SIAM J. FINANCIAL MATH.
Vol. 13, No. 3, pp. 702–744

Principal Eigenportfolios for U.S. Equities*

Marco Avellaneda†, Brian Healy‡, Andrew Papanicolaou§, and George Papanicolaou¶

Abstract. We analyze portfolios constructed from the principal eigenvector of the equity returns' correlation matrix and compare these portfolios with the capitalization weighted market portfolio. It is well known empirically that principal eigenportfolios are a good proxy for the market portfolio. We quantify this property through the large-dimensional asymptotic analysis of a spike model with diverging top eigenvalue, comprising a rank-one matrix and a random matrix. We show that, in this limit, the top eigenvector of the correlation matrix is close to the vector of market betas divided componentwise by returns standard deviation. Historical returns data are generally consistent with this analysis of the correspondence between the top eigenportfolio and the market portfolio. We further examine this correspondence using eigenvectors obtained from hierarchically constructed tensors where stocks are separated into their respective industry sectors. This hierarchical approach results in a principal factor whose portfolio weights are all positive for a greater percentage of time compared to the weights of the vanilla eigenportfolio computed from the correlation matrix. Returns from hierarchical construction are also more robust with respect to the duration of the time window used for estimation. All principal eigenportfolios that we observe have returns that exceed those of the market portfolio between 1994 and 2020. We attribute these excess returns to the brief periods where short holdings are more than a small percentage of portfolio weight.

Key words. eigenportfolios, principal component analysis, tensor decompositions

MSC codes. 62P05, 91B28

DOI. 10.1137/20M1383501

1. Introduction.

A principal eigenportfolio is commonly constructed using the top eigenvector of the equity returns' correlation matrix, and it has been known for some time (Plerou et al. (1999); Potters, Bouchaud, and Laloux (2005); Avellaneda and Lee (2010)) to be a good proxy for the capitalization-weighted market portfolio. In this paper, we provide a theoretical analysis for why this is so using a diverging spike model for the returns. This spike model consists of a deterministic rank-one matrix plus a random matrix whose bulk spectrum lies well below the rank-one eigenvalue. In the asymptotic limit where the number of stocks and observations tends toward infinity, we show that a stock's market beta divided by its return variance is proportional to its principal eigenportfolio weight (see Proposition 2.2). This theoretically obtained result is consistent with historical daily returns data of S&P500 stocks.

*Received by the editors December 1, 2020; accepted for publication (in revised form) March 6, 2022; published electronically July 18, 2022.
https://doi.org/10.1137/20M1383501
Funding: The third author was partially supported by NSF grant DMS-1907518.
†Courant Institute of Mathematical Sciences, New York University, New York, NY 10012-1185 USA (avellaneda@courant.nyu.edu).
‡Institute of Finance and Technology, University College London, Gower Street, London WC1E 6BT, UK (brian@decisionsci.net).
§Department of Mathematics, North Carolina State University, Raleigh, NC 27695 USA (apapani@ncsu.edu).
¶Department of Mathematics, Stanford University, Stanford, CA 94305 USA (papanicolaou@stanford.edu).

Specifically, the eigenportfolio weights computed from the data are close to the weights predicted by the spike model; this is the main result of the paper. To gain a sense for why this result is perhaps surprising and of broader interest, we may say informally that the principal eigenvector of the correlation matrix of equity returns, an entirely algebraic construct, discovers market betas.

We further consider different structures for organizing the returns data and the effect they have on the relationship between principal eigenportfolios and the market portfolio. These data structures are based on industry sector classifications, which are determined by either economic or market considerations. Using this additional structure we build a tensor whose decomposition yields a principal eigenportfolio that behaves like the "vanilla" eigenportfolio (i.e., the eigenportfolio computed from the standard correlation matrix without any tensor structure), but with some key differences. First, principal eigenportfolio weights from our tensor have fewer short positions than the other eigenportfolio constructions, which is a good property to have in a primary market factor. Second, the tensor eigenportfolio's weights are not as close to the market portfolio betas, but when we regress these eigenportfolio returns onto the market portfolio returns we obtain residuals that have lower kurtosis. We see this reduced kurtosis for the tensor eigenportfolios consistently as we vary the length of the rolling time window used for building the tensor. Thus, tensor eigenportfolios are a more robust proxy for the market portfolio than their vanilla counterpart. One explanation is that the additional structure in the tensors creates restrictions which render a close fit less likely while acting as a regularizer that provides robustness. At present, these latter findings are empirical, without theoretical analysis.

In this paper we perform backtests for three different principal eigenportfolios: the vanilla construction, the hierarchical principal component analysis (HPCA) construction in Avellaneda (2020) and in Kakushadze (2015), and the tensor construction that we propose in what follows. We observe that these eigenportfolios have significant excess returns between 1994 and 2020, but we also find evidence that these excess returns are caused by short positions. In fact, if we censor most short positions, that is, if we regress the principal eigenportfolio onto the market portfolio only on days where the eigenportfolio has short positions amounting to less than 3%, then we don't find significant excess returns. The finding of short positions in a principal factor portfolio runs contrary to conventional portfolio theory, but this is precisely the issue discussed in Brennan and Lo (2010) wherein it is argued that so-called impossible frontiers are quite common and that mean-variance frontiers with at least one long-only portfolio are an exception rather than the rule. Nonetheless, it remains valid to have a preference for a long-only principal factor, which is why the tensor eigenportfolio we've constructed is useful, as it has considerably fewer negative weights in comparison to the vanilla and HPCA eigenportfolios.

1.1. Background. The literature on principal component analysis (PCA) and its application to financial and other data is very extensive. Our interest here is in asymptotic methods that explore consistency of eigenvalues and eigenvectors of large covariance and correlation matrices when the data comes from an underlying model with a prominent low-rank structure (i.e., a spike model). The two most important parameters in PCA are the number of features N, i.e., the number of stocks here, and the number of time samples T, i.e., the number of trading days here; both are taken to be large.

When N is fixed and T tends to infinity, which is classical large sample theory, consistency of PCA for sample covariances was carried out in Anderson (1963), and later came a more involved analysis for correlations in Konishi (1978) and Kollo and Neudecker (1993). A central limit theorem (CLT) for fluctuations is given as well.

When N and T are both large with N/T of order one, then we are in the regime of random matrix theory (RMT). When the data matrix is populated with independent, identically distributed random variables, then the spectrum of the normalized covariance converges to the law shown in Marchenko and Pastur (1967) (which we denote as MP law). This basic result was used with equity returns data and PCA (Plerou et al. (1999, 2002); Potters, Bouchaud, and Laloux (2005); Avellaneda and Lee (2010)) in order to separate market information from noise. That is, after projecting out roughly 10 to 20 eigenvectors corresponding to the largest eigenvalues of the data correlation matrix, the remainder should have a spectrum close to the MP law. The effective dimension of the market subspace, obtained empirically, varies with time. It is high in normal times, indicating that there is a lot of market diversity, and it is low when the markets are stressed.

To account for information-plus-noise data matrices, spike models were introduced (Johnstone (2001)), wherein covariance estimators are composed of a random part plus a finite-rank deterministic part. In the limit, the deterministic eigenvalues will usually be above the MP spectrum of the random part, thus allowing for principal eigenvalue detection. One important result of the asymptotic analysis is that if the true eigenvalue is above the critical threshold, then the sample eigenvalues of the deterministic part do not converge to those of the model, that is, they are not consistent estimators (Bai and Silverstein (1998, 1999); Baik, Ben Arous, and Péché (2005)). A more complete asymptotic theory for spike models is presented in Paul (2007), Baik and Silverstein (2006), Benaych-Georges and Nadakuditi (2011), and Shi (2013), including a CLT for the fluctuations of the eigenvalues and eigenvectors.

It is clear in empirical studies on equities data that the principal eigenvalue of the correlation matrix is much larger than the rest (Potters, Bouchaud, and Laloux (2005)), and this is what we study in this paper. An asymptotic theory of spike covariance models with diverging principal eigenvalues was carried out by Shen, Shen, and Marron (2016) and Wang and Fan (2017). One of the main results of their theory is that, as expected, the sample spike eigenvalues and eigenvectors are asymptotically consistent with the model. An analogous result for correlation matrices is provided in this paper, but without requiring a detailed model for the correlations; such models are usually needed when a fluctuation theory is carried out in addition to proving consistency (Paul (2007)).

The theoretical literature on spike models almost always works with covariances—not correlations—the idea being that the results are not so different with correlations, but the analysis is more involved as in the classical large sample theory of PCA that we have already cited. The empirical studies, however, are almost always done with correlations because inhomogeneities in financial data, for example, are incompatible with the homogeneity assumptions of the theory. The differences between PCA with covariances versus correlations for very large financial data sets are studied empirically in Conlon, Ruskin, and Crane (2010) and Zumbach (2011). The overall conclusion is that the results with correlations are more stable, especially for the principal eigenvalue and eigenvector, which interest us here. The results of Zumbach (2011) provide empirical evidence that both supports the use of the asymptotic theory and helps in interpreting it correctly when applied to real data. An analysis of the asymptotic properties of correlation matrices for spike models with nondiverging eigenvalues is carried out in Morales-Jimenez et al. (2021) and one of the main results there is that the variance of the fluctuations of the spike eigenvalues is smaller for correlations. This quantifies the stability seen in the empirical study of Zumbach (2011).

So far we have cited literature mostly from statistics and econophysics. There is very extensive econometrics literature addressing the issues that concern us here. Consistency and fluctuation theory for eigenvalues and eigenvectors of sample covariances, or factors more generally, is analyzed in Bai (2003) and Bai and Ng (2008, 2019). The large N and T limit is studied but the ratio N/T is not fixed and the distinction of diverging eigenvalues in spike covariance models is not applicable in their framework. The models can be time dependent although the analysis is most often static, not evolutionary, and time is a parameter. Time varying behavior more generally is treated extensively in the econometrics literature (Su and Wang (2017)). The issue of estimating the number of significant factors is also analyzed (Onatski (2010, 2012)), although it is a side issue here. It should be noted that in Onatski (2012) there is an extensive discussion and assessment of the literature on spike covariance models and there is a comparison with models in the econometrics literature such as Bai and Ng (2008). The model proposed and analyzed in Onatski (2012) is different and is particularly well suited for estimating the number of significant eigenvectors or factors when the top eigenvalues are not very large, which is not the case for the principal eigenvector for equities data but it is in general an important issue.

A basic question to ask is whether or not the principal eigenportfolio is mean-variance efficient (Markowitz (1952)). If instead of the principal eigenportfolio we use the market portfolio, then this question goes back to Sharpe (1964) and Lintner (1965), who argue that the capitalization weighted portfolio should be mean-variance efficient or equivalently the tangency portfolio. In Roll (1980) portfolios orthogonal to the market portfolio are studied and in Boyle (2014) the principal eigenportfolio of correlations is taken as a proxy for the market portfolio, because their returns are comparable, and orthogonal eigenportfolios are discussed. The relation between PCA and mean-variance portfolios is considered in Steele (1995) using a data set of monthly returns for about half of the Dow Jones 30 industrials and it is found that with two principal components the weights of the mean-variance portfolio appear to be unstable. The prevailing view in the 1990s and before was that PCA was to be used for denoising before doing a mean-variance optimization and that it did not work well in practice. The theory of PCA as denoising before a mean-variance optimization is considered in Chen and Yuan (2016). What changed the outlook in denoising is the introduction of shrinkage methods (Ledoit and Wolf (2004, 2012); Ledoit and Péché (2011)). These methods are now being used extensively in econophysics (Bun, Bouchaud, and Potters (2017)). As for using only the principal eigenvector of the correlation matrix when the principal eigenvalue is large, then the mean-variance optimal portfolio is, asymptotically, the principal eigenportfolio. This is not in Chen and Yuan (2016) but can be obtained from there.

Principal component methods for tensor data structures should be introduced when there is a lot of diversity in the data, not just strong heterogeneity, and when it is not a good idea to string out all the information in one long features vector (Tucker (1964, 1966)). The main difficulty is that there is no analogue of the singular value decomposition (SVD) for tensors of dimension three or more, and reducing a tensor to a sum of rank-one tensors can only be done approximately in general. This is the canonical polyadic decomposition (CPD) and the computations are much more involved than the corresponding ones for matrices (Kolda and Bader (2009)). There is a decomposition for tensors that is similar to SVD (De Lathauwer, De Moor, and Vandewalle (2000); Cichocki et al. (2015)), the multilinear SVD (MLSVD), which is relatively easy to implement because it uses the matrix SVD repeatedly, but it does not produce a representation in the form of a sum of rank-one tensors. However, when there is a dominant rank-one approximation of the tensor, then it can be obtained from the top rank-one truncation of MLSVD (Cichocki et al. (2015)). When the top rank-one approximation is dominant, then we may use a spike tensor model and try to show that as the dominance increases the approximation becomes exact; this is done in Montanari and Richard (2014) with an emphasis on the computational complexity of the approximation algorithms. What is important in our empirical study of tensor methods here is the differences that emerge when compared to the vanilla matrix PCA approach. The differences are not very big because the data set is not overly heterogeneous, but we identify and discuss how these differences affect portfolio performance.

1.2. Contributions of this paper. In this paper we identify and analyze mathematically the relationship between the principal eigenportfolio and the market portfolio. That the market portfolio is a good proxy for the eigenportfolio in that their returns over time are close is well known (Potters, Bouchaud, and Laloux (2005); Boyle (2014)). We introduce an asymptotic theory and we demonstrate, using historical market data, that the theory's predictions consistently model the behavior of the data regarding the eigenportfolio. The significance of this theory is that we have shown the principal eigenportfolio's connection with well-known financial quantities, namely, the market betas.

To place the theoretical contribution into the broad context of PCA research that we just reviewed, we note that we are dealing with correlations, with the principal eigenvalue being of order N, the number of features (stocks), and with N and T, the time interval, going to infinity with N/T of order one. We show that in this diverging eigenvalue, large data, or RMT limit the principal eigenvalue and eigenvector are consistent with a spike model. The corresponding result for covariances is known (Shen, Shen, and Marron (2016); Wang and Fan (2017)). Going to correlations is more involved, just as it is in classical large sample theory for PCA as already noted, but the methodology is similar. The assumptions needed for the diverging spike eigenvalue model are simpler here because we only show consistency without a CLT for the fluctuations. The way the consistency is applied is to (a) calculate the principal eigenvalue and eigenvector of the correlation matrix of the excess returns data, then (b) calculate the betas of these returns with the market portfolio, the S&P500 index, and (c) show that the betas divided by returns variance are close to the principal eigenvector. The importance of using correlations and not covariances is clearly seen in the analysis of large financial data sets in Zumbach (2011): the principal eigenvalue and the principal eigenvector of the correlation matrix are more stable over time and the gap to the rest of the eigenvalues is wider during periods of market stress and narrower during normal periods. Our empirical results for U.S. equity returns are consistent with this behavior.

Additionally, we consider alternative eigenportfolios based on hierarchical structures that include each stock's industry sector classification. We consider two principal eigenportfolio constructions: one that is derived using multilinear tensor PCA and a second that uses the HPCA method. We observe empirically that vanilla PCA (i.e., standard PCA without sector classifications) and HPCA eigenportfolios have portfolio weights that align closely with the market betas, as expected from the theoretical analysis of vanilla PCA. However, we also observe vanilla and HPCA to have higher residual kurtosis when we regress the eigenportfolios' returns onto the market returns. In contrast, the tensor PCA eigenportfolio has a significantly noisier relationship with the market betas, but less kurtosis when we regressed the tensor eigenportfolio returns onto the market returns. This is consistently the case as we adjust the sliding window length between 20 and 40 days. In addition, the tensor approach more frequently yields a long-only portfolio, which is a preferred property for a principal factor. In summary, these findings indicate that there is a general robustness gained by using the tensor structure and multilinear PCA.

The rest of the paper is organized as follows. In section 2 we introduce notation, present vanilla PCA eigenportfolios, and use asymptotic theory to show the principal eigenportfolio's convergence under a spike model with diverging eigenvalue (see Theorem 2.1). In section 3 we define tensors using sector classification, discuss the use of polyadic decompositions (PDs), and present results obtained using the MLSVD. Also in section 3, we present the HPCA method, provide analysis to confirm that the matrix form is in fact positive definite, and give a condition under which the eigenvectors of the HPCA correlation matrix can be computed from the sectorwise components (see Proposition 3.2). Section 4 concludes. In Appendix A we discuss in some detail the data that we use and their structure. Appendix B contains the proof for the main theorem of section 2, and C a brief review of Markowitz portfolio theory.

2. Eigenportfolios.

We consider a market consisting of N stocks where the return on stock i at time t is denoted by $r_i(t)$ for $1 \le i \le N$ and $1 \le t \le T$, which we write as a vector,

$$r(t) = \begin{pmatrix} r_1(t) \\ r_2(t) \\ \vdots \\ r_N(t) \end{pmatrix} , \tag{1}$$

for each t. The population mean and variance for asset i are denoted by

$$\mu_i = \mathbb{E} r_i(t) \ , \qquad \sigma_i^2 = \mathbb{E}(r_i(t) - \mu_i)^2 \ , \tag{2}$$

where $\sigma_i > 0$ for all i. In matrix/vector form

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_N \end{pmatrix} , \qquad \sigma = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_N \end{pmatrix} , \tag{3}$$

where σ is a diagonal matrix. The returns are standardized to remove the effects of the differing magnitudes of volatility across the portfolio of stocks and then stored in an $N \times T$

matrix,

$$R = \left[\frac{r_i(t) - \overline{r}_i}{h_i} \right]_{1 \le i \le N,\ 1 \le t \le T} , \tag{4}$$

where

$$\overline{r}_i = \frac{1}{T} \sum_{t=1}^{T} r_i(t) , \qquad h_i = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (r_i(t) - \overline{r}_i)^2} . \tag{5}$$

In matrix/vector form

$$\bar{r} = \begin{pmatrix} \bar{r}_1 \\ \bar{r}_2 \\ \vdots \\ \bar{r}_N \end{pmatrix}, \qquad h = \begin{pmatrix} h_1 & & & \\ & h_2 & & \\ & & \ddots & \\ & & & h_N \end{pmatrix} , \tag{6}$$

where h is a diagonal matrix. The empirical covariance matrix $\widehat{\Sigma}$ can be computed as

$$\widehat{\Sigma} = \frac{1}{T-1} \sum_{t=1}^{T} (r(t) - \bar{r})(r(t) - \bar{r})^* , \tag{7}$$

and the empirical correlation matrix

$$\widehat{\rho} = \frac{1}{T-1} RR^* = h^{-1} \widehat{\Sigma} h^{-1} . \tag{8}$$

Rather than calculating the eigenvalue decomposition of $\widehat{\rho}$, we extract the principal components of R using the SVD. This SVD represents R as

$$R = USV^*, \tag{9}$$

where

$$U = [U^1, U^2, \dots, U^N] \in \mathbb{R}^{N \times N}, \tag{10}$$

with orthonormal columns, S is an $N \times T$ diagonal matrix with the nonzero entries being R's singular values $S_{11} \ge S_{22} \ge \cdots \ge 0$, and

$$V = [V^1, V^2, \dots, V^T] \in \mathbb{R}^{T \times T}, \tag{11}$$

with orthonormal columns. The eigenvalues of $\hat{\rho}$ are then related to the singular values by

$$\lambda_i = \frac{1}{T-1} S_{ii}^2 \tag{12}$$

for $i \le N$. Note also that since $\text{Trace}(\hat{\rho}) = N$ we have that $\sum_1^N \lambda_i = N$.

