Multiperiod Collection of Results

Prof. Serhiy Kozak & James Zhang

January 27, 2025

Contents

Updates

- Small things like annualizing of Sharpe ratios in graphs, removing industry and constant factors in datasets
- Implemented formulas in the paper for mean and covariance matrix of basis assets in our primary tensor approach, results significantly improved
- Implemented benchmark models (precise tensor approach, pooled PCA approach, multiperiod (aggregate) PCA)
- Incorporating daily data into this framework proved a little bit cumbersome $\implies$ did not provide much more performance improvement.
- To-Do: a plausible explanation is due to the data imputation, which could help uncover patterns in the data. Specifically, these patterns may be driven by stocks with missing observations, perhaps because market participants trading these strategies only arbitraged away mispricing in stocks for which data was readily available. We should investigate this further at some point. TL;DR - results for OOS Characteristic Anomalies are unsatisfactory, and overall, Full-Sample + OOS PooledPCA and OOS Multiperiod PCA WRDS results are interesting, how much imputation here?
- To-Do: experiment with the mean-variance procedure (SCS, RP-PCA) as right now we are currently using the Markowitz formula with 0 regularization
- To-Do: implement T-FF6 and compare to FF6?

1 Setup

All R tensors are *log* returns, not simple and not excess of the one-lag return [1]. The out-of-sample period for all datasets is post January 2005. [2] In our primary specification,

- Rolling window size W is fixed at 120 months.
- Maximum horizon S is fixed at 36 months.
- Maximum lag (lagged characteristics) is 36 months.
- For each dataset and setting, we experiment with $K = 1, 3, 5, 10, 15, 20, 25$, where K represents the rank of PARAFAC and PCA ie. the number of factors. extracted from the tensor or matrix, respectively.

[1] Clarify that returns are excess of the risk-free rate, they should be, I know it is for the Characteristics Anomalies dataset

[2] We should standardize this to be 01-31-2005 to 2020-08-01, but in the preliminary, should not be a huge difference across the datasets.

- Default $\gamma = 0$, we do not yet experiment with RP-PCA in the MVE procedure.

For robustness purposes, we present multiperiod results for each of the following 3 datasets:

1. Characteristics Anomalies dataset: 1972-06-01 (ie. $\min(t-L)$) to 2020-08-01 . - 43 characteristic-sorted portfolios (SortedFactors)
2. SCS dataset: 1972-07-31 to 2021-12-31

   - 44 characteristic-based portfolios (UnivariateFactors). All industry and constant factors are now removed.
3. WRDS dataset: 1975-01-31 to 2020-12-31

   - 107 characteristic-based portfolios (UnivariateFactors). I believe this dataset still currently has imputed data points, so we should get a version of this with 0 imputed data points.

For the Characteristic Anomalies dataset, factors are decile-sorted long-short portfolios. For the SCS and WRDS datasets, each stock is weighted by its characteristic signal value and signals are based on cross-sectional ranks, centered and normalized.

2 Models

2.1 Tensor Factor Model (Approximate)

This is the primary model outlined in the paper. For time T and horizon S, slice the window of three-dimensional data $R \in \mathbb{R}^{W \times L \times N}$, apply PARAFAC to obtain

$$R_{t,i,l} = \sum_{k=1}^{K} \lambda_k F_{t,k} B_{t,k} W_{l,k}$$

or in tensor notation

$$R = \sum_{k=1}^{K} \lambda_k \cdot F \circ B_k \circ W_k \quad \text{for } F \in \mathbb{R}^{T \times K}, B \in \mathbb{R}^{N \times K}, W \in \mathbb{R}^{L \times K} \tag{1}$$

For clarity, let us denote the original returns R simply put as assets or returns and the K latent factors as the basis assets or factors. From here, for an S-period investment, we can compute the vector of multihorizon basis asset means and covariance matrix using the formulas in the paper,

$$\mu^{F,S} = \left(\sum_{l=1}^{S} W_l \right) * \mathbb{E}_t[F_t] \in \mathbb{R}^K \tag{2}$$

$$\Sigma^{F,S} = \left(\sum_{l=1}^{S} W_l W_l^\top \right) * \mathrm{Var}_t(F_t) \in \mathbb{R}^{K \times K} \tag{3}$$

where $*$ represents element-wise multiplication. Importantly, we assume that basis asset returns are uncorrelated over time, which is an empirically reasonable assumption. SDF weights are constructed using the simple Markowitz formula

$$w^S = \left(\Sigma^{F,S}\right)^{-1} \mu^{F,S} = \left(\left(\sum_{l=1}^{S} W_l W_l^\top\right) * \mathrm{Var}_t(F_t)\right)^{-1} \mathrm{diag}\left(\sum_{l=1}^{S} W_k\right) \mathbb{E}_t[F_t] \tag{4}$$

To test our estimators, $\forall\ s \in [1, S]$ ie. for each of the next S months, we solve the regression

$$F_{t+s}\left(\lambda * (W \odot B)\right)^\top = R_{t+s}$$

$$F_{t+s} = \mathrm{flat}(R_{t+s})((\lambda * (W \odot B))^\top)^\dagger \in \mathbb{R}^K \tag{5}$$

Completing this for all S months effectively yields the OOS time series of the basis assets $F' \in \mathbb{R}^{S\times K}$. Then calculate multiperiod returns by combining it with W such that

$$\sum_{s=1}^{S} F_{t+s,k} W_{s,k} \tag{6}$$

and apply the mean-variance weights w_S to compute our returns in the OOS S-month period, which we can then use to compute the Sharpe Ratio.

2.2 Tensor Factor Model (Precise)

This model is also based on our tensor framework; however, it *does not* make the assumption that basis asset returns are uncorrelated over time. For time T and horizon S, slice the same three-dimensional window of data $R \in \mathbb{R}^{W\times L\times N}$ as above. Apply 1 to obtain the time series of basis assets returns, lag components, and cross-sectional loadings. Next, construct a matrix of *overlapping* multiperiod returns that we will denote $FW \in \mathbb{R}^{(W-S+1)\times K}$ such that

$$FW_{t,k} = \sum_{s=1}^{S} F_{t+s-1,k} W_{s,k}\ \forall\ t \in 1 \cdots W - S + 1 \text{ and } k \in 1 \dots K \tag{7}$$

where $FW_{t,k}$ represents the S-month return of basis asset k beginning at time t. From here, we can easily compute the vector of observed sample means of basis asset k as

$$\mu^{F,S} = \frac{1}{W - S + 1} \sum_{t=1}^{W-S+1} FW_t = \overline{FW_t} \in \mathbb{R}^K \tag{8}$$

and we use the Newey-West Covariance estimator because the matrix FW contains overlapping returns

$$\Sigma^{F,S} = \text{Newey-West}(FW, \text{overlap} = S - 1) \in \mathbb{R}^{K\times K} \tag{9}$$

Construct SDF weights w^S using the Markowitz formula. Then, use 5 and 6 to obtain OOS basis asset multiperiod returns and combine with SDF weights to get the OOS return. If the approximate model in 2.1 out-performs this naive model, then we empirically justify our assumption that basis asset returns in separate months are uncorrelated.

