Identifying Broad and Narrow Financial Risk Factors with Convex Optimization

Alexander D. Shkolnik * Lisa R. Goldberg † and Jeffrey R. Bohn ‡

August 20, 2016§

Abstract

Factor analysis of security returns aims to decompose a return covariance matrix into systematic and specific risk components. To date, most commercially successful factor analysis has been based on fundamental models, although there is a large academic literature on statistical models. While successful in many respects, traditional statistical approaches like principal component analysis and maximum likelihood suffer from several drawbacks. These include a lack of robustness, strict assumptions on the underlying model of returns, and insensitivity to narrow factors such as industries and currencies, which affect only a small number of securities, but in an important way.

We apply convex optimization methods to decompose a security return covariance matrix into its low rank and sparse parts. The low rank component includes the market and other broad factors that affect most securities. The sparse component includes narrow factors and security specific effects.

We measure the variance forecasting accuracy of a low rank plus sparse covariance matrix estimator on an equally weighted portfolio of 125 securities simulated from a model with two broad factors and 25 narrow factors. We find that the low rank plus sparse estimators are more accurate than estimates made with classical principal component analysis, in particular, at forecasting risk due to narrow factors. Finally, we illustrate a low rank plus sparse decomposition of an empirical covariance matrix of 125 equities drawn from 25 countries.

*Berkeley Associates LLC and University of California, Berkeley. ads2@berkeley.edu

†Berkeley Associates LLC and University of California, Berkeley, lrg@berkeley.edu

‡State Street and University of California, Berkeley, jbohn@statestreet.com

§We thank the conference participants at the 2016 UC Berkeley-Columbia Meeting in Engineering and Statistics for their comments.

1 Introduction

It is a standard assumption in financial economics that security returns are driven by a relatively small number of risk factors, plus security specific returns. This assumption is at the heart of asset pricing models and portfolio risk analysis. Quantitative methods for identifying financial risk factors fall into two categories. In a structural or *fundamental* model, factors are identified by humans, while factor attributes such as volatility and correlation are estimated with statistical methods. In a data-driven or *latent factor* model, statistical methods are used both to identify factors and to estimate their attributes.

The strengths and weaknesses of fundamental and latent factor models are complementary. Fundamental models are intuitive, and practitioners use them for risk forecasting, risk reporting, exposure analysis and portfolio construction. On the negative side, the inclusion of new factors typically requires a structural change, so fundamental models are prone to mistaking an emerging factor for noise. Latent factors are not readily interpretable, and that may be the primary reason why they have been less popular with practitioners. Yet, there is a vast academic literature on latent factor models, which rely on the rapidly developing field of data science. Latent factor models respond dynamically to new information. This feature gives latent factor models the potential to detect emerging factors, but it predisposes them to mistake chance correlation for signal.

Many latent factor models are estimated with principal component analysis (PCA), which extracts factors from a sample covariance matrix of security returns. PCA relies on spectral analysis (Jolliffe 2002), which identifies factors as sample covariance matrix eigenvectors with the largest eigenvalues. In financial return data, these eigenvectors tend to represent broad risk drivers (such as sensitivity to the market or to interest rates), each of which affects most securities. However, they may omit narrow risk drivers (such as industry or country membership), each of which affects only a few securities, but has a pronounced impact on security return correlation nevertheless. To be detected by classical PCA, a factor must affect a significant fraction of the securities (Miller 2006). As PCA requires no distributional assumptions, it is appropriate only when the broad factors are assumed to be variance maximizing (latent) variables. A negative consequence of this fact is that the analysis is highly sensitive to outliers.

In this article, we develop a factor identification scheme that relies on the low-rank plus sparse (SLR) matrix decompositions of covariance matrices, which were proposed as a *robust* alternative to PCA (see Chandrasekaran, Sanghavi, Parrilo & Willsky (2011), Candès, Li, Ma & Wright (2011), and Agarwal, Negahban, Wainwright et al. (2012)). To date, these methods have been successfully applied to image processing, latent semantic indexing, collaborative filtering, graphical model learning, gene expression, and

other problems. Some experimentation with SLR decompositions of financial return covariance matrices has been attempted by academics in Chandrasekaran, Parrilo & Willsky (2012) and discussant papers. More, recently SLR formulations have been investigated in the context of portfolio risk in Luo (2013) and Fan, Han & Liu (2014).

SLR recovery methods differ from PCA in that they do not rely directly on spectral methods to extract broad factors from the sample covariance of returns. Instead they have much more in common with factor analysis (FA) Jolliffe (2002, see Chapter 7). A model of security returns for which the covariance matrix (or the concentration matrix) is a sum of a low rank matrix and a sparse matrix is hypothesized. The low rank component corresponds to broad factors and the sparse component corresponds to the narrow factors and security specific returns. Surprisingly, under quite mild assumptions such a decomposition is unique. The aim of SLR recovery is to formulate a (typically convex) programming problem that may be solved efficiently, and that yields the true decomposition. The methods turn out to be robust to outliers. They can incorporate various statistical modeling assumptions, but may remain distribution-free as well.

We apply a method for SLR recovery proposed in Chandrasekaran, Parrilo & Willsky (2010) and analyzed in Chandrasekaran et al. (2012). While their distributional assumptions are not met by financial return data, the results are encouraging. We develop metrics to assess risk forecasting accuracy similar to those of Bender, Lee, Stefek & Yao (2009). We benchmark the SLR method against classical PCA on simulated data. We also apply SLR to empirical data to illustrate the promise of the method for practical applications.

The rest of the article is organized as follows. In Section 2 we formulate our model of security returns and justify the SLR structure of the security covariance matrix. In Section 3, we relate SLR decompositions of a covariance matrix and its associated concentration matrix. We state the convex program for SLR recovery in Section 4. An outline of an application SLR recovery to portfolio risk forecasting is in Section 5. This section also reviews measures of the accuracy of SLR risk forecasts on simulated financial return data. Section 6 presents results of simulation experiments and Section 7 illustrates an SLR recovery from a sample covariance matrix estimated from empirical data. Section 8 concludes. Supporting ideas and technical details are described in our appendices. Appendix A describes the graphical structure associated with a SLR decomposition of the concentration matrix of Gaussian variables. The implementation of our SLR algorithm is described in Appendix B.

2 Factor model and problem formulation

While the relationship between security returns and factor returns may be non-linear in general, we explore a linear model, which has proven to be a useful specification for public equities and sovereign bonds. Specifically, we assume the generating process[1] for the N-vector of security returns R is given by

$$R = Z\zeta + \epsilon, \tag{1}$$

where Z is an $N \times M$ matrix of factor exposures, ζ is an M-vector of factor returns and ϵ is an N-vector of diversifiable specific returns. While the returns (R, ζ, ϵ) are random, we treat the exposures is Z as constants to be estimated. The factor returns and specific returns are taken to be mean zero and uncorrelated, i.e. $\mathbf{E}(\zeta) = 0_K$, $\mathbf{E}(\epsilon) = 0_N$ and $\mathbf{E}(\zeta\epsilon^\top) = 0_{M\times N}$. Under these assumptions, the security covariance matrix $\Sigma = \mathbf{E}(RR^\top)$ can be expressed as

$$\Sigma = ZHZ^\top + \Delta \tag{2}$$

where H is an $M \times M$ factor covariance matrix and $\Delta = \mathbf{E}(\epsilon\epsilon^\top)$ is a diagonal $N \times N$ covariance matrix of specific returns.

In anticipation of the latent factor analysis that we will explore, it is useful to refine (1) to take account of the division between broad and narrow factors. We assume that returns may be attributed to K broad factors and κ narrow factors. A broad factor affects most of the securities while a narrow factor may influence only a "small" subset.

$$R = Y\psi + X\phi + \epsilon, \tag{3}$$

where $\zeta = (\psi, \phi)^\top$. In (3), Y is an $N \times K$ matrix of broad factor exposures, ψ is a K-vector of broad factor returns, X is an $N \times \kappa$ matrix of narrow factor exposures, ϕ is a κ-vector of narrow factor returns, and $M = K + \kappa$. We assume $\mathbf{E}(\psi\phi^\top) = 0_{K\times\kappa}$. In light of the above assumptions on the factors, the matrix Y is considered dense (few zero entries) while X is a sparse matrix (a relatively large number of zero entries).