2.1. Construction of eigenportfolios. Using the SVD computed in (9), the left singular vectors determine the portfolio's weights $\pi^i \in \mathbb{R}^N$ for $i = 1, 2, \ldots, N$, with

$$\pi^i \propto h^{-1} U^i \ ,$$

where h is the diagonal matrix in (6), and U^i is the ith column of U. The principal eigenportfolio is π^1,

$$\pi^1 = \frac{1}{c} h^{-1} U^1 \qquad \text{with} \qquad c = \sum_{i=1}^{N} U_i^1 / h_i \ , \tag{13}$$

where c normalizes so that the portfolio sums to unity and is assumed to be positive. For example, if the covariance matrix has all positive entries, then the Perron–Frobenius theorem ensures $c > 0$. For U.S. equities, only a small proportion of stocks have negative correlation, and these lead to negative weights in the principal eigenportfolio, that is, shorting, which can have a significant effect on the returns.

The occurrence of short positions in mean-variance portfolios is considered in Brennan and Lo (2010). When the principal eigenvalue is large compared to the rest, then the principal eigenportfolio is close to a mean-variance portfolio (Chen and Yuan (2016)) and so it will very likely have negative weights as noted by Brennan and Lo (2010). The assumption $c > 0$ is added here for completeness.

Eigenportfolios are orthogonal in the sense that there is zero empirical covariance among these portfolios' returns,

$$(\pi^i)^* \widehat{\Sigma} \pi^j \propto (U^i)^* \widehat{\rho} U^j = \lambda_i (U^i)^* U^j = 0 \qquad \text{for } i \neq j.$$

We denote by $f(t)$ the principal eigenportfolio return at time t,

$$f(t) = \sum_{i=1}^{N} r_i(t) \pi_i^1 \ , \tag{14}$$

which has mean $\bar{f} = \sum_i \bar{r}_i \pi_i^1$ and variance

$$\begin{aligned} h_f^2 &= \frac{1}{T-1} \sum_{t=1}^{T} (\pi^1)^* (r(t) - \bar{r})(r(t) - \bar{r})^* \pi^1 \\ &= \frac{1}{(T-1)c^2} (U^1)^* R R^* U^1 \\ &= \frac{\lambda_1}{c^2} \ . \end{aligned} \tag{15}$$

The principal eigenportfolio captures the most variance among all eigenportfolios and explains the most variation in the cross section of returns. Letting r_0 denote the risk-free rate, we introduce the cross-sectional regression,

$$r_i(t) = r_0 + b_i (f(t) - r_0) + \xi_i(t) \ , \tag{16}$$

where $\xi_i(t)$ is the residual left over after fitting with least squares. We can further compute the empirical covariance,

$$\begin{aligned}\widehat{\text{cov}}(r_i(t), f(t)) &= \frac{1}{T-1}\sum_{t=1}^{T}(r_i(t)-\bar{r})(f(t)-\bar{f}) \\ &= \frac{1}{T-1}\sum_{t=1}^{T} e_i^*(r(t)-\bar{r})(r(t)-\bar{r})^*\pi^1 \\ &= \frac{h_i}{c(T-1)}e_i^* RR^* U^1 \\ &= \frac{h_i\lambda_1}{c}U_i^1 \ ,\end{aligned}$$

where e_i is the ith standard unit vector. This yields the regression coefficients,

$$b_i = \frac{\widehat{\text{cov}}(r_i(t), f(t))}{h_f^2} = ch_i U_i^1 \ . \tag{17}$$

We also note that the principal eigenportfolio's standardized returns over the time interval are proportional to the first right-hand singular vector,

$$\left[\frac{f(t)-\bar{f}}{h_f}\right]_{1\le t\le T} = \frac{\sqrt{T-1}}{S_{11}}R^*U^1 = \sqrt{T-1}\ V^1 \ , \tag{18}$$

where V^1 is the first column of the $T \times T$ matrix V of right singular vectors in (11).

Remark 2.1. Equation (18) is derived as follows. Using (13), the standardized factor returns are

$$\frac{f(t)-\bar{f}}{h_f} = \frac{1}{h_f}\sum_i (r_i(t)-\bar{r}_i)\pi_i^1 = \frac{1}{h_f}\frac{(R^*U^1)_t}{c} \ .$$

Using (12) and (15), we have $h_f^2 = \lambda_1/c^2 = S_{11}^2/(c^2(T-1))$, and it follows that

$$\frac{f(t)-\bar{f}}{h_f} = \frac{(R^*U^1)_t}{c\sqrt{\lambda_1}}|c| = \text{sign}(c)\frac{(R^*U^1)_t}{S_{11}}\sqrt{T-1} \ .$$

Then using the SVD in (9), $R = USV^*$, we have $(R^*U^1)_t = S_{11}V_t^1$, and the formula in (18) follows since we have assumed that $c > 0$.

2.2. Principal eigenportfolio, market returns, and the spike model. The representation of the returns given by (16) is the one-factor PCA of the normalized data. An alternative representation of the returns relates them to the market portfolio,

$$r_i(t) = r_0 + \beta_i(r_m(t) - r_0) + \epsilon_i(t) \ , \tag{19}$$

where now $r_m(t)$ is the return on the capitalization-weighted, or market, portfolio and $\epsilon_i(t)$ is a mean-zero noise that is uncorrelated from $r_m(t)$. Equation (19) is the capital asset pricing

Table 1

Regression (23) for the principal eigenportfolio's daily returns, $f(t) - r_0 = \alpha + \beta(r_m(t) - r_0) + \varepsilon(t)$. The principal eigenvectors are computed using a 24-day sliding window with rebalancing every 10 days, from January 3, 1994, to December 31, 2020; the cap-weighted portfolio is also rebalanced every 10 days. The null hypotheses for the t-statistics above are $H_0 : \alpha = 0$ and $H_0 : \beta = 1$. The principal eigenportfolio has a β of .9588 which is significantly different from 1, and an α of approximately 2.4% per year which is significantly different from zero at the 5% level.

Variable	Estimate	Standard error	t-statistic	p-value
α	9.6132e-05	4.9327e-05	1.9489	0.051351
β	0.9588	0.0041345	-9.9647	3.1383e-23

Number of observations: 6799, *Error degrees of freedom:* 6797
Root mean squared error: 0.00407
R^2: 0.888; *Adjusted* R^2: 0.888
F-statistic vs. constant model: $5.38e+04$, *p-value* $= 0$
Residual excess kurtosis: 31.9361

model, which is reviewed briefly in C. The principal eigenportfolio returns and the market portfolio returns share significant colinearity, as seen from Figure 13 and the linear regression in Table 1 with the historical data. This relationship is also discussed in Avellaneda and Lee (2010) and in Potters et al. (2005). As already noted in the introduction, the principal eigenportfolio is constructed from the returns algebraically, without any financial insight, while construction of the capitalization-weighted portfolio is warranted by economic theory (see Lintner (1965); Ross (1976); Sharpe (1964)).

Letting Σ represent the model covariance for the stock returns,

$$\Sigma_{ij}(t) = \text{cov}(r_i(t), r_j(t)),$$

using (19), we have the representation

$$\Sigma = \text{var}(r_m(t))\beta\beta^* + \Omega \ , \tag{20}$$

where Ω is the covariance matrix of the model residuals with $\Omega_{ij}(t) = \text{cov}(\epsilon_i(t), \epsilon_j(t))$; this is the same covariance matrix factorization that was considered in spike models (Johnstone (2001)). We are interested in the case in which the principal eigenvalue dominates, as is the case with the equities returns data where λ_1 in (12) does in fact stick out. This can be seen in Figure 1.

We denote the estimator for the variance of the market portfolio $r_m(t)$ in (19) by

$$h_m^2 = \frac{1}{T-1}\sum_{t=1}^{T}(r_m(t) - \bar{r}_m)^2 \ , \qquad \bar{r}_m = \frac{1}{T}\sum_{t=1}^{T} r_m(t). \tag{21}$$

We want to show that as N, T tend toward infinity, with N/T tending to a constant, the large data limit, and when the top eigenvalue λ_1 is large, then it and the top eigenvector of $\widehat{\rho}$ (defined in (8)) will be "close" to the normalized market betas, $h_m^2\|\sigma^{-1}\beta\|^2$ and $\sigma^{-1}\beta$, respectively. For this we introduce the following assumptions in the form of conditions to be met.

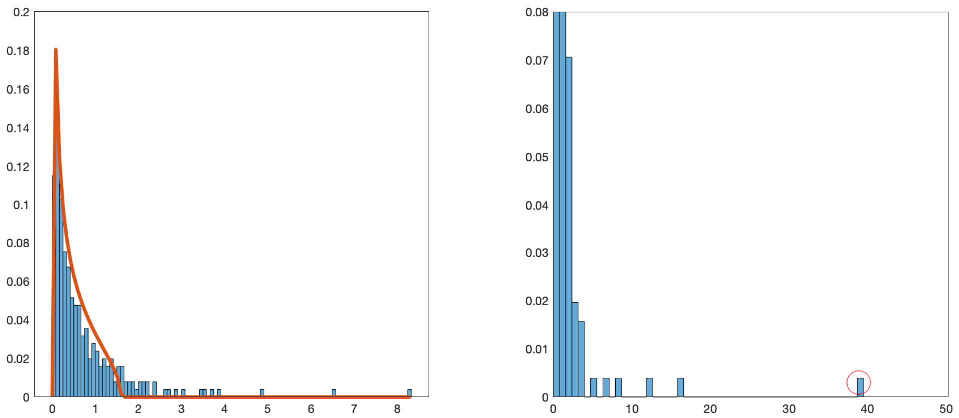


Figure 1. *The left panel shows the histogram of the low, bulk eigenvalues of the correlation of S&P* 500 *daily returns for the year* 2017*. An empirically fitted Marchenko–Pastur density of the bulk is shown with a red line. The right panel shows the full histogram, including the market (spike) eigenvalue on the far right, with a red circle around it.*

Condition 2.1. *The population parameters of the returns processes $r_i(t)$ and the standard deviation estimates h_m and the h_i's satisfy the following conditions:*

1. $\frac{1}{N}\|\sigma^{-1}\beta\|^2$ *has a finite positive limit as N and T tend toward infinity with $N/T \to \eta \in (0,\infty)$,*
2. *for noise matrix $\epsilon \in \mathbb{R}^{N\times T}$, $\limsup_N \frac{1}{T}\mathbb{E}[\sup_{\|u\|=1} u^*\sigma^{-1}\epsilon\epsilon^*\sigma^{-1}u] < \infty$ as $N \to \infty$ with $N/T \to \eta \in (0,\infty)$,*
3. *the estimator h_m has a finite positive limit in mean square as N and T tend toward infinity with $N/T \to \eta \in (0,\infty)$,*
4. *the diagonal matrix $h\sigma^{-1}$ converges to the identity in mean square (i.e., $\mathbb{E}\|h\sigma^{-1} - I\|^2 \to 0$, where I is the identity in $\mathbb{R}^N$) as N and T tend toward infinity with $N/T \to \eta \in (0,\infty)$.*

Items 1 and 2 of these conditions stipulate a large gap between the spike eigenvalue proportional to $\|\sigma^{-1}\beta\|^2$ in (20) and the spectrum of the residual, in this limit. For the S&P500 equity returns there is a large gap between the top eigenvalue and the other eigenvalues of the correlation matrix, as seen in the right panel of Figure 1. The outlier behavior of the top eigenvalue of the S&P500 correlations is well known, and in Potters, Bouchaud, and Laloux (2005) this top eigenvalue is called the market eigenvalue because the principal eigenportfolio captures the movements of the S&P500 index. In particular, item 2 in Condition 2.1 is shown to be true for large covariance matrices of purely random matrices with independent, identically distributed elements (Yin, Bai, and Krishnaiah (1988)). Here we assume this condition without introducing a model for the noise matrix. Items 3 and 4 are clearly consistency conditions with the needed uniformity.

We have the following theorem that connects the market betas with the principal eigenportfolio.

Theorem 2.1. *Assume Condition 2.1. Then for the first eigenvalue/eigenvector pair, $\lambda_1 = S_{11}^2/(T-1)$ and U^1, for the empirical correlation matrix of (8) as calculated using the SVD of (9), we have the limits*

$$\mathbb{E}\left|\frac{\lambda_1}{\|\sigma^{-1}\beta\|^2} - h_m^2\right| \to 0,$$
$$\mathbb{E}\left\|U^1 - \frac{1}{\|\sigma^{-1}\beta\|}\sigma^{-1}\beta\right\|^2 \to 0 \ ,$$

as $N, T \to \infty$ with $N/T \to \eta \in (0, \infty)$.

Proof. See Appendix B. ∎

Remark 2.2 (main idea of Theorem 2.1). From (20) the population correlation matrix is

$$\rho = \mathrm{var}(r_m(t))\sigma^{-1}\beta\beta^*\sigma^{-1} + \sigma^{-1}\Omega\sigma^{-1} \ .$$

If we take $u^1 = \sigma^{-1}\beta/\|\sigma^{-1}\beta\|$, then

$$(u^1)^*\rho u^1 = \mathrm{var}(r_m(t))\|\sigma^{-1}\beta\|^2 + (u^1)^*\sigma^{-1}\Omega\sigma^{-1}u^1 \ .$$

For large markets of equities, it is reasonable to assume $\|\sigma^{-1}\beta\|^2$ is of order N. Moreover, if we assume the $\epsilon_i(t)$'s are independent and identically distributed with bounded variance, then Ω is diagonal so that $\sup_{\|u\|=1} u^*\sigma^{-1}\Omega\sigma^{-1}u \le 1$. Therefore, we have the following heuristic:

$$\frac{1}{N}(u^1)^*\rho u^1 \approx \frac{\mathrm{var}(r_m(t))}{N}\|\sigma^{-1}\beta\|^2 \ .$$

And, if we take any $\tilde{u} \perp u^1$ with $\|\tilde{u}\| = 1$, we have

$$\tilde{u}^*\rho\tilde{u} = \tilde{u}^*\sigma^{-1}\Omega\sigma^{-1}\tilde{u} \le 1 \ .$$

This suggests that for large N, ρ's top eigenvector is "close" to u^1, that the top eigenvalue behaves like $\mathrm{var}(r_m(t))\|\sigma^{-1}\beta\|^2$, and that higher-order eigenvalues are significantly smaller than N. The effect we see is that the principal eigenvalue will "stick out" from the bulk.

This heuristic argument for the population model is what motivates us to write Condition 2.1, parts 1, 2. Condition 2.1, parts 3 and 4, are consistency conditions that are used to carry through the proof for the empirical correlation matrix, as done in Appendix B.

The implications of Theorem 2.1 are perhaps better captured by the following proposition.

Proposition 2.2. *Assuming Condition 2.1 and that the population standard deviations satisfy $\limsup_N \|\sigma^{-1}\| < \infty$, then for the principal eigenportfolio weights π^1 given in (13) we have*

$$\left\|\pi^1 - \frac{\sigma^{-2}\beta}{\sum_i \beta_i/\sigma_i^2}\right\|^2 \to 0 \ ,$$

in probability as $N \to \infty$, where π^1 is the vector of principal eigenporfolio weights given by (13).

Proof. Condition 2.1, part 4, has $\|h\sigma^{-1} - I\|$ converging to zero in mean square, $\mathbb{E}\|h\sigma^{-1} - I\|^2 \to 0$, and we have that

$$\|h^{-1}\sigma - I\| = \|h^{-1}\sigma(I - \sigma^{-1}h)\| \leq \|h^{-1}\sigma\|\|I - \sigma^{-1}h\| \ .$$

From here, by using the reverse triangle inequality on $\|h^{-1}\sigma - I\|$, we have

$$0 \leq \left|1 - \frac{1}{\|h^{-1}\sigma\|}\right| \leq \frac{\|h^{-1}\sigma - I\|}{\|h^{-1}\sigma\|} \leq \|I - \sigma^{-1}h\| \to 0$$

in probability. Therefore, $\|h^{-1}\sigma\| \to 1$ in probability, and $\|h^{-1}\sigma - I\| \to 0$ in probability. In addition, if $\limsup_N \|\sigma^{-1}\| < \infty$, then $\|h^{-1} - \sigma^{-1}\| \leq \|\sigma^{-1}\|\|h^{-1}\sigma - I\| \to 0$, in probability as $N \to \infty$.

From (13) and (17) we see that

$$\pi^1 = \frac{h^{-2}b}{\sum_i b_i/h_i^2} = \frac{h^{-1}U^1}{\sum_i U_i^1/h_i} \ ,$$

where in the numerators we have $h^{-1}U^1$, for which

$$\begin{aligned} &\left\|h^{-1}U^1 - \sigma^{-1}\frac{\sigma^{-1}\beta}{\|\sigma^{-1}\beta\|}\right\| \\ &\leq \left\|h^{-1}\left(U^1 - \frac{\sigma^{-1}\beta}{\|\sigma^{-1}\beta\|}\right)\right\| + \left\|(h^{-1} - \sigma^{-1})\frac{\sigma^{-1}\beta}{\|\sigma^{-1}\beta\|}\right\| \\ &\leq \|h^{-1}\sigma\| \, \|\sigma^{-1}\| \left\|U^1 - \frac{\sigma^{-1}\beta}{\|\sigma^{-1}\beta\|}\right\| + \|h^{-1} - \sigma^{-1}\| \\ &\to 0 \end{aligned}$$

in probability as $N \to \infty$, which follows from Theorem 2.1 provided that $\limsup_N \|\sigma^{-1}\| < \infty$. Convergence of the denominator follows from similar calculations, and hence convergence of the ratio follows and proves the statement of the proposition. ■

Proposition 2.2 tells us that, in the large data limit, the principal eigenportfolio weights behave like the market betas divided by variance and normalized. We will see in the next section, and as seen in Figures 3 through 9, that the equity returns data give results that are consistent with Theorem 2.1 and Proposition 2.2, with Condition 2.1 being sufficient for this theory.

