Covariance Implied Risk Factors

Mohammed Mehdi Kaebi[1]
Insper

Abstract

This paper examines the role of heteroskedasticity in extracting latent risk factors from asset returns. I show that standard principal component analysis suffers from distortions when assets exhibit heterogeneous idiosyncratic variances, leading factors to reflect clusters of idiosyncratic risk rather than true systematic risk. Building on recent developments in the statistics literature, I apply heteroskedastic PCA (heteroPCA) to correct for this bias by iteratively replacing the diagonal of the sample covariance matrix with estimates implied by the off-diagonal structure. This approach delivers superior out-of-sample cross-sectional pricing performance compared to standard PCA, with higher Sharpe ratios and lower average pricing errors across multiple equity portfolios. The identified factors exhibit clear economic interpretability, and the implied stochastic discount factor achieves lower Hansen-Jagannathan distances. These results highlight that accounting for heteroskedasticity in idiosyncratic variance substantially improves factor estimation in asset pricing applications.

JEL Classification: G11, G12, C38

Keywords: Factor models, principal component analysis, heteroskedasticity, asset pricing, latent factors.

---

[1] Quatá Street, 300 – Vila Olímpia, São Paulo – SP, Brazil. Email: mohammedmk@al.insper.edu.br.

1. Introduction

A fundamental insight in asset pricing theory is that expected returns should be explained by exposure to systematic risk factors. Identifying these factors has become a central pursuit in empirical finance, leading to what Harvey et al., 2015 characterize as a "factor zoo", with over 300 proposed risk factors. Recent work has sought to bring order to this chaos by summarizing pricing information with a small number of latent factors (Kozak et al., 2020; Lettau and Pelger, 2020; He et al., 2023; Bryzgalova et al., 2023). The goal is to preserve interpretability and stability while capturing the priced sources of risk.

Many of these approaches rely on principal component analysis (PCA), which is appealing because it is nonparametric, computationally tractable, and scales easily to large cross-sections and time series. However, standard PCA operates on the sample covariance matrix and implicitly assumes that idiosyncratic errors are homoskedastic. When assets exhibit heterogeneous idiosyncratic variances, as is common in equity, foreign exchange, and other markets, PCA factors become distorted. Because PCA maximizes total explained variance, it loads disproportionately on high-variance assets even when their variance reflects idiosyncratic noise rather than systematic risk (Zhang et al., 2022; Bai and Ng, 2002). Consequently, the estimated factors may reflect clusters of idiosyncratic risk instead of the underlying systematic components.

This paper addresses this problem by applying heteroskedastic PCA (heteroPCA), a recently developed method from the statistics literature (Zhang et al., 2022), to the estimation of latent risk factors in finance. HeteroPCA corrects for heteroskedasticity in idiosyncratic errors by iteratively replacing the diagonal of the sample covariance matrix with estimates implied by a low-rank approximation of the off-diagonal structure. The key insight is that systematic information resides primarily in the covariances (off-diagonal elements), while heteroskedastic noise inflates the variances (diagonal elements). By downweighting the diagonal and focusing on the systematic covariance structure, heteroPCA delivers more accurate factor estimates.

My contribution is threefold. First, I demonstrate that heteroPCA yields superior out-of-sample cross-sectional pricing performance compared to standard PCA. Across multiple equity portfolios sorted on size, book-to-market, accruals, investment, operating profitability, and other characteristics, heteroPCA delivers higher Sharpe ratios and lower root mean squared pricing errors. These improvements are economically meaningful: the method identifies factors with clearer economic content and achieves better risk-return tradeoffs.

Second, I show that the factors identified by heteroPCA exhibit stronger economic interpretability. In cross-sections with known factor structures, such as portfolios sorted on size and accruals, heteroPCA cleanly identifies market, size, and accrual factors, while standard PCA struggles to isolate these dimensions. By filtering out idiosyncratic noise, heteroPCA uncovers systematic risk that remains hidden under standard methods.

Third, I examine the stochastic discount factor (SDF) implied by the estimated factors. HeteroPCA consistently delivers lower Hansen-Jagannathan distances, indicating better cross-sectional pricing accuracy. The SDF-mimicking portfolios exhibit intuitive patterns: for instance, in size-accrual portfolios, heteroPCA identifies a clear quality-versus-junk tilt, shorting small high-accrual stocks and holding low-accrual positions, whereas standard PCA produces less interpretable ”size barbell” patterns.

My findings highlight the practical importance of accounting for heteroskedasticity when extracting latent factors. Heteroskedastic idiosyncratic variances are pervasive across asset classes. In equity markets, firm size, industry composition, leverage, and growth opportunities all contribute to heterogeneous volatility (Fama and French, 1992; Campbell et al., 2001; Ang et al., 2006). In foreign exchange markets, interest rate differentials, sensitivity to global macro shocks, and liquidity differences generate varying levels of idiosyncratic risk across currencies (Menkhoff et al., 2012; Andersen et al., 2003; Della Corte et al., 2016). Even when assets are grouped into portfolios, as is standard practice, characteristic-driven volatility persists. My results demonstrate that correcting for this heteroskedasticity leads to more reliable and interpretable risk factors.

