $K = K_C K_P K_Q$ be the total number of factors. The matrix of factor weights of the K-factor PCA model has $CPQ\,K = NK$ elements. Orthonormality implies $K(K+1)/2$ restrictions, so there are $NK - K(K+1)/2$ free parameters. Factor weights of the 3D-PCA model are generated by $\mathbf{V}^{(C)^\top}$,$\mathbf{V}^{(P)^\top}$, and $\mathbf{V}^{(Q)^\top}$. For example, $\mathbf{V}^{(C)^\top}$ has CK_C elements and $K_C(K_C+1)/2$ orthonormality restrictions. Thus, the total number of free parameters in the three matrices is $CK_C + PK_P + QK_Q - (K_C(K_C+1)/2 + K_P(K_P+1)/2 + K_Q(K_Q+1)/2)$. For example, the data set used in the next section is of size $C = 11, P = Q = 5$. For $K_C = K_P = K_Q = 3$, PCA has 7,047 parameters compared to only 45 for 3D-PCA. As we will see in the empirical part of the paper, this stark difference drastically affects the estimation of both models.

In matrix-PCA, the contribution of a factor to the overall variation in the data is given by the associated eigenvalue of the eigenvector that is used to construct the factor. A similar result can be derived for the 3D-PCA model. Consider first the full Tucker decomposition (12) with $K_T = T, K_C = C, K_P = P, K_Q = Q$, so that it holds with an equality. Since the $\mathbf{V}^{(i)}$ matrices are orthonormal, the overall variation of the data tensor $\boldsymbol{\mathcal{X}}$ is equal to the squared norm of the core tensor $\boldsymbol{\mathcal{G}}$:

$$\|\boldsymbol{\mathcal{X}}\|^2 = \sum_{t,c,p,q} x_{tcpq}^2 = \sum_{t,c,p,q} g_{tcpq}^2 = \|\boldsymbol{\mathcal{G}}\|^2. \tag{20}$$

Therefore, the squared elements of the core tensor $\boldsymbol{\mathcal{G}}$ can be interpreted as the contribution of the corresponding factor to the overall variation of the data and play a similar role as eigenvalues in matrix-PCA.

In the partial Tucker decomposition (12), $\boldsymbol{\mathcal{F}}$ corresponds to the core tensor, so that $\|\boldsymbol{\mathcal{X}}\|^2 = \|\boldsymbol{\mathcal{F}}\|^2$. Recall that $\mathbf{X}_{(T)} = \text{mat}_T(\boldsymbol{\mathcal{X}})$ is the $(T \times CPQ)$ matrix with columns that are time series of portfolio

[7] See Figure C.6 for an illustration.

returns. Similarly, $\mathbf{F}_{(T)} = \text{mat}_T(\boldsymbol{\mathcal{F}})$ is the $(T \times K_C K_P K_Q)$ matrix of time series of factors. Hence,

$$\|\boldsymbol{\mathcal{X}}\|^2 = \|\mathbf{X}_{(T)}\|^2 = \|\mathbf{F}_{(T)}\|^2 \tag{21}$$

$$= \sum_{c,p,q} \left(\sum_t F_{cpq,t}^2 \right), \tag{22}$$

which implies that the contribution of a factor F_{cpq} to the overall variation in the data is given by its time series sum of squares, $\sum_t F_{cpq,t}^2$.

4.3. Estimation

Given (K_C, K_P, K_Q), the Tucker decomposition (13) can be written as

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{F}} \times_2 \mathbf{V}^{(C)} \times_3 \mathbf{V}^{(P)} \times_4 \mathbf{V}^{(Q)} + \boldsymbol{\mathcal{E}}, \tag{23}$$

where $\boldsymbol{\mathcal{E}}$ is the approximation error. The objective is to find $\boldsymbol{\mathcal{F}}, \mathbf{V}^{(C)}, \mathbf{V}^{(P)}$, and $\mathbf{V}^{(Q)}$ that minimize the l^2-norm of $\boldsymbol{\mathcal{E}}$:

$$\widehat{\boldsymbol{\mathcal{F}}}, \hat{\mathbf{V}}^{(C)}, \hat{\mathbf{V}}^{(P)}, \hat{\mathbf{V}}^{(Q)} = \text{argmin} \|\boldsymbol{\mathcal{E}}\|, \tag{24}$$

subject to the restriction that $\hat{\mathbf{V}}^{(C)}, \hat{\mathbf{V}}^{(P)}, \hat{\mathbf{V}}^{(Q)}$ are orthonormal. There is no closed-form solution to (24), so the problem has to be solved numerically. One candidate solution is given by the Higher-Order SVD (HOSVD) as suggested by De Lathauwer et al. (2000). HOSVD solves a series of 2-dimensional eigenvalue problems that yield estimates of the $\mathbf{V}^{(i)}$ matrices followed by the computation of $\boldsymbol{\mathcal{F}}$. In the 3-dimensional case considered here, HOSVD is given by the following procedure:

1. Unfold $\boldsymbol{\mathcal{X}}$ along the C dimension: $\mathbf{X}_{(C)} = \text{mat}_C(\boldsymbol{\mathcal{X}})$. Compute the SVD of $\mathbf{X}_{(C)}$ and set $\hat{\mathbf{V}}^{(C)}$ to the first K_C left singular vectors, i.e. the eigenvectors of $\mathbf{X}_{(C)}\mathbf{X}_{(C)}^\top$.
2. Unfold $\boldsymbol{\mathcal{X}}$ along the P dimension: $\mathbf{X}_{(P)} = \text{mat}_P(\boldsymbol{\mathcal{X}})$. Compute the SVD of $\mathbf{X}_{(P)}$ and set $\hat{\mathbf{V}}^{(P)}$ to the first K_P left singular vectors, i.e. the eigenvectors of $\mathbf{X}_{(P)}\mathbf{X}_{(P)}^\top$.
3. Unfold $\boldsymbol{\mathcal{X}}$ along the Q dimension: $\mathbf{X}_{(Q)} = \text{mat}_Q(\boldsymbol{\mathcal{X}})$. Compute the SVD of $\mathbf{X}_{(Q)}$ and set $\hat{\mathbf{V}}^{(Q)}$ to the first K_Q left singular vectors, i.e. the eigenvectors of $\mathbf{X}_{(Q)}\mathbf{X}_{(Q)}^\top$.
4. Set $\widehat{\boldsymbol{\mathcal{F}}} = \boldsymbol{\mathcal{X}} \times_2 \hat{\mathbf{V}}^{(C)^\top} \times_3 \hat{\mathbf{V}}^{(P)^\top} \times_4 \hat{\mathbf{V}}^{(Q)^\top}$.

In general, HOSVD does not minimize $\|\boldsymbol{\mathcal{E}}\|$ and is therefore not optimal. Kroonenberg and de Leeuw (1980) suggested an iterative procedure called Higher-Order Orthogonal Iteration (HOOI) that has proven to yield efficient solutions in most applications.[8] They show that one $\hat{\mathbf{V}}^{(i)}$ can be solved if the other $\hat{\mathbf{V}}^{(j)}$ and $\hat{\mathbf{V}}^{(k)}$ are known, see the Appendix Appendix A.3 and Kolda and Bader (2009) for more details. Each step solves a linear problem so that HOOI is computationally feasible even for large tensors. Typically, HOOI uses the HOSVD estimates as starting values and iterates until a convergence criterion is satisfied.[9]

[8] HOOI is also known as Alternate Least Square (ALS).

[9] There are other solution methods, but HOOI remains the most widely used method, see Kolda and Bader (2009).

5. Empirical results

In this section, I estimate the 3D-PCA model using the 3-dimensional data set of double-sorted quintile portfolios of 11 characteristics described in section 3. In addition to the in-sample estimation that uses the entire sample, I also construct out-of-sample estimations that are not subject to a look-ahead bias using rolling samples of length h. I estimate the 3D-PCA model in each subsample ending in $t' = h, ..., T-1$, and construct factor returns in $t'+1$ by multiplying the portfolio returns in period $t'+1$ by the factor weights from the subsample ending in t'. The benchmark window length is $h = 120$, but I also consider other values for h.

The 3D-PCA model approximates the data tensor with 275 portfolios when $K_C < C, K_P < P, K_Q < Q$. Therefore, I first compute the proportion of the variation in $\boldsymbol{\mathcal{X}}$ captured by the model for different combinations of (K_C, K_P, K_Q). The R^2 is defined as $R^2 = 1 - \|\boldsymbol{\mathcal{E}}\| / \text{Var}(\boldsymbol{\mathcal{X}})$, where $\boldsymbol{\mathcal{E}}$ is the approximation error in (23). Figure 2 shows the heatmap of the R^2 for all possible combinations of (K_C, K_P, K_Q). The columns correspond to K_C and the rows to (K_P, K_Q) combinations. The model with a single factor in each dimension in the upper left cell captures 83% of the data variation. Increasing K_C while keeping $K_P = K_Q = 1$ has a marginal impact on the R^2, suggesting that higher values of (K_P, K_Q) are required. The most parsimonious models with an R^2 of at least 90% and 95% are $K_C = K_P = K_Q = 2$ with a total of 8 factors and $K_C = 5, K_P = K_Q = 3$ with 45 factors, respectively. Increasing the number of factors improves the fit only marginally.

As we will see in the next section, the first three P and Q factors are relevant for capturing the cross-section of mean returns. Combining $K_P = K_Q = 3$ with $K_C = 3$ yields a parsimonious model with 27 factors that captures 93% of the variation in the data. In rolling samples, the R^2 varies between 89% and 96%, which indicates that the model is stable over time. In the rest of the paper, the specification with $K_C = K_P = K_Q = 3$ will be the benchmark, but I will also consider other cases.

5.1. The cross-section of returns

I first present the critical asset pricing results before investigating the structure and properties of the estimated factors in the next section. I compare the 3D-PCA results to those obtained from factor models obtained from PCA, Lettau and Pelger (2020a)'s Risk Premium-PCA (RP-PCA), and models with Fama-French factors. Consider an asset pricing model that consists of L factors, $\mathbf{F}_t^L$, that 3D-PCA, PCA, or RP-PCA estimate. Given that 3D-PCA and PCA factors are linear combinations of excess portfolio returns, they are themselves excess returns. I therefore run N time-series regressions of excess returns of portfolios $i = 1, \dots, N$ on a constant and factor returns $\mathbf{F}_t^L$:

$$R_{i,t+1}^e = \alpha_i + \boldsymbol{\beta}_i^\top \mathbf{F}_t^L + e_{i,t+1}, \quad i = 1, ..., N, \tag{25}$$

where $R_{i,t+1}^e$ is the excess return of portfolio i in quarter $t+1$. The pricing error of portfolio i is the intercept α_i. I evaluate the performance of the model by the (pseudo) cross-sectional R^2:

$$R_{xs}^2 = 1 - \frac{\frac{1}{N}\sum_{i=1}^N \alpha_i^2}{\text{Var}_{xs}(\bar{R}_i)}, \tag{26}$$

where $\bar{R}_i$ is the mean excess return of portfolio i and $\text{Var}_{xs}(\bar{R}_i)$ is the cross-sectional variance of mean returns. Note that the pricing errors in the nominator of (26) are not demeaned so that the mean pricing error is taken into account. As a result, R^2_{xs} is not a "proper" R^2 and does not have to be positive. I also compute the Sharpe ratio that is generated by the factors $\mathbf{F}_t$:

$$\text{SR} = \sqrt{\bar{\mathbf{F}}^\top \boldsymbol{\Sigma}_F^{-1} \bar{\mathbf{F}}}, \tag{27}$$

where $\bar{\mathbf{F}} = (\bar{F}_1, ..., \bar{F})^\top$ is the vector of factor means and $\boldsymbol{\Sigma}_F$ is the variance-covariance matrix of factor returns.

To obtain parsimonious factor models, I select subsets of the K factors as follows. I start with the first factor $\mathbf{F}^1_t = [F_{1t}]$ and estimate (25). I then add each of the remaining $K-1$ factors individually and re-estimate (25). I add the factor that yields the highest R^2_{xs} to $\mathbf{F}^1_t$ and obtain $\mathbf{F}^2_t$. I add the third, fourth, ... factors using the same recursive procedure and stop when the incremental R^2_{xs} becomes small. Note that the chosen factors for a given $L > 2$ might not yield the highest R^2_{xs} among all possible combinations of L factors, but this procedure avoids searching over a large set of factor combinations.

Table 2 reports results for 3D-PCA, Fama-French, PCA, and RP-PCA factors for up to $L = 6$. The table shows the cross-sectional R^2_{xs}, the annualized Sharpe ratio of the included factors, and the factor added at step L. Panel A and B report results for in-sample and out-of-sample factors, respectively. Results for models with in-sample 3D-PCA factors are in the first three columns of Panel A. F^{3D}_{cpq} is the 3D-PCA factor that is created by the c-th, p-th, and q-th columns of the $\mathbf{V}^{(C)^\top}$,$\mathbf{V}^{(P)^\top}$, and $\mathbf{V}^{(Q)^\top}$ matrices, see (17) and (18). The in-sample 1-factor 3D-PCA model that only includes F^{3D}_{111} has a negative R^2_{xs} and thus yields a poor fit. As we will see in the next section, the first factor, F^{3D}_{111}, is a long-only factor that is highly correlated with the CRSP-VW. Hence, it is not surprising that the R^2_{xs} of the factor model with only F^{3D}_{111} is similar to that of the CAPM (-0.58). The model with F^{3D}_{111} and F^{3D}_{112} as factors improves the fit considerably, as indicated by the higher cross-sectional R^2 of 0.37. Adding additional 3D-PCA factors improves the fit further. For example, the model with $L = 3$ factors captures 65% of the variation in mean returns and the specification with six factors captures 76%s of the variation in mean returns. The third column of the table list the factors that are included in the respective specifications. I will return to the interpretation of the factors in the next section.

Next, consider factor models based on the CRSP-VW index (MKT) and Fama-French factors SMB, HML, CMA, RMW, and MOM. The in-sample R^2_{xs} of the CAPM is -0.58, confirming the well-established result that the CAPM cannot capture the cross-section of returns. The negative value of R^2_{xs} arises because the mean pricing error of 1.21% is non-zero. Adding CMA raises R^2_{xs} to 0.01. However, at least five factors are required to render an R^2_{xs} above 50%. The model with MKT, CMA, MOM, SMB, and RMW yields an R^2_{xs} of 0.57. Adding the last factor, HML, does not improve the fit of the model further.

Note that the in-sample Sharpe ratios of 3D-PCA factor models are significantly larger than those of comparable Fama-French models. For example, the 3D-PCA model with $L = 3$ factors generates a Sharpe ratio of 1.57, while the model with MKT, CMA, and MOM as factors has a maximum Sharpe

ratio of 1.00. The Sharpe ratio of the model with six 3D-PCA factors is 1.95 compared to 1.21 for the model with the MKT and all five Fama-French factors. Hence, factor models based on 3D-PCA factors yield smaller pricing errors and higher Sharpe ratios than comparable models with Fama-French models.

The last six columns of Panel A report results for PCA and RP-PCA factors.[10] Given L, the R^2_{xs} and Sharpe ratios are lower than those of 3D-PCA models but higher than those of Fama-French factors. The Sharpe ratios of RP-PCA models are generally higher than those of PCA models, but their cross-sectional R^2s are similar.

Since in-sample factors are estimated using the entire sample, they are subject to look-ahead bias. Panel B reports results for factors that are estimated out-of-sample and do not use future information.[11] The out-of-sample R^2_{xs} of models with 3D-PCA factors are similar to those with in-sample 3D-PCA factors but larger for some specifications. For example, the model with $L = 4$ and $L = 5$ factors capture 72% and 76% of the variation in mean returns of the 275 portfolios compared to 67% and 68% for the corresponding in-sample models. The out-of-sample Sharpe ratios of these two models are also higher than those for in-sample factors. However, the in-sample Sharpe ratio of the $L = 6$ model is significantly higher than its out-of-sample Sharpe ratio. These results suggest that 3D-PCA estimations are stable over time and that the rolling estimation picks up changes in the dependence structure of portfolio returns over times that the full-sample estimation ignores. I will explore this possibility further below.

In contrast to 3D-PCA models, the models with out-of-sample PCA and RP-PCA factors perform considerably worse than the corresponding in-sample models in most cases. Even 6-factor models explain no more than 51% and 48%, respectively, of the variation in mean returns, which is lower than the R^2_{xs} of the 6-factor Fama-French model. The Sharpe ratios of out-of-sample PCA and RP-PCA models are also considerably lower than their in-sample counterparts.

Next, I run horse races between multiple factor types. The first row of Table 3 reports results for in-sample 3D-PCA factors. The '**F**' column shows the R^2_{xs} of the model with $L = 5$ 3D-PCA factors, see Table 2. The '+FF' column reports the cross-sectional R^2 when four Fama-French factors are added to the regression. The added factors are those included in the 5-factor specification in Table 2: CMA, MOM, SMB, RMW. I exclude MKT because it is essentially co-linear with F^{3D}_{111}.[12] The resulting R^2_{xs} of 0.65 is slightly lower than 0.68 suggesting that Fama-French factors do not add explanatory power to 3D-PCA factors. The '+PCA' and '+RP-PCA' columns add four PCA and RP-PCA factors, respectively, and also yield larger pricing errors. The results for out-of-sample factors in Panel B are similar. In contrast, adding 3D-PCA factors to benchmark factors generally improves the fit, although the effect is small for in-sample factors. For out-of-sample factors the improvement is more significant. The R^2_{xs} for the Fama-French model rises from 0.57 to 0.65 when 3D-PCA factors are added. For PCA and RP-PCA models, the R^2_{xs} improve from 0.50 to 0.66, and from 0.46 to 0.60,

[10] The tuning parameter γ is set to 15.

[11] Since the rolling windows are of length $h = 120$, the effective sample starts in July 1973.

[12] Because of co-linearity, none of the first factors are added in any specification reported in Table 3. However, the results are similar when the first factors are included.

respectively. I conclude that none of the benchmark factors improve the fit of models with only 3D-PCA factors but the additions of 3D-PCA factors reduce pricing errors of Fama-French as well as PCA and RP-PCA factor models.

The results in Tables 2 and 3 are based on models that use all 27 factors. Table 4 reports cross-sectional R^2_{xs} for models with subsets of the 27 3D-PCA factors for $L = 4$ and $L = 6$. K_C, K_P, and K_Q are the number of factors in the characteristic (C), size-quintile (P), and characteristic-quintile (Q)dimensions, respectively. For example, the case $K_C = 1, K_P = 3, K_Q = 2$ has a single characteristic factor, two size-quintile factors, and three characteristics. For each combination shown in Table 4, I use the same recursive procedure to pick which L factors are included. When the number of possible factors is lower than L, I report the R^2_{xs} of the model with all factors. Consider first the results for in-sample factors in Panel A and only one C factor, $K_C = 1$, reported in the first three columns. The R^2_{xs} for all models that include only the first Q factor, $K_Q = 1$, are negative, indicating that at least two Q-factors are required to achieve a reasonable fit. On the other hand, models with only one P-factor ($K_P = 1$) have R^2_{xs} of 37% and 46%, respectively, when the second and third Q-factors are included ($K_Q = 2, 3$). Combining multiple P and Q factors improves the fit substantially. For example, the R^2_{xs} of the specification with $K_C = 1, K_Q = 2, K_P = 3$ is 55% for $L = 4$ and 65% for $L = 6$. However, it has only eight factors compared to 27 of the full model. The fit of this specification for out-of-sample factors is slightly higher with 64% and 69%.

I conclude that parsimonious models with 3D-PCA factors successfully capture the cross-section of mean returns of the 275 portfolios in the sample. They outperform Fama-French and PCA-based models by a considerable margin. Moreover, the out-of-sample 3D-PCA factors yield as good a fit as in-sample factors. The rest of this section analyzes the fit in more detail. Since the in-sample estimations are subject to a look-ahead bias, I focus on the results of out-of-sample estimations of models with $L = 6$ factors.

Since it is difficult to visualize the pricing errors along all three dimensions, I aggregate them by dimensions. For example, the pricing error for a characteristic is the root-mean-square pricing error (RMSPE) of the errors of the 25 ($P \times Q$) portfolios of that characteristic. I also compute the RMS pricing errors for all 25 size and characteristic quintile combinations. Figure 3 plots the annualized RMSPE for each characteristic as well as the mean RMSPE across all 11 characteristics for the 3D-PCA (in blue), the Fama-French (in orange), and the PCA (in black) models. The average RMSPE of 3D-PCA is 1.23% compared to 1.60% and 1.72% for the Fama-French and PCA models. For some characteristics, such as OP, REV, and BETA the pricing errors of the 3D-PCA, Fama-French, and PCA models are similar. For the other characteristics, the 3D-PCA model has the lowest pricing errors and yields relatively small errors for those characteristics that have a poor fit for the PCA and Fama-French models. For example, the 3D-PCA pricing errors for VAR and RVAR are below 1.2% while the errors of the PCA and Fama-French model are around 2% or higher. The highest pricing errors of the PCA and Fama-French models are for MOM and SREV, respectively. SREV has the highest average pricing error across all three models.

Next, consider the pricing errors by ($P \times Q$) portfolios shown in Figure 4. The heatmaps indicate

higher RMSPE in darker shades. Consider first the 3D-PCA model in Panel A. The four corner portfolios (and P4Q5) are associated with the highest pricing errors, which is the typical pattern in the literature. Moreover, the small stock portfolios P5Q1 and P5Q5 have larger errors than the large stock portfolios P1Q1 and P1Q5. The mean RMSPE across all 25 portfolios is 1.20%. Pricing errors of the Fama-French, PCA, and RP-PCA models are generally higher than those of the 3D-PCA model. The difference is particularly large for small/low characteristic quintile portfolios. The RMSPE of the P5Q1 portfolio are very high and range from 4.82% for PCA to 5.57% for RP-PCA. The pricing errors of the P4Q1 portfolio are also large and significantly higher than for the 3D-PCA model.