Denote the asymptotic market beta of the principal eigenportfolio as β_π, which from Proposition 2.2 is

$$\beta_\pi = \frac{\sum_i (\beta_i/\sigma_i)^2}{\sum_i \beta_i/\sigma_i^2} \ . \tag{22}$$

Figure 2 shows a time series plot of the β_π from 1994 to 2020, estimated using a 24-day sliding window, with reestimation every 10 days. As can be seen in the plot, β_π appears to be near 1, but the distribution of estimators is skewed upward, which means that there is likely to be

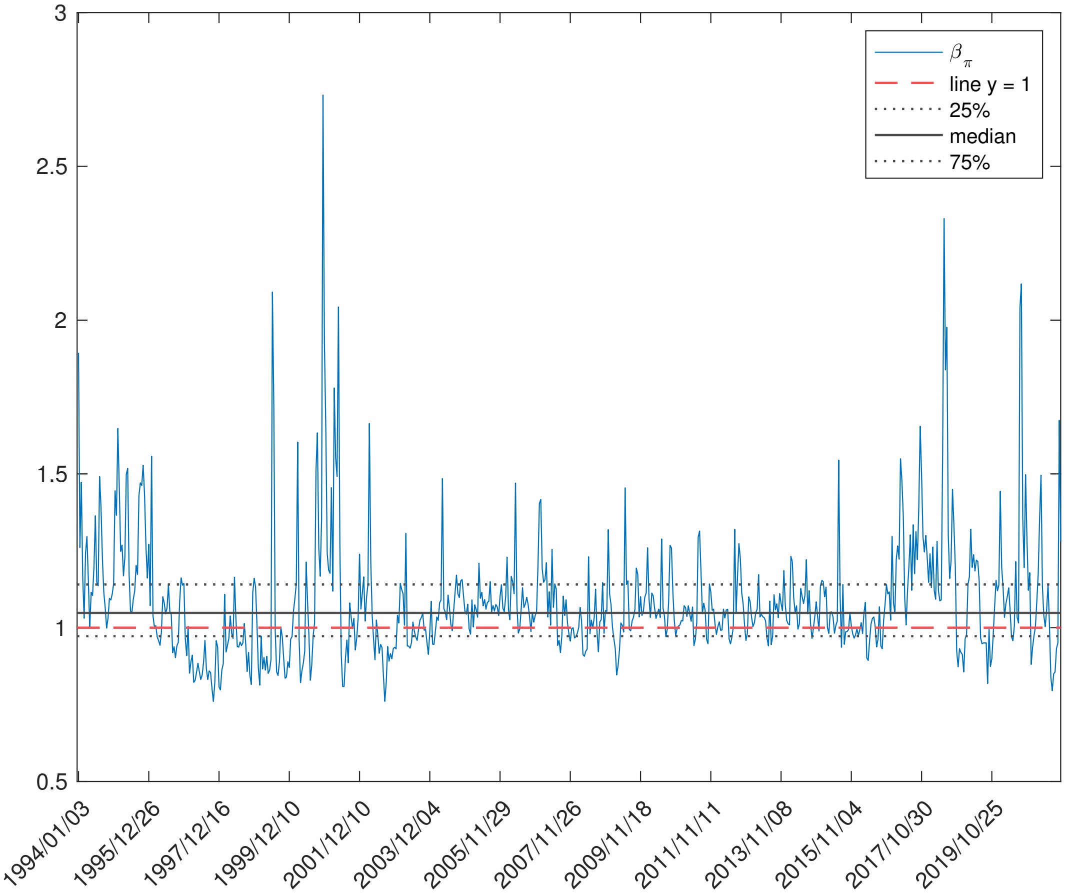


Figure 2. *The time series of the principal eigenportfolio's estimated market beta. The theoretical value β_π is given by (22). The estimators are computed using a 24-day sliding window, with reestimation every 10 days, from January 1994 to December 2020. The distribution's median is near to 1, but the upward skew is an indication that there will be some tracking error with the market portfolio.*

some tracking error between the principal eigenportfolio and the market portfolio. To explain the relationship between eigenportfolios and the market portfolio, we paraphrase a similar comment from Avellaneda et al. (2020) for implied volatilities. Suppose that market betas are estimated using data points $\{1, 2, \dots, T\}$, and assume for any $t' \notin \{1, 2, \dots, T\}$ that $\epsilon(t')$ is independent of the estimation data. Using Proposition 2.2, and assuming $\limsup_N \| \sigma^{-1} \| < \infty$, the residual between the principal eigenportfolio and the market portfolio is

$$\begin{aligned}
&\sum_{i=1}^{N} \pi_i^1 r_i(t') - \beta_\pi r_m(t') \\
&\quad = \sum_{i=1}^{N} \left(\pi_i^1 \beta_i - \frac{(\beta_i/\sigma_i)^2}{\sum_i \beta_i/\sigma_i^2} \right) r_m(t') + \sum_{i=1}^{N} \pi_i^1 \epsilon_i(t') \\
&\quad \to \varepsilon(t'),
\end{aligned}$$

where $\varepsilon(t') = \lim_{N\to\infty} \sum_{i=1}^N \pi_i^1 \epsilon_i(t')$, in probability. Thus, in the limit, the principal eigenportfolio has a market beta of β_π given by (22), plus a residual. In this paper, we will see empirically that this residual will be small, that is, the R^2 after regressing the principal eigenportfolio onto the market portfolio will be around .9.

Similarly, returns of orthogonal portfolios, i.e., portfolios such that $\widetilde{\pi} \propto h^{-1}\widetilde{U}$ with $\widetilde{U} \in \text{span}(U^2, U^3, \ldots, U^N)$, tend toward a limit that has zero correlation with the market portfolio (i.e., the limit is market neutral). That is, any portfolio

$$\widetilde{\pi} \in \text{span}(h^{-1}U^2, h^{-1}U^3, \ldots, h^{-1}U^N)$$

(i.e., any orthogonal portfolio) will tend toward market neutrality,

$$\sum_{i=1}^N \widetilde{\pi}_i r_i(t') = \left(\sum_{i=1}^N \widetilde{\pi}_i \beta_i\right) r_m(t') + \sum_{i=1}^N \widetilde{\pi}_i \epsilon_i(t') \to \widetilde{\varepsilon}(t'),$$

where $\widetilde{\varepsilon}(t') = \lim_{N\to\infty} \sum_{i=1}^N \widetilde{\pi}_i \epsilon_i(t')$.

We end this section with the following remark that connects the principal eigenportfolio to the market betas in another way.

Remark 2.3. From C it is apparent that the market betas are given by $\beta_i = \frac{\mu_i - r_0}{\mu_m - r_0}$ where r_0 is the risk free rate, μ_i is given by (2), and $\mu_m = \mathbb{E} r_m(t)$. It follows from Proposition 2.2 that for a fixed $i < \infty$,

$$\pi_i^1 \propto \frac{\mu_i - r_0}{\sigma_i^2} + \text{“noise”},$$

where "noise" tends toward zero as $N \to \infty$. In other words, the entries of the principal eigenvector U^1 are proportional to the optimal allocation ratios $(\mu_i - r_0)/\sigma_i^2$ as N grows (recall from (13) that $\pi^1 \propto h^{-1}U^1$). We may say that for N large and fixed, the principal eigenportfolio is equivalent to solving N-many scalar optimizations for each asset and then combining them for a global portfolio. Another way to say this is that the principal eigenportfolio "discovers" the optimal allocation ratios of the equity returns in the large data limit of Proposition 2.2.

2.3. Principal eigenportfolio regression onto market returns. In the previous sections we took a theoretical, modeling point of view that led to some interesting asymptotic relations between the principal eigenportfolio of returns and the market portfolio. We now consider how these theoretical results hold up for the U.S. equities daily returns data as described in Appendix A.

Figures 3 through 9 show the sorted principal eigenportfolio weights π^1, red line, given by (13) and plotted alongside the estimated β_i/σ_i^2's, blue dots, sorted according to the ranking indices of the eigenportfolio. The estimator of each β_i is

$$\widehat{\beta}_i = \frac{1}{h_m^2(T-1)} \sum_{t=1}^T (r_i(t) - \bar{r}_i)(r_m(t) - \bar{r}_m) \ ,$$

with $T = 24$ days, where π^1, $\widehat{\beta}$, and h are recalculated every 10 days. In a given plot, each blue dot shown in Figures 3–9 is proportional to the $\widehat{\beta}_i/h_i^2$ for some i. According to Proposition 2.2,

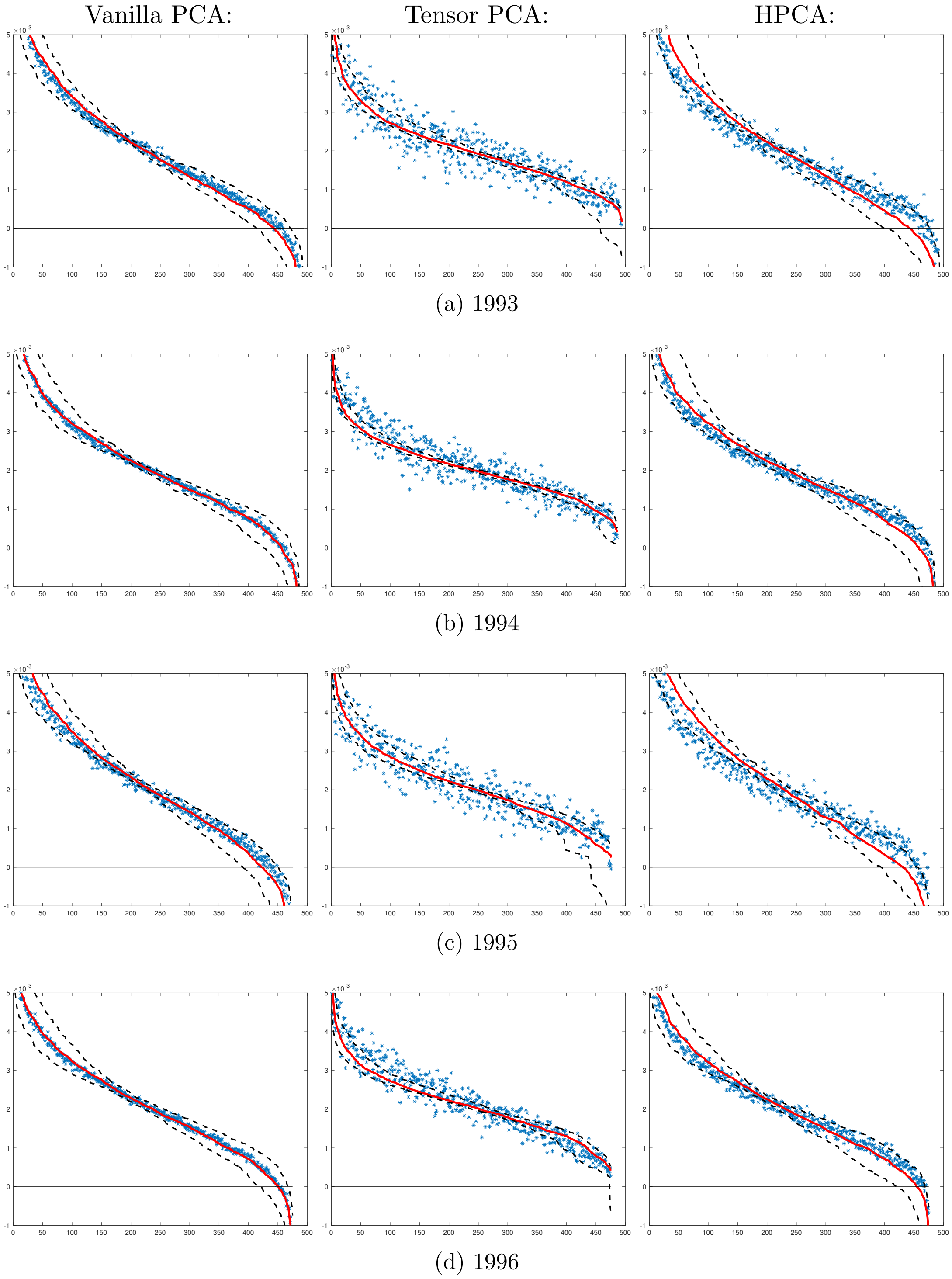


Figure 3. *Median of sorted eigenportfolio weights (solid line), 10% and 90% quantiles of sorted eigenportfolio weights (dashed lines), and the median of* β_i/σ_i^2 *'s sorted with the same index as the sorted eigenportfolios.*

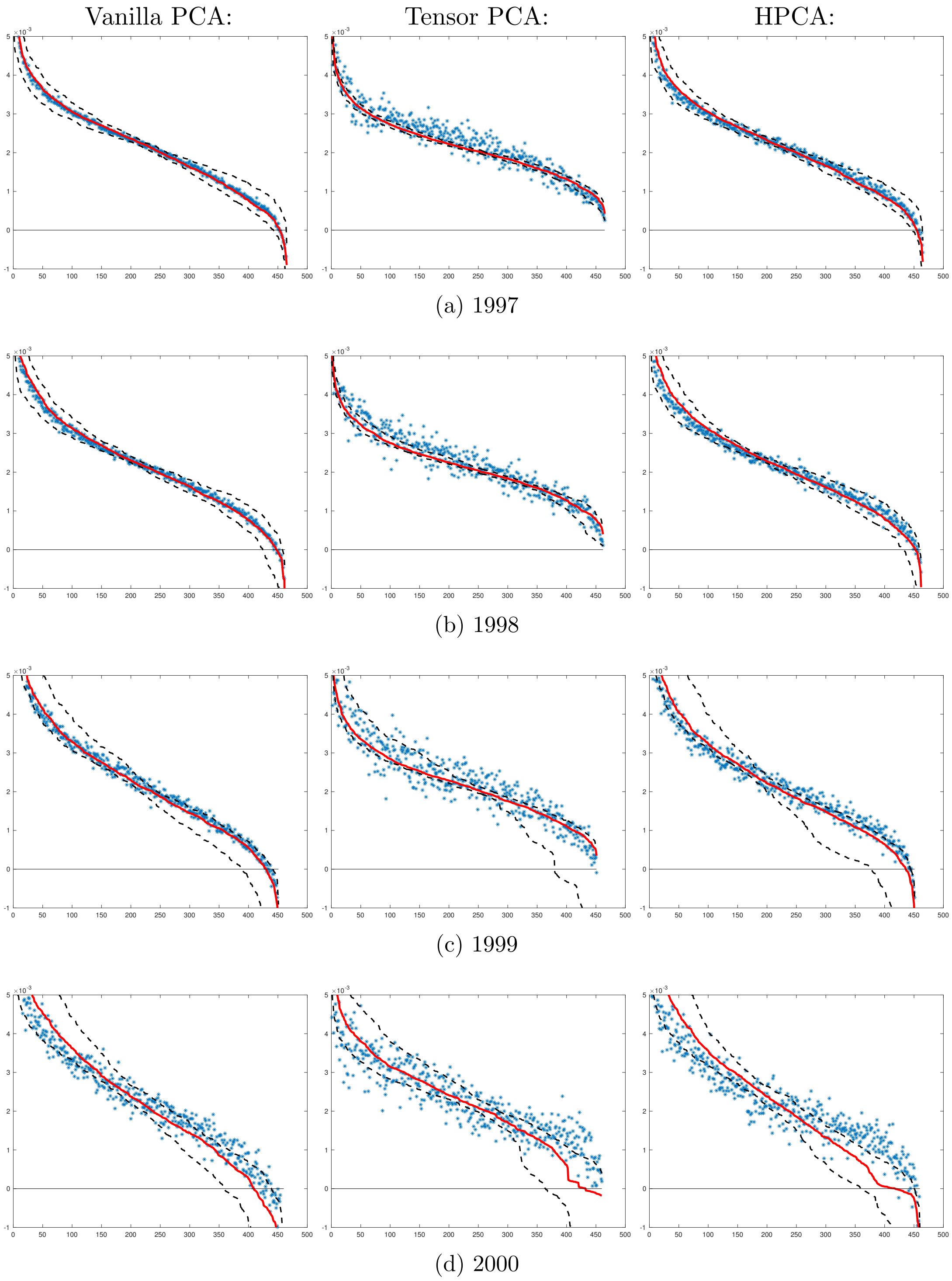
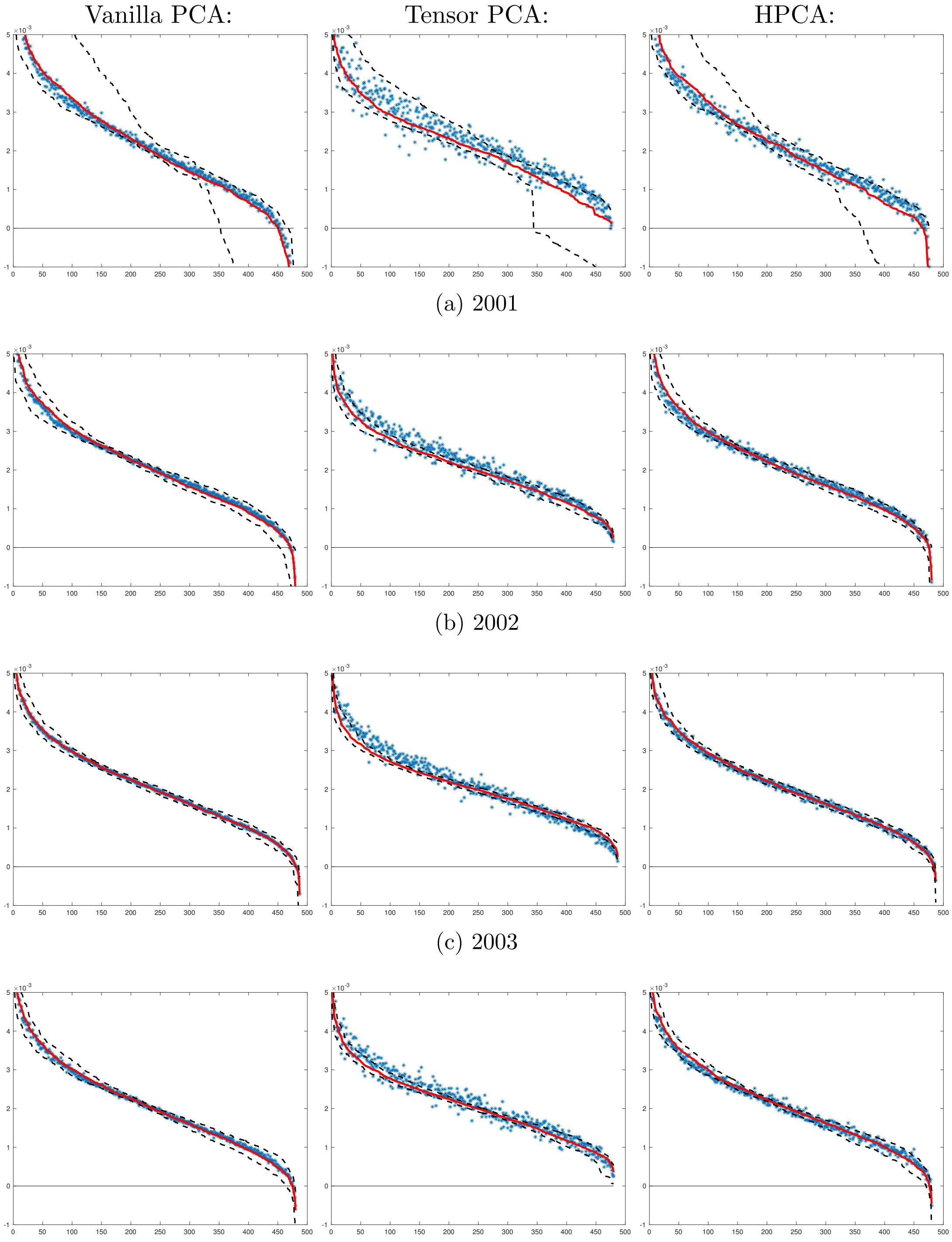