2.3 Pooled PCA

Given time T and horizon S, [3]once again slice the window of three-dimensional data $R \in \mathbb{R}^{W \times L \times N}$ Let $\overrightarrow{R}$ be

$$\overrightarrow{R} = \begin{pmatrix} R_{1,1,:} & \cdots & R_{1,L,:} \\ \vdots & \vdots & \vdots \\ R_{W,1,:} & \cdots & R_{W,L,:} \end{pmatrix} \in \mathbb{R}^{W \times NL}.$$

The Pooled PCA estimator interprets the factor returns conditioned on lagged characteristics as new cross-sectional observations and applies traditional PCA to this matrix to obtain basis assets and loadings

$$\overrightarrow{R} = \widehat{F}\widehat{\Lambda}^\top \tag{10}$$

where $\widehat{F} \in \mathbb{R}^{W \times K}$ and $\widehat{\Lambda} \in \mathbb{R}^{NL \times K}$. Calculate basis asset moments

$$\mu^{F,S} = \overline{\widehat{F}} \in \mathbb{R}^K \text{ and } \Sigma^{F,S} = \frac{1}{W}\widehat{F}^\top \widehat{F} - \overline{\widehat{F}}\,\overline{\widehat{F}}^\top \in \mathbb{R}^{K \times K} \tag{11}$$

and the subsequent SDF weights using the Markowitz formula. OOS factor returns are estimated by a regression of the returns on the estimated loadings

$$\widehat{F}_{OOS} = \overrightarrow{R}_{OOS}\widehat{\Lambda}\left(\widehat{\Lambda}^\top\widehat{\Lambda}\right)^{-1} \in \mathbb{R}^S \tag{12}$$

and combined with the mean-variance weights yields the OOS return. This method freely learns the structure of both the lag components and the cross-sectional loadings ie. $\lambda(B \circ W)$; on the other hand, applying PARAFAC forces this decomposition. The tensor based approaches in 2.1 and 2.2 impose a more structured, economically-intuitive approach and should theoretically outperform this relatively nonparametric method.

2.4 Multihorizon PCA

Given time T and horizon S, consider the tensor $R \in \mathbb{R}^{W \times L \times N}$. Construct a matrix of *overlapping* S-month horizon returns denoted as $R^S \in \mathbb{R}^{(W-S+1) \times N}$ such that $\forall\ t \in 1 \ldots W - S + 1$ and $i \in 1 \ldots N$

$$R^S_{t,i} = \sum_{l=1}^{S} R_{t+l,l,i} = \sum_{l=1}^{S} r_{t+l,i}|C_t \tag{13}$$

[3]This method is in a sense independent of the horizon, is this correct?

or in plain English, the S-month buy-and-hold return of characteristic-based portfolio i conditioned on characteristics from the current month t. Apply PCA to R^S obtain basis assets $\widehat{F}$ and loadings $\widehat{\Lambda}$. Compute basis asset moments

$$\mu^{F,S} = \frac{1}{W-S+1} \sum_{t=1}^{W-S+1} \widehat{F}_t = \overline{\widehat{F}} \in \mathbb{R}^K$$
$$\Sigma^{F,S} = \text{Newey-West}(\widehat{F}, \text{overlap} = S-1) \in \mathbb{R}^{K \times K}$$

Construct SDF weights and then use 12 to find OOS basis asset values and the model's subsequent OOS returns.

2.5 Multihorizon Model-Free

Once more, given time T and horizon S, obtain the window of data and construct the same matrix of overlapping S-month returns $R^S \in \mathbb{R}^{(W-S+1)\times N}$. Rather than applying PCA to this panel, compute the moments of the factors as is.

$$\mu^{R,S} = \frac{1}{W-S+1} \sum_{t=1}^{W-S+1} R_t^S = \overline{R^S} \in \mathbb{R}^N$$
$$\Sigma^{R,S} = \text{Newey-West}(R^S, \text{overlap} = S-1) \in \mathbb{R}^{N \times N}$$

Construct mean-variance weights and combine with the OOS factor values to obtain the S-month return. In theory, as $K \to N$, the Sharpe Ratios from 2.4 should approach this Model-Free method.