Figure 5 plots the fitted mean returns on the x-axis and the actual mean returns on the y-axis, so that the distance to the 45-degree line is the pricing error of a particular portfolio. To make the plots readable, they only include the four corner portfolios P1Q1, P1Q5, P5Q1, and P5Q5 of the 11 characteristics. The plots also show the R^2_{xs} of only the 44 corner portfolios (instead of the R^2_{xs} of all 275 portfolios in the previous tables and figures) to assess the fit of these portfolios that are most difficult to price. The fitted returns of the 6-factor 3D-PCA model are shown in Panel A. The plot confirms the good fit of the model as all portfolios are located relatively close to the 45-degree line. The portfolios with the largest pricing errors are SREV-P5Q5 (4.13%) and SREV-P5Q1 (-3.63%). The cross-sectional R^2 of the 44 corner portfolios is 76%. Recall from Table 2 that the 3D-PCA with $L = 4$ factors has almost as good a fit as the model with six factors, which is confirmed by the plot in Panel B. The fitted returns of the 4-factor model are virtually identical to those of the 6-factor model, and the $R^2_{xs} = 73\%$ is only slightly lower.

The fitted returns of the Fama-French and PCA models with $L = 6$ factors show that their fit is significantly worse. The R^2 of the Fama-French and PCA models are 39% and 30%, respectively, and an order of magnitude lower than those of the 3D-PCA models. The plots show that some portfolios in both models have large pricing errors (in absolute values). The P5Q1 RVAR, VAR, SREV, NSI, and BM have pricing errors of at least -4%, and the MOM and SREV P5Q5 portfolios have errors of above 3.5%. The fit of the PCA model is worse. The pricing errors of six portfolios is lower than -4% and four portfolios have errors above 3.5%.[13]

The out-of-sample results reported above are based on rolling windows of length $h = 120$. The rows of Table 5 show the R^2_{xs} of out-of-sample 3D-PCA and PCA models for different lengths of the rolling windows: $h = 12, 36, 60, 120, 180, 360$. The columns show models with $L = 2, 4, 6, 8, 10$ factors. Note that the results across different values of h are not comparable since the effective sample starts in period $h + 1$, so the time series regressions (25) span different samples. However, comparisons of the 3D-PCA and PCA R^2_{xs} for a given h are valid.

Consider first the case with only three time series observations, $h = 3$. The R^2_{xs}'s of all PCA models are negative, indicating that estimation in such short samples is infeasible. In contrast, 3D-PCA models yield a reasonably good fit with R^2_{xs} around 50% for all $L > 2$. Increasing h to 6 and 12 yields R^2_{xs} between 50% and 70%. 3D-PCA yields similar fits for window lengths between 12 and 180 while increasing h to 360 deteriorates the R^2_{xs} slightly. PCA requires at least 60 time series

[13]The portfolios are P5Q1 of VAR, RVAR, MOM, SREV, NSI, INV, and the MOM and INV P5Q5 and AC P1Q5 portfolios.

observations to yield reasonable results; however, the R^2_{xs} is lower than that of 3D-PCA for all cases. The reason why 3D-PCA estimation is feasible in very short samples is, of course, that the number of free parameters, in contrast to PCA, is small. Recall from the previous section that the number of free parameters in PCA is an order of magnitude higher than that of 3D-PCA. In the case considered here, 3D-PCA has only 48 free parameters compared to 7,372 for PCA, which implies that the factor weights are more stable in short samples, as I will show in the next section,

In summary, the 3D-PCA models provide robust in-sample and out-of-sample fits of the cross-section of mean returns of the 275 portfolios in the sample. The model is parsimonious with relatively few free parameters, and can be successfully estimated in very short samples. The specifications with 3D-PCA factors consistently outperform popular benchmark in-sample, particularly out-of-sample estimations. They provide substantially better fits of the extreme size and characteristics corner quintiles, which are a challenge for many asset pricing models. The following section studies the estimates of the components of the 3D-PCA model (13), the properties of the 3D-PCA factors, and the construction of the portfolio weights that underlie the factors.

5.2. Factor weights

The estimation of the 3D-PCA model yields the $(T\times K_C\times K_P\times K_Q)$-dimensional tensor of factors and three matrices $\mathbf{V}^{(C)}, \mathbf{V}^{(P)}$, and $\mathbf{V}^{(Q)}$ of dimensions $(K_C\times C), (K_P\times P)$, and $(K_Q\times P)$, respectively. Recall that the weights that generate the factors are determined by the transposes of $\mathbf{V}^{(C)}, \mathbf{V}^{(P)}$, and $\mathbf{V}^{(Q)}$. I first present the in-sample estimates of the three $\mathbf{V}$ matrices and then study the implications for factor weights. Weights derived from the out-of-sample estimation are discussed in the following subsection.

Figure 6 shows the estimated matrices $\widehat{\mathbf{V}}^{(i)}$ as heatmaps with the point estimates displayed in each cell. Positive values are in blue, and negative values are in red. While the columns have a unit norm, their signs are not identified. I set the signs so that the interpretations of factors line up with standard portfolio sorts and risk factors whenever possible. The rows and columns of $\widehat{\mathbf{V}}^{(C)}$ in Panel A correspond to the 11 characteristics and C-factors, $k_C = 1, 2, 3$. All elements of the first columns, $\mathbf{v}_1^{(C)}$, are positive and represent "long-only" portfolio weights. The weights are between 0.29 and 0.31, thus similar across characteristics, creating almost an equal-weighting scheme. The second and third columns include positive and negative values and, therefore, correspond to "long-short" weights. The largest positive weights of the second factor, $\mathbf{v}_2^{(C)}$, are for VAR and RVAR, while the weight of BETA is by far the lowest, followed by SREV, REV, and AC. The elements of the other characteristics are small. Hence, the second factor has the interpretation of an (approximately) long-VAR and RVAR and short-BETA weighting scheme. The third factor, $\mathbf{v}_3^{(C)}$, is dominated by MOM with a value of 0.88, followed by BM (-0.36) and NSI (-0.25) and can be interpreted as a long-MOM and short-BM factor.

Characteristic weights in the second and third factors are related to the volatilities of long-short portfolios. Results in Table 1 showed that volatilities of SMB_c portfolios by characteristic are similar across characteristics; however, the volatilities of characteristic-specific HML_c portfolios vary

significantly. For example, the HML_c portfolio created by a size/accruals double-sort has a volatility of 4.91%, whereas HML_c of several other characteristics have standard deviations over 15%. In fact, the characteristics with the highest HML_c volatilities have the highest weights in $\mathbf{v}_2^{(C)}$ and $\mathbf{v}_3^{(C)}$ (in absolute value), namely, VAR (19.55%), RVAR (17.93%), MOM (17.22%), and BETA (16.86%), followed by SREV (12.96%) and BM (12.59%). Of course, factor weights depend on the complex correlation structure of the test assets; however, we will see below that factors with high-minus-low structures of these factors play an essential role.

The $\mathbf{V}^{(P)}$ and $\mathbf{V}^{(Q)}$ matrices determine factor weights along the size and characteristic dimensions, respectively. Both follow similar patterns that have a familiar structure. The first P-factor, $\mathbf{v}_1^{(P)}$, is long-only, but the weights increase in P-quintiles, so that the corresponding factor overweighs the small-stock quintiles and underweight the big-stock quintiles. $\mathbf{v}_2^{(P)}$ is monotonically increasing in size quintiles ranging from -0.66 for the P1 to 0.62 for the P5 quintile and has, therefore, a small-minus-big pattern. In contrast, the P1 and P5 elements of $\mathbf{v}_3^{(P)}$ are positive, while the P2, P3, and P4 elements are negative, creating a long-small/big and short-medium pattern.

The shape of the $\mathbf{V}^{(Q)}$ matrix is similar to that of $\mathbf{V}^{(P)}$. The first Q-factor, $\mathbf{v}_1^{(Q)}$, is positive and thus long-only. The second column, $\mathbf{v}_2^{(Q)}$, is increasing from Q1 to Q5 and has the interpretation of a high-minus-low factor. Like the third P-factor, the third Q-factor is positive in Q1 and Q5 and negative in Q2, Q3, and Q4. The columns of $\mathbf{V}^{(P)}$ and $\mathbf{V}^{(Q)}$ have therefore a level-slope-curvature pattern often seen in PCA application. The first "level" factor is long-only with approximately similar values, the second factor has a "slope" pattern that is monotonically increasing in the quintile portfolios, and the third factor is convex and resembles "curvature".

As argued in section 4 the columns of $\mathbf{V}^{(C)}, \mathbf{V}^{(P)}$, and $\mathbf{V}^{(Q)}$ form building blocks for factor weights. Next, I study the portfolios that can be constructed from the estimated $\hat{\mathbf{V}}^{(i)}$. Consider first the P and Q-dimensions. The outer product of the p-th and q-th columns of $\mathbf{V}^{(P)}$, and $\mathbf{V}^{(Q)}$ is the (5×5) matrix $\mathbf{W}_{pq}^{(PQ)} = \mathbf{v}_p^{(P)} \circ \mathbf{v}_q^{(Q)}$. The elements of $\mathbf{W}_{pq}^{(PQ)}$ are weights for the 25 possible $(P\times Q)$ combinations of quintile portfolios. Since $K_P = K_Q = 3$, there are nine possible combinations of the columns of $\mathbf{V}^{(P)}$, and $\mathbf{V}^{(Q)}$. Each of the nine combinations represents a different weighting scheme for the P and Q dimensions that can be constructed from the estimates $\hat{\mathbf{V}}^{(P)}$ and $\hat{\mathbf{V}}^{(Q)}$.

Figure 7 shows all nine $\mathbf{W}_{pq}^{(PQ)}$ matrices that are generated by the estimated $\hat{\mathbf{V}}^{(P)}$ and $\hat{\mathbf{V}}^{(Q)}$ shown in Figure 6. The first matrix $\mathbf{W}_{11}^{(PQ)}$ combines the first columns $\mathbf{v}_1^{(P)}$ and $\mathbf{v}_1^{(Q)}$. Since both columns are long-only, $\mathbf{W}_{11}^{(PQ)}$ is a positive matrix. Since $\mathbf{v}_1^{(P)}$ underweights P1, the P1 weights of $\mathbf{W}_{11}^{(PQ)}$ are somewhat lower than those P2 to P5 weights. Next, consider combining the first column of $\hat{\mathbf{V}}^{(P)}$ and and the second column of $\hat{\mathbf{V}}^{(Q)}$. $\mathbf{v}_1^{(P)}$ is long only while $\mathbf{v}_2^{(Q)}$ is monotonic in the Q1 to Q5 quintiles. Hence, the columns of $\mathbf{W}_{12}^{(PQ)}$ (Panel B) retain the monotonic structure in the Q-dimension while its rows, representing the P-dimension, are almost identical. The smaller values (in absolute terms) of the P1 row is again due to the underweighting of the P1 portfolio in $\mathbf{v}_1^{(P)}$. Therefore, $\mathbf{W}_{12}^{(PQ)}$ has a high-minus-low structure that is long in high-characteristic quintile Q5 and short in the low-characteristic quintile Q1 while combined with equally weighing the size dimension. By the same token, the outer product of the second P-factor and the first Q-factor, $\mathbf{W}_{21}^{(PQ)}$, has a similar structure but with the P

and Q dimensions reversed since $\mathbf{v}_1^{(P)}$ is monotonic in P1 to P5 and $\mathbf{v}_1^{(Q)}$ is long-only. Thus, $\mathbf{W}_{21}^{(PQ)}$ has a small-minus-big pattern with (approximately) equal weights in the Q dimension. In other words, $\mathbf{W}_{12}^{(PQ)}$ and $\mathbf{W}_{21}^{(PQ)}$ are similar to HML and SMB but are estimated rather than constructed from fixed weights.

Combining the first and third columns of $\hat{\mathbf{V}}^{(P)}$ and $\hat{\mathbf{V}}^{(Q)}$ generates $\mathbf{W}_{13}^{(PQ)}$, shown in Panel C, retains the approximately equal weighting in the P dimension with the convex structure of $\mathbf{v}_3^{(Q)}$. Hence, all P1 to P5 rows have positive Q1 and Q5 and negative Q2, Q3, and Q4 weights. As for $\mathbf{W}_{21}^{(PQ)}$, the pattern is reversed for $\mathbf{W}_{31}^{(PQ)}$, so that each column has the curvature pattern in the size dimension.

The $\mathbf{W}_{jk}^{(PQ)}$ weight matrices mentioned so far combine a 'long-only' vector with a vector in the other dimension that has a level, slope, or curvature structure. Thus, all portfolios that can be generated retain a level, slope, or curvature structure in one dimension with approximately equal weighting in the other dimension. The outer product of the second columns, $\mathbf{v}_2^{(P)}$ and $\mathbf{v}_2^{(Q)}$ combine two monotonic (slope) vectors. The resulting matrix $\mathbf{W}_{22}^{(PQ)}$ is shown in Panel E. The weights of the diagonal P1Q1 and P5Q5 corner portfolios are positive, while the weights of the off-diagonal corner portfolios P1Q5 and P5Q1 are negative. The other elements of $\mathbf{W}_{22}^{(PQ)}$ are an order of magnitude smaller. Hence, the pattern is similar to the CROSS portfolios introduced in section 3 (see Table 1). Almost all results in this paper are based on these six $\mathbf{W}_{jk}^{(PQ)}$ matrices. For example, the optimal 3D-PCA models reported in Table 4 include only factors based on $\mathbf{W}_{11}^{(PQ)}, \mathbf{W}_{12}^{(PQ)}, \mathbf{W}_{13}^{(PQ)}, \mathbf{W}_{22}^{(PQ)}$, and $\mathbf{W}_{31}^{(PQ)}$. The remaining higher order matrices $\mathbf{W}_{23}^{(PQ)}, \mathbf{W}_{32}^{(PQ)}$, and $\mathbf{W}_{33}^{(PQ)}$ will play no further role. These matrices combine slope and curvature vectors and create complex P and Q combinations.

Since the $\mathbf{W}_{jk}^{(PQ)}$ are based on combinations of the same column vectors, they all share common structures that combine the level, slope, and curvature patterns of the column vectors. Hence, all $\mathbf{W}_{jk}^{(PQ)}$ have straightforward economic interpretations. Moreover, some weight matrices exhibit patterns that resemble long-short portfolios that have been studied in the literature. Rather than using ad-hoc weights, the weights are estimated from the data.

To obtain weights for all $N = 275$ portfolios, $\mathbf{W}_{jk}^{(PQ)}$ matrices are combined with the columns of the $\hat{\mathbf{V}}^{(C)}$ matrix (see Panel A in Figure 6). The outer product of a (11×1)-dimensional column vector $\mathbf{v}_i^{(C)}$ and a (25×25)-dimensional matrix $\mathbf{W}_{jk}^{(PQ)}$ yields a $(11\times25\times25)$-dimensional tensor of weights $\boldsymbol{\mathcal{W}}_{ijk}$. Since $K_C = K_P = K_P = 3$, each of the 27 combinations of $i, j, k = 1, 2, 3$ represents the weight tensor $\boldsymbol{\mathcal{W}}_{ijk}$ of a factor F_{ijk}^{3D}. As for the $\mathbf{W}_{jk}^{(PQ)}$ matrices, the $\boldsymbol{\mathcal{W}}_{ijk}$ tensors retain the structure of the vectors and matrices on which they are built. Recall that the first column of $\hat{\mathbf{V}}^{(C)}$ is long-only with similar weights for all characteristics. Hence, all factor weights based on $\mathbf{v}_1^{(C)}$ will also have this property. The second column, $\mathbf{v}_2^{(C)}$, is long in VAR and RVAR, and short in BETA, while the third column is long in MOM and short in BM and NSI. As before, factor weights derived from these columns retain their structures.

The factor weight matrix $\mathbf{W}^{\text{3D}}$ is (275×27)-dimensional matrix $\mathbf{W}^{\text{3D}}$. Figure 8 shows a subset of rows of $\mathbf{W}^{\text{3D}}$ as a heatmap. For each characteristic, I plot weights of the P1Q1, P1Q3, P1Q5, P3Q1, P3Q3, P3Q5, P51Q1, P5Q3, and P5Q5 portfolios. Each column of the heatmap corresponds to the weights of a factor F_{ijk}^{3D}. Positive weights are in blue, and negative weights are in red. While the

heatmap is complex, common structures emerge from the figure. The first nine columns show factors $F^{3D}_{1\bullet\bullet}$ that are based on the first column of $\mathbf{v}^{(C)}_1$, the next nine columns correspond to factors generated by $\mathbf{v}^{(C)}_2$, and the last nine columns are computed from $\mathbf{v}^{(C)}_3$. The heatmap shows that factors $F^{3D}_{2\bullet\bullet}$ are dominated by VAR, RVAR, and BETA, while factors $F^{3D}_{3\bullet\bullet}$ depend on MOM, BM, and NSI.

By the same token, factors $F^{3D}_{\bullet 12}$ and $F^{3D}_{\bullet 21}$ retain the HML and SMB structures of $\mathbf{W}^{(PQ)}_{12}$ and $\mathbf{W}^{(PQ)}_{21}$. For example, the Q5 weights of $F^{3D}_{112}, F^{3D}_{212}$ and F^{3D}_{312} are positive, the Q3 weights are close to 0, and the Q1 weights are negative. Similarly, $F^{3D}_{121}, F^{3D}_{221}$ and F^{3D}_{321} have the same structure but for size quintiles P5, P3, and P1. The other factors have similar interpretations.

For comparison, Figure 9 plots the corresponding weight matrix derived from a PCA estimation with 27 factors. Aside from the first long-only factor, PCA weights show little common structure. For example, factors are not related to particular characteristics but involve portfolios of many characteristics. In contrast, 3D-PCA factors are associated with all characteristics ($F^{3D}_{1\bullet\bullet}$) or small subsets of characteristics ($F^{3D}_{2\bullet\bullet}$ and $F^{3D}_{3\bullet\bullet}$). Moreover, the weights of most PCA factors do not have clear patterns across size and characteristic quintiles. Figure C.9 shows heatmaps of factor weights of the first nine PCA factors. Only two factors have easily interpretable patterns.[14] The weights of the second and fifth factors are related to size. F^{PC}_2 reflect big-minus-small since its weights are positive for P1 and negative for P5 quintiles. The weights of the fifth factor are positive for P1 and P5 quintiles and negative for P3, creating a "size-curvature" pattern.

5.3. Time-variation in factor weights

In the out-of-sample estimation, factor models are estimated in rolling windows with h data points. In the benchmark specification, h is set to 120 months. Figure 10 plots the first four 3D-PCA factor weights (left column) and PCA weights (right column). Each panel shows the time series of all $N = 275$ portfolio weights of a given factor and reports the average time series standard deviation in the top left corner. The estimates of the weights of the first (long-only) 3D-PCA factors in Panel A are stable over time; however, they "fan" out between 2000 and 2010. The pattern of the weights of the first PCA factor (Panel B) is similar, but the estimates are not as stable over time. The average standard deviation is 0.58% compared to 0.36% for the first 3D-PCA factor.

The behaviors of the second, third, and fourth 3D-PCA factors in Panels C, E, and G are similar. The weights show little variation over time, and their average standard deviations are between 0.62% and 0.79%. As for the first factor, the 2000 to 2010 period is associated with somewhat higher instability. In contrast to the 3D-PCA estimates, the weights of the second to fourth PCA factors are significantly more unstable throughout the sample. Their mean standard deviations range from 2.32% to 4.73% and are an order of magnitude higher than those of 3D-PCA factors. On average, the standard deviations of the 27 PCA factors is 5.05% compared to 1.84% for 3D-PCA factors.

[14] Since the signs of PCA factor weights are not identified, I normalize the signs so that the mean returns of PCA factors are positive.

The results in Figure 10 are based on estimations in rolling windows of length $h = 12$ months. Figure 11 plots the means standard deviation of all 27 factor weights of 3D-PCA (in blue) and PCA (in orange) for different values of h ranging from 12 to 600 months. 3D-PCA weights are more stable than PCA weights for any choice of h. The mean standard deviation of 3D-PCA is 4.07% for subsamples with only 12 time series observations decreases to below 2% for $h = 120$ and below 1% for $h = 360$. For PCA, the average standard deviation is 5.28% for $h = 12$ and declines slowly for longer subsamples. Windows with about 360 observations are required for a standard deviation that compares to 3D-PCA estimations with only 12 time series observations. These results suggest that the estimation of 3D-PCA factor weights in shorter samples is more reliable than that of PCA estimation.

As noted above, the period between 2000 and 2010 showed different behavior of subsample weights in Figure 10. To investigate this behavior, I compute pairwise correlations of all $N = 275$ portfolios in rolling windows of length 24. Panel A of Figure 12 plots the mean correlation across time. Until the early 1990s, the average correlation tends to be above 0.8 but then declines to between 0.7 and 0.8. In early 2000, there is a precipitous drop to below 0.5 before correlations rise again until the end of the sample. The decline in the average correlation coincides with the fanning out of portfolio weights in subsamples that include the early 2000s. To pinpoint the change in the behavior of returns, I compute the cross-sectional standard deviation of portfolio returns in each month. The time series of standard deviations are plotted in Panel B in orange while blue line shows an MA(12). The cross-sectional standard deviation spikes in February 2000 to a value of 12.33; by far the highest value in the sample.[15] This change in the behavior of returns affects the estimation of factors in all subsamples, including February 2000. Since subsamples include 120 months, the effect vanishes after 2000.