Figure 4. *Median of sorted eigenportfolio weights (solid line), 10% and 90% quantiles of sorted eigenportfolio weights (dashed lines), and the median of* β_i/σ_i^2 *'s sorted with the same index as the sorted eigenportfolios.*



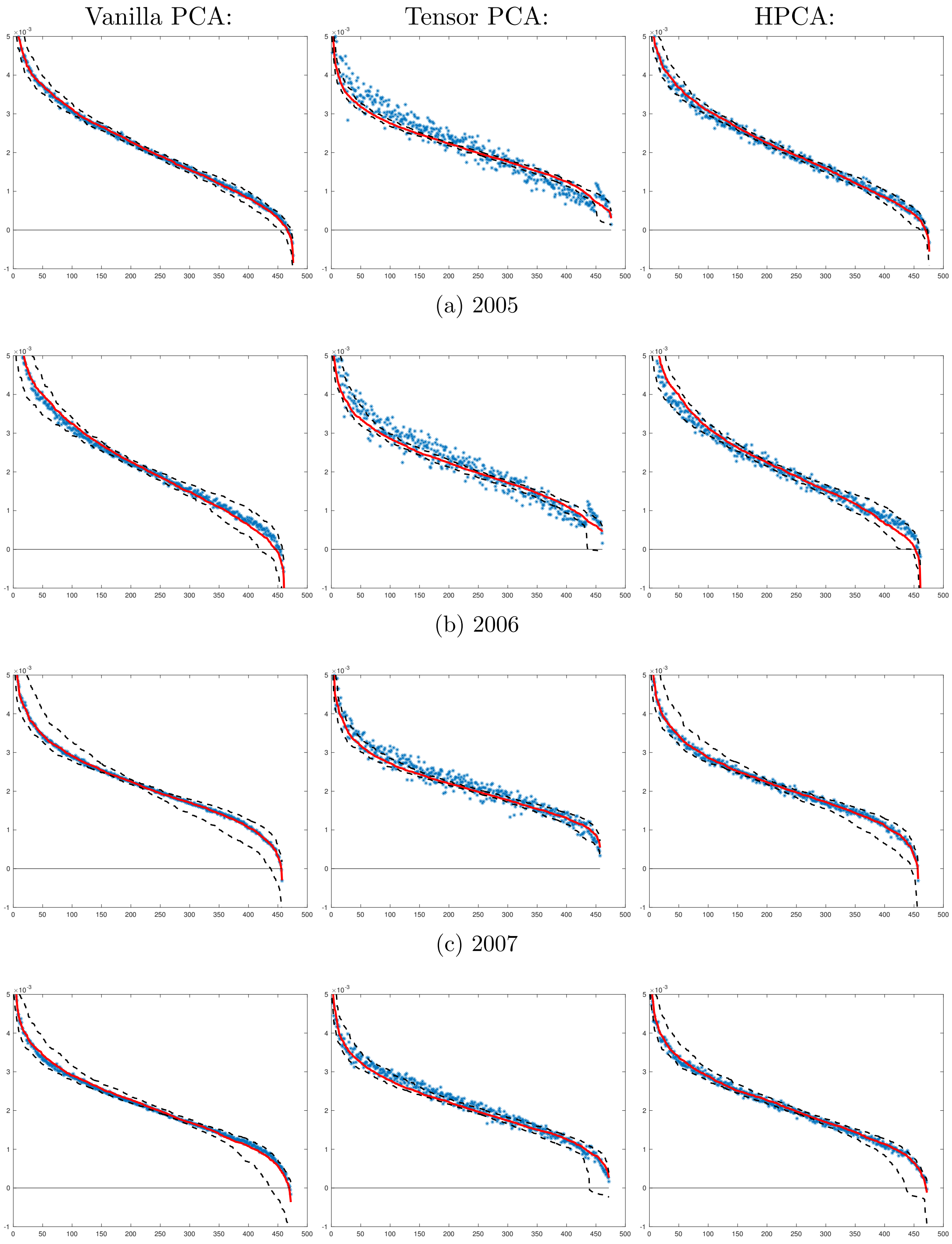
(a) 2001

(b) 2002

(c) 2003

(d) 2004

Figure 5. *Median of sorted eigenportfolio weights (solid line), 10% and 90% quantiles of sorted eigenportfolio weights (dashed lines), and the median of β_i/σ_i^2 's sorted with the same index as the sorted eigenportfolios.*



(a) 2005

(b) 2006

(c) 2007

(d) 2008

Figure 6. *Median of sorted eigenportfolio weights (solid line),* 10% *and* 90% *quantiles of sorted eigenportfolio weights (dashed lines), and the median of* β_i/σ_i^2 *'s sorted with the same index as the sorted eigenportfolios.*

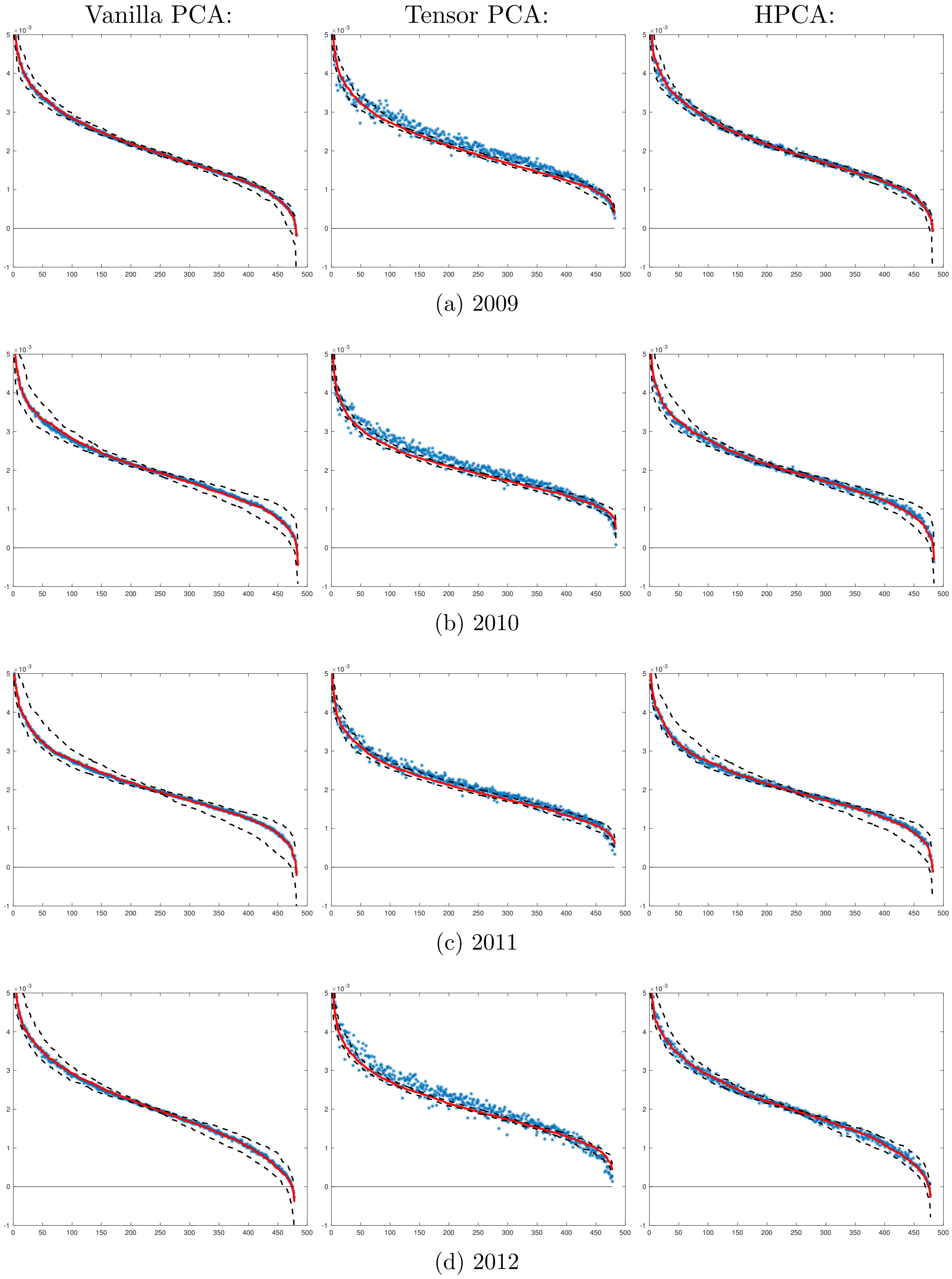


Figure 7. *Median of sorted eigenportfolio weights (solid line),* 10% *and* 90% *quantiles of sorted eigenportfolio weights (dashed lines), and the median of* β_i/σ_i^2 *'s sorted with the same index as the sorted eigenportfolios.*

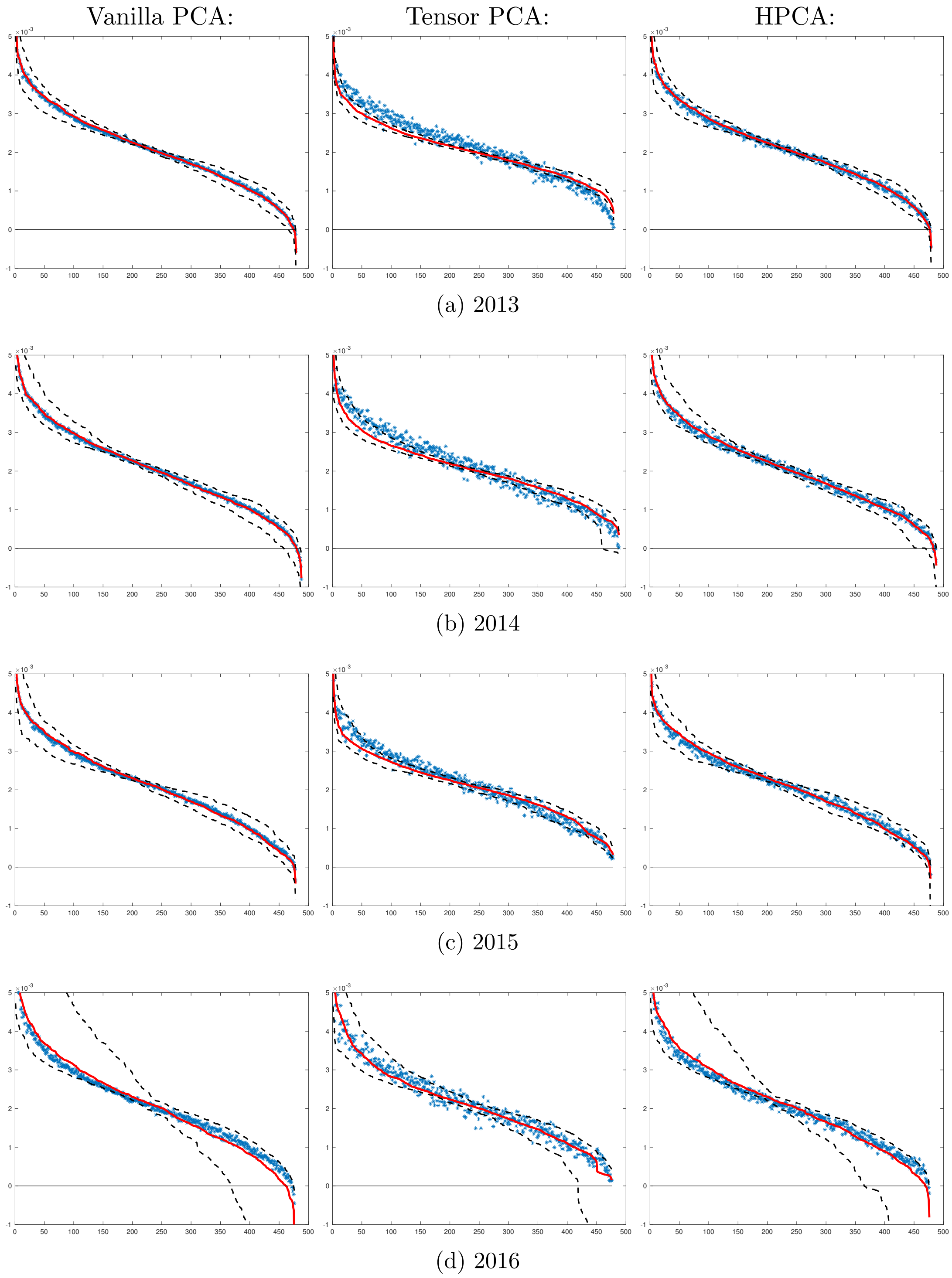


Figure 8. *Median of sorted eigenportfolio weights (solid line), 10% and 90% quantiles of sorted eigenportfolio weights (dashed lines), and the median of* β_i/σ_i^2 *'s sorted with the same index as the sorted eigenportfolios.*

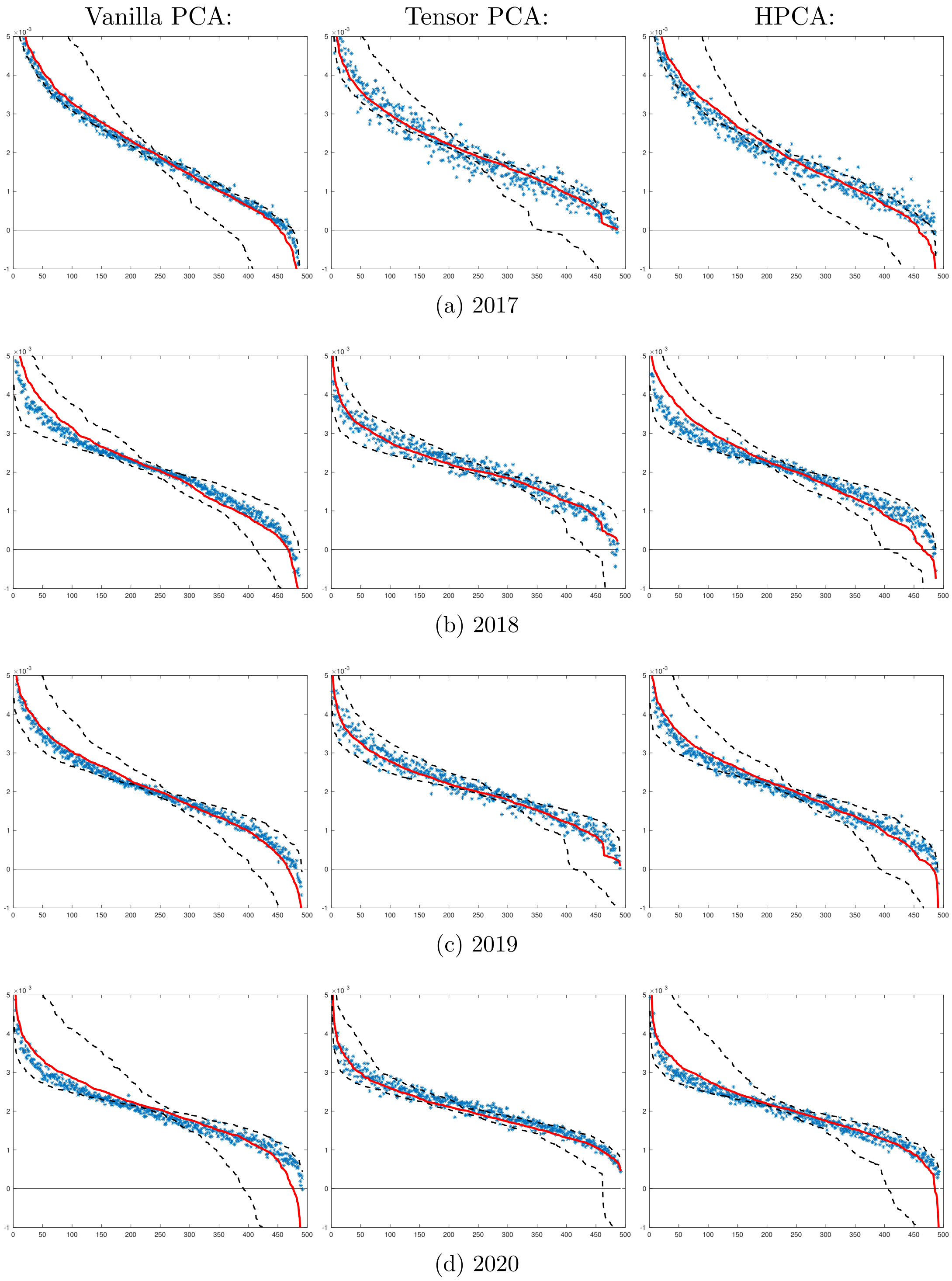


Figure 9. *Median of sorted eigenportfolio weights (solid line), 10% and 90% quantiles of sorted eigenportfolio weights (dashed lines), and the median of β_i/σ_i^2 's sorted with the same index as the sorted eigenportfolios.*

the eigenportfolio weights and their respective β_i/σ_i^2's should have similar rank. Figures 3–9 show orderings of the principal eigenportfolio weights that are consistent with the orderings of the β_i/σ_i^2's,

Figure 11 plots the cumulative returns of the eigenportfolio, the cap-weighted portfolio, and the SPY ETF (with dividends reinvested) from January 3, 1994, through to December 31, 2020. The SPY ETF with reinvested dividends is the benchmark to measure portfolio returns. It is clear from these time series that the principal eigenportfolios have a systemic excess return. The significance of these excess returns is evaluated with a linear regression of the principal eigenportfolio returns $f(t)$ onto the market returns,

$$f(t) - r_0 = \alpha + \beta(r_m(t) - r_0) + \varepsilon(t) \ , \tag{23}$$

where $r_m(t)$ is the cap-weighted market portfolio from the data. Table 1 shows the results of regression (23), where a 24-day sliding-window of data was used to calculate each eigenportfolio, with rebalancing occurring every 10 days. The table shows an α that is significantly different from zero at the 5% level and a β that is significantly different from 1 at the 10^{-23} level.