3 Collection of Results

List of Figures

List of Tables

3.1 Primary Specification: Full-Sample

3.1.1 Full-Sample Characteristic Anomalies Dataset

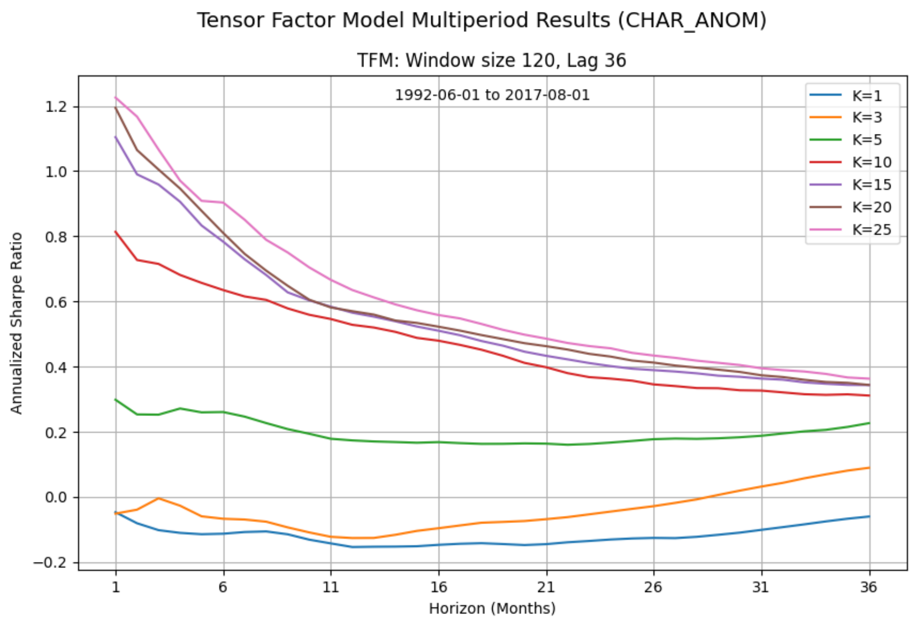

Figure 1: (CA) Tensor Factor Model: Full-Sample Annualized SR

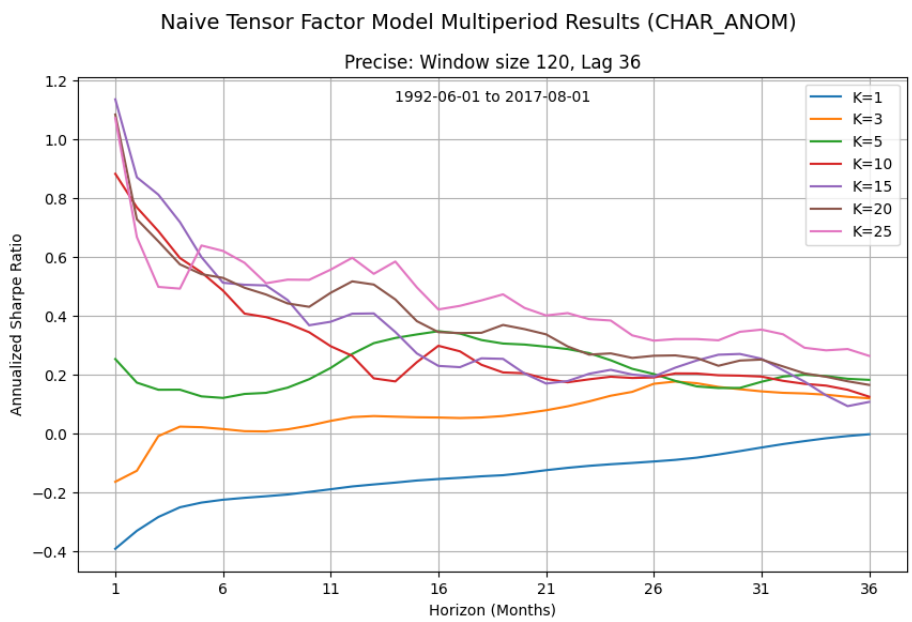

Figure 2: (CA) Precise Tensor Factor Model: Full-Sample Annualized SR

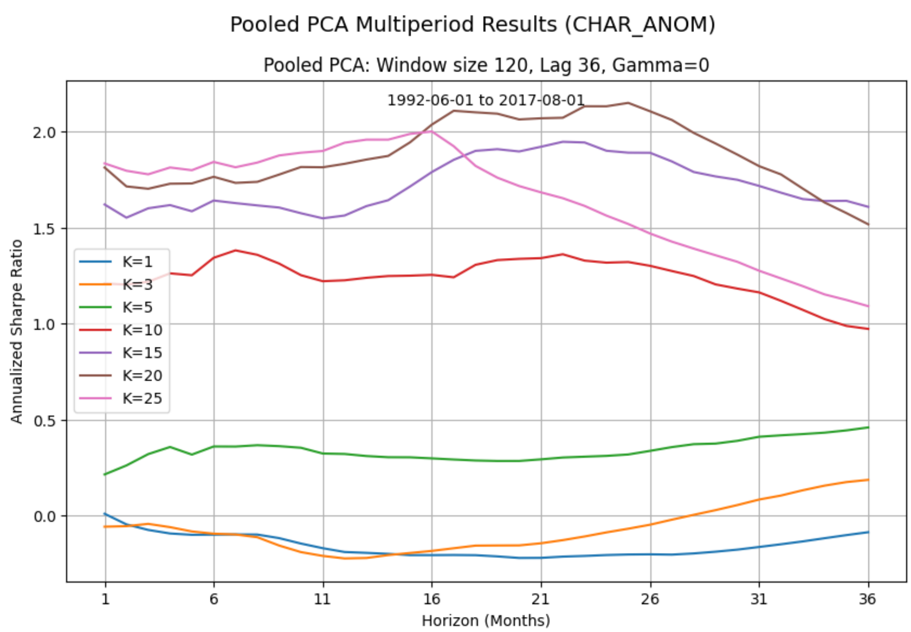

Figure 3: (CA) Pooled PCA Model: Full-Sample Annualized SR

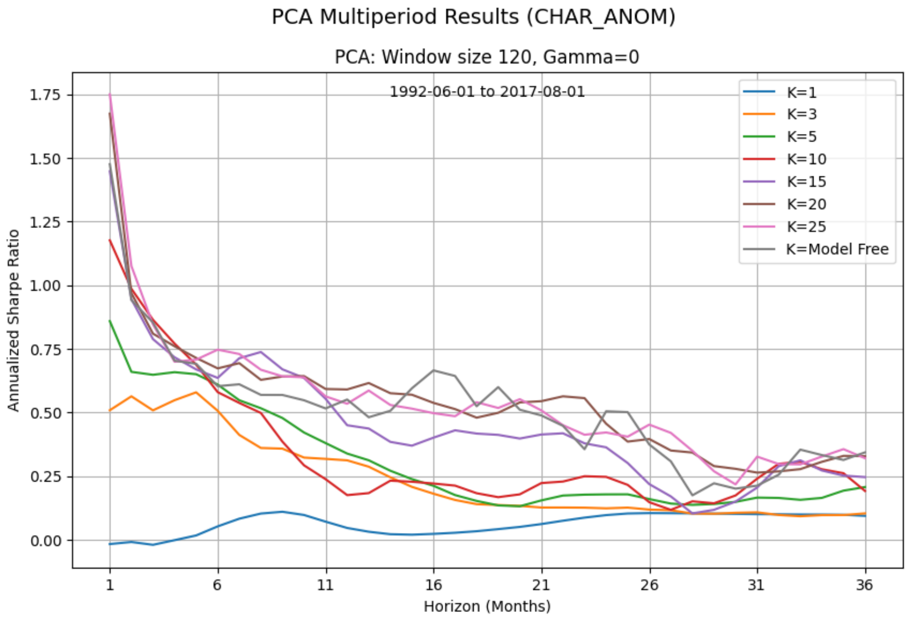

Figure 4: (CA) Multiperiod PCA Model: Full-Sample Annualized SR

still need to add the pca graph for CA dataset + ensure that the model-free sharpes are included for all datasets