5.4. Factor returns

Table 6 shows the 3D-PCA and PCA factors with the highest annualized Sharpe ratios (in absolute value). The annualized Sharpe ratios of the CRSP-VW index and Fama-French factors SMB, HML, RMW, CMA, and MOM are included as benchmarks. The Sharpe ratio of the CRSP-VW index is 0.42. MOM has the highest Sharpe ratio of 0.50 among the Fama-French factors, followed by RMW and CMA with 0.46 and 0.42, respectively. The Sharpe ratios of HML and SMB are 0.32 and 0.20.

The 3D-PCA factor with the highest Sharpe ratios is F^{3D}_{112}. Its in-sample SR is 1.23 is higher than the SR of the CRSP-VW index by a factor of three and higher than the highest SR of the Fama-French factors (MOM) by a factor of 2.5. Recall that F^{3D}_{112} combines the first column of $\hat{\mathbf{V}}^{(P)}$ and the second column of $\hat{\mathbf{V}}^{(Q)}$, which corresponds to "level" in the size dimension and "slope" in the characteristic dimension. The factor F^{3D}_{122} has the second highest SR with 0.90 F^{3D}_{122} combines the second columns of $\hat{\mathbf{V}}^{(P)}$ and $\hat{\mathbf{V}}^{(Q)}$ to create a "cross" factor that is long in the P1Q1 and P5Q5 quintiles and short in P1Q5 and P5Q1 quintiles. This strategy has a positive return for most characteristics (see Table 1).

[15] Portfolio returns exhibit an unusually large spread across most characteristic quintiles in February 2000, causing the cross-sectional spread to spike.

Combined with the long-only first column of $\hat{\mathbf{V}}^{(C)}$, the resulting factor has a positive mean return and a high Sharpe ratio. The third highest Sharpe ratio (in absolute terms) of -0.72 is due to F^{3D}_{113}. F^{3D}_{113} combines the first column of and third columns of $\hat{\mathbf{V}}^{(P)}$ and $\hat{\mathbf{V}}^{(Q)}$ to create a "'curvature" factor that is long in the Q1 and Q5 quintiles and short in Q2, Q3, and Q4 quintiles. On average, the return of this factor is negative.

Note that the four factors with the highest Sharpe ratios are based on the first column of $\hat{\mathbf{V}}^{(C)}$, which implies an (approximately) equal weighting of all characteristics. The factor with the fifth highest Sharpe is F^{3D}_{222} and thus based on the second columns of $\hat{\mathbf{V}}^{(C)}$, $\hat{\mathbf{V}}^{(P)}$, and $\hat{\mathbf{V}}^{(Q)}$. Recall from Figure 6 that the second column of $\hat{\mathbf{V}}^{(C)}$ is dominated by positive weights of VAR and RVAR and a large negative weight of BETA. Hence, F^{3D}_{222} has the interpretation of a "cross" strategy in the size/characteristic dimensions that is long in VAR and RVAR and short in BETA. The F^{3D}_{223} and F^{3D}_{213} factors in Table 6 are also based on the second column of $\hat{\mathbf{V}}^{(C)}$ and also represent long VAR/RVAR-short BETA strategies.

The results for out-of-sample 3D-PCA factors are reported in Panel B. Note that except F^{3D}_{222}, the same factors have high Sharpe ratios out-of-sample as in-sample. Second, the out-of-sample Sharpe ratios are comparable to their in-sample counterparts, which shows that the high in-sample Sharpe ratios are not due to a look-ahead bias. In contrast, the Sharpe ratios of out-of-sample PCA factors are generally lower than those of in-sample PCA factors. Three PCA factors have Sharpe ratios around 0.7, but only F^{PC}_3 has a comparable Sharpe ratio out-of-sample.

Panel A of Table 7 shows the pairwise correlations of out-of-sample factors with their in-sample counterparts. Although out-of-sample 3D-PCA models are estimated in relatively short rolling windows with 120 monthly observations, their returns are generally closely aligned with returns of full-sample factors. Except for F^{3D}_{213}, the correlations of factors that are constructed using the first and second columns of $\hat{\mathbf{V}}^{(C)}$ shown in the first two rows of Panel A are over 0.9. The correlations of 3D-PCA factors that stem from the third column of $\hat{\mathbf{V}}^{(C)}$ are somewhat lower but still at least 77%. The average correlation across all 27 factors is 0.91. In contrast, correlations of most out-of-sample and in-sample PCA and RP-PCA factors are substantially lower. Only the correlations of the first two factors are above 90%. The correlations drop off quickly for higher-order factors, below 0.2 beyond the seventh factor. These results suggest that the relative stability of out-of-sample 3D-PCA factor weights shown in Figure 10 implies that out-of-sample factors are highly correlated with in-sample factors. On the other hand, the instability of out-of-sample PCA weights yields out-of-sample factors that do not resemble the in-sample factors.

Recall from (22) that the contribution of a 3D-PCA factor to the overall variation in the data is given by its sum of squares, $\sum_t F^2_{cpq}$. Table 8 reports the sum of squares for the 27 3D-PCA factors. Consider first the results for in-sample factors in Panel A. As in many PCA applications, the first factor captures the majority of the data variation. The sum of squares of F^{3D}_{111} is 573.92, which is an order of magnitude larger than the second largest sum of squares, which is 28.52 for F^{3D}_{121}. It is instructive to compare the two factors that are created by slope factors in the P and Q dimensions with the level factors in the other dimensions, i.e. F^{3D}_{121}, which has a small-minus-big structure, and F^{3D}_{112}, which has a high-minus-low structure (see Figure 6). The sum of squares of F^{3D}_{121} is 28.52 and

considerably higher than that of F_{112}^{3D} (5.45). Therefore, F_{121}^{3D} captures a higher portion of the data variation than F_{112}^{3D}. On the other hand, F_{112}^{3D} has a higher Sharpe ratio (see Table 6) and plays an essential role in cross-sectional factor models (see Table 2). Hence, these two factors both play important but different roles. Other factors that contribute to the variation in the data are F_{212}^{3D},F_{131}^{3D}, and F_{312}^{3D}. The pattern is similar for out-of-sample factors; the factors that capture the most variation are identical, see Panel B.

6. Extension to (n,m)D-PCA

The 3D-PCA model can be extended to tensors of arbitrary dimensions Let $\boldsymbol{\mathcal{X}}$ be a m-dimensional tensor of size $(I_1 \times I_2 \times \cdots \times I_m)$. Suppose we are interested in factoring only a subset $\mathcal{N}$ of the dimensions. Let $\mathcal{N}'$ be the set of unfactored dimensions. Without loss of generality, assume that $\boldsymbol{\mathcal{X}}$ is arranged so that the first n dimensions are factored. Hence $\mathcal{N} = \{1,\ldots,n\}$ and $\mathcal{N}' = \{n+1,\ldots,m\}$. The implied factor model is called (n,m)D-PCA, where n of m dimensions are factored.

Let K_j, $j \in \mathcal{N}$ be the number of factors for the factored dimensions. Then, the partial-$\mathcal{N}$ Tucker decomposition is given by

$$\boldsymbol{\mathcal{X}} \approx \boldsymbol{\mathcal{F}}_{\mathcal{N}} \times_1 \mathbf{V}_{\mathcal{N}}^{(1)} \times_2 \mathbf{V}_{\mathcal{N}}^{(2)} \cdots \times_n \mathbf{V}_{\mathcal{N}}^{(n)} \tag{28}$$

where $\mathbf{V}_{\mathcal{N}}^{(j)}$ are $(I_j \times K_j)$-dimensional matrices and $\boldsymbol{\mathcal{F}}_{\mathcal{N}}$ is a $(K_1 \times \cdots \times K_n \times I_{n+1} \times \cdots \times I_m)$-dimensional tensor. Hence, the factored dimensions of $\boldsymbol{\mathcal{F}}_{\mathcal{N}}$ are of length K_j while the unfactored dimensions have the same size as $\boldsymbol{\mathcal{X}}$, I_j. As for 3D-PCA, the factors of the (n,m)D-PCA model are given by $\boldsymbol{\mathcal{F}}_{\mathcal{N}}$ and can be constructed as linear combinations of the elements of $\boldsymbol{\mathcal{X}}$:

$$\boldsymbol{\mathcal{F}}_{\mathcal{N}} = \boldsymbol{\mathcal{X}} \times_1 \mathbf{V}_{\mathcal{N}}^{(1)^\top} \times_2 \mathbf{V}_{\mathcal{N}}^{(2)^\top} \cdots \times_n \mathbf{V}_{\mathcal{N}}^{(n)^\top}. \tag{29}$$

As before, the columns of $\mathbf{V}_{\mathcal{N}}^{(1)^\top},\ldots,\mathbf{V}_{\mathcal{N}}^{(n)^\top}$ matrices are the building blocks of factor weights. The difference of (n,m)D-PCA compared to 3D-PCA is that the factor weights are based on a subset of the m dimensions of $\boldsymbol{\mathcal{X}}$.

The factor tensor $\boldsymbol{\mathcal{F}}_{\mathcal{N}}$ can be written as a matrix by folding the unfactored dimensions in $\mathcal{N}'$ as rows and the factored dimensions in $\mathcal{N}$ as columns: $\mathbf{F}_{\mathcal{N}} = \text{mat}_{\mathcal{N}}(\boldsymbol{\mathcal{F}}_{\mathcal{N}})$. The matrix of factors $\mathbf{F}_{\mathcal{N}}$ has $\prod_{k\in\mathcal{N}'} I_k$ rows and $\prod_{j\in\mathcal{N}} K_j$ columns. Each column of $\mathbf{F}_{\mathcal{N}}$ corresponds to one of the $\prod_{j\in\mathcal{N}} K_j$ factors.

We can apply this model to the data set used in this paper. $\boldsymbol{\mathcal{X}}$ is a $m = 4$-dimensional panel with dimensions time, characteristics, size-quintiles, and characteristic-quintiles. The 3D-PCA model in section 4 factored all dimensions but time; hence it is equivalent to a $(3,4)$D-PCA model. As an alternative, consider the case where only the size-quintile and characteristic-quintile dimensions are factored, yielding a $(2,4)$D-PCA model. $\boldsymbol{\mathcal{F}}_{\mathcal{N}}$ is $(T \times C \times K_P \times K_C)$-dimensional factor tensor and the factor matrix $\mathbf{F}_{\mathcal{N}} = \text{mat}_{\mathcal{N}}(\boldsymbol{\mathcal{F}}_{\mathcal{N}})$ has TC rows and $K_P K_Q$ columns. In the 3D-PCA model, the factor matrix $\mathbf{F}_{(T)}$ is $(T \times K_C K_P K_Q)$-dimensional so that each row corresponds to factors in a given month. In contrast, in the $(2,4)$D-PCA model, each row of $\mathbf{F}_{\mathcal{N}}$ corresponds to the $K_P K_Q$ factors of a characteristic in a month. In other words, factors are estimated for each characteristic in each month. The interpretation of the factors is as follows. Suppose one factor's weights have a high-minus-low struc-

ture across characteristic quintiles. Then, the corresponding column of $\mathbf{F}_{\mathcal{N}}$ includes monthly factor observation of this high-minus-low factor for all characteristics. For example, the BM elements represent the time series of the estimated high-minus-low factor for the book-to-market ratio, which is related to the Fama-French HML factor. The difference is that factor weights are estimated by the $(2,4)$D-PCA model rather than fixed. By the same token, RMW, CMA, and MOM are related to the OP, INV, and MOM entries of the columns of $\mathbf{F}_{\mathcal{N}}$.

Figure 13 shows the estimates of the $\widehat{\mathbf{V}}_{\mathcal{N}}^{(P)}$ and $\widehat{\mathbf{V}}_{\mathcal{N}}^{(Q)}$ matrices, which are almost identical to the corresponding matrices of the 3D-PCA shown in Figure 6. Hence, they have the same level, slope, and curvature interpretation, so that factor weights, given by the outer products of combinations of columns of $\widehat{\mathbf{V}}_{\mathcal{N}}^{(P)}$ and $\widehat{\mathbf{V}}_{\mathcal{N}}^{(Q)}$, are similar to those of 3D-PCA shown in Figure 7. Table 9 shows the Sharpe ratios of the nine in-sample and out-of-sample factors for each characteristic. The $F_{11}^{\mathcal{N}}$ factor is a combination of the first columns of $\widehat{\mathbf{V}}_{\mathcal{N}}^{(P)}$ and $\widehat{\mathbf{V}}_{\mathcal{N}}^{(Q)}$. Since both columns are long-only and (approximately) equal-weighted, $F_{11}^{\mathcal{N}}$ is (approximately) an equal-weighted average of the 25 double-sorted portfolios for each characteristic. The Sharpe ratios range from 0.41 for NSI to 0.51 for REV. The combination of the first column of $\widehat{\mathbf{V}}_{\mathcal{N}}^{(P)}$ and the second column of $\widehat{\mathbf{V}}_{\mathcal{N}}^{(Q)}$ yields a factor, $F_{12}^{\mathcal{N}}$, with weights that have the familiar high-minus-low pattern along the characteristic dimension. As mentioned above, this factor is related to the Fama-French factors, whose Sharpe ratios are reported in Table 6. The in-sample Sharpe of BM-$F_{12}^{\mathcal{N}}$ is 0.48, which is substantially higher than that of HML (0.32). In fact, Sharpe ratios of all $F_{12}^{\mathcal{N}}$ factors are higher than those of the corresponding Fama-French factors: 0.52 for OP-$F_{12}^{\mathcal{N}}$ compared to 0.46 for RMW, 0.67 for INV-$F_{12}^{\mathcal{N}}$ compared to 0.45 for CMA, and 0.69 for MOM-$F_{12}^{\mathcal{N}}$ compared to 0.50 for MOM.

The Sharpe ratios of the $F_{13}^{\mathcal{N}}$ are all negative. Since the weights underlying $F_{13}^{\mathcal{N}}$ have a convex curvature structure along characteristic quintiles, the inverse weights that are long in the middle quintiles and short in the top and bottom quintiles generate a factor with a positive mean return. The three highest (absolute) in-sample Sharpe ratios in Panel A are due to $F_{13}^{\mathcal{N}}$ factors. The NSI-$F_{13}^{\mathcal{N}}$ factor has the highest (absolute) Sharpe ratio among all factors with -0.94, followed by RVAR-$F_{13}^{\mathcal{N}}$ and VAR-$F_{13}^{\mathcal{N}}$ with Sharpe ratios of -0.86 and -0.78, respectively. Sharpe ratios of the other $(2,4)$D-PCA factors are generally smaller except for VAR and RVAR $F_{22}^{\mathcal{N}}$ and $F_{23}^{\mathcal{N}}$ factors, which are intersection of slopes in size and characteristic quintiles and slope in size and curvature in characteristic quintiles, respectively. The Sharpe ratios of out-of-sample $(2,4)$D-PCA follow the same patterns as their in-sample counterparts, confirming that the recursive estimation in short rolling windows yields stable results.

7. Conclusion

This paper considers an estimation method, 3D-PCA, for latent factors in multidimensional panels that exploits the structure of the data set. Factor weights are constructed from a small number of dimension-specific building blocks. The implied factor weights are subject to a set of constraints. Consequently, 3D-PCA has a relatively small number of free parameters compared to standard models based on PCA. The model is an implication of the Tucker tensor decomposition, which general-

izes the matrix SVD to higher dimensions. The weights and factors can be estimated by numerical recursive methods.

I estimate 3D-PCA using a 4-dimensional panel of time series of size and characteristic-sorted portfolio returns for a set of characteristics. The estimated building blocks that determine the factor weights have familiar level, slope, and curvature patterns. The factor weights are given by interactions of level, slope, and curvature across size, and characteristic quintiles. The factor weights can be estimated in short samples and are stable over time. In contrast, PCA requires longer samples and exhibits instability over time.

3D-PCA have substantially higher in-sample and out-of-sample Sharpe ratios than Fama-French and PCA factors. The 3D-PCA factors with the highest Sharpe ratios are related to slope and curvature along characteristic quintiles. Moreover, 3D-PCA factor models capture the cross-section of expected return in in-sample and out-of-sample estimations. Parsimonious models with 3D-PCA factors yield smaller pricing errors than larger models with Fama-French or PCA factors.

References

Andersson, Claus A. and Rasmus Bro (1998) "Improving the speed of multi-way algorithms: Part I. Tucker3," *Chemometrics and Intelligent Laboratory Systems*, 42 (1-2), 93–103.

Babii, Andrii, Eric Ghysels, and Junsu Pan (2022) "Tensor Principal Component Analysis," SSRN Electronic Journal.

Bryzgalova, Svetlana, Jiantao Huang, and Christian Julliard (2023) "Bayesian Solutions for the Factor Zoo: We Just Ran Two Quadrillion Models," *Journal of Finance*, 78 (1), 487–557.

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

Chen, Luyang, Markus Pelger, and Jason Zhu (2024) "Deep Learning in Asset Pricing," *Management Science*, 20 (2), 714–750.

Cochrane, John H. (2011) "Presidential Address: Discount Rates," *Journal of Finance*, 66 (4), 1047–1108.

Connor, Gregory, Matthias Hagmann, and Oliver Linton (2012) "Efficient Semiparametric Estimation of the Fama–French Model and Extensions," *Econometrica*, 80 (2), 713–754.

Connor, Gregory and Robert A. Korajczyk (1986) "Performance measurement with the arbitrage pricing theory: A new framework for analysis," *Journal of Financial Economics*, 15 (3), 373–394.

——— (1988) "Risk and return in an equilibrium APT: Application of a new test methodology," *Journal of Financial Economics*, 21 (2), 255–289.

De Lathauwer, Lieven, Bart De Moor, and Joos Vandewalle (2000) "A multilinear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, 21 (4), 1253–1278.

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

Elden, Lars and Berkant Savas (2009) "A Newton–Grassmann Method for Computing the Best Multilinear Rank (r1, r2,r3) Approximation of a Tensor," *SIAM Journal on Matrix Analysis and Applications*, 31 (2), 248–271.

Fama, Eugene F. and Kenneth R. French (1993) "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33 (1), 3–56.

Fan, Jianqing, Yuan Liao, and Weichen Wang (2016) "Projected principal component analysis in factor models," *The Annals of Statistics*, 44 (1), 219–254.

Giglio, Stefano and Dacheng Xiu (2021) "Asset pricing with omitted factors," *Journal of Political Economy*, 129 (7), 1947–1990.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu (2021) "Autoencoder asset pricing models," *Journal of Econometrics*, 222 (1), 429–450.

Huberman, Gur (1982) "A Simple Approach to Arbitrage Pricing Theory," *Journal of Economic Theory*, 28 (1), 183–191.

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.

Kim, Soohun, Robert A Korajczyk, and Andreas Neuhierl (2021) "Arbitrage Portfolios," *The Review of Financial Studies*, 34 (6), 2813–2856.

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

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh (2020) "Shrinking the Cross-Section," *Journal of Financial Economics*, 135 (2), 271–292.

Kroonenberg, Pieter M. and Jan de Leeuw (1980) "Principal component analysis of three-mode data by means of alternating least squares algorithms," *Psychometrika*, 45 (1), 69–97.

Lettau, Martin (2023) "High-Dimensional Factor Models and the Factor Zoo," *National Bureau of Economic Research*, No. 31719.

Lettau, Martin and Markus Pelger (2020a) "Factors That Fit the Time Series and Cross-Section of Stock Returns," *The Review of Financial Studies*, 33 (5), 2274–2325.

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

Ross, Stephen A. (1976) "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, 13 (3), 341–360.

Subrahmanyam, Avanidhar (2010) "The Cross-Section of Expected Stock Returns: What Have We Learnt from the Past Twenty-Five Years of Research?," *European Financial Management*, 16 (1), 27–42.

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

Table 1: Long-short portfolios

	Mean			Std. Dev.		
c	SMB_c	HML_c	$CROSS_c$	SMB_c	HML_c	$CROSS_c$
BM	2.23	4.93	9.10	15.27	12.59	17.61
OP	3.91	3.86	0.22	14.19	9.16	14.30
INV	2.58	4.17	4.97	14.58	8.05	13.99
MOM	2.91	10.43	8.30	13.31	17.22	16.21
REV	2.42	4.06	3.17	13.18	10.41	15.83
SREV	1.30	7.37	9.26	13.84	12.96	15.51
AC	3.10	2.61	-1.28	15.68	4.91	12.86
BETA	3.33	0.70	-0.00	13.42	16.86	16.86
NSI	3.28	5.22	4.72	15.82	9.28	13.64
VAR	2.58	4.48	13.70	14.69	19.55	17.52
RVAR	2.60	5.31	13.10	14.47	17.93	16.97
Mean	2.75	4.83	5.93	14.40	12.63	15.57

Note: The table reports annualized means and standard deviations of long-short portfolios. SMB_c is the difference of the small stock portfolios (P5) and the big stock portfolios (P1) of the double sort of characteristic c averaged of all characteristic quintiles. HML_c is constructed accordingly but is the difference of Q5 and Q1 portfolios averaged over all size quintiles. $CROSS_c$ is the is the average of the P1Q1 and P5Q5 portfolios minus the average of the P1Q5 and P5Q1 portfolios. The sample is from July 1967 to October 2023.