We define the rank K, broad factor approximation to the covariance matrix Σ in (2) as

$$L = YFY^\top, \tag{4}$$

where F is the $K \times K$ covariance matrix of the broad factor returns ψ. We assume $K \ll N$, i.e. a relatively small number of broad factors drive the security returns.

[1] To ease notation we omit the dependence on time in equation (1). That is, (R_t, Y_t, ϵ_t) would denote the returns at some time $t \geq 0$ but Y is typically treated as constant. Analogous comments apply to (3).

We do not impose such an assumption on the number of narrow factors. The residual covariance S is the sum

$$S = XGX^\top + \Delta, \tag{5}$$

where G is the $\kappa \times \kappa$ covariance matrix of the narrow factor returns ϕ. Sparsity of the exposures X to the narrow factors often induces sparsity in S as we shall see in Section 6. Thus, the covariance matrix Σ is expressed as a sum of a low-rank matrix and a residual matrix,

$$\Sigma = L + S. \tag{6}$$

In what follows, we adopt the standard assumption that security returns follow the return generating process (3), and we explore statistical methods for estimating the low rank and residual components of the decomposition (6). These methods distinguish broad factor return from the sum of narrow factor return and specific return, putting us closer to a fully automated identification of the factors that drive financial risk.

3 Sparse and low rank decompositions of a covariance matrix or its inverse

The problem of decomposing a covariance matrix into low rank and residual components is complicated by the question of identifiability. Under mild technical conditions, the decomposition (6) is unique when the residual matrix S is sparse and the low rank matrix L is not sparse. The low rank condition on L is natural in view of the assumption that the number of broad factors K is small relative to N. The sparsity assumption on S hinges on the structure of the narrow factors.

In many statistical applications, however, sparse representations of sample data are imposed on the concentration matrix, Σ^{-1} (Dempster 1972, and others). For example, this is particularly appropriate in the context of Gaussian data where zeros in the concentration matrix dictate conditional independence relationships known as graphical models (Anderson 2003, Chapter 15). While imposing sparsity on Σ^{-1} is not appropriate in the broad factor setting of Section 2, one could aim to decompose this concentration matrix as

$$\Sigma^{-1} = \mathcal{S} - \mathcal{L}. \tag{7}$$

where $\mathcal{L}$ is low rank. There is a simple connection between (7) and (6).

LEMMA 3.1. *Let* Σ *and* S *in (6) be invertible. Then* (7) *holds with* $\mathcal{S} = S^{-1}$ *and* $\mathcal{L} = S^{-1}L\Sigma^{-1}$. *Conversely, given* (7)*, we have* $S = \mathcal{S}^{-1}$ *and* $L = \mathcal{S}^{-1}\mathcal{L}\Sigma$ *in* (6). *Moreover, rank*$(\mathcal{L})$ = *rank*(L).

Proof. Multiplying Σ from the left by $S^{-1} - S^{-1}L\Sigma^{-1}$ and substituting $L = \Sigma - S$ verifies $\mathcal{S} - \mathcal{L}$ is the left inverse of Σ (hence, the inverse). As S and Σ^{-1} are full rank, the second claim is a basic property of rank (Horn & Johnson 2012, 0.4.6 (b)). The reverse direction follows by similar arguments. □

Lemma 3.1 allows us to pass between decompositions (6) and (7) and says that when L is low rank so is $\mathcal{L}$ and vice versa. However, it does not provide information on which decomposition yields sparsity in the remaining matrices S and $\mathcal{S}$. Note, that the inverse of a sparse matrix is not in general sparse.[2] Sparsity in S implies most of the security returns to the narrow factors (i.e., $X\phi$) are uncorrelated. The matrix $\mathcal{S}$, on the other hand, is the concentration matrix of the narrow factor and specific risk covariance. Sparsity in this context may have several interpretations. For instance, when (R, ψ) are jointly Gaussian, sparsity of $\mathcal{S}$ implies the security returns to the narrow factor depend on each other not directly, but through the other securities (see Appendix A).

4 Estimating the decomposition

Low rank plus sparse decompositions may be identified via solutions to certain convex optimization problems (see Candès et al. (2011) and Chandrasekaran et al. (2011)). A standard formulation is known as *principal component pursuit* (PCP), which minimizes the trace of one variable subject to a sparsity constraint on another. When this approach is adapted to decomposing the covariance matrix Σ of security returns as in (6), PCP minimizes the convex objective function

$$\mathcal{P}(L, S \mid \gamma) = \|L\|_* + \gamma\|S\|_1 \tag{8}$$

subject to the constraint $\Sigma = L + S$. Here, $\|\cdot\|_*$ denotes the nuclear norm and $\|\cdot\|_1$ denotes the vector ℓ_1-norm. The nuclear norm computes the sum of singular values of L which when symmetric is equal to the sum of its eigenvalues. The $\|\cdot\|_1$ norm regards S as a long N^2-vector and encourages sparsity in this variable.

[2]Howerever, as one example, a block matrix S yields a block matrix $\mathcal{S}$ with at least as many off-block zero entries.

A difficulty with the application of PCP to return covariance matrix decomposition is that in practical applications, Σ is not known.[3] Instead, we are given a sample covariance matrix $\hat{\Sigma}$ estimated from observed returns. Hence, the decomposition we look for may be phrased as

$$\hat{\Sigma} = L + S + W, \tag{9}$$

where W is a noise term. A method of dealing with such estimation errors was proposed by Chandrasekaran et al. (2012). The authors replace the constraint (9) by penalizing[4] (8) with a Gaussian likelihood function ℓ to solve the optimization problem,

$$\begin{aligned} \min_{(S,\mathcal{L})} \quad & \lambda \mathscr{P}(\mathcal{L}, S \,|\, \gamma) - \ell(S - \mathcal{L} \,|\, \hat{\Sigma}) \\ \text{subject to} \quad & S - \mathcal{L} \succ 0, \mathcal{L} \succeq 0. \end{aligned} \tag{10}$$

The constraints $(\succeq)\ \succ$ impose symmetry and positive (semi)definiteness. The constants $\gamma, \lambda > 0$ are parameters that control the rank and sparsity of the solution $(S, \mathcal{L})$. The Gaussian likelihood function ℓ is a proxy for the constraint (9) and is defined as

$$\ell(Q \,|\, \hat{\Sigma}) = \log\det Q - \langle Q^{\top}, \hat{\Sigma} \rangle \tag{11}$$

where the inner product $\langle X_1, X_2 \rangle = \operatorname{trace}(X_1^{\top} X_2)$. The maximizer of (11) (the ML estimate) is the sample concentration matrix $\hat{\Sigma}^{-1}$ and consequently the Gaussian penalty pushes the objective function (10) into the neighborhood of this matrix.

It is important to observe that solving the aforementioned PCP problem and program (10) decompose two entirely different matrices. PCP decomposes the covariance matrix Σ (with an implied assumption that this matrix is known), while (10) decomposes the concentration matrix Σ^{-1} by relying only on knowledge of the sample covariance $\hat{\Sigma}$. The conversion between these two decompositions in (6) and (7) is addressed by Lemma 3.1. Neither decomposition is perfect for our aims since PCP relies on knowledge of Σ and (10) assumes the observations are i.i.d. Gaussian. In this article, we pursue the latter with the algorithm described in Appendix B.

5 Measuring the accuracy of SLR decompositions

We evaluate SLR decompositions in terms of the accuracy of their portfolio return variance forecasts.

[3] There are other difficulties. For instance while it is easy to show that PCP preserves symmetry, it does not ensure (semi)positive definiteness of the two variables.

[4] In the literature it is more common to treat the term $\mathscr{P}$ as the penalty, and that provides an alternative interpretation for this convex program.

