Instrumented Principal Component Analysis

Nhi Truong, Yu Wang

September 30, 2019

1 Introduction

One of the most fundamental result in asset pricing theory is that the cross-sectional variation in asset prices (expected returns) should be explained by the corresponding exposure to risk factors. The theory of arbitrage pricing developed by Ross (1976), Rubinstein (1976), Harrison and Kreps (1977) and Hansen and Richard (1987) states that, with the absence of no-arbitrage, there exists a stochastic discount factor m_t such that, the excess return $r_{i,t}$ of any stock i at time t must satisfy

$$\mathbb{E}_t[m_{t+1}r_{i,t+1}] = 0.$$

This implies that

$$\mathbb{E}_t[r_{i,t+1}] = \underbrace{\frac{\mathrm{Cov}(m_{t+1}, r_{i,t+1})}{\mathrm{Var}_t(m_{t+1})}}_{\beta_{i,t}} \underbrace{\left(-\frac{\mathrm{Var}_t(m_{t+1})}{\mathbb{E}_t[m_{t+1}]}\right)}_{\lambda_t}. \tag{1.1}$$

Clearly, we can interpret $\beta_{i,t}$ as exposure to systematic risk factors, and λ_t as the risk premium of factors. This theoretical equation highlights the importance of approximate factor modeling in asset pricing theory.

The paper "Characteristics are Covariances" by Kelly-Pruitt-Su (2017) develops a statistical method called Instrumented Principal Component Analysis (IPCA) that allows for incorporating observation characteristics to estimate latent factors in a time-varying fashion. As we shall see, their approach could be essentially thought of as first projecting individual stock returns into managed-porfolios based on characteristics (which provides dimension reduction), and then performing PCA on that.

2 Model

2.1 Individual Stock Interpretation

The IPCA model specifies that

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} \underbrace{f_{t+1}}_{K\times 1} + \epsilon_{i,t+1} \tag{2.1}$$
$$\alpha_{i,t} = z'_{i,t} \underbrace{\Gamma_\alpha}_{L\times 1} + \nu_{\alpha,i,t}, \quad \beta_{i,t} = z'_{i,t} \underbrace{\Gamma_\beta}_{L\times K} + \nu_{\beta,i,t}$$

Here, $r_{i,t}$ is the excess return of some stock i at time t. The vectors $z_{i,t} \in \mathbb{R}^L$ contains information on observable characteristics. Thus, we can see that the model specifies that the factor loading $\beta_{i,t}$ is, up to the noise term $\nu_{a,i,t}$, linearly dependent on the observable characteristic $z_{i,t}$. Thus, we expect an improvement over the standard PCA approach, where the (latent) factor structure is extracted solely from return data because the IPCA method allows for instrumenting on characteristics.

Another feature of IPCA is its reduction of dimensions: the matrix Γ_β can be interpreted as a linear map from the high-dimensional space of observable characteristics to some low dimensional space of risk factors, assuming that $L \gg K$.

As we can see, the model also specifies an analogous setting for the alpha $\alpha_{i,t}$ of stocks. In particular, $\Gamma_\alpha = 0$ corresponds to the case where we assume no relationship between the characteristics and stock alphas. On the other hand, in an unrestricted model where $\Gamma_\alpha \neq 0$, if the relationship between characteristics and expected stock returns proves to be different from that between characteristics and risk factors, then IPCA will give a nonzero Γ_α estimate.

Note that IPCA is a subcase of the Projected-PCA and Projected RP-PCA approaches introduced by Fan, Liao, Wang (2016) and Lettau, Pelger (2019), respectively.

2.2 Managed Portfolio Interpretation

A different way to interpret the IPCA model is on the level of managed portfolios. In particular, let's look at the vector form of (2.1) in the case where $\Gamma_\alpha = 0$:

$$r_{t+1} = Z_t \Gamma_\beta f_{t+1} + \epsilon^*_{t+1}, \tag{2.2}$$

where $r_{t+1} \in \mathbb{R}^N$ is the vector of returns of N assets at time $t+1$, $Z_t \in \mathbb{R}^{N\times L}$ is the matrix of stacked characteristics, and $\epsilon^*_{t+1} \in \mathbb{R}^N$ is a vector of model residual.