Table 2: Cross-sectional R^2_{xs} and Sharpe ratios

	3D-PCA			Fama-French			PCA			RP-PCA		
L	R^2_{xs}	SR	F_i	R^2_{xs}	SR	F_i	R^2_{xs}	SR	F_i	R^2_{xs}	SR	F_i
						A: In-sample						
1	-0.45	0.46	F^{3D}_{111}	-0.58	0.42	MKT	-0.45	0.45	F^{PC}_1	-0.45	0.46	F^{RP}_1
2	0.37	1.33	F^{3D}_{112}	0.01	0.78	CMA	0.18	0.85	F^{PC}_3	-0.04	0.93	F^{RP}_3
3	0.65	1.57	F^{3D}_{313}	0.06	1.00	MOM	0.39	1.04	F^{PC}_4	0.38	1.03	F^{RP}_2
4	0.67	1.59	F^{3D}_{312}	0.19	1.01	SMB	0.53	1.07	F^{PC}_2	0.58	1.24	F^{RP}_4
5	0.68	1.60	F^{3D}_{131}	0.57	1.21	RMW	0.62	1.31	F^{PC}_7	0.65	1.68	F^{RP}_7
6	0.76	1.95	F^{3D}_{122}	0.58	1.21	HML	0.70	1.37	F^{PC}_5	0.74	1.74	F^{RP}_5
						B: Out-of-sample						
1	-0.53	0.50	F^{3D}_{111}	-0.70	0.46	MKT	-0.55	0.48	F^{PC}_1	-0.54	0.49	F^{RP}_1
2	0.43	1.17	F^{3D}_{112}	0.02	0.85	CMA	0.27	0.92	F^{PC}_3	0.08	0.88	F^{RP}_3
3	0.65	1.29	F^{3D}_{312}	0.09	1.02	MOM	0.42	1.03	F^{PC}_4	0.28	1.03	F^{RP}_4
4	0.72	1.65	F^{3D}_{122}	0.17	1.02	SMB	0.49	1.04	F^{PC}_2	0.43	1.04	F^{RP}_2
5	0.76	1.67	F^{3D}_{131}	0.57	1.23	RMW	0.50	1.04	F^{PC}_5	0.46	1.13	F^{RP}_{10}
6	0.76	1.68	F^{3D}_{133}	0.57	1.23	HML	0.51	1.04	F^{PC}_6	0.48	1.18	F^{RP}_{19}

Note: The table shows cross-sectional R^2_{xs} and annualized Sharpe ratios of models with 3D-PCA factors (F^{3D}_{cpq}), Fama-French factors (MKT, SMB, HML, RMW, CMA, MOM), PCA factors (F^{PC}_j), and RP-PCA factors (F^{RP}_j). Pricing errors α_i are the intercepts in time series regressions of portfolio excess returns on factors. L is the number of factors and $R^2_{xs} = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}\alpha_i^2}{\text{Var}_{xs}(R_i)}$. 3D-PCA factors are based on a partial Tucker decomposition with $K_C = K_P = K_Q = 3$ factors. PCA and RP-PCA factors are based on models with $K_C K_P K_Q = 27$ factors. The columns denoted F_i show the L-th factor added to the model with $L-1$ factors. Panel A shows results for in-sample factors, and Panel B reports results for out-of-sample factors estimated in rolling samples of length $h = 120$ months. The sample is from July 1967 to October 2023.

Table 3: Cross-sectional R^2_{xs} – Horse races

	F	+3D-PCA	+FF	+PCA	+RP-PCA
			A: In-sample		
3D-PCA	0.68		0.65	0.65	0.61
FF	0.57	0.65		0.59	0.57
PCA	0.62	0.65	0.58		0.54
RP-PCA	0.65	0.61	0.56	0.54	
			B: Out-of-sample		
3D-PCA	0.76		0.71	0.73	0.66
FF	0.57	0.65		0.59	0.57
PCA	0.50	0.66	0.56		0.53
RP-PCA	0.46	0.60	0.54	0.54	

Note: This table reports cross-sectional R^2_{xs} of models with multiple factor types. The column **F** reports the R^2_{xs} for models with a single factor for $L = 5$, see Table 2. The remaining columns report the R^2_{xs} when other factor types are added. The added factors in the column +3D-PCA are those that are included in the 3D-PCA 5-factor model reported in Table 2: $F^{3D}_{112}, F^{3D}_{313}, F^{3D}_{312}, F^{3D}_{131}$. The added factors for the other types are chosen accordingly: +FF (CMA, MOM, SMB, RMW), +PCA ($F^{PC}_3, , F^{PC}_4, F^{PC}_2$), +RP-PCA ($F^{RP}_3, F^{RP}_4, F^{RP}_2$). I exclude the first factors (F^{3D}_{111}, MKT, F^{PC}_1, F^{RP}_1) because these four factors are essentially colinear. Panel A shows results for in-sample factors, and Panel B reports results for out-of-sample factors estimated in rolling samples of length $h = 120$ months. The sample is from July 1967 to October 2023.

Table 4: Cross-sectional R^2_{xs} for subsets of 3D-PCA factors

	$K_C=1$			$K_C=2$			$K_C=3$		
	$K_P=1$	$K_P=2$	$K_P=3$	$K_P=1$	$K_P=2$	$K_P=3$	$K_P=1$	$K_P=2$	$K_P=3$
A: In-sample $L=4$									
$K_Q=1$	-0.45	-0.41	-0.36	-0.52	-0.35	-0.23	-0.52	-0.07	0.15
$K_Q=2$	0.37	0.47	0.55	0.56	0.56	0.58	0.67	0.67	0.67
$K_Q=3$	0.46	0.60	0.56	0.56	0.56	0.58	0.67	0.67	0.67
B: In-sample $L=6$									
$K_Q=1$	-0.45	-0.41	-0.36	-0.52	-0.35	-0.22	-0.52	-0.09	0.26
$K_Q=2$	0.37	0.47	0.65	0.56	0.60	0.67	0.70	0.70	0.76
$K_Q=3$	0.46	0.62	0.70	0.58	0.65	0.67	0.67	0.71	0.76
C: Out-of–sample $L=4$									
$K_Q=1$	-0.53	-0.50	-0.40	-0.54	-0.44	-0.08	-0.56	-0.37	0.03
$K_Q=2$	0.43	0.57	0.64	0.51	0.58	0.64	0.65	0.72	0.72
$K_Q=3$	0.43	0.57	0.64	0.51	0.58	0.64	0.66	0.72	0.72
D: Out-of–sample $L=6$									
$K_Q=1$	-0.53	-0.50	-0.40	-0.54	-0.44	-0.04	-0.56	-0.39	0.11
$K_Q=2$	0.43	0.57	0.69	0.51	0.60	0.69	0.64	0.72	0.76
$K_Q=3$	0.43	0.60	0.69	0.48	0.62	0.69	0.66	0.73	0.76

Note: The table shows cross-sectional R^2_{xs} for models with subsets of 3D-PCA factors for different combinations of K_C,K_P, and K_Q. The numbers of factors in the time series regressions are $L=4$ and $L=6$. Panels A and B show results for in-sample factors, and Panels C and D report results for out-of-sample factors estimated in rolling samples of length $h=120$ months. The sample is from July 1967 to October 2023.

Table 5: Cross-sectional R^2_{xs} – Rolling windows of length h

	Number of factors L				
h	2	4	6	8	10
	A: 3D-PCA				
3	0.42	0.48	0.51	0.53	0.54
6	0.46	0.55	0.58	0.59	0.60
12	0.45	0.65	0.67	0.68	0.69
36	0.40	0.66	0.75	0.76	0.78
60	0.39	0.61	0.69	0.76	0.78
120	0.43	0.72	0.76	0.76	0.76
180	0.27	0.50	0.60	0.67	0.70
360	0.01	0.33	0.46	0.53	0.55
	B: PCA				
3	-0.33	-0.23	-0.19	-0.17	-0.17
6	-0.23	-0.11	-0.07	-0.05	-0.04
12	-0.13	-0.10	-0.09	-0.08	-0.08
36	-0.12	0.04	0.09	0.12	0.12
60	0.09	0.35	0.41	0.43	0.44
120	0.26	0.48	0.50	0.51	0.51
180	-0.13	0.34	0.46	0.48	0.49
360	-0.51	0.18	0.29	0.34	0.37

Note: The table shows cross-sectional R^2_{xs} of models with out-of-sample 3D-PCA factors and PCA factors for different window lengths h. L is the number of factors. 3D-PCA factors are based on a partial Tucker decomposition with $K_C = K_P = K_Q = 3$ factors. PCA and RP-PCA factors are based on models with $K_C K_P K_Q = 27$ factors. The sample is from July 1967 to October 2023.

Table 6: 3D-PCA factors – Sharpe ratios

3D-PCA		PCA		Fama-French	
F^{3D}_{cpq}	SR	F^{PC}_{i}	SR	F^{FF}_{i}	SR
		A: In-sample			
F^{3D}_{112}	1.23	F^{PC}_{16}	0.71	MOM	0.50
F^{3D}_{122}	0.90	F^{PC}_{3}	0.71	RMW	0.46
F^{3D}_{113}	−0.72	F^{PC}_{7}	0.70	CMA	0.45
F^{3D}_{123}	−0.58	F^{PC}_{4}	0.57	MKT	0.42
F^{3D}_{222}	0.46	F^{PC}_{21}	0.49	HML	0.32
F^{3D}_{111}	0.46	F^{PC}_{11}	0.47	SMB	0.20
F^{3D}_{223}	−0.43	F^{PC}_{1}	0.45		
F^{3D}_{213}	−0.41	F^{PC}_{19}	0.40		
		B: Out-of-sample			
F^{3D}_{112}	1.08	F^{PC}_{3}	0.80	CMA	0.50
F^{3D}_{122}	0.85	F^{PC}_{1}	0.48	RMW	0.48
F^{3D}_{113}	−0.79	F^{PC}_{4}	0.40	MKT	0.46
F^{3D}_{123}	−0.72	F^{PC}_{27}	0.34	MOM	0.43
F^{3D}_{223}	−0.55	F^{PC}_{10}	−0.26	HML	0.30
F^{3D}_{111}	0.50	F^{PC}_{7}	0.23	SMB	0.20
F^{3D}_{213}	−0.45	F^{PC}_{2}	0.21		
F^{3D}_{132}	0.43	F^{PC}_{20}	−0.20		

Note: The table shows the 3D-PCA and PCA factors with the highest annualized Sharpe ratios (in absolute value) and the Sharpe ratios of the Fama-French factors SMB, HML, RMW, CMA, and MOM. The 3D-PCA factors are derived from a partial Tucker decomposition with $K_C = K_P = K_Q = 3$. The sample is from July 1967 to October 2023.

Table 7: Correlation of in-sample and out-of-sample factors

	$F^{3D}_{\bullet 11}$	$F^{3D}_{\bullet 12}$	$F^{3D}_{\bullet 13}$	$F^{3D}_{\bullet 21}$	$F^{3D}_{\bullet 22}$	$F^{3D}_{\bullet 23}$	$F^{3D}_{\bullet 31}$	$F^{3D}_{\bullet 32}$	$F^{3D}_{\bullet 33}$
A: 3D-PCA									
$F^{3D}_{1\bullet\bullet}$	1.00	0.92	0.93	0.99	0.96	0.97	0.98	0.97	0.98
$F^{3D}_{2\bullet\bullet}$	0.91	0.98	0.87	0.95	0.94	0.92	0.92	0.94	0.92
$F^{3D}_{3\bullet\bullet}$	0.89	0.84	0.83	0.77	0.80	0.82	0.86	0.79	0.80
B: PCA									
F^{PC}_{1} to F^{PC}_{9}	1.00	0.95	0.83	0.59	0.49	0.49	0.31	0.18	0.05
F^{PC}_{10} to F^{PC}_{20}	0.16	0.09	0.16	0.04	-0.01	-0.01	0.09	0.14	0.14
F^{PC}_{19} to F^{PC}_{27}	0.11	0.01	0.03	-0.03	-0.02	0.00	0.06	0.07	0.08
C: RP-PCA									
F^{RP}_{1} to F^{RP}_{9}	1.00	0.90	0.75	0.53	0.66	0.39	0.39	0.08	0.23
F^{RP}_{10} to F^{RP}_{20}	0.18	0.15	0.20	0.15	0.06	-0.01	0.02	0.14	0.07
F^{RP}_{19} to F^{RP}_{27}	0.14	0.07	0.06	-0.02	0.07	0.13	0.13	0.01	0.01

Note: The tables reports correlations of in-sample and corresponding out-of-sample 3D-PCA factors (Panel A), PCA factors (Panel B), and RP-PCA factors (Panel C). F^{3D}_{cpq} factor corresponding to the c-th column of $\mathbf{V}^{(C)^\top}$, the p-th column of $\mathbf{V}^{(P)^\top}$, and the q-th column of $\mathbf{V}^{(Q)^\top}$. The $\bullet$ symbol can take values of 1, 2, or 3. The first three rows in Panels B and C report correlations of the first nine PCA and RP-PCA factors, the second and third rows show correlations for the 10th to 18th and 19th to 27th factors, respectively. The sample is from July 1967 to October 2023.

Table 8: 3D-PCA factors – Sums of F_t^2

	$F^{3D}_{\bullet 11}$	$F^{3D}_{\bullet 12}$	$F^{3D}_{\bullet 13}$	$F^{3D}_{\bullet 21}$	$F^{3D}_{\bullet 22}$	$F^{3D}_{\bullet 23}$	$F^{3D}_{\bullet 31}$	$F^{3D}_{\bullet 32}$	$F^{3D}_{\bullet 33}$
				A: In-sample					
$F^{3D}_{1\bullet\bullet}$	573.92	5.45	3.48	28.52	0.54	0.50	5.80	0.35	0.33
$F^{3D}_{2\bullet\bullet}$	0.50	15.15	1.05	0.35	1.07	0.26	0.11	0.53	0.20
$F^{3D}_{3\bullet\bullet}$	0.68	5.03	0.73	0.33	0.70	0.22	0.13	0.29	0.14
				B: Out-of-sample					
$F^{3D}_{1\bullet\bullet}$	488.71	5.05	3.05	24.67	0.51	0.45	5.47	0.31	0.28
$F^{3D}_{2\bullet\bullet}$	0.35	13.33	0.82	0.29	0.99	0.22	0.10	0.44	0.18
$F^{3D}_{3\bullet\bullet}$	1.17	3.90	0.86	0.33	0.60	0.22	0.14	0.29	0.15

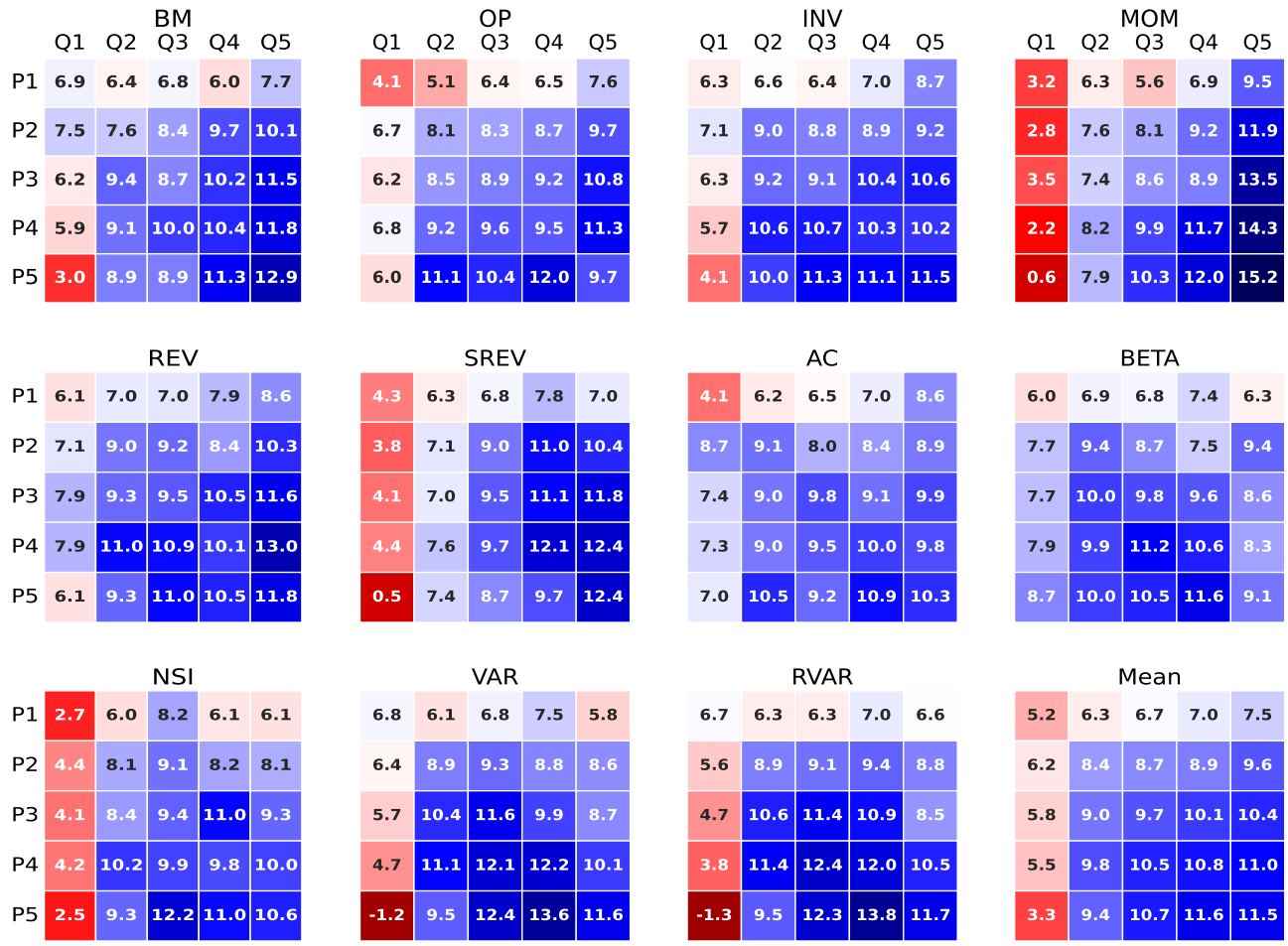
Note: The table reports means of squared factors of the 27 in-sample and out-of-sample 3D-PCA factors that are derived from a partial Tucker decomposition with $K_C = K_P = K_Q = 3$. F^{3D}_{cpq} factor corresponding to the c-th column of $\mathbf{V}^{(C)^\top}$, the p-th column of $\mathbf{V}^{(P)^\top}$, and the q-th column of $\mathbf{V}^{(Q)^\top}$. The $\bullet$ symbol can take values of 1, 2, or 3. The sample is from July 1967 to October 2023.

Table 9: $(2,4)$D-PCA – Sharpe ratios of factors

	$F_{11}^{\mathcal{N}}$	$F_{12}^{\mathcal{N}}$	$F_{13}^{\mathcal{N}}$	$F_{21}^{\mathcal{N}}$	$F_{22}^{\mathcal{N}}$	$F_{23}^{\mathcal{N}}$	$F_{31}^{\mathcal{N}}$	$F_{32}^{\mathcal{N}}$	$F_{33}^{\mathcal{N}}$
A: In-sample									
BM	0.47	0.48	-0.29	-0.09	0.43	-0.22	-0.13	0.08	0.03
OP	0.46	0.52	-0.36	0.04	-0.03	-0.34	-0.07	0.04	-0.31
INV	0.48	0.67	-0.44	-0.07	0.26	-0.46	-0.02	0.17	0.09
MOM	0.43	0.69	-0.18	-0.02	0.27	-0.20	-0.02	0.11	-0.06
REV	0.51	0.47	-0.22	-0.07	0.12	-0.13	-0.11	0.09	-0.09
SREV	0.42	0.67	-0.34	-0.13	0.30	0.06	-0.24	-0.07	-0.08
AC	0.45	0.74	-0.42	-0.03	-0.06	-0.18	-0.06	0.30	-0.04
BETA	0.49	0.11	-0.53	-0.01	0.01	-0.16	-0.02	-0.00	0.00
NSI	0.41	0.74	-0.94	0.02	0.24	-0.07	-0.06	0.15	-0.18
VAR	0.46	0.35	-0.78	-0.06	0.70	-0.61	-0.15	0.37	-0.05
RVAR	0.46	0.43	-0.86	-0.06	0.72	-0.68	-0.14	0.29	0.15
B: Out-of-sample									
BM	0.50	0.43	-0.33	-0.10	0.40	-0.20	-0.15	0.17	-0.11
OP	0.49	0.58	-0.45	0.05	0.06	-0.30	-0.13	-0.05	-0.29
INV	0.51	0.65	-0.45	-0.07	0.14	-0.49	-0.05	0.33	0.03
MOM	0.47	0.63	-0.28	-0.01	0.37	-0.14	-0.13	0.18	-0.05
REV	0.55	0.38	-0.19	-0.09	0.02	-0.18	-0.13	0.10	-0.06
SREV	0.46	0.58	-0.45	-0.12	0.26	0.06	-0.35	-0.24	-0.06
AC	0.48	0.56	-0.42	-0.02	-0.05	-0.26	-0.09	0.32	-0.07
BETA	0.53	0.12	-0.52	-0.05	0.03	-0.18	-0.05	0.05	0.12
NSI	0.45	0.65	-0.91	0.04	0.25	-0.25	-0.07	0.18	-0.17
VAR	0.50	0.31	-0.80	-0.07	0.70	-0.77	-0.27	0.44	-0.19
RVAR	0.50	0.39	-0.91	-0.07	0.72	-0.82	-0.25	0.35	0.02

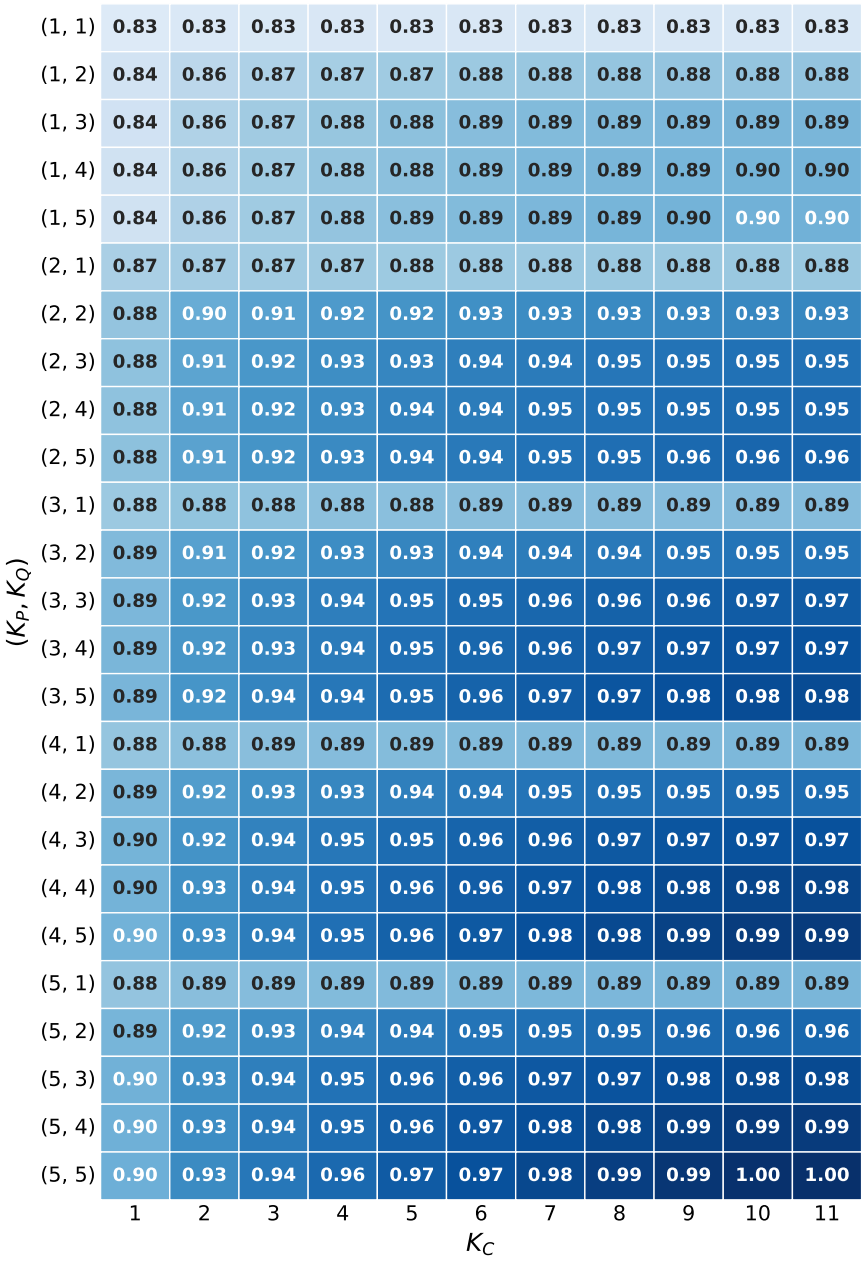
Note: The table reports annualized Sharpe ratios of in-sample and out-of-sample $(2,4)$D-PCA factors The sample is from July 1967 to October 2023.