5.1 Forecasting Portfolio Return Variance

We identify a portfolio with its vector of weights w. If security returns follow factor model (3), then the security return covariance matrix Σ admits a low rank plus sparse decomposition (6) and the variance of returns to portfolio w can be decomposed into low rank and sparse components,

$$\mathbf{Var}_{\Sigma}(w) = w^{\top}\Sigma w \tag{12}$$

We identify an instance of convex program (10) with a parameter $\theta = (\lambda, \gamma)$. The output of convex program θ is the pair of matrices $\mathcal{L}(\theta)$ and $\mathcal{S}(\theta)$. By Lemma 3.1, we form the estimators $S(\theta) = \mathcal{S}(\theta)^{-1}$ and $L(\theta) = \mathcal{S}(\theta)^{-1}\mathcal{L}(\theta)(\mathcal{S}(\theta) - \mathcal{L}(\theta))^{-1}$. Then,

$$\Sigma(\theta) = L(\theta) + S(\theta) \tag{13}$$

is an estimator of Σ. Thus, each convex program (10) can be used to forecast return variance for portfolio w as

$$\mathbf{Var}_{\Sigma(\theta)}(w) = \mathbf{Var}_{L(\theta)}(w) + \mathbf{Var}_{S(\theta)}(w). \tag{14}$$

5.2 Measuring Forecast Accuracy

We assess the accuracy of variance forecasts (14) with performance measures developed in Bianchi, Goldberg & Rosenberg (2016). The portfolio variance forecasting ratio $\mathscr{R}_{\Sigma}$ for portfolio w is the quotient of the θ-forecast of variance by true variance,

$$\mathscr{R}_{\Sigma}(w \,|\, \theta) = \frac{\mathbf{Var}_{\Sigma(\theta)}(w)}{\mathbf{Var}_{\Sigma}(w)}. \tag{15}$$

Since our covariance matrix estimates are based on the decomposition (6), we measure accuracy on components L and S separately. The broad variance forecasting ratio, $\mathscr{R}_{L}$, is the quotient of the θ-forecast of variance due to broad factors by true variance,

$$\mathscr{R}_{L}(w \,|\, \theta) = \frac{\mathbf{Var}_{L(\theta)}(w)}{\mathbf{Var}_{L}(w)}. \tag{16}$$

Similarly, the sparse variance forecast ratio, $\mathscr{R}_{S}$, is the quotient of the θ-forecast of variance due to narrow factors and specific return by true variance,

$$\mathscr{R}_{S}(w \,|\, \theta) = \frac{\mathbf{Var}_{S(\theta)}(w)}{\mathbf{Var}_{S}(w)}. \tag{17}$$

For all of these ratios, a value greater than 1 means that estimate θ overforecast variance for w and a value less than 1 means that estimate θ underforecast variance for w.

An additional performance measure is the recovered rank $\mathscr{K}$, the rank of $L(\theta)$, or equivalently, the estimated number of broad factors. As we illustrate below, however, the ranks of L and $L(\theta)$ need not be close necessary in order for the measures $\mathscr{R}_\Sigma$, $\mathscr{R}_L$ and $\mathscr{R}_S$ to be close to 1.

6 Simulation experiments

We use the performance measures developed in Section 5.2 to evaluate the accuracy of SLR variance forecasts for equally weighted portfolios of simulated securities.

We assume that security returns follow generating process (3), so that the time t observation for security return R_t is given by

$$R_t = Y\psi_t + X\phi_t + \epsilon_t, \tag{18}$$

where $(\psi_t, \phi_t, \epsilon_t)$ are the time t factor returns and specific returns. We assume returns are time independent. The broad and narrow factor exposures are constants. It follows that the security covariance matrix Σ admits a low rank plus sparse decomposition (6), and we investigate how accurately convex program (10) can recover the decomposition from a sample covariance matrix $\hat{\Sigma}$ estimated from simulated data.

To make the experiment precise, we need to calibrate the return generating process (18), and to specify the number of securities N and the number of observations T. Our choices take account of empirically observed properties of financial return data. We also need to specify the number of simulation paths ρ, and to identify the parameters $\theta = (\lambda, \gamma)$ of convex program (10). Since we are still developing criteria and algorithms for the selection of these parameters, we examine several candidates.

6.1 Calibrating the Return Generating Process

Return generating process (18) is fully specified by

- the number of broad factors, K, and the number of narrow factors, κ,
- the $K \times N$ matrix Y of exposures of securities to broad factors,
- the $\kappa \times N$ matrix X of exposures of securities to narrow factors,

- distributions of the K-vector of broad factor returns, ψ, the κ-vector of narrow factor returns, ϕ, and the N-vector of specific returns, ϵ. The returns ψ, ϕ and ϵ are independent.

We assume broad factor returns, narrow factor returns and specific returns are jointly Gaussian, which is consistent with convex program (10).

- Factor 1 is market-like, meaning that most factors have positive exposure and the factor has an annualized volatility of 16%.
- Factor 2 is long/short and value-like, with an annualized volatility of 4%.

In other words, the factor covariance matrix F in (4) is given by

$$F = \frac{1}{250} \begin{pmatrix} 0.16 & 0 \\ 0 & 0.04 \end{pmatrix}^2. \tag{19}$$

The average exposure of security to the market-like factor is 1, while the value exposures average to zero. So we calibrate our model by drawing the rows of the broad factor exposure matrix Y from a normal distribution with mean $(1, 0)^\mathrm{T}$ and covariance matrix

$$\begin{pmatrix} 0.25 & 0 \\ 0 & 0.75 \end{pmatrix}.$$

Factor exposures (unlike factor returns) are not random in our model. Here, we are using the normal distribution as a convenient means of constructing exposures of securities to factors.

Narrow factors are indicators of sector or country. In our simulations, the number of narrow factors κ depends on the number of securities N in the estimation universe, $\kappa = \lfloor N / \log N \rfloor$. Each of the N securities has exposure of 1 to a single narrow factor and an exposure of 0 to all other narrow factors. Each narrow factor supports five non-zero exposures. Narrow factors have annualized volatilities drawn uniformly from $(10\%, 25\%)$. The specific covariance matrix Δ is diagonal, and annualized specific volatilities are drawn uniformly from $(20\%, 50\%)$.

6.2 Experimental design

For a universe of $N = 125$ securities, we use return generating process (18) to simulate a year's worth of daily returns, $T = 250$ observations. We use this simulated data

to generate a sample covariance matrix $\hat{\Sigma}$, which we submit to an instance of convex program (10) specified by parameters $\theta = (\lambda, \gamma)$. For each $\hat{\Sigma}$, we try several different parameter specifications. The parameter γ is set to $1/\sqrt{N}$ as recommended in Candès et al. (2011), and the values of λ are chosen by trial and error. We use recovered covariance matrices $\Sigma(\theta) = L(\theta) + S(\theta)$ to forecast variance for the equally weighted portfolio.

For each $\theta = (\lambda, \gamma)$, we run $\rho = 400$ simulation paths,[5] and we report the average and standard deviation of the performance accuracy measures $\mathscr{R}_\Sigma$, $\mathscr{R}_L$, $\mathscr{R}_S$ and $\mathscr{K}$.

We benchmark our results in two ways. First, we run a noiseless base case where convex program (10) is given the true covariance matrix Σ. In effect, we have set the number of observations T to ∞. By construction, Σ has an exact SLR decomposition, and we are asking how effectively convex program (10) can recover it.[6]

Second, we compare SLR forecasts to those generated by a principal component analysis estimator, PCA. The dominant eigenvectors of the sample covariance matrix are the broad factors that generate the low rank component. We endow PCA with the correct number of broad factors. There are no narrow factors, and the sparse matrix is a diagonal populated by variances of the residuals of security returns to broad factors. Stretching notation, we allow the symbol θ to refer to PCA.

6.3 Results

Accuracy statistics for the noiseless base case are reported in Table 1. Columns 1–3 report accuracy results for SLR variance forecasts with three different values of the calibration parameter λ, which controls fidelity to the SLR decomposition of the concentration matrix (7). There are four rows in Table 1. The first three rows report the accuracy measures $\mathscr{R}_\Sigma$, $\mathscr{R}_L$ and $\mathscr{R}_S$. The fourth row reports the recovered rank $\mathscr{K}$, which is the estimated number of broad factors. In all cases, $\mathscr{K} = 2$. The ratios $\mathscr{R}_\Sigma$ and $\mathscr{R}_L$ indicate near perfect forecasting of overall variance and variance due to broad factors. The sparse variance forecasting ratio, $\mathscr{R}_S$, indicates underforecasting. The results tend to worsen as λ increases. Table 1 column 4 reports the performance of PCA. On both the low rank and sparse components of variance, PCA is less accurate than any of the SLR models, ostensibly because PCA neglects narrow factors. This is true even

[5] With $\rho = 400$ simulation paths, 97% Monte-Carlo confidence intervals on the reported forecast ratios implied by the Central Limit Theorem are obtained by dividing the standard deviation by 10.