In particular, if we are to define

$$X_t = Z'_t r_{t+1},$$

Then our model becomes

$$X_t = Z'_t Z_t \Gamma_\beta f_{t+1} + \tilde{\epsilon}_{t+1}$$

Note that if $Z'_t Z_t$ is not too volatile, then the combined loading $\Lambda_t = Z'_t Z_t \Gamma_\beta$ is more or less constant, and then the result of IPCA is more or less the same with that of doing PCA on managed portfolios. We will explain this in more details in section 3.1.1

2.3 Noise Strcture of the Two Approaches

Recall that in the IPCA model (2.2), the noise ϵ^*_{t+1} is assumed to be i.i.d. Thus, from the view of portfolio management:

$$X_t = Z_t' Z_t \Gamma_\beta f_{t+1} + \tilde{\epsilon}_{t+1}, \tag{2.3}$$

we have $\tilde{\epsilon}_{t+1} = Z_t' \epsilon^*_{t+1}$. Thus, the noise term is now dependent with covariance structure

$$\mathrm{Cov}(\tilde{\epsilon}_{t+1}, \tilde{\epsilon}_{t+1}) = Z_t' Z_t.$$

In particular, in the static approach when one assumes $Z_t' Z_t \approx I$ across time, the main difference is that one assumes that the noise structure is i.i.d. On the other hand, the dynamic IPCA assumes that the i.i.d. noise structure only comes from the individual stock level, and so when we form portfolios, we have to realize that the noise structure is now dependent on the formation of portfolios given by Z_t.

On the other hand, we can perform noise-whitening on the model by multiplying both sides of equation (2.4) by $(Z_t' Z_t)^{-1/2}$. This gives

$$(Z_t' Z_t)^{-1/2} X_t = (Z_t' Z_t)^{1/2} \Gamma_\beta f_{t+1} + \epsilon'_{t+1}, \tag{2.4}$$

where ϵ'_{t+1} is again i.i.d. Note that the left hand side can be written as $(Z_t' Z_t)^{-1/2} Z_t' R_t$, so we can view this as a mapping and rotation of R_t to a L-dimensional space because the matrix $M_t = (Z_t' Z_t)^{-1/2} Z_t'$ satisfies $M_t M_t' = I_{L \times L}$.

3 Estimation

3.1 Restricted Model: $\Gamma_\alpha = 0$

Given the vector form of our model in (2.2), our objective is to minimize square loss:

$$\min_{\Gamma_\beta, f_t} \sum_{t=1}^{T-1} (r_{t+1} - Z_t \Gamma_\beta f_{t+1})' (r_{t+1} - Z_t \Gamma_\beta f_{t+1}) \tag{3.1}$$

Solving for first order conditions, we get that the solution $\hat{\Gamma}_\beta$ and $\hat{f}_t$, $1 \le t \le T$ of this minimization problem satisfies the system

$$\hat{f}_{t+1} = (\hat{\Gamma}_\beta' Z_t' Z_t \hat{\Gamma}_\beta)^{-1} \hat{\Gamma}_\beta' Z_t' r_{t+1} \tag{3.2}$$

$$\mathrm{vec}(\hat{\Gamma}_\beta) = \left(\sum_{t=1}^{T-1} Z_t' Z_t \otimes \hat{f}_{t+1} \hat{f}_{t+1}' \right) \left(\sum_{t=1}^{T-1} [Z_t \otimes \hat{f}_{t+1}']' r_{t+1} \right) \tag{3.3}$$

In this general setting, there is no closed form solution. However, the authors proposed 2 methods: numerical approximation via alternating least square, which retains some of the dynamic aspect of the model, and a closed-form scenario where the result is indeed applying PCA to managed portfolios as we have discussed in section 2.2.

We will first discuss the case corresponding to managed portfolios because it gives some intuition to the initialization step in the alternating least square case.

