Dimension Reduction and Coefficient Estimation in Multivariate Linear Regression

Zirui Pan[1] Haole Tian[2]

[1]School of Mathematics, University of Science and Technology of China
pzr201663@mail.ustc.edu.cn

[2]School of Mathematics, University of Science and Technology of China

December 22, 2023

About the author

Yuan Ming
Professor, Dept. of Statistics
Assoc. Director, Data Science Institute
Columbia University

Contact: 1255 Amsterdam Ave, NY, NY 10027
Email: ming.yuan@columbia.edu

Education:

- Ph.D. in Statistics, UW-Madison (2004)
- M.S. in CS, UW-Madison (2003)
- M.S. in Stats, USTC (2000)
- B.S. in EEIS, USTC (1997)

The main issues and contributions of this paper are as follows:

- Contribution 1: Introduction of a general formulation for dimension reduction and coefficient estimation in the multivariate linear model, highlighting that many existing methods can be expressed within this framework but have various limitations.

The main issues and contributions of this paper are as follows:

- Contribution 1: Introduction of a general formulation for dimension reduction and coefficient estimation in the multivariate linear model, highlighting that many existing methods can be expressed within this framework but have various limitations.
- Contribution 2: Proposal of a more flexible and universally applicable method, a novel penalized least squares estimate. The penalty used is the Ky Fan norm of the coefficient matrix.

The main issues and contributions of this paper are as follows:

- Contribution 1: Introduction of a general formulation for dimension reduction and coefficient estimation in the multivariate linear model, highlighting that many existing methods can be expressed within this framework but have various limitations.
- Contribution 2: Proposal of a more flexible and universally applicable method, a novel penalized least squares estimate. The penalty used is the Ky Fan norm of the coefficient matrix.
- Contribution 3: Introduction of a generalized cross-validation criterion for tuning parameter selection in the penalized least squares.

The main issues and contributions of this paper are as follows:

- Contribution 1: Introduction of a general formulation for dimension reduction and coefficient estimation in the multivariate linear model, highlighting that many existing methods can be expressed within this framework but have various limitations.
- Contribution 2: Proposal of a more flexible and universally applicable method, a novel penalized least squares estimate. The penalty used is the Ky Fan norm of the coefficient matrix.
- Contribution 3: Introduction of a generalized cross-validation criterion for tuning parameter selection in the penalized least squares.
- Contribution 4: Demonstration of the competitive performance of the new method through simulations and an application in financial econometrics.

The main issues and contributions of this paper are as follows:

- Contribution 1: Introduction of a general formulation for dimension reduction and coefficient estimation in the multivariate linear model, highlighting that many existing methods can be expressed within this framework but have various limitations.
- Contribution 2: Proposal of a more flexible and universally applicable method, a novel penalized least squares estimate. The penalty used is the Ky Fan norm of the coefficient matrix.
- Contribution 3: Introduction of a generalized cross-validation criterion for tuning parameter selection in the penalized least squares.
- Contribution 4: Demonstration of the competitive performance of the new method through simulations and an application in financial econometrics.
- Contribution 5: Discussion of an extension to non-parametric factor models, further broadening the applicability of the method.

Multivariate Linear Regression Models

In general multivariate linear regression, we have n observations on q responses $y = (y_1, \ldots, y_q)$ and p explanatory variables $x = (x_1, \ldots, x_p)$.

$$Y = XB + E \quad (1)$$

where:

- Y is an $n \times q$ matrix.
- X is an $n \times p$ matrix.
- B is a $p \times q$ coefficient matrix.
- $E = (e_1, \ldots, e_n)$ is the regression noise, and the es are independently sampled from $N(0, \Sigma)$.

Throughout the paper, we center each input variable to remove the intercept in equation (1) and scale each input variable so that the observed standard deviation is 1.

Estimating Coefficient Matrix B

The standard approach to estimating the coefficient matrix B is through ordinary least squares or maximum likelihood estimation methods (Anderson, 2003). The resulting estimates are equivalent to regressing each response on the explanatory variables separately.