The significance of this α raises an eyebrow because the theory we've presented thus far has characterized the principal eigenportfolio as a factor—not as an α-producing portfolio. However, we can explain this α by censoring the short positions. That is, on any given day if $\sum_i |\pi_i^1| \mathbf{1}_{\pi_i^1 < 0} > 3\%$, then we discard the eigenportfolio and use the capitalization weights instead (i.e., we censor returns from negative positions). The results from the censored regression are shown in Table 4 and the counts of days with over 3% shorting per year are shown in Table 5. The α of the censored regression is no longer significant, which is an indication that short positions were responsible for the excess return. Visually, the cumulative censored returns can be seen in Figure 12, wherein the days with more than 3% shorting the portfolio uses the cap-weighted portfolio instead.

The impact of negative positions in the eigenportfolio (along with the significance of the α) raises the question of whether or not the principal eigenportfolio is in fact meaningful as a principal factor. Traditional mean-variance theory says that a principal factor should be a long-only portfolio, but as shown in the Brennan and Lo (2010), it is entirely possible that all frontier portfolios have at least one short position. Therefore, the negative weights we are finding in the principal eigenportfolio are not a misspecification, but are within the framework of a modernized portfolio theory that admits a tangency portfolio with short positions.

3. Portfolios from hierarchical data structures.

In this section we want to consider the same questions as in section 2 but with industry sector information taken into account. Our approach is to separate the stocks into sectors and form tensor structures, and then construct principal eigenportfolios from these structures. The relation of these eigenportfolios to the market portfolio is essentially the same as that of the vanilla PCA results in the previous sections but there are some interesting differences. Another reason that tensor structures are of interest is because we may have other data in addition to returns, such as traded volumes. This is the case with options data considered in Avellaneda et al. (2020), where the natural data structures are tensors.

To introduce the tensor data structures, we recall that the data is described in Appendix A and note that the CRSP database contains an industry classification for each stock. We will

Sector		Industry Group	
10	Energy	1010	Energy
15	Materials	1510	Materials
20	Industrials	2010	Capital Goods
		2020	Commercial & Professional Services
		2030	Transportation
25	Consumer Discretionary	2510	Automobiles & Components
		2520	Consumer Durables & Apparel
		2530	Consumer Services
		2550	Retailing
30	Consumer Staples	3010	Food & Staples Retailing
		3020	Food, Beverage & Tobacco
		3030	Household & Personal Products
35	Health Care	3510	Health Care Equipment & Services
		3520	Pharmaceuticals, Biotechnology & Life Sciences
40	Financials	4010	Banks
		4020	Diversified Financials
		4030	Insurance
45	Information Technology	4510	Software & Services
		4520	Technology Hardware & Equipment
		4530	Semiconductors & Semiconductor Equipment
50	Communication Services	5010	Telecommunication Services
55	Utilities	5510	Utilities
60	Real Estate	6010	Real Estate

Figure 10. *GICS sectors and industry groups as provided by MSCI. These are the* 11 *sectors that we use for the hierarchical tensor and the HPCA methods in section* 3. *These sectors are assigned by market mediators and while they may be somewhat arbitrary they are used throughout the finance industry and therefore provide a useful reference for the purposes of this paper. Eigenportfolios based on sector/clusters statistically estimated from the data were not considered here.*

use this to assign each stock/company to a sector. While these assignments are not always perfect due to some companies having much in common with companies in more than one sector, they nonetheless represent a reasonable allocation of stocks to industry sectors.

For the case of U.S. equities there are 11 industry sectors, details of which are provided in Figure 10, and we index these sectors by $k = 1, 2, \ldots, M$ (e.g., $M = 11$). Let sector k contain N_k stocks so that the total number of assets $N = N_1 + N_2 + \cdots + N_M$. Once again let T be the number of trading days in the sample. We can define a coordinate mapping from the space of indices for a given sector k to the space of indices for the entire set of stocks as

$$\tau_k : \{1, \ldots, N_k\} \to \{1, \ldots, N\},$$

i.e., the ith stock in the kth sector is the $\tau_k(i)$th stock in the whole space. Let R^k be the $N_k \times T$ matrix of standardized returns in sector k,

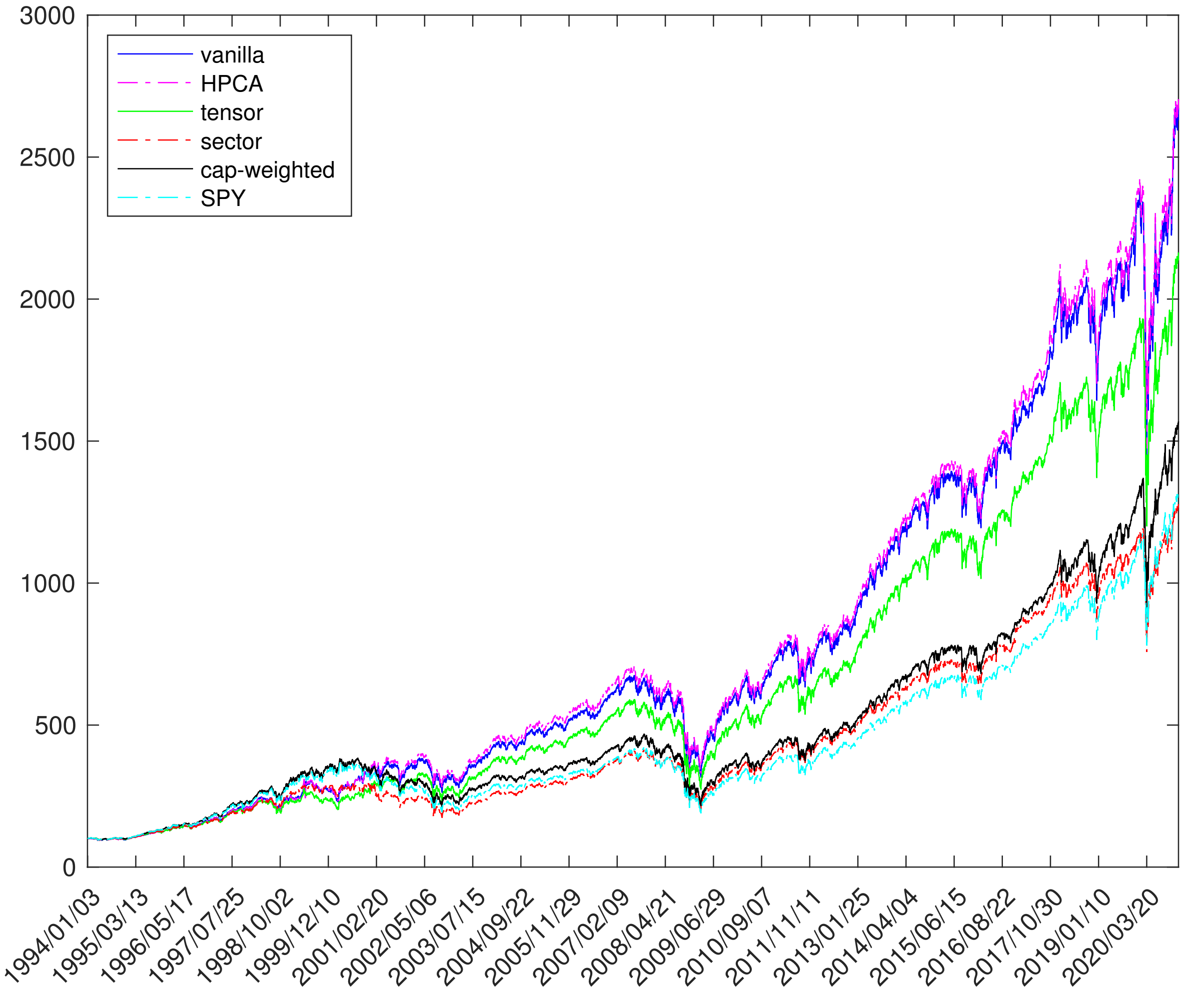


Figure 11. *The eigenportfolios obtained from vanilla PCA, the hierarchical tensor, HPCA, and a vanilla PCA on the* 11*-sector ETFs. The portfolio weights are computed using a* 24*-day sliding window, from* 1994 *through* 2020*, with rebalancing every* 24 *days. Also shown is the cap-weighted portfolio, which is also rebalanced every* 10 *days.*

$$R_{it}^k = \frac{r_{\tau_{k(i)}}(t) - \overline{r}_{\tau_{k(i)}}}{h_{\tau_{k(i)}}}$$

for $1 \leq i \leq N_k$ and $1 \leq t \leq T$, where $\overline{r}_{\tau_{k(i)}} = \frac{1}{T}\sum_{t=1}^{T} r_{\tau_{k(i)}}(t)$ and $h_{\tau_{k(i)}}^2 = \frac{1}{T-1}\sum_{t=1}^{T}(r_{\tau_{k(i)}}(t) - \overline{r}_{\tau_{k(i)}})^2$. For each k let $T_k = \min(T, N_k)$. The compact SVD of R^k is

$$R^k = U^k S^k (V^k)^*, \tag{24}$$

where U^k and V^k are $N_k \times T_k$ and $T \times T_k$ matrices, respectively, given by

$$U^k = [U^{1k}, U^{2k}, \ldots, U^{T_k k}],$$
$$V^k = [V^{1k}, V^{2k}, \ldots, V^{T_k k}],$$

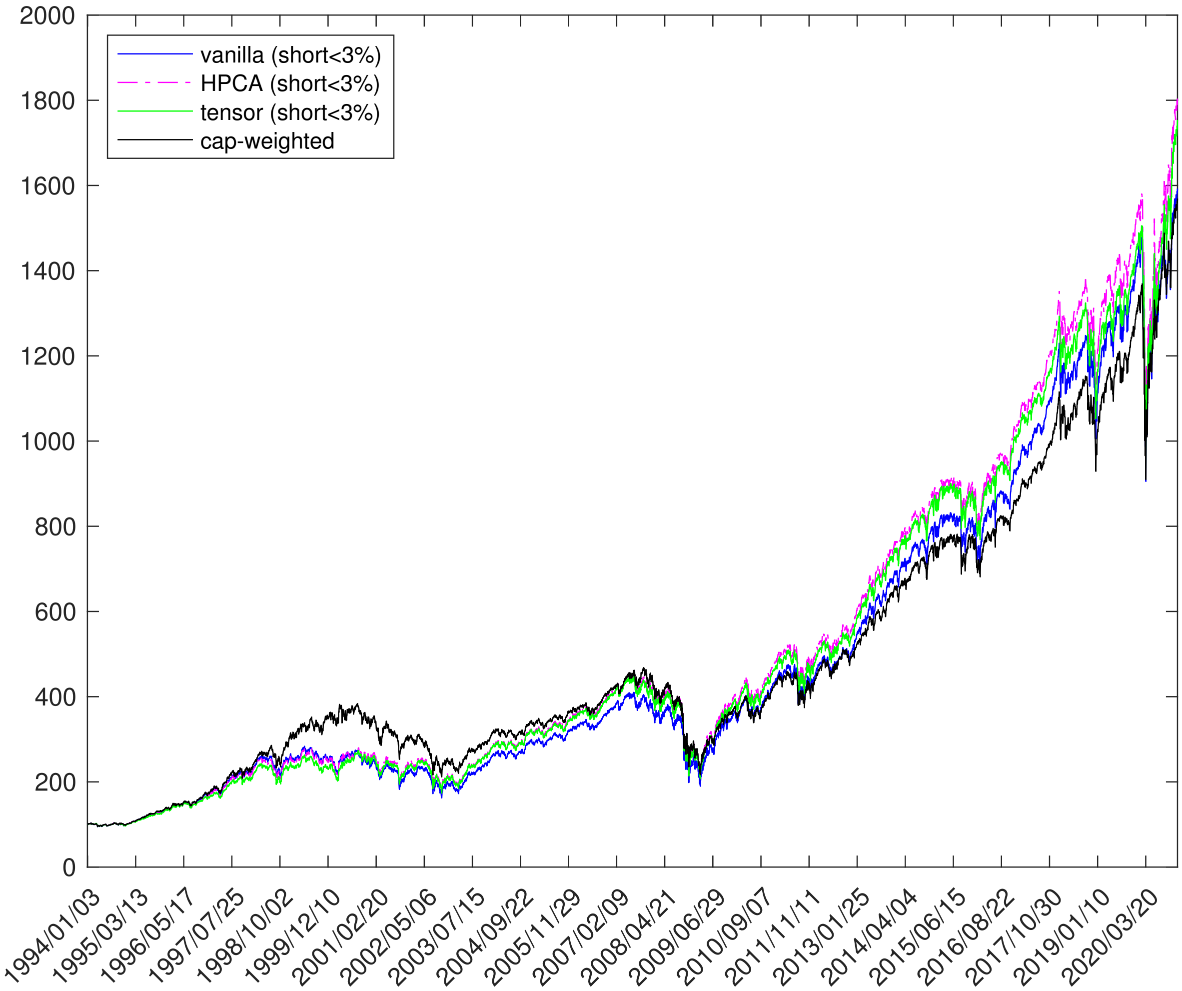


Figure 12. *The same as Figure 11, but with an eigenportfolio taking capitalization weights on the days where its short positions exceed 3% of total portfolio weight. These returns are similar to those analyzed in Table 4, except that days exceeding 3% are replaced with the returns of the cap-weighted portfolio, as opposed to being discarded like they are in the regression tables.*

each with columns orthonormal to one another (note that columns of U^k are not assumed to be orthogonal to columns of V^k), and S^k is a $T_k \times T_k$ matrix with nonzero entries only along the main diagonal with those being the sorted singular values, $S^k_{11} \geq S^k_{22} \geq \cdots \geq S^k_{T_kT_k} \geq 0$.

Let us define the following sector index functions:

$$I(i) = k \text{ if the } i\text{th stock is in the } k\text{th sector.} \tag{25}$$

We extend the sectorwise singular vectors to the larger space $\mathbb{R}^N$, defining them as

$$W_j^{ik} = \begin{cases} U^{ik}_{\tau^{-1}_{k(j)}} & \text{if } I(j) = k, \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

There is orthogonality among these extended vectors,

$$(W^{ik})^* W^{j\ell} = \begin{cases} 1 & \text{if } i = j \text{ and } k = \ell, \\ 0 & \text{otherwise }, \end{cases}$$

which is easy to check. Within each sector we have a principal factor portfolio with returns,

$$f^k(t) = \frac{1}{c^k} \sum_{i=1}^{N} r_i(t)(h^{-1}W^{1k})_i \quad \text{with} \quad c^k = \sum_{i=1}^{N} W_i^{1k}/h_i \tag{27}$$

for $k = 1, 2, \ldots, M$.

We will next construct normalized returns tensors using the sector notation and quantities we have introduced.

3.1. The hierarchical returns tensor.

We define an $N \times T \times M$ tensor $\mathbf{R}$ such that

$$\mathbf{R}(i,t,k) = \begin{cases} R^k_{\tau^{-1}_{k(i)},t} & \text{if } I(i) = k, \\ \dfrac{f^{I(i)}(t) - \bar{f}^{I(i)}}{h^{I(i)}} & \text{otherwise}, \end{cases} \tag{28}$$

where $\bar{f}^k = \frac{1}{T}\sum_{t=1}^{T} f^k(t)$, $h^k = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(f^k(t) - \bar{f}^k)^2}$, and where $I(i)$ is the sector index function defined in (25). This is referred to as the "hierarchical tensor." This tensor has a sector number k as the third index, and if the sector of stock i is not equal to k, then the tensor's entry is the standarized returns of sector $I(i)$'s principal factor portfolio. We recall that a tensor's PD is its representation as the sum of rank-one tensors. Utilizing the SVD in (24) and the formula from (18) for each sectorwise factor's standardized returns, a PD for the returns tensor is

$$\mathbf{R} = \sum_{k=1}^{M} \sum_{i=1}^{T_k} S^k_{ii} \left(W^{ik} \circ V^{ik} \circ e^k \right) + \sqrt{T-1} \sum_{k=1}^{M} \sum_{\substack{\ell=1 \\ \ell \neq k}}^{M} \text{sign}(c^\ell) \left(\mathbf{1}^\ell \circ V^{1\ell} \circ e^k \right) \ , \tag{29}$$

where $e^k \in \mathbb{R}^M$ is the kth canonical basis vector, "$\circ$" denotes vector outer product, and

$$\mathbf{1}^k_i = \begin{cases} 1 & \text{if } I(i) = k, \\ 0 & \text{otherwise}. \end{cases}$$

The expression in (29) is a sum of $\sum_{k=1}^{M}(T_k + M - 1)$ rank-one tensors, and it is possible that the rank of $\mathbf{R}$ is $\sum_{k=1}^{M}(T_k + M - 1)$, but in general the rank of a tensor with a decomposition as given in (29) will be less.

Definition 3.1 (tensor rank). *Let $\mathbf{X}$ denote a tensor in $\mathbb{R}^{N\times T\times M}$. The rank of $\mathbf{X}$ is the minimal integer d needed to write $\mathbf{X}$ as a sum of rank-one tensors,*

$$\mathbf{X} = \sum_{i=1}^{d} X^{i1} \circ X^{i2} \circ X^{i3}, \tag{30}$$

where $X^{i1} \in \mathbb{R}^N$, $X^{i2} \in \mathbb{R}^T$, and $X^{i3} \in \mathbb{R}^M$ for each i.

A PD as given by Definition 3.1 is referred to as canonical (a CPD) when d is equal to the rank of $\mathbf{X}$ (Kolda and Bader (2009)). In general it is an NP-hard problem to determine the rank of a tensor of dimension 3 or greater and so we cannot be sure that the PD for $\mathbf{R}$ in (29) is minimal and hence (29) is unlikely to be a CPD. Nonetheless, (29) is still useful to us when we search for low-rank decompositions of the tensor, which is done by solving the following optimization:

$$\begin{aligned} &\min_{\mathbf{X}\in\mathcal{R}^n} \|\mathbf{R}-\mathbf{X}\|^2_{fro} \\ &\qquad \text{s.t} \\ &\mathcal{R}^n = \left\{\mathbf{X}\in\mathbb{R}^{N\times T\times M} \middle| \text{rank}(X)\le n\right\}, \end{aligned} \tag{31}$$

where n is the predetermined rank.