maybe a table that's like

model $\implies$ K value (or model-free) $\implies$ sharpes for all 36 horizons

3.1.2 Full-Sample SCS Dataset

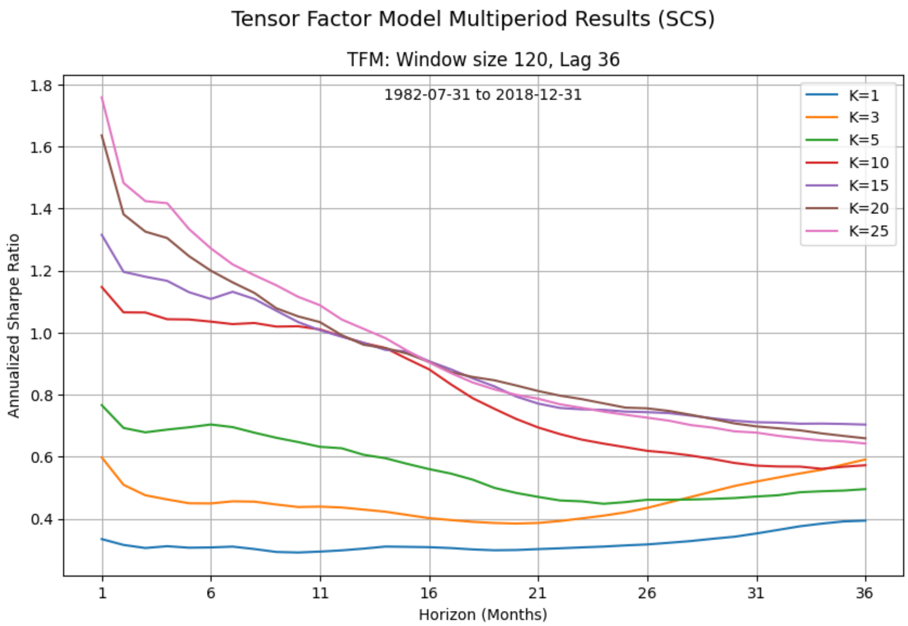

Figure 5: (SCS) Tensor Factor Model: Full-Sample Annualized SR

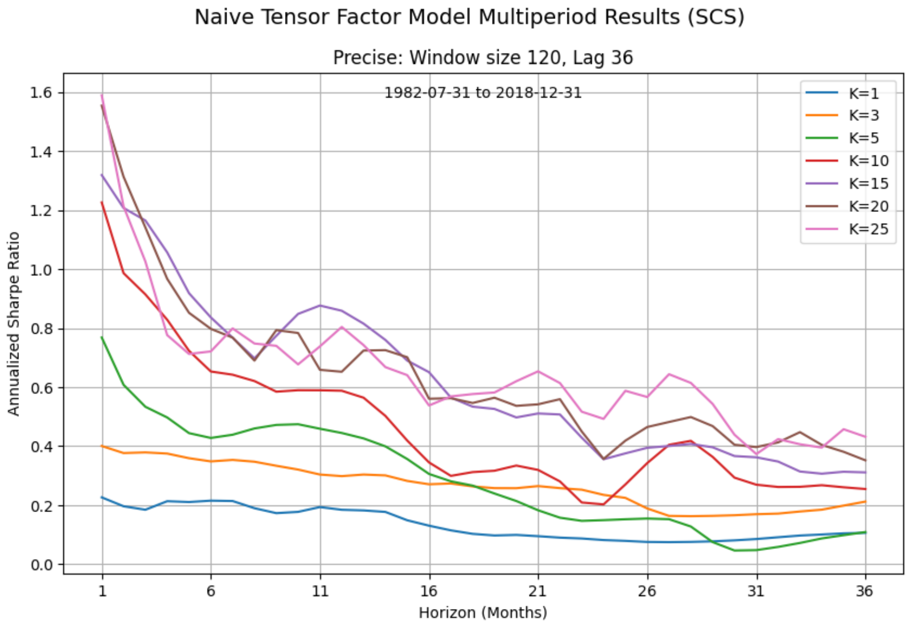

Figure 6: (SCS) Precise Tensor Factor Model: Full-Sample Annualized SR

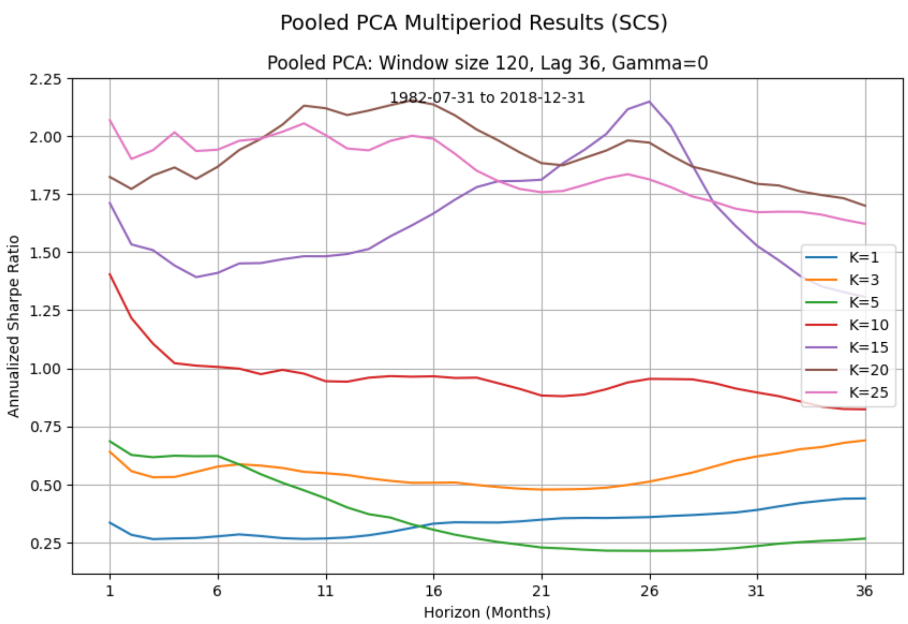

Figure 7: (SCS) Pooled PCA Model: Full-Sample Annualized SR