[6] This benchmarking exercise conflicts with the theoretical basis of (10), which expects a noisy input. The results are interesting nevertheless.

though we endow PCA with knowledge of the true number of broad factors.[7] Column 5 of Table 1 reports the true annualized volatility of the equally weighted portfolio. Columns 6 and 7 report the true numbers of broad and narrow factors.

N	Metric	SLR: ($\lambda \times 10^4$)			PCA	Ground Truth		
		1.0	2.0	4.0		Ann. Vol.	K	κ
125	$\mathscr{R}_\Sigma$	1.00	1.00	0.99	1.04	0.17	2	25
	(stdev)	(0.00)	(0.00)	(0.00)	(0.00)			
	$\mathscr{R}_L$	1.00	1.00	1.00	1.08	0.16		
	(stdev)	(0.00)	(0.00)	(0.00)	(0.00)			
	$\mathscr{R}_S$	0.98	0.96	0.92	0.53	0.05		
	(stdev)	(0.00)	(0.00)	(0.00)	(0.00)			
	$\mathscr{K}$	2.00	2.00	2.00	2.00			
	(stdev)	(0.00)	(0.00)	(0.00)	(0.00)			

Table 1: Accuracy of SLR and PCA decompositions of the true covariance matrix. We report portfolio variance forecasting ratios for an equally weighted portfolio as well as the number of broad factors recovered. Since there is no sample noise, we run a single path for each decomposition. The last three columns report the true annualized volatility of the equally weighted portfolio along with the true number of broad and narrow factors that generate security returns.

Results of simulation experiments are in Table 2, whose layout is identical to that of Table 1. For each $\theta = (\lambda, \gamma)$ and PCA, we report the average and standard deviation of the performance measures over $\rho = 400$ simulation paths. To improve performance, values of λ are greater in Table 2 than in Table 1.

[7] In experiments not reported in this note, we found the performance of PCA degraded when we extracted different numbers of broad factors.

N	Metric	SLR: ($\lambda \times 10^4$)			PCA	Ground Truth		
		6.0	8.0	10.0		Ann. Vol.	K	κ
125	$\mathscr{R}_\Sigma$	1.00	0.99	0.98	1.06	0.17		
	(stdev)	(0.09)	(0.09)	(0.09)	(0.09)			
	$\mathscr{R}_L$	1.01	1.00	0.99	1.11	0.16		
	(stdev)	(0.10)	(0.10)	(0.10)	(0.10)			
	$\mathscr{R}_S$	0.82	0.81	0.80	0.52	0.05		
	(stdev)	(0.03)	(0.02)	(0.02)	(0.00)			
	$\mathscr{K}$	12.25	8.48	5.59	2.00		2	25
	(stdev)	(0.64)	(0.64)	(0.71)	(0.00)			

Table 2: Accuracy of SLR and PCA decompositions of sample covariance matrices. We report portfolio variance forecasting ratios for an equally weighted portfolio as well as the number of broad factors recovered. Averages and standard deviations are estimated from $\rho = 400$ simulated paths. The last three columns report the true annualized volatility of the equally weighted portfolio along with the true number of broad and narrow factors that generate security returns.

While the accuracy of the portfolio variance forecasts and broad factor variance forecasts persists, noise in $\hat{\Sigma}$ exacerbates the underforecasting of narrow and specific variance. For the range of parameters considered, accuracy ratios are relatively consistent although the number of broad factors estimated, $\mathscr{K}$, varies materially. A deeper investigation indicates that higher rank solutions erroneously identify noise as broad factors will small eigenvalues. As in the noiseless base case, PCA is less accurate than any of the SLR models, even though we endow PCA with the knowledge of the true number of broad factors.

In Figures 1–4, we show the largest eigenvalues of sample covariance matrices as well as the true and recovered covariance matrices and their decompositions for a universe of $N = 125$ securities. We illustrate both SLR and PCA decompositions, and both noiseless benchmarks and simulations.

Figure 1 shows eigenvalues of the true, sample and recovered covariance matrices, Σ, $\hat{\Sigma}$ and $\Sigma(\theta)$. In the noiseless examples shown in the top two panels, the true and sample covariance matrices are identical. SLR (top left panel) replicates the largest eigenvalues of Σ better than PCA (top right panel), which overestimates the top two eigenvalues. This overestimate points to an intrinsic bias in basic PCA estimators. According to the first Horn inequality, the largest eigenvalue Σ is bounded above by the sum of the

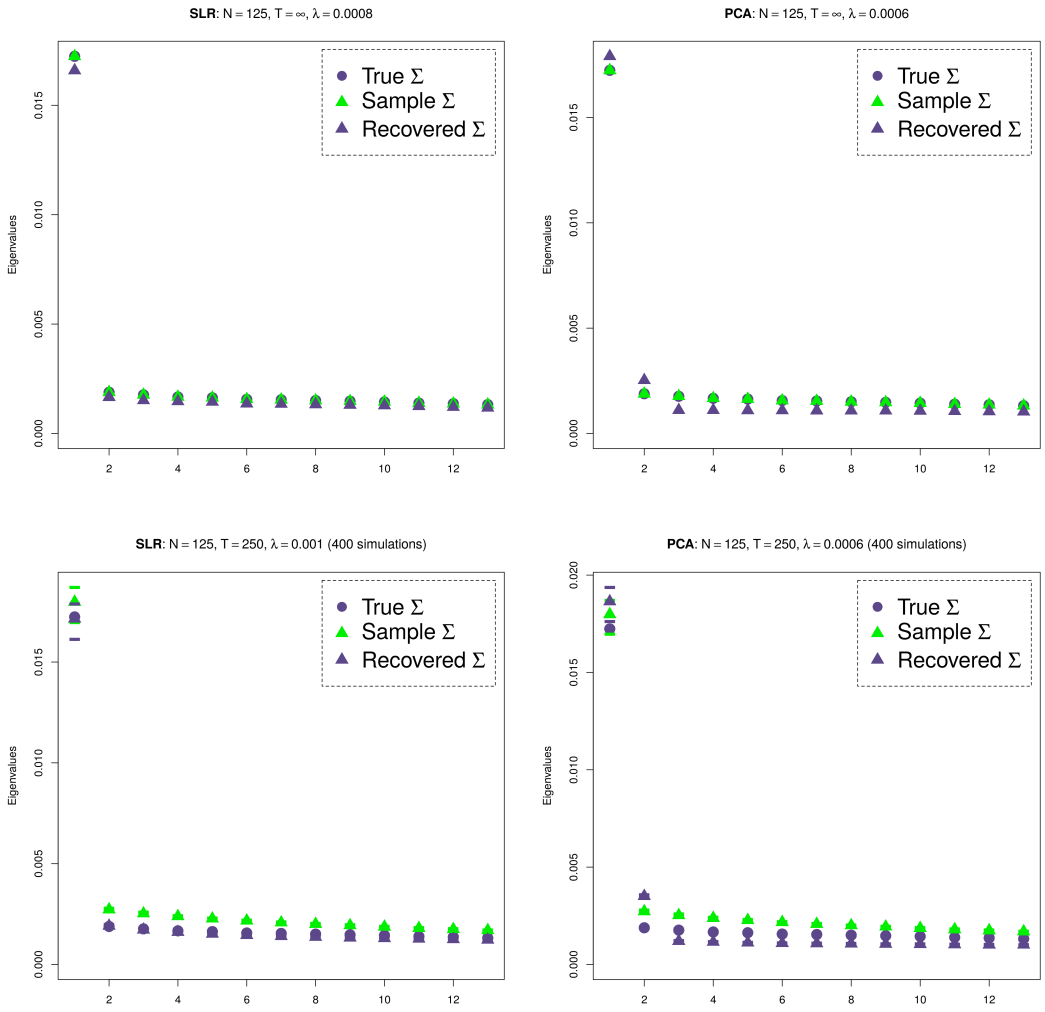

Figure 1: Eigenvalues of true, sample and recovered covariance matrices for $N = 125$ securities. Top panels correspond to the noiseless base case, where the true and sample covariance matrices coincide. Bottom panels correspond to $\rho = 400$ simulations, each with $T = 250$ daily observations. Left panels correspond to SLR (top: $\theta = (\lambda, \gamma) = (0.0002, 0.089)$, Table 1 column 2; bottom: $\theta = (\lambda, \gamma) = (0.0008, 0.089)$, Table 2 column 2). Right panels correspond to PCA (top: Table 1 column 4; bottom: Table 2 column 4).

largest eigenvalues of L and S.[8] By construction, however, the largest eigenvalue of the PCA estimate of Σ is bounded below by the largest eigenvalue of $\hat{\Sigma}$.