The best rank-one approximation can be used for a principal eigenportfolio. We prefer to use the MSVD of the next section, however, even though for the U.S. equities data differences between CPD and MLSVD are insignificant when it comes to constructing principal eigenportfolios.

Multilinear singular value decomposition. The tensor analogue of SVD is the MLSVD, which is a decomposition of the form

$$\mathbf{R} = \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{M} \mathbf{S}(i,t,k)\left(U^{(i1)}\circ U^{(t2)}\circ U^{(k3)}\right), \tag{32}$$

where $U^{(1)} = [U^{(11)}, U^{(21)}, \dots, U^{(N1)}] \in \mathbb{R}^{N\times N}$ is a matrix with orthonormal columns, $U^{(2)} = [U^{(12)}, U^{(22)}, \dots, U^{(T2)}] \in \mathbb{R}^{T\times T}$ is a matrix with orthonormal columns, and $U^{(3)} = [U^{(13)}, U^{(23)}, \dots, U^{(M3)}] \in \mathbb{R}^{M\times M}$ is a matrix also with orthonormal columns. The decomposition in (32) is an MLSVD with basis elements $U^{(1)}$, $U^{(2)}$, $U^{(3)}$ and tensor core $\mathbf{S} \in R^{N\times T\times M}$. The third-order tensor S is not diagonal anymore, in general. However, it has the property of *all orthogonality*,

$$\sum S(i,t,k)S(i',t',k') = 0, \tag{33}$$

where the indices (i,t,k) and (i',t',k') are equal except for one of the three components and the sum is over the two equal components. This means that distinct "slices" of each orientation are orthogonal. Moreover, the indices can be permuted so that the sums of squares over all except for one index are ordered by size. The MLSVD representation (32) is exact and is obtained from the application of SVDs to all the possible flattenings of the tensor and organizing the output suitably to get the result; see Kolda and Bader (2009), De Lathauwer, De Moor, and Vandewalle (2000), and Cichocki et al. (2015). The MLSVD can be used for solving problems like (31) for low n (e.g., $n = 1$ or $n = 11$). From the MLSVD the top component for constructing a principal eigenportfolio should be $X = U^{(11)}$, and hence the portfolio weights are

$$\pi^1 = \frac{h^{-1}U^{(11)}}{\sum_i h_i^{-1}U_i^{(11)}} .$$

This is the same method for computing an eigenportfolio that was used in Avellaneda et al. (2020) for eigenportfolios of implied volatilities.

Results with equity returns data. The middle columns of Figures 3–9 display the eigenportfolio weights for the hierarchical tensor against the estimated β_i/σ_i^2 as predicted by Theorem 2.1; the leftmost column in the same plots are the analogous plot for vanilla PCA. There is a good correspondence between the weights predicted by the theory, assumed to carry over to tensors, and those extracted from the data, but there is more dispersion evident in the cloud-like formation of the estimates around the sorted eigenportfolio.

Compared with the vanilla eigenportfolio, the MLSVD rank-one estimation results in an eigenportfolio that is more relaxed. One possible explanation for this is that the tensor approach adds noise to weights as a way to create a stabilizing effect, similar to the addition of noise in an underdetermined linear system for a stable solution. In comparison, the vanilla PCA weights are in closer alignment with the β_i/σ_i^2's.

We also observe that the MLSVD eigenportfolio is more frequently long only compared to the vanilla eigenportfolio; Table 5 displays the counts of days with over 3% shorting per year, which are noticeably lowered for the MLSVD eigenportfolio. Similar to the vanilla regression shown in Table 1, there is a significant excess return when the MLSVD-constructed eigenportfolio is regressed onto the market returns (see Table 2), but this significance disappeared when we ran the 3% censored regression (see Table 4). Indeed, the cumulative returns of the hierarchical-tensor eigenportfolio can be seen in Figure 11 and the cumulative 3% censored returns in Figure 12, and in the latter there is no noticeable excess return for the MLSVD-constructed eigenportfolio.

Finally, we observe that returns from the MLSVD eigenportfolio have systemically less kurtosis in their residuals after regression onto the market portfolio. That is, we regressed the returns from the eigenportfolios constructed using the various methodologies onto the market portfolio's returns. We did this for varying sliding-window lengths and rebalancing intervals. What we see in Table 3 is that the MLSVD construction systematically shows residuals with the lowest excess kurtosis among the different methodologies. We can also see the residual outliers visually in Figure 13 for the eigenportfolios computed using a 24-day sliding window with rebalancing every 10 days. Bearing in mind that each block of returns is an out-of-sample test, it therefore may be the case that vanilla PCA is overfitting, whereas the tensor PCA, because it considers interdependence at the sector level, is producing a coarse-grained vector that works better out-of-sample.

3.2. Hierarchical PCA.

HPCA is used in Avellaneda (2020) for a similar equities data set. The HPCA method uses the natural separation by sector to organize the data matrix in a way that focuses on the principal component of each sector. The idea is to impose the following block correlation structure to the returns:

$$\widehat{\text{corr}}(r_i(t), r_j(t)) = \begin{cases} \frac{1}{T-1}\Big(R^{I(i)}(R^{I(j)})^*\Big)_{ij} & \text{if} \quad I(i) = I(j), \\ \frac{b_i b_j}{h_i h_j}\widehat{\text{cov}}\Big(f^{I(i)}(t), f^{I(j)}(t)\Big) & \text{if} \quad I(i) \neq I(j), \end{cases} \tag{34}$$

Table 2

Regression (23) for tensor PCA, HPCA, and sector ETF eigenportfolios' daily returns, $f(t) - r_0 = \alpha + \beta(r_m(t) - r_0) + \varepsilon(t)$. The principal vectors are computed using a 24-day sliding window with rebalancing every 10 days, from January 3, 1994, to December 31, 2020; the cap-weighted portfolio is also rebalanced every 10 days. The null hypotheses for the t-stats above are $H_0 : \alpha = 0$ and $H_0 : \beta = 1$. This table should be compared with Table 1, which contains the results for vanilla PCA.

Tensor PCA regressions				
Variable	Estimate	Standard error	t-statistic	p-value
α	6.7672e-05	4.007e-05	1.6889	0.091291
β	0.93555	0.0033586	-19.189	5.7629e-80

Number of observations: 6799; *error degrees of freedom:* 6797
Root mean squared error: 0.0033
R^2: 0.919; *adjusted* R^2: 0.919
F-statistic vs. constant model: $7.76e+04$, $p - value = 0$
Residual excess kurtosis: 18.3381

HPCA regression				
Variable	Estimate	Standard error	t-statistic	p-value
α	0.00010382	6.0133e-05	1.7265	0.084303
β	0.9504	0.0050402	-9.8411	1.0616e-22

Number of observations: 6799; *error degrees of freedom:* 6797
Root mean squared error: 0.00496
R^2: 0.84; *adjusted* R^2: 0.839
F-statistic vs. constant model: $3.56e+04$, $p - value = 0$
Residual excess kurtosis: 177.6654

Sector ETF PCA regression				
Variable	Estimate	Standard error	t-statistic	p-value
α	1.3242e-05	3.8467e-05	0.34423	0.73068
β	0.90728	0.0032243	-28.756	1.0494e-171

Number of observations: 6799; *error degrees of freedom:* 6797
Root mean squared error: 0.00317
R^2: 0.921; *adjusted* R^2: 0.921
F-statistic vs. constant model: $7.92e+04$, $p - value = 0$
Residual excess kurtosis: 58.3926

where each f^k is the sectorwise factor given by (27), the ith return's loading on its respective factor is $b_i = h_i c^k W_i^{1k}$ (recall the b_i computed in (17)) with W^{1k} given by (26). The structure of the HPCA correlation matrix (34) is a block matrix with the same index separations as the tensor (28), but differs because it imposes a specific covariance structure between stocks across sectors. Using the SVD $R^k = U^k S^k (V^k)^*$ from (24), the covariance between the kth and the ℓth factors is

$$\widehat{\text{cov}}(f^k(t), f^\ell(t)) = \frac{S_{11}^k S_{11}^\ell (V^{1k})^* V^{1\ell}}{(T-1)c^k c^\ell},$$

where $V^{1k} \in \mathbb{R}^T$ and $V^{1\ell} \in \mathbb{R}^T$ are each proportional to the time series of their respective standardized factor returns, as given in (18) for the entire market factor $f(t)$. The off-block structure leads to the following block-matrix structure for the correlation matrix:

Table 3

Robustness: tensor PCA eigenportfolio never results in high excess kurtosis because it has no outliers relative to the cap-weighted returns.

		Excess kurtosis of residuals			
Training window	Rebalancing period	Vanilla	Tensor	HPCA	Sector ETFs
20	5	444.81	21.96	1495.36	125.44
20	10	69.36	25.56	506.85	28.89
20	20	73.78	26.49	155.13	26.21
24	5	29.85	18.67	150.40	122.92
24	10	31.94	18.34	177.67	58.39
24	20	23.17	18.24	41.51	29.14
28	5	228.58	14.83	1572.81	53.73
28	10	207.84	15.08	1171.70	31.51
28	20	13.34	13.30	14.87	25.10
32	5	15.45	18.46	20.36	44.01
32	10	16.82	18.73	22.07	34.64
32	20	18.74	20.19	27.31	42.48
36	5	16.16	13.88	1367.43	27.54
36	10	16.72	13.85	885.74	26.05
36	20	22.22	13.45	598.87	31.33
40	5	14.62	17.55	15.36	20.91
40	10	14.38	17.89	15.45	20.85
40	20	14.23	18.58	15.51	23.71

$$\widehat{\rho}^{\mathrm{H}} = \begin{bmatrix} \frac{1}{T-1}R^1(R^1)^* & G_{12}U^{11}(U^{12})^* & \dots & G_{1M}U^{11}(U^{1M})^* \\ G_{21}U^{12}(U^{11})^* & \frac{1}{T-1}R^2(R^2)^* & \dots & G_{2M}U^{12}(U^{1M})^* \\ \vdots & & & \vdots \\ \vdots & & \ddots & \vdots \\ G_{M1}U^{1M}(U^{11})^* & G_{M2}U^{1M}(U^{12})^* & \dots & \frac{1}{T-1}R^M(R^M)^* \end{bmatrix}, \tag{35}$$

where $G \in \mathbb{R}^{M\times M}$ with

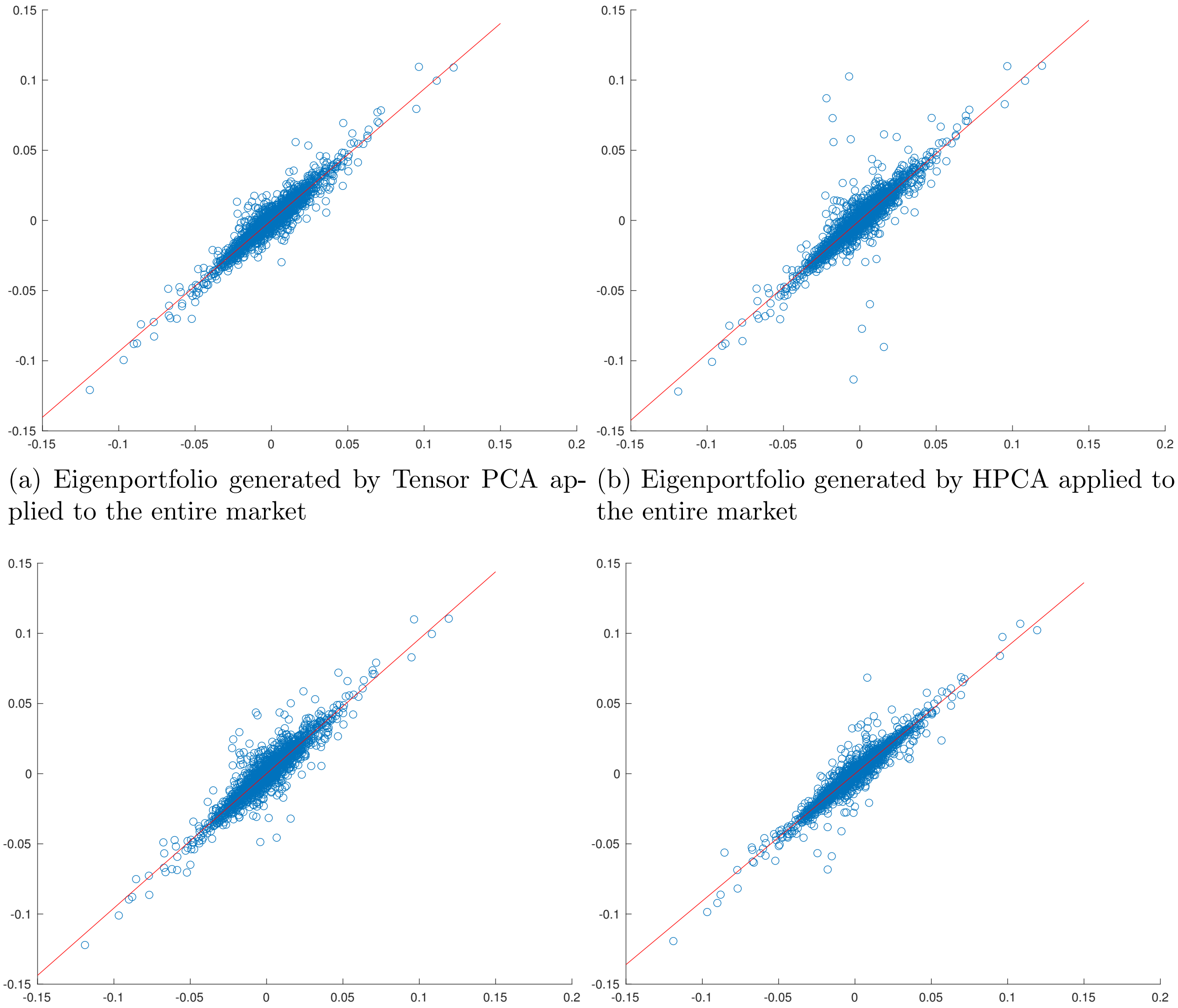
$$G_{k\ell} = \widehat{\mathrm{cov}}(f^k(t), f^\ell(t))c^k c^\ell = \tfrac{1}{T-1} S^k_{11} S^\ell_{11} (V^{1k})^* V^{1\ell}\ .$$

The matrix G will be used the prove the following proposition.

Proposition 3.2. *The block matrix $\widehat{\rho}^H$ given in (35) is positive semidefinite, and strictly positive definite if and only if $S^k_{ii} > 0$ for all $i \leq T_k$ and for all k. If $\frac{1}{M}\sum_{k=1}^M (S^k_{11})^2 > \max_{\ell\leq M}(S^\ell_{22})^2$, then the principal eigenvector of $\widehat{\rho}^H$ is $\sum_{k=1}^M a_k W^{1k}$, where $a \in \mathbb{R}^M$ is the principal (normalized) eigenvector of G.*

Proof. Consider any vector $a \in \mathbb{R}^M$ with $\|a\| = 1$ that is an eigenvector of matrix G. Then $\sum_{k=1}^M a_k W^{1k}$ is an eigenvector of $\widehat{\rho}^H$,

$$\widehat{\rho}^{\mathrm{H}} \sum_{k=1}^M a_k W^{1k} = \sum_{k=1}^M a_k \sum_{\ell=1}^M G_{k\ell} W^{1\ell} = \lambda \sum_{k=1}^M a_k W^{1k}\ ,$$



(a) Eigenportfolio generated by Tensor PCA applied to the entire market

(b) Eigenportfolio generated by HPCA applied to the entire market

(c) Eigenportfolio generated by Vanilla PCA applied to the sector ETFs

(d) Eigenportfolio generated by Vanilla PCA applied to the entire market

Figure 13. *Returns of eigenportfolios versus returns of the cap-weighted portfolio, for each of the individual methods for principal eigenportfolio construction. The tensor PCA in panel* (a) *clearly provides a closer fit to the market than the alternative approaches and this finding is consistent across estimation windows of different lengths*

where λ is the eigenvalue such that $Ga = \lambda a$. The matrix G has all nonnegative eigenvalues because it is the covariance matrix of the $c^k f^k(t)$'s constructed from (27). Hence, $\lambda \geq 0$. For the remaining $\sum_{k=1}^{M}(T_k - 1)$ orthogonal W^{ik}'s, we have

$$\widehat{\rho}^{\mathrm{H}} W^{ik} = \frac{1}{T-1}(S_{ii}^k)^2 W^{ik} \qquad \text{for } i \geq 2,$$

which shows that any W^{ik} with $i \geq 2$ is an eigenvector with nonnegative eigenvalue. Hence, $\widehat{\rho}^{\mathrm{H}}$ is positive semidefinite, and strictly positive definite if and only if $S_{ii}^k > 0$ for all $1 \leq i \leq T_k$

Table 4

Regression (23) with 3% censored data for vanilla PCA, tensor PCA, and HPCA eigenportfolios' daily returns, $f(t) - r_0 = \alpha + \beta(r_m(t) - r_0) + \varepsilon(t)$. Censoring at 3% means the regression excludes days where short portfolio exceeds 3%, if $\sum_i |\pi_i| \mathbf{1}_{\pi_i<0} > 3\%$. The principal eigenvectors are computed using a 24-day sliding window with rebalancing every 10 days, from January 3, 1994, to December 31, 2020; the cap-weighted portfolio is also rebalanced every 10 days. The null hypotheses for the t-stats above are $H_0 : \alpha = 0$ and $H_0 : \beta = 1$.

Vanilla PCA regression (short less than 3%)				
Variable	Estimate	Standard error	t-statistic	p-value
α	1.4178e-05	4.0001e-05	0.35445	0.72301
β	0.97226	0.0032337	-8.4557	3.4758e-17
Number of observations: 5791; *error degrees of freedom:* 5789				
Root mean squared error: 0.00304				
R^2: 0.94; *adjusted* R^2: 0.94				
F-statistic vs. constant model: $9.05e+04$, $p-value = 0$				
Residual excess kurtosis: 16.1060				
Tensor PCA regression (short less than 3%)				
Variable	Estimate	Standard error	t-statistic	p-value
α	3.7959e-05	3.7609e-05	1.0093	0.31287
β	0.93859	0.0031097	-19.748	2.6789e-84
Number of observations: 6366; *error degrees of freedom:* 6364				
Root mean squared error: 0.003				
R^2: 0.935; *adjusted* R^2: 0.935				
F-statistic vs. constant model: $9.11e+04$, $p-value = 0$				
Residual excess kurtosis: 15.5670				
HPCA Regression (short less than 3%)				
Variable	Estimate	Standard error	t-statistic	p-value
α	3.8705e-05	4.1533e-05	0.93191	0.35142
β	0.95955	00.0033675	-12.012	7.2178e-33
Number of observations: 6044; *error degrees of freedom:* 6042				
Root mean squared error: 0.00323				
R^2: 0.931; *adjusted* R^2: 0.931				
F-statistic vs. constant model: $8.12e+04$, $p-value = 0$				
Residual excess kurtosis: 18.5770				

and for all k. Finally, if $\frac{1}{M}\sum_{k=1}^{M}(S_{11}^k)^2 > \max_{\ell \le M}(S_{22}^\ell)^2$, then for λ the top eigenvalue of G we have

$$\lambda \ge \frac{1}{M}\text{trace}(G) = \frac{1}{M(T-1)}\sum_{k=1}^{M}(S_{11}^k)^2 \ge \frac{1}{T-1}(S_{22}^\ell)^2 \quad \forall \ell$$

and thus λ is the principal eigenvalue of $\widehat{\rho}^{\mathrm{H}}$. ■

Remark 3.1. In all the backtesting runs with equity returns that we did in this paper, taking $M = 11$ for the 11 industry sectors, the condition $\frac{1}{M}\sum_k (S_{11}^k)^2 > \max_k (S_{22}^k)^2$ from Proposition 3.2 never failed to be true.