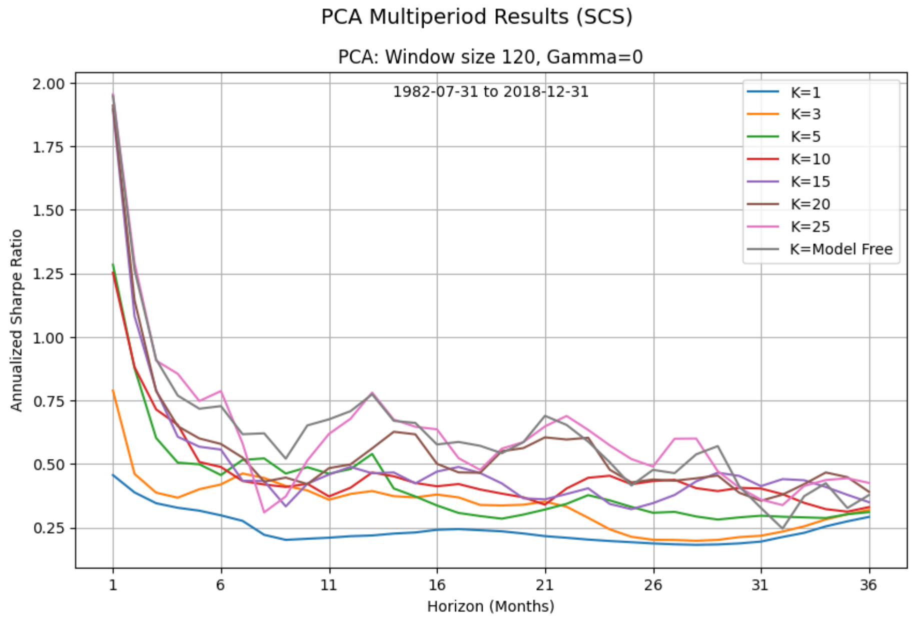

Figure 8: (SCS) Multiperiod PCA Model: Full-Sample Annualized SR

3.1.3 Full-Sample WRDS Dataset

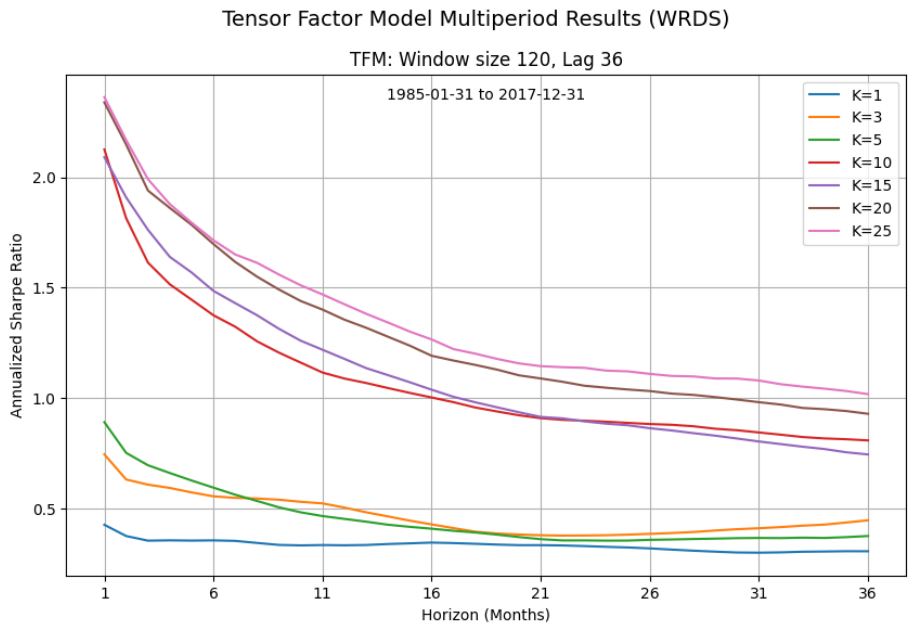

Figure 9: (WRDS) Tensor Factor Model: Full-Sample Annualized SR

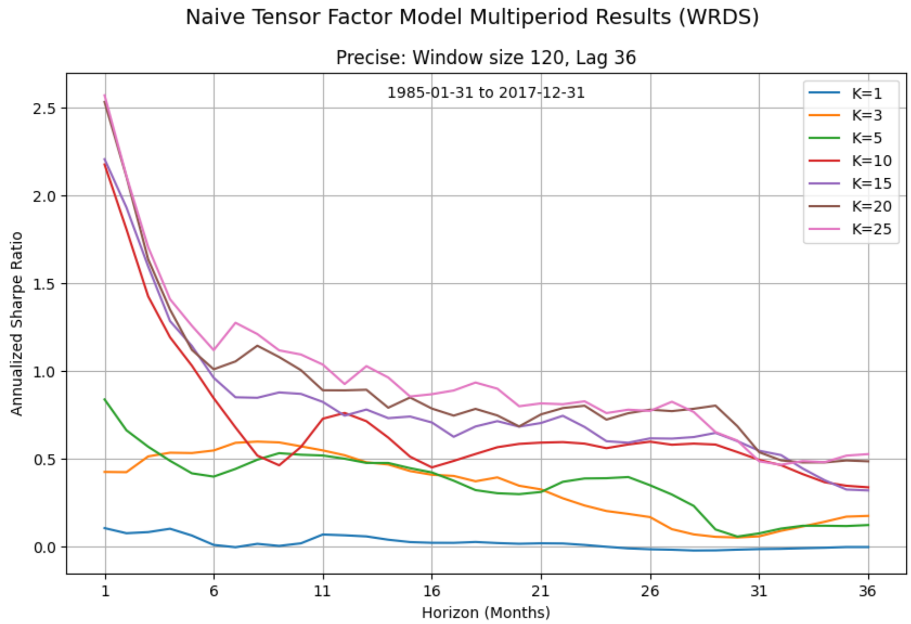

Figure 10: (WRDS) Precise Tensor Factor Model: Full-Sample Annualized SR

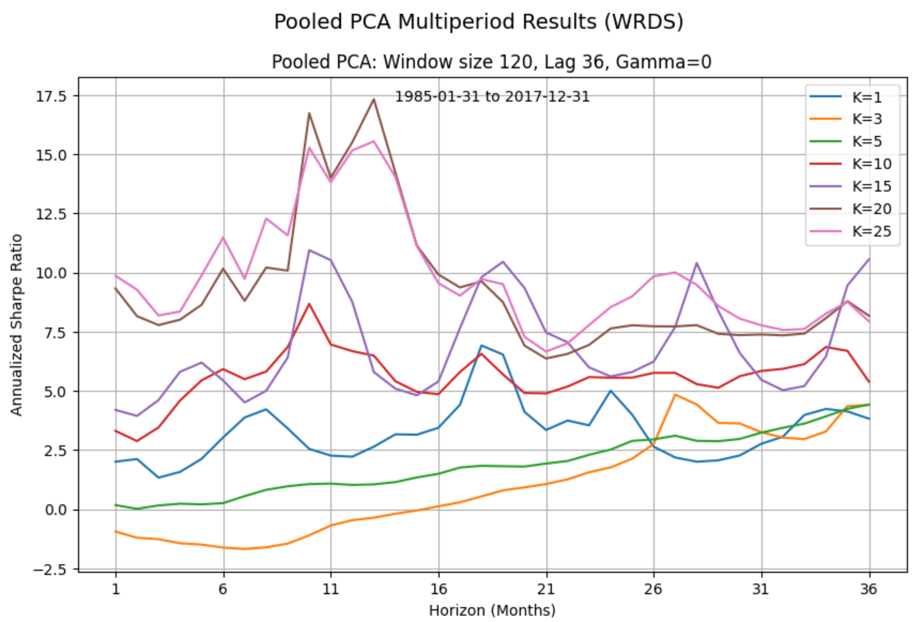

Figure 11: (WRDS) Pooled PCA Model: Full-Sample Annualized SR

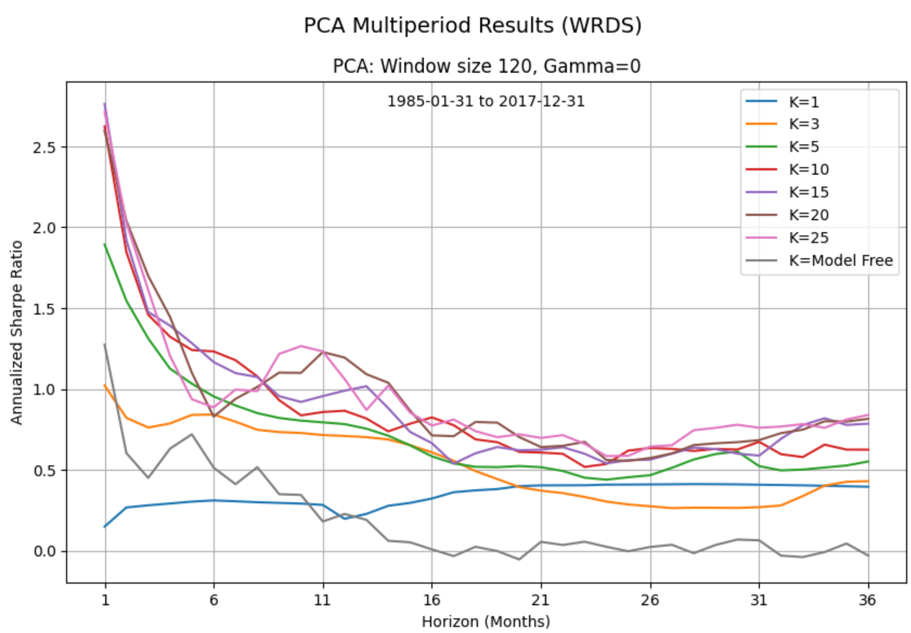