Figure 1: Means of portfolio returns



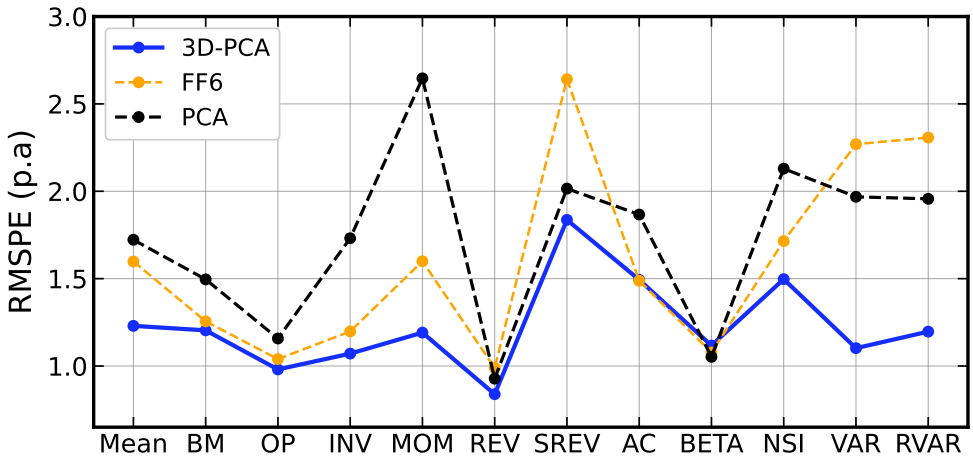
Notes: The figure shows heatmaps of annualized mean returns of the 275 portfolios. Each panel shows the 5×5-double sorted portfolios of a characteristic. The bottom-right panel shows the mean across all 11 characteristics. Portfolios with means that are lower (higher) than the mean of the CRSP-VW return (6.61%) are in red (blue). The sample is from July 1967 to October 2023.

Figure 2: Time series R^2 of 3D-PCA



Notes: This figure plots R^2 of 3D-PCA estimated in rolling windows of length $h = 120$. The R^2 is defined as $R^2 = 1 - \|\boldsymbol{\varepsilon}\| / \mathrm{Var}(\boldsymbol{\mathcal{X}})$, where $\boldsymbol{\varepsilon} = \boldsymbol{\mathcal{X}} - \widehat{\boldsymbol{\mathcal{X}}}$ and $\widehat{\boldsymbol{\mathcal{X}}}$ is the approximation of $\boldsymbol{\mathcal{X}}$ given by a 3D-PCA model with $K_C = K_P = K_Q = 3$. The figure plots the R^2 in each subsample. The R^2 for the in-sample estimation is in parentheses. The sample is from July 1967 to October 2023.

Figure 3: Pricing errors – Characteristics



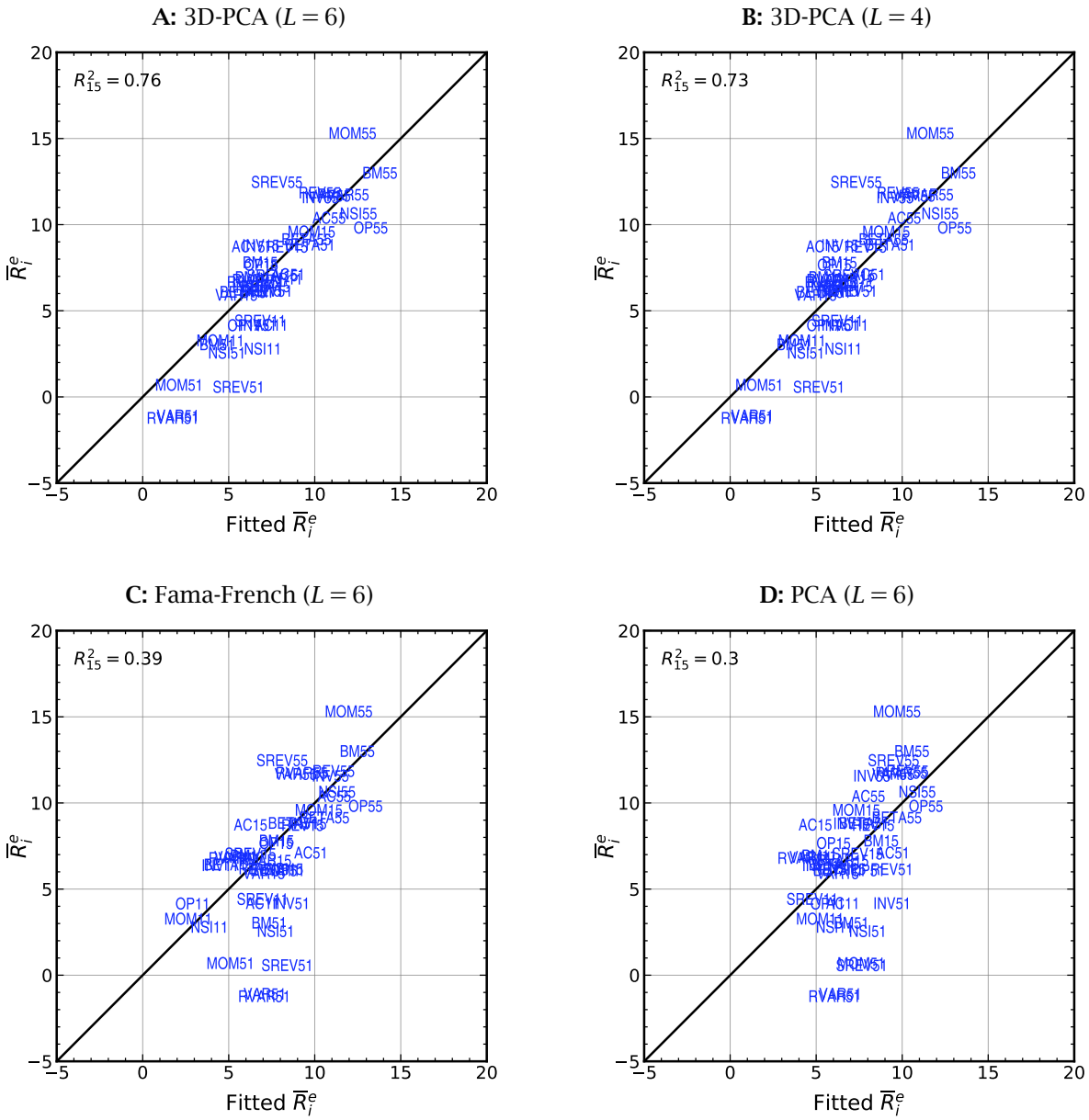
Notes: The figure shows aggregated pricing errors based on a model with 3D-PCA (in blue), Fama-French (in orange), and PCA (in black) factors. The number of factors in the cross-sectional models is set to $L = 6$. Pricing errors are aggregated by the 11 characteristics by computing the RMSPE of the 5×5-double sorted portfolios of each characteristic. Factors are derived from out-of-sample estimations of a 3D-PCA model with $K_C = K_P = K_Q = 3$ and a PCA model with 27 factors. The sample is from July 1967 to October 2023.

Figure 4: Pricing errors – Quintile portfolios



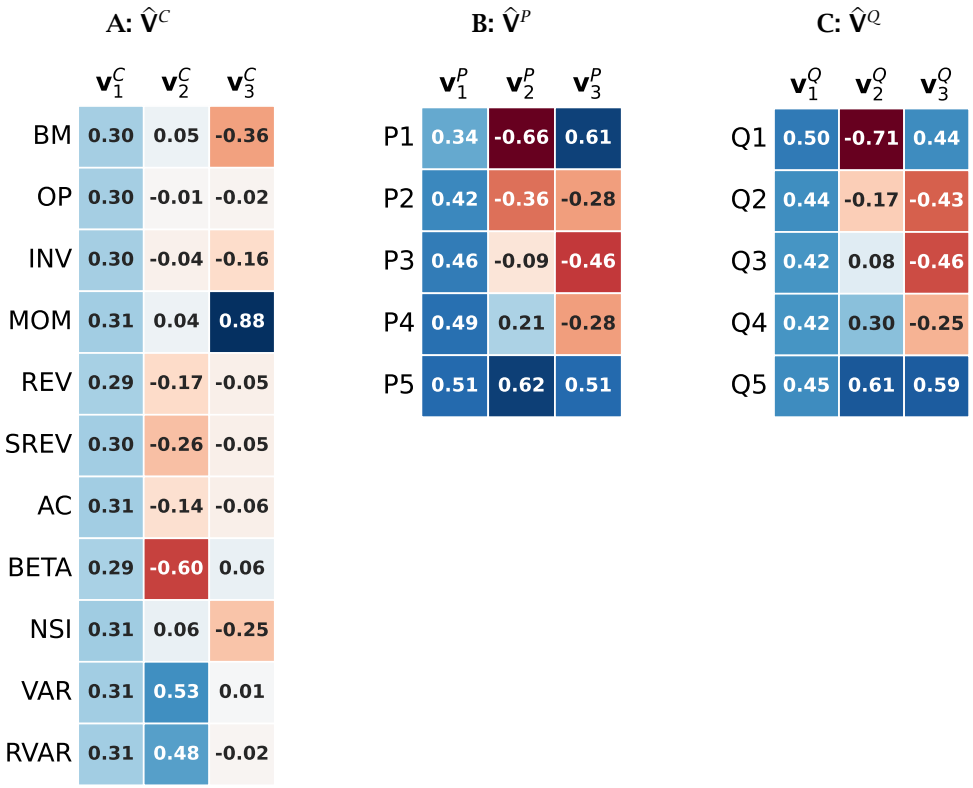
Notes: The figure shows heatmaps of pricing errors aggregated by the 25 size/characteristic quantile combinations. The number of factors in the cross-sectional models is set to $L = 6$. Factors are derived from out-of-sample estimations of a 3D-PCA model with $K_C = K_P = K_Q = 3$ and a PCA model with 27 factors. The sample is from July 1967 to October 2023.

Figure 5: Fitted vs. actual means of corner portfolios



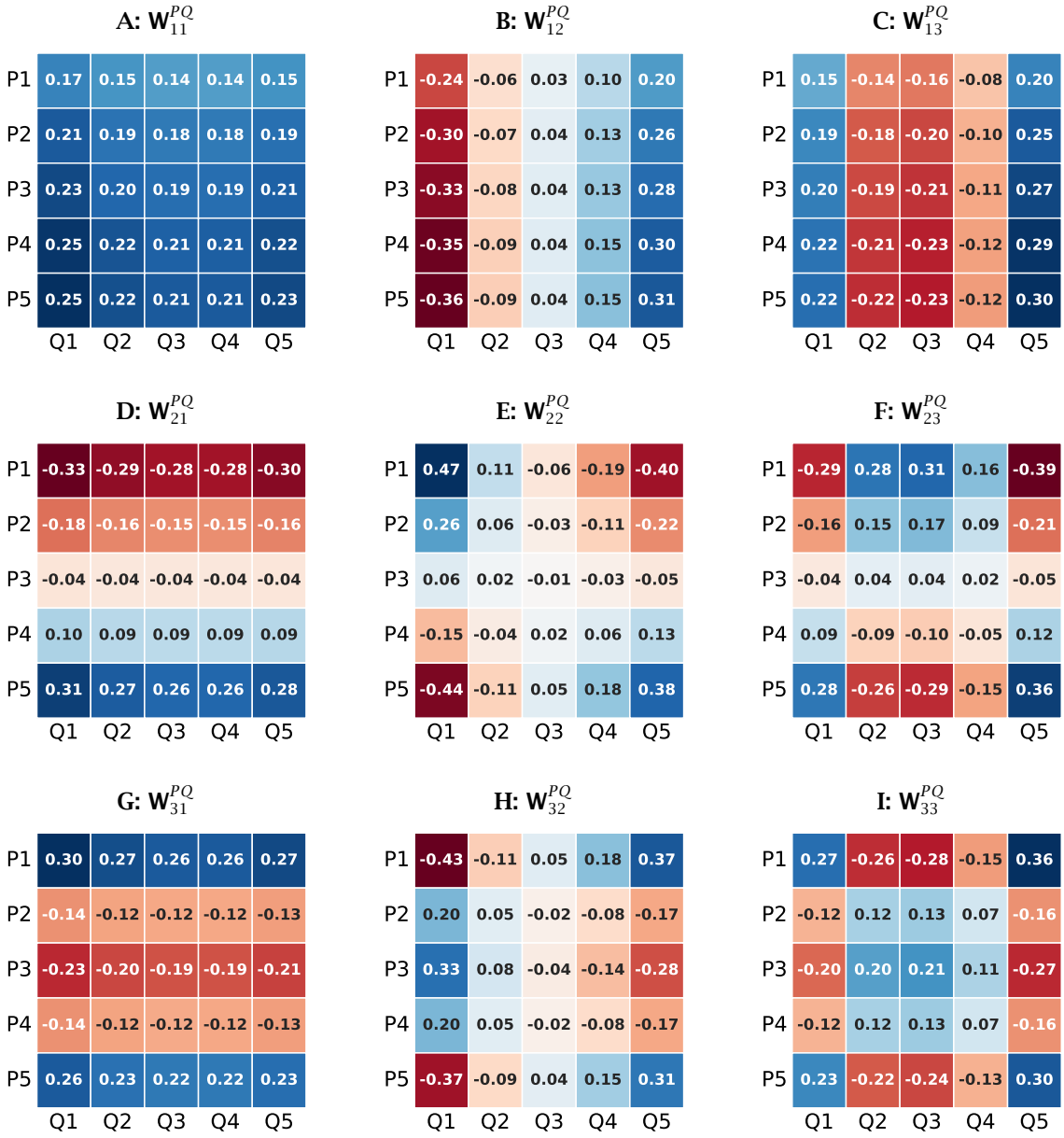
Notes: The figure shows fitted mean returns on the x-axis and means of observed returns on the y-axis. The plots include the corner portfolios P1Q1, P1Q5, P5Q1, and P5Q5 of all 11 characteristics. Factors are derived from out-of-sample estimations of a 3D-PCA model with $K_C = K_P = K_Q = 3$ and a PCA model with 27 factors. The number of factors in the cross sectional models is set to $L = 6$ in Panels A, C, and D to $L = 4$ in Panel B. The sample is from July 1967 to October 2023.

Figure 6: 3D-PCA – $\hat{\mathbf{V}}^i$ matrices



Notes: The heatmaps shows estimates of the $\hat{\mathbf{V}}^C, \hat{\mathbf{V}}^P$, and $\hat{\mathbf{V}}^Q$ matrices of the 3D-PCA model (14) with $K_C = K_P = K_Q = 3$. Negative values are plotted in red and positive ones in blue. The model is estimated by HOOI. The sample is from July 1967 to October 2023.

Figure 7: 3D-PCA weights $\mathbf{W}_{pq}^{PQ}$



Notes: The heatmaps shows the (5×5) matrices $\mathbf{W}_{pq}^{PQ}$, $p,q=1,2,3$ that are given by the outer product of the column vectors of $\hat{\mathbf{V}}^P$, and $\hat{\mathbf{V}}^Q$: $\mathbf{W}_{pq}^{PQ}=\mathbf{v}_p^P\circ\mathbf{v}_q^Q$. $\hat{\mathbf{V}}^P$, and $\hat{\mathbf{V}}^Q$ are from the estimation of a partial Tucker model (13) with $K_C=K_P=K_Q=3$. Negative values are plotted in red and positive ones in blue. The model is estimated by HOOI. The sample is from July 1967 to October 2023.