3.1.1 PCA on Managed Portfolios

In this setting, one standardizes by performing cross-sectional orthonormalization on the characteristics per time period. In other words, we assume that $Z_t'Z_t = \mathbb{I}_L$. Then equation (3.2) becomes

$$\hat{f}_{t+1} = (\hat{\Gamma}_\beta'\hat{\Gamma}_\beta)^{-1}\hat{\Gamma}_\beta' Z_t' r_{t+1} = (\hat{\Gamma}_\beta'\hat{\Gamma}_\beta)^{-1}\hat{\Gamma}_\beta' X_t,$$

i.e. giving $\hat{\Gamma}_\beta$, we are simply regressing on managed portfolios to find the factors. Now, substituting this into the objective minimization problem (3.1), we have

$$\hat{\Gamma}_\beta = \arg\min \mathrm{Tr}\left(\sum_{t=1}^{T-1}(\Gamma_\beta'\Gamma_\beta)^{-1}\Gamma_\beta' X_t X_t'\Gamma_\beta\right)$$

Note that the solution to this is exactly choosing $\hat{\Gamma}_\beta$ to be formed by stacking the eigenvectors corresponding to the largest K eigenvalues of the covariance matrix $\sum_{t=1}^{T} X_t X_t'$. In other words, we are indeed performing the standard static PCA approach to factor modelling with a fixed number of factors K.

3.1.2 Alternating Least Square

To have any meaningful improvement over doing static PCA on managed portfolios, we have to find a way to retain some dynamic structure in solving this problem. On this end, the author proposed to use numerical approximation via alternating least square (ALS).

The algorithm is as follows

Algorithm 1: Alternating Least Square

initialization: $\hat{\Gamma}_\beta$ as eigenvectors corresponding to K largest eigenvalues of $\sum_{t=1}^{T} X_t X_t'$.

while *not convergent* **do**

$$\hat{f}_{t+1} \leftarrow (\hat{\Gamma}_\beta' Z_t' Z_t \hat{\Gamma}_\beta)^{-1}\hat{\Gamma}_\beta' Z_t' r_{t+1} \tag{3.4}$$

$$\mathrm{vec}(\hat{\Gamma}_\beta) \leftarrow \left(\sum_{t=1}^{T-1} Z_t'Z_t \otimes \hat{f}_{t+1}\hat{f}_{t+1}'\right)\left(\sum_{t=1}^{T-1}[Z_t \otimes \hat{f}_{t+1}']' r_{t+1}\right) \tag{3.5}$$

end

There is a crucial assumption here: the term $Z_t'Z_t$ is not too volatile. That is, we are asking the covariance of characteristics of stocks to stay more or less constant. This is a needed assumption because it ensures that the initialization $\hat{\Gamma}_\beta$ (which is in fact the solution to the static PCA problem shown in section 3.1.1) is closed to the true value Γ_β. This slight relaxation that allows $Z_t'Z_t$ to fluctuate in a small neighbordhood around a constant value is the one edge that allows IPCA to outperform the standard static PCA on managed portfolio.

3.1.3 Identification

Since we have to estimate both Γ_β and f_{t+1} in our model (2.2), there arises an identification issue: given a $\hat{\Gamma}_\beta$ and $\hat{f}_{t+1}$, we can always rotate both into an equivalent pair of solution $\Gamma_\beta R$

and $R^{-1}\hat{f}_{t+1}$. In order to resolve this, we restrict to have $\hat{\Gamma}_\beta'\hat{\Gamma}_\beta = I_{K\times K}$. Further, we impose that the factor mean is nonnegative, and also sort them in descending order of variance.

3.2 Unrestricted Model: $\Gamma_\alpha \neq 0$

3.2.1 Adapting ALS

The unrestricted model does not eliminate the possibility that the alpha's of stocks $\alpha_{i,t}$ in (2.1) might also depend (here modeled as linearly) on characteristics. Rewriting (2.1) explicitly in terms of Γ_α and in vector form, we get

$$r_{t+1} = Z_t\Gamma_\alpha + Z_t\Gamma_\beta f_{t+1} + \epsilon^*_{t+1}.$$

Thus, letting $\tilde{\Gamma} = [\Gamma_\alpha, \Gamma_\beta]$, and $\tilde{f}_t = [1, f_t]^T$, we can rewrite this in a form similar to equation (2.2) in the restricted case:

$$r_{t+1} = Z_t\tilde{\Gamma}\tilde{f}_{t+1} + \epsilon^*_{t+1}.$$

Hence, we have that the system of equations (3.2) and (3.3) still hold but with $\hat{\Gamma}_\beta$ replaced by $\tilde{\Gamma}$, and $\hat{f}_{t+1}$ replaced by $\tilde{f}_{t+1}$. We use this to backout

$$f_{t+1} = (\hat{\Gamma}_\beta' Z_t' Z_t \hat{\Gamma}_\beta)^{-1}\hat{\Gamma}_\beta' Z_t' \left(r_{t+1} - Z_t\hat{\Gamma}_\alpha\right).$$