This paper contributes to several strands of the literature. First, it relates to the growing body of work on latent factor models and PCA-based methods in asset pricing, including instrumented PCA (Kelly et al., 2019), projected PCA (Fan et al., 2016), and risk premium PCA (Lettau and Pelger, 2020). My approach is complementary: by addressing heteroskedasticity, heteroPCA can be integrated into these existing frameworks to improve factor estimation. Second, my work connects to the literature on heteroskedasticity in asset returns and its implications for estimation (Zhang et al., 2022; Bai and Ng, 2002). I show that the statistical insights from this literature have direct and substantial implications for empirical asset pricing. Third, my results relate to the literature on stochastic discount factors and Hansen-Jagannathan bounds (Hansen and Jagannathan, 1991), demonstrating that better factor estimation translates into improved SDF performance.

2. Theoretical framework and methodology

This section describes the approximate factor model underlying my analysis, explains how heteroskedasticity distorts standard PCA, and presents the heteroPCA methodology for correcting this bias.

2.1. The approximate factor model

I assume that the excess returns of N assets over T time periods follow an approximate factor model with K latent factors:

$$r_{nt} = \lambda_n^\top f_t + \varepsilon_{nt}, \qquad n = 1, \ldots, N, \;\; t = 1, \ldots, T, \tag{1}$$

where r_{nt} is the excess return on asset n at time t, $f_t \in \mathbb{R}^K$ is the vector of common factors, $\lambda_n \in \mathbb{R}^K$ is the vector of factor loadings for asset n, and ε_{nt} is the idiosyncratic error with $\mathbb{E}[\varepsilon_{nt}] = 0$ and $\mathrm{Var}(\varepsilon_{nt}) = \sigma_n^2$. In matrix form:

$$\underbrace{R}_{T\times N} = \underbrace{F}_{T\times K}\underbrace{\Lambda^\top}_{K\times N} + \underbrace{E}_{T\times N}, \tag{2}$$

where R contains asset returns, F the factor realizations, Λ the factor loadings, and E the idiosyncratic errors.

The approximate factor model allows for weak cross-sectional dependence in the idiosyncratic errors, but requires that the common component $\Lambda\Sigma_F\Lambda^\top$ dominates asymptotically as $N \to \infty$, where $\Sigma_F = \mathrm{Var}(f_t)$ is the factor covariance matrix. Under the assumption that factors and errors are uncorrelated ($\mathbb{E}[\varepsilon_t f_t^\top] = 0$), the covariance matrix of returns decomposes into systematic and idiosyncratic components:

$$\Sigma = \underbrace{\Lambda\,\Sigma_F\,\Lambda^\top}_{\text{systematic}} + \underbrace{\Sigma_E}_{\text{idiosyncratic}}, \qquad \Sigma_E = \mathrm{Var}(E_t). \tag{3}$$

The sample covariance matrix is given by

$$\widehat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(r_t - \bar{r})(r_t - \bar{r})^\top, \tag{4}$$

where $\bar{r}$ is the time-series mean. Standard PCA extracts factors by computing the top K eigenvectors of $\widehat{\Sigma}$, which estimates the loadings Λ. The factor realizations F are then obtained by projecting returns onto these estimated loadings.

2.2. The problem of heteroskedasticity

The validity of standard PCA for factor extraction relies on the implicit assumption that idiosyncratic variances are homogeneous across assets. When this assumption holds, the largest eigenvalues of $\widehat{\Sigma}$ are driven primarily by the systematic component $\Lambda\Sigma_F\Lambda^\top$, and the top eigenvectors of $\widehat{\Sigma}$ converge to those of the true systematic covariance structure.

However, when assets exhibit heteroskedastic idiosyncratic errors, this convergence breaks down. Because PCA seeks to maximize total explained variance, it loads disproportionately on assets with high total variance, even when that variance is primarily idiosyncratic. This creates two distinct problems. First, the estimated factor loadings become distorted, with excessive weight placed on high-variance assets. Second, the factors themselves reflect a mix of systematic risk and clusters of idiosyncratic volatility, rather than capturing pure common variation.

Zhang et al., 2022 and Bai and Ng, 2002 formalize this problem and show that heteroskedasticity in Σ_E introduces bias in PCA-based factor estimation. The intuition is straightforward: the diagonal elements of $\widehat{\Sigma}$ contain both systematic variance (from $\Lambda\Sigma_F\Lambda^\top$) and idiosyncratic variance (from Σ_E). When idiosyncratic variances are heterogeneous and large, they dominate the diagonal and pull the eigenvectors away from the true factor space.

A common response in applied work is to standardize returns to have unit variance before applying PCA. While this eliminates heteroskedasticity, it also distorts the covariance structure by forcing all assets to contribute equally to total variance, regardless of their true systematic importance. This approach trades one bias for another and is not guaranteed to improve factor estimation.

2.3. Heteroskedastic PCA (heteroPCA)

HeteroPCA addresses heteroskedasticity directly by recognizing that systematic information resides primarily in the off-diagonal elements of $\widehat{\Sigma}$ (the covariances), while heteroskedastic noise inflates the diagonal elements (the variances). The key insight is to replace the observed diagonal with an estimate implied by a low-rank approximation of the off-diagonal structure.