Challenges:

- Suboptimal Performance: Estimates may underperform by neglecting related response information.
- Poor Performance: Especially with highly correlated variables or large p.

Addressing Challenges:

- Dimension Reduction: Methods use techniques to reduce dimensionality.
- Linear Factor Regression: Attractive approach regressing Y against a few transformed predictors (factors).

Factor Models

A compelling approach in multivariate linear regression is **linear factor regression**, where the response Y is regressed against a few linearly transformed predictors, often called **factors**.

This can be expressed as:

$$Y = F\Omega + E \quad (2)$$

where $F = X\Gamma$, Γ is a $p \times r$ matrix ($r \leq \min(p, q)$), and Ω is an $r \times q$ matrix. The columns of F, F_j for $j = 1, \ldots, r$, represent these factors.

Estimation in Linear Factor Regression

Estimation in linear factor regression typically involves two steps:

Step 1: *Factor Estimation*

- Factors (Γ) are estimated first.
- Common methods include canonical correlation, reduced rank, principal components, partial least squares, and joint continuum regression.
- These methods differ in how they determine the factors.

Step 2: *Parameter Estimation*

- Parameter Ω is estimated using least squares for the linear factor regression equation $Y = F\Omega + E$.

Estimation of r

Selecting the Number of Factors (r):

- Crucial for accuracy and model complexity.
- Determined separately through hypothesis testing or cross-validation.
- Coefficient matrix estimated based on the selected factors.
- Because of its discrete nature, this type of procedure can be very unstable in the sense of Breiman (1996): small changes in the data can result in very different estimates.

Dimension Reduction in Multivariate Regression

Modeling the jth Response:

- Denote by Y_j, B_j, and E_j the jth columns of Y, B, and E.
- $Y_j = XB_j + E_j$ for $B_j \in \mathbb{R}^p$, $j = 1, \ldots, q$.

Dimension Reduction Idea:

- Regression coefficients $B_1, B_2, \ldots, B_q$ are from a linear space B of lower dimension than p.
- Approach involves basis elements $\{\eta_1, \ldots, \eta_p\}$ for $\mathbb{R}^p$ and a subset $A \subseteq \{1, \ldots, p\}$ such that $B \subseteq \text{span}\{\eta_i : i \in A\}$
- In variable selection,$\eta_i = e_i$; and we want to estimate A. In the case of linear factor regression, the ith factor is given by $F_i = X\eta_i$, and A takes the form $\{1, 2, \ldots, r\}$, where r is to be estimated.

Our Proposal:

- Propose a procedure allowing simultaneous estimation of $\{\eta_i\}$ and A.

Simultaneous Estimation of Factors and Selection

Simultaneous Estimation Approach:

- Begin with factor selection, assuming $\{\eta_1, \ldots, \eta_p\}$ are known up to a permutation.
- Write $F = [X\eta_1, \ldots, X\eta_p]$, $Y = F\Omega + E$, where Ω is a $p \times q$ matrix such that $\{\eta_1, \ldots, \eta_p\}\Omega = B$.
- Factor selection is reformulated as a variable selection problem for $Y = F\Omega + E$.

As pointed out by Turlach et al. (2005):

- Minimize $tr\{(Y - F\Omega)^T W(Y - F\Omega)\}$ subject to $\sum_{i=1}^{p} \|\omega_i\|_\alpha \leq t$, where W is a weight matrix, ω_i is the i-th row of Ω, $t > 0$ is a regularization parameter, and $\|\cdot\|_\alpha$ is the l_α-norm for some $\alpha > 1$.
- Weight matrix choices include Σ^{-1} and I; here, we assume $W = I$.
- α most obvious choices include 2 and ∞; here, we assume $\alpha = 2$.