Figure 8: 3D-PCA weight matrices $\hat{\mathbf{W}}^{3D}$

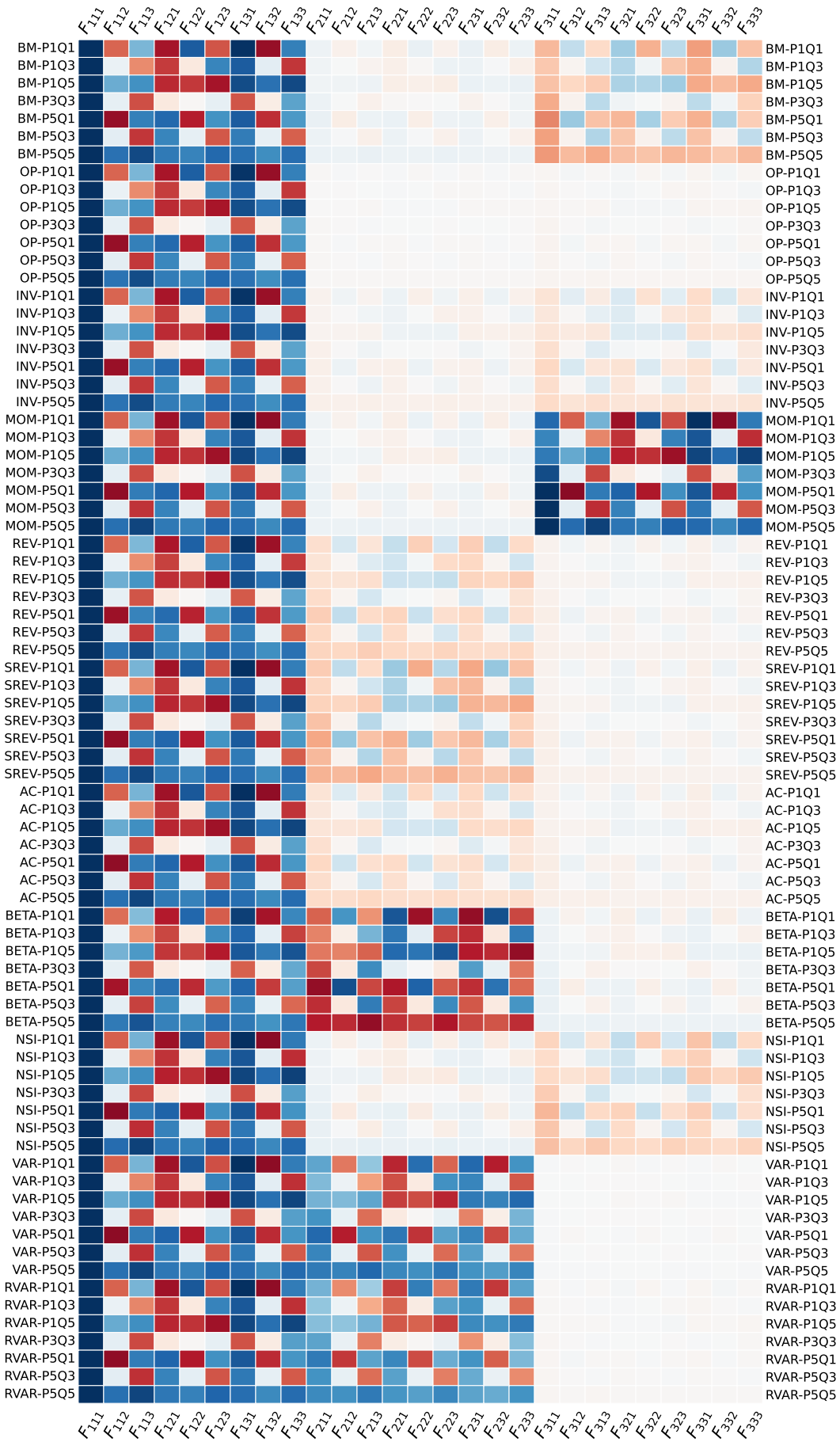
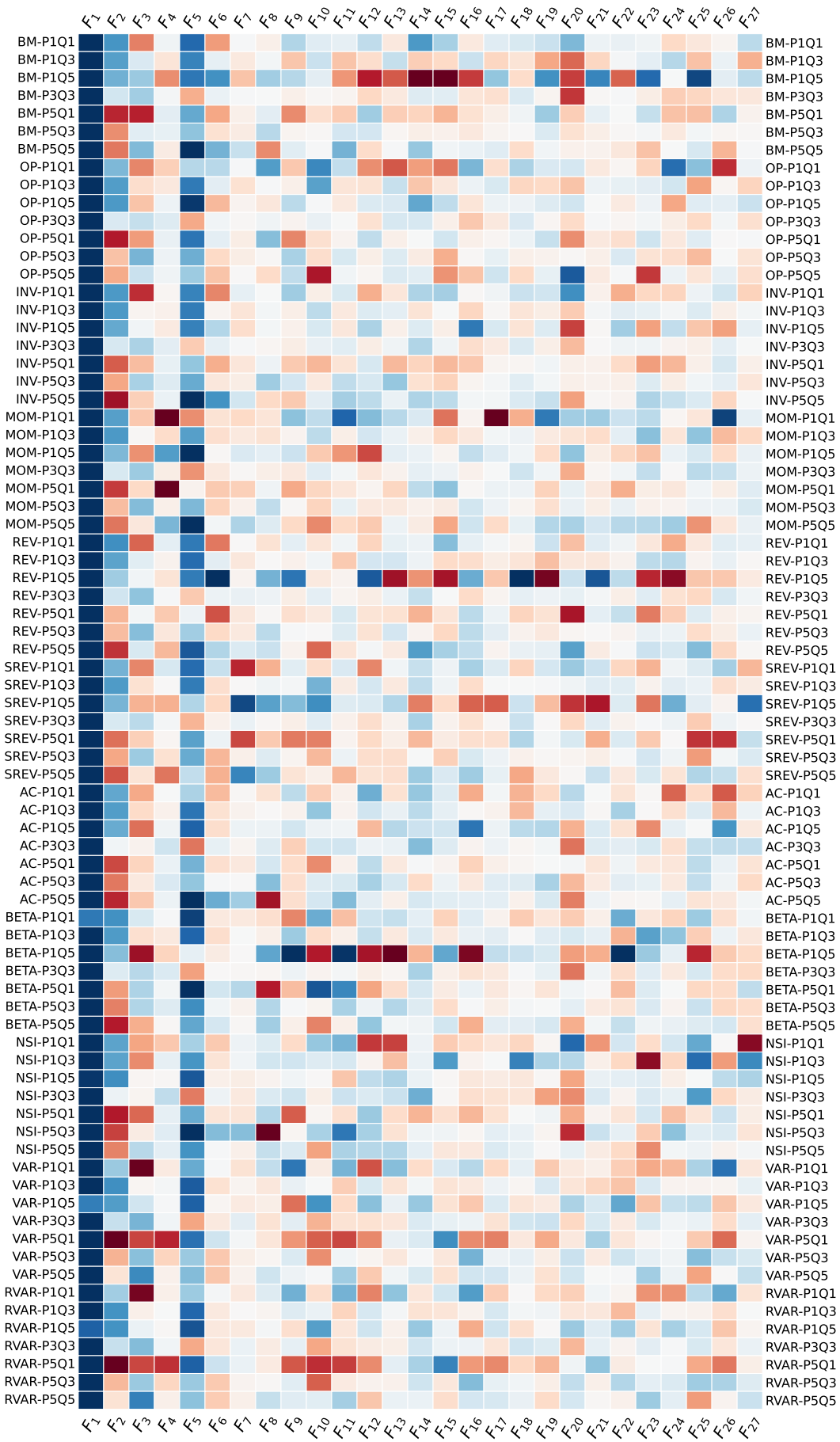
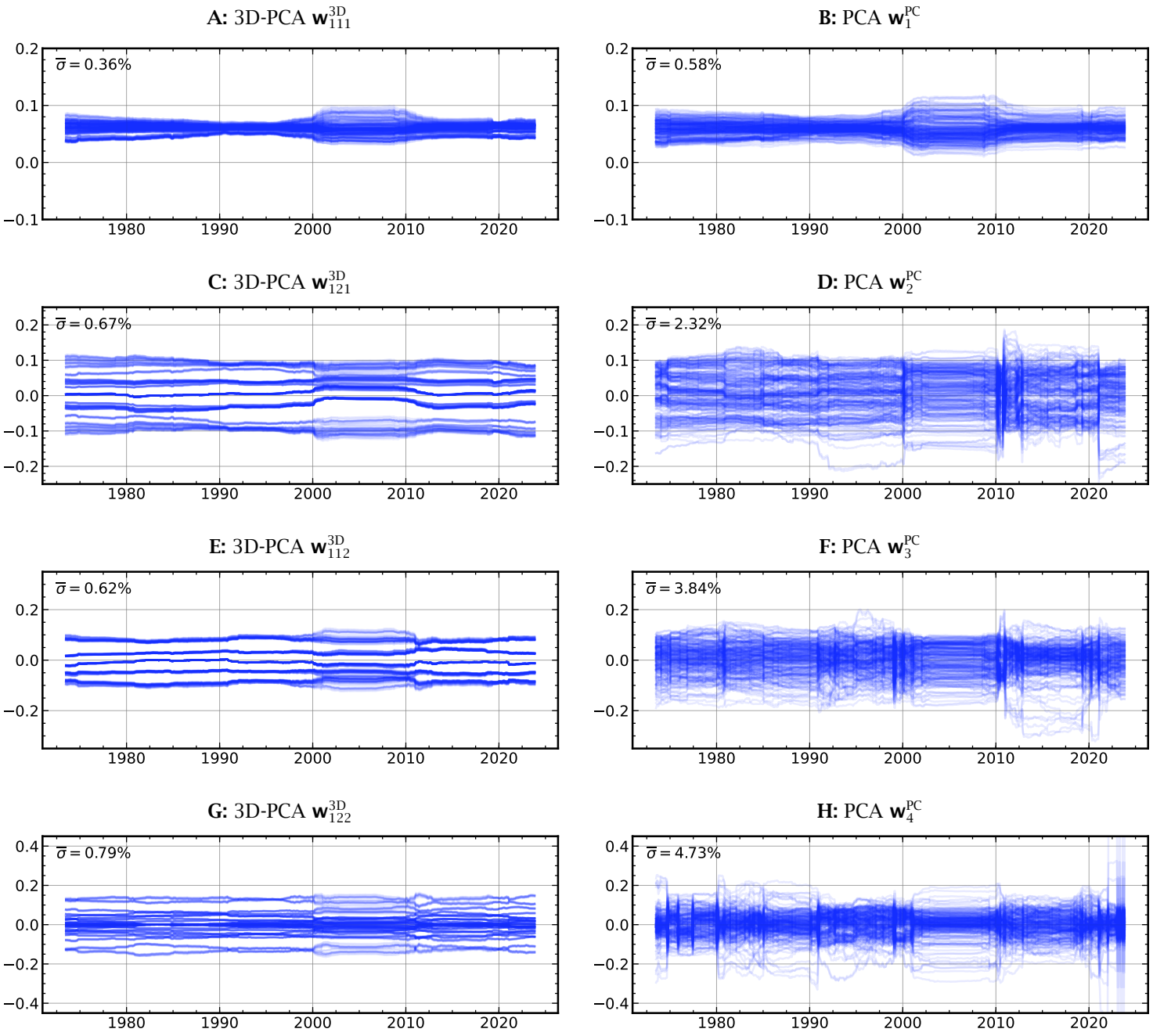


Figure 9: PCA weight matrices $\hat{\mathbf{W}}^{\text{PC}}$



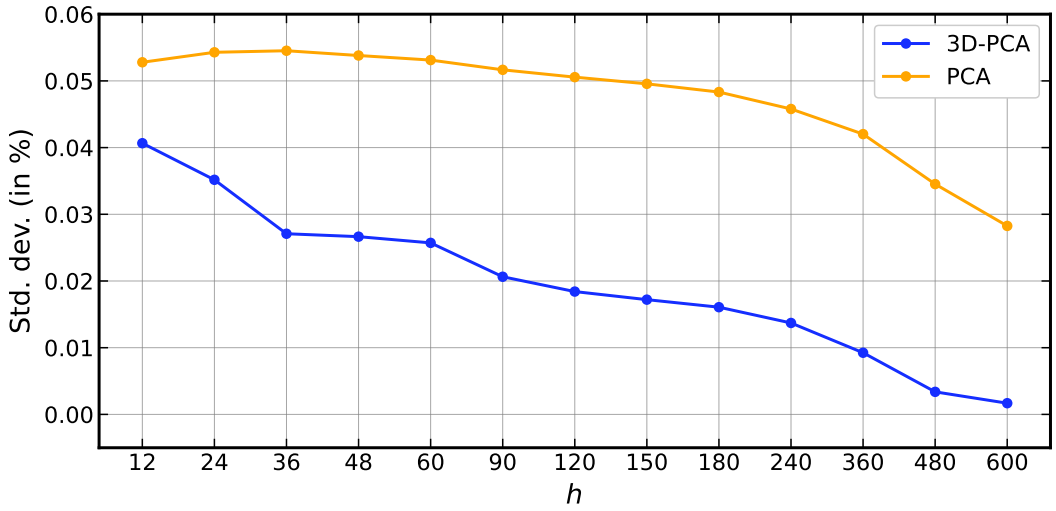
Notes for Figures 8 and 9: The figures show heatmaps of factor weights of P1Q1, P1Q3, P1Q5, P3Q1, P3Q3, P3Q5, P51Q1, P5Q3, and P5Q5 portfolios of the 11 characteristics. Figures 8 shows weights of the 27 3D-PCA factors based ona partial Tucker decomposition with $K_C = K_P = K_Q = 3$ and Figures 9 shows weights based on PCA with 27 factors. The 3D-PCA factor F_{cpq} is based on the c-th column of $\mathbf{V}^{(C)^\top}$, the p-th column of $\mathbf{V}^{(P)^\top}$, and the q-th column of $\mathbf{V}^{(Q)^\top}$. Negative values are plotted in red and positive ones in blue. The sample is from July 1967 to October 2023.

Figure 10: Weights – Rolling windows



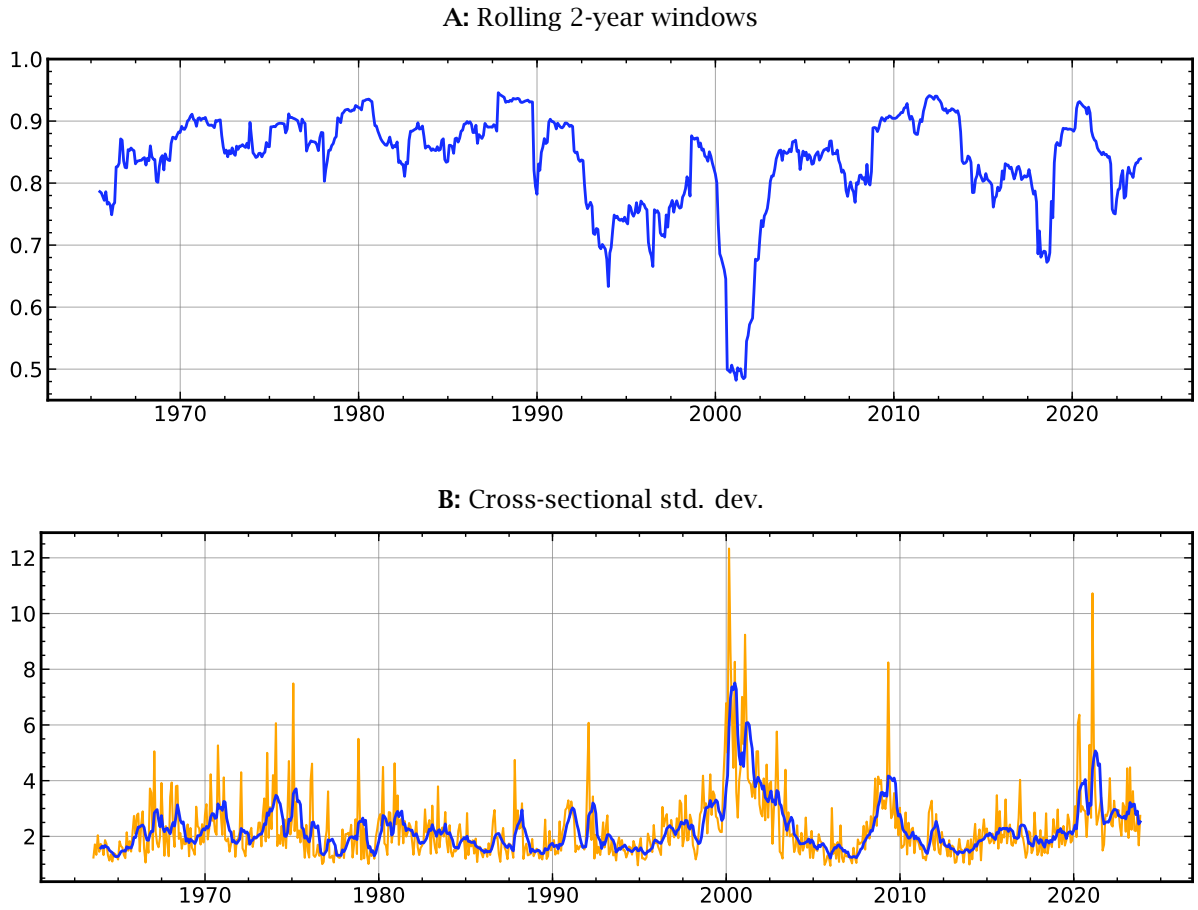
Notes: The figure plots the time series of factor weights estimated in rolling samples of length $h = 120$ of 3D-PCA factors $\mathbf{w}_{111}^{3D}, \mathbf{w}_{112}^{3D}, \mathbf{w}_{121}^{3D}$ and $\mathbf{w}_{122}^{3D}$ (left panels) and the first four PCA (right panels) factors. The 3D-PCA factor F_{cpq} is based on the c-th column of $\mathbf{V}^{(C)^\top}$, the p-th column of $\mathbf{V}^{(P)^\top}$, and the q-th column of $\mathbf{V}^{(Q)^\top}$. 3D-PCA factors are based on a partial Tucker decomposition with $K_C = K_P = K_Q = 3$ while PCA factor are based on an estimation with 27 factors. Each plot shows the time series of the factor weights of the 275 portfolios. $\bar{\sigma}$ is the average standard deviation of the 275 time series in each plot. The sample is from July 1967 to October 2023.

Figure 11: Std. dev. of weights in rolling windows of length h



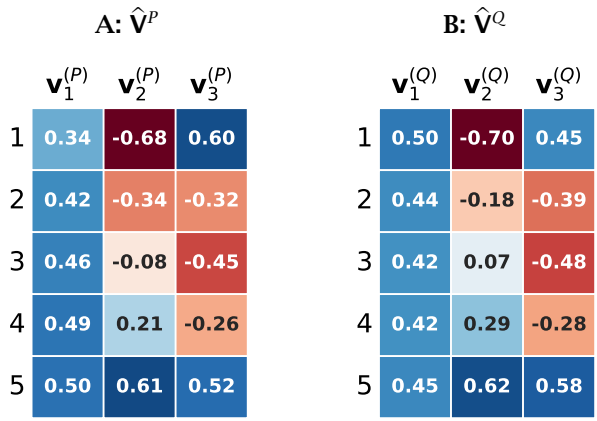
Notes: The figure plots the standard deviations of out-of-sample 3D-PCA (in blue) and PCA (in orange) factors for different window lenghts h ranging from 12 to 720 months. 3D-PCA factors are based on a partial Tucker decomposition with $K_C = K_P = K_Q = 3$ while PCA factor are based on an estimation with 27 factors. The sample is from July 1967 to October 2023.

Figure 12: Correlations and standard deviations over time



Notes: The top panel plots correlations of portfolios in rolling 24-month samples. In each subsample I compute the correlation matrix of all 275 portfolios and compute the mean correlation by characteristic. The plot shows the time series of mean correlation of the 11 characteristics. The bottom panel shows the 12-month moving average of the cross-sectional standard deviation across all 275 portfolios in each month. The sample is from July 1967 to October 2023.

Figure 13: $(2,4)$D-PCA – $\hat{\mathbf{V}}^i$ matrices



Notes: The heatmaps shows estimates of the $\hat{\mathbf{V}}^P$ and $\hat{\mathbf{V}}^Q$ matrices of the $(2,4)$D-PCA model (29) with $K_P = K_Q = 3$. Negative values are plotted in red and positive ones in blue. The model is estimated by HOOI. The sample is from July 1967 to October 2023.

Appendix A. Appendix

Appendix A.1. Tensor operations

Let $\boldsymbol{\mathcal{X}}$ be a $(T\times N\times C)$-dimensional tensor. A 3-dimensional tensor can expressed as collections of one-dimensional *fibers* and 2-dimensional *slices*. Fibers are vectors and correspond to rows and columns of a matrix, while slices are matrices. Fibers are defined by fixing every index but one so that $\boldsymbol{\mathcal{X}}$ has fibers along each mode, denoted by $\mathbf{x}_{(nc)t}, \mathbf{x}_{(tc)n}$, and $\mathbf{x}_{(tn)c}$, respectively.[16] Slices are created by fixing all but two indices and are written as $\mathbf{X}_{(t)nc}, \mathbf{X}_{(n)tc}, \mathbf{X}_{(c)tn}$.[17]

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

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

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

and the *norm* of $\boldsymbol{\mathcal{X}}$ is $\|\boldsymbol{\mathcal{X}}\| = \langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}} \rangle^{1/2}$. The *outer product* $\circ$ of two vectors $\mathbf{a} \in \mathbb{R}^T$ and $\mathbf{b} \in \mathbb{R}^N$ is defined as[18]

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

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

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

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

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

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

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

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

---

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

[17] See Panels E, F, and G of Figure C.2.

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

The 1-mode product of a $(2\times2\times3)$ tensor with a (5×2) matrix is illustrated in Panel A of Figure C.4. 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.[19]

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

Appendix A.2. The Singular Value Decomposition (SVD) of a matrix

Let $\mathbf{X}$ be a $(T\times N)$ data matrix with TN observations x_{tn}.[20] The SVD of $\mathbf{X}$ is given by

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

$$= \sum_{i=1}^{\min(M,N)} h_i\,\mathbf{u}_i^{(1)}\,\mathbf{u}_i^{(2)\top}, \tag{A.3}$$

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

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

$$\mathbf{X} = \mathbf{F}_N\,\mathbf{B}_N^\top, \tag{A.4}$$

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 (A.4) are not unique and can be rotated by any nonsingular $(N\times N)$ matrix $\mathbf{S}$: $\mathbf{X} = \mathbf{F}_N\mathbf{S}\mathbf{S}^{-1}\mathbf{B}_N^\top$. The standard normalization assumes that $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ are orthonormal and that $\mathbf{H}$ is diagonal.

It is also possible to compute the SVD of the $(N\times T)$ matrix $\mathbf{X}^\top$ instead of $\mathbf{X}$. The isomorphic factor representation for $\mathbf{X}^\top$ is given by $\mathbf{X}^\top = \mathbf{F}_T\,\mathbf{B}_T^\top$, where $\mathbf{F}_T = \mathbf{X}^\top\mathbf{U}^{(1)} = \mathbf{U}^{(2)}\mathbf{H}^\top$ and $\mathbf{B}_T = \mathbf{U}^{(1)}$. The representations are equivalent, but the roles of $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ are reversed so that factors of the SVD of $\mathbf{X}$ become factor loadings in the SVD of $\mathbf{X}^\top$, and *vice versa*.

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

$$\mathbf{X} = \widehat{\mathbf{X}}_K + \mathbf{E}_K, \tag{A.5}$$

$$\text{where } \widehat{\mathbf{X}}_K = \mathbf{U}_K^{(1)}\,\mathbf{H}_K\,\mathbf{U}_K^{(2)\top}, \tag{A.6}$$

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

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

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

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

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

$$\hat{\mathbf{X}}_K = \mathbf{U}_K^{(1)} \mathbf{H}_K \mathbf{U}_K^{(2)\top}. \tag{A.7}$$

The truncated SVD (A.7) is equivalent to the K-factor model

$$\mathbf{X} = \mathbf{F}_K \mathbf{B}_K^\top + \mathbf{E}_K, \tag{A.8}$$

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 equals the first K principal components of $\mathbf{X}^\top\mathbf{X}$. I will refer to this model as SVD-PCA throughout the paper.