Thus, the unrestricted model seeks to optimally allocate variation in returns to stocks' alphas, and then perform cross-sectional linear regression of returns excess of alphas on $Z_t\Gamma_\beta$, i.e. the systematic, characteristic-driven component of the loadings $\beta_{i,t}$.

With this setup, we can employ the methods of the previous section on the restricted model (either alternating least square or PCA on managed portfolios), with an additional step to back out the factors and Γ_α, Γ_β.

3.2.2 Identification

In addition to the previous identification restrictions we imposed on $\hat{\Gamma}_\beta$ and $\hat{f}_{t+1}$, we need to impose further conditions as we now have to estimate a third parameters. In particular, note that from a triple of solution $(\hat{\Gamma}_\alpha, \hat{\Gamma}_\beta, \hat{f}_{t+1})$, we can use a constant vector h to get to an equivalent triple of solution $(\hat{\Gamma}_\alpha - \hat{\Gamma}_\beta h, \hat{\Gamma}_\beta, \hat{f}_{t+1} + h)$. To resolve this, we impose that $\hat{\Gamma}_\alpha'\hat{\Gamma}_\beta = 0_{1\times K}$. This can be done by regressing Γ_α on Γ_β, and choose the orthogonal residual as the estimate for Γ_α.

4 Empirical Results

4.1 Data

We use data from Freyberger, Neuhierl, Weber (2017). There are a total of 7593 unique stocks from 07/1962 to 05/2014. However, there are, on average, only about 2000 stocks per

month. The beginning period from 1962 to 1971 has a lot of stock instability, so we only use data from 01/1971 to 05/2014. The number of stocks for each date is shown in Figure 1.