The heteroPCA algorithm proceeds as follows. Let $\Delta(A)$ denote the operator that sets the diagonal of a matrix A to zero while preserving the off-diagonals. Given the sample covariance matrix $\widehat{\Sigma}$, the number of factors K, and the number of iterations T_0, heteroPCA iteratively constructs a sequence of matrices $N^{(0)}, N^{(1)}, \ldots, N^{(T_0)}$:

```
1. Initialize: $N^{(0)} = \Delta(\widehat{\Sigma})$ (set diagonal to zero, keep off-diagonals)
2. For $t = 0, 1, \ldots, T_0 - 1$:
    (a) Compute the best rank-$K$ approximation of $N^{(t)}$:
            $\widetilde{N}^{(t)} = \sum_{k=1}^{K} \lambda_k^{(t)} u_k^{(t)} {u_k^{(t)}}^\top,$
        where $\lambda_k^{(t)}$ and $u_k^{(t)}$ are the $k$-th eigenvalue and eigenvector of $N^{(t)}$.
    (b) Update the matrix by replacing the diagonal:
            $N_{ij}^{(t+1)} = \begin{cases} \widetilde{N}_{ii}^{(t)}, & i = j, \\ \widehat{\Sigma}_{ij}, & i \neq j. \end{cases}$
3. Output: The top $K$ eigenvectors of $N^{(T_0)}$.
```

Algorithm 1: HeteroPCA

The algorithm preserves the off-diagonal covariances from the data while iteratively replacing the diagonal with values consistent with a low-rank factor structure. This downweights idiosyncratic variance and focuses estimation on the systematic covariance structure. Zhang et al., 2022 prove that under heteroskedasticity, heteroPCA reduces bias in eigenvector estimation and delivers more accurate recovery of the factor space compared to standard PCA.

The intuition behind heteroPCA is that the off-diagonals contain information about how assets co-move due to common factors, while the diagonals are contaminated by heteroskedastic noise. By enforcing consistency between the diagonal and a rank-K approximation of the off-diagonal structure, heteroPCA effectively estimates what the diagonal should be if the data were generated purely by a low-rank factor model. This approach avoids the pitfalls of standardization (which distorts covariances) while directly addressing the source of the bias.

3. Data and empirical implementation

3.1. Data

The empirical analysis uses multiple cross-sections of equity portfolios to assess the performance of heteroPCA relative to standard PCA. I consider three main datasets.

First, I use double- and triple-sorted equity portfolios from Kenneth French's Data Library. These portfolios sort stocks based on pairs or triples of firm characteristics including size, book-to-market (BM), accruals (ACC), investment (INV), operating profitability (OP), short-term reversal (STREV), momentum (MOM), idiosyncratic volatility (IVOL), and total volatility (VOL). The double-sorted portfolios contain $N = 25$ portfolios, while the triple-sorted portfolios contain $N = 32$ portfolios. The sample period runs from November 1963 to June 2025 at monthly frequency.

Second, I use the Asset Pricing Tree (AP-Tree) portfolios from Bryzgalova et al., 2025. These portfolios are constructed from ten firm characteristics using the AP-Tree algorithm, which generates tree-based partitions designed to capture cross-sectional variation in expected returns. I use both the Tree10 and Tree40 versions, which contain $N = 60$ portfolios each but reflect different pruning strategies. The sample period runs from January 1964 to December 2016.

Third, I construct a balanced panel of $N = 233$ individual stocks from CRSP. The sample period runs from May 1972 to December 2024. I emphasize that this balanced panel reflects strong survivorship bias, as it includes only firms that survived the entire sample period. These firms are therefore likely to be large, mature, and relatively homogeneous, which limits the scope for heteroskedasticity corrections.

For all datasets, I compute monthly excess returns over the risk-free rate. Following standard practice in the factor model literature, I demean returns before applying PCA or heteroPCA.

3.2. Implementation of PCA and heteroPCA

I implement both standard PCA and heteroPCA to extract $K = 3$ latent factors from each cross-section. For PCA, I compute the sample covariance matrix $\widehat{\Sigma}$ and extract the top three eigenvectors, which serve as estimates of the factor loadings $\widehat{\Lambda}$. The factor realizations $\widehat{F}$ are obtained by projecting returns onto these loadings.

For heteroPCA, I follow the algorithm described in Section 2.3, setting the number of iterations to $T_0 = 5$. This choice balances computational efficiency with convergence of the iterative procedure. As with PCA, the estimated loadings $\widehat{\Lambda}$ are given by the top three eigenvectors of the final matrix $N^{(T_0)}$, and the factor realizations are obtained by projection.

To ensure comparability across methods, I normalize the signs of the factors so that each has a positive mean. This normalization is standard in the factor model literature and facilitates interpretation.

3.3. Out-of-sample evaluation framework

To assess the performance of PCA and heteroPCA, I conduct an out-of-sample evaluation using rolling estimation windows. The procedure is as follows.

I use a rolling window of $T = 240$ months (20 years) to estimate factors and loadings. For each window ending at time t, I estimate $\widehat{F}_t$ and $\widehat{\Lambda}_t$ using data from $t - 239$ to t. I then run time-series regressions for each asset n:

$$r_{t,n} = \alpha_n + \widehat{F}_t B_n^\top + e_{t,n}, \tag{5}$$

yielding estimated loadings $\widehat{B} = [B_1, \ldots, B_N]^\top$ and pricing errors $\widehat{\alpha} = [\alpha_1, \ldots, \alpha_N]^\top$.

Next, I construct the projection matrix onto the factor span:

$$P_t = B_t (B_t^\top B_t)^{-1} B_t^\top. \tag{6}$$

I use this projection to decompose the realized next-period returns R_{t+1} into a component explained by the factors and an out-of-sample pricing error:

$$\widehat{R}_{t+1|t} = R_{t+1} P_t, \qquad \widehat{\alpha}_{t+1} = R_{t+1}(I - P_t). \tag{7}$$

The out-of-sample pricing error $\widehat{\alpha}_{t+1}$ measures the portion of realized returns that the model fails to explain when loadings are treated as constant one step ahead.