Figure 2 extends the investigation in Figure 1 by showing true and recovered eigenvalues of the low rank and sparse components of Σ. The top right panel of Figure 2 shows that the overestimation of the top two eigenvalues of Σ by PCA can be traced to the low rank component. The top left panel of Figure 2 shows that the overall agreement of recovered and true eigenvalues of the covariance matrix for SLR extends to the low rank and sparse components.

The bottom two panels of Figures 1 and 2 show results for simulations, where the true covariance matrix Σ is not observable, so matrix decomposition is based on the sample covariance matrix $\hat{\Sigma}$. For sample and recovered quantities, we show the median and interquartile range. For the SLR example (Figure 1, bottom left panel), the median top recovered eigenvalue is closer to the truth than to the median sample, even though the truth is not observable. For PCA(2) (Figure 1, bottom right panel), both the true and median sample top eigenvalues are outside the interquartile range of the recovered top eigenvalue.

In Figure 3, we show the eigenvalues of Σ and its low rank and sparse components on the same plot. The image suggests how the eigenvalues of the components combine to generate the eigenvalues of the sum.[9]

[8] More information about Horn's inequalities is in Knutson & Tao (2001).

[9] More information about eigenvalue sums is in Knutson & Tao (2001).

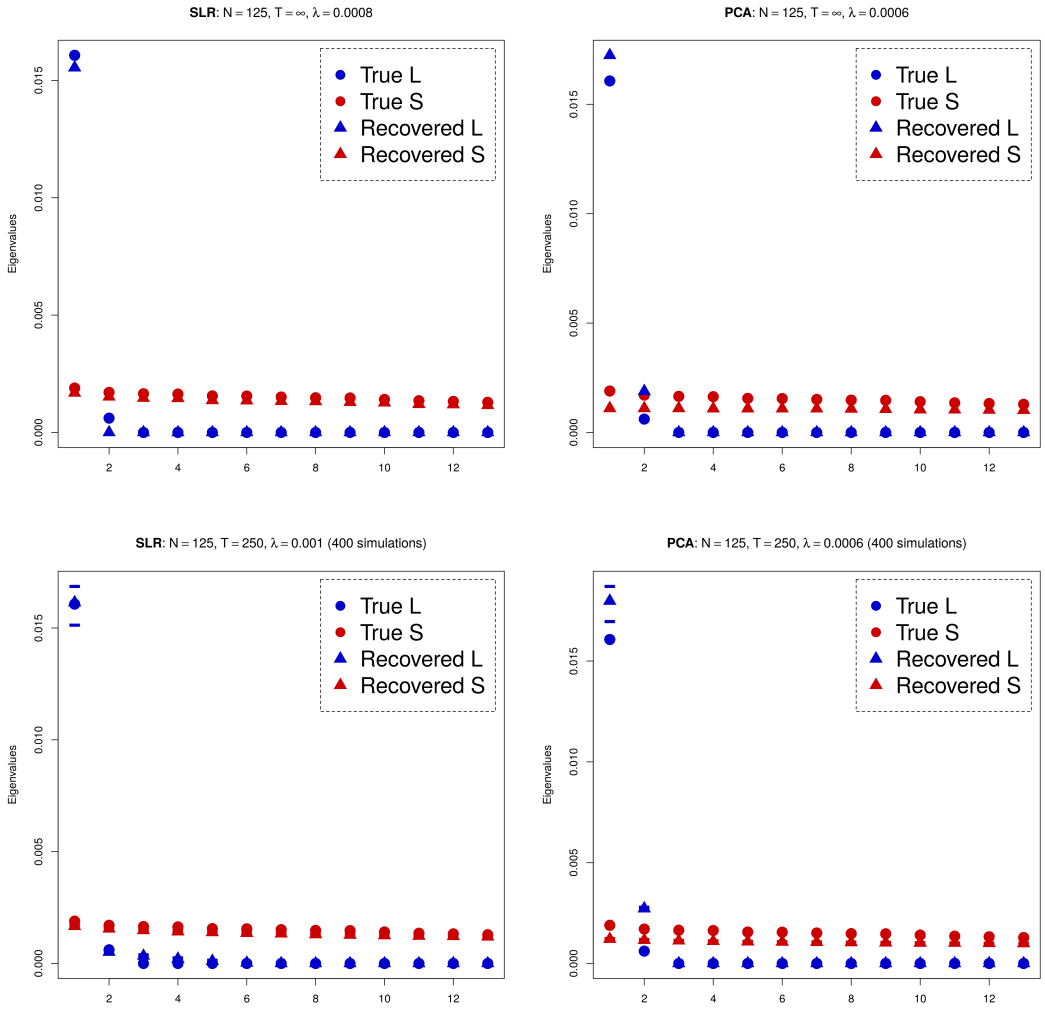

Figure 2: Eigenvalues of low rank and sparse components of true and recovered covariance matrices for $N = 125$ securities. Top panels correspond to the noiseless base case, where the true and sample covariance matrices coincide. Bottom panels correspond to $\rho = 400$ simulations, each with $T = 250$ daily observations. Left panels correspond to SLR (top: $\theta = (\lambda, \gamma) = (0.0002, 0.089)$, Table 1 column 2; bottom: $\theta = (\lambda, \gamma) = (0.0008, 0.089)$, Table 2 column 2). Right panels correspond to PCA (top: Table 1 column 4; bottom: Table 2 column 4).

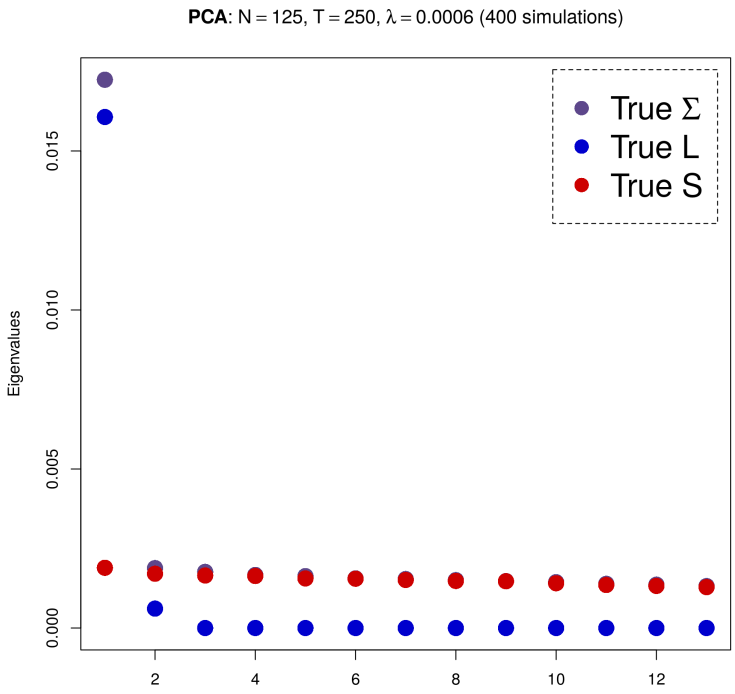

Figure 3: Eigenvalues of the true covariance matrix for $N = 125$ securities along with those for its low rank and sparse components.