Factor Selection with Sparse Representation

Factor Selection Approach:

- Factor selection is most powerful when responses can be predicted by a small subset of common factors.
- Ideally, $\{\eta_1, \ldots, \eta_p\}$ should contain a set of basis for B to allow the sparsest representation of B in the factor space.
- In our method, we choose ηs to be the eigenvectors of BB^T.

Factor Selection without Factor Estimation:

- Choose $\{\eta_1, \ldots, \eta_p\}$ to be the eigenvectors of BB^T.
- write $U = [\eta_1, \ldots, \eta_p]$. The singular value decomposition of B is $B = UDV^T$ for some $q \times q$ orthonormal matrix V and a $p \times q$ matrix D with $D_{ij} = 0$ for $i \neq j$ and $D_{ii} = \sigma_i(B)$, where $\sigma_i(\cdot)$ is the i-th largest singular value. Now $\Omega = DV^T$ and $\omega_i = \sigma_i(B)V_i$, where V_i is the i-th column of V, implying that $\|\omega_i\|_2 = \sigma_i(B)$.

Factor Selection with Sparse Representation

Expression (5), with $\alpha = 2$, is given by:

$$\min \left[\mathrm{tr}\{(Y - XB)(Y - XB)^T\} \right] \text{ subject to } \sum_{i=1}^{\min(p,q)} \sigma_i(B) \leq t, \quad (5)$$

where $\sigma_i(B)$ represents the i-th largest singular value of B.

$\Sigma_{\min(p,q)}^{i=1} \sigma_i(B)$ is known as the Ky Fan (p or q) norm of B.

No knowledge of η is required in this expression, and we use the minimizer of (5) as our final estimate of B. In Appendix A, we show that expression (5) is equivalent to a conic program and can be computed efficiently.

Connection with Other Methods

The proposed estimate, defined as the minimizer of expression (5), is closely connected with several other popular methods. Specifically, expression (5), reduced rank regression, and ridge regression can all be viewed as minimizing

$$\mathrm{tr}\{(Y - XB)(Y - XB)^T\} \text{ subject to } \Big\{ \sum_{i=1}^{\min(p,q)} \sigma_i(B)^{\alpha} \Big\}^{\frac{1}{\alpha}} \le t,$$

with different choices of α.

- Ridge regression ($\alpha = 2$): $\mathrm{tr}\{(Y - XB)(Y - XB)^T\} + \lambda \mathrm{tr}(B^T B)$.
- Proposed estimate ($\alpha = 1$): enjoys a similar shrinkage property.
- Reduced rank regression ($\alpha = 0^+$).

To understand further the statistical properties of the method proposed, we consider the special case of orthogonal design. The following lemma gives an explicit expression for the minimizer of expression (5) in this situation.

Lemma

Lemma 1: Let $\hat{U}^{LS}$, $\hat{D}^{LS}$, and $\hat{V}^{LS}$ be the singular value decomposition of the least squares estimate $\hat{B}^{LS}$. Then, under the orthogonal design where $X^T X = nI$, the minimizer of expression (5) is

$$\hat{B} = \hat{U}^{LS}\hat{D}(\hat{V}^{LS})'$$

where $\hat{D}_{ij} = 0$ if $i \neq j$, $\hat{D}_{ii} = \max(\hat{D}^{LS}_{ii} - \lambda, 0)$, and $\lambda > 0$ is a constant such that $\sum_i \hat{D}_{ii} = \min(t, \sum_i \hat{D}^{LS}_{ii})$.

Proof of Lemma 1

Proof. Expression (5) can be written in a Lagrange form:

$$Q_n(B) = \frac{1}{2}\mathrm{tr}\{(Y - XB)^T(Y - XB)\} + n\lambda \sum_{i=1}^{min(p,q)} \sigma_i(B). \quad (7)$$

Simple algebra yields

$$\begin{aligned}&\mathrm{tr}\{(Y - XB)^T(Y - XB)\}\\ &\quad = \mathrm{tr}\{(Y - X\hat{B}^{LS})^T(Y - X\hat{B}^{LS})\} + n\mathrm{tr}\{(\hat{B}^{LS} - B)^T(\hat{B}^{LS} - B)\}.\end{aligned}$$

Together with the fact that $\mathrm{tr}(B^TB) = \sum_i \sigma_i^2(B)$, equation (7) equals

$$\frac{1}{2}\sum_{i=1}^{q}\sigma_i^2(B) - \mathrm{tr}(B^T\hat{B}^{LS}) + \lambda\sum_{i=1}^{q}\sigma_i(B) + C,$$

C not depending on B.