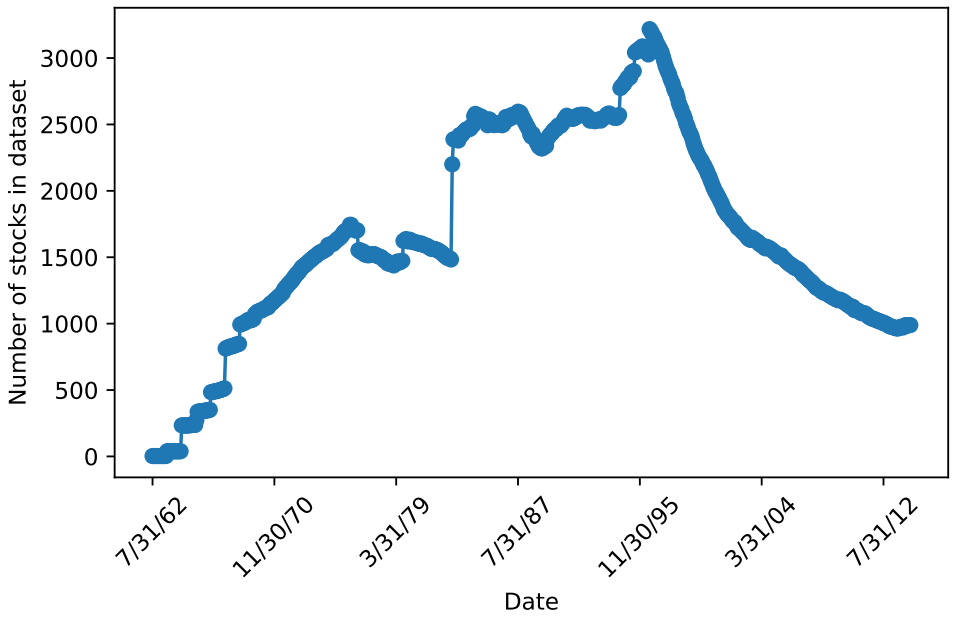

Figure 1: Number of stocks for all dates

We use 36 characteristics: market beta (beta), assets-to-market (`a2me`), total assets (`assets`), sales-to-assets (`ato`), book-to-market (`bm`), cash-to-short-term-investment (`c`), capital turnover (`cto`), capital intensity (`d2a`), ratio of change in PP&E to change in total assets (`dpi2a`), earnings-to-price (`e2p`), fixed costs-to-sales (`fc2y`), cash flow-to-book (`freecf`), idiosyncratic volatility with respect to the FF3 model (`idiovol`), investment (`invest`), leverage (`lev`), market capitalization (`mktcap`), turnover (`turn`), net operating assets (`noa`), operating accruals (`oa`), operating leverage (`ol`), price-to-cost margin (`pcm`), profit margin (`pm`), gross profitability (`prof`), Tobins Q (`q`), price relative to its 52-week high (`w52h`), return on net operating assets (`rna`), return on assets (`roa`), return on equity (`roe`), momentum (`cum-return-12-2`), intermediate momentum (`cum-return-12-7`), short-term reversal (`cum-return-1-0`), long-term reversal (`cum-return-36-13`), sales-to-price (`s2p`), SG&A-to-sales (`sga2s`), bid-ask spread (`bidask`), and unexplained volume (`suv`).

4.2 Results

4.2.1 Performance Measure

Kelly, Pruitt, Su (2019) suggested using two types of R^2 as their performance measure. The first one is called total R^2 and is defined by

$$\text{Total } R^2 = 1 - \frac{\sum_{i,t}(r_{i,t+1} - z'_{i,t}(\hat{\Gamma}_\alpha + \hat{\Gamma}_\beta \hat{f}_{t+1}))^2}{\sum_{i,t} r^2_{i,t+1}}.$$

The predictive R^2 is defined by

$$\text{Predictive } R^2 = 1 - \frac{\sum_{i,t}(r_{i,t+1} - z'_{i,t}(\hat{\Gamma}_\alpha + \hat{\Gamma}_\beta \hat{\lambda}))^2}{\sum_{i,t} r^2_{i,t+1}}.$$

The total R^2 measures the fraction of variance that is explained by the model when using realized factors. On the other hand, the predictive R^2 is the fraction of variance explained by the model using conditional expected factors.

4.2.2 In-sample R^2

We report our findings of in-sample R^2's in Table 1. This is done for both restricted and unrestricted models with the number of factors K ranging from 1 to 5.

		$K=1$	$K=2$	$K=3$	$K=4$	$K=5$
Total	$\Gamma_\alpha = 0$	0.00105	0.00181	0.00251	0.00316	0.00379
	$\Gamma_\alpha \neq 0$	0.00109	0.00186	0.00254	0.00320	0.00382
Pred.	$\Gamma_\alpha = 0$	-9e-6	-7e-6	-9e-6	-1e-5	-1e-5
	$\Gamma_\alpha \neq 0$	4e-5	4e-5	3e-5	4e-5	2e-5

Table 1: Total and predictive R^2 for restricted and unrestricted models with $K = 1, 2, 3, 4, 5$

The original paper suggested that the total R^2 is on the order of 10^{-1}, and the predictive R^2 is on the order of 10^{-3}. However, our results reported in table 1 are two orders of magnitude smaller.

4.2.3 Out-of-Sample R^2

We also computed out-of-sample R^2 using their suggested procedure for the case where $\Gamma_\alpha = 0$. In particular, we use a rolling window of 120 months. That is, given information of the previous 120 months, estimate $\hat{\Gamma}_\beta$ and also factors f_{t-119} to f_t. Then, the total R^2 is computed using

$$\text{Total } R^2 = 1 - \frac{\sum_{i,t}(r_{i,t+1} - z'_{i,t}(\hat{\Gamma}_\alpha + \hat{\Gamma}_\beta \hat{f}_{t+1,t})}{\sum_{i,t} r^2_{i,t+1}}.$$

Here, $\hat{f}_{t+1,t} = (\hat{\Gamma}'_\beta Z'_t Z_t \hat{\Gamma}_\beta)^{-1} \hat{\Gamma}_\beta Z'_t r_{t+1}$ is the realized factor given the estimation of $\hat{\Gamma}_\beta$ from historical value in the rolling window and observation of z_{t+1}. On the other hand, predictive R^2 is computed with $\hat{\lambda}$ being the average of f_{t_j}, with $t - 119 \leq t_j \leq t$.