Figure 12: (WRDS) Multiperiod PCA Model: Full-Sample Annualized SR

4 Appendix A

4.1 Primary Specification: OOS Results

4.1.1 OOS Characteristic Anomalies Dataset

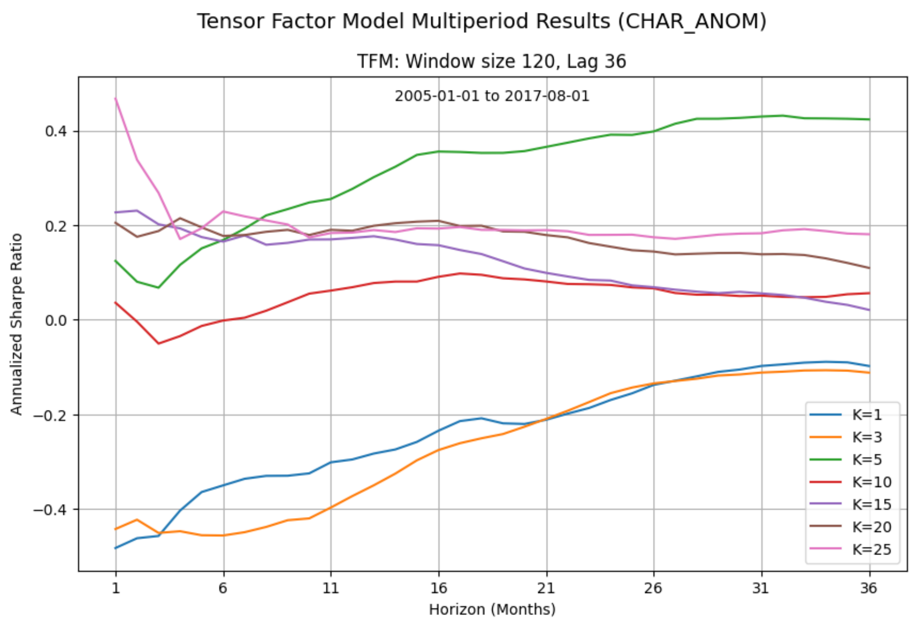

Figure 13: (CA) Tensor Factor Model: OOS Annualized SR

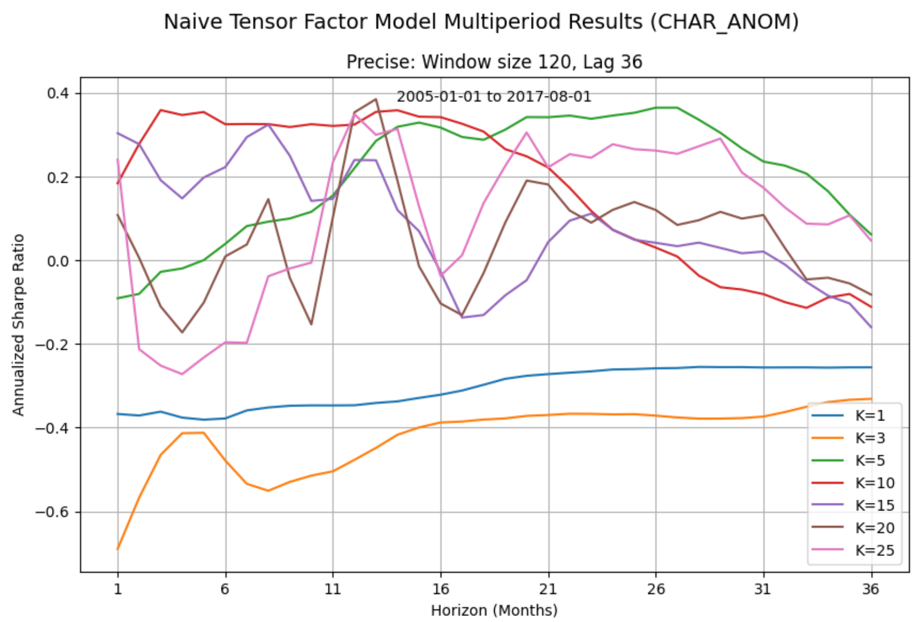

Figure 14: (CA) Precise Tensor Factor Model: OOS Annualized SR

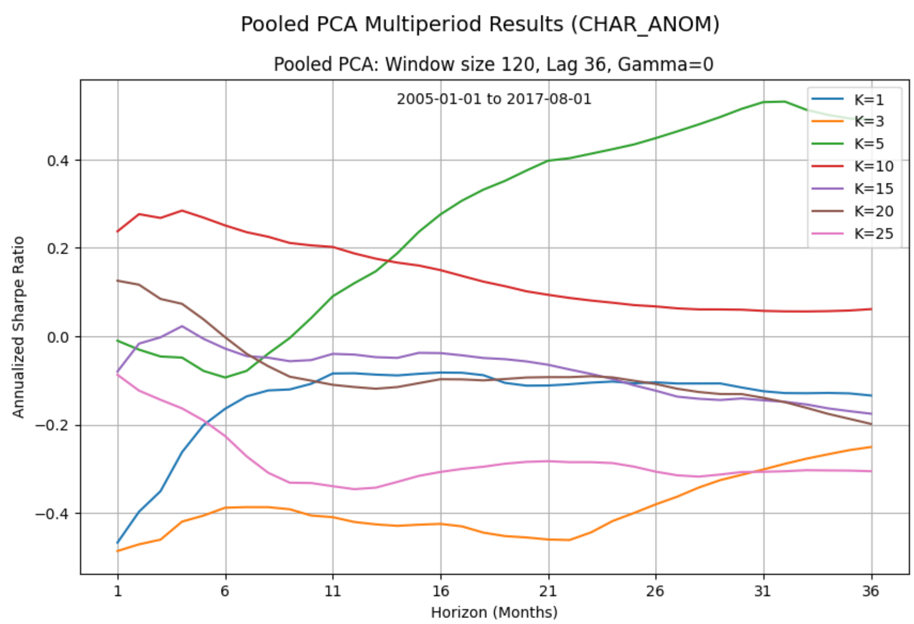

Figure 15: (CA) Pooled PCA Model: OOS Annualized SR