Proof of Lemma 1 (Contd.)

Now an application of von Neumann ' s trace inequality, which states that for matrices A and B,

$$\mathrm{tr}(AB) \leq \sum_{i=1}^{\min(m,n)} \sigma_i(A)\sigma_i(B),$$

where $\sigma_i(A)$ and $\sigma_i(B)$ are the singular values of matrices A and B, yields

$$Q_n(B) \geq \frac{1}{2}\sum_{i=1}^{q} \sigma_i^2(B) - \sum_{i=1}^{q} \sigma_i(B)\hat{D}_{ii}^{LS} + \lambda \sum_{i=1}^{q} \sigma_i(B).$$

Note that $\sigma_i(B) \geq 0$. The right-hand side is minimized at $\sigma_i(B) = \max(\hat{D}_{ii}^{LS} - \lambda, 0)$. The proof is now completed by noting that $\hat{B}$ achieves the lower bound for Q_n. ■

Specifically, the following lemma indicates that we can always find an appropriate tuning parameter such that the non-zero singular values of B are consistently estimated and the rest are set to 0 with probability 1.

Lemma

Lemma 2: Suppose that $\max(p, q) = o(n)$. *Under the orthogonal design, if* $\lambda \to 0$ *in such a fashion that* $\max(p, q) = n = o(\lambda^2)$, *then* $|\sigma_i(\hat{B}) - \sigma_i(B)| \to_p 0$ *if* $\sigma_i(B) > 0$, *and* $P\{\sigma_i(\hat{B}) = 0\} \to 1$ *if* $\sigma_i(B) = 0$.

Proof of Lemma 2

Proof : Note that

$$\hat{B}^{LS} = (X^T X)^{-1} X^T Y = B + \frac{X^T E}{n},$$

where E follows $N(0, \Sigma)$. Since $X^T X = nI$ and the rows of E are independent observations from $N(0, \Sigma)$, each entry of $\frac{X^T E \Sigma^{-1/2}}{\sqrt{n}}$ follows $N(0, 1)$ and is independent of each other.

Applying the result from Johnstone (2001), we have

$$\sigma_1(X^T E \Sigma^{-1/2})/\sqrt{n} \sim (\sqrt{p} + \sqrt{q})/\sqrt{n}.$$

Therefore,

$$\frac{\sigma_1(X^T E)}{n} \leq \sigma_1(\frac{X^T E \Sigma^{-1/2}}{n})\sigma_1(\Sigma^{1/2}) \sim \sigma_1^{1/2}(\Sigma)\frac{\sqrt{p} + \sqrt{q}}{\sqrt{n}}$$

Proof of Lemma 2 (Contd.)

Now, an application of Theorem 3.3.16 of Horn and Johnson (1991) yields

$$|\sigma_i(B) - \sigma_i(\hat{B}^{LS})| \leq \frac{\sigma_1(X^T E)}{n} = O_p\left(\frac{\sqrt{p} + \sqrt{q}}{\sqrt{n}}\right).$$

Therefore, if $\lambda \to 0$ at a slower rate than the right-hand side of the equation (10), the proposed estimate can provide consistent estimates of the non-zero singular values of B and simultaneously shrink the remaining singular values to 0.
Lemma 2 also indicates that the singular values of the method proposed are shrunk in a similar fashion to the lasso under orthogonal designs.

Tuning Parameter Selection: GCV-type Statistic

Like any other regularization method, choosing an appropriate tuning parameter t in expression (5) is crucial. Before this ,We first characterize the equivalence between expression (5) and its Lagrange form (7).