The truncated SVD has an alternative representation useful for understanding tensor decompositions. $\mathbf{U}_K^{(1)} \mathbf{H}_K \mathbf{U}_K^{(2)\top}$ is equivalent to the weighted sum of the outer products of the column vectors of $\mathbf{U}_K^{(1)}$ and the row vectors of $\mathbf{U}_K^{(2)\top}$. This can be seen by writing (A.7) as

$$\hat{\mathbf{X}}_K = \sum_{t=1}^{K} \sum_{n=1}^{K} h_{tn} \underbrace{\mathbf{u}_t^{(1)} \mathbf{u}_n^{(2)\top}}_{T \times N} \tag{A.9}$$

$$= \sum_{k=1}^{K} h_{kk} \mathbf{u}_k^{(1)} \mathbf{u}_k^{(2)\top}. \tag{A.10}$$

The second equality follows from the fact that $\mathbf{H}_k$ is a diagonal matrix. $\hat{\mathbf{X}}_K$ is the weighted sum of K matrices with dimensions $(T \times N)$, which are the outer vector product of the eigenvectors $\mathbf{u}_k^{(1)}$ and $\mathbf{u}_k^{(2)\top}$ of $\mathbf{X}\mathbf{X}^\top$ and $\mathbf{X}^\top\mathbf{X}$, respectively. Each k in the summation corresponds to a factor in the K-factor representation (A.8). The advantage of representation (A.9) is that it shows the contribution of each of the K factors in the model's fit. Since the eigenvectors are normalized, the K outer vector products $\mathbf{u}_k^{(1)} \mathbf{u}_k^{(2)\top}$ are of the same magnitude, so the weight of the contribution of each factor k is approximately equal to the k-th eigenvalue.

In the truncated SVD (A.5)-(A.7) the number of factors is K. Note that we could define an asymmetric SVD that has different numbers of factors for the two dimensions:

$$\hat{\mathbf{X}}_{(K_1,K_2)} = \mathbf{U}_{K_1}^{(1)} \mathbf{H}_{K_1,K_2} \mathbf{U}_{K_1}^{(2)\top}, \tag{A.11}$$

where $\mathbf{H}_{K_1,K_2}, \mathbf{U}_{K_1}^{(1)}, \mathbf{U}_{K_2}^{(2)}$ are $(K_1 \times K_2), (N \times K_1), (N \times K_2)$ matrices. However, since $\mathbf{H}_{K_1,K_2}$ is diagonal, the asymmetric SVD reduces to a K-factor SVD where $K = \min(K_1, K_2)$. In contrast to the 2-dimensional matrix SVD, the core tensor $\boldsymbol{\mathcal{G}}$ in the Tucker decomposition is *not* diagonal. Consequently, the number of factors can differ by dimension.[21]

Appendix A.3. Higher-Order Orthogonal Iteration (HOOI)

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

$$\|\boldsymbol{\mathcal{E}}\| = \|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{F}} \times_2 \mathbf{V}^{(C)} \times_3 \mathbf{V}^{(P)} \times_4 \mathbf{V}^{(Q)}\|$$

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

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

$$\boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{X}} \times_2 \mathbf{V}^{(C)\top} \times_3 \mathbf{V}^{(P)\top} \times_4 \mathbf{V}^{(Q)\top}. \tag{A.12}$$

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

$$\begin{aligned}
\|\boldsymbol{\mathcal{E}}\|^2 &= \|\boldsymbol{\mathcal{X}}\|^2 - 2\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{F}} \times_2 \mathbf{V}^{(C)} \times_3 \mathbf{V}^{(P)} \times_4 \mathbf{V}^{(Q)} \rangle + \|\boldsymbol{\mathcal{F}} \times_2 \mathbf{V}^{(C)} \times_3 \mathbf{V}^{(P)} \times_4 \mathbf{V}^{(Q)}\|^2 && \text{(A.13)} \\
&= \|\boldsymbol{\mathcal{X}}\|^2 - 2\langle \boldsymbol{\mathcal{X}} \times_2 \mathbf{V}^{(C)\top} \times_3 \mathbf{V}^{(P)\top} \times_4 \mathbf{V}^{(Q)\top}, \boldsymbol{\mathcal{F}} \rangle + \|\boldsymbol{\mathcal{F}}\|^2 && \text{(A.14)} \\
&= \|\boldsymbol{\mathcal{X}}\|^2 - 2\langle \boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{F}} \rangle + \|\boldsymbol{\mathcal{F}}\|^2 && \text{(A.15)} \\
&= \|\boldsymbol{\mathcal{X}}\|^2 - \|\boldsymbol{\mathcal{F}}\|^2 && \text{(A.16)} \\
&= \|\boldsymbol{\mathcal{X}}\|^2 - \|\boldsymbol{\mathcal{X}} \times_2 \mathbf{V}^{(P)\top} \times_3 \mathbf{V}^{(P)\top} \times_4 \mathbf{V}^{(Q)\top}\|^2. && \text{(A.17)}
\end{aligned}$$

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

$$\max_{\mathbf{V}^{(Q)}} \|\boldsymbol{\mathcal{X}} \times_2 \mathbf{V}^{(C)\top} \times_3 \mathbf{V}^{(P)\top} \times_4 \mathbf{V}^{(Q)\top}\|. \tag{A.18}$$

This maximization problem can be rewritten in matrix form as

$$\max_{\mathbf{V}^{(Q)}} \|\mathbf{V}^{(Q)\top} \mathbf{W}_Q\| \tag{A.19}$$

$$\text{where } \mathbf{W}_Q = \mathbf{X}_{(Q)}(\mathbf{V}^{(C)} \otimes \mathbf{V}^{(P)}). \tag{A.20}$$

The optimal $\mathbf{V}^{(Q)}$ is given by the first K_Q left singular vectors of $\mathbf{W}_Q$. Since one $\mathbf{V}^{(i)}$ can be computed if the other two are known, we can use the following recursive algorithm known as Higher-Order Orthogonal Iteration (HOOI):

1. Pick initial values for $\mathbf{V}^{(C)}, \mathbf{V}^{(P)}$.
2. Compute $\mathbf{V}^{(Q)}$ as the first K_Q left singular vectors of $\mathbf{X}_{(Q)}(\mathbf{V}^{(C)} \otimes \mathbf{V}^{(P)})$.
3. Compute $\mathbf{V}^{(C)}$ as the first K_C left singular vectors of $\mathbf{X}_{(C)}(\mathbf{V}^{(P)} \otimes \mathbf{V}^{(Q)})$.
4. Compute $\mathbf{V}^{(P)}$ as the first K_P left singular vectors of $\mathbf{X}_{(P)}(\mathbf{V}^{(C)} \otimes \mathbf{V}^{(Q)})$.
5. Repeat Steps 2. to 4. recursively until a convergence criterion is satisfied.
6. Compute $\boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{X}} \times_2 \mathbf{V}^{(C)\top} \times_3 \mathbf{V}^{(P)\top} \times_4 \mathbf{V}^{(Q)\top}$.

The literature has developed numerous numerical improvements of the HOOI estimator; see, for example, Andersson and Bro (1998). Several other algorithms exist, including nonlinear Newton-Grassmann optimization (Elden and Savas (2009)). Starting values of the $\mathbf{V}^{(i)}$ matrices can be set by HOSVD.

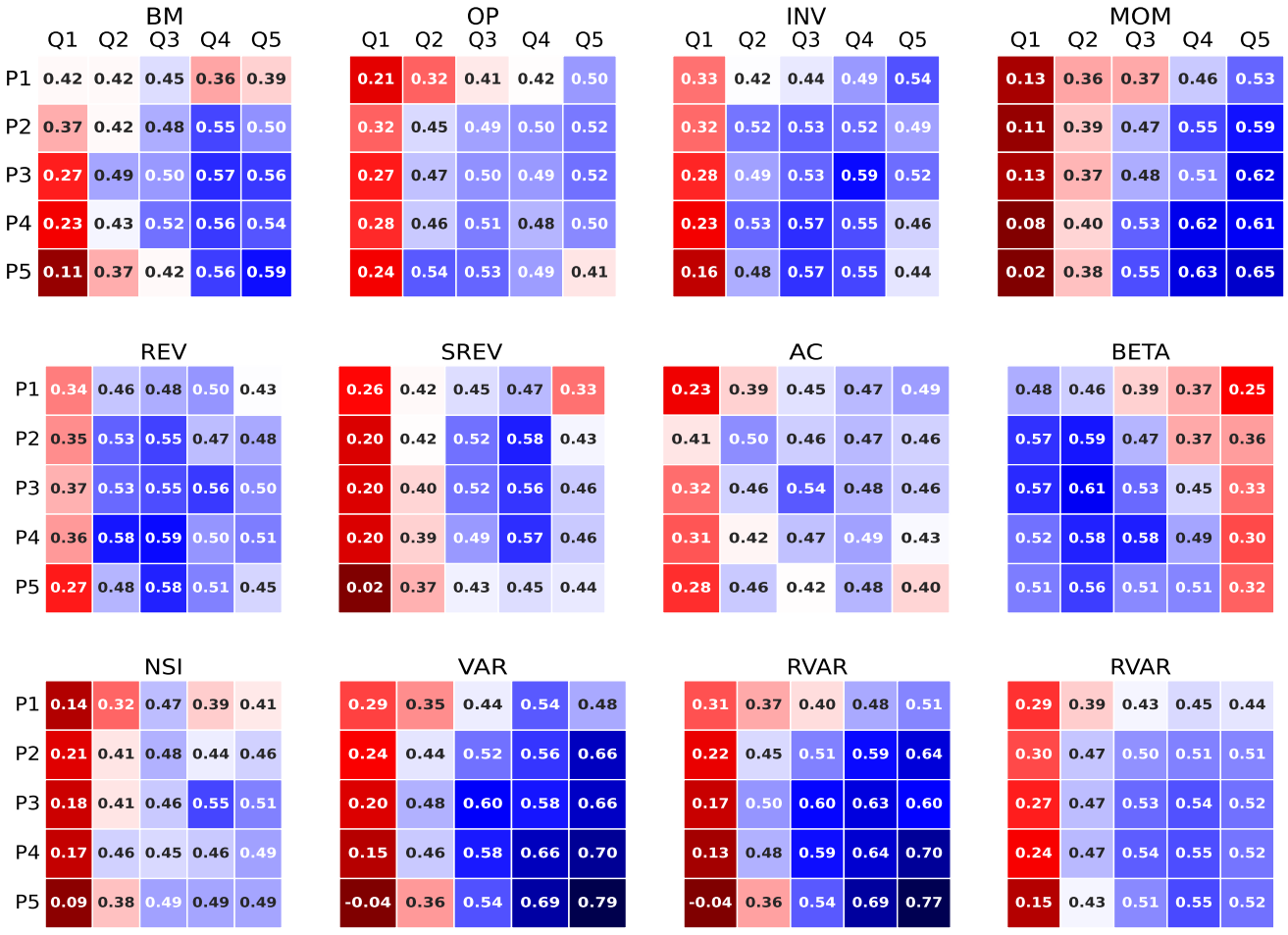
The HOOI estimator converges after 20 to 40 iterations for the data set used in this paper. In addition to setting the initial $\mathbf{V}^{(i)}$ using the method described above, I also choose initial values randomly. The numerical computations are robust and converge to the same optimum.

Table B.1: Cross-sectional R^2_{xs} for subsets of 3D-PCA factors – All combinations

	$K_C=1$			$K_C=2$			$K_C=3$		
	$K_P=1$	$K_P=2$	$K_P=3$	$K_P=1$	$K_P=2$	$K_P=3$	$K_P=1$	$K_P=2$	$K_P=3$
				A: In-sample $L=4$					
$K_Q=1$	-0.45	-0.41	-0.36	-0.52	-0.35	-0.22	-0.52	-0.07	0.15
$K_Q=2$	0.37	0.47	0.55	0.56	0.56	0.58	0.67	0.67	0.67
$K_Q=3$	0.46	0.60	0.60	0.56	0.60	0.60	0.67	0.67	0.67
				B: In-sample $L=6$					
$K_Q=1$	-0.45	-0.41	-0.36	-0.52	-0.35	-0.22	-0.52	-0.09	0.26
$K_Q=2$	0.37	0.47	0.65	0.56	0.60	0.69	0.70	0.70	0.76
$K_Q=3$	0.46	0.62	0.70	0.58	0.65	0.70	0.70	0.73	0.76
				C: Out-of–sample $L=4$					
$K_Q=1$	-0.53	-0.50	-0.40	-0.54	-0.44	-0.08	-0.56	-0.37	0.03
$K_Q=2$	0.43	0.57	0.64	0.51	0.58	0.64	0.65	0.72	0.72
$K_Q=3$	0.43	0.57	0.64	0.51	0.58	0.64	0.66	0.72	0.72
				D: Out-of–sample $L=6$					
$K_Q=1$	-0.53	-0.50	-0.40	-0.54	-0.44	-0.04	-0.56	-0.39	0.11
$K_Q=2$	0.43	0.57	0.69	0.51	0.60	0.69	0.64	0.72	0.76
$K_Q=3$	0.43	0.60	0.69	0.48	0.64	0.69	0.66	0.73	0.76

Note: The table shows cross-sectional R^2_{xs} for models with subsets of 3D-PCA factors derived from a partial Tucker decomposition with $K_C = K_P = K_Q = 3$ factors for $L = 5$. Ci, Pj, and Qk denote the factors F^{3D}_{ijk} that are included in the time-series estimations. For example, the combination (C1, P12, Q123) includes factors F^{3D}_{ijk} where $i = 1, j = 1,2, k = 1,2,3$. Panel A shows results for in-sample factors, and Panel B reports results for out-of-sample factors estimated in rolling samples of length $h = 120$ months. The sample is from July 1967 to October 2023.

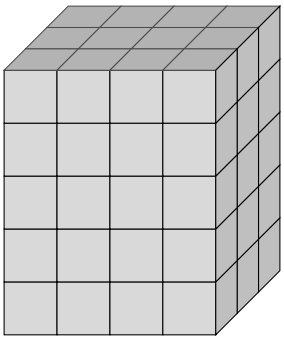
Figure C.1: Sharpe ratios of portfolio returns



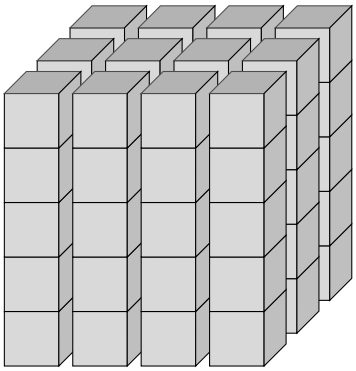
Notes: The figure shows heatmaps of annualized Sharpe ratios of the 275 portfolios. Each panel displays the 5×5-double sorted portfolios of a characteristic. The bottom-right panel shows the mean Sharpe ratio across all 11 characteristics. Portfolios with Sharpe ratios that are lower (higher) than the Sharpe ratio of the CRSP-VW return (0.42) are in red (blue). The sample is from July 1967 to October 2023.

Figure C.2: Tensor fibers and slices

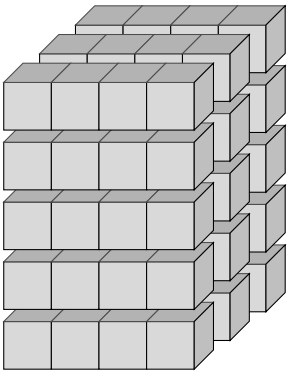
A: Tensor $\boldsymbol{\mathcal{X}}$: $(5 \times 4 \times 3)$



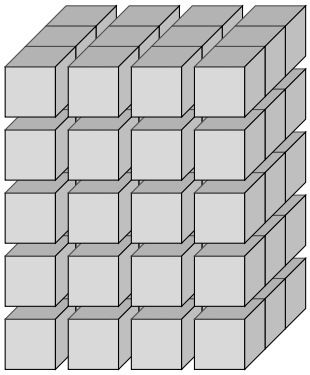
B: Mode-1 fibers $\mathbf{x}_{(nc)t}$



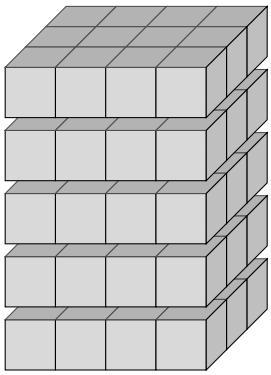
C: Mode-2 fibers $\mathbf{x}_{(tc)n}$



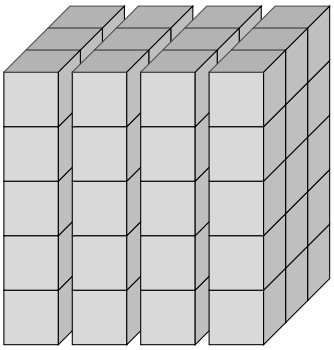
D: Mode-3 fibers $\mathbf{x}_{(tn)c}$



E: Horizontal slices $\mathbf{X}_{(i)jk}$



F: Lateral slices $\mathbf{X}_{(j)ik}$



G: Frontal slices $\mathbf{X}_{(k)ij}$

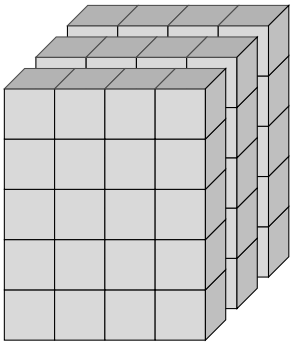
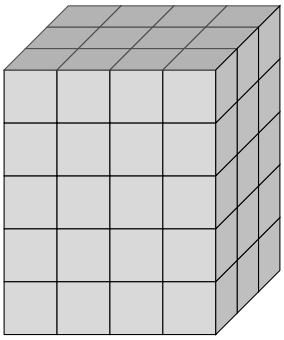
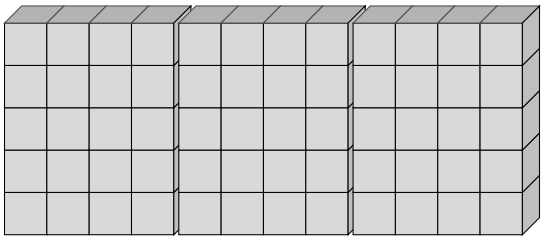


Figure C.3: Tensor as matrices

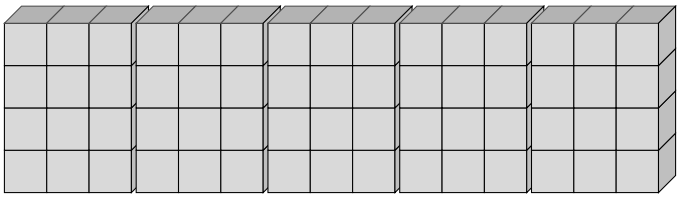
A: Tensor $\boldsymbol{\mathcal{X}}$: $(5 \times 4 \times 3)$



B: $\mathbf{X}_{(1)}$: (5×12)



C: $\mathbf{X}_{(2)}$: (4×15)



D: $\mathbf{X}_{(3)}$: (3×20)

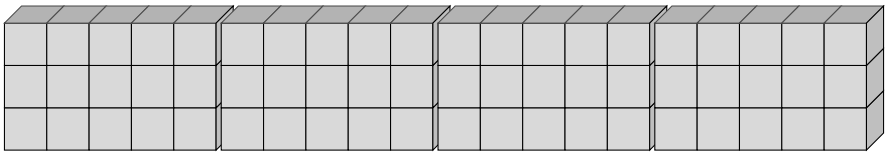


Figure C.4: n-mode Tensor multiplication

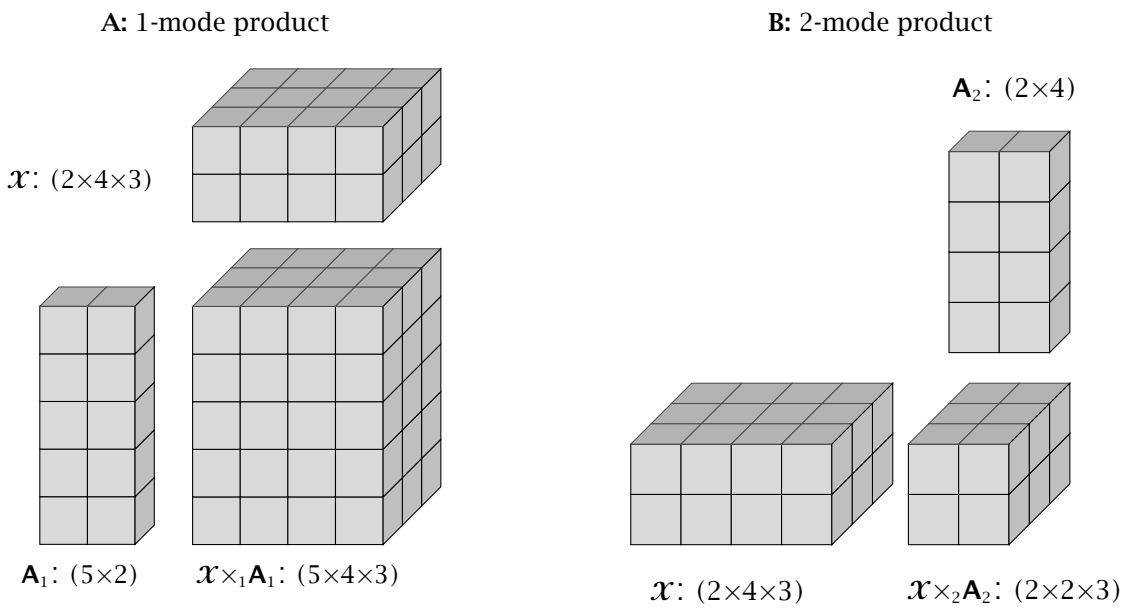


Figure C.5: Tucker Decomposition $\mathcal{X} = \mathcal{G} \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \times_3 \mathbf{V}_3$