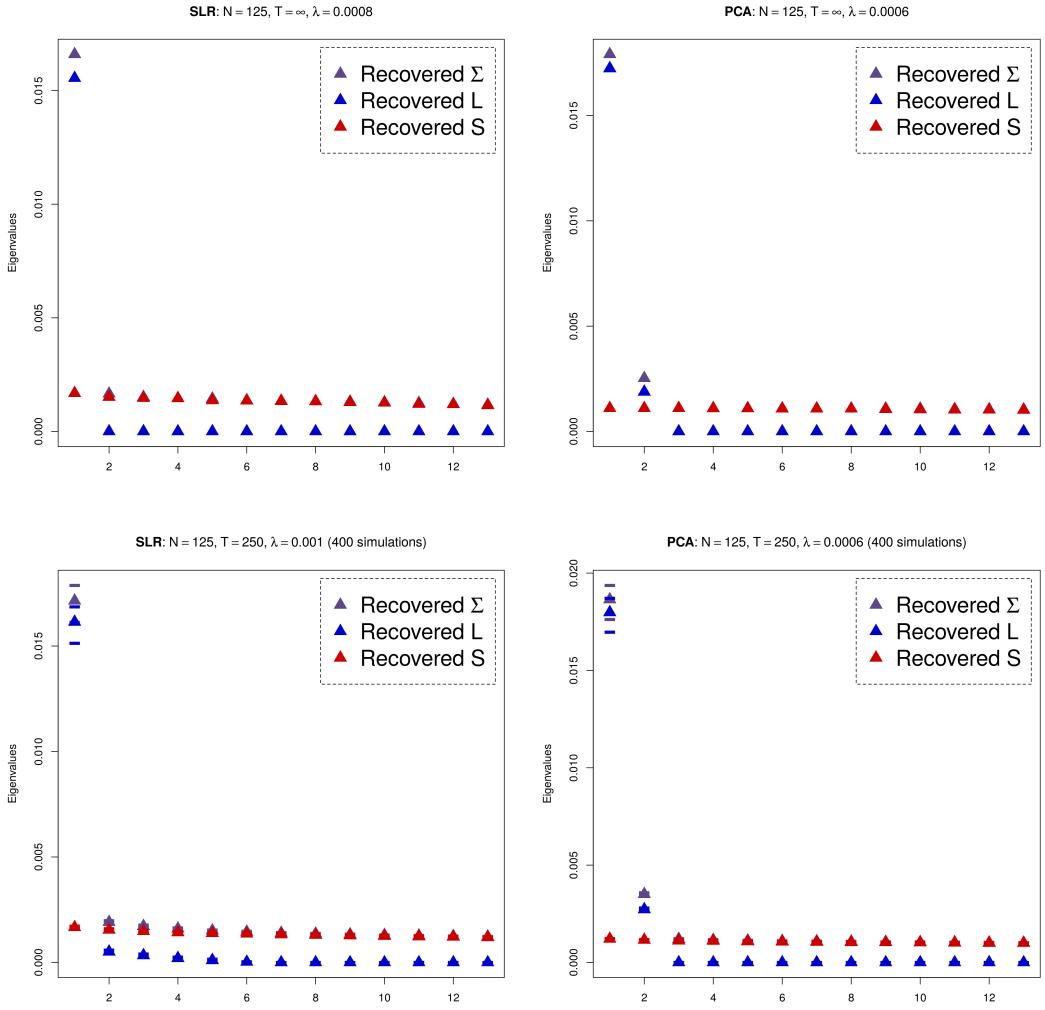

Figure 4: Eigenvalues of recovered covariance matrices for $N = 125$ securities along with their low rank and sparse components. Top panels correspond to the noiseless base case, where the true and sample covariance matrices coincide. Bottom panels correspond to $\rho = 400$ simulations, each with $T = 250$ daily observations. Left panels correspond to SLR (top: $\theta = (\lambda, \gamma) = (0.0002, 0.089)$, Table 1 column 2; bottom: $\theta = (\lambda, \gamma) = (0.0008, 0.089)$, Table 2 column 2). Right panels correspond to PCA (top: Table 1 column 4; bottom: Table 2 column 4).

Analogous plots for the four examples are in Figure 4. Visually, the eigenvalue sums for the SLR noiseless base case (top left panel) and simulation (bottom left panel) resemble the true eigenvalue sum more faithfully than do the PCA noiseless base case (top right

panel) and simulation (bottom right panel).

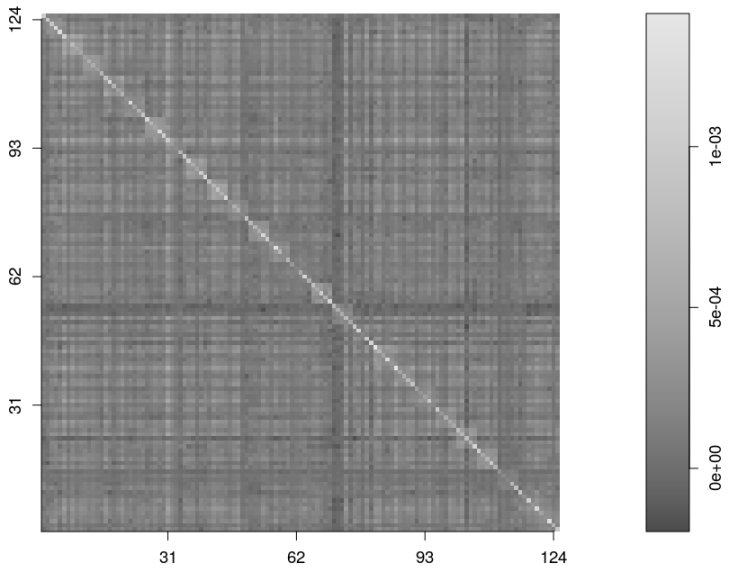

Figure 5: Sample covariance matrix, $N = 125$ securities, $T = 250$ daily observations.

To conclude the analysis of our simulations, we look at heat maps of the input and output correlations matrices of a single run of SLR.[10] Figure 5 shows a sample correlation matrix estimated from $T = 250$ simulated observations of $N = 125$ securities. Apart from the ones along the diagonal, the most striking feature is the block diagonal structure, which corresponds to the 25 narrow factors. Figure 6 shows the true and recovered low rank correlation matrices, which are structurally similar to the eye. Both the true and recovered sparse components, shown in Figure 7, feature a block diagonal structure.

[10]The SLR paramaters are $\theta = (\lambda, \gamma) = (0.0008, 0.089)$, and the results correspond to the bottom left panels of Figures 1, 2 and 4, and to Table 2 column 3.

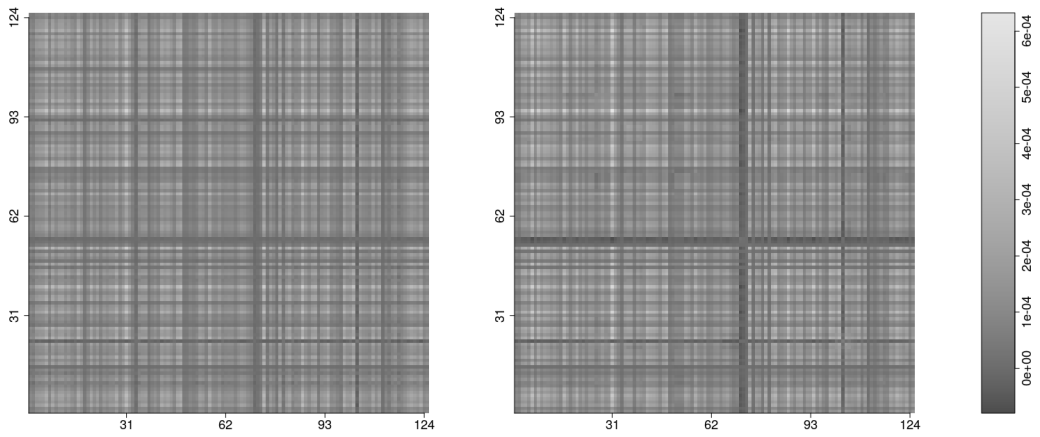

Figure 6: Low rank covariance matrices, $N = 125$ securities, $T = 250$ daily observations. Left panel: True. Right panel: Recovered, SLR with $\theta = (\lambda, \gamma) = (0.0008, 0.089)$.