Lemma

Lemma 3: Denote $\hat{B}$ as the minimizer of expression (5) and $\hat{U}\hat{D}\hat{V}^T$ its singular value decomposition. Write $\hat{d}_i = \hat{D}_{ii}$ for $i = 1, \ldots, \min(p, q)$. For any $t \leq \sum_i \hat{d}_i$, the minimizer of equation (7) coincides with the minimizer of expression (5), $\hat{B}$, if

$$n\lambda = \frac{1}{\mathrm{card}(\hat{d}_i > 0)} \sum_{\hat{d}_i > 0} \tilde{X}_i^T \tilde{Y}_i - \tilde{X}_i^T \tilde{X}_i \hat{d}_i. \quad (11)$$

where $\mathrm{card}(\cdot)$ stands for the cardinality of a set, $\tilde{Y}_i$ is the ith column of $\tilde{Y} = Y\hat{U}$, and $\tilde{X}_i$ is the ith column of $\tilde{X} = XV$.

Proof of Lemma 3

Proof. Note that

$$\sum_{i=1}^{\min(p,q)} \sigma_i(\hat{B}) = \sum_{i=1}^{\min(p,q)} \hat{D}_{ii} = \sum_{i=1}^{p} \sigma_i(\hat{B}K\hat{B}^T) = \mathrm{tr}(\hat{B}K\hat{B}^T),$$

where

$$K = \sum_{\hat{D}_{ii}>0} \frac{1}{\hat{D}_{ii}} \hat{V}_i \hat{V}_i^T,$$

and $\hat{V}_i$ is the i-th column of V. Therefore, $\hat{B}$ is also the minimizer of

$$\frac{1}{2}\mathrm{tr}\{(Y - XB)^T(Y - XB)\} + n\lambda \mathrm{tr}(BKB^T). \quad (12)$$

Proof of Lemma 3 (Contd.)

From expression (12), $\hat{d}$ is the minimizer of

$$\frac{1}{2}\sum_{i=1}^{\min(p,q)}(\tilde{Y}_i - \tilde{X}d_i)^T(\tilde{Y}_i - \tilde{X}d_i) + n\lambda\sum_{i=1}^{\min(p,q)} d_i,$$

subject to the constraint that $d_i \geq 0$. The first-order optimality condition for this expression yields

$$n\lambda = \tilde{X}_i^T\tilde{Y}_i - \tilde{X}_i^T\tilde{X}_i\hat{d}_i,$$

for any $\hat{d}_i > 0$. The proof is now completed by taking an average of the above expression over all i such that $\hat{d}_i > 0$.

Implementation Steps for Tuning Parameter

Since $\hat{B}$ is the minimizer of expression (12), it can be expressed as

$$\hat{B} = (X^T X + 2n\lambda K)^{-1} X^T Y.$$

GCV Score Definition(Golub et al., 1979) :
Now the Generalized Cross Validation (GCV) score is given by

$$GCV(t) = \frac{\mathrm{tr}\{(Y - XB)^T (Y - \hat{B})\}}{qp - \mathrm{df}(t)},$$

where $\mathrm{df}(t)$ is the degrees of freedom defined as

$$\mathrm{df}(t) = q\mathrm{tr}\{X(X^T X + 2n\lambda K)^{-1} X^T\}.$$

Implementation Steps for Tuning Parameter

Tuning Parameter Selection: The tuning parameter is chosen by minimizing $GCV(t)$. The implementation steps are:

1. For each candidate t-value:
    1. Compute the minimizer of expression (5) (denote the solution as $\hat{B}_t$).
    2. Evaluate λ by using equation (11).
    3. Compute the GCV score.
2. Denote t^* as the minimizer of the GCV score obtained in step 1. Return $\hat{B}_{t^*}$ as the estimate of B.