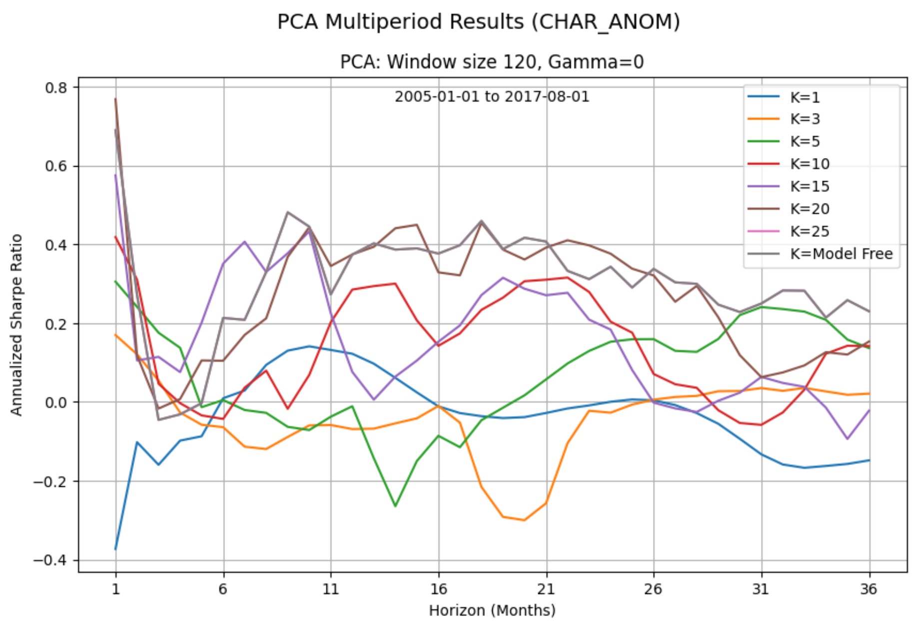

Figure 16: (CA) Multiperiod PCA Model: OOS Annualized SR

4.1.2 OOS SCS Dataset

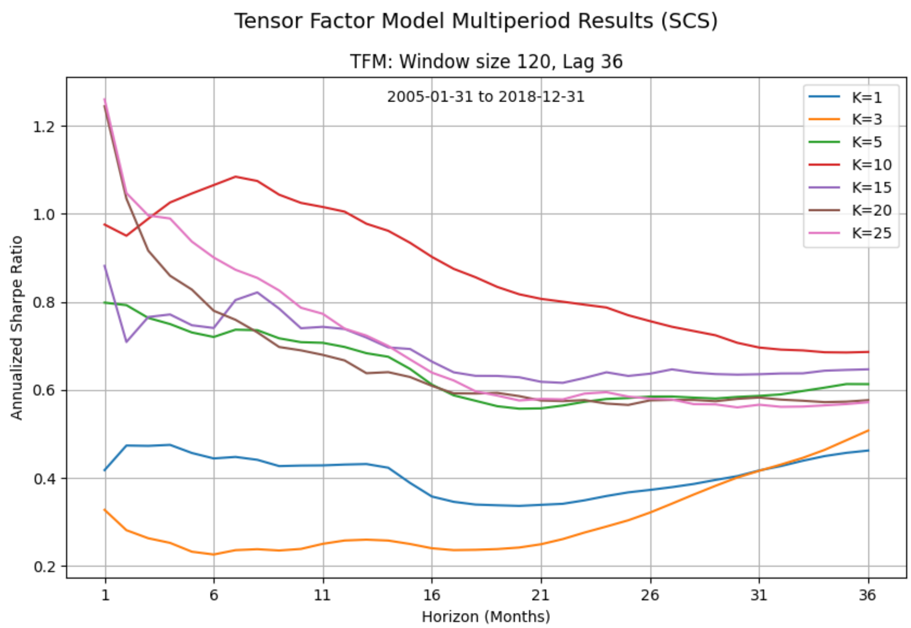

Figure 17: (SCS) Tensor Factor Model: OOS Annualized SR

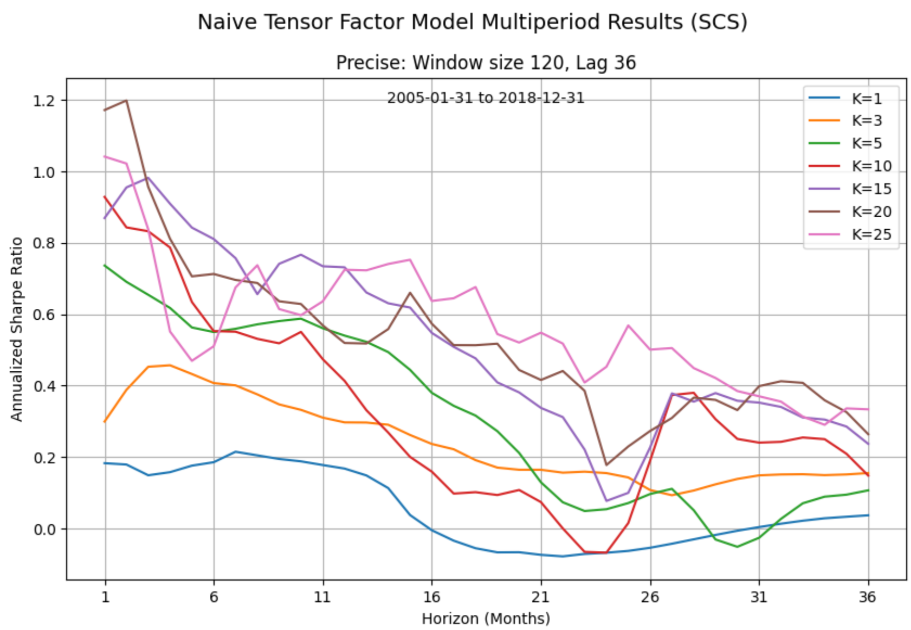

Figure 18: (SCS) Precise Tensor Factor Model: OOS Annualized SR

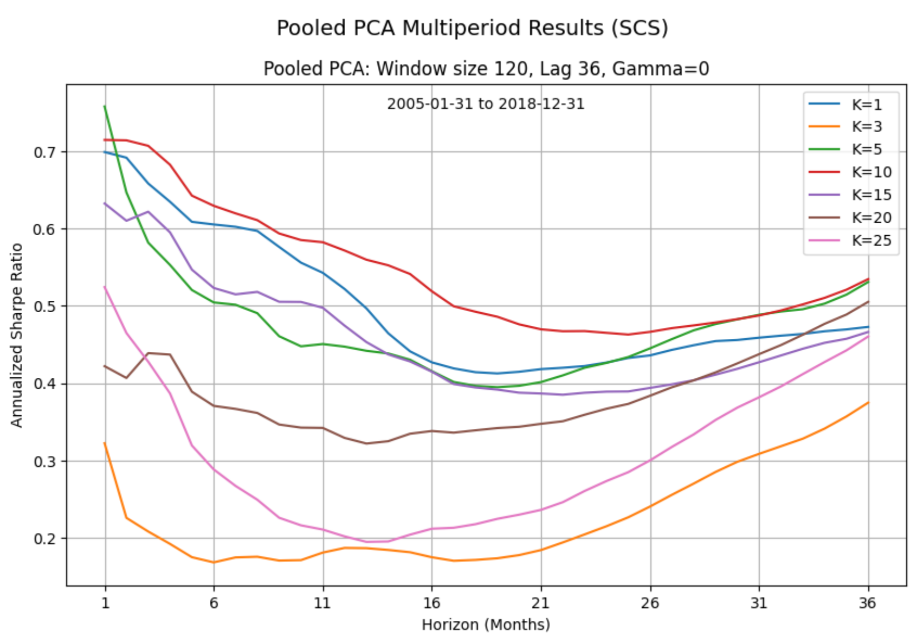

Figure 19: (SCS) Pooled PCA Model: OOS Annualized SR

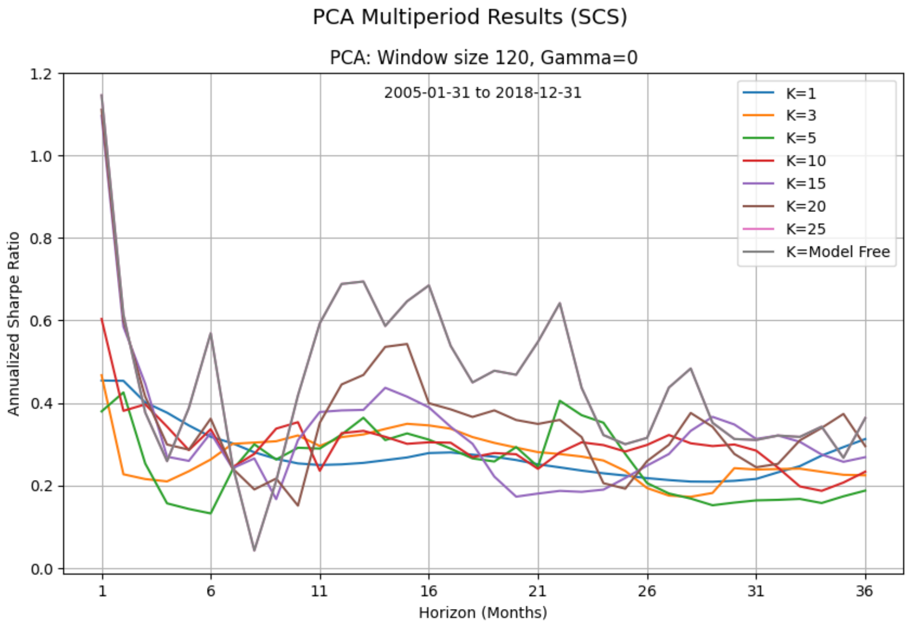