	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$
Total R^2	0.0004	0.0007	0.0013	0.0015	0.0021
Pred. R^2	-0.0002	-0.0004	-0.0007	-0.0009	-0.0015

Table 2: Out of sample R^2 for restricted model with $K = 1, 2, 3, 4, 5$

Again, out of sample R^2 is lower than what the paper suggested, with an order of magnitude difference.

4.2.4 Out-of-Sample Sharpe Ratio

Next, we computed the Sharpe Ratio of the restricted model for different numbers of factors K and compare that with the Sharpe ratio gotten from doing only PCA. Surprisingly, the Sharpe ratio from IPCA dominated that of PCA for $K = 1$ through 4, while the Sharpe ratio of PCA is higher for $K = 5$. This also differs from the original paper where IPCA dominates overall. Note that the paper reported Sharpe ratio for unrestricted model.

	$K = 1$	$K = 2$	$K = 3$	$K = 4$	$K = 5$
IPCA	0.223	0.185	0.265	0.291	0.316
PCA	0.131	0.141	0.252	0.279	0.328

Table 3: Sharpe Ratio of Restricted IPCA vs PCA

4.2.5 In-sample Factor Decomposition

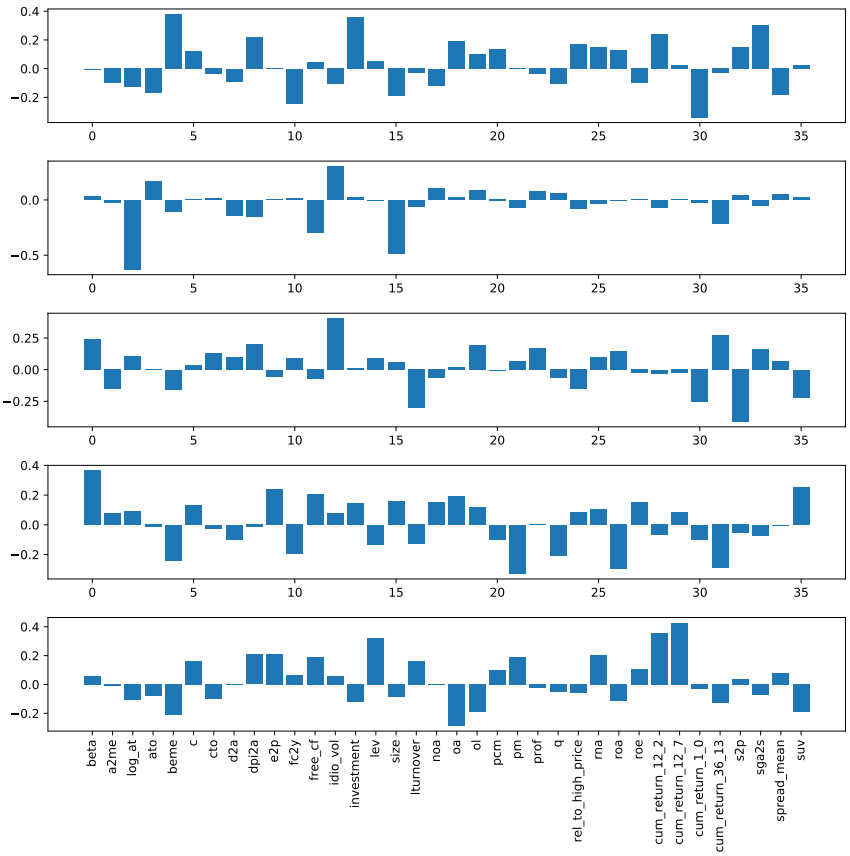

Figure 2: Factor loading dependence on characteristics

The first factor mainly goes long in book to market ratio, investment (growth rate in total assets), and short on short term reversals. The second factor trades mainly size and assets. The third factor trades on volatility and reversals. The fourth has no clear trend but

trades largely on beta of stocks. The last factor trades mainly on momentum (denoted as `cum-return-12-2`) and intermediate momentum (`cum-return-12-7`).

4.2.6 Algorithm Runtime

As can be seen from Figure 3, the alternating least square converges quickly to within the tolerance of 10^{-5} in about 70 iterations. Typical convergence takes more than 10 iterations as claimed in the original paper.

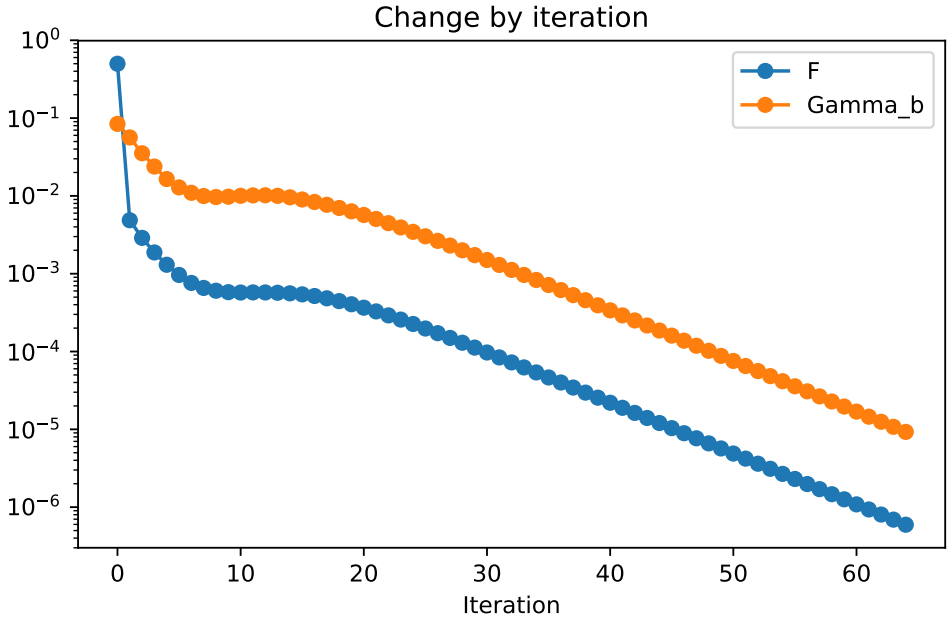

Figure 3: Change in F and Γ_β per iteration

4.3 Explore $Z_t' Z_t$

In order to understand the difference in performance between the ALS algorithm in dynamic IPCA and the static PCA on managed portfolio approach, we look more closely at the characteristic covariance matrix $Z_t' Z_t$. In particular, we looked at the singular values of $Z_t' Z_t$ shown in Figure Figure 4.

Contrary to the assumption in static PCA that $Z_t' Z_t$ is approximately constant, there is a lot of volatility, especially in the beginning period. The largest 5 singular values steadily increase after 1993.

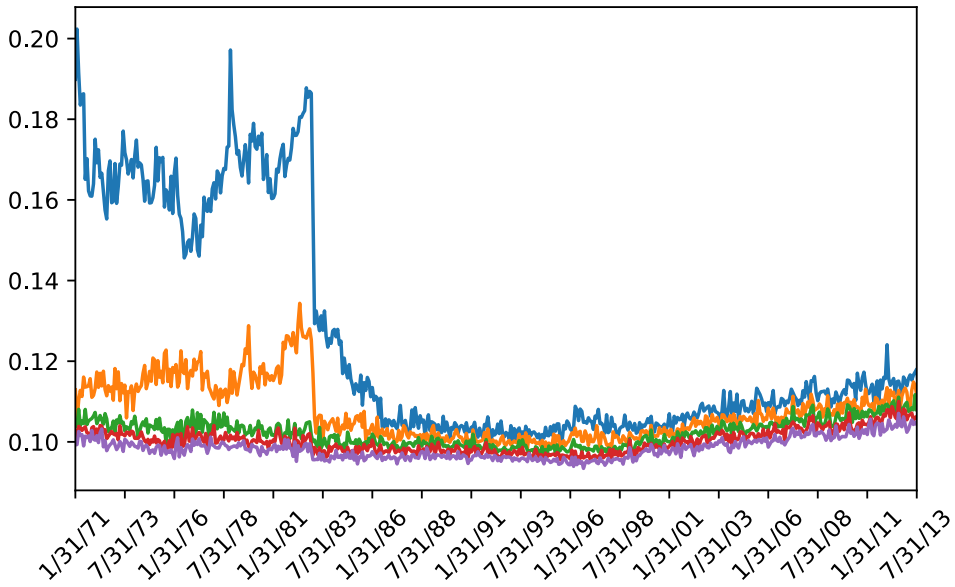

Figure 4: Largest 5 singular values of $Z_t'Z_t$

We also created a animated heatmap for viewing the change in characteristics correlation through time as reflected in the changing matrix $Z_t'Z_t$. The animated heatmap in Figure 5 can be viewed using Adobe Reader.

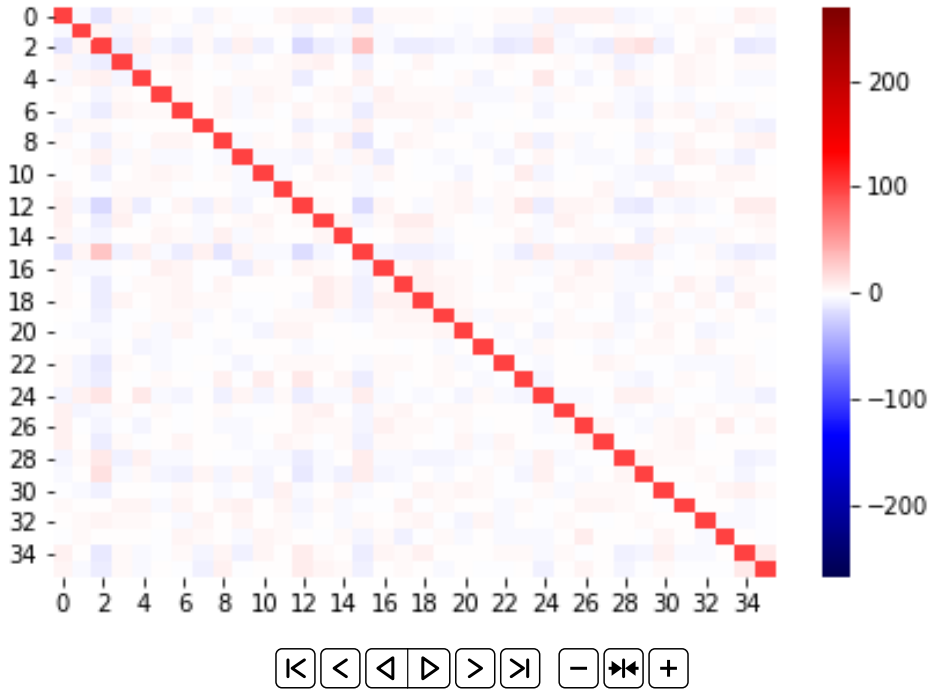

Figure 5: Heatmap showing characteristic correlation reflected by $Z_t'Z_t$

5 Conclusion

Despite our best efforts to ensure that our procedure follows the paper's closely, the results we receive are very different from what the papers reported, from runtime to R^2 to Sharpe ratio. We have double-checked our codes many times through a simulation where we use fixed, known Γ_β and fixed, known times series for f_{t+1}, but with the Z_t we extracted from the data. The simulation yields the expected results, which is a sanity check for our code. However, the discrepancies when performing on real data are not resolved.

Given more time, it will also be interesting to explore the characteristic covariance matrix $Z_t' Z_t$ in more details because this is the main difference between the IPCA approach and the static PCA on managed portfolio approach.

References

[1] Jianqing Fan, Yuan Liao, and Weichen Wang. Projected principal component analysis in factor models. 2017.

[2] Bryan Kelly, Seth Pruitt, and Yinan Su. Characteristics are covariances: A unified model of risk and return. 2019.

[3] Martin Lettau and Markus Pelger. Factors that fit the time series and cross-section of stock returns. 2019.