Simulation Models

1. **Model I:** $p = q = 8$, A random 8×8 matrix with singular values $(3, 2, 1.5, 0, 0, 0, 0, 0)$ was first generated as the true coefficient matrix,Predictor $\mathbf{x}$ is generated from a multivariate normal distribution with correlation between x_i and x_j being $0.5^{|i-j|}$, $\mathbf{y}$ is generated from $N(\mathbf{x}B, I)$,The sample size for this example is $n = 20$.
2. **Model II:** Same as Model I, but singular values $\sigma_1 = \ldots = \sigma_8 = 0.85$.
3. **Model III:** Same set-up, singular values $(5, 0, 0, 0, 0, 0, 0, 0)$.
4. **Model IV:** $p = 20$, $q = 20$, first 10 singular values 1, last 10 singular values 0, $n = 50$.

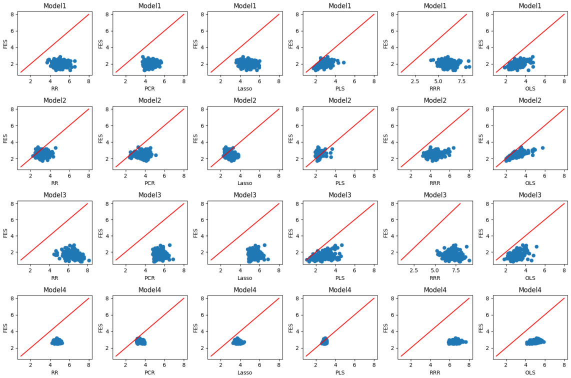
Comparisons on the Simulated Data Sets

Table: Results for the following methods: FES, RR, PCR, LASSO, PLS, RRR, OLS

Model	FES	RR	PCR	LASSO	PLS	RRR	OLS
I	2.01	5.10	4.39	4.85	2.88	6.09	3.15
	(0.09)	(0.27)	(0.23)	(0.24)	(0.26)	(0.39)	(0.31)
II	2.51	3.21	3.66	2.90	2.28	4.45	3.11
	(0.09)	(0.16)	(0.26)	(0.10)	(0.06)	(0.30)	(0.32)
III	1.55	6.59	5.37	5.37	2.76	7.35	3.10
	(0.15)	(0.36)	(0.16)	(0.16)	(0.51)	(0.42)	(0.26)
IV	2.75	4.68	3.34	3.76	2.97	6.50	4.82
	(0.01)	(0.04)	(0.04)	(0.04)	(0.01)	(0.10)	(0.10)

Comparison and Prediction Accuracy

To gain further insight into the comparison, we provide a pairwise prediction accuracy comparison between method FES and the other methods



Real Example: Financial Econometrics

To demonstrate the utility of the proposed method, we consider a real example in financial econometrics.

Vector Autoregressive Model:

$$y_t = y_{t-1}B + E \quad (15)$$

Model (15) is a special case of the multivariate linear model. Accurate estimation of B in model (15) leads to good forecasts, serving as instruments for efficient portfolio allocation and identifying arbitrage opportunities. Identification of factors in model (15) is crucial for constructing benchmark portfolios and diversifying investments.

Application: Stock Prices

To illustrate our method, we select leading enterprises in different industries in China and fit their stock prices.