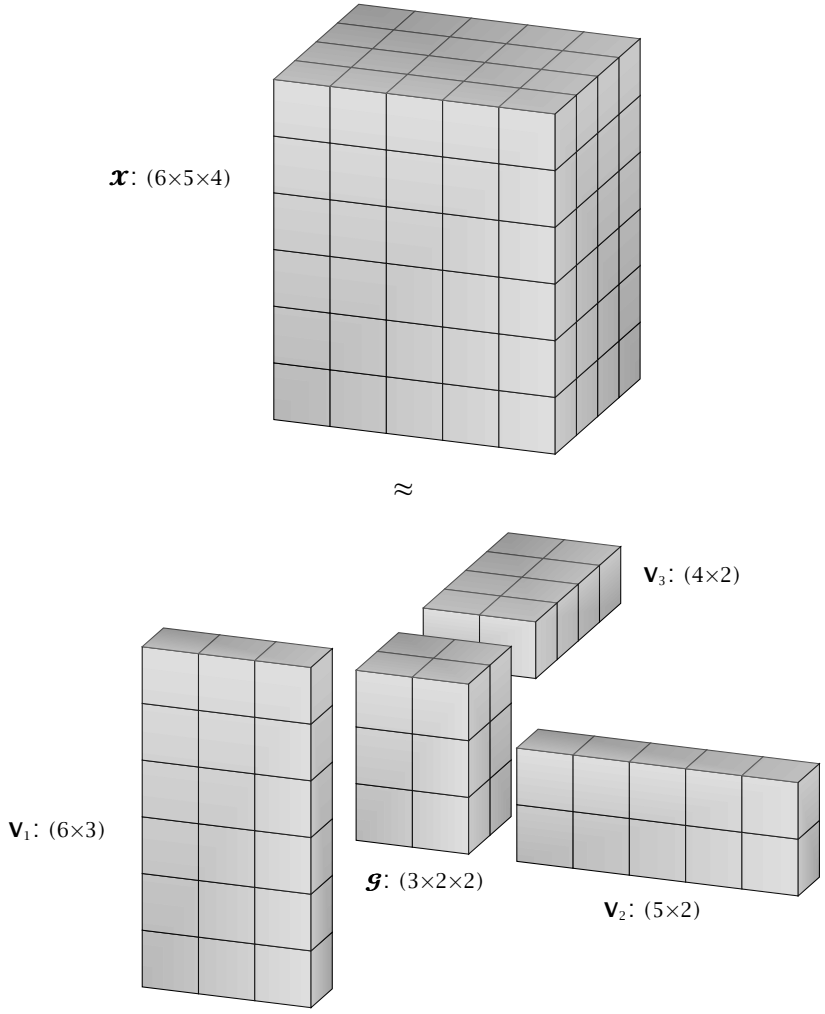


Figure C.6: Proportionality of C-slices of weight tensor $\boldsymbol{\mathcal{W}}_{cpq} = \mathbf{v}_c^{(C)} \circ \mathbf{v}_p^{(P)} \circ \mathbf{v}_q^{(Q)}$

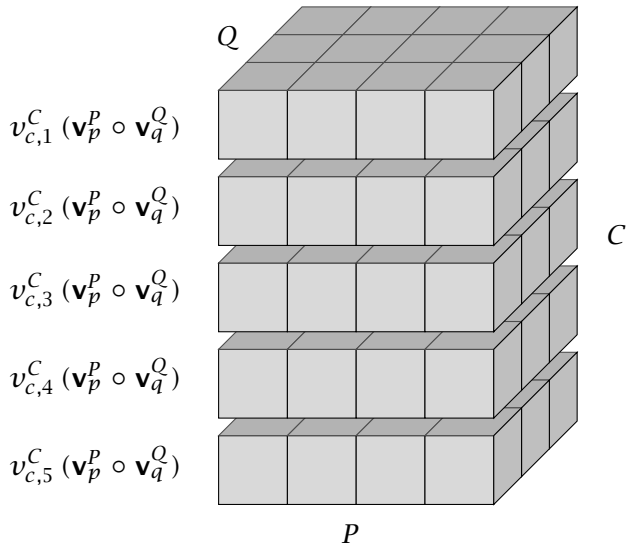
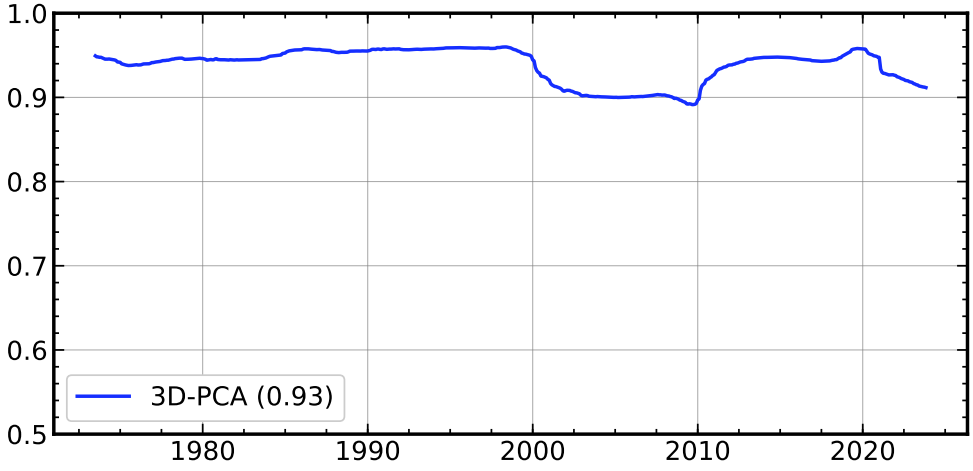
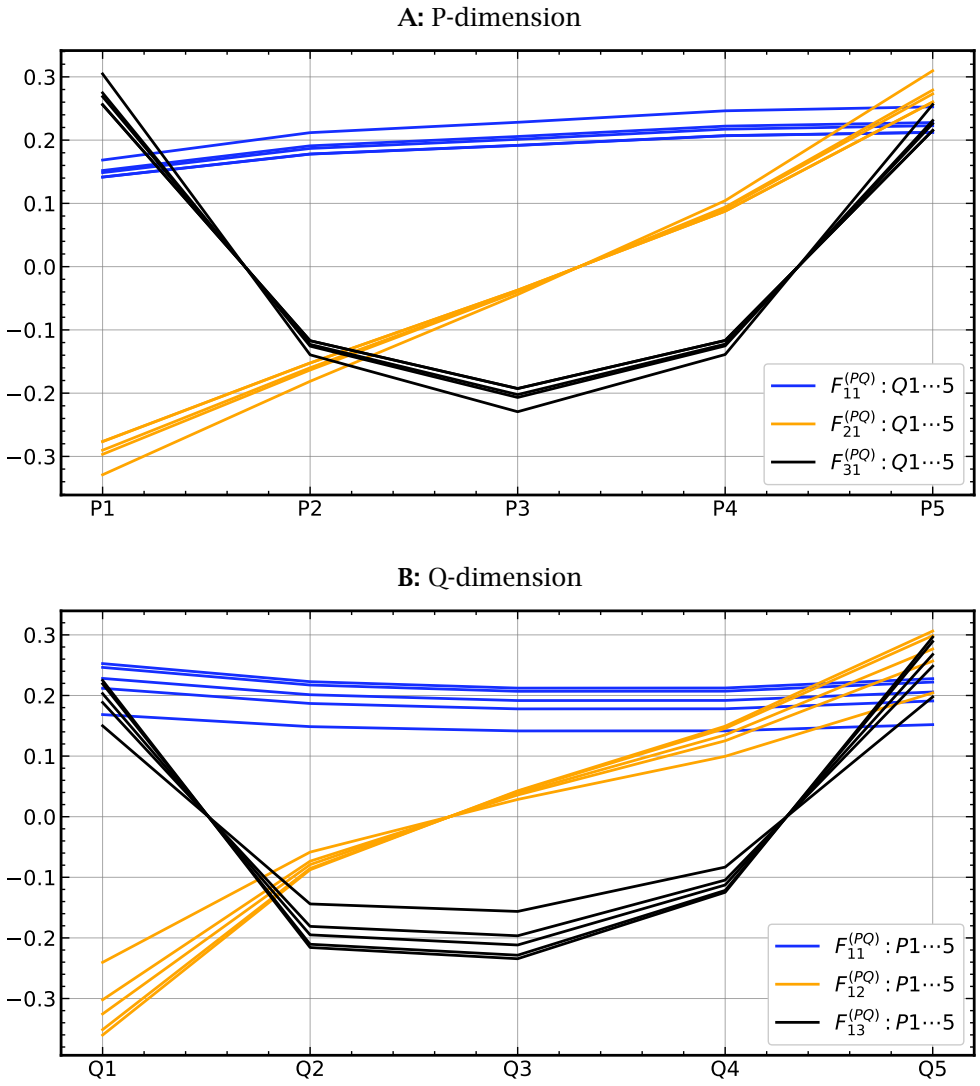


Figure C.7: Time series R^2 of 3D-PCA rolling windows



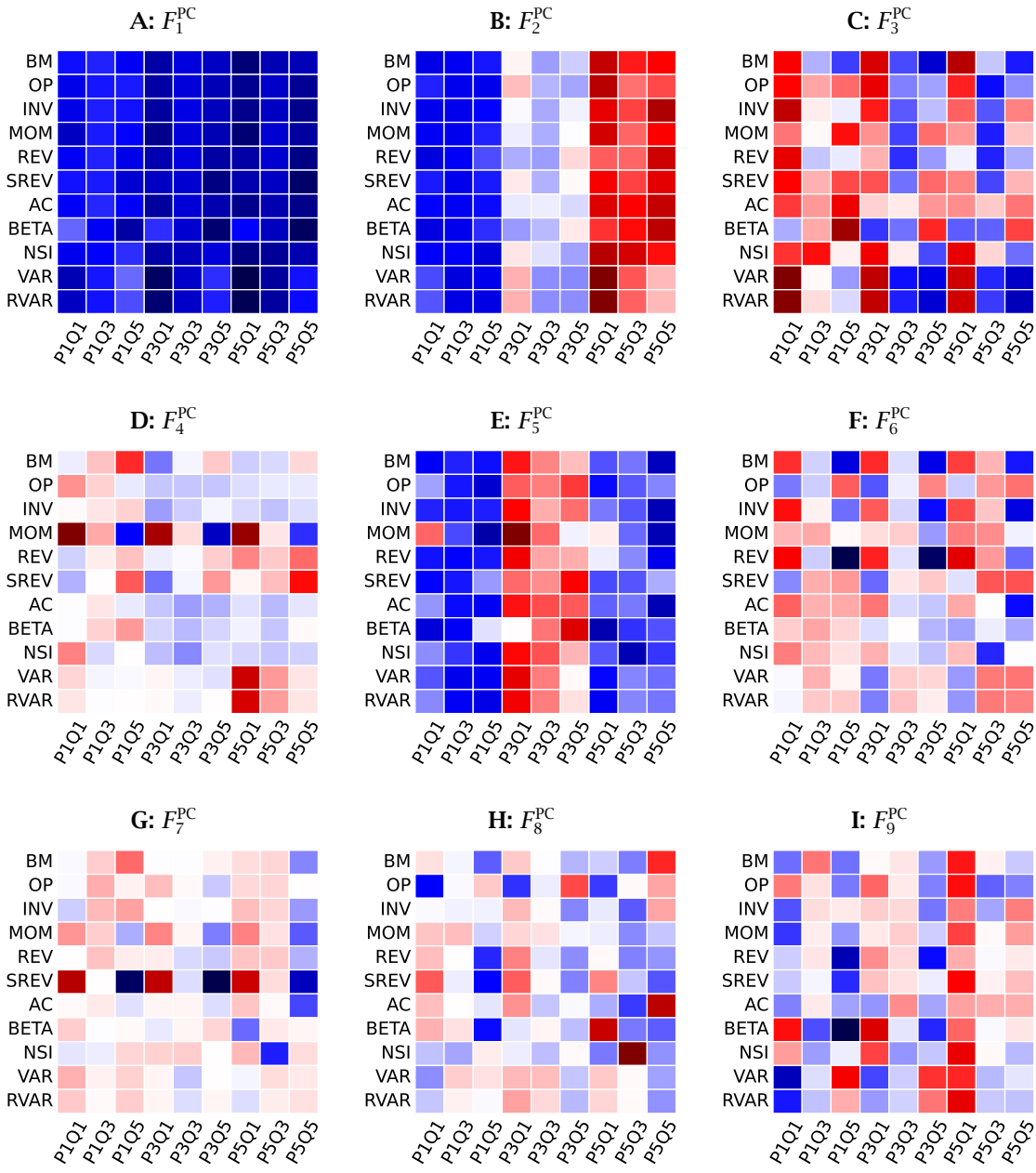
Notes: This figure plots the time series R^2 of 3D-PCA estimated in rolling windows of length $h = 120$. The R^2 is defined as $R^2 = 1 - \|\boldsymbol{\varepsilon}\| / \text{Var}(\boldsymbol{x})$, where $\boldsymbol{\varepsilon} = \boldsymbol{x} - \widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{x}}$ is the approximation of $\boldsymbol{x}$ given by a 3D-PCA model with $K_C = K_P = K_Q = 3$. The figure plots the R^2 in each subsample. The R^2 for the in-sample estimation is in parentheses. The sample is from July 1967 to October 2023.

Figure C.8: Level, slope, curvature



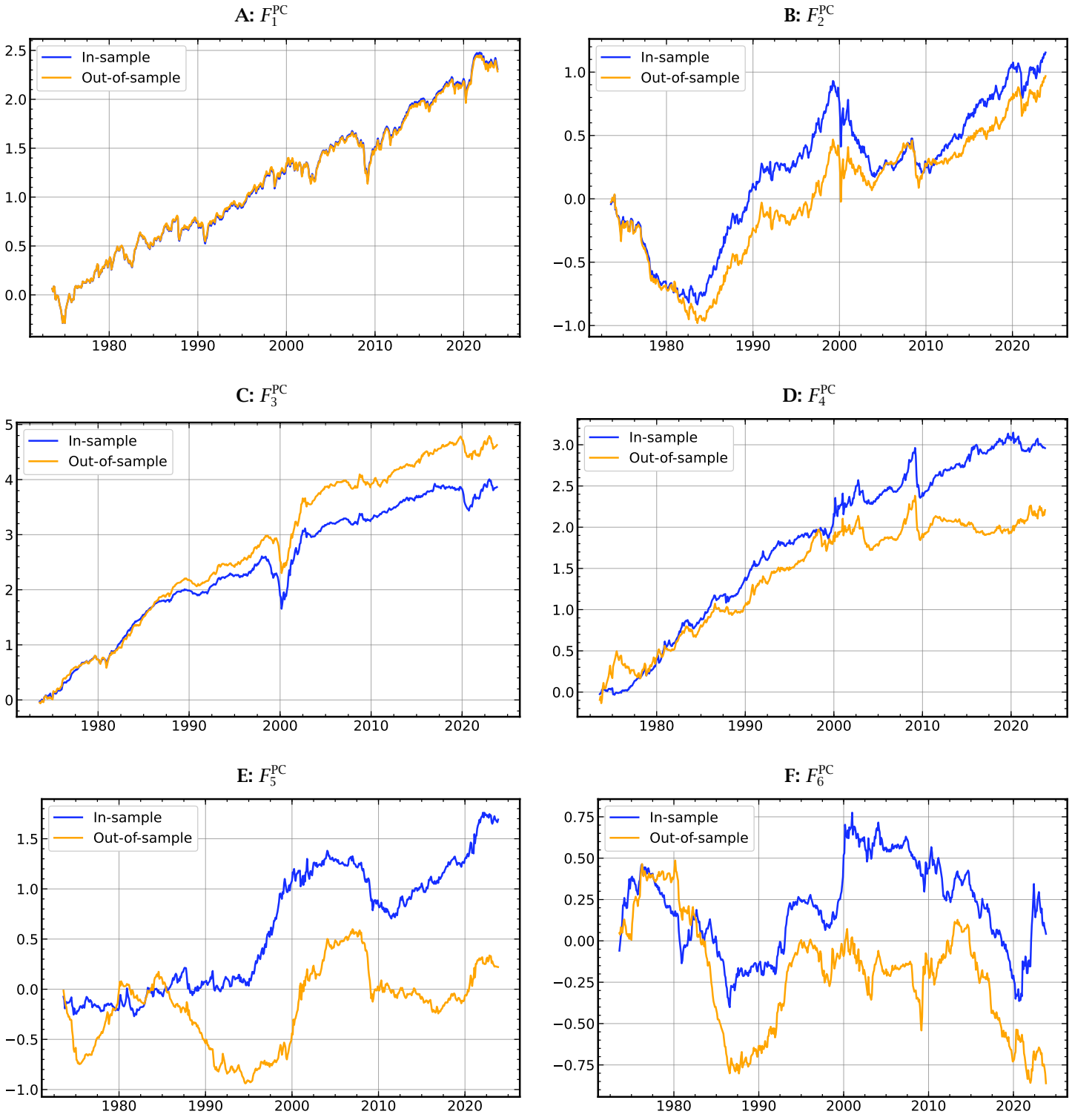
Notes: The heatmaps show the (5×5) matrices $\mathbf{W}_{pq}^{(PQ)}$, $p,q=1,2,3$ that are given by the outer product of the column vectors of $\hat{\mathbf{V}}^{(P)^\top}$, and $\hat{\mathbf{V}}^{(Q)^\top}$: $\mathbf{W}_{pq}^{(PQ)}=\mathbf{v}_p^{(P)}\circ\mathbf{v}_q^{(Q)}$. $\hat{\mathbf{V}}^{(P)^\top}$, and $\hat{\mathbf{V}}^{(Q)^\top}$ are from the estimation of a partial Tucker model (13) with $K_C=K_P=K_Q=3$. Negative values are plotted in red and positive ones in blue. The model is estimated by HOOI. The sample is from July 1967 to October 2023.

Figure C.9: PCA – Factor weights



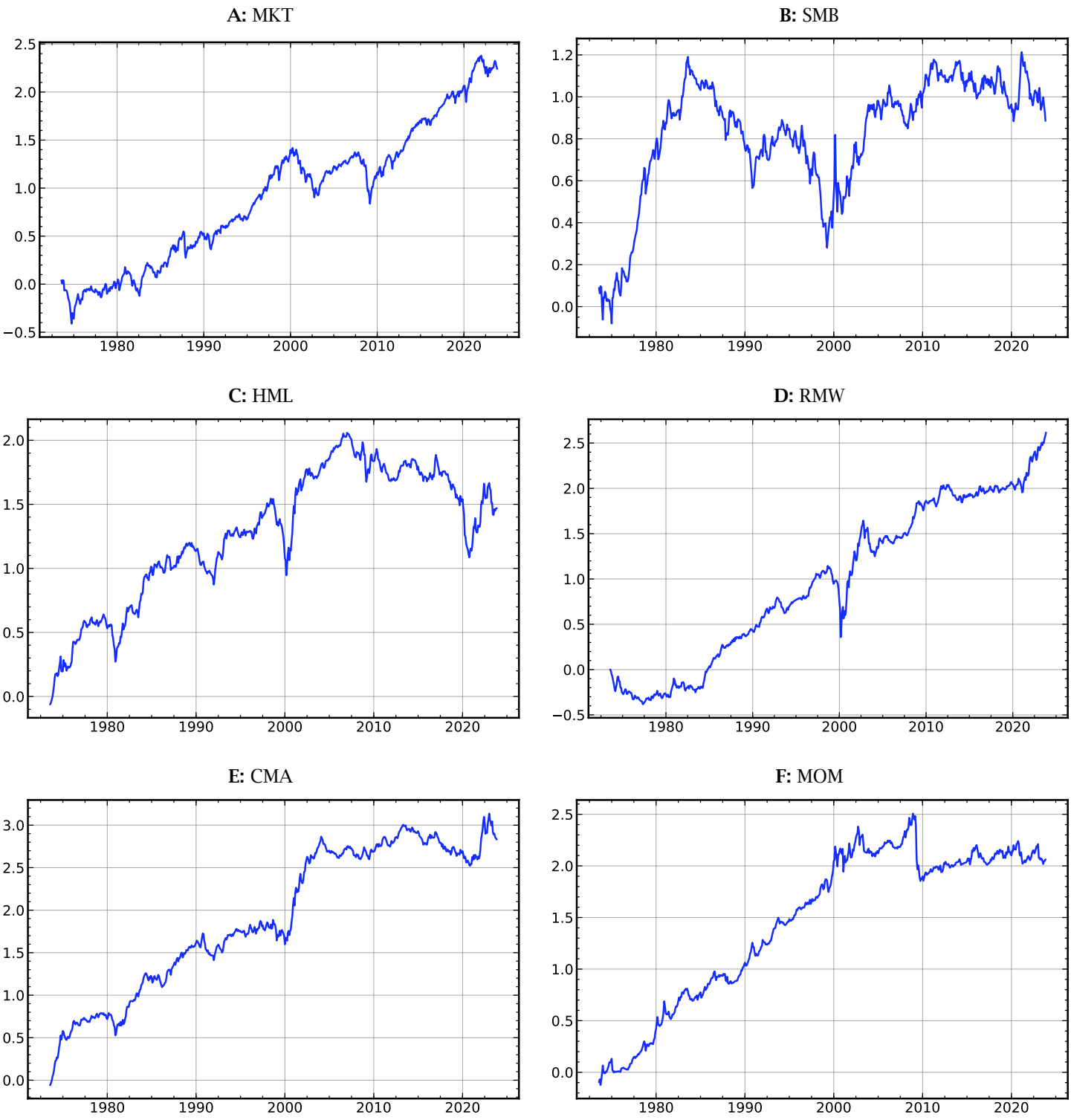
Notes:

Figure C.10: Cumulative log returns of PCA factors



Notes: The figure shows cumulative log-returns of first six PCA factors. In-sample factors are plotted in blue, and out-of-sample-factors are in orange. PCA factors are based on an estimation with 27 factors. The sample is from July 1967 to October 2023.

Figure C.11: Cumulative log returns of Fama-French factors



Notes: The figure shows cumulative log returns of Fama-French factors. The sample is from July 1967 to October 2023.