Results with equity returns data. The rightmost columns of Figures 3–9 show the sorting of the HPCA eigenportfolio weights along with the estimated β_i/σ_i^2's. The sortings are tighter than the noisy-banded sorting for tensors in the middle columns, but are not as

Table 5

The number of days in each year for which the proportion of the tensor eigenportfolio with short positioning was greater than or equal to 5%/3%.

Short positions				
	Vanilla EP		Tensor EP	
Year	Days > 5%	Days > 3%	Days > 5%	Days > 3%
1994	30	70	0	0
1995	110	160	10	20
1996	20	40	10	20
1997	0	0	0	0
1998	10	30	0	10
1999	42	52	20	40
2000	112	142	40	62
2001	60	88	50	70
2002	10	10	0	10
2003	0	0	0	0
2004	0	0	0	0
2005	0	0	0	0
2006	0	20	0	0
2007	0	0	0	0
2008	0	10	0	0
2009	0	0	0	0
2010	0	0	0	0
2011	0	0	0	0
2012	0	0	0	0
2013	0	0	0	0
2014	0	10	0	0
2015	0	0	0	0
2016	50	70	30	30
2017	81	91	40	71
2018	60	80	10	30
2019	62	62	20	40
2020	63	73	30	30

tight as the sortings for vanilla PCA in the leftmost columns. Our conclusion after obervation of the HPCA eigenportfolio for varying sliding-window and reblancing lengths is that it is the least stable among the methodologies we examined. We ran the regression of the HPCA returns onto the market portfolio's, for which we see the results in Table 2 for 24-day sliding window with 10-day rebalancing. The HPCA eigenportfolio behaved similarly to the vanilla and the tensor eigenportfolios: there is a significant excess return that becomes insignificant when we censor with a 3% shorting threshold (see censored regression in Table 4 and the censored cumulative returns in Figure 12). But what is most striking is what we see in Table 3, namely, the consistently outsized kurtosis of the HPCA regression residuals compared to the other methodologies. In addition, the scatter plot for HPCA in Figure 13 (upper right) shows that the HPCA has outliers that are roughly of the same magnitude as vanilla PCA and the sector ETFs' eigenportfolio.

To summarize, both for this section and the results of sections 2.3 and 3.1, we have examined four methodologies, vanilla PCA, HPCA, tensor PCA, and vanilla PCA on the 11 sector

ETFs described in Appendix A. We find that vanilla, tensor, and HPCA eigenportfolios have an excess return that can be attributed to short positions, which we censor using a 3% threshold and we rerun regressions to find insignificant α's. Overall, each of these methodologies agrees with the predicted portfolio-weight behavior of Theorem 2.1, but it is perhaps the tensor portfolio that is the best candidate to use as a factor, as our empirical analyses indicate that it is more robust than the others. Moreover, if there is a concern over the amount of shorting in the factor portfolio, then the tensor eigenportfolio is better because it seems to result in the fewest short positions over time. In any case, the tensor construction is certainly better than the HPCA construction, the latter of which has considerably more kurtosis in its residuals after regression onto the market returns.

4. Summary and conclusions. We have analyzed the relationship between the principal eigenportfolio of equity returns and the associated capitalization weighted market portfolio. In section 2.2 we introduced a spike model for the returns data and showed in Theorem 2.1 that in a large data limit and large principal eigenvalue the weights of the principal eigenportfolio are close to the market betas divided by variance. The significance of this is that the results with the U.S. equities returns data are consistent with the asymptotic theory of the spike model, as seen in Figures 3 through 9. It is also significant in that this result provides a clear and somewhat surprising connection between the capitalization weighted portfolio and the principal eigenportfolio that comes from a purely algebraic construction. In the second part of the paper in section 3 the principal eigenportfolio is constructed from tensor forms of the data. The main differences with the vanilla PCA results of section 2 are looser consistency between the tensor eigenportfolio weights and the market betas over variance, a more frequently long-only portfolio, and when regressed onto the market portfolio the tensor has residuals with lower kurtosis. We may say that the tensor structure of the data relaxes the in-sample relation between the principal eigenportfolio weights and the market betas over variance, and is more robust over time. Quantifying this empirical observation remains a challenge. We should point out that option eigenportfolios can also be constructed from natural tensor structures as shown is Avellaneda et al. (2020). An important issue there is to find an analogue of the capitalization weighted portfolio of equities, which turns out to be an open-interest weighted portfolio of options.

Appendix A. The data. For our analysis we created 27 data sets, each one consisting of the constituents of the S&P500 index in January for each year from 1994 to 2020. From the Center for Research in Security Prices (CRSP) database we then retrieved, for each year, the sector information and the time series of adjusted daily returns from the first trading day in January to the last in December for each constituent. When creating the data set for a particular year we dropped any names that did not have returns available for all of the trading days in that year. For U.S. equities there are 11 sectors (see sector classifications in Figure 10).

Additionally, we obtained the returns for the 11 major sector ETFs, for the period 2005–2020, which we computed using daily adjusted close price provided by Yahoo! Finance. These ETFs represent the 11 GICS[1] sectors of the S&P500 (energy, materials, industrials, consumer

[1]General Industry Classification Standard.

discretionary, consumer staples, health care, financials, technology, communication services, utilities, and real estate). For the years 1994 to 2004 we computed synthetic ETFs using the returns and GICs in our CRSP data.

Using the 11 sector ETFs we can perform PCA to obtain an eigenportfolio using the same approach as we used for individual stocks. If we plot this eigenportfolio alongside the SPY we will see that it tracks the SPY and a regression reveals an α not significantly different from zero (see Table 2). Figure 11 shows the eigenportfolio and cap-weighted portfolios alongside the SPY ETF. The eigenportfolio is computed using a 24-day sliding window, which is slightly longer than one month of business days, and is rebalanced every 10 days. The same frequency is also used to rebalance the cap-weighted portfolio.

Appendix B. Proof of Theorem 2.1. Based on the model in (19), the empirical covariance matrix $\widehat{\Sigma}$ given by (7) can be written as

$$\widehat{\Sigma} = h_m^2\beta\beta^* + \frac{1}{T-1}\sum_{t=1}^{T}\nu(t)\nu^*(t) + \frac{1}{T-1}\sum_{t=1}^{T}\Big((r_m(t)-\bar{r}_m)\beta\nu^*(t) + (r_m(t)-\bar{r}_m)\nu(t)\beta^*\Big) , \tag{36}$$

where $\nu(t) = \epsilon(t) - \frac{1}{T}\sum_{t=1}^{T}\epsilon(t)$ and $\bar{r}_m$ is the market returns' sample mean as given in (21). Using $\widehat{\Sigma}$, the empirical correlation matrix is $\widehat{\rho} = h^{-1}\widehat{\Sigma}h^{-1}$, where h is the diagonal matrix of empirical standard deviations given by (5).

Throughout we will denote

$$u^1 = \frac{\sigma^{-1}\beta}{\|\sigma^{-1}\beta\|} , \tag{37}$$

which are a deterministic sequence of unit vectors indexed by N, with $u^1 \in \mathbb{R}^N$ for each N. The proof follows from some lemmas beginning with the next one.

Lemma B.1. *For u^1 given by (37) we have the limit*

$$\mathbb{E}\left|\frac{1}{\|\sigma^{-1}\beta\|^2}(u^1)^*\sigma^{-1}\widehat{\Sigma}\sigma^{-1}u^1 - h_m^2\right| \to 0 , \tag{38}$$

as $N \to \infty$.

Proof. For any deterministic $u \in \mathbb{R}^N$ with $\|u\| = 1$, using (36) we have

$$\begin{aligned}
&\left|u^*\sigma^{-1}\widehat{\Sigma}\sigma^{-1}u - h_m^2(u^*\sigma^{-1}\beta)^2\right| \\
&= \left|\frac{1}{T-1}u^*\sigma^{-1}\nu\nu^*\sigma^{-1}u + \frac{2}{T-1}\sum_{t=1}^{T}(r_m(t)-\bar{r}_m)u^*\sigma^{-1}\beta\nu^*(t)\sigma^{-1}u\right| \\
&\le \frac{1}{T-1}u^*\sigma^{-1}\nu\nu^*\sigma^{-1}u + \frac{2}{T-1}\left|\sum_{t=1}^{T}(r_m(t)-\bar{r}_m)u^*\sigma^{-1}\beta\nu^*(t)\sigma^{-1}u\right| \\
&\le \frac{1}{T-1}u^*\sigma^{-1}\nu\nu^*\sigma^{-1}u + \frac{2T}{T-1}\sqrt{\frac{1}{T}\sum_{t=1}^{T}(r_m(t)-\bar{r}_m)^2}\sqrt{\frac{1}{T}\sum_{t=1}^{T}(u^*\sigma^{-1}\beta\nu^*(t)\sigma^{-1}u)^2}
\end{aligned}$$

$$= \frac{1}{T-1} u^* \sigma^{-1} \nu \nu^* \sigma^{-1} u + \frac{2T|u^*\sigma^{-1}\beta|}{T-1} \sqrt{\frac{1}{T}\sum_{t=1}^{T}(r_m(t)-\bar{r}_m)^2} \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\nu^*(t)\sigma^{-1}u)^2}$$
$$= \frac{1}{T-1} u^* \sigma^{-1} \nu \nu^* \sigma^{-1} u + \frac{2Th_m|u^*\sigma^{-1}\beta|}{T-1} \sqrt{\frac{1}{T} u^* \sigma^{-1} \nu \nu^* \sigma^{-1} u}.$$

If we take u^1 defined by (37), then it follows that

$$\left| \frac{1}{\|\sigma^{-1}\beta\|^2} (u^1)^* \sigma^{-1} \widehat{\Sigma} \sigma^{-1} u^1 - h_m^2 \right|$$
$$\leq \frac{1}{(T-1)\|\sigma^{-1}\beta\|^2} \sigma^{-1}(u^1)^* \nu \nu^* \sigma^{-1} u^1 + \frac{2Th_m}{(T-1)\|\sigma^{-1}\beta\|} \sqrt{\frac{1}{T}(u^1)^* \sigma^{-1} \nu \nu^* \sigma^{-1} u^1}$$
$$\leq \frac{1}{(T-1)\|\sigma^{-1}\beta\|^2} \sigma^{-1}(u^1)^* \nu \nu^* \sigma^{-1} u^1 + \frac{2Th_m}{(T-1)\|\sigma^{-1}\beta\|} \sqrt{\frac{1}{T}(u^1)^* \sigma^{-1} \epsilon \epsilon^* \sigma^{-1} u^1}$$
$$\to 0 \,, \tag{39}$$

in probability as $N \to \infty$, where in (39) for the term with the product we have

$$\mathbb{P}\left(\frac{h_m}{\|\sigma^{-1}\beta\|} \sqrt{\frac{1}{T}(u^1)^* \sigma^{-1} \epsilon \epsilon^* \sigma^{-1} u^1} \geq \delta \right)$$
$$\leq \frac{1}{\delta} \mathbb{E}\left[\frac{h_m}{\|\sigma^{-1}\beta\|} \sqrt{\frac{1}{T}(u^1)^* \sigma^{-1} \epsilon \epsilon^* \sigma^{-1} u^1} \right]$$
$$\leq \frac{1}{\delta} \sqrt{\mathbb{E}[h_m^2] \mathbb{E}\left[\frac{1}{\|\sigma^{-1}\beta\|^2} \frac{1}{T}(u^1)^* \sigma^{-1} \epsilon \epsilon^* \sigma^{-1} u^1 \right]}$$
$$\to 0 \,,$$

which follows from Condition 2.1. ∎

Lemma B.2. *For u^1 given by (37) we have the limit*

$$\mathbb{E}\left| \frac{1}{\|\sigma^{-1}\beta\|^2} (u^1)^* \widehat{\rho} u^1 - h_m^2 \right| \to 0 \,, \tag{40}$$

as $N \to \infty$.

Proof. First notice that $\widehat{\rho} = h^{-1}\sigma\sigma^{-1}\widehat{\Sigma}\sigma^{-1}\sigma h^{-1}$, and then using similar steps as those for proving (38), we have

$$\frac{1}{\|\sigma^{-1}\beta\|^2} (u^1)^* \left(\widehat{\rho} - h_m^2 \|\sigma^{-1}\beta\|^2 I \right) u^1$$
$$= \frac{1}{\|\sigma^{-1}\beta\|^2} ((I - h\sigma^{-1} + h\sigma^{-1})u^1)^* \left(\widehat{\rho} - h_m^2 \|\sigma^{-1}\beta\|^2 I \right) (I - h\sigma^{-1} + h\sigma^{-1}) u^1$$
$$\leq \frac{2}{\|\sigma^{-1}\beta\|^2} ((I - h\sigma^{-1})u^1)^* \widehat{\rho} (I - h\sigma^{-1}) u^1$$

$$
\begin{aligned}
&\qquad + \frac{2}{\|\sigma^{-1}\beta\|^2}((h\sigma^{-1})u^1)^*\left(\widehat{\rho} - h_m^2\|\sigma^{-1}\beta\|^2 I\right)(h\sigma^{-1})u^1 \\
&= \frac{2}{\|\sigma^{-1}\beta\|^2}((I - h\sigma^{-1})u^1)^*\widehat{\rho}(I - h\sigma^{-1})u^1 \\
&\qquad + \frac{2}{\|\sigma^{-1}\beta\|^2}(u^1)^*\left(\sigma^{-1}\widehat{\Sigma}\sigma^{-1} - h_m^2\|\sigma^{-1}\beta\|^2\sigma^{-1}h^2\sigma^{-1}\right)u^1 \\
&\le \frac{2N\|I - h\sigma^{-1}\|^2}{\|\sigma^{-1}\beta\|^2} + \frac{2}{\|\sigma^{-1}\beta\|^2}(u^1)^*\left(\sigma^{-1}\widehat{\Sigma}\sigma^{-1} - h_m^2\|\sigma^{-1}\beta\|^2 I\right)u^1 \\
&\qquad\qquad + 2h_m^2(u^1)^*\left(I - \sigma^{-1}h^2\sigma^{-1}\right)u^1 \\
&= \frac{2N\|I - h\sigma^{-1}\|^2}{\|\sigma^{-1}\beta\|^2} + \frac{2}{\|\sigma^{-1}\beta\|^2}(u^1)^*\left(\sigma^{-1}\widehat{\Sigma}\sigma^{-1} - h_m^2\|\sigma^{-1}\beta\|^2 I\right)u^1 \\
&\qquad\qquad + 2h_m^2(1 - \|\sigma^{-1}h\|^2)\ . \qquad (41)
\end{aligned}
$$

The first and third terms in (41) converge to zero in expectation because of Condition 2.1, and the second term converges to zero in mean using (38). ■

It is also useful to note that for any sequence of random vectors $\widetilde{U}$ such that $\widetilde{U} \perp u^1$ and $\|\widetilde{U}\| = 1$ for all N, using (36) we have

$$
\begin{aligned}
0 &\le \frac{1}{\|\sigma^{-1}\beta\|^2}\mathbb{E}\left[(h\sigma^{-1}\widetilde{U})^*\widehat{\rho}h\sigma^{-1}\widetilde{U}\right] \\
&\le \frac{1}{\|\sigma^{-1}\beta\|^2}\mathbb{E}\left[\sup_{\substack{\tilde{u}\perp u^1\\ \|\tilde{u}\|=1}}(h\sigma^{-1}\tilde{u})^*\widehat{\rho}h\sigma^{-1}\tilde{u}\right] \\
&= \frac{1}{(T-1)\|\sigma^{-1}\beta\|^2}\mathbb{E}\left[\sup_{\substack{\tilde{u}\perp u^1\\ \|\tilde{u}\|=1}}\tilde{u}^*\left[\sigma^{-1}\nu\nu^*\sigma^{-1}\right]\tilde{u}\right] \\
&\le \frac{1}{(T-1)\|\sigma^{-1}\beta\|^2}\mathbb{E}\left[\sup_{\|u\|=1}(\sigma^{-1}u)^*\nu\nu^*(\sigma^{-1}u)\right] \\
&\le \frac{1}{(T-1)\|\sigma^{-1}\beta\|^2}\mathbb{E}\left[\sup_{\|u\|=1}(\sigma^{-1}u)^*\epsilon\epsilon^*(\sigma^{-1}u)\right] \\
&\to 0\ , \qquad (42)
\end{aligned}
$$

as $N \to \infty$, where $\frac{1}{(T-1)\|\sigma^{-1}\beta\|^2}\mathbb{E}[\sup_{\|u\|=1}(\sigma^{-1}u)^*\epsilon\epsilon^*(\sigma^{-1}u)] \to 0$ follows from Condition 2.1. Note that because the quantity is positive, (42) implies that

$$
\frac{1}{\|\sigma^{-1}\beta\|^2}(h\sigma^{-1}\widetilde{U})^*\widehat{\rho}h\sigma^{-1}\widetilde{U} \to 0 \qquad (43)
$$

in probability as $N \to \infty$.

Lemma B.3. *For any sequence of random vectors $\widetilde{U}$ such that $\widetilde{U} \perp u^1$ and $\|\widetilde{U}\| = 1$ for all N, we have*

$$\lim_N \mathbb{E}\left|(U^1)^* h\sigma^{-1}\widetilde{U}\right| = 0 \ , \tag{44}$$

where $U^1 \in \mathbb{R}^N$ is the top eigenvector of $\widehat{\rho}$.