I summarize out-of-sample performance using three metrics. First, I compute the root mean squared pricing error:

$$\text{RMS}_\alpha = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \bar{\alpha}_n^2}, \tag{8}$$

where $\bar{\alpha}_n$ is the time-series average of the out-of-sample pricing errors for asset n. This metric captures the average magnitude of mispricing across assets.

Second, I compute the average share of unexplained variance:

$$\bar{\sigma}_e^2 = \frac{1}{N} \sum_{n=1}^{N} \frac{\text{Var}(e_{t,n})}{\text{Var}(r_{t,n})}, \tag{9}$$

which measures the fraction of total return variance not explained by the factors. Lower values indicate better time-series fit.

Third, I evaluate the out-of-sample Sharpe ratio of the tangency portfolio implied by the factors. In each window, I estimate the mean and covariance of the factors and compute the Markowitz tangency portfolio in factor space. I then map this portfolio to asset space using the factor-mimicking loadings $\widehat{\Lambda}_t$ and evaluate the realized return at $t+1$. The Sharpe ratio is computed from the full out-of-sample time series of realized portfolio returns.

3.4. Stochastic discount factor diagnostics

To assess the economic content of the estimated factors, I examine the implied stochastic discount factor (SDF). For factors $\widehat{F}$ and loadings $\widehat{\Lambda}$ estimated over the full sample, I compute the vector of risk prices:

$$\widehat{\lambda} = \widehat{\Sigma}_F^{-1}\widehat{\mu}_F, \tag{10}$$

where $\widehat{\mu}_F$ and $\widehat{\Sigma}_F$ are the sample mean and covariance of the factors. These risk prices correspond to the weights of the mean-variance tangency portfolio in the factor space.

The SDF time series is then constructed as:

$$\widetilde{m}_t = \widehat{F}_t^\top \widehat{\lambda}. \tag{11}$$

I also compute the SDF-mimicking portfolio in asset space:

$$\widehat{w}_{\text{SDF}} = \widehat{\Lambda}(\widehat{\Lambda}^\top \widehat{\Lambda})^{-1}\widehat{\lambda}. \tag{12}$$

This portfolio represents the asset-space implementation of the SDF and provides economic intuition about which assets the model views as risky.

To quantify the pricing accuracy of the SDF, I compute the Hansen-Jagannathan distance (Hansen and Jagannathan, 1991):

$$d_{\text{HJ}} = \sqrt{\widehat{\alpha}^\top \widehat{\Sigma}_e^{-1}\widehat{\alpha}}, \tag{13}$$

where $\widehat{\alpha}$ is the vector of in-sample pricing errors from the time-series regressions and $\widehat{\Sigma}_e = \widehat{\text{Cov}}(e_t)$ is the covariance matrix of the residuals. The HJ distance measures how far the SDF is from correctly pricing all assets: smaller values indicate better cross-sectional fit.

4. Empirical results

4.1. Out-of-sample pricing performance

Table 1 presents the out-of-sample performance of PCA and heteroPCA across all equity portfolios. The results reveal a consistent pattern: heteroPCA delivers superior cross-sectional pricing performance compared to standard PCA, as measured by higher Sharpe ratios and lower root mean squared pricing errors (RMS_α). However, heteroPCA exhibits slightly worse time-series fit, reflected in higher average unexplained variance ($\bar{\sigma}_e^2$).

[Table 1 about here.]

For the AP-Tree portfolios, heteroPCA achieves out-of-sample Sharpe ratios of 0.46 (Tree10) and 0.55 (Tree40), substantially higher than the 0.18 and 0.26 delivered by standard PCA. Similarly, heteroPCA reduces RMS_α to 0.72 and 0.80, compared to 0.85 and 0.90 for PCA. These improvements are economically meaningful and indicate that heteroPCA identifies factors with stronger cross-sectional pricing power.

Across the double-sorted portfolios, heteroPCA outperforms PCA in the majority of cases. For instance, in the Size & Book-to-Market portfolios, heteroPCA delivers a Sharpe ratio of 0.28 versus 0.15 for PCA, while reducing RMS_α from 0.16 to 0.12. Similar patterns emerge for Size & Accruals (Sharpe ratio 0.21 vs 0.13), Size & Investment (0.32 vs 0.20), and Size & Idiosyncratic Volatility (0.35 vs 0.21). The improvements are robust across most characteristic sorts, with exceptions primarily in portfolios where both methods perform similarly.

The time-series fit, measured by $\bar{\sigma}_e^2$, is consistently better for standard PCA. This is expected because PCA maximizes total explained variance, whereas heteroPCA maximizes a low-rank approximation of the covariance structure after filtering out heteroskedastic diagonal elements. Thus, PCA naturally explains more of the time-series variation, but at the cost of loading on idiosyncratic noise. HeteroPCA trades off some time-series fit to focus on the systematic covariance structure, resulting in better cross-sectional pricing.

For individual stocks, heteroPCA does not outperform PCA. The Sharpe ratios are similar (0.14 vs 0.15), and RMS_α is only marginally lower (0.38 vs 0.39). This result is consistent with the strong survivorship bias in the balanced panel of stocks: firms that survived from 1972 to 2024 are likely large, mature, and relatively homogeneous in their idiosyncratic volatility. Consequently, there is limited scope for heteroskedasticity corrections to improve factor estimation.

Taken together, these results demonstrate that accounting for heteroskedasticity delivers substantial gains in cross-sectional pricing performance. HeteroPCA identifies factors with stronger economic content and achieves better out-of-sample Sharpe ratios, albeit with a modest reduction in time-series fit, a tradeoff that appears economically favorable given the improvements in risk-adjusted returns.

4.2. Factor interpretability

Beyond statistical performance, a key question is whether heteroPCA delivers factors with clearer economic interpretation. To examine this, I focus on the Size & Accruals portfolios, which have a well-understood factor structure corresponding to market, size, and accrual risk.

Figure 1 presents the factor loadings estimated by PCA and heteroPCA. For both methods, the first factor exhibits similar loadings across all size-accrual portfolios, consistent with a market factor. The second factor captures a size effect under both methods, with positive loadings on small portfolios and negative loadings on large portfolios, similar to the Fama-French SMB factor.

[Figure 1 about here.]

The key difference emerges in the third factor. For standard PCA, the third factor exhibits no clear pattern across the size-accrual space. In contrast, heteroPCA's third factor loads positively on low-accrual portfolios and negatively on high-accrual portfolios, cleanly identifying an accrual factor analogous to the Fama-French accrual anomaly. This pattern is consistent across size quintiles, indicating that heteroPCA successfully isolates the accrual dimension that standard PCA fails to detect.

The improved interpretability of heteroPCA stems directly from its treatment of heteroskedasticity. Small, extreme-accrual portfolios exhibit high idiosyncratic volatility, which causes standard PCA to avoid loading heavily on them. By downweighting this idiosyncratic variance, heteroPCA reveals the systematic covariance structure underlying these portfolios, allowing the third factor to capture the accrual premium.

Figure 2 illustrates the diagonal adjustment performed by heteroPCA. The figure plots the original diagonal elements of $\widehat{\Sigma}$ (total variance) against the covariance-implied diagonal from the final iteration of heteroPCA. The adjustment is most pronounced for small, extreme-accrual portfolios, whose variances are substantially reduced. Importantly, these are precisely the portfolios that receive large loadings on the third heteroPCA factor. Once the excess idiosyncratic variance is stripped away, the shared co-movement of these portfolios becomes visible, and heteroPCA loads on this systematic component rather than the noise.

[Figure 2 about here.]

These results demonstrate that heteroPCA is not merely loading on noise, it uncovers latent systematic dimensions that remain hidden under standard PCA. Yet, the same portfolios with reduced variance receive large factor loadings after correction, because once their excess variance is stripped away, their shared co-movement becomes visible. Standard PCA misses this systematic dimension. By focusing on the off-diagonal covariance structure, heteroPCA identifies factors with clearer economic meaning and stronger alignment with known risk premia.

4.3. Stochastic discount factor

I now examine the stochastic discount factor (SDF) implied by the estimated factors. Table 2 reports the Hansen-Jagannathan distance for PCA and heteroPCA across all portfolios. HeteroPCA delivers lower d_{HJ} in 10 out of 15 panels, indicating better cross-sectional pricing accuracy.

[Table 2 about here.]

The improvement is particularly pronounced in portfolios sorted on Size & Book-to-Market (0.27 vs 0.35), Size & Accruals (0.26 vs 0.31), Size & Momentum (0.29 vs 0.35), and Size & Idiosyncratic Volatility (0.34 vs 0.38). For these cross-sections, heteroPCA reduces the HJ distance by 15–30%, suggesting that the factors more accurately capture the sources of systematic risk.

The economic interpretation of the SDF becomes clearer when I examine the SDF-mimicking portfolio weights. Figure 3 presents these weights for the Size & Accruals portfolios under both PCA and heteroPCA.

[Figure 3 about here.]

Under standard PCA, the SDF exhibits a "size barbell" pattern: it is long in both big and small stocks (with the strongest positions in low-accrual portfolios) and short in mid-cap stocks. This pattern lacks clear economic interpretation and suggests that PCA is capturing a mix of systematic and idiosyncratic risk.

In contrast, heteroPCA delivers a much cleaner pattern. The SDF is concentrated short in small, high-accrual portfolios and maintains small long positions elsewhere, particularly in low-accrual stocks. Big stocks are nearly neutral. This represents a clear "quality versus junk" trade: shorting small stocks with extreme accruals (which tend to have poor future performance) and holding positions in higher-quality stocks with moderate accruals.

The difference between the two approaches reflects their treatment of idiosyncratic variance. Standard PCA loads on total variance, leading to complex portfolio weights that reflect both systematic risk and idiosyncratic volatility clusters. HeteroPCA replaces the noisy diagonal with a rank-K imputation from the off-diagonal covariance structure, reorienting the SDF toward the systematic low-versus-high accrual dimension. This yields a more intuitive economic interpretation: a premium for bearing accrual risk concentrated in small-cap stocks—and delivers lower pricing errors.

These results reinforce the main findings from the out-of-sample analysis: by filtering out heteroskedastic noise and focusing on the systematic covariance structure, heteroPCA identifies factors and SDFs that better reflect the underlying sources of risk in the cross-section of asset returns.

5. Conclusion

This paper demonstrates that accounting for heteroskedasticity in idiosyncratic errors substantially improves the estimation of latent risk factors in asset pricing applications. Building on recent developments in the statistics literature, I apply heteroskedastic PCA (heteroPCA) to correct for the distortions that arise when standard PCA is applied to returns with heterogeneous idiosyncratic variances. This approach iteratively replaces the diagonal of the sample covariance matrix with estimates implied by the off-diagonal structure, downweighting idiosyncratic noise while preserving the systematic covariance information.

The empirical results demonstrate three main findings. First, heteroPCA delivers superior out-of-sample cross-sectional pricing performance compared to standard PCA. Across multiple equity portfolios sorted on size, book-to-market, accruals, investment, profitability, and other characteristics, heteroPCA achieves higher Sharpe ratios and lower average pricing errors. These improvements are economically meaningful, with Sharpe ratio gains often exceeding 50% relative to standard PCA. While heteroPCA exhibits slightly worse time-series fit, an expected consequence of filtering out idiosyncratic variance—the tradeoff appears favorable given the substantial improvements in risk-adjusted returns.

Second, heteroPCA identifies factors with clearer economic interpretation. In cross-sections with well-known factor structures, such as portfolios sorted on size and accruals, heteroPCA cleanly recovers market, size, and accrual factors, whereas standard PCA struggles to isolate these dimensions. The improved interpretability stems from heteroPCA's ability to uncover systematic covariance patterns that remain hidden beneath idiosyncratic noise in standard PCA. By downweighting high-variance assets when their variance reflects idiosyncratic risk, heteroPCA focuses on true systematic co-movement.

Third, the stochastic discount factors implied by heteroPCA factors exhibit better pricing accuracy and more intuitive economic content. HeteroPCA consistently delivers lower

Hansen-Jagannathan distances, and the SDF-mimicking portfolios display coherent patterns aligned with known risk premia. For instance, in size-accrual portfolios, heteroPCA identifies a clear quality-versus-junk trade, whereas standard PCA produces less interpretable "size barbell" patterns.

My findings have practical implications for empirical asset pricing. Heteroskedastic idiosyncratic variances are pervasive across asset classes: in equities, firm size, leverage, and growth opportunities generate varying levels of idiosyncratic risk; in foreign exchange markets, interest rate differentials and liquidity differences create heterogeneous volatility across currencies; similar patterns arise in bonds, commodities, and other markets. Even when assets are grouped into portfolios, as is standard practice, characteristic-driven volatility persists. My results show that correcting for this heteroskedasticity leads to more reliable and interpretable risk factors.

The heteroPCA methodology can be integrated into existing PCA-based approaches in asset pricing, including instrumented PCA, projected PCA, and risk premium PCA. By addressing heteroskedasticity as a first step, these methods may achieve improved factor estimation and better pricing performance. Future research could explore the application of heteroPCA to other asset classes, examine its performance in larger cross-sections with hundreds or thousands of assets, and investigate its interaction with other methodological innovations in factor modelling.

References

Andersen, Torben G et al. (2003). “Micro effects of macro announcements: Real-time price discovery in foreign exchange”. In: *American Economic Review* 93.1, pp. 38–62.

Ang, Andrew et al. (2006). “The Cross-Section of Volatility and Expected Returns”. In: *The Journal of Finance* 61.1, pp. 259–299. DOI: `https://doi.org/10.1111/j.1540-6261.2006.00836.x`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2006.00836.x`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2006.00836.x`.

Bai, Jushan and Serena Ng (2002). “Determining the Number of Factors in Approximate Factor Models”. In: *Econometrica* 70.1, pp. 191–221. DOI: `https://doi.org/10.1111/1468-0262.00273`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-0262.00273`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00273`.

Bryzgalova, Svetlana, Markus Pelger, and Jason Zhu (2025). “Forest through the Trees: Building Cross-Sections of Stock Returns”. In: *The Journal of Finance* 80.5, pp. 2447–2506. DOI: `https://doi.org/10.1111/jofi.13477`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.13477`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.13477`.

Bryzgalova, Svetlana et al. (2023). “Asset-Pricing Factors with Economic Targets”. Working paper / Preprint (SSRN 4344837). URL: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4344837`.

Campbell, John Y. et al. (2001). “Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk”. In: *The Journal of Finance* 56.1, pp. 1–43. DOI: `https://doi.org/10.1111/0022-1082.00318`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/0022-1082.00318`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/0022-1082.00318`.

Della Corte, Pasquale, Steven J Riddiough, and Lucio Sarno (2016). “Currency premia and global imbalance risk”. In: *Journal of Finance* 71.6, pp. 3671–3716.

Fama, Eugene F and Kenneth R French (1992). “The Cross-Section of Expected Stock Returns”. In: *The Journal of Finance* 47.2, pp. 427–465. DOI: `https://doi.org/10.1111/j.1540-6261.1992.tb04398.x`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1992.tb04398.x`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1992.tb04398.x`.

Fan, Jianqing, Yuan Liao, and Weichen Wang (Feb. 2016). “Projected principal component analysis in factor models”. In: *The Annals of Statistics* 44.1, pp. 219–254. DOI: `10.1214/15-AOS1364`. URL: `https://doi.org/10.1214/15-AOS1364`.

Hansen, Lars Peter and Ravi Jagannathan (1991). "Implications of Security Market Data for Models of Dynamic Economies". In: *Journal of Political Economy* 99.2, pp. 225–262. ISSN: 00223808, 1537534X. URL: `http://www.jstor.org/stable/2937680` (visited on 01/24/2026).

Harvey, Campbell R., Yan Liu, and Heqing Zhu (Oct. 2015). "… and the Cross-Section of Expected Returns". In: *The Review of Financial Studies* 29.1, pp. 5–68. ISSN: 0893-9454. DOI: `10.1093/rfs/hhv059`. eprint: `https://academic.oup.com/rfs/article-pdf/29/1/5/24450794/hhv059.pdf`. URL: `https://doi.org/10.1093/rfs/hhv059`.

He, Ai et al. (2023). "Shrinking Factor Dimension: A Reduced-Rank Approach". In: *Management Science* 69.9, pp. 5501–5522. DOI: `10.1287/mnsc.2022.4563`.

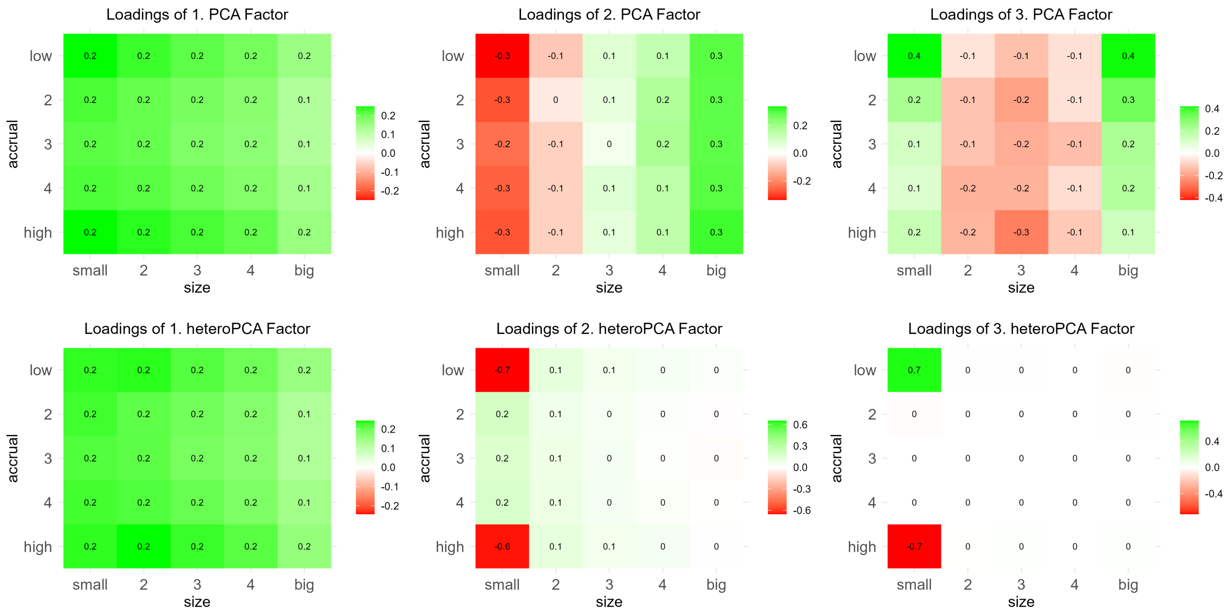
Kelly, Bryan T., Seth Pruitt, and Yinan Su (2019). "Characteristics are covariances: A unified model of risk and return". In: *Journal of Financial Economics* 134.3, pp. 501–524. ISSN: 0304-405X. DOI: `https://doi.org/10.1016/j.jfineco.2019.05.001`. URL: `https://www.sciencedirect.com/science/article/pii/S0304405X19301151`.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh (Feb. 2020). "Shrinking the cross-section". en. In: *Journal of Financial Economics* 135.2, pp. 271–292. ISSN: 0304405X. DOI: `10.1016/j.jfineco.2019.06.008`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0304405X19301655` (visited on 09/08/2025).

Lettau, Martin and Markus Pelger (May 2020). "Factors That Fit the Time Series and Cross-Section of Stock Returns". en. In: *The Review of Financial Studies* 33.5. Ed. by Stijn Van Nieuwerburgh, pp. 2274–2325. ISSN: 0893-9454, 1465-7368. DOI: `10.1093/rfs/hhaa020`. URL: `https://academic.oup.com/rfs/article/33/5/2274/5756219` (visited on 08/29/2025).

Menkhoff, Lukas et al. (2012). "Carry trades and global foreign exchange volatility". In: *Journal of Finance* 67.2, pp. 681–718.

Zhang, Anru R., T. Tony Cai, and Yihong Wu (Feb. 2022). "Heteroskedastic PCA: Algorithm, optimality, and applications". en. In: *The Annals of Statistics* 50.1. ISSN: 0090-5364. DOI: `10.1214/21-AOS2074`. URL: `https://projecteuclid.org/journals/annals-of-statistics/volume-50/issue-1/Heteroskedastic-PCA-Algorithm-optimality-and-applications/10.1214/21-AOS2074.full` (visited on 08/29/2025).



Loadings of 1. PCA Factor

Loadings of 2. PCA Factor

Loadings of 3. PCA Factor

Loadings of 1. heteroPCA Factor

Loadings of 2. heteroPCA Factor

Loadings of 3. heteroPCA Factor

Figure 1: Factor loadings for Size & Accruals portfolios. Top panel: PCA (top 3 factors). Bottom panel: heteroPCA (bottom 3 factors).

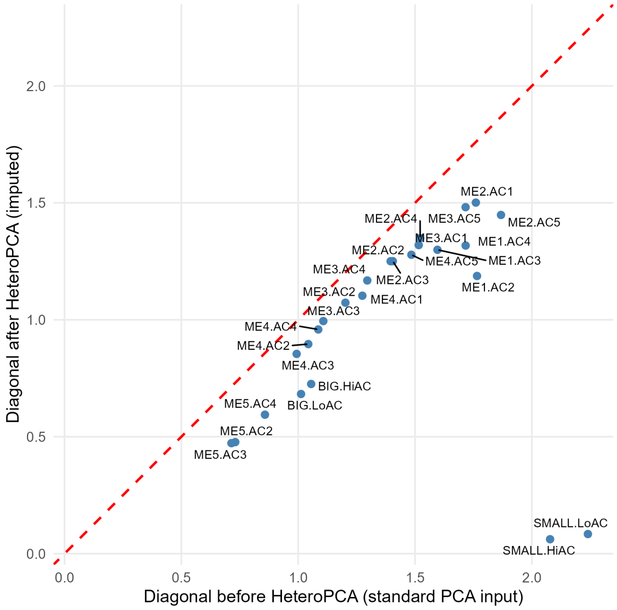
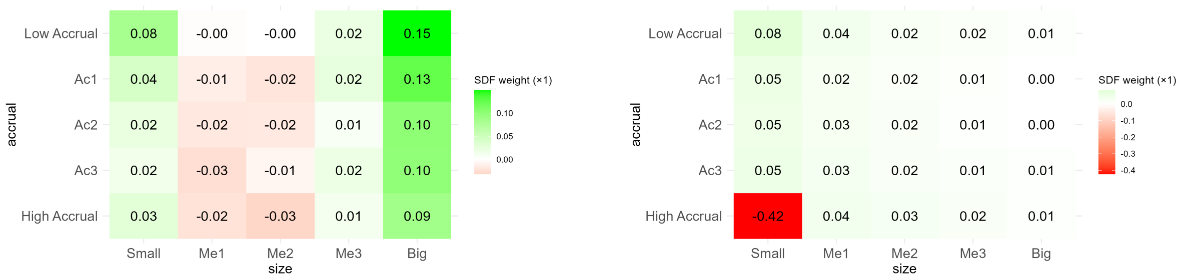


Figure 2: Diagonal adjustment for Size & Accruals portfolios. The x-axis shows the original diagonal (total variance), and the y-axis shows the covariance-implied diagonal from heteroPCA. HeteroPCA flattens the diagonal of the covariance matrix, especially for small, extreme-accrual portfolios.



PCA

heteroPCA

Figure 3: SDF-mimicking portfolio weights for Size & Accruals portfolios. Left panel: PCA exhibits a "size barbell" pattern. Right panel: heteroPCA shows a clear "quality versus junk" pattern.

Table 1: Out-of-sample fit: $K = 3$ factors. T_0 set to 5 in heteroPCA case.

	SR		RMS_α		$\bar{\sigma}_e^2$ (%)	
	heteroPCA	PCA	heteroPCA	PCA	heteroPCA	PCA
AP-Tree (Tree10)	**0.46**	0.18	**0.72**	0.85	10.01	**9.62**
AP-Tree (Tree40)	**0.55**	0.26	**0.80**	0.90	10.97	**10.48**
Individual stocks	0.14	**0.15**	**0.38**	0.39	72.17	**71.96**
SIZE&BM	**0.28**	0.15	**0.12**	0.16	9.10	**7.58**
SIZE&ACC	**0.21**	0.13	**0.09**	0.10	7.26	**6.83**
SIZE&INV	**0.32**	0.20	**0.12**	0.14	8.74	**7.25**
SIZE&OP	**0.17**	0.11	0.13	**0.12**	10.61	**8.70**
SIZE&STREV	0.04	**0.12**	**0.17**	0.18	9.94	**7.65**
SIZE&MOM	**0.28**	0.18	**0.13**	0.17	9.44	**8.04**
SIZE&IVOL	**0.35**	0.21	**0.16**	0.17	8.02	**6.09**
SIZE&VOL	**0.30**	0.22	0.15	0.15	7.81	**6.03**
OP&INV	**0.22**	0.21	0.13	0.13	19.09	**17.37**
SIZE&BM&INV	**0.24**	0.11	**0.14**	0.18	12.53	**12.40**
SIZE&BM&OP	**0.24**	0.16	**0.13**	0.20	20.63	**18.03**
SIZE&OP&INV	**0.32**	0.18	**0.12**	0.18	13.35	**12.67**

Table 2: Hansen–Jagannathan distance. $K = 3$ factors. T_0 set to 5 in heteroPCA case.

	heteroPCA	PCA
AP-Tree (Tree10)	1.42	1.47
AP-Tree (Tree40)	1.42	1.47
Individual stocks	0.57	0.57
SIZE&BM	**0.27**	0.35
SIZE&ACC	**0.26**	0.31
SIZE&INV	0.41	**0.37**
SIZE&OP	0.24	**0.23**
SIZE&STREV	0.33	**0.31**
SIZE&MOM	**0.29**	0.35
SIZE&IVOL	**0.34**	0.38
SIZE&VOL	**0.33**	0.38
OP&INV	0.24	0.24
SIZE&BM&INV	**0.25**	0.34
SIZE&BM&OP	**0.23**	0.30
SIZE&OP&INV	**0.30**	0.36