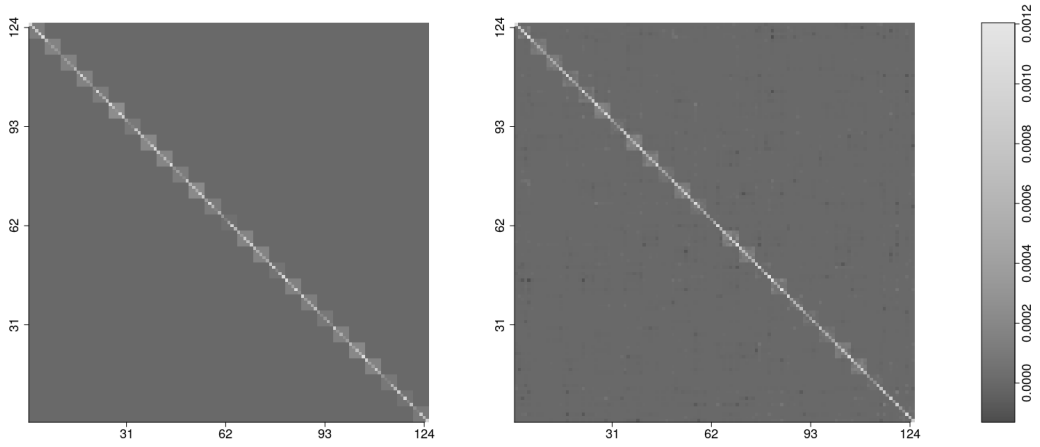

Figure 7: Sparse components of the covariance matrix, $N = 125$ securities, $T = 250$ daily observations. Left panel: True. Right panel: Recovered, SLR with $\theta = (\lambda, \gamma) = (0.0008, 0.089)$.

7 Empirical experiments

In this exploratory section, we illustrate the performance of SLR on empirical data with heat maps of low rank and sparse return correlation matrices recovered from an empirical covariance matrix, $\hat{\Sigma}$, estimated from $T = 250$ daily returns to $N = 125$ global equities. The sample period ended 31 October 2015. We randomly selected securities from $\kappa = 25$ countries. The block diagonal structure of the input correlation matrix shown in Figure 8 indicates that country effects contribute to empirical correlations.

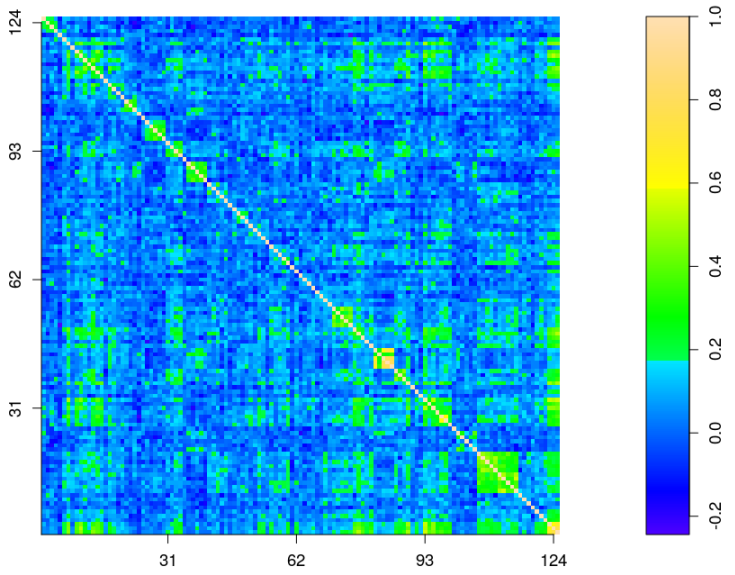

Figure 8: Empirical correlation matrix for $\hat{\Sigma}$. Data source: State Street GX Labs.

Figure 9 shows heat maps of the low rank (left panel) and sparse (right panel) correlation matrices of $L(\theta)$ and $S(\theta)$ recovered by SLR with $\theta = (\lambda, \gamma) = (0.001, 0.089)$. In this recovery, there are 15 broad factors, and the block diagonal structure is visible in the sparse component. Figure 10, illustrates the eigenvalues of $L(\theta)$ and $S(\theta)$ recovered by the SLR decomposition. Unlike PCA, the positive eigenvalues of the low rank (broad factor) component can be interspersed with the eigenvalues of the sparse component. This key feature distinguishes SLR methods from PCA as the latter assumes the broad factors are necessarily variance maximizing latent variables, while the latter does not. Section 6, Figure 2 (top right panel) illustrates this distinction in a simulation

where the eigenvalues of the low rank and sparse components of the true covariance matrix are interspersed. This intertwining is correctly captured by the SLR estimate, but it is not (and cannot be) captured by PCA.

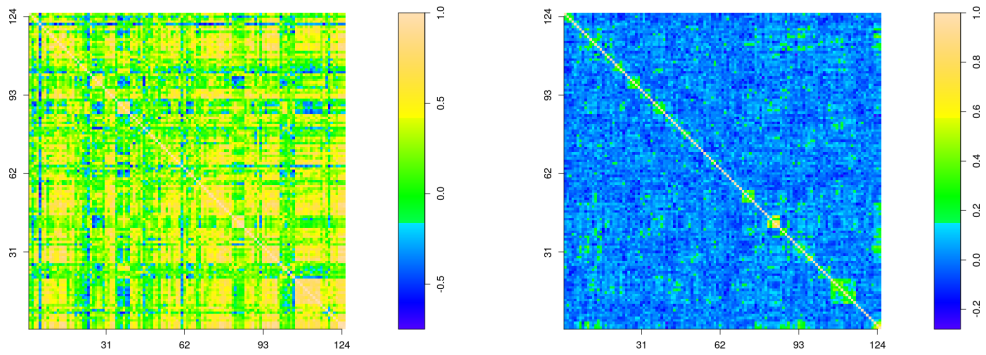

Figure 9: Low rank and sparse correlation matrices recovered from $\hat{\Sigma}$ by SLR with $\theta = (\lambda, \gamma) = (0.001, 0.089)$. The low rank component is based on two broad factors. Data source: State Street GX Labs.

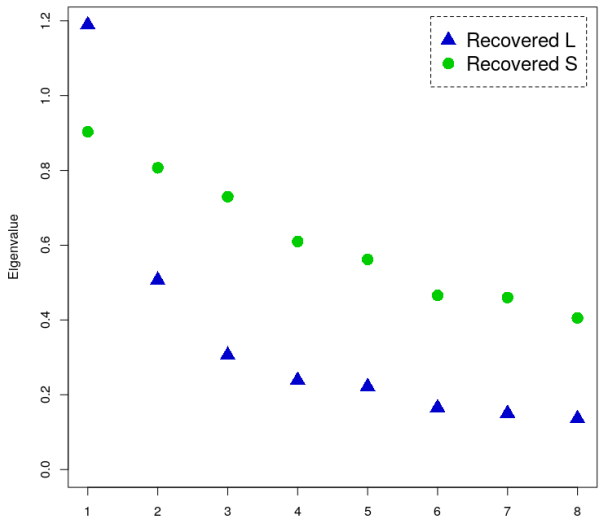

Figure 10: Low rank and sparse eigenvalues for SLR with $\theta = (\lambda, \gamma) = (0.001, 0.089)$. Data source: State Street GX Labs.

8 Conclusion

The use of convex optimization to identify the risk factors that drive returns to financial securities is a recent innovation. In this article, we begin to explore the efficacy of a particular convex program that decomposes a sample concentration matrix into low rank component that includes broad factors affecting most securities, and a sparse component that includes both narrow factors such as industries and countries as well as specific return. On a theoretical basis, there are advantages to convex optimization over traditional approaches based on principal component analysis. For example, convex optimization does not rely on spectral analysis to identify factors, and it is less sensitive to outliers than classical principal component analysis.

To illustrate the efficacy of convex optimization, we measure the variance forecasting accuracy of a SLR covariance matrix estimate on an equally weighted portfolio of 125 securities simulated from a model with two broad factors and 25 narrow factors. Our accuracy measures indicate a tendency to underforecast variance due to narrow factors and idiosyncratic effects, but not as severely as classical PCA. Further, the way in which the spectra of SLR recovered low rank and sparse components combine is closer to the true combination that classical PCA.

Our analysis highlights the fact that SLR avoids an intrinsic bias present in classical PCA. According to the simplest Horn inequality, the largest eigenvalue of a true low rank plus sparse covariance matrix Σ is bounded above by the sum of the largest eigenvalues of the summands. By construction, however, the largest eigenvalue of a PCA estimate of the covariance matrix is bounded below by the largest eigenvalue of the sample covariance matrix.

In a final experiment, we illustrate a SLR decomposition of an empirical covariance matrix of 125 equities drawn from 25 countries. The block diagonal structure associated with country effects is clearly visible in the sparse component.

The analysis in this article is just a small indication of the possibilities that convex optimization holds for the identification of financial risk factors, which, to date, has been largely the province of fundamental analysts.

A Gaussian latent graphical model

We justify the application of convex problem (10) of Chandrasekaran et al. (2012) to decompose the inverse of Σ in (6). The mathematical setting of Chandrasekaran et al. (2012) concerns observed and latent variables that jointly Gaussian. To this end, consider (R, ψ) ordered as a zero-mean Gaussian random vector in $\mathbb{R}^{N+K}$ with a symmetic, positive definite covariance block matrix

$$\begin{bmatrix} \Sigma & YF \\ FY^\top & F \end{bmatrix}. \tag{20}$$

In particular, there are N observed variables R and K latent (or hidden) variables ψ. It is well known that in this setting (for example see Anderson (2003)) the marginal covariance matrix Σ of the observed variables R has the following representation.

$$\Sigma = YFY^\top + \mathbf{Var}(R \mid \psi) \tag{21}$$

Here, $\mathbf{Var}(R \mid \psi)$ is the conditional (on the latent variables) covariance matrix of the observed variables. Due to the Gaussian assumption, it does not depend on the realization of the latent variables ψ. Setting $S = \mathbf{Var}(R \mid \psi)$ in (5), the model

$$R = Y\psi + \delta \tag{22}$$

serves as a representation for (R, ψ) with covariance matrix in (20). The broad factors in (1) are the latent variables and the residuals are the conditional (on the latent variables) returns. In our Gaussian setting, entries of S depict the pairwise covariances of the security returns $\delta = X\phi + \epsilon$ (see (3)) conditional on the broad factor returns. Indeed, $\mathbf{E}(\delta\delta^\top) = S$.

B Algorithm

We summarize the "Alternating Direction Method of Multipliers" suggested by Stephen Boyd and developed by Ma, Xue & Zou (2013). Setting aside the derivation we detail Algorithm 2 of the latter. This algorithm solves the convex program (10) to decompose the concentration matrix of the security returns as in (7).

Define variables $Q^k, S^k, \mathcal{L}^k, M^k \in \mathbb{R}^{N\times N}$. The superscript k indicates the iterate. The matrix $\hat{\Sigma}$ denotes the input sample covariance matrix. Constants $\tau \in (0, 0.5)$ and $\mu > 0$ are the step size and the augmented Lagrangian parameter. The range of τ guarantees global convergence (Ma et al. 2013). Define the shrinkage operator (extending it to matrices by application entrywise) $\mathscr{H}_z(x) = \operatorname{sgn}(x)\max\{|x - z|, 0\}$. With some $\mathcal{L}^0, S^0, Q^0, M^0$, for $k = 1, 2, \ldots$ until convergence, we perform the steps:

1. Compute the SVD of the matrix $\mu\hat{\Sigma} - S^k + \mathcal{L}^k - \mu M^k$ and denote it by $U\,\mathrm{diag}(\sigma)U^\top$. Set the next iterate for Q as

$$Q^{k+1} = U\,\mathrm{diag}(\gamma)U^\top; \qquad \gamma_i = \frac{1}{2}\left(-\sigma_i + \sqrt{\sigma_i^2 + 4\mu}\right). \tag{23}$$

2. Set $D = Q^k - S^k + \mathcal{L}^k - \mu M^k$ and apply the shrinkage operator

$$S^{k+1} = \mathscr{H}_{\gamma\mu\lambda\tau}(S^k + \tau D). \tag{24}$$

3. Compute the SVD of $\mathcal{L}^k - \tau D$ and denote it by $U\,\mathrm{diag}(\sigma)U^\top$, and

$$\mathcal{L}^{k+1} = U\,\mathrm{diag}(\gamma)U^\top; \qquad \gamma_i = \max\left\{\sigma_i - \mu\lambda, 0\right\}. \tag{25}$$

4. Update the Lagrange multipliers $M^{k+1} = M^k - (1/\mu)(Q^k - S^k + \mathcal{L}^k)$.

It is recommended to "symmetrize" each of the iterates after each step, e.g., to apply the map $X \to \frac{1}{2}(X + X^\top)$ to $Q^k, S^k, \mathcal{L}^k$ and M^k. Initialization depends on our best guess of the solution. In the absence of such a guess, we set $S^0 = \mathcal{L}^0 = Q^0 = M^0 = 0_{N\times N}$. Stopping criteria are subjective. Ma et al. (2013) propose to stop either when

$$\delta_k^Z < 10^{-4}; \qquad \delta_k^Z \triangleq \frac{\|Q^k - S^k + \mathcal{L}^k\|_F}{\max\{1, \|Q^k\|_F, \|S^k\|_F, \|\mathcal{L}^k\|_F\}} \tag{26}$$

or when the difference in the objective function values in two successive steps is less than 10^{-6}. Here, $\|\cdot\|_F$ is the Frobenius norm. The objective function is computed as

$$O(Q, \mathcal{L}, S) = -\ell(Q; \hat{\Sigma}) + \lambda\left(\sum_{i=1}^{N}\sigma_i + \gamma\sum_{i=1}^{N}\sum_{j=1}^{N}|S_{ij}|\right)$$

where $\{\sigma_i\}$ are the singular (eigenvalues) of L. Define, the difference

$$\delta_k^O = |O(Q^{k+1}, \mathcal{L}^{k+1}, S^{k+1}) - O(Q^k, \mathcal{L}^k, S^k)| \tag{27}$$

and so the above condition says $\delta_k^F < 10^{-6}$. Ma et al. (2013) suggest taking $\mu = 10$ and $\tau = 0.6$ to improve convergence rates. We are more interested in accuracy and select $\tau = 0.1$ and $\mu = 10 \times \|\hat{\Sigma}\|_F / N$.

References

Agarwal, A., Negahban, S., Wainwright, M. J. et al. (2012), 'Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions', *The Annals of Statistics* **40**(2), 1171–1197.

Anderson, T. T. W. (2003), *An introduction to multivariate statistical analysis*, John Wiley & Sons.

Bender, J., Lee, J.-H., Stefek, D. & Yao, J. (2009), 'Forecast risk bias in optimized portfolios'.

Bianchi, S. W., Goldberg, L. R. & Rosenberg, A. (2016), 'The impact of estimation error on latent factor models of portfolio risk', *forthcoming in The Journal of Portfolio Management* .

Candès, E. J., Li, X., Ma, Y. & Wright, J. (2011), 'Robust principal component analysis?', *Journal of the ACM (JACM)* **58**(3), 11.

Chandrasekaran, V., Parrilo, P. A. & Willsky, A. S. (2010), Latent variable graphical model selection via convex optimization, *in* 'Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on', IEEE, pp. 1610–1613.

Chandrasekaran, V., Parrilo, P. A. & Willsky, A. S. (2012), 'Latent variable graphical model selection via convex optimization', *The Annals of Statistics (with discussion)* **40**(4), 1935–1967.

Chandrasekaran, V., Sanghavi, S., Parrilo, P. A. & Willsky, A. S. (2011), 'Rank-sparsity incoherence for matrix decomposition', *SIAM Journal on Optimization* **21**(2), 572–596.

Dempster, A. P. (1972), 'Covariance selection', *Biometrics* pp. 157–175.

Fan, J., Han, F. & Liu, H. (2014), Page: Robust pattern guided estimation of large covariance matrix, Technical report, Technical report, Princeton University.

Horn, R. A. & Johnson, C. R. (2012), *Matrix analysis*, Cambridge university press.

Jolliffe, I. (2002), *Principal component analysis*, Wiley Online Library.

Knutson, A. & Tao, T. (2001), 'Honeycombs and sums of hermitian matrices', *Notices Amer. Math. Soc* **48**(2).

Luo, X. (2013), 'Recovering model structures from large low rank and sparse covariance matrix estimation', *arXiv preprint arXiv:1111.1133* .

Ma, S., Xue, L. & Zou, H. (2013), 'Alternating direction methods for latent variable gaussian graphical model selection', *Neural computation* **25**(8), 2172–2198.

Miller, G. (2006), 'Needles, haystacks, and hidden factors', *Journal of Portfolio Management* **32**(2), 25.