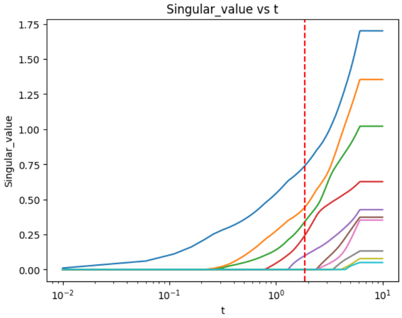


Figure: Singular Value vs t

Application: Stock Prices

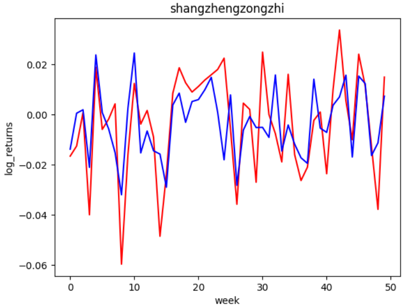


Figure: Chart of Shanghai Composite Index

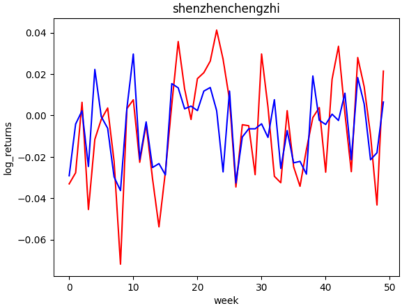


Figure: Chart of Shenzhen Component Index

Prediction Error Comparison

To compare the proposed method with others, we evaluated their prediction errors on data from the second half of the year.

Table: Out-of-sample mean-squared error for various methods

Company	FES	OLS	RR	PCR	Lasso	PLS	RRR
贵州茅台	48758.05	50110.59	124239.45	113472.23	112721.87	110825.09	12906.61
比亚迪	1885.64	16737.83	2354.53	2755.39	3110.51	3095.4	1420.48
中国平安	20.1	52.99	18.68	27.21	22.45	22.34	30.24
格力电器	4.21	7.9	4.67	4.4	4.39	4.46	24.15
海天味业	75.28	48.56	216.67	195.2	200.38	195.08	5150.7
恒瑞药业	5.39	9.31	4.82	5.35	5.79	5.87	85.58
科大讯飞	28.96	63.2	26.39	47.52	34.19	33.8	10.35
万科 A	12.44	75.29	12.2	13.23	13.64	13.47	13.22
伊利股份	46.54	34.37	59.95	52.03	50.5	50.66	549.53
云南白药	31.27	395.25	52.19	54.04	33.83	31.44	18.74

Original Results

Table 2. Factor loadings for the stocks example

Company	*Loadings for the following factors:*			
	1	*2*	*3*	*4*
Walmart	−0.47	−0.42	−0.30	0.19
Exxon	0.20	−0.68	0.07	−0.40
GM	0.05	0.19	−0.61	−0.31
Ford	0.18	0.22	−0.42	−0.13
GE	−0.35	0.13	−0.03	−0.44
ConocoPhillips	0.42	0.04	0.05	−0.52
Citigroup	−0.45	0.13	−0.26	−0.17
IBM	−0.24	0.43	0.49	−0.21
AIG	−0.38	−0.22	0.22	−0.39

Original Results

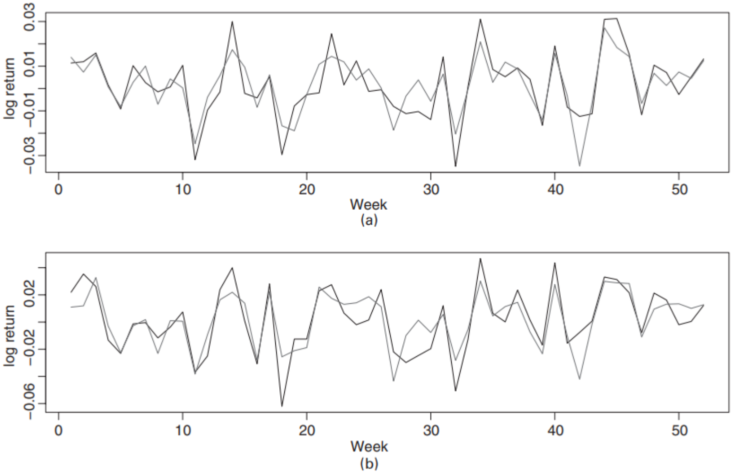


Fig. 3. (a) S&P500 and (b) NASDAQ indices (———) together with their approximations in the factor space (·······)

Solution Approach

To solve expression (5), we leverage recent advancements in convex optimization. We demonstrate that expression (5) is equivalent to a second-order cone program (SOCP) and can be effectively solved using standard solvers such as SDPT3 .

Notