The Review of Financial Studies

Factors That Fit the Time Series and Cross-Section of Stock Returns*

Martin Lettau
Haas School of Business, University of California at Berkeley, NBER and CEPR

Markus Pelger
Department of Management Science & Engineering, Stanford University

We propose a new method for estimating latent asset pricing factors that fit the time series and cross-section of expected returns. Our estimator generalizes principal component analysis (PCA) by including a penalty on the pricing error in expected returns. Our approach finds weak factors with high Sharpe ratios that PCA cannot detect. We discover five factors with economic meaning that explain well the cross-section and time series of characteristic-sorted portfolio returns. The out-of-sample maximum Sharpe ratio of our factors is twice as large as with PCA with substantially smaller pricing errors. Our factors imply that a significant amount of characteristic information is redundant. (*JEL* C14, C52, C58, G12)

Received August 21, 2018; editorial decision January 29, 2020 by Editor Stijn Van Nieuwerburgh. Authors have furnished an Internet Appendix, which is available on the Oxford University Press Web site next to the link to the final published paper online.

The fundamental insight of asset pricing theory is that the cross-section of expected returns should be explained by exposure to systematic risk factors. Finding the "right" factors has become the central question of asset pricing. Harvey, Liu, and Zhu (2016) document that more than 300 published candidate factors have predictive power for the cross-section of expected returns. As argued by Cochrane (2011), this "factor zoo" leads to the question of which risk factors are important and which factors are subsumed by others.

This paper develops a new statistical method to find the most important factors for explaining asset returns and bringing order to the chaos of

*We thank Svetlana Bryzgalova, John Cochrane, Jianqing Fan, Kay Giesecke, Bob Hodrick, Serhiy Kozak (discussant), Per Mykland, Serena Ng, Shrihari Santosh, Yinan Su (discussant), Viktor Todorov, Dacheng Xiu and seminar participants at Columbia, Chicago, Stanford, UC Berkeley, Harvard, MIT, Zürich, Toronto, Boston University, Humboldt University, Frankfurt, Ulm, Bonn and the conference participants at the NBER Asset Pricing Conference, SFS Cavalcade, AFA Annual Meeting, NBER-NSF time series Conference, NBER-NSF SBIES, SoFiE, SITE, GEA, Western Mathematical Finance Conference and INFORMS. Send correspondence to Markus Pelger, Stanford University, Department of Management Science & Engineering, 475 Via Ortega, Stanford, CA 94305; telephone: +1(510) 859-5151. E-mail: mpelger@stanford.edu.

The Review of Financial Studies 33 (2020) 2274–2325

doi:10.1093/rfs/hhaa020 Advance Access publication March 17, 2020

factors. Our methodology uses large financial data sets to identify factors that simultaneously explain the time series and cross-section of stock returns. The estimation approach is a generalization of the widely utilized principal component analysis (PCA), for example, in Connor and Korajczyk (1986, 1988). Statistical factor analysis based on PCA extracts factors that capture comovement but does not incorporate any information contained in the means of the data. Therefore, it is not surprising that PCA factors do not capture the differences in mean risk premia of assets. We propose an alternative estimator, risk-premium PCA (RP-PCA), that incorporates information in the first and second moments of the data yielding a more efficient estimator than standard PCA. The risk-premium PCA estimator can be interpreted as a generalized PCA with an additional penalty term that accounts for cross-sectional pricing errors, thus combining PCA factor analysis with the arbitrage pricing theory (APT) of Ross (1976). The objective of finding factors that can explain comovement and the cross-section of expected returns simultaneously is based on fundamental insights of APT: Systematic time-series factors also determine cross-sectional risk premia. The RP-PCA exploits this connection explicitly.[1]

This paper focuses on the empirical estimation using RP-PCA, while the econometric asymptotic theory for the estimator is developed in Lettau and Pelger (Forthcoming). That paper shows that RP-PCA dominates conventional estimation based on PCA along a number of dimensions, especially when factors are “weak.” We define “weak” factors as factors that affect only a subset of the underlying assets. Weak factors are harder to detect than “strong” factors that affect all assets (such as the “market” factor). Many anomaly-based factors are more likely to be “weak” factors, and RP-PCA can find weak factors with high Sharpe ratios, which cannot be detected with PCA, even if an infinite amount of data is available. Lettau and Pelger (Forthcoming) show theoretically that the RP-PCA estimator is (asymptotically) more efficient than standard PCA in the sense that the stochastic discount factor (SDF) and factors estimated by RP-PCA are more highly correlated with the true SDF and factors than those estimated by PCA.

Our empirical analysis uses excess returns of smaller cross-sections of double-sorted portfolios as well as a large cross-section of single-sorted decile portfolios constructed from 37 anomaly characteristics. The empirical findings can be summarized as follows. First, PCA is not a reliable method to estimate latent asset pricing factors and is dominated by RP-PCA. We show that even for 25 double-sorted portfolios that follow a clear factor structure, PCA can fail to detect the underlying factor structure, while RP-PCA reliably finds all relevant asset-pricing factors. Second, we show that a small number of factors is sufficient to fit the first and second moments of the 370 anomaly portfolios.

[1] Koijen, Lustig, and Van Nieuwerburgh (2017) study the relationship of time-series factors constructed from bond yields and the cross-section of equity returns. They find that differences in the exposure to bond factors across value-growth portfolios are consistent with differences in risk premia.

The RP-PCA method extracts five significant factors that together yield a high Sharpe ratio (SR), small pricing errors, and capture most of the time-series variation in the data. The first factor is long-only in all portfolios and is highly correlated with the CRSP-VW return. Two additional factors capture time-series variation but play no role in the cross-section of returns or the Sharpe ratio of the implied SDF. The remaining two factors are relevant for the cross-section and SR but are less critical for the time-series variation. Hence, an RP-PCA model with three factors captures the cross-sectional differences in expected returns of 37 decile portfolios, while a model with three (different) factors captures 85% to 90% of the time-series variation. All results hold in-sample as well as out-of-sample, suggesting that RP-PCA is stable and robust. These results show the importance of using not only information in second moments (as in standard PCA) but also information in the means of the data when constructing factor models. The RP-PCA estimator achieves this parsimoniously and efficiently without adding any computational burden.

An analysis of the composition of factors and the implied SDF in terms of their portfolio weights shows that RP-PCA and PCA factors differ significantly. The exception is the first factors, which are long-only factors and highly correlated with the market return. Factors estimated by RP-PCA fall into two categories: Factors that capture comovement and factors with high Sharpe ratios. The weights of portfolios in RP-PCA factors with high Sharpe ratios are highly correlated with the mean return of the portfolios. In other words, factors that capture the cross-section are composed mostly of portfolios with either high returns (and positive weights) or low returns (and negative weights). Moreover, the SDF that is implied by RP-PCA factors shows the same tilt toward portfolios with high or low returns. In contrast, the structure of PCA factors and the implied SDF is less clear. In particular, their portfolio weights are only weakly related to mean returns.

We also estimate RP-PCA and PCA models using a sample of 270 large individual stocks. Both models fit reasonably well in-sample, but their out-of-sample fit is poor. The reason is that the underlying factor structure is more unstable than that of portfolios suggesting that the assumption of a constant factor structure is not appropriate for individual stock data.[2]

Our paper contributes to an emerging literature that uses new econometric techniques in asset pricing for high-dimensional data, including machine-learning methods. Kozak, Nagel, and Santosh (Forthcoming) (KNS) also exploit economic restrictions relating expected returns to the covariance to extract the factors spanning the stochastic discount factor. Their estimator is based on an elastic net that shrinks the contributions of low-variance principal components of candidate factors. Their approach selects factors that explain variation and mean returns, based on a similar insight as RP-PCA.

[2] Pelger (Forthcoming) estimates factor models with high-frequency data on individual stock returns and confirms that the factor loadings of individual stock returns are time-varying.

Kelly, Pruitt, and Su (2017) introduce instrumented PCA (IPCA) to perform dimensionality reduction of the characteristic space. Their work is closely related to the projected PCA of Fan, Liao, and Wang (2016) that allows for time-varying factor loadings.[3] Both methods first project the individual stock returns into managed portfolios based on observed characteristics and then apply PCA to the projected data.

Both methods involve a computation of factors using PCA but it is straightforward to use RP-PCA instead to extract factors. We compare the fit of the KNS and IPCA models combined with RP-PCA to the PCA specification in the original papers. Replacing PCA with RP-PCA yields higher Sharpe ratios and lower pricing errors in both models. Thus, RP-PCA is not an alternative to the KNS and IPCA methods but can easily be used in combination with both estimators to further improve the fit. RP-PCA estimation can also be combined with other PCA-based methods, e.g. the three-pass model in Giglio and Xiu (2019).

Our paper is also related to recent applications of machine-learning techniques in asset pricing. For example, DeMiguel et al. (2017), Freyberger, Neuhierl, and Weber (2020) and Feng, Giglio, and Xiu (Forthcoming) employ factor selection with Lasso-style L^1-norm penalties. These papers have in common that they assume that the stochastic discount factor has a sparse exposure to the characteristics. In contrast, we assume a low-dimensional factor structure without imposing any sparsity constraints on how the characteristics can affect the factors and, hence, the stochastic discount factor. Our work is part of the emerging econometrics literature that combines latent factor extraction with a form of regularization. Bai and Ng (2019) develop the statistical theory for robust principal components. Their estimator can be understood as performing iterative ridge instead of least squares regressions, which shrinks the eigenvalues of the common components to zero. They combine their shrinked estimates with a clean-up step that sets the small eigenvalues to zero. Their estimates have less variation at the cost of a bias. Our approach also includes a penalty that in contrast is based on economic information and does not create a bias-variance trade-off. Our work is related to that of Daniel, Mota, Rottke, and Santos (2020). Under the assumption of priced and unpriced systematic factors, they create characteristic efficient portfolios that are only exposed to the priced risk and line up with characteristics. Similarly, our focus is as well the estimation of the priced risk factors.

1. The Risk-Premium PCA Estimator

1.1 Methodology

We assume that excess returns follow a standard approximate factor model and the assumptions of the arbitrage pricing theory are satisfied. This means

[3] A related approach is that of Fama and French (2020), who extract factors by cross-sectional regressions on prespecified characteristics and model time-varying loadings through time-varying characteristics.

that excess returns of an asset n, X_{nt}, have a systematic component captured by K factors and a nonsystematic, idiosyncratic component capturing asset-specific risk. The approximate factor structure allows the nonsystematic risk to be weakly dependent. We observe the excess return of N assets over T time periods:

$$X_{nt} = \mathrm{F}_t \Lambda_n^\top + e_{nt} \qquad n=1,...,N,\ t=1,...,T \tag{1}$$

$$\iff \underbrace{\mathrm{X}}_{T\times N} = \underbrace{\mathrm{F}}_{T\times K} \underbrace{\Lambda^\top}_{K\times N} + \underbrace{\mathrm{e}}_{T\times N}, \tag{2}$$

where the unknown latent factors F and loadings (or betas) Λ have to be estimated. $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_F$ are the variance-covariance matrices of returns and factors, respectively, and $\boldsymbol{\Sigma}_e$ is the variance-covariance matrix of e. We will work in a large dimensional panel—that is, the number of cross-sectional observations N and the number of time-series observations T are both large, and we study the asymptotics as they jointly go to infinity, but N/T converges to a finite limit. Under the assumption that the factors and residuals are uncorrelated, the covariance matrix of the returns consists of a systematic and idiosyncratic part:

$$\mathrm{Var(X)} = \Lambda \mathrm{Var(F)} \Lambda^\top + \mathrm{Var(e)}.$$

Since the largest eigenvalues of Var(X) are driven by the factors, principal component analysis (PCA) can be used to estimate the loadings and factors. Note that standard PCA estimators of latent factors use the information contained in the second moments but ignore information that is contained in the first moment.

Factor models, for example, Connor and Korajczyk (1988, 1993), Bai (2003), Bai and Ng (2002), and Stock and Watson (2002), typically assume that the mean of the data matrix X is equal to zero. As we will see later, this assumption is restrictive if the means contain information about the factor structure, as is the case in applications to financial data. The RP-PCA estimator exploits this information, so it is crucial to allow the means to be unrestricted. We therefore allow that $\overline{\mathrm{X}} = \mathrm{E[X]} \neq 0$ and $\overline{\mathrm{F}} = \mathrm{E[F]} \neq 0$.

If X contains only excess returns, the role of means is explicitly given by Ross' arbitrage pricing theory (APT), which implies that expected excess returns are explained by the exposure to the risk factors multiplied by the risk premium of the factors. If the factors are excess returns, the APT implies[4]

$$\mathrm{E}[\mathrm{X}_n] = \Lambda_n \mathrm{E[F]}.$$

Factors identified by standard PCA explain as much time variation as possible. Conventional statistical factor analysis applies PCA to the sample covariance

4 We assume a strong form of APT, where residual risk has a risk premium of zero. In its more general form APT requires only the risk premium of the idiosyncratic part of well-diversified portfolios to go to zero. As most of our analysis will be based on portfolios, there is no loss of generality by assuming the strong form.

matrix $\frac{1}{T}\mathrm{X}^\top \mathrm{X}-\overline{\mathrm{X}}\,\overline{\mathrm{X}}^\top$ where $\overline{\mathrm{X}}$ denotes the sample mean of excess returns. The case when the test assets are normalized by their standard deviation, so that $\mathrm{diag}(\mathbf{\Sigma}_X)=1$, is equivalent to PCA applied to the correlation matrix.

Standard PCA analysis proceeds in two steps. First, the N test assets X are rotated using an eigendecomposition of the variance-covariance matrix $\mathbf{\Sigma}_X$ to obtain N orthogonal factors. The rotated factors are sorted according to their variances, which are equal to the corresponding eigenvalues. Second, the K factors with the K largest variances are chosen as estimates of the true factors F in Equation (2). Since factor loadings in Equation (2) are proportional to the eigenvectors associated with the K largest eigenvalues, loadings are estimated by the $K\times N$ matrix of eigenvectors defined as $\widehat{\Lambda}_{\mathrm{PCA}}$. The $T\times K$ matrix of rotated orthogonal factors is given by $\widehat{\mathrm{F}}_{\mathrm{PCA}}=\mathrm{X}\widehat{\Lambda}_{\mathrm{PCA}}\left(\widehat{\Lambda}_{\mathrm{PCA}}^\top\widehat{\Lambda}_{\mathrm{PCA}}\right)^{-1}$.

It is straightforward to express the PCA loadings and factors as solutions to the minimization of the objective function (Stock and Watson (2002)):

$$\text{PCA:}\qquad \widehat{\mathrm{F}}_{\mathrm{PCA}},\widehat{\Lambda}_{\mathrm{PCA}}=\underset{\Lambda,\mathrm{F}}{\operatorname{argmin}}\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\left((X_{nt}-\overline{\mathrm{X}}_n)-(\mathrm{F}_t-\overline{\mathrm{F}})\Lambda_n^\top\right)^2. \quad (3)$$

The objective function in Equation (3) shows that PCA loadings and factors do not depend on the means of the test assets. In practice, the data matrix X is usually demeaned before PCA is applied.[5]

Our risk-premium-PCA (RP-PCA) estimator modifies the objective function so that cross-sectional pricing errors are taken into account. The RP-PCA objective function minimizes a weighted average of the unexplained variation and cross-sectional pricing errors:

$$\text{RP}-\text{PCA:}\qquad \widehat{\mathrm{F}}_{\mathrm{RP}},\widehat{\Lambda}_{\mathrm{RP}}=\underset{\Lambda,\mathrm{F}}{\operatorname{argmin}}\ \underbrace{\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}(X_{nt}-\mathrm{F}_t\Lambda_n^\top)^2}_{\text{unexplained TS variation}}+\gamma\underbrace{\frac{1}{N}\sum_{n=1}^{N}\left(\overline{\mathrm{X}}_n-\overline{\mathrm{F}}\Lambda_n^\top\right)^2}_{\text{XS pricing error}}, \quad (4)$$

where $\gamma\geq -1$ is the weight of the average cross-sectional pricing error relative to the times-series error in standard PCA. It is straightforward to show that minimizing Equation (4) is equivalent to applying PCA to the matrix

$$\mathbf{\Sigma}_{\mathrm{RP}}=\frac{1}{T}\mathrm{X}^\top\mathrm{X}+\gamma\overline{\mathrm{X}}\,\overline{\mathrm{X}}^\top. \quad (5)$$

Note that $\mathbf{\Sigma}_{\mathrm{RP}}$ is equal to the variance-covariance matrix of X if $\gamma=-1$. Setting $\gamma=0$ implies that $\mathbf{\Sigma}_{\mathrm{RP}}$ is the second moment matrix of X and corresponds to applying PCA to the second moment matrix as opposed to the covariance

[5] An alternative is to apply PCA to the second moment matrix $\frac{1}{T}\mathrm{X}^\top\mathrm{X}$ instead of the variance-covariance matrix; see later for more details.

matrix. Thus, standard PCA using the variance-covariance matrix or the second-moment matrix is a special case of RP-PCA. RP-PCA with $\gamma > -1$ can be understood as PCA applied to a matrix that "overweights" the means. As in standard PCA, the eigenvectors of the K largest eigenvalues of $\mathbf{\Sigma}_{\mathrm{RP}}$ are proportional to the loadings $\widehat{\Lambda}_{\mathrm{RP}}$. In PCA, the eigenvalues are equal to factor variances, while eigenvalues in RP-PCA are equal to a more generalized notion of "signal strength" of a factor that is defined later. RP-PCA factors are estimated by a regression of the returns on the estimated loadings, that is, $\widehat{\mathrm{F}}_{\mathrm{RP}} = \mathrm{X}\widehat{\Lambda}_{\mathrm{RP}} \left(\widehat{\Lambda}_{\mathrm{RP}}^{\top}\widehat{\Lambda}_{\mathrm{RP}}\right)^{-1}$.[6]

For notational convenience, we use $\widehat{\Lambda}$ and $\widehat{\mathrm{F}}$ to denote the PCA/RP-PCA estimators of loadings and factors in the rest of the paper with the understanding that the estimates depend on the number of factors K and the choice of γ.

Given factors $\widehat{\mathrm{F}}$, we can compute the maximal Sharpe ratio from the tangency portfolio of the mean-variance frontier that is spanned by $\widehat{\mathrm{F}}$ as

$$\widehat{\mathrm{b}}_{\mathrm{MV}} = \mathbf{\Sigma}_F^{-1}\mu_F, \tag{6}$$

where μ_F and $\mathbf{\Sigma}_F$ are the mean and variance-covariance matrix of $\widehat{\mathrm{F}}$, respectively. If $\mathbf{\Sigma}_F$ is a diagonal matrix, then

$$\widehat{\mathrm{b}}_{\mathrm{MV},i} = \frac{\mu_{F,i}}{\sigma_{F,i}^2}. \tag{7}$$

The implied SDF is given by

$$M_t = 1 - \widehat{\mathrm{b}}_{\mathrm{MV}}^{\top}\left(\widehat{\mathrm{F}}_t - \mathrm{E}[\widehat{\mathrm{F}}_t]\right). \tag{8}$$

1.2 RP-PCA as ordinary least squares estimator

The RP-PCA loadings $\widehat{\Lambda}$ are related to ordinary least squares (OLS) betas of time-series regressions of excess returns on factors. Suppose that factors $\widehat{\mathrm{F}}$ are estimated by RP-PCA, and consider the time-series regression

$$X_{nt} = \widehat{\mathrm{F}}_t \mathrm{B}_n^{\top} + e_{nt}. \tag{9}$$

The factor model in Equation (2) implies that there is no intercept, so excluding a constant in the regression imposes this condition and, therefore, the mean of the regression errors e_{nt} is not necessarily zero. Alternatively, one can estimate

$$X_{nt} = \alpha_n + \widehat{\mathrm{F}}_t \mathrm{B}_n^{\top} + e_{nt}, \tag{10}$$

and evaluate the model by the magnitude of the pricing errors α_n.

[6] In latent factor models only the product $\mathrm{F}\Lambda^{\top}$ is identified. For any full rank $K \times K$ matrix H the factors $\mathrm{F}\mathrm{H}^{-1}$ and loadings $\Lambda\mathrm{H}^{\top}$ yield the same factor model. We use the standard convention to normalize the loadings $\Lambda^{\top}\Lambda/N = \mathrm{I}_K$ and assume that the factors are uncorrelated. That means that a factor with a variance of $\sigma_F^2 = 0.5$ could be interpreted as affecting only half of the assets with an average loading of 1. Alternatively, we could normalize the covariance matrix of the factors to an identity matrix, $\mathbf{\Sigma}_F = \mathrm{I}$, so that the norm of the loading vectors measures the strength of the factors.

The first part of the RP-PCA objective function in Equation (4) is the same as the OLS objective function of Equation (9), while the second part captures the difference between the mean excess returns $\overline{X}$ and the fitted model-implied mean $E[\widehat{F}_t]\widehat{B}_n^\top$. Consider first the case $\gamma=0$, so that the second part of Equation (4) drops out and the RP-PCA objective is identical to the OLS objective function. Thus, the OLS estimate $\widehat{B}_n$ in Equation (9) is equal to the RP-PCA estimator $\widehat{\Lambda}_n$. The case $\gamma=-1$ is similar but the regression in Equation (9) is applied to demeaned X_n and $\widehat{F}$, so that $E[e_{nt}]=0$. However, the cross-sectional pricing error $E[X_{nt}]-E[\widehat{F}_t]\widehat{B}_n^\top$ applies to data that is not demeaned and is not necessarily zero. As for $\gamma=0$, $\widehat{B}_n=\widehat{\Lambda}_n$.

A similar argument can be applied in general for any value of γ by incorporating the cross-sectional error directly into the regression (9). Define $\tilde{\gamma}=\sqrt{\gamma+1}-1$ and $\tilde{X}_{nt}=X_{nt}+\tilde{\gamma}\,\overline{X}_{nt}, \tilde{F}_{nt}=\widehat{F}_t+\tilde{\gamma}\,\overline{F}_t$ and specify the time-series regression in terms of $\tilde{X}_{nt}$ and $\tilde{F}_t$:

$$\tilde{X}_{nt}=\tilde{F}_t B_{\gamma,n}^\top+\tilde{e}_{nt}. \tag{11}$$

The OLS objective function in terms of $\tilde{X}_{nt}$ and $\tilde{F}_{nt}$ is equal to the RP-PCA objective in Equation (4), so that $\widehat{\Lambda}_n=\widehat{B}_{\gamma,n}$. In other words, $\widehat{\Lambda}$ can always be interpreted as regression coefficients that are based on a specification with transformed data.[7]

To properly address the pricing implication of a factor model, pricing errors are computed based on Equation (10) using X_n and F_t, without any mean adjustment. Hence, after computing RP-PCA factors $\widehat{F}$ (given γ), we use Equation (10) to estimate $\widehat{\alpha}_n, \widehat{B}_n$ and $\widehat{e}_n$ for each asset n. If $\gamma\neq-1,0$, the estimated $\widehat{B}_n$ differ from the RP-PCA $\widehat{\Lambda}_n$, but in in all data sets that we consider below, the difference is small and has no meaningful impact on the results. The overall performance of the model is evaluated by computing the root-mean-squared pricing error $\text{RMS}_\alpha=\sqrt{\widehat{\alpha}^\top\widehat{\alpha}/N}$, the magnitude of idiosyncratic variance, $\overline{\sigma}_e^2=\text{avg}(\text{Var}(\widehat{e}_n)/\text{Var}(X_n))$, and the maximal Sharpe ratio implied by $\widehat{F}$.

To summarize, the RP-PCA estimation involves the following steps:

1. Apply PCA to the matrix $\frac{1}{T}X^\top X+\gamma\overline{X}\overline{X}^\top$ to obtain $\widehat{\Lambda}$.
2. Construct factors $\widehat{F}=X\widehat{\Lambda}\left(\widehat{\Lambda}^\top\widehat{\Lambda}\right)^{-1}$.
3. Estimate $X_{nt}=\alpha_n+\widehat{F}_t B_n^\top+e_{nt}$, $n=1,\dots,N$ by OLS to obtain $\widehat{\alpha}, \widehat{B}$, and $\widehat{e}$.
4. Compute RMS_α and $\overline{\sigma}_e^2$.
5. Compute the maximum Sharpe ratio SR that can be obtained by estimated factors $\widehat{F}$.

[7] See Lettau and Pelger (Forthcoming) for details.

1.3 Properties of RP-PCA

Lettau and Pelger (Forthcoming) develop a formal statistical theory for the RP-PCA model and show that the asymptotic results hold in finite samples of the size typically encountered in practice. The advantage of RP-PCA compared with PC is that RP-PCA uses additional information contained in first moments, whereas PCA only uses information in the second moments. As we show in this section, the estimation of factors depends on the "signal strength" of factors relative to the variance of idiosyncratic "noise." Adding information in the mean increases the "signal strength" of factors and therefore enhances the "signal-to-noise" ratio, resulting in more efficient estimators. The rest of this section focuses on the intuition of the theoretical results. Appendix A presents a simplified version of the model in Lettau and Pelger (Forthcoming) and includes formal statements of the results described in this section.

The key question is whether given factors can be "detected" and how precisely factors and loading can be estimated. One important implication of the theory in Lettau and Pelger (Forthcoming) is that the properties of the RP-PCA (and standard PCA) estimators depend critically on the "signal strength" of the factors. "Strong" factors affect a large number of assets and/or have large variances. The market portfolio is an example of a strong factor. In contrast, "weak" factors affect only a small subset of assets or affect all assets only weakly.[8] Under a suitable normalization, the strength of a factor depends on the structure of the loadings (that is, betas) Λ. Formally, in a strong factor model $\Lambda^\top \Lambda / N \xrightarrow{p} \mathrm{I}_K$, while in a weak factor model $\Lambda^\top \Lambda \xrightarrow{p} \mathrm{I}_K$.[9] Therefore, the associated eigenvalues of Λ explode in strong factor models but are finite for weak factors.[10] This difference is crucial for the behavior of PCA-based estimators in factor models.

In practice, the spectrum of estimated eigenvalues can be used to discern how many factors are relevant and whether these factors are strong or weak (see Appendix A for more details). The eigenvalue spectrum in the equity return data used in this paper suggests a combination of strong and weak factors. The first eigenvalue of the sample covariance matrix is large, typically around ten times the size of the rest of the spectrum. The next eigenvalues usually stand out, but have magnitudes only around twice or three times the average of the residual spectrum, which would be more in line with a weak factor interpretation. The first statistical factor in our data sets is always strongly correlated with an equal-weighted market factor. Hence, if we are interested in learning more about factors besides the market, the weak factor model is likely to be appropriate.

---

8 Lettau and Pelger (Forthcoming) also consider models in which some factors are strong and others are weak.

9 One example of a strong factor is $\lambda_{nk} = N(1, 1/\sqrt{N})\, \forall n, k$, that is, the loading of each asset on each factor converge to 1 as N grows, and $\lambda_{nk} = N(0, 1/\sqrt{N})$ is an example of a weak factor.

10 The eigenvalues in strong factor models are finite when normalized by N.

Lettau and Pelger (Forthcoming) define a measure of the "signal strength" of a factor that depends on the eigenvalues of loadings matrix Λ and the factor means μ_F, as well as the RP-PCA parameter γ, since γ affects how efficiently the information contained in the factors is used. Whether a factor can be detected and estimated by RP-PCA depends on its strength and the variance of the "noise" that is due to the idiosyncratic component e in Equation (2). If the factor strength relative to the noise variance is below a threshold, then the factor cannot be detected by RP-PCA (or PCA). If the factor is strong enough and above the threshold, it can be detected and the correlation of the estimated factor with the true factor is strictly positive but potentially less than one. Given moments of the factors, an appropriate choice of γ can increase the strength of a factor, which makes detection more likely and estimates more precise compared with standard PCA with $\gamma = -1$.

Consider first the strong factor model, in which standard methods apply. Factors and loadings are estimated consistently, and loadings estimates are asymptotically normal, irrespective of the choice of γ. The optimal γ that minimizes the asymptotic variance is $\gamma = 0$, but the efficiency gain is relatively small. See Proposition 1 in Appendix A for a formal statement. In other words, strong factors can be reliably estimated with PCA-based methods.

The case of weak factors is more interesting. First, while standard statistical theory applies to strong factors, it breaks down for weak factors. In addition, allowing for nonzero factor means introduces further complications. Lettau and Pelger (Forthcoming) generalize the spiked covariance models from random matrix theory and properties in Onatski (2012) and Benaych-Georges and Nadakuditi (2011) to analyze the asymptotic behavior of the RP-PCA estimator for weak factors. The results can be summarized as follows (see Proposition 2 in Appendix A for details):

1. Weak factors can only be estimated with a bias even as N and T grow.
2. If a factor is too weak, that is, its signal strength is below a threshold (that is independent of γ), then it cannot be detected at all.
3. The signal strength is increasing in γ, so that RP-PCA is more likely to detect factors with lower signal strength than PCA with $\gamma = -1$.
4. The correlations of estimated factors with true factors depend on γ. Typically, there is a finite γ that maximizes the correlation of estimated factors with true factors.
5. In particular, RP-PCA with $\gamma > -1$ yields estimated factors that are more highly correlated with true factors than PCA with $\gamma = -1$. In this sense, the RP-PCA estimator is more efficient than the PCA estimator.

These results show that that RP-PCA dominates standard PCA in terms of detection and estimation of weak factors. How should the optimal γ be chosen? Based on our asymptotic theory, there are two possible criteria. First, a very large γ maximizes the probability of detecting a weak factor but lowers the correlation with the true factor. Alternatively, a moderately high γ maximizes this correlation. Another aspect of choosing γ is the out-of-sample performance of the RP-PCA estimator. As we will see later, RP-PCA estimation deteriorates out-of-sample, if γ is chosen too high. In practice, we start with $\gamma = -1$ and then increase γ until the OOS fit is no longer improved.

Appendix B reports results from a simulation study that is designed to match the moments of the data used below. As an illustration of how γ affects the estimation, consider Figure 1. The figure shows the correlation of the estimated and true factors for parameters used in the simulation exercise described in the Appendix. The variance of the factor in the top panel is set to 0.05 and to 0.1 in the bottom panel. Ceteris paribus, a higher factor variance increases the signal strength of the factor. We also consider three positive values for the Sharpe ratio (or, equivalently, the mean) of the factor, as well as a case with a factor of mean 0. The vertical lines indicate the γ that maximizes the correlation. In all cases with positive Sharpe ratios, the correlation increases sharply with γ for small γ and then flattens out as γ increases. For a fixed factor variance, the gains in the correlation of the estimated and true factors increase with the Sharpe ratio (and mean) of the factor. Comparing the two panels shows that the optimal γ is higher if the signal strength of the factor is lower (that is, the factor variance decreases) and if the Sharpe ratio increases. If the Sharpe ratio (or the mean) of a factor is zero, then the correlation is decreasing in γ and the optimal γ is equal to -1. The reason is that the factor mean does not contain any useful information, so a higher γ increases the weight on the uninformative cross-sectional errors. Hence, in this case RP-PCA performs strictly worse than PCA.

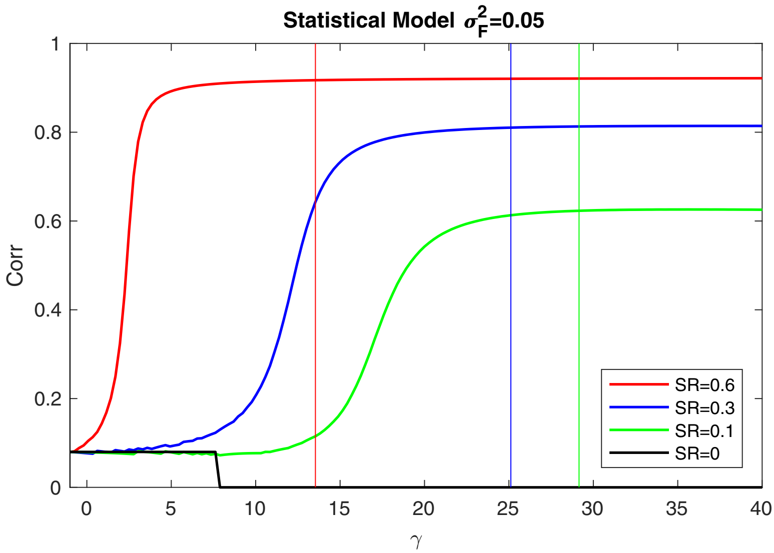
Finally, we note that standard tests of the null hypotheses that the pricing errors are jointly zero are invalid. The reason is that the asymptotic theory developed in Lettau and Pelger (Forthcoming) is based on $N, T \to \infty, N/T \to c$ while GRS, Wald and Fama-MacBeth-type tests are based on the assumption that N is constant and $T \to \infty$. For fixed N and $T \to \infty$ the inverse of the sample covariance matrix converges to the population matrix. This is no longer the case of $N, T \to \infty$. In fact, the inverse of the sample covariance matrix converges to a matrix that is different from the population counterpart. Hence, the test statistics are biased even in large samples.

1.4 Robust SDF estimation

Kozak, Nagel, and Santosh (Forthcoming) (KNS) suggest an alternative to the SDF weights in (6). The first step of their procedure is a rotation of the matrix of test assets. KNS apply the standard PCA rotation but a rotation based on RP-PCA can also be used. We will compare both cases in the next section. The $T \times N$ matrix $\tilde{\mathrm{F}}$ denotes the matrix of all N rotated factors with mean and variance-covariance matrix $\mu_{\tilde{F}}$ and $\mathbf{\Sigma}_{\tilde{F}}$, respectively.

The objective of KNS is a robust estimation of the SDF and is based on two principles. First, the SDF representation should be sparse in the sense that

A Low factor variance



B High factor variance

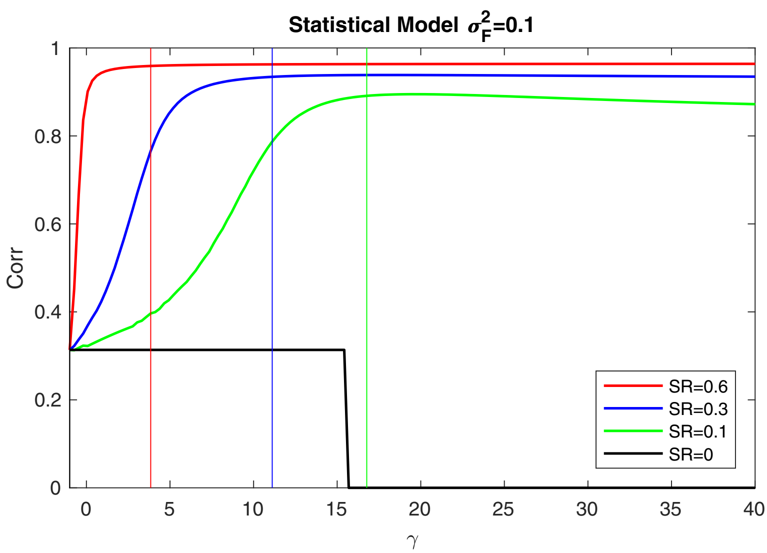


Figure 1
Weak factors: Correlation of estimated factor with the true factor
This plot shows the correlation of the estimated factor with the true factor as a function of γ for different parameter settings implied by the theoretical results of the weak factor model ($N=370$ and $T=650$).

only K SDF weights should be nonzero, which is equivalent to choosing the K factors with the largest eigenvalues in PCA and RP-PCA. Second, the SDF weights are shrinked toward zero compared with the mean-variance weights in (6). The KNS estimator can be written as

$$\widehat{\mathrm{b}}_{\mathrm{KNS}}=\arg\min_{\mathrm{b}}\left[\left(\mu_{\tilde{F}}-\boldsymbol{\Sigma}_{\tilde{F}}\mathrm{b}\right)^{\top}\boldsymbol{\Sigma}_{\tilde{F}}^{-1}\left(\mu_{\tilde{F}}-\boldsymbol{\Sigma}_{\tilde{F}}\mathrm{b}\right)+2\nu_1\sum_{i=1}^{N}|b_i|+\nu_2\mathrm{b}^{\top}\mathrm{b}\right], \quad (12)$$

where ν_1 and ν_2 are two penalty parameters. KNS solve Equation (12) numerically, but a closed-form solution of $\widehat{\mathrm{b}}_{\mathrm{KNS}}$ exists if $\boldsymbol{\Sigma}_{\tilde{F}}$ is a diagonal matrix, which is the case in KNS since $\tilde{\mathrm{F}}$ consists of PCA factors: [11]

$$\widehat{b}_{\mathrm{KNS},i}=\begin{cases}\dfrac{\mu_{\tilde{F},i}-\nu_1}{\sigma^2_{\tilde{F},i}+\nu_2} & \text{if } \mu_{\tilde{F},i}\geq\nu_1,\\ 0 & \text{if } \mu_{\tilde{F},i}<\nu_1.\end{cases} \quad (13)$$

The closed-form solution in Equation (13) illustrates the differences between KNS and PCA. First, consider the role of ν_1. Only factors with means larger than ν_1 have a nonzero weight in the SDF. Therefore, ν_1 controls how many factors have nonzero weight and can be interpreted as a sparsity parameter. KNS consider models with different number of factors and choose ν_1 such that the implied SDF has $K=1,2,3...$ factors. Since there is a one-to-one mapping between K and ν_1, sparsity in KNS is equivalent to choosing the number of factors in PCA or RP-PCA. The second difference of KNS compared to PCA is that the factors are ranked and chosen according to their means instead of their variances. Only factors with a mean that is larger than ν_1 are included in the SDF; others have SDF weights of zero. Hence, given factors $\tilde{\mathrm{F}}$, the set of K factors that have nonzero SDF weights differs in PCA and KNS. The K factors in a PCA model are those with the K largest variances, while the K factors in KNS are those with the highest means.

Furthermore, the KNS weights in Equation (13) are shrinked toward zero compared with the mean-variance weights in Equation (6). Given the sparsity parameter ν_1, larger values of ν_2 lower the weights toward zero and thus have the interpretation of shrinkage parameters. Since the effects of ν_1 and ν_2 can be separated, the KNS estimator can be applied sequentially in three steps:

1. Construct factors $\tilde{\mathrm{F}}$ by rotating the N test assets X using PCA (normalized so that $\mu_{\tilde{F},i}\geq 0\ \forall i$).

---

[11] The solution to Equation (12) is an elastic net regression of $\boldsymbol{\Sigma}_{\tilde{F}}^{1/2}$ on $\boldsymbol{\Sigma}_{\tilde{F}}^{-1/2}\mu_{\tilde{F}}$ with an analytical solution if $\boldsymbol{\Sigma}_{\tilde{F}}$ is a diagonal matrix. The proof follows the same logic as the elastic net regression with orthonormal regressors in Zou and Hastie (2005). The first-order condition on the active set, that is, for the nonzero values of $\widehat{b}_{\mathrm{KNS},i}$, equals

$$\left(\sigma^2_{\tilde{F},i}+\nu_2\right)\widehat{b}_{\mathrm{KNS},i}=\mu_{\tilde{F},i}-\nu_1\,\mathrm{sign}(\mu_{\tilde{F},i}),$$

with the solution in (13). Without loss of generality, we have normalized the means $\mu_{\tilde{F},i}$ to be positive.

2. Rank the the factors $\tilde{F}_i$ according to their means.
3. Select K factors with means that exceed ν_1.
4. Construct the SDF M_t using weights (13).

The same procedure can be applied when $\mathbf{\Sigma}_{\tilde{F}}$ is not diagonal. The only difference is that the factor selection in step 2 and the estimation of SDF weights in step 3 are combined by numerically solving (12).[12]

Note that $\mu_{\tilde{F},i} = \sigma_{\tilde{F},i} \mathrm{SR}_{\tilde{F},i}$, so the factor selection depends on the factor variances as well as Sharpe ratios. Factors with a low variance and high Sharpe ratio can be included in the KNS SDF but are excluded in the PCA SDF. In this sense, KNS has a similar motivation as RP-PCA, namely to incorporate information in the first moment. The difference is that RP-PCA changes the construction of factors, while KNS affects the factor selection and SDF weighting for a given set of factors. Both methods are therefore complementary and can be combined. In the empirical section, we compare results for the KNS estimator based on PCA factors to results for the KNS estimator based on RP-PCA factors.

Finally, we consider a case in which K factors are chosen according to their means, as in KNS, but SDF weights $\widehat{b}_{\mathrm{SMV},i}$ are not shrinked but set to their mean-variance values. We label this estimator SMV (sparse mean variance) which takes the following form for PCA factors:[13]

$$\widehat{b}_{\mathrm{SMV},i} = \begin{cases} \dfrac{\mu_{\tilde{F},i}}{\sigma^2_{\tilde{F},i}} & \text{if } \mu_{\tilde{F},i} \geq \nu_1, \\ 0 & \text{if } \mu_{\tilde{F},i} < \nu_1. \end{cases} \tag{14}$$

To maintain consistency throughout the paper, we apply all estimators to excess returns of the test assets. KNS first regress excess returns of all assets on the excess return of the CRSP-VW index and use the residuals as test assets. Hence the empirical results are not directly comparable to those in KNS.

2. Empirical Results

We proceed with the empirical application in two steps. First, we consider double-sorted portfolios with $N = 25$ test assets. The RP-PCA methodology is designed to handle much larger cross-sections, but the cases with smaller N are helpful in demonstrating the methodology. Next, we consider the larger cross-section of anomaly portfolios used in Kozak, Nagel, and Santosh (Forthcoming)

[12] We follow KNS and solve the optimization in Equation (12) numerically using a least angle regression (LARS) algorithm. In the PCA space the separation of Equation (13) holds exactly and gives identical results to the numerical elastic net regression. When applying equation (13) in the RP-PCA space, $\mathbf{\Sigma}_{\tilde{F}}$ is close to but not exactly a diagonal matrix. However, in all cases that we consider, the KNS estimator selects the K factors with the largest means.

[13] In the case of RP-PCA based factors we use Equation (12) to select K RP-PCA factors and apply Equation (6) to obtain the SDF weights $\widehat{\mathrm{b}}_{\mathrm{SMV}}$.

that is based on single-sorts of 37 different characteristics.[14] The sample span is November 1963 to December 2017 in all cases.

We compare models using three criteria: (i) the maximum Sharpe ratio that can be obtained by a linear combination of the factors, (ii) the root-mean-square pricing error RMS_α, and (iii) the average idiosyncratic variance $\overline{\sigma}_e^2$. All reported results are out-of-sample using the following procedure: Factors and loadings are estimated in rolling windows of 20 years ($T=240$). With these estimated loadings including information up to time t, we predict the $t+1$ return and obtain the out-of-sample pricing error at $t+1$. The mean and variance of the out-of-sample errors are used to calculate the average pricing error and the idiosyncratic variation, respectively. We use the optimal portfolio weights for the maximum Sharpe ratio portfolio estimated in the rolling window period to create an out-of-sample optimal return, giving us the maximum Sharpe ratio portfolio out-of-sample.

As mentioned in Section 1, only the product of factors and loadings, $\text{F}\Lambda^\top$, is identified. For the theoretical derivations in Section 1, it was convenient to normalize the loadings in terms of $\Lambda^\top\Lambda$. When presenting empirical results, it turns out to be more useful to normalize the variance-covariance matrix of the factors instead; hence we set $\mathbf{\Sigma}_F=\text{I}_K$, unless otherwise noted. The reason is that we want to compare different factor models and their composition in terms of the original portfolios. To do so, the "units" of estimated loadings must be comparable across models. Normalizing $\mathbf{\Sigma}_F=\text{I}_K$ achieves this objective. For some results, normalizing the loadings, that is, $\Lambda^\top\Lambda=\text{I}_k$, is more informative. Of course, the overall model is not affected by any normalization.

2.1 Double-sorted portfolios

Table 1 reports results for eight different sets of 25 portfolios sorted on size and book-to-market, accruals, investment, profitability, momentum, short-term reversal, volatility, and idiosyncratic volatility. We compare RP-PCA with PCA as well as models with three Fama-French factors (the market, small-minus-big (SMB), and high-minus-low (HML), where SMB and HML are constructed from sorts on size and on the second characteristic). We set γ to 20, but all results are not sensitive to this choice (robustness checks are reported in the Online Appendix).

The RP-PCA Sharpe ratio is larger than the PCA and FF Sharpe ratios in six of the eight cases while the RMS alphas are smaller than those of RP-PCA than those of PCA. For the two volatility cases, FF model yields RMS_α that are slightly smaller than those of RP-PCA, but the RP-PCA cross-sectional errors are smaller in all other cases. PCA minimizes, by construction, the in-sample idiosyncratic variance, but the out-of-sample $\overline{\sigma}_e^2$ is larger for PCA than for RP-PCA in five of the eight double-sorts, and both models capture more covariation than FF models in all cases.

---

[14] We thank the authors for sharing their data.

Table 1
Out-of-sample fit of RP-PCA, PCA and FF – Double-sorted portfolios

	SR			RMS_{α}			$\overline{\sigma}_e$		
	RP-PCA	PCA	FF	RP-PCA	PCA	FF	RP-PCA	PCA	FF
SIZE&BM	**0.21**	0.18	0.16	**0.17**	0.18	0.18	7.73%	**7.69%**	7.79%
SIZE&ACC	**0.22**	0.13	0.11	**0.10**	0.12	0.12	6.86%	**6.52%**	7.10%
SIZE&INV	**0.26**	0.23	0.17	**0.13**	0.15	0.16	**6.99%**	7.04%	7.95%
SIZE&OP	0.13	0.13	**0.20**	**0.10**	0.10	0.16	**6.91%**	7.04%	9.38%
SIZE&ST-REV	**0.16**	0.10	0.08	**0.19**	0.20	0.20	7.89%	**7.87%**	7.90%
SIZE&MOM	0.19	0.16	**0.23**	**0.20**	0.21	0.22	**8.21%**	8.31%	12.03%
SIZE&IVOL	**0.30**	0.24	0.28	0.17	0.18	**0.16**	**6.21%**	6.23%	6.61%
SIZE&VOL	**0.28**	0.22	0.27	0.18	0.20	**0.17**	**6.22%**	6.25%	6.63%

OOS maximal Sharpe ratios, root-mean-squared pricing errors and unexplained idiosyncratic variation for panels of double-sorted portfolios with $N=25$. Bold numbers indicate the best-performing models.

Why does RP-PCA outperform PCA in most cases? The estimated RP-PCA model can differ from the estimated PCA model in three ways: the detection of factors, the factor compositions, and the order of factors. To illustrate the mechanism, we study two cases in more detail: size/accruals and size/short-term reversal. The Fama-French model consists of three factors: the market, SMB, and HML (based on size and second characteristic). As we will see later, both methods extract a long-only portfolio that is almost perfectly correlated with the market portfolio and a long-short size factor. However, PCA is not able to detect the long/short accrual factor, while RP-PCA does. In the size/short-term reversal case, the order of the factors changes. The second RP-PCA factor is a high/low reversal factor, and the third factor is a size factor. In PCA, the second factor is related to size, while the third factor is related to reversals. The reason is that the reversal factor captures average return differences and hence is given a higher weight in RP-PCA estimation.

Figure 2 plots the Sharpe ratio in the top panel, the RMS time-series pricing errors α_n, and the unexplained time-series variation (as a fraction of total variation) as a function of the RP-PCA weight γ for the size/accruals portfolios in the left panels and size/reversals in the right. To illustrate the effect of each factor, we consider models with one, two, and three factors. For size/accrual portfolios, γ has little effect on the Sharpe ratios, α's, and unexplained variances in models with one and two factors. Hence, RP-PCA and PCA are essentially equivalent. This is not the case when a third factor is added (solid green lines). The out-of-sample SR increases and the RMS α decreases when increasing γ to 5, then maintains this level. The out-of-sample SR is more than doubled, and RMS α is cut in half compared with the PCA case of $\gamma=-1$.

What causes the difference between PCA and RP-PCA when a third factor is added? Figure 3 shows a heat map of the loadings of the three factors for PCA with $\gamma=-1$ and RP-PCA with $\gamma=20$. Positive loadings are in green and negative loadings in red. The figure shows that the loadings of the first two factors are similar for PCA and RP-PCA. The first factor is a "long" factor with positive weights on all portfolios with a tilt toward small-stock portfolios.

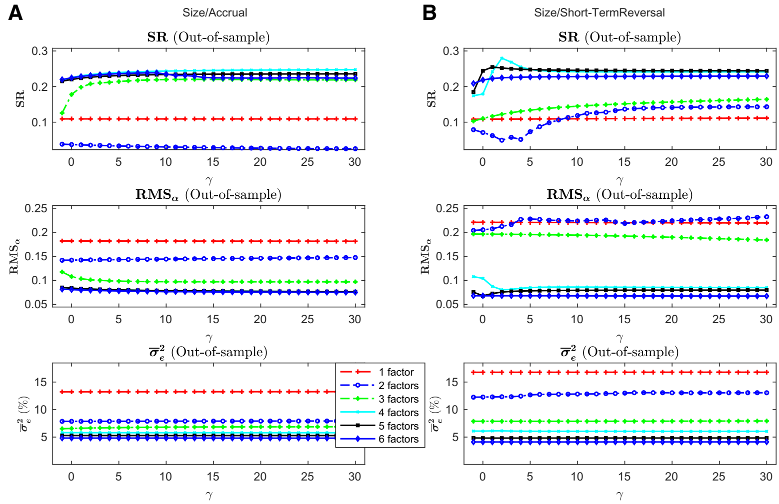


Figure 2
Out-of-sample results as a function of γ
Out-of-sample maximal Sharpe ratios, root-mean-squared pricing errors, and unexplained idiosyncratic variation for different number of factors and γ. Left: Size/Accrual. Right: Size/Short-Term Reversal.

The second factor is a long-short factor with positive weights on small-stock portfolios and short in large-stock portfolios. These factors are similar to the market and SMB factors in Fama-French models. The composition of the third factor is different for PCA and RP-PCA. The PCA factor has no clear pattern and adds very little information, as shown in Figure 2. In contrast, the third RP-PCA factor has positive loadings on low-accrual stocks and negative loadings on high-accrual stocks, similar to a Fama-French-type factor. This pattern explains the higher Sharpe ratio and lower α's in the three-factor RP-PCA model shown in Figure 2.

Consider next the results for the 25 size/short-term reversal portfolios shown in the right panels in Figure 2. The SR and RMS α for models with one factor are almost unaffected by γ and change only moderately for the three-factor model. In contrast, the statistics of two-factor models change substantially for different γ. The OOS-SR rises and OOS-RMSE α decreases once γ exceeds 5 and they settle down at $\gamma \approx 20$. Figure 3 compares the loadings for the three factors for PCA and RP-PCA with $\gamma = 20$. As in the case of the size/accruals portfolios, the loadings of the first factor is very similar for both models. The second PCA factor has positive loadings of small-stock portfolios and negative ones of large-stock portfolios, it is hence a "small-minus-big" portfolio. The third PCA factor is long in low-reversal portfolios and short in high-reversal stocks and is therefore a "long-minus-short" reversal factor. These two factors are reversed in RP-PCA. The second RP-PCA factor is a "long-minus-short" reversal factor, while the third factor is "small-minus-big" factor. The reason for

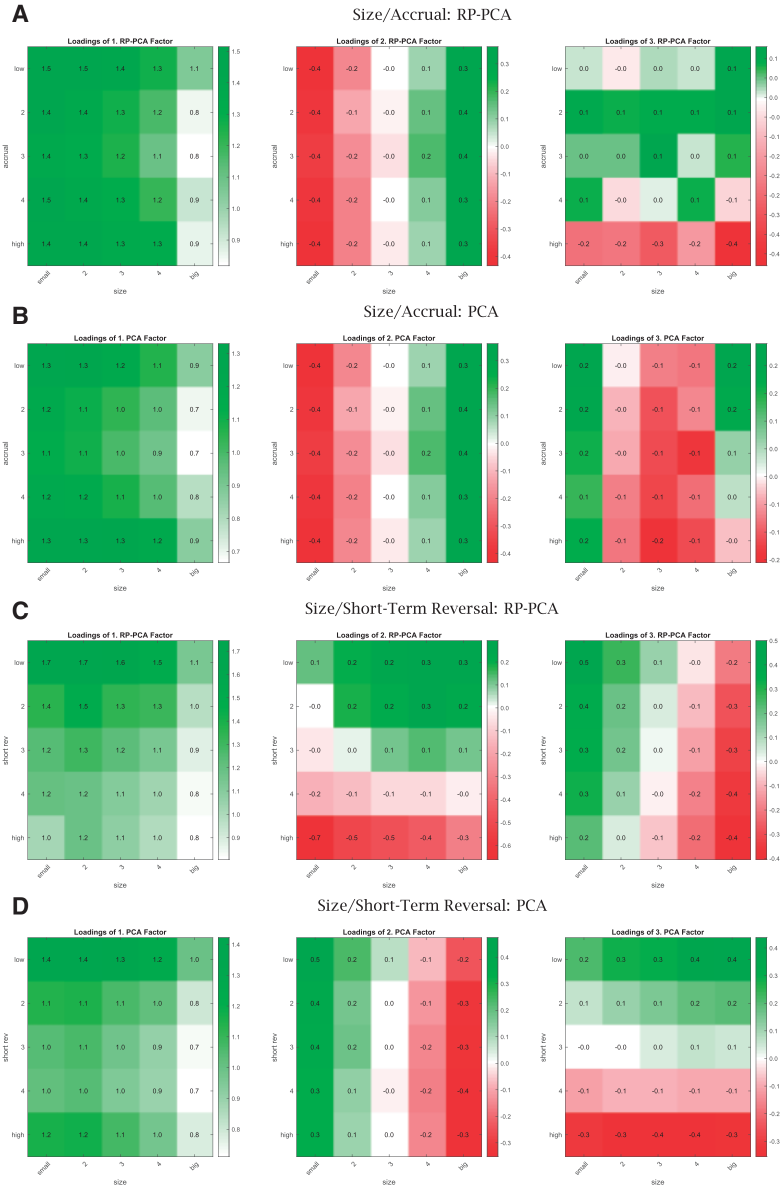


Figure 3
Statistical factors for size/accrual and size/short-term reversal portfolios
Heat map of loadings of $K=3$ statistical factors and RP weight $\gamma=20$.

this switch is that the return spread is much larger along the reversal dimension than along the size dimension since RP-PCA gives high Sharpe ratio factors additional weight. A "small-minus-big" factor contributes more towards the common time-series variation and is this favored by PCA, whereas the "long-minus-short" reversal factor contributes more to the cross-sectional dimension and is thus favored by RP-PCA.

2.2 Large cross-section of single-sorted portfolios

Next, we apply the RP-PCA estimator to a larger cross-section of portfolios. We select all 37 characteristics from the data set used in Kozak, Nagel, and Santosh (Forthcoming) that are available as of November 1963. Our sample ends in December 2017 ($T = 650$). As a robustness check, we also estimate the models using a broader set of 49 characteristics, albeit over a shorter span of 530 months starting in 1973. The data set includes 10 single-sorted decile portfolios for each characteristic for a total of 370 test portfolios. While the RP-PCA estimator is suited for large cross-sections, the presentation of results with the full set of 370 portfolios is cumbersome. As we will see, most of the relevant information is contained in the extreme first and tenth decile portfolios. Therefore we consider a smaller cross-section with the 74 decile-1 and decile-10 portfolios as test assets as a benchmark. The paper also reports some results for the full panel of 370 portfolios, while the Online Appendix includes a complete set of results. The results for both sets of portfolios are similar, and all findings are robust.

Table C.1 shows the mean returns, the standard deviations, the SRs of the long-short portfolios, as well as the mean return of the extreme decile 1 and 10 portfolios. The portfolios are sorted according to their Sharpe ratios. The means of the top 22 long-short portfolios are statistically significantly different from zero while 15 are insignificant.[15] The moments of the portfolio returns are described in more detail in Kozak, Nagel, and Santosh (Forthcoming).

2.2.1 Choice of γ and number of systematic factors. We start the analysis by investigating how many factors are "systematic," how many are "idiosyncratic," and how the parameter γ affects the detection of systematic factors. We use two approaches to determine the number of factors: First, we apply a diagnostic criterion based on the pattern of eigenvalues associated with the eigenvectors that define factors, as described in Section 1. Second, we use cross-validation to study the effect of γ and the number of factors on the performance of the model.

Figure 4 plots the differences of consecutive eigenvalues of $\frac{1}{T}\mathrm{X}^\top\mathrm{X}+\gamma\bar{\mathrm{X}}\bar{\mathrm{X}}^\top$ for different values of γ for the sample with only deciles 1 and 10 in panel A and the sample with all deciles in panel B. The red line indicates the critical values of the Onatski criterion. For $\gamma < 10$, the fifth eigenvalue difference is below the critical value in both samples, indicating that the first four factors

[15] The t-statistic is $t_\mu = \sqrt{T}\,\mathrm{SR} = 25.5\,\mathrm{SR}$.

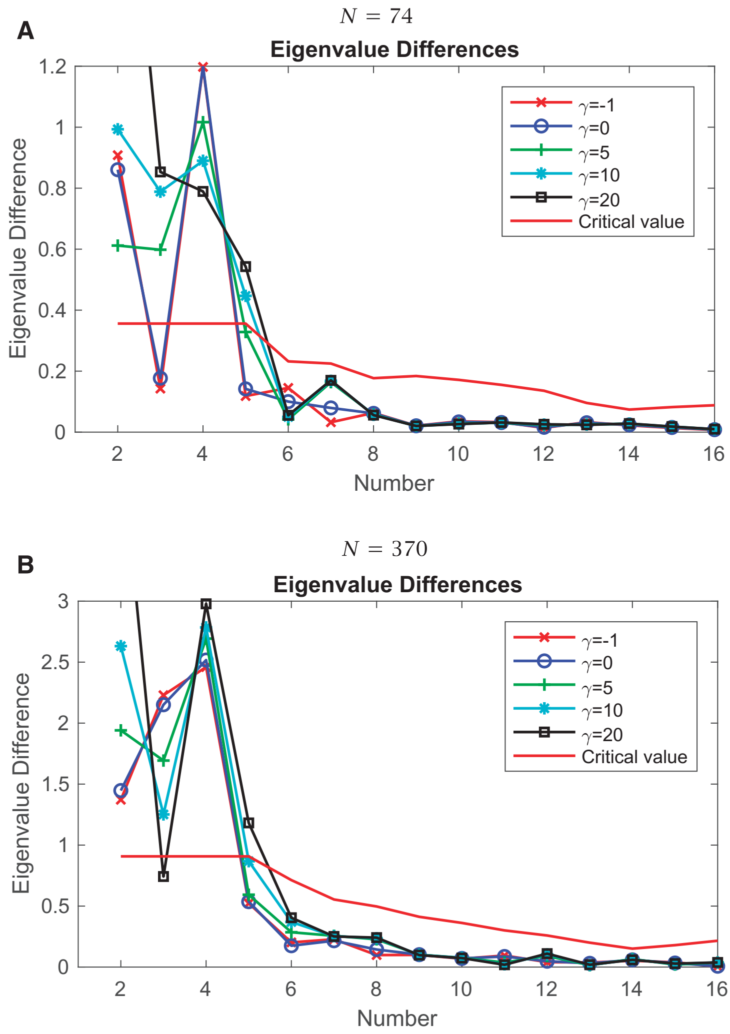


Figure 4
Eigenvalue differences for single-sorted portfolios
Differences of consecutive eigenvalues of the matrix $\left(\frac{1}{T}X^{\top}X+\gamma\bar{X}\bar{X}^{\top}\right)$ for different RP weights γ. The red line indicates the critical value of the Onatski test. First and last deciles of 37 single-sorted portfolios ($N=74$).

are systematic. However, for $\gamma \geq 10$, the fifth eigenvalue difference is above the critical value, indicating that there are five systematic factors. This suggests that the fifth factor is weak and below the detection threshold for PCA and RP-PCA with $\gamma < 10$. Once γ is chosen high enough, RP-PCA is able to detect this factor. Note that the optimal number of systematic factors is the same for the sample with 74 decile-1 and decile-10 portfolios and the full sample with

370 portfolios. Adding 296 decile-2 to decile-9 portfolios does not change the number of systematic factors, suggesting that most of the relevant information is contained in the decile-1 and decile-10 portfolios. Results reported later confirm this interpretation.

As a second diagnostic criterion, we use a cross-validation estimation for different combinations of γ and number of factors k. We divide the sample into three subsamples of equal length. For each subsample $\upsilon = 1, 2, 3$, we estimate the model using only data in subsample υ and compute the "out-of-sample" properties on the data not included in subsample υ. For each γ and k, we average the results for the three subsamples.

Figure 5 shows the heat map for the maximum Sharpe ratio for $k = 1, \ldots, 10$ factors on the x-axis and $\gamma = -1, 0, 1, \ldots, 50$ on the y-axis. We report results in the sample with the decile-1 and 10 portfolios ($N = 74$) but the results for the full sample are similar. Yellow cells indicate low Sharpe ratios while red cell indicate high SRs. The Sharpe ratio for models with only one factor is around 0.1 to 0.2, no matter the value of γ. For low values of γ, adding additional factors gradually increases the Sharpe ratio and even higher-order factors improve the performance of the model. For PCA with $\gamma = -1$, adding factors 2 to 5 raises the Sharpe ratio to 0.38, while including factors 6 to 10 yields an SR of 0.47. In contrast, for $\gamma \geq 10$, the first five factors have a large incremental effect on the Sharpe ratio while the effect of higher-order factors is minor. For $\gamma = 10$, the RP-PCA model with five factors yields an SR of 0.5. Adding factors 6 to 10 only increases the SR to 0.58. Increasing γ beyond 10 does not change the performance of the model substantially. These results are consistent with the Onatski criterion and suggest that RP-PCA is able to extract five factors. We set $\gamma = 10$ as a benchmark and show results for different values in the appendix.

2.2.2 Estimation results: RP-PCA versus PCA. Figure 6 compares the SR, RMS α and unexplained variance in-sample and out-of-sample for the sample with 74 decile-1 and decile-10 portfolios. For each specification we start with one factor and successively add factors (up to a total of 10 factors). The RP-PCA weight γ is set to 10. The top two panels show the Sharpe ratios of the stochastic discount factors that are implied by the estimated factors. As expected, adding factors increases the ins-sample (IS) Sharpe ratios of the SDFs for all models, but the magnitudes of the incremental results differ. First, the overall SRs are significantly higher for RP-PCA than those for PCA. The IS-SR for the RP-PCA model with five factors is 0.57, about twice as high as the IS-SR for PCA. The second and fifth RP-PCA factors have a particularly large incremental effect on the IS-SR while the SR pattern for PCA is more gradual. The out-of-sample (OOS) SRs exhibit a similar pattern but there are some important differences. The OOS-SRs are only slightly lower than the IS-SRs suggesting that the RP-PCA model does not suffer from excessive overfitting. The out-of-sample SR increases more gradually for the first five factors but adding additional factors has only a minor effect on the OOS-SR, which confirms that RP-PCA extracts

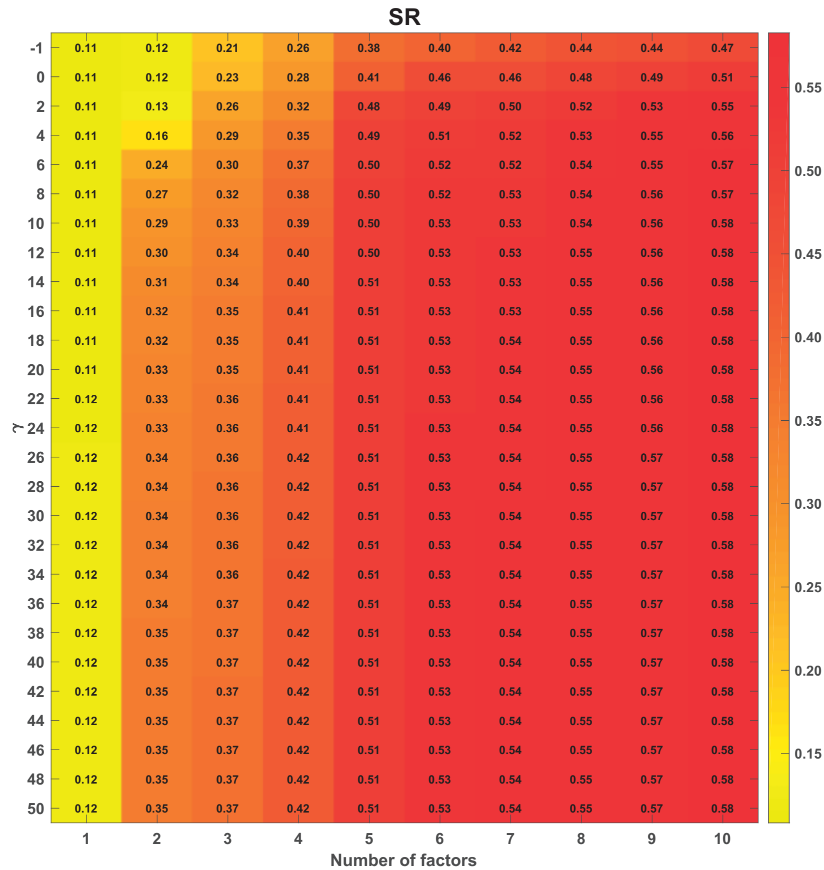


Figure 5
Cross-validation estimation for decile-1 and decile-10 portfolios (N=74)
SR as a function of the number of factors k and γ for decile-1 and decile-10 portfolios (N=74).

five systematic factors. The OOS-SRs for PCA models are significantly lower than those for RP-PCA. A 10-factor PCA specification yields an OOS-SR of 0.36, while the OOS-SR of a five-factor RP-PCA model is 0.50. These OOS results are consistent with the cross-validation estimation in Figure 5.

The middle row shows the root-mean-squared cross-sectional pricing errors, RMS_α. First, note that adding incremental factors in all models reduces pricing errors not only in-sample but also out-of-sample, again confirming that the factor structure is stable and the models are not overfitting. Second, RP-PCA pricing errors are smaller than those for PCA in all cases but the incremental effects of the factors are different. As for Sharpe ratios, the second and fifth RP-PCA factors have a substantial effect on the IS-RMS_α but the effects are

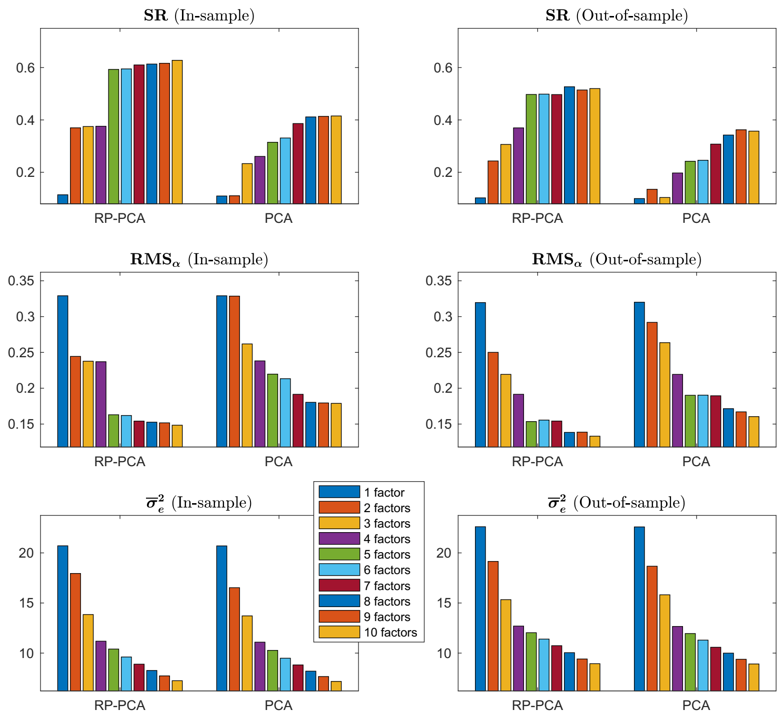


Figure 6
Fit for 74 decile-1 and decile-10 portfolios
Maximal Sharpe ratios, root-mean-squared pricing errors, and unexplained idiosyncratic variation for different numbers of factors. RP weight $\gamma = 10$.

mitigated out-of-sample. The reduction of pricing errors beyond the fifth RP-PCA factor is minimal in-sample as well as out-of-sample. Moreover, the pricing errors of the five-factor PCA model are about 50% larger than those of the five-factor RP-PCA model with $\gamma = 10$.

The plots for the unexplained idiosyncratic variations in the bottom row are almost identical for comparable RP-PCA and PCA models. Recall that PCA minimizes the in-sample unexplained variation $\overline{\sigma}_e^2$ but does not necessarily yield the smallest out-of-sample variation. The differences between the out-of-sample $\overline{\sigma}_e^2$ of RP-PCA and PCA with the same number of factors is negligible and negative in some cases. Hence, RP-PCA yields higher OOS-SRs and smaller OOS pricing errors than PCA, yet captures a similar amount of OOS covariation as PCA.

Panel A of Table 2 compares the Sharpe ratios, pricing errors and idiosyncratic variance for RP-PCA and PCA models with three and five factors for our benchmark sample. We also consider the three-factor Fama-French models (market, SMB, and HML) and the five-factor model (market, SMB, HML, RMW, and CMA) as comparison. Consider first the three-factor models.

Table 2
Fit of RP-PCA, PCA, and Fama-French models

	In-sample			Out-of-sample		
Model (K)	SR	RMS_α	$\overline{\sigma}_e$	SR	RMS_α	$\overline{\sigma}_e$
Panel A: 74 portfolios						
RP-PCA (3)	**0.37**	**0.24**	13.85%	**0.31**	**0.22**	**15.33%**
PCA (3)	0.23	0.26	**13.72%**	0.10	0.26	15.82%
Fama-French (3)	0.21	0.31	17.48%	0.16	0.25	16.53%
RP-PCA (5)	**0.59**	**0.16**	10.40%	**0.50**	**0.15**	12.04%
PCA (5)	0.31	0.22	**10.27%**	0.24	0.19	**11.95%**
Fama-French (5)	0.32	0.26	15.96%	0.30	0.19	13.68%
Panel B: 370 portfolios						
RP-PCA (3)	**0.23**	0.17	12.78%	**0.18**	**0.15**	**14.57%**
PCA (3)	0.17	**0.17**	**12.71%**	0.13	0.15	14.66%
Fama-French (3)	0.21	0.18	14.60%	0.16	0.16	14.90%
RP-PCA(5)	**0.53**	**0.14**	10.79%	**0.45**	**0.12**	12.70%
PCA (5)	0.24	0.14	**10.68%**	0.16	0.14	**12.56%**
Fama-French (5)	0.32	0.16	13.56%	0.30	0.13	13.66%
Panel C: 98 portfolios						
RP-PCA (3)	**0.34**	**0.23**	13.58%	**0.19**	**0.23**	**16.34%**
PCA (3)	0.24	0.25	**13.52%**	0.15	0.24	16.78%
Fama-French (3)	0.21	0.31	17.02%	0.14	0.26	17.73%
RP-PCA (5)	**0.54**	**0.17**	10.39%	**0.42**	**0.16**	13.05%
PCA (5)	0.33	0.20	**10.30%**	0.23	0.19	**12.90%**
Fama-French (5)	0.34	0.23	15.15%	0.22	0.20	14.70%
Panel D: 270 individual stocks						
RP-PCA (3)	0.22	**0.32**	64.37%	0.13	**0.40**	68.74%
PCA (3)	0.18	0.34	**64.25%**	0.16	0.41	**68.60%**
Fama-French (3)	**0.22**	0.43	70.94%	**0.17**	0.43	70.68%
RP-PCA(5)	0.26	**0.27**	59.95%	0.02	0.43	**65.72%**
PCA (5)	0.23	0.28	**59.92%**	0.15	**0.41**	66.11%
Fama-French (5)	**0.34**	0.39	69.29%	**0.25**	0.42	68.32%

Maximal Sharpe ratios, root-mean-squared pricing errors, and unexplained idiosyncratic variation. Bold numbers indicate the best-performing models for a given K. The samples include 74 decile-1 and decile-10 portfolios in panel A, all 370 decile portfolios in panel B, 98 decile-1 and decile-10 portfolios in panel B for a shorter sample from November 1973 to December 2017 ($T=530$), and 270 individual stocks for a sample from May 1972 to December 2013. Unless otherwise noted, the sample starts in November 1963 and ends in December 2017 ($T=650$).

RP-PCA outperforms PCA and Fama-French in all criteria: highest SR, smallest pricing errors, and residual variance. The in-sample residual variance is by construction minimized by PCA (13.72%), but $\overline{\sigma}_e^2$ is only slightly higher for RP-PCA (13.85%). The results for the three-factor Fama-French model are comparable to those for PCA but are worse than those for RP-PCA. The pattern is similar for five-factor models with the exception that out-of-sample PCA $\overline{\sigma}_e^2$ is slightly lower than that of RP-PCA. Note that the out-of-sample Sharpe ratios for the 3 and 5 factor Fama-French models are higher than those of the corresponding PCA models but substantially lower than those of RP-PCA models. Results for the full sample with 370 portfolios are reported in panel B and panel C shows the results for 98 portfolios ($T=530$) as robustness checks.

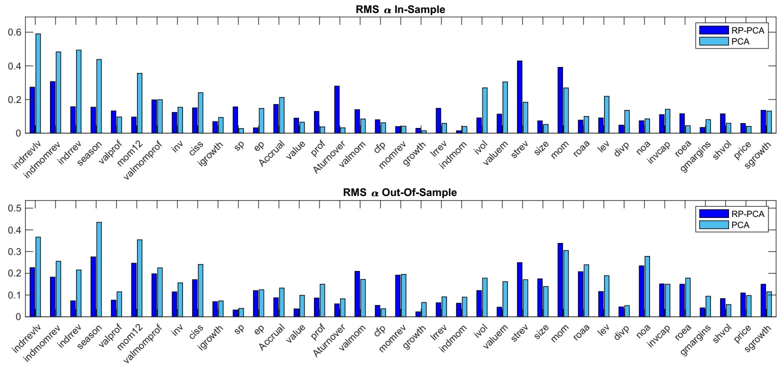


Figure 7
RMS of time-series α's by characteristic
First and last deciles of 37 single-sorted portfolios ($N=74$). Root-mean-squared pricing errors in- and out-of-sample for five RP-PCA and PCA factors.

The results are similar to those in the benchmark sample. We conclude that a five-factor RP-PCA model is the preferred model for all three data sets.

Figure 7 shows the mean-squared cross-sectional pricing errors (α's) for each anomaly based on models with five factors. The anomalies are ranked by the Sharpe ratio of the decile-10 return minus the return of decile-1 ("10-1"). The dark blue bars are RP-PCA α's, and the light blue bars are those for PCA. The IS and OOS PCA pricing errors are largest for high-SR anomalies, such as industry relative reversals (*indrrevlv*), industry momentum-reversal (*indmomrev*), and industry relative reversals (*indrrev*). The pricing errors of the RP-PCA model for these portfolios are substantially smaller, for some anomalies by an order of magnitude (e.g. 12-month momentum (*mom12*) and industry relative reversals (*indrrev*)). The RP-PCA model improves the fit, in particularly for the portfolios for which PCA performs worst. The in-sample PCA alphas of asset turnover (*aturnover*), short-term reversal (*strev*), and momentum (*mom*) are lower than those of RP-PCA, but the difference is much smaller out-of-sample. We conclude that RP-PCA factor models lower the pricing errors of most portfolios, in particular out-of-sample, relative to PCA models, especially for the high-SR portfolios that are most mispriced by PCA.

2.2.3 Time-series versus cross-sectional factors. So far, we have studied the overall performance of various models. Next, we investigate the properties of individual factors. Table 3 reports the mean, variance, and Sharpe ratios of the first ten RP-PCA and PCA factors. The factors are normalized so that their mean returns are positive. Factor means that are significantly different from zero are starred. To make factors comparable, we normalize all factor loadings to a norm of one. As is common in factor modeling of asset returns, the variance of the first factor is a magnitude larger than the variances of the other factors.

Table 3
Individual factors

	RP-PCA				PCA			
Factor	Mean	Variance	SR	Mean rank	Mean	Variance	SR	Mean rank
1	5.02*	1940.89	0.11	**1**	4.83*	1950.76	0.11	**1**
2	2.33*	65.86	0.35	**2**	0.15	102.60	0.01	9
3	0.24	101.54	0.06	**4**	1.71*	69.19	0.21	**2**
4	0.09	65.50	0.03	6	0.94*	64.54	0.12	**3**
5	0.74*	26.68	0.46	**3**	0.79	20.16	0.18	**5**
6	0.04	19.52	0.05	9	0.45*	19.27	0.10	7
7	0.12*	18.05	0.14	**5**	0.81*	16.50	0.20	**4**
8	0.05	15.59	0.06	8	0.56*	15.26	0.14	6
9	0.04	13.27	0.06	10	0.15	13.22	0.04	8
10	0.08	12.04	0.12	7	0.12	11.99	0.04	10

The table shows the mean, variance of original factors, and the Sharpe ratio of the incremental uncorrelated factor component. Significant means are starred. The table also shows the rank of the factor means. The five factors with the highest mean are indicated in bold.

As we will see later, the first factor is highly correlated with aggregate stock indices and thus akin to the "market." For PCA, the factors are ranked according to their variances. The variance of the second PCA factor is 102.60, which is substantially higher than the variances of higher-order factors, but its mean of 0.15 is lower than the means of factors three to eight, suggesting that this factor captures comovement but might not be priced. Four of the first five factors have significant means, and only the mean of the second factor is insignificant. Note that factors six, seven, and eight also have a significant mean, although they are not selected by the model selection criteria.

The ranking of RP-PCA factors depends on their variances but also on their means and thus is not necessarily monotonic in either moment. In contrast to PCA, the second RP-PCA factor has a high and significant mean, but its variance is lower than that of the third factor. Therefore, this factor has a high Sharpe-ratio. The third RP-PCA factor has an insignificant mean but higher variance. Therefore, the second factor captures less comovement but is priced, while the third factor captures comovement and is not priced. As we have seen, the preferred RP-PCA model includes five factors according to the selection criteria. Among those five factors, factors 2 and 5 have high and significant means and Sharpe ratios and are therefore likely to be priced and important for capturing the cross-section of returns, while factors 3 and 4 capture comovement but are likely unpriced.

To further investigate the roles of RP-PCA factors, we compute the fit for models with subsets of factors; see Table 4. In addition to the case with all five factors as benchmark, we consider specifications with only factors $[1],[1,2,5],[2,5],[1,3,4]$, and $[3,4]$. Recall that the first factor is long-only in all portfolios and is highly correlated with the market return. Thus, the specification with only the first factor is similar to the CAPM with only the market return. The Sharpe ratio is only 0.11 (compare to the five-factor model) and the RMS pricing error and $\overline{\sigma}_e^2$ are about twice as high. However, the first

Table 4
Fit for RP-PCA model with subset of factors

	In-sample			Out-of-sample		
Factors	SR	RMS_α	$\overline{\sigma}_e$	SR	RMS_α	$\overline{\sigma}_e$
[1,2,3,4,5]	0.59	0.16	10.40%	0.50	0.15	12.04%
[1]	0.11	0.33	20.71%	0.10	0.32	22.57%
[1,2,5]	0.58	0.21	17.09%	0.51	0.18	18.16%
[2,5]	0.42	1.73	74.00%	0.38	0.34	21.67%
[1,3,4]	0.12	0.33	13.89%	0.01	0.31	15.80%
[3,4]	0.03	0.66	93.10%	−0.09	0.51	55.60%

factor by itself captures about 80% of the time-series variation, and adding factors 2 to 5 results in a marginal increase of 10 percentage points.

Next, we consider specifications that include the high mean and SR factors 2 and 5. The Sharpe ratio of the model with factors 1, 2, and 5 is as high as as for the specification with all five factors, 0.58 versus 0.59 in-sample and 0.51 versus 0.50 out-of-sample. The RMS pricing errors are 0.21 (IS) and 0.18 (OOS) compared with 0.16 and 0.15, respectively, and the IS and OOS idiosyncratic variances are about 50% higher than those of the five-factor model. Further removing the first factor increases the RMS_α and $\overline{\sigma}_e^2$, but the Sharpe ratios are still sizable at around 0.4. Note that the in-sample RMS_α and $\overline{\sigma}_e^2$ are orders of magnitude higher than their out-of-sample counterparts. The reason for this difference in RMS_α is that factors are orthogonal when computed in-sample, and hence higher-order factors only capture variation that is not captured by lower-order factors. Thus, removing the first factor, which captures a large portion of variation, reduces the systematic variance drastically. In contrast, factors are not necessarily orthogonal when applied out-of-sample, and thus higher-order factors can also capture some of the variation that is due to the first factor. A similar effect explains the much lower IS-RMS_α than OOS-RMS_α. We conclude that the three-factor model with factors 1, 2, and 5 produces Sharpe ratios and pricing errors similar to the five-factor model but captures less of the time-series variation.

The results for specifications that include factors 3 and 4 differ from those for models with factors 2 and 5. Consider first the three-factor model [1,3,4]. The Sharpe ratios and RMS_α are close to those of the model with only the first factor. In fact, the OOS-SR is only 0.01 and much lower than that of the one-factor model. Hence, the factors 3 and 4 play only minor marginal roles for the SR and RMS_α. On the other hand, the IS (OOS) idiosyncratic variances of the three-factor model are 13.89% (15.80%), only about 3.5 percentage points higher than those of the model with five factors. Removing the first factors deteriorates the fit of the model along all dimensions. The OOS Sharpe ratio of −0.09 is negative and the lowest among all specifications we consider while RMS_α and $\overline{\sigma}_e^2$ are the highest.

We conclude that the RP-PCA factors can be characterized as follows. The first factor is relevant for the time-series and cross-sectional fit. Factors 2 and

5 are relevant for the cross-sectional fit and SR but not for the time-series. And factors 3 and 4 are relevant for the time-series fit but play no role in the cross-section or Sharpe-ratio. Moreover, three factors $[1,2,5]$ are sufficient to capture the cross-section of expected returns and yield a high Sharpe-ratio, while a different three-factor model $[1,3,4]$ captures most of the time-series variation.

2.2.4 Composition of SDF and factors. Next, we study the composition of the estimated SDF and individual factors. Recall that a given set of factors implies an SDF that is perfectly negatively correlated with the tangency portfolio created by the factors, which is given in Equation (6). The SDF is a linear combination of factors, which are in turn linear combinations of the test assets X. Therefore, the SDF is also a linear combination of the test assets X. Next, we study the composition of the SDFs that are implied by the RP-PCA and PCA factors. Figure 8 shows the composition of the RP-PCA SDF in panel A and PCA SDF in panel B in terms of the 74 test portfolios. The weights of decile-10 portfolios are in dark blue bars, while the weights of decile-1 portfolios are in light blue bars. The portfolios are ranked by their Sharpe-ratio. First, we note that for both SDFs the high-returns decile-10 weights are mostly positive, while the weights of the low-return decile-1 portfolios are mostly negative. Hence, the SDF is composed of long-short portfolios but the weights do not necessarily sum to zero, as is the case for Fama-French style factors. Second, the largest loadings (in absolute value) of the RP-PCA SDF are related to anomalies with the highest Sharpe ratio (for example, *indrrev* and *indmomrev*), while the magnitudes of the PCA weights are not linked to Sharpe ratios. In other words, the high Sharpe ratio that is created by the RP-PCA factors reported above, is due to the fact that the RP-PCA SDF is to a large degree composed of high SR portfolios. In contrast, the PCA-SDF weights of the three characteristics with the highest Sharpe ratio are close to zero.

Figure 9 shows the composition of the first five RP-PCA and PCA factors. For presentation purposes, we group related anomalies into categories following Hou, Xue, and Zhang (2015) and Freyberger, Neuhierl, and Weber (2020). Table 5 lists the categories, the portfolios included in each category, the mean returns of the decile-1 and decile-10 portfolios and the average SR of the long-short portfolios in each category. The categories are listed according to their average Sharpe ratios. On average, reversal portfolios have the highest Sharpe ratio, followed by value-interaction portfolios. The factor weights of decile-1 portfolios are shown in red while those of decile-10 portfolios are in blue. Each bar shows the total weight of a category with the contribution of each portfolio in the categories indicated by lines.

The weights of the first RP-PCA and PCA factors are in the top row. The weights are all positive, so both factors are long-only “level” factors. Moreover, the weights of the first RP-PCA and PCA factors are similar and the factors have a correlation of 0.999, suggesting that the first factor is “strong” and

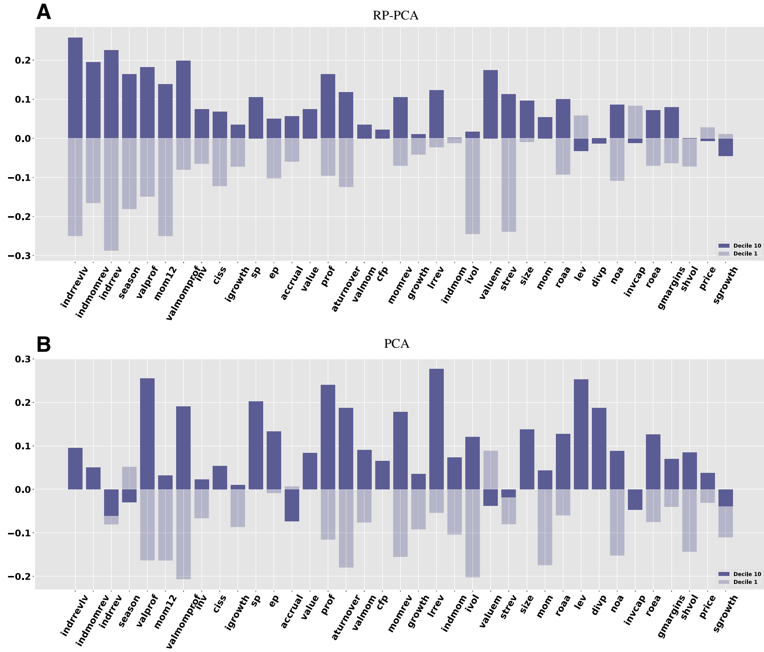


Figure 8
Portfolio weights in RP-PCA and PCA SDFs
Portfolio loadings of SDF estimated by RP-PCA and PCA for the first and last deciles of 37 single-sorted portfolios. The anomalies on the x-axis are sorted by their SR.

relatively easy to identify by PCA-based methods. The correlations of the first RP-PCA and PCA factors with the CRSP-VW index return are 0.993 and 0.995, respectively, confirming their interpretation as "level" factors that proxy for the market return. The monthly Sharpe ratio of the first RP-PCA and PCA factors is 0.11 (see Table 3), which is close to the SR of the CRSP-VW index of 0.12.

The second row of Figure 9 shows the composition of the second factors. In contrast to the long-only structure of the first factor, the second factors are composed of long and short positions of individual portfolios. Moreover, the decile-10 and decile-1 loadings are (mostly) of different signs; hence this factor is akin to long-short factors in which the high-return factors are of opposite signs to the low-return factors. Consider first the RP-PCA factors on the left of the figure. The second RP-PCA factor is composed of portfolios in all groups and thus appears not to be related to any particular anomaly category. In contrast, value portfolios dominate the third factor and can therefore be interpreted as "value factors". The composition of the fourth RP-PCA factor differs from the other figures because some decile-10 (decile-1) portfolios have positive (negative) weights while others have negative (positive) weights. Recall that this factor captures comovement but is unpriced since the mean of this factor is only 0.29 and statistically insignificant (see Table 3). Portfolios related to

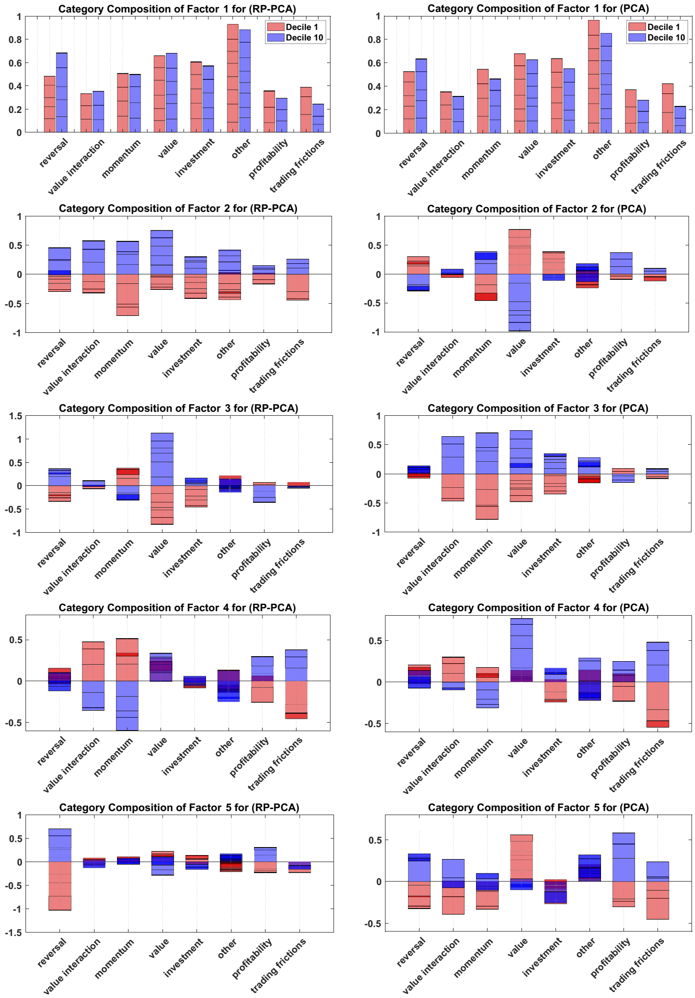


Figure 9
Factor loadings RP-PCA categories
Factor loadings of factors estimated by RP-PCA and PCA for first and last deciles of 37 single-sorted portfolios. Factors are grouped into anomaly categories.

momentum, value interaction and trading frictions have the largest (in absolute value) weights. The fifth RP-PCA factor is almost exclusively composed of reversal portfolios. This factor has a high and significant mean of 0.74 and its Sharpe ratio of 0.46 is the highest among all factors. This is due to the fact the reversal portfolios have particularly high Sharpe ratios, in fact three individual long/short portfolios with the highest SR are all in the reversal category. We will investigate the relationship of the high Sharpe ratio factors and returns further below.

Table 5
Descriptive statistics of categories

Category	Portfolios	Mean Dec. 1	Mean Dec. 10	SR 10-1
reversal	*lrrev*, *strev*, *indmomrev*, *indrrev*, *indrrevlv*	0.14	1.01	0.24
value interaction	*valmom*, *valmomprof*, *valprof*	0.31	1.01	0.16
momentum	*mom*, *mom12*, *indmom*, *momrev*	0.24	0.84	0.10
value	*value*, *valuem*, *divp*, *ep*, *cfp*, *sp*	0.41	0.87	0.09
investment	*inv*, *invcap*, *igrowth*, *growth*, *noa*	0.39	0.71	0.09
other	*size*, *price*, *accruals*, *ciss*, *gmargins*, *lev*, *season*, *sgrowth*	0.37	0.65	0.08
profitability	*prof*, *roaa*, *roea*	0.44	0.66	0.06
trading frictions	*ivol*, *shvol*, *aturnover*	0.39	0.65	0.05

The plots on the right in Figure 9 show the composition of PCA factors. The second factor is mostly composed of value portfolios but the high-return decile-10 value portfolios have a negative weight, while the weight of the decile-1 portfolios is positive. Recall that the mean of this factor is close to zero and insignificant, which implies that the signs of the weights are not informative. The third factor is mostly composed of portfolios related to value, value interactions, and momentum while the fourth factor is predominantly long in value and short in portfolios related to trading frictions. The pattern of the fifth factor is somewhat different in the sense that the weights of decile-10 portfolios related to profitability and frictions are positive, yet the weights of the decile-10 value portfolios are negative. In other words, this factor is not a traditional factor that is long in high-return decile-10 portfolios and short in low-return decile-1 portfolios.

2.2.5 Portfolio weights and returns. To further investigate the differences between RP-PCA and PCA, we study the relationship between portfolio weights and returns. Figure 10 shows scatterplots of average returns of the 74 portfolios on the x-axis and SDF loadings on the y-axis. Loadings of the RP-PCA SDF are in panel A and loadings of the PCA SDF are in panel B. The loadings of decile-10 portfolios are displayed in blue while decile-1 loadings are in red. Panel A shows that the composition of the SDF estimated by RP-PCA is strongly linked to the returns of the individual portfolios. The portfolios with the largest weights in the SDF are those with the highest returns, while portfolios with the smallest (negative) weights are those with the lowest returns. The correlation of returns and weights is 0.88.

While loadings of the PCA-SDF, shown in panel B, are also related to returns, the link is weaker. The weights of some portfolios with negative mean returns are positive, and some portfolios with high returns have a negative weight. The correlation of 0.53 is substantially lower than that for the RP-PCA SDF. Given that SDF loadings for RP-PCA are more strongly linked to returns, it is not surprising that the implied Sharpe ratio of the RP-PCA SDF of 0.50 is more than twice as high as the SR of the PCA-SDF of 0.24 (see Table 2).

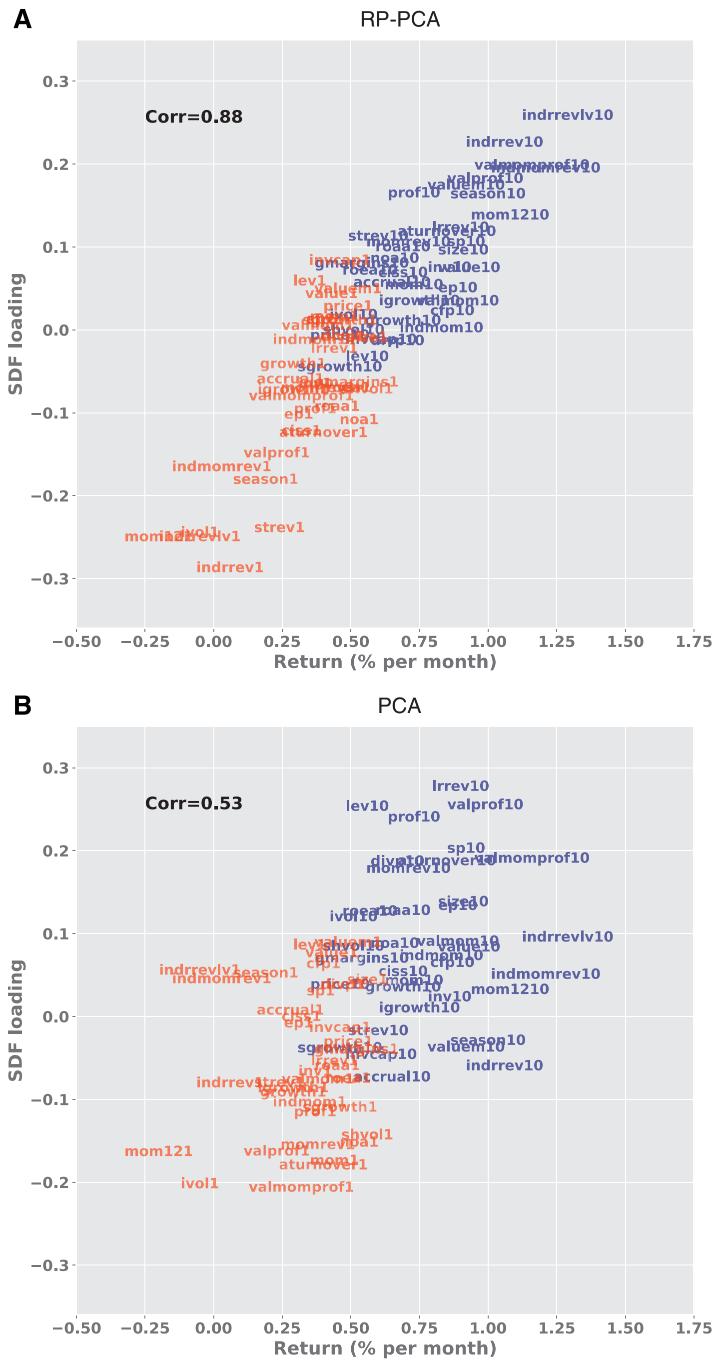


Figure 10
Portfolio returns and loadings of SDFs
This figure shows scatter plots of portfolio mean returns on the x-axis and SDF loadings on the y-axis for the first and last deciles of 37 single-sorted portfolios.

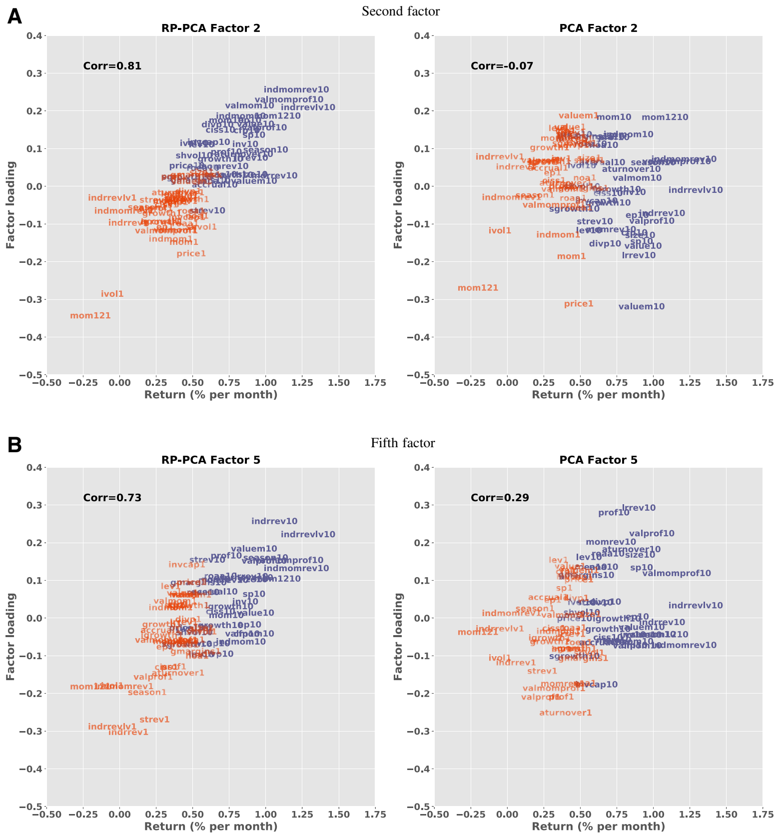


Figure 11
Portfolio returns and loadings of high-SR factors
This figure shows scatter plots of loadings of the second and fifth factors with portfolio mean returns for the first and last deciles of 37 single-sorted portfolios.

Recall from Table 3 that the second and fifth RP-PCA factors have high returns and Sharpe ratios. Figure 11 shows the same scatterplots for these factors as well as the second and fifth PCA factors for comparison. There is a strong link between portfolio returns and factor loadings of both factors, with a correlation of 0.81 for the second factor and 0.73 for the fifth factor. In contrast, PCA factor loadings show little correlation with returns, in particular for the second factor, which has a negative correlation.

These results illustrate the difference of factors extracted by RP-PCA versus PCA factors. First, RP-PCA factors have well-defined roles, some factors capture comovement but have low means and Sharpe ratios and the other factors have high means and Sharpe ratios but capture less comovement. In contrast, this separation is less clear for PCA factors. Second, the Sharpe ratio of SDF based on RP-PCA factors is much higher than that of the PCA SDF because the

RP-PCA weights are highly correlated with portfolio returns. The correlation is much lower for the PCA SDF.

It is important to stress that RP-PCA does not create factors and SDFs that mechanically result in high Sharpe ratios since the objective function depends on cross-sectional pricing errors rather than explicitly maximizing Sharpe ratios or mean returns. Instead, factors that reduce cross-sectional pricing errors turn out to also have high Sharpe ratios and means that are large and significantly different from zero.

2.2.6 Proxy factors and ad hoc long-short portfolios. To interpret latent factor models, Pelger and Xiong (2018) propose the use of proxy factors. The proxy factors use only the largest portfolio weights of the latent factors and set the smaller portfolio weights to zero. Pelger and Xiong (2018) show that the largest factor portfolio weights already contain most of the information signal even if the true factor itself is not sparse. We approximate the first latent factor by an equally weighted portfolio of all assets and use the ten largest portfolio weights (in absolute value) to approximate the second to fifth RP-PCA and PCA factors. Table 6 shows these portfolios along with their weights. Their interpretation is in line with Figure 9.

One of the major problems when comparing two different sets of factors is that a factor model is only identified up to invertible linear transformations. Two sets of factors represent the same factor model if the factors span the same vector space. When trying to interpret estimated factors by comparing them with economic factors, we need a measure to describe how close two vector spaces are to each other. Bai and Ng (2006) propose the generalized correlation (GC) as a natural candidate measure. Intuitively, GC measures the correlation between the latent factors and candidate factors after rotating them appropriately. If the qth generalized correlation is high, then the two sets of factors have (at least) q common factors.[16]

Table 7 shows the generalized correlations of the original factors with the proxy factors, that is, the correlations after rotating the latent factors. The generalized correlation of the first four proxy factors with the estimated RP-PCA and PCA factors is 0.91 and higher, confirming that the proxy factors provide a good approximation to the latent factors. The fifth GC is somewhat lower; 0.81 for RP-PCA and 0.85 for PCA. The next question is how well the proxy factors price the portfolios. Panel A of Table 8 shows the pricing statistics

[16] Let F be the K latent factors, and G are K_G candidate factors. To what degree can a linear combination of the candidate factors G replicate some or all of the factors F? The first generalized correlation is the highest correlation that can be achieved through a linear combination of the factors F and the candidate factors G. For the second generalized correlation we first project out the subspace that spans the linear combination for the first generalized correlation and then determine the highest possible correlation that can be achieved through linear combinations of the remaining $K-1$, respectively K_G-1 dimensional subspaces. This procedure continues until we have calculated the $\min(K,K_G)$ generalized correlation. Mathematically the generalized correlations are the square root of the $\min(K,K_G)$ largest eigenvalues of the matrix $\mathrm{Cov}(F,G)\mathrm{Var}(F)^{-1}\mathrm{Cov}(F,G)\mathrm{Var}(G)^{-1}$. If $K=K_G=1$, it is simply the correlation.

Table 6
Proxy factors: Largest 10 portfolio weights of factors 2 to 5

RP-PCA							
indmomrev10	2.78	valuem10	3.25	indmom1	1.71	indrrev10	1.48
valmomprof10	2.48	price1	2.87	mom1	1.66	invcap1	1.08
valmom10	2.31	mom121	2.19	valmom1	1.63	strev10	1.04
indrrevlv10	2.24	lrrev10	2.00	valmomprof1	1.54	valuem10	0.97
indmom10	2.00	value10	1.92	indmomrev1	1.54	indmomrev1	−1.02
mom1210	1.98	divp10	1.87	valmomprof10	−1.47	divp10	−1.04
ep10	1.87	sp10	1.74	mom10	−1.50	season1	−1.10
price1	−2.13	lev1	−1.63	price1	−1.76	strev1	−1.56
ivol1	−3.34	value1	−1.67	size10	−2.04	indrrev1	−1.63
mom121	−3.96	valuem1	−1.91	ivol1	−2.27	indrrevlv1	−1.71
PCA							
valuem1	1.89	valmom10	2.38	lev10	1.91	lrrev10	1.31
mom10	1.85	indmom10	2.09	divp10	1.88	prof10	1.25
mom1210	1.84	valmomprof10	1.89	ivol10	1.63	valprof10	1.00
value1	1.58	mom10	1.77	indmomrev1	1.57	momrev10	0.90
value10	−1.61	value10	1.51	shvol10	1.40	aturnover10	0.81
lrrev10	−1.87	valmomprof1	−1.56	roea1	−1.41	invcap10	−0.80
mom1	−1.89	indmom1	−1.85	shvol1	−1.69	valmomprof1	−0.84
mom121	−2.74	valmom1	−1.93	size10	−1.77	prof1	−0.94
price1	−3.17	mom1	−2.23	price1	−2.24	valprof1	−0.95
valuem10	−3.24	mom121	−2.25	ivol1	−2.67	aturnover1	−1.13

Portfolio composition of second to fifth proxy factors based on $N=74$ extreme deciles. The portfolio weights of the proxy factors are the 10 largest loadings of the latent factors. RP-weight $\gamma=10$.

Table 7
Correlation of proxy factors and long-short factors with original factors

	Proxy factors	
	RP-PCA	PCA
1.GC	1.00	1.00
2.GC	0.99	0.99
3.GC	0.97	0.97
4.GC	0.91	0.94
5.GC	0.81	0.85

The table shows the generalized correlation of the RP-PCA and PCA factors with the corresponding proxy factors.

for models with proxy factors in comparison with the statistics of the original factors in panel A. Overall, the results are very similar, confirming that the proxy factors retain most of the pricing information of the original factors.

How do the statistical RP-PCA and PCA factors relate to simple long-short factors (that is, with weights of 1 for decile 10, -1 of decile 1, and 0 otherwise)? To answer this question, we compute generalized correlations of statistical factors with long-short portfolios. The first LS factor is the market factor. LS factors are added incrementally based on the largest accumulative absolute loading of the anomaly in the portfolio weights of the statistical factors. The results are shown in Figure 12. Each panel shows five lines for the five statistical factors in each model/sample. The number of incremental factors is on the

Table 8
Alternative models

		In-sample				Out-of-sample		
Model (K)		SR	RMS_α	$\overline{\sigma}_e$		SR	RMS_α	$\overline{\sigma}_e$
Panel A: Proxy models								
RP-PCA (3)		**0.40**	**0.24**	**14.27%**		**0.34**	**0.22**	**15.47%**
PCA (3)		0.21	0.28	14.31%		0.14	0.29	16.01%
RP-PCA (5)		**0.54**	**0.18**	11.18%		**0.44**	**0.19**	11.94%
PCA (5)		0.29	0.22	**11.04%**		0.23	0.19	**11.89%**
Panel B: SDF based on factors sorted by means								
Model (K)	SR_{SMV}	SR_{KNS}	RMS_α	$\overline{\sigma}_e$	SR_{SMV}	SR_{KNS}	RMS_α	$\overline{\sigma}_e$
RP-PCA SMV (3)	**0.58**	0.47	**0.21**	17.09%	**0.35**	0.32	**0.21**	**16.90%**
PCA SMV (3)	0.26	0.22	0.24	**15.26%**	0.26	0.18	0.23	18.18%
RP-PCA SMV (5)	**0.66**	0.51	**0.16**	**12.65%**	**0.48**	0.39	**0.15**	**13.43%**
PCA SMV (5)	0.37	0.18	0.20	13.77%	0.33	0.24	0.18	15.23%

Panel A reports the Sharpe ratios, root-mean-squared pricing errors, and unexplained idiosyncratic variation of proxy models. Panel B reports results for models with factors chosen according to their means instead of their variances. SR_{SMV} is the Sharpe ratio for the case when SDF weights are set to their mean-variance values in Equation (14), while SR_{KNS} is based on the KNS estimator in Equation (13). Given K, pricing errors and the idiosyncratic variances are not affected by which SDF method is used. The KNS parameters are set to obtain the SR prior of 0.3, which is approximately the optimal choice in Kozak, Nagel, and Santosh (Forthcoming). The sample consists of 74 decile-1 and decile-10 portfolios.

x-axis, and the generalized correlation is on the y-axis. Consider first the RP-PCA model shown in the left panel. The purple line corresponding to the first RP-PCA factor shows a generalized correlation of close to one with the first LS factor (the market return) confirming the results described above that the first RP-PCA factor is essentially a market proxy. This is true for the first factors in all cases displayed in Figure 12. The generalized correlation of the second to fifth statistical factors with LS portfolios is generally significantly lower than one. For example, five LS portfolios are required to generate a generalized correlation with the second RP-PCA factor of 0.8 while about 10 LS portfolios are required for factors three and four. The fifth factor is the most difficult to proxy for with LS portfolios; a combination of 18 LS portfolios yields a generalized correlation of 0.8. These patterns are broadly similar for PCA factors. Statistical factors clearly yield more parsimonious representations of pricing factors than simple LS factors.

2.2.7 Results for robust SDF estimation. Next, we present results for robust SDF estimators introduced in Section 1.4, KNS in Equation (13) and the sparse mean-variance estimator (SMV) (14). Both methods select factors from a set of given PCA or RP-PCA factors according to their means instead of their variances. The SDF weights of the KNS estimators are shrunk toward zero, while weights of chosen factors are set to their mean-variance values in the

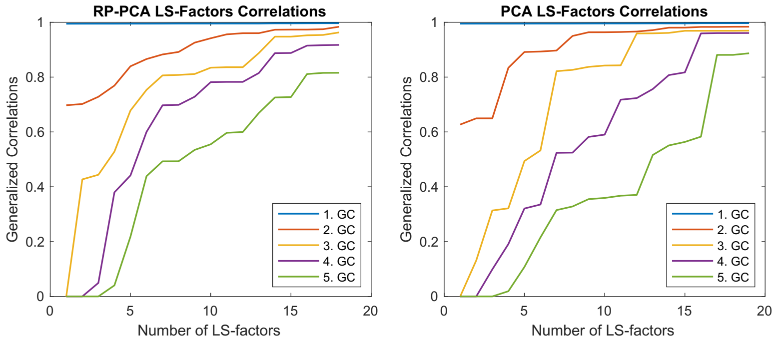


Figure 12
Generalized correlations or RP-PCA and PCA factors with long-short portfolios
First and last deciles of 37 single-sorted portfolios ($N=74$). Generalized correlation of $K=5$ statistical factors ($\gamma=10$) and an increasing number of long-short anomaly factors. The first LS factor is the market factor, and LS factors are added incrementally based on the largest accumulative absolute loading of the anomaly in the portfolio weights of the statistical factors.

SMV estimator.[17] Table 3 shows the first 10 RP-PCA and PCA factors as well as their rank when sorted by their means. Lower-order factors typically have high means, but there are exceptions. The seventh RP-PCA factor has a higher mean than the second factor. The second PCA has a particularly low mean, while the fifth factor has the fourth largest mean. The bold numbers indicate whether a factor is selected by five-factor KNS/SMV models and show that the factors with the highest means are selected.

Panel B in Table 8 reports the fit of three- and five-factor models with RP-PCA and PCA factors. In our sample, the Sharpe ratios based on sparse mean-variance estimation is higher in all cases than the KNS estimator that incorporates sparsity and shrinkage. This is in contrast to the results of KNS who find that shrinkage improves Sharpe ratios. The difference might be due to their use of market-orthogonalized returns, while we use excess returns.

When compared with the standard SR estimation in Table 2, we find that selecting factors based on means, as in SMV, improves the SR. All PCA-SMV specifications significantly outperform their PCA counterparts in terms of out-of-sample Sharpe ratios and pricing errors. The Sharpe ratio of three-factor PCA-SMV is more than twice as high as that of the PCA model (0.26 vs. 0.10) while the RMS pricing error is also lower. The same pattern is true for the five-factor models. The effect of SMV for RP-PCA is not as pronounced. The three-factor RP-PCA-SMV model outperforms the three-factor RP-PCA model, the difference is reversed for five-factor models. The Online Appendix collects additional results for KNS and SMV and confirms that these findings are robust.

17 As mentioned earlier, KNS use residuals from CAPM regressions as test assets, while we use simple excess returns.

The reason for the smaller effect for RP-PCA than for PCA is as follows. SMV chooses factors based on the means (and, implicitly, by their SR) rather than according to their variances. PCA ranks factors by their variance irrespective of their means and Sharpe ratios. With PCA factors as inputs, SMV selects the highest mean factors, so that the resulting SDF has a significantly higher mean and Sharpe ratio than the PCA-SDF for which factor means play no role. In contrast, RP-PCA factors are constructed to lower pricing errors and have higher means and SRs than PCA factors. Hence, SMV selects among factors that already have a high mean, and the marginal effect is much smaller than for PCA.

2.2.8 Individual stocks. Next, we briefly explore the ability of RP-PCA and PCA to capture returns of individual stocks rather than returns of portfolios. Recall that both methods assume that factor loadings are constant over time, and we showed that this assumption is satisfied in the case of portfolio returns. Intuitively, the assumption of constant loadings appears problematic in the case of individual stocks. For example, the turnover in sorts of some anomaly characteristics is very high. For example, a given stock might be in the highest momentum portfolio at some period but in a lower momentum portfolio in another period. If momentum is a priced factor, then the momentum loading of stocks is likely to change over time. To explore whether this is indeed the case, we estimate RP-PCA and PCA on a panel of large stocks.

Our sample consists of monthly excess returns of a balanced panel of stock returns from May 1972 to December 2013, which results in $N=270$ stocks with $T=500$ monthly returns. The data is obtained from CRSP. We include only stocks that have been constituents of the S&P 500 index at some point during the sample span and have no missing values during the time period that we consider. Extending to a longer time period would drastically reduce the number of stocks in our panel. By construction the stocks in our sample are mainly large cap stocks, and the results are not driven by small stocks.

Panel D of Table 2 and Figure 13 show the in-sample and out-of-sample SR, pricing errors and idiosyncratic variance $\overline{\sigma}_e^2$ for the stock sample. Recall that the corresponding results for the sample with decile-1 and decile-10 portfolio returns are in Figure 6. There are several important differences between the results for the stock versus portfolio samples. First, while the out-of-sample fits were close to their in-sample counterparts, the out-of-sample Sharpe ratios and pricing errors deteriorate compared with the in-sample results, especially for RP-PCA. In-sample, the IS Sharpe ratios for RP-PCA models with four factors or more are between 0.25 and 0.28, but are close to zero out-of-sample. The difference between IS and OOS Sharpe ratios is significant but less extreme for PCA models. The OOS RMS_α for RP-PCA and PCA are about 50% higher than in-sample (≈ 0.25 compared to ≈ 0.41) while the difference is smaller for $\overline{\sigma}_e^2$. Lastly, the three- and five-factor Fama-French models yield higher OOS Sharpe ratios than both PCA models and similar RMS_α and $\overline{\sigma}_e^2$.

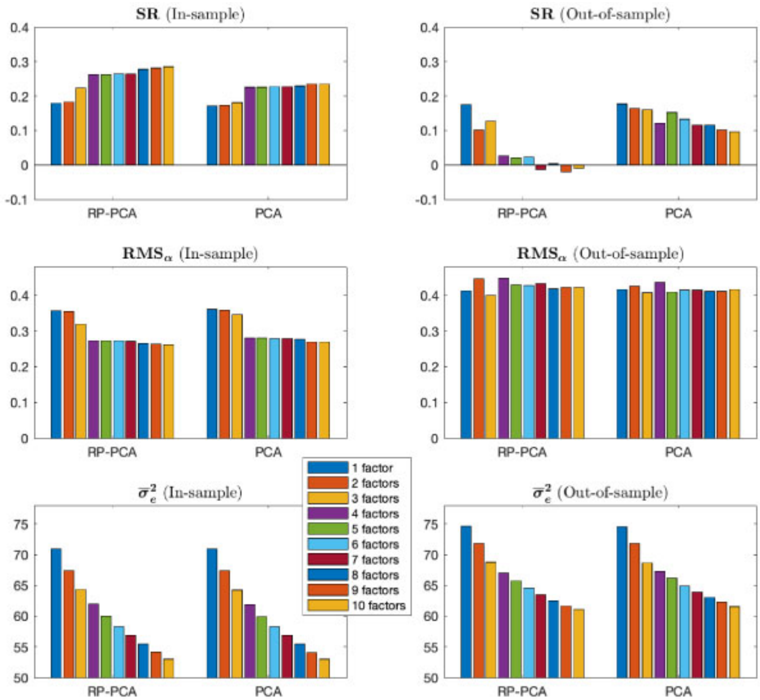


Figure 13
Fit for individual stocks
Maximal Sharpe ratios, root-mean-squared pricing errors, and unexplained idiosyncratic variation for different numbers of factors. RP weight $\gamma = 10$.

Clearly, the RP-PCA and PCA models do not perform well in the sample with individual stocks. RP-PCA models, in particular, extract factors and SDFs with high Sharpe ratios and low pricing errors in-sample and out-of-sample when applied to portfolio data but fit poorly out-of-sample. There are several possible reasons for this difference. First, stock returns are more volatile than portfolio returns, which makes estimation of means more imprecise. Since the RP-PCA estimator is based on means, it is likely more sensitive to noisy first moments than PCA models. Second, the signal-to-noise ratio is lower for stock returns than for portfolio returns, which makes the identification of common factors more difficult. Third, the factor structure is likely to be less stable for stocks than for portfolios. For example, the loading of an individual stock on a momentum factor changes over time, while loadings of rebalanced portfolios might be more stable.

To investigate the the stability of factors over time, we compute the generalized correlation of factors estimated using the whole sample with factors based on rolling subsamples. Figure 14 traces the generalized correlations across our portfolio (panel A) and individual stocks (panel B) samples. The first

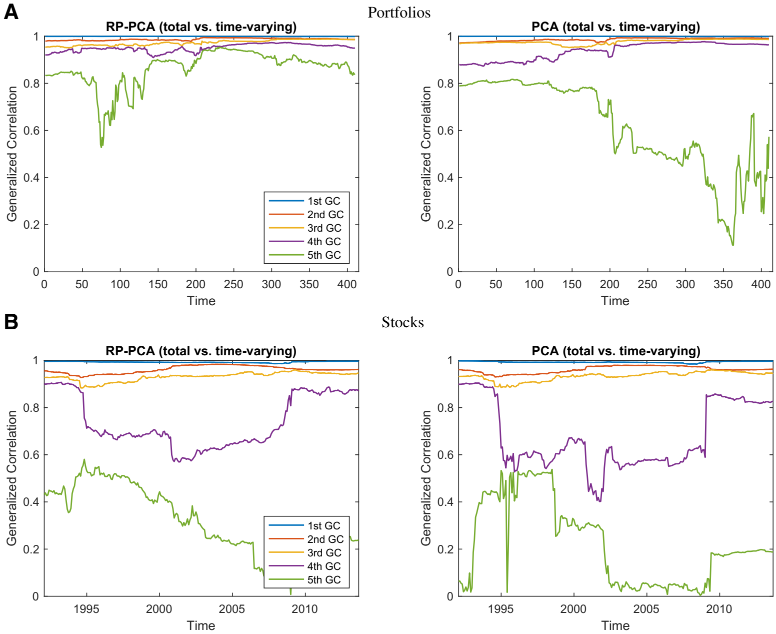


Figure 14
Generalized correlations of local loading estimates
Generalized correlations between loadings estimated on the whole sample and rolling windows with lengths of 240 months.

generalized correlations of the first four RP-PCA and PCA factors are above 0.9, indicating that the factor structure is stable over time. The fifth factor is less stable over time, in particular for PCA. The fifth generalized correlation drops to 0.15 toward the end of the sample. The fifth RP-PCA generalized correlation is above 0.9 in the second half of our sample but is lower in the first half. For individual stocks in panel B, the generalized correlations of the first three RP-PCAs and PCAs are also stable over time, in contrast to the correlations of the fourth and fifth factors. The generalized correlations of the fourth RP-PCA and PCA factors are between 0.5 and 0.9 but below 0.5 for the fifth factors, indicating significant instability of the factor structure over time. Hence, it is not surprising that the out-of-sample fit for the sample with individual stocks is worse than that for the sample with portfolio returns. In contrast, the Fama-French factors are based on fixed weights and need not to be estimated and perform as well as the RP-PCA and PCA models out-of-sample.

3. Conclusion

We propose a new estimator, risk premium PCA (RP-PCA), for latent asset pricing factors that combines information in the first two moments of asset returns. The explicit objective of RP-PCA is to extract factors that capture common time-series comovement as well as yield small cross-sectional pricing errors. RP-PCA can be expressed as a generalization of standard PCA with an additional penalty term on the cross-sectional pricing errors that is motivated by Ross' s APT. The RP-PCA estimator can be implemented using a standard eigenvalue composition of the variance-covariance matrix of asset returns after a simple transformation.

Our main conclusions are as follows. The time-series and cross-sectional moments of a sample of 370 single-sorted decile portfolios based on 37 different characteristics can be summarized by five RP-PCA factors. RP-PCA factor models yield higher Sharpe ratios and smaller pricing errors, in-sample as well out-of-sample, than PCA models without a significant drop in the captured time-series comovement. The five RP-PCA factors have clear economic interpretations. The first factor is a long-only and highly correlated with the CRSP-VW index with high mean and variance. Factors 2 and 5 have high means and Sharpe ratios, a relatively low variance with weights that are highly correlated with the mean returns of the test assets. In contrast, factors 3 and 4 have low means and Sharpe ratios, but high variances and are linked to value/growth portfolios (factor 3) and momentum and value/growth-interaction portfolios (factor 4). A model with the three high-mean/SR factors 1, 2, and 5 is sufficient to capture most of the cross-sectional return differences with pricing errors that are as small as those for the five-factor model. Moreover, a three-factor model with the three high-variance RP-PCA factors 1, 3, and 4 captures most of the common time-series comovement. The Sharpe ratio of the RP-PCA model is significantly higher than the Sharpe ratio of PCA factors. The reason is that the addition of the cross-sectional penalty in RP-PCA tilts the portfolio weights of the SDF toward characteristics with high return premia, whereas the SDF weights of PCA models are not correlated with return premia. We find that the empirical results are robust across a variety of samples and specifications.

The RP-PCA method can be easily combined with other PCA-based methods, such as the SDF estimation proposed by Kozak, Nagel, and Santosh (Forthcoming) and the instrumented-PCA estimator in Kelly, Pruitt, and Su (2017).[18] Both methods include an estimation of factors by PCA that can be substituted by RP-PCA without affecting other parts of the estimator. Replacing PCA with RP-PCA improves the fit substantially in both cases. In other words, RP-PCA is a complement to these methods rather than an alternative.

The separation of RP-PCA factors that capture comovement with low means and factors with high means but capture less comovement has implications for asset pricing theory. Low-mean but high-comovement factors indicate the presence of some, potentially unobserved, variable that drives comovement in returns but is only weakly correlated with priced shocks in the SDF. In contrast, high- mean but low-comovement factors must be driven by priced shocks that

[18] The Online Appendix includes results for IPCA combined with RP-PCA.

are largely orthogonal to systematic factors of returns. For example, differences in the exposure to priced shocks to higher- order moments can cause differences in expected returns while at the same time play little role in the comovement of returns. A detailed investigation of the implications for asset pricing models goes beyond the scope of this paper, but some potential ingredients of a model that matches the dichotomy of factors might include (i) an aggregate priced risk factor, say consumption growth, (ii) asset-specific cash flows that depend on the aggregate risk factor, a common factor that is uncorrelated with the the priced factor, plus an idiosyncratic shock, and (iii) priced stochastic volatility. The common but unpriced common component of cash flows corresponds to a low-mean factor that captures return comovement. In contrast, risk premia that are due to exposure to stochastic volatility shocks could yield high-mean factors that capture little return comovement. More generally, the investigation of the theoretical implications of the empirical results from the newly emerging literature on factor modeling presents a promising research agenda.

Appendix A. A Factor Model

In this paper we consider a simplified factor model to make the weak and strong factor models comparable. The assumptions of the strong factor models can be relaxed considerably without affecting the results, as shown in Lettau and Pelger (Forthcoming).

Assumption 1. Factor Model: The factor model (2) holds with the following assumptions:

(i) **Convergence rates:** $T, N \longrightarrow \infty$, $N/T \longrightarrow c$ with $0 < c < \infty$.

(ii) **Factors:** The factors F are uncorrelated, are independent of e and Λ and have bounded first two moments:

$$\hat{\mu}_F = \frac{1}{T}\sum_{t=1}^{T} \mathrm{F}_t \xrightarrow{p} \mu_F \qquad \hat{\mathbf{\Sigma}}_F = \frac{1}{T}\mathrm{F}_t \mathrm{F}_t^\top - \hat{\mu}_F \hat{\mu}_F^\top \xrightarrow{p} \mathbf{\Sigma}_F = \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_K^2 \end{pmatrix}$$

(iii) **Loadings:** The loadings Λ are independent of the factors and residuals.

(iv) **Residuals:** The matrix of residuals can be represented as $\mathrm{e} = \epsilon\, \mathbf{\Sigma}_e$ with $\epsilon_{t,i} \sim i.i.d.\, N(0,1)$.

A.1. Strong factor model

Traditional factor models as in Connor and Korajczyk (1988, 1993), Bai (2003), Bai and Ng (2002) and Stock and Watson (2002) are based on the assumption that factors are "strong" in the following sense. The "strengths" of systematic factors are given by their corresponding eigenvalues of the matrix $\Lambda^\top \Lambda$. The eigenvalues of strong factors diverge to infinity (while in the weak factor models considered later the eigenvalues are bounded). This is captured by the assumption that $\frac{1}{N}\Lambda^\top \Lambda \to \mathbf{\Sigma}_\Lambda$ where $\mathbf{\Sigma}_\Lambda$ is a full-rank matrix. Without loss of generality we will use the normalization $\mathbf{\Sigma}_\Lambda = \mathrm{I}_K$. The interpretation of the assumption $\frac{1}{N}\Lambda^\top \Lambda \to \mathrm{I}_K$ is that strong factors affect an infinite number of assets with non-vanishing loadings. One example of this factor structure is $\lambda_{nk} = N(1, 1/\sqrt{N})\, \forall n, k$, that is, the loading of each asset on each factor converges to 1 as N grows. The results show that in a strong factor model RP-PCA and PCA provide consistent estimators for the loadings and factors but the RP-PCA estimator of the loadings is more efficient than the PCA estimator. Furthermore, the assumptions required for the strong factor model are weaker than those imposed in the weak factor model.

Assumption 2. Simplified Strong Factor Model
Assume Assumption 1 holds and in addition:

(i) **Loadings:** $\Lambda^\top \Lambda / N \xrightarrow{p} \mathrm{I}_K$ and all loadings are bounded.
(ii) **Residuals:** All elements and all row sums of $\boldsymbol{\Sigma}_e$ are bounded.

The following result is proven in Lettau and Pelger (Forthcoming):

Proposition 1. Simplified Strong Factor Model Suppose that Assumption 2 holds. The RP-PCA estimator that minimizes the objective function (4) has the following properties:

(i) The factors and loadings estimators are consistent.
(ii) The asymptotic distribution of the factor estimates is independent of γ.
(iii) The asymptotic distribution of the estimates of the loadings of factor k is given by

$$\sqrt{T}\left(\mathrm{H}^\top \hat{\Lambda}_k - \Lambda_k\right) \longrightarrow N(0, \boldsymbol{\Omega}_k).$$

Closed-form expressions of $\boldsymbol{\Omega}_k$ and the rotation matrix H are given in Lettau and Pelger (Forthcoming).
(iv) The optimal γ that minimizes the asymptotic variance is $\gamma = 0$. Choosing $\gamma = -1$, that is, PCA with covariance matrix for factor estimation, is not efficient.

Simulations results here and in Lettau and Pelger (Forthcoming) show that efficiency gains of choosing $\gamma = 0$ are relatively minor. We conclude that strong factors are relatively easy to estimate by PCA-based methods. We now turn to the more interesting case when factors "weak."

A.2. Weak factor model

In contrast to strong factors, weak factors affect only a smaller fraction of the assets (or have a small variance). Formally, in a weak factor model, the matrix $\Lambda^\top \Lambda$ is bounded and converges to a fixed limit, in contrast to a strong factor model where $\frac{1}{N}\Lambda^\top \Lambda$ converges. Without loss of generality, we use the normalization $\Lambda^\top \Lambda \xrightarrow{p} \mathrm{I}_K$. Lettau and Pelger (Forthcoming) give a general characterization of weak factors. In this paper, we focus on an easily interpretable class of weak factors of the form $\Lambda_{nk} = N(0, 1/\sqrt{m})$ with probability m/N, $\Lambda_{nk} = 0$ with probability $1 - m/N$. For $m = N$, the factor k affects all assets, but the loadings converge to zero. If m is small, fewer assets are affected by the factors (in expectation) but with larger weights. A weak factor can thus be interpreted either as a factor with only a weak effect on many assets or a factor with strong effects on a few assets. Moreover for $m > 1$, factors of this form have the interpretation of long-short portfolios since the mean of all Λ_{nk} is zero, so that in expectation half the loadings are positive and half are negative.

The statistical model for analyzing weak factor models is considerably more complicated than that of the strong factor model and is based on spiked covariance models from random matrix theory (see Lettau and Pelger (Forthcoming) for details).[19] The distinction of strong and weak

19 The intuition is as follows. Under the assumptions of random matrix theory, the eigenvalues of a sample covariance matrix separate into two areas: (i) the bulk spectrum with the majority of the eigenvalues that are clustered together and (ii) some spiked large eigenvalues separated from the bulk. Under appropriate assumptions the bulk spectrum converges to the generalized Marčenko-Pastur distribution. The largest eigenvalues are estimated with a bias that is characterized by the Stieltjes transform of the generalized Marčenko-Pastur distribution. If the largest population eigenvalues are below some critical threshold, a phase transition phenomena occurs. The estimated eigenvalues will vanish in the bulk spectrum and the corresponding estimated eigenvectors will be orthogonal to the population eigenvectors. Onatski (2012) studies weak factor models and shows the phase transition phenomena for weak factors estimated with PCA. Our paper provides a solution to this factor-detection problem. It is important to notice that essentially all models in random matrix theory work with processes with mean zero. However, this assumption is not appropriate for factor models of asset returns. Lettau and Pelger (Forthcoming) develop new methods when this assumption is relaxed.

factors matters in practice since the estimation of weak factors is more difficult than the estimation of strong factors. Recall that strong factors and their loadings can be estimated consistently. In contrast, PCA-based estimators of weak factors are generally biased, even in the limit as T and N grow. For example, the correlation of an estimated weak factor with the "true" weak factor typically converges to a number less than one (see Proposition 2 below). Lettau and Pelger (Forthcoming) derive a set of results that show that RP-PCA dominates standard PCA in a number of dimensions. The optimal choice of γ is generally larger than -1 and in most cases larger than 0. We present some simulation evidence here and refer to Lettau and Pelger (Forthcoming) for more results. The bias reduction of choosing $\gamma > -1$ is often substantial.

Moreover, in the weak factor model RP-PCA can detect factors that are missed by conventional PCA. If a weak factor affects "too few" assets, if the loadings are "too small," or if the variance of the factor is "too low," it cannot be detected by PCA. However, the signal of RP-PCA depends on the mean and the variance of the factors. Thus, RP-PCA can detect weak factors with a high Sharpe ratio even if the factor is below the critical detection value for PCA. The results in Lettau and Pelger (Forthcoming) are summarized in the following proposition.

Assumption 3. Weak Factor Model Assume Assumption 1 holds and in addition:

(i) **Loadings:** $\Lambda^\top \Lambda \xrightarrow{p} \mathrm{I}_K$ and the column vectors of the loadings Λ are orthogonally invariant, *e.g.* $\Lambda_{1k} = N(0, 1/\sqrt{m})$ with prob. m/N, $\Lambda_{nk} = 0$ with probability $1 - m/N$.[20]

(ii) **Residuals:** The empirical eigenvalue distribution function of $\mathbf{\Sigma}_e$ converges almost surely weakly to a nonrandom spectral distribution function with compact support. The supremum of the support is b and the largest eigenvalues of $\mathbf{\Sigma}_e$ converge to b.

As mentioned earlier, factors of the form considered in Assumption 3 (i) have the interpretation of long-short portfolios that either affect many assets weakly, or few assets more strongly. Assumption 3 (ii) is a standard assumption in random matrix theory. The assumption allows for nontrivial weak cross-sectional correlation in the residuals, but excludes serial correlation.

The following proposition conveys the intuition of the results that are formally stated and proven in Lettau and Pelger (Forthcoming).

Proposition 2. Risk-Premium PCA under weak factor model
Suppose that Assumption 3 holds.

Let $\theta_1, \ldots, \theta_K$ be the first K largest eigenvalues of an appropriately defined "signal matrix" given in Lettau and Pelger (Forthcoming). The eigenvalues $\theta_k = \theta_k(\gamma)$ are strictly increasing functions of γ and have the interpretation of "signal strengths" of individual factors k. Let $\hat{F}_k$ denote the RP-PCA estimator of factor F_k. Then the correlation of $\hat{F}_k$ with the true factor F_k converges to

$$\mathrm{Corr}(\hat{F}_k, F_k) \xrightarrow{p} \begin{cases} \tau_k(\gamma) > 0 & \text{if } \theta_k(\gamma) > \theta_{\text{crit}} \\ 0 & \text{otherwise.} \end{cases}$$

θ_{crit} is a threshold that depends not on γ or the factors but only on the noise distribution.

$\tau_k(\gamma)$ has the following properties:

(i) For generic cases $\tau_k(\gamma) < 1$.

(ii) Typically, $\tau_k(\gamma)$ has a unique finite maximum $\gamma^* = \operatorname{argmax} \tau_k(\gamma)$ that yields the highest correlation of the estimated factor with the true factor.

[20] The normality assumption is not essential and can be relaxed to allow for a wider class of distributions (*e.g.* all spherical distributions)

(iii) If $\mu_F \neq 0$, then $\gamma^* > -1$ and RP-PCA yields factor estimates that are more highly correlated with the true factors than PCA with $\gamma = -1$. If $\mu_F = 0$, then $\gamma^* = -1$.

Proposition 2 states that weak factors can only be estimated with a bias even as N and K grow. If a factor is too weak, that is, its signal strength $\theta_k(\gamma)$ is below the threshold θ_{crit}, then it cannot be detected at all. Since $\theta_k(\gamma)$ is increasing in γ, RP-PCA is more likely to detect factors with lower signal strength than PCA with $\gamma = -1$.

γ also affects the correlations of estimated factors with true factors. Typically, there is a finite γ that maximizes the correlation of detected factors with true factors. In particular, RP-PCA with $\gamma > -1$ yields estimated factors that are more highly correlated with true factors than PCA with $\gamma = -1$.[21] The factor means μ_F play a crucial role for this result. If the factors have means of zero, $\mu_F = 0$, PCA with $\gamma = -1$ is optimal. If $\mu_F \neq 0$, then RP-PCA with $\gamma > -1$ is optimal. This distinction is important since asset pricing factors have nonzero means.

Figure 1 plots the function $\tau_k(\gamma)$ for parameters used in the simulation exercise described below.[22] The variance of the factor in the top panel is set to 0.05 and to 0.1 in the bottom panel. Ceteris paribus, a higher factor variance increases the signal strength of the factor. We also consider three positive values for the Sharpe ratio (or, equivalently, the mean) of the factor, as well as a case with mean-0 factor. The vertical lines indicate the γ that maximizes $\tau_k(\gamma)$[23]. In all cases with positive Sharpe ratios, $\tau_k(\gamma)$ increases sharply with γ for small γ and then flattens out as γ increases. For fixed factor variance, the gains in the correlation of the estimated and true factors increase with the Sharpe ratio (and mean) of the factor. Comparing the two panel shows that the optimal γ is higher if the signal strength of the factor is lower (*that is,* the factor variance decreases) and if the Sharpe ratio increases. If the Sharpe ratio (or the mean) of a factor is zero, then $\tau_k(\gamma)$ is decreasing in γ and the optimal weight $\gamma = -1$, as stated in Proposition 2. Note that in this case $\tau_k(\gamma) = 0$ if γ is large enough. The reason is that the factor mean does not contain any useful information, but more and more weight is put on the uninformative cross-sectional errors. Hence, in this case RP-PCA performs strictly worse than PCA.

How should the optimal γ be chosen? Based on our asymptotic theory, there are two possible criteria. First, a very large γ maximizes the probably of detecting a weak factor but lowers the correlation with the true factor. Alternatively, a moderately high γ maximizes this correlation. Another aspect of choosing γ is the out-of-sample performance of the RP-PCA estimator. As we will show in the next section, RP-PCA estimation deteriorates out-of-sample if γ is chosen too high. In practice, we start with $\gamma = -1$ and then increase γ until the OOS fit is no longer improved.

Appendix B. Simulations

We use a Monte Carlo simulation to study the behavior of the RP-PCA estimator in samples of the size typically encountered in practice. We focus on the effects of γ, the signal strength and excess returns of factors (Lettau and Pelger (Forthcoming) conduct a more comprehensive simulation study).

We consider a two-factor model, where the first factor mimics a strong "market" factor that affects all assets with an average loading of one. The second factor is a weak long-short factor with

[21] Intuitively, the phase transition phenomena that hides weak factors can be avoided by putting some weight on the information captured by the risk premium.

[22] We have estimated the sparse residual correlation matrix with a thresholding approach based on $N = 370$ deciles sorted portfolios as described in the empirical Section 4. The other parameters are the same as in the simulation section.

[23] We calculate the smallest value of γ such that $\tau_k(\gamma)$ is less than 0.5% from its maximum value as $\tau_k(\gamma)$ can be essentially flat for large values of γ and there is practically no gain in increasing γ.

an average loading of zero: $\Lambda_{n,1}=1, \Lambda_{n,2} \sim N(0,1)$. The mean and standard deviation of the market factor are $\mu_1=0.5\%, \sigma_1=4.5\%$ and the Sharpe ratio is 0.11. The signal strength of the weak second factors depends on its variance σ_2^2 and Sharpe ratio. We consider a variety of different values for σ_F^2 and SR_F to assess the RP-PCA estimation of weak factors.

The performance of RP-PCA is evaluated by the following criteria: (i) the correlation between the estimated and "true" factors, $\rho(\widehat{F}_k, F_k)$, (ii) the Sharpe ratios of the estimated factors, $\widehat{\mathrm{SR}}_k$, and (iii) the root-mean-squared pricing errors of time-series regressions of returns on estimated factors, $\mathrm{RMSE}=\sqrt{1/N\sum_n \alpha_n^2}$. We normalize the RMSE of the model with estimated factors by the RMSE for the true factors. To conserve space we report only out-of-sample results where we first estimate the loading vector in-sample and then obtain the out-of-sample factor estimates by projecting the out-of-sample returns on the estimated loadings. The sample size of $N=74$ and $T=650$ observations is based on our empirical study. Finally, the correlation matrix of our simulated residuals is set to the empirical correlation that we observe in the data.[24]

We first consider estimation of the strong first factor. Figure D.1 shows the evaluation criteria as a function of γ. Panel A shows that the correlation of the estimated factor and the true factor $\rho(\widehat{F}_1, F_1)$ is close to one, indicating that unobservable strong factor is estimated with high precision for all values of γ. Panel B shows that the estimated Sharpe ratio $\widehat{\mathrm{SR}}_1$ is close to the true SR of the strong factor (0.11) for all values of γ. This confirms the theoretical results in Proposition 1 that stated that strong factors can be estimated consistently and that γ only affects the efficiency of the estimator. We conclude that strong factors are relatively easy to estimate by PCA-based methods. We next consider the more interesting case of weak factors.

Figure D.2 has the same layout as Figure D.1 but considers a range of values for the variance and Sharpe ratio of the weak factor: $\sigma_2^2=0.05, 0.1, 0.5$ and $\mathrm{SR}_2=0.1, 0.3, 0.6$. The panel columns have different values of σ_2^2, and each panel has four lines corresponding to four different values of SR_2. Recall that the signal strength of a factor increases in σ_2^2 and SR_2, so a factor with a small variance but a high Sharpe ratio is detectable by RP-PCA as long as γ is sufficiently high. Consider a factor with a low variance $\sigma_2^2=0.05$ (left column in Figure D.2). The top panel shows that the correlation $\rho(F_2, \widehat{F}_2)$ is essentially zero if the Sharpe ratio is also low ($\mathrm{SR}_2=0.1$). This is an example of an undetectable factor that is below the signal threshold in Proposition 2. As the Sharpe ratio of the weak factor increases, the factor becomes detectable as long as γ is set sufficiently high. The intuition is that a nonzero Sharpe ratio contains information that can be exploiting by putting more weight on the mean, that is, increasing γ. Hence a factor that is missed by standard PCA with $\gamma=-1$ can be detected by RP-PCA. However, $\rho(F_2, \widehat{F}_2)$ is below one in all cases showing that in contrast to strong factors, weak factors are not fully identifiable as shown theoretically in Proposition 2.

The panel shows that the estimated Sharpe ratio $\widehat{\mathrm{SR}}_2$ of the second factor increases with γ. This is intuitive since a higher γ allows for a better identification of the weak factor. Finally, the bottom panel shows that the RMSE time-series error decreases slightly with γ if $\mathrm{SR}=0.6$. If the Sharpe ratio is lower, coupled with a low variance of the second factor relative to the variance of the first factor (0.05 vs. 0.21), the true model is dominated by the market factor and identifying the second factor decreases the RMSE only marginally.

The middle column of Figure D.2 considers a larger σ_2^2 of 0.1. The patterns are similar as for $\sigma_2^2=0.05$, but the effect of a higher γ is more pronounced. The effect of a higher γ is especially

---

24 In more detail, we have estimated the residual correlation matrix based on $N=74$ extreme deciles sorted portfolios as described in the Section 2. In each case we have first regressed out the systematic factors and then estimated the residual covariance matrix with a hard thresholding approach setting small values to zero, see Bickel and Levina (2008) and Fan, Liao and Mincheva (2013). This provides a consistent estimator of the residual population covariance matrix. We have regressed out the first 6 PCA factors. Our results remain unchanged when we calculate residuals based on more PCA factors or using RP-PCA factors. The additional results are available upon request. The remaining correlation structure in the residuals is sparse. In particular, the estimated eigenvalues of the simulated residuals coincide with the empirical estimates of the eigenvalues. The average idiosyncratic noise level is estimated from the data and set to $\sigma_e^2=4$.

pronounced if the Sharpe ratio of the weak factor is high. For $\mathrm{SR}_2=0.6$, the improvements along all three evaluation criteria are substantial.

Since the signal strength of the second factor is higher, $\rho(F_2, \widehat{F}_2)$ and $\widehat{\mathrm{SR}}_1$ are larger than in the left column as long as γ is sufficiently high. The effect of a higher γ on the normalized RMSE is larger than in the left panel with smaller factor variance. Note that standard PCA misses detecting this factor completely, and thus RP-PCA strictly dominates PCA for $\gamma \gtrsim 5$.

In the right columns σ_2^2 is increased further to 0.5. Now the signal strength is sufficiently high that the second factor is better characterized as strong rather than weak and thus can be estimated regardless of γ. This example shows that RP-PCA is particularly powerful for factors in the intermediate range between very weak factors that are undetectable and factors that are strong enough to be detectable by standard PCA. We argue in the empirical section that in the data the market index is a strong factor while any additional factors are better characterized as weak, so they might be missed by standard PCA but strong enough to be estimated by RP-PCA.

In Figure D.3, we plot sample paths of one representative simulation run. Each panel shows the paths of cumulative returns of the true factor as well as the cumulative returns of estimated RP-PCA factors for a range of γ's. The estimated sample paths of the strong factor in the first panel are close to the sample path of the true factor, which is not surprising since the correlations of the estimated factors with the true factor is essentially equal to one, as discussed previously. The pattern is different for the second (weak) factor in the second panel since the sample paths of the factor estimates diverge from the path of the true factor. As predicted by Proposition 2, the paths are closer to the path of the true factor for higher γ. In particular, the estimated PCA factor with $\gamma=-1$ has no resemblance to the path of the true factor and thus the weak factor is unidentified by PCA.

Appendix C. Tables

Table C.1
Single-sorted portfolios

Portfolio	Abbreviation	Mean	SD	*SR*	Mean		RP-PCA		PCA	
					Dec1	Dec10	Dec1	Dec10	Dec1	Dec10
Ind. Rel. Rev. (L.V.)	indrrevlv	1.28	3.05	0.42	0.00	1.28	−0.29	0.20	0.07	0.09
Industry Mom. Rev.	indmomrev	1.17	3.44	0.34	0.07	1.24	−0.18	0.12	0.07	0.06
Industry Rel. Reversals	indrrev	1.01	4.12	0.24	0.08	1.09	−0.29	0.22	−0.07	−0.05
Seasonality	season	0.79	3.94	0.20	0.22	1.01	−0.19	0.15	0.04	−0.01
Value-Profitability	valprof	0.75	3.81	0.20	0.26	1.01	−0.17	0.15	−0.17	0.28
Momentum (12m)	mom12	1.25	6.88	0.18	−0.16	1.08	−0.17	0.10	−0.14	0.07
Value-Mom-Prof.	valmomprof	0.83	4.83	0.17	0.35	1.18	−0.07	0.15	−0.19	0.20
Investment/Assets	inv	0.47	3.06	0.15	0.40	0.86	−0.06	0.04	0.02	0.06
Composite Issuance	ciss	0.48	3.30	0.14	0.30	0.78	−0.12	0.00	0.02	0.07
Investment Growth	igrowth	0.38	2.71	0.14	0.33	0.71	−0.05	0.02	−0.08	0.00
Sales/Price	sp	0.52	4.28	0.12	0.41	0.93	0.04	0.06	−0.03	0.24
Earnings/Price	ep	0.57	4.69	0.12	0.33	0.90	−0.07	−0.03	−0.10	0.18
Net Operating Assets	noa	0.38	3.28	0.12	0.26	0.64	−0.12	0.10	−0.11	0.00
Accrual	accrual	0.35	3.14	0.11	0.31	0.66	−0.02	0.05	0.06	−0.08
Value (A)	value	0.48	4.56	0.11	0.45	0.93	0.09	−0.01	0.04	0.10
Gross Profitability	prof	0.36	3.38	0.11	0.38	0.75	−0.17	0.19	−0.23	0.19
Asset Turnover	aturnover	0.41	3.83	0.11	0.30	0.71	−0.20	0.12	−0.32	0.24
Value-Momentum	valmom	0.51	5.05	0.10	0.40	0.91	0.07	−0.06	−0.04	0.12
Cash Flows/Price	cfp	0.43	4.37	0.10	0.42	0.85	0.06	−0.05	0.04	0.12
Momentum-Reversals	momrev	0.46	4.84	0.10	0.42	0.88	−0.06	0.09	−0.06	0.08
Asset Growth	growth	0.29	3.46	0.08	0.41	0.70	−0.02	−0.03	−0.04	0.02
Long Run Reversals	lrrev	0.41	5.07	0.08	0.47	0.89	0.00	0.11	0.00	0.18
Industry Momentum	indmom	0.47	6.21	0.08	0.34	0.81	0.03	−0.06	−0.06	0.10
Idiosyncratic Volatility	ivol	0.54	7.16	0.08	−0.01	0.53	−0.18	−0.04	−0.19	0.07
Value (M)	valuem	0.40	5.86	0.07	0.52	0.92	0.08	0.14	0.06	−0.05
Short-Term Reversals	strev	0.36	5.27	0.07	0.26	0.62	−0.26	0.12	−0.06	−0.02
Size	size	0.29	4.80	0.06	0.46	0.75	−0.03	0.09	0.00	0.16
Momentum (6m)	mom	0.35	6.25	0.06	0.58	0.93	0.07	0.00	−0.12	0.07
Leverage	lev	0.26	4.63	0.06	0.48	0.73	0.10	−0.09	0.04	0.22
Return on Assets (A)	roaa	0.21	4.08	0.05	0.37	0.58	−0.08	0.12	−0.16	0.09
Dividend/Price	divp	0.18	5.08	0.04	0.49	0.67	0.02	−0.11	0.08	0.13
Investment/Capital	invcap	0.12	5.01	0.02	0.57	0.68	0.15	−0.09	−0.06	−0.01
Return on Book Equity (A)	roea	0.08	4.39	0.02	0.51	0.59	−0.07	0.08	−0.12	0.13
Sales Growth	sgrowth	0.05	3.64	0.01	0.58	0.53	0.03	−0.08	−0.07	−0.06
Gross Margins	gmargins	0.01	3.37	0.00	0.55	0.56	−0.10	0.10	−0.01	−0.01
Share Volume	shvol	0.02	5.96	0.00	0.49	0.47	−0.03	−0.05	−0.08	0.06
Price	price	0.01	6.80	0.00	0.50	0.49	0.07	−0.03	−0.05	0.01

Long-short portfolios of extreme deciles of 37 single-sorted portfolios from November 1963 to December 2017: Mean, standard deviation and Sharpe ratio, mean of low and high portfolios and contribution to the SDF based on RP-PCA and PCA with Equation (5) applied to the extreme decile portfolios ($N=74$).

Appendix D. Figures

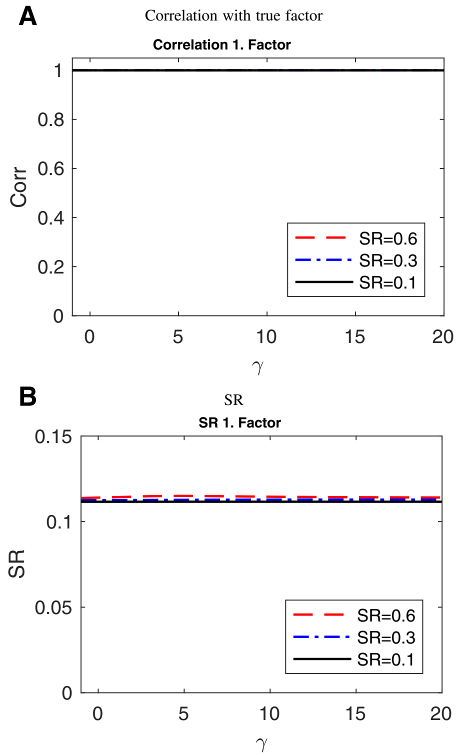


Figure D.1
Simulation – first factor
Correlations and Sharpe ratios as a function of γ for different variances and Sharpe ratios for the first (strong) factor.

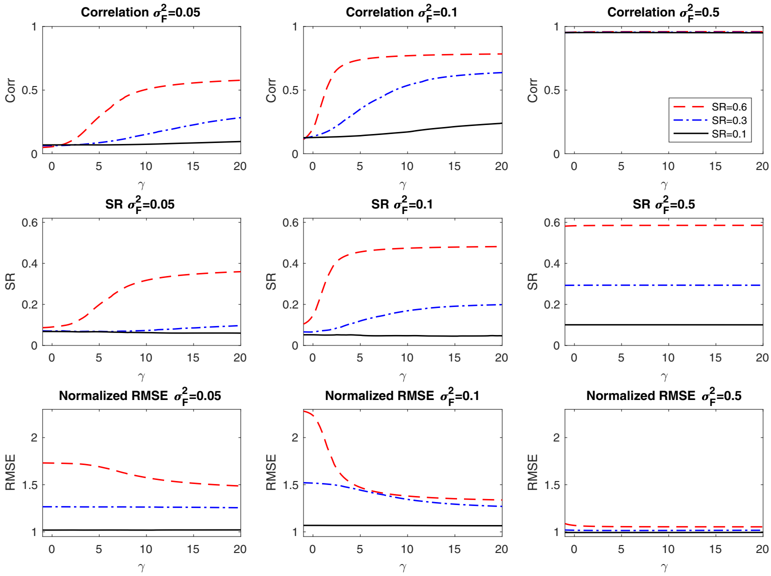


Figure D.2
Simulation – second factor
Out-of-sample correlations, Sharpe ratios, and normalized RMSE for the second (weak) factor as a function of γ for different variances and Sharpe ratios.

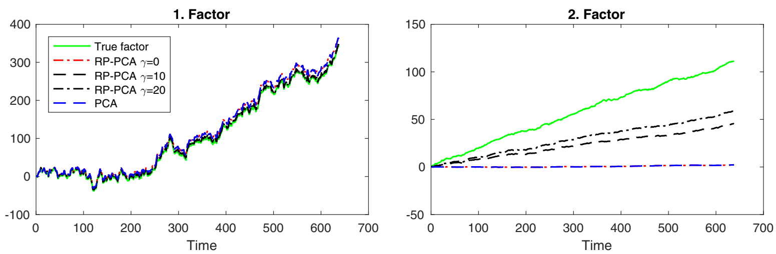


Figure D.3
Simulation - factor realizations
Sample path of one representative simulation for the cumulative returns of the true factor as well as estimated factors for different values of γ.

References

Bai, J. 2003. Inferential theory for factor models of large dimensions. *Econometrica* 71:135–71.

Bai, J., and S. Ng. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70:191–221.

———. 2006. Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics* 131:507–37.

———. 2019. Rank regularized estimation of approximate factor models. *Journal of Econometrics* 212:78–96.

Benaych-Georges, F., and R. R. Nadakuditi. 2011. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics* 227:494–521.

Cochrane, J. H. 2011. Presidential address: Discount rates. *Journal of Finance* 66:1047–108.

Connor, G., and R. Korajczyk. 1988. Risk and return in an equilibrium APT: Application to a new test methodology. *Journal of Financial Economics* 21:255–89.

———. 1993. A test for the number of factors in an approximate factor model. *Journal of Finance* 58:1263–91.

Connor, G., and R. A. Korajczyk. 1986. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics* 15:373–94.

Daniel, K., L. Mota, S. Rottke, and T. Santos. 2020. The cross-section of risk and return. *Review of Financial Studies* 33:1927–79.

DeMiguel, V., L. Garlappi, F. Nogales, and R. Uppal. 2017. A portfolio perspective on the multitude of firm characteristics. Working Paper.

Fama, E. F., and K. R. French. 2020. Comparing cross-section and time-series factor models. *Review of Financial Studies* 33:1891–1927.

Fan, J., Y. Liao, and W. Wang. 2016. Projected principal component analysis in factor models. *The Annals of Statistics* 44:219–54.

Feng, G., S. Giglio, and D. Xiu. Forthcoming. Taming the factor zoo. *Journal of Finance*.

Freyberger, J., A. Neuhierl, and M. Weber. 2020. Dissecting characteristics nonparametrically. *Review of Financial Studies* 33:2326–77.

Giglio, S., and D. Xiu. 2019. Asset pricing with omitted factors. Working Paper.

Harvey, C. R., Y. Liu, and H. Zhu. 2016. ... and the cross-section of expected returns. *Review of Financial Studies* 29:5–68.

Hou, K., C. Xue, and L. Zhang. 2015. Digesting anomalies: An investment approach. *Review of Financial Studies* 650–705.

Kelly, B., S. Pruitt, and Y. Su. 2017. Instrumented principal component analysis. Working Paper.

Koijen, R. S., H. Lustig, and S. Van Nieuwerburgh. 2017. The cross-section and time series of stock and bond returns. *Journal of Monetary Economics* 88:50–69.

Kozak, S., S. Nagel, and S. Santosh. Forthcoming. Shrinking the cross section. *Journal of Financial Economics*.

Lettau, M., and M. Pelger. Forthcoming. Estimating latent asset pricing factors. *Journal of Econometrics*.

Onatski, A. 2012. Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics* 244–58.

Pelger, M. Forthcoming. Understanding systematic risk: A high-frequency approach. *Journal of Finance*.

Pelger, M., and R. Xiong. 2018. Interpretable proximate factors for large dimensions. Working Paper.

Ross, S. A. 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13:341–60.

Stock, J. H., and M. W. Watson. 2002. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97:1167–79.

Zou, H., and T. Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67:301–20.