Figure 20: (SCS) Multiperiod PCA Model: OOS Annualized SR

4.1.3 OOS WRDS Dataset

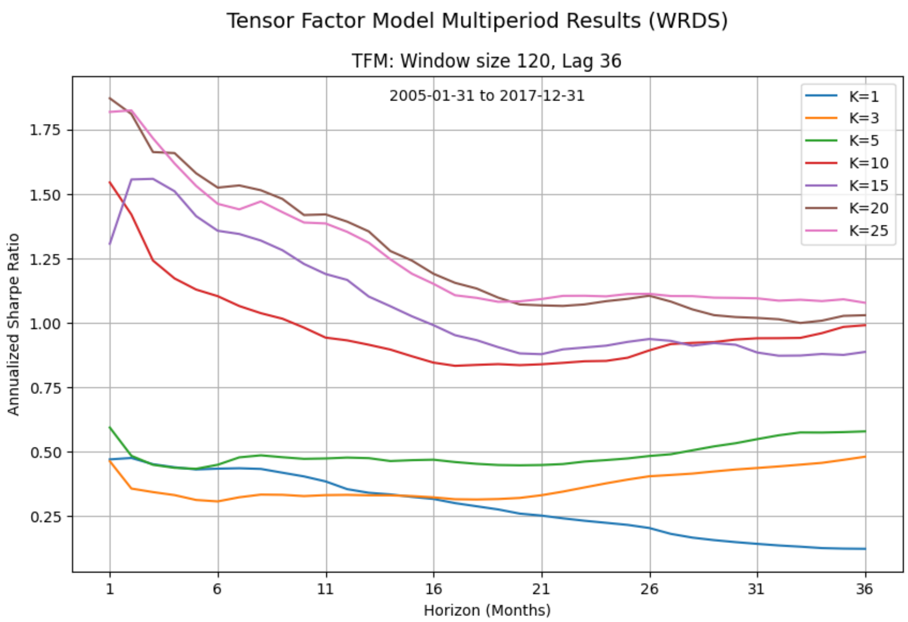

Figure 21: (WRDS) Tensor Factor Model: OOS Annualized SR

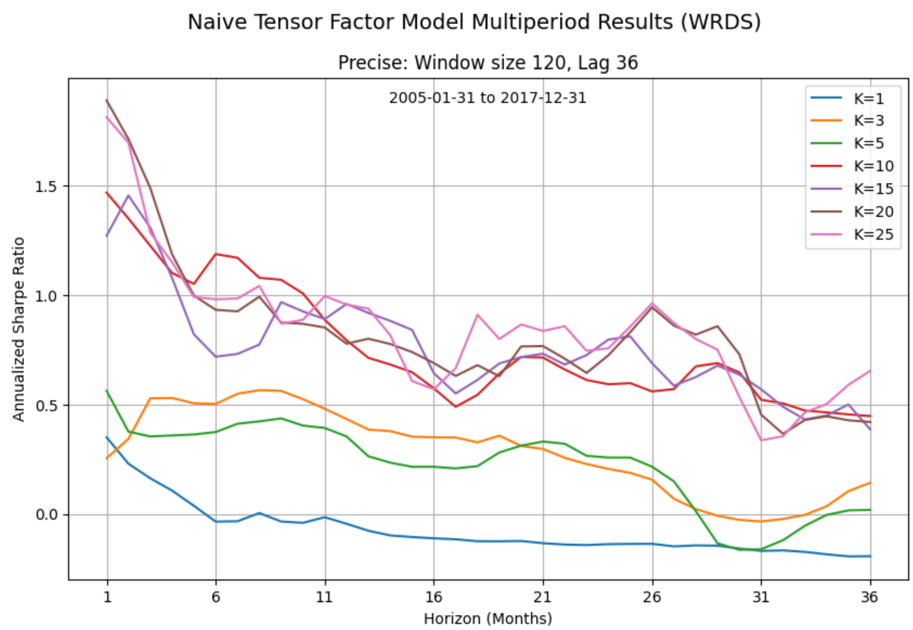

Figure 22: (WRDS) Precise Tensor Factor Model: OOS Annualized SR

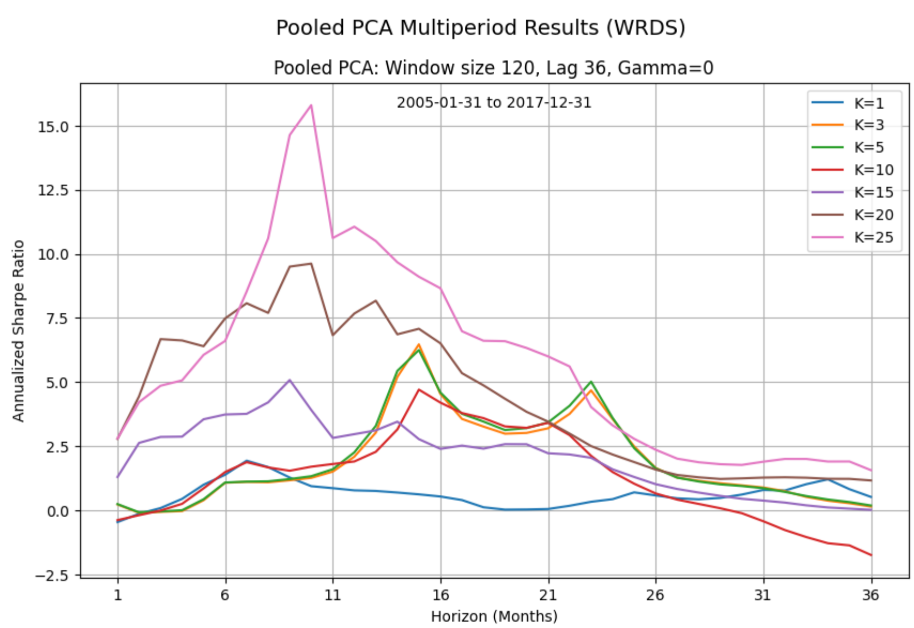

Figure 23: (WRDS) Pooled PCA Model: OOS Annualized SR

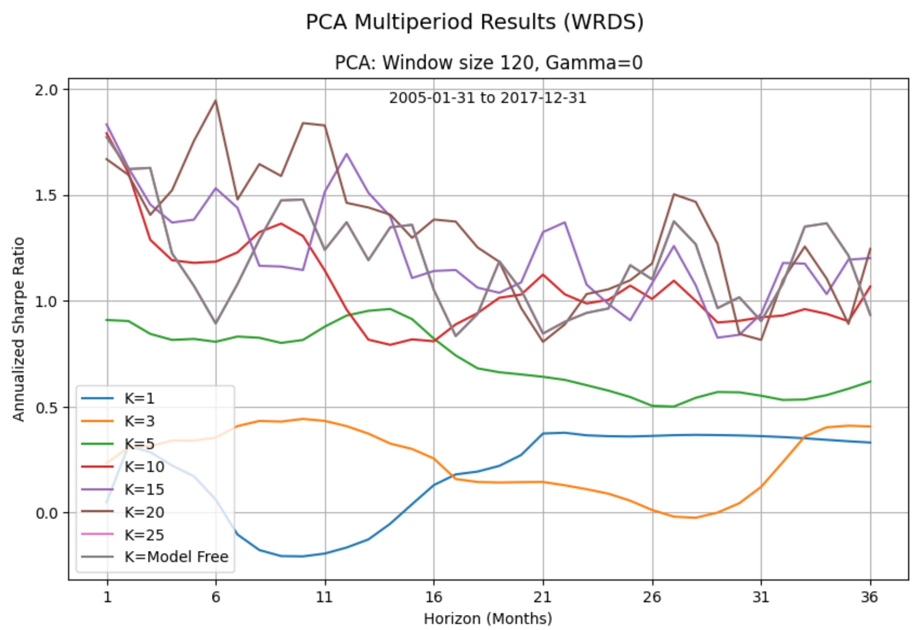

Figure 24: (WRDS) Multiperiod PCA Model: OOS Annualized SR