Proof. For U^1 and any sequence of random vectors $\widetilde{U}$ such that $\widetilde{U} \perp u^1$ and $\|\widetilde{U}\| = 1$ for all N, we have

$$\begin{aligned}
0 &\le \frac{\lambda_1}{\|\sigma^{-1}\beta\|^2}\left|(U^1)^* h\sigma^{-1}\widetilde{U}\right| \\
&= \frac{1}{\|\sigma^{-1}\beta\|^2}\left|(U^1)^*\widehat{\rho} h\sigma^{-1}\widetilde{U}\right| \\
&\le \frac{1}{\|\sigma^{-1}\beta\|^2}\sqrt{(U^1)^*\widehat{\rho}U^1}\sqrt{(h\sigma^{-1}\widetilde{U})^*\widehat{\rho}h\sigma^{-1}\widetilde{U}} \\
&= \frac{\sqrt{\lambda_1}}{\|\sigma^{-1}\beta\|^2}\sqrt{(h\sigma^{-1}\widetilde{U})^*\widehat{\rho}h\sigma^{-1}\widetilde{U}} \\
&\le \frac{\sqrt{N}}{\|\sigma^{-1}\beta\|^2}\sqrt{(h\sigma^{-1}\widetilde{U})^*\widehat{\rho}h\sigma^{-1}\widetilde{U}} \\
&\to 0 \ , \qquad (45)
\end{aligned}$$

in probability as $N \to \infty$, where $\lim_N \frac{\sqrt{N}}{\|\sigma^{-1}\beta\|} \in (0,\infty)$ by Condition 2.1 and

$$\sqrt{\frac{1}{\|\sigma^{-1}\beta\|^2}(h\sigma^{-1}\widetilde{U})^*\widehat{\rho}h\sigma^{-1}\widetilde{U}} \to 0 \ ,$$

in probability follows from (43).

Now we proceed by contradiction. Suppose there is a subsequence N_n for which $|(U^1)^* h\sigma^{-1}\widetilde{U}|$ does not converge to zero in probability. There exists a further subsequence N_{n_j} under which the limit of h_m holds almost surely, and where the limits in (40) and (45) hold almost surely. Then, for this subsequence we have

$$\begin{aligned}
0 &= \limsup_{N_{n_j}} \frac{\lambda_1}{\|\sigma^{-1}\beta\|^2}\left|(U^1)^* h\sigma^{-1}\widetilde{U}\right| \\
&\ge \liminf_{N_{n_j}} \left(\frac{\lambda_1}{\|\sigma^{-1}\beta\|^2} - h_m^2\right)\left|(U^1)^* h\sigma^{-1}\widetilde{U}\right| + \limsup_{N_{n_j}} h_m^2\left|(U^1)^* h\sigma^{-1}\widetilde{U}\right| \\
&= \liminf_{N_{n_j}} \left(\frac{(U^1)^*\widehat{\rho}U^1}{\|\sigma^{-1}\beta\|^2} - h_m^2\right)\left|(U^1)^* h\sigma^{-1}\widetilde{U}\right| + \limsup_{N_{n_j}} h_m^2\left|(U^1)^* h\sigma^{-1}\widetilde{U}\right| \\
&\ge \liminf_{N_{n_j}} \left(\frac{(u^1)^*\widehat{\rho}u^1}{\|\sigma^{-1}\beta\|^2} - h_m^2\right)\left|(U^1)^* h\sigma^{-1}\widetilde{U}\right| + \limsup_{N_{n_j}} h_m^2\left|(U^1)^* h\sigma^{-1}\widetilde{U}\right| \\
&= \limsup_{N_{n_j}} h_m^2\left|(U^1)^* h\sigma^{-1}\widetilde{U}\right| \\
&\ge 0 \ . \qquad (46)
\end{aligned}$$

This contradicts our supposed nonconvergence to zero, and thus, $\limsup_{N_{n_j}} |(U^1)^* h\sigma^{-1}\widetilde{U}| = 0$ because $\limsup_{N_{n_j}} h_m^2 > 0$ by Condition 2.1, and thus, $|(U^1)^* h\sigma^{-1}\widetilde{U}| \to 0$ in probability. Meansquare integrability of $h\sigma^{-1}$ is given by Condition 2.1, and hence, convergence in mean follows for $|(U^1)^* h\sigma^{-1}\widetilde{U}|$. ∎

Finally, we are ready to prove the two limits stated in Theorem 2.1. Using Lemma B.3 we have

$$\begin{aligned}
&\lim_N \mathbb{E}\left|(U^1)^*\widetilde{U}\right|^2 \\
&= \lim_N \mathbb{E}\left|(U^1)^*(I - h\sigma^{-1})\widetilde{U} + (U^1)^* h\sigma^{-1}\widetilde{U}\right|^2 \\
&\leq 2\lim_N \mathbb{E}\|I - h\sigma^{-1}\|^2 + 2\lim_N \mathbb{E}\left|(U^1)^* h\sigma^{-1}\widetilde{U}\right|^2 \\
&= 0\ ,
\end{aligned}$$

as $N \to \infty$, where $h\sigma^{-1} \to I$ in meansquare from Condition 2.1. Hence, the top eigenvector converges close to u^1,

$$\mathbb{E}\|U^1 - u^1\|^2 \to 0\ ,$$

which along with (40) shows the top eigenvalue convergence,

$$\begin{aligned}
&\left|\frac{\lambda_1}{\|\sigma^{-1}\beta\|^2} - h_m^2\right| \\
&= \left|\frac{1}{\|\sigma^{-1}\beta\|^2}(U^1)^*\widehat{\rho}U^1 - h_m^2\right| \\
&= \frac{1}{\|\sigma^{-1}\beta\|^2}\left\|(\widehat{\rho} - h_m^2\|\sigma^{-1}\beta\|^2 I)^{1/2}U^1\right\|^2 \\
&\leq \frac{2}{\|\sigma^{-1}\beta\|^2}\left\|(\widehat{\rho} - h_m^2\|\sigma^{-1}\beta\|^2 I)^{1/2}(U^1 - u^1)\right\|^2 + \underbrace{\frac{2}{\|\sigma^{-1}\beta\|^2}\left\|(\widehat{\rho} - h_m^2\|\sigma^{-1}\beta\|^2 I)^{1/2}u^1\right\|^2}_{\longrightarrow 0 \quad \text{by (40)}} \\
&\leq \frac{2\|\widehat{\rho} - h_m^2\|\sigma^{-1}\beta\|^2 I\|}{\|\sigma^{-1}\beta\|^2}\left\|U^1 - u^1\right\|^2 + \frac{2}{\|\sigma^{-1}\beta\|^2}\left\|(\widehat{\rho} - h_m^2\|\sigma^{-1}\beta\|^2 I)^{1/2}u^1\right\|^2 \\
&\leq \frac{2(N + h_m^2\|\sigma^{-1}\beta\|^2)}{\|\sigma^{-1}\beta\|^2}\left\|U^1 - u^1\right\|^2 + \frac{2}{\|\sigma^{-1}\beta\|^2}\left\|(\widehat{\rho} - h_m^2\|\sigma^{-1}\beta\|^2 I)^{1/2}u^1\right\|^2 \\
&= 0
\end{aligned}$$

in probability as $N \to \infty$; this final limit also holds in expectation via a generalized dominated convergence theorem.

Appendix C. Review of Markowitz porfolio theory.

Consider a market with n-many assets with returns r_i such that

$$\mathbb{E}r_i = \mu_i \qquad \text{and} \qquad \mathrm{cov}(r_i, r_j) = \Sigma_{ij}$$

for all $1 \leq i, j \leq N$. Markowitz theory (see Markowitz (1952)) looks to optimize mean-variance preferences,

$$\min_{\pi} \left(\frac{1}{2}\pi^* \Sigma \pi - \frac{1}{\gamma}\pi^* \mu \right)$$
$$\text{s.t.} \quad \pi^* \mathbf{1} = 1 \ ,$$

where $\gamma > 0$ is investor risk aversion and $\mathbf{1}$ is an $\mathbb{R}^N$ vector of all 1's. Solutions to this problem form a hyperbola in the two-dimensional plane, opening to the right, with portfolio standard deviation along the horizontal axis and portfolio expected return along the vertical. The top half of this hyperbola is the efficient frontier. Mean-variance optimization can also be formulated with a risk-free rate and without constraints,

$$\min_{\pi} \left(\frac{1}{2}\pi^* \Sigma \pi - \frac{1}{\gamma}(\pi^* \mu + r_0(1 - \pi^* \mathbf{1})) \right) \ ,$$

where r_0 is the risk-free rate. The solution is

$$\pi = \frac{1}{\gamma}\Sigma^{-1}(\mu - r_0 \mathbf{1}) \ .$$

The variance of one of these optimal portfolios is

$$\sigma_\gamma^2 = \frac{1}{\gamma^2}(\mu - r_0 \mathbf{1})^* \Sigma^{-1}(\mu - r_0 \mathbf{1}),$$

and the expected return is

$$\begin{aligned} m_\gamma &= r_0 + \frac{1}{\gamma}(\mu - r_0 \mathbf{1})^* \Sigma^{-1}(\mu - r_0 \mathbf{1}) \\ &= r_0 + \sqrt{(\mu - r_0 \mathbf{1})^* \Sigma^{-1}(\mu - r_0 \mathbf{1})} \times \sigma_\gamma. \end{aligned}$$

Hence, expected return is a linear function of standard deviation, and forms the so-called capital market line (CML). The portfolio that is both on the CML and on the efficient frontier (i.e., the portfolio at the point where CML is tangent to the frontier) is called the tangency portfolio. The 1-fund theorem points out that any optimal portfolio will be some linear combination of the risk-free asset and the tangency portfolio. Equilibrium theory suggests that the tangency portfolio is the capitalization-weighted market portfolio (see Lintner (1965), Sharpe (1964)).

The capital asset pricing model (CAPM) is derived from the Markowitz theory in the following way. Let $X_m = \sum_i \pi_m^i (r_i - r_0) = \pi_m^*(r - r_0 \mathbf{1})$ denote the excess return on the tangency portfolio, and let $X = \pi^*(r - r_0 \mathbf{1})$ denote excess returns on some other (possibly suboptimal) portfolio. The variance of X_m is

$$\sigma_m^2 = \frac{1}{\gamma_m^2}(\mu - r_0 \mathbf{1})^* \Sigma^{-1}(\mu - r_0 \mathbf{1}) = \frac{1}{\gamma_m}\mathbb{E}X_m \ ,$$

where γ_m is the risk aversion of the tangency investor. The covariance between X and X_m is

$$\text{cov}(X, X_m) = \frac{1}{\gamma_m}\pi^*(\mu - r_0 \mathbf{1}) = \frac{\sigma_m^2}{\mathbb{E}X_m}\mathbb{E}X \ .$$

Rearranging this expression yields

$$\mathbb{E}X = \frac{\text{cov}(X, X_m)}{\sigma_m^2}\mathbb{E}X_m \ .$$

This suggests the CAPM model

$$X = \beta X_m + \epsilon \ ,$$

where ϵ is a mean-zero noise uncorrelated with X_m, and $\beta = \frac{\text{cov}(X,X_m)}{\sigma_m^2}$. For the individual assets the CAPM model is the expression in (19).

REFERENCES

T. W. Anderson. *Asymptotic theory for principal component analysis*, Ann. Math. Statist., 34 (1963), pp. 122–148.

M. Avellaneda, *Hierarchical PCA and applications to portfolio management*, Rev. Mex. Economía Finanzas, 15 (2020), pp. 1–16.

M. Avellaneda and J.-H. Lee, *Statistical arbitrage in the US equities market*, Quant. Finance, 10 (2010), pp. 761–782.

M. Avellaneda, B. Healy, A. Papanicolaou, and G. Papanicolaou, *Pca for implied volatility surfaces*, J. Financial Data Sci., 2 (2020), pp. 85–109.

J. Bai, *Inferential theory for factor models of large dimensions*, Econometrica, 71 (2003), pp. 135–171.

J. Bai and S. Ng, *Large Dimensional Factor Analysis*, Now Publishers, 2008.

J. Bai and S. Ng, *Rank regularized estimation of approximate factor models*, J. Econometrics, 212 (2019), pp. 78–96, https://doi.org/10.1016/j.jeconom.2019.04.021.

Z-D. Bai and J. W. Silverstein, *No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices*, Ann. Probab., 26 (1998), pp. 316–345.

Z.-D. Bai and J. W. Silverstein, *Exact separation of eigenvalues of large dimensional sample covariance matrices*, Ann. Probab., 1999, pp. 1536–1555.

J. Baik and J. W. Silverstein, *Eigenvalues of large sample covariance matrices of spiked population models*, J. Multivariate Anal., 97 (2006), pp. 1382–1408.

J. Baik, G. Ben Arous, and S. Péché, *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*, Ann. Probab., 33 (2005), pp. 1643–1697, https://doi.org/10.1214/009117905000000233.

F. Benaych-Georges and R. R. Nadakuditi, *The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices*, Adv. Math., 227 (2011), pp. 494–521.

P. Boyle, *Positive weights on the efficient frontier*, North American Actuarial J., 18 (2014), pp. 462–477.

T. J. Brennan and A. W. Lo, *Impossible frontiers*, Management Sci., 56 (2010), pp. 905–923.

J. Bun, J. P. Bouchaud, and M. Potters, *Cleaning large correlation matrices: Tools from random matrix theory*, Phys. Rep., 666 (2017), pp. 1–109, https://doi.org/10.1016/j.physrep.2016.10.005.

J. Chen and M. Yuan, *Efficient portfolio selection in a large market*, J. Financial Econometrics, 14 (2016), pp. 496–524.

A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. Phan, *Tensor decompositions for signal processing applications: From two-way to multiway component analysis*, IEEE Signal Processing Magazine, 32 (2015), pp. 145–163.

T. Conlon, H. Ruskin, and M. Crane, *Cross-correlation dynamics in financial time series*, Phys. A, 388 (2010), pp. 705–714, https://doi.org/10.1016/j.physa.2008.10.047.

L. De Lathauwer, B. De Moor, and J. Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.

I. M. Johnstone, *On the distribution of the largest eigenvalue in principal components analysis*, Ann. Statist., 29 (2001), pp. 295–327, https://doi.org/10.1214/aos/1009210544.

Z. Kakushadze, *Heterotic risk models*, Wilmott, 2015 (80), pp. 40–55.

T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500.

T. Kollo and H. Neudecker, *Asymptotics of eigenvalues and unit-length eigenvectors of sample variance and correlation matrices*, J. Multivariate Anal., 47 (1993), pp. 283–300.

S. Konishi, *Asymptotic expansions for the distributions of statistics based on a correlation matrix*, Canad. J. Statist., 6 (1978), pp. 49–56.

O. Ledoit and S. Péché, *Eigenvectors of some large sample covariance matrix ensembles*, Probab. Theory Related Fields, 151 (2011), pp. 233–264.

O. Ledoit and M. Wolf, *A well conditioned estimator for large dimensional covariance matrices*, J. Multivariate Anal., 88 (2004), pp. 365–411.

O. Ledoit and M. Wolf, *Nonlinear shrinkage estimation of large-dimensional covariance matrices*, Ann. Statist., 40 (2012), pp. 1024–1060.

J. Lintner, *The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets*, Rev. Economics Statistics, 47 (1965), pp. 13–37.

H. Markowitz, *Portfolio selection*, J. Finance, 7 (1952), pp. 77–91.

V. A. Marchenko and L. Pastur, *Distribution of eigenvalues for some sets of random matrices*, Math USSR Sb., 1 (1967), pp. 457–483.

A. Montanari and E. Richard, *A statistical model for tensor PCA*, In Proceedings of the 27th International Conference on Neural Information Processing Systems, Vol. 2, MIT Press, Cambridge, MA, 2014, pp. 2897–2905.

D. Morales-Jimenez, I. M. Johnstone, R. M. McKay, and J. Yang, *Asymptotics of eigenstructure of sample correlation matrices for high-dimensional spiked models*, Statist. Sinica, 31 (2021), pp. 571–601.

A. Onatski, *Determining the number of factors from empirical distribution of eigenvalues*, Rev. Economics Statistics, 92 (2010), pp. 1004–1016.

A. Onatski, *Asymptotics of the principal components estimator of large factor models with weakly influential factors*, J. Econometrics, 168 (2012), pp. 244–258. https://doi.org/10.1016/j.jeconom.2012.01.034.

D. Paul, *Asymptotics of sample eigenstructure for a large dimensional spiked covariance model*, Statist. Sinica, 17 (2007), pp. 1617–1642.

V. Plerou, P. Gopikrishnan, B. Rosenow, L. Amaral, and E. H. Stanley, *Universal and nonuniversal properties of cross correlations in financial time series*, Phys. Rev. Lett., 83 (1999), 1471.

V. Plerou, P. Gopikrishnan, B. Rosenow, L. Amaral, T. Guhr, and E. H. Stanley, *A random matrix approach to cross-correlations in financial data*, Phys. Rev. E, 65 (2002), 066126, https://doi.org/10.1103/PhysRevE.65.066126.

M. Potters, J.-P. Bouchaud, and L. Laloux, *Financial applications of random matrix theory: Old laces and new pieces*, Acta Phys. Polon. B, 36 (2005).

R. Roll, *Orthogonal portfolios*, J. Financial Quantitative Anal., 15 (1980), pp. 1005–1023.

J. Ross, *The arbitrage theory of capital asset pricing*, J. Economic Theory, 13 (1976), pp. 341–360.

W. Sharpe, *Capital asset prices: A theory of market equilibrium under conditions of risk*, J. Finance, 19 (1964), pp. 425–442.

D. Shen, H. Shen, and J. S. Marron, *A general framework for consistency of principal component analysis*, J. Mach. Learn. Res., 17 (2016), pp. 150:1–150:34.

D. Shi, *Asymptotic Joint Distribution of Extreme Sample Eigenvalues and Eigenvectors in the Spiked Population Model*, preprint, https://arxiv.org/abs/1304.6113, 2013.

A. Steele, *On the eigen structure of the mean variance efficient set*, J. Business Finance Accounting, 22 (1995), pp. 245–255.

L. Su and X. Wang, *On time-varying factor models: Estimation and testing*, J. Econometrics, 198 (2017), pp. 84–101, https://doi.org/10.1016/j.jeconom.2016.12.004.

L. R. Tucker, *The extension of factor analysis to three-dimensional matrices*, in Contributions to Mathematical Psychology, Holt, Rinehart and Winston, 1964, pp. 110–182.

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

W. Wang and J. Fan, *Asymptotics of empirical eigenstructure for high dimensional spiked covariance*, Ann. Statist., 45 (2017), pp. 1342–1374, https://doi.org/10.1214/16-AOS1487.

Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, *On the limit of the largest eigenvalue of the large dimensional sample covariance matrix*, Probab. Theory Related Fields, 78 (1988), pp. 509–521, https://doi.org/10.1007/BF00353874.

G. Zumbach, *The empirical properties of large covariance matrices*, Quant. Finance, 11 (2011), pp. 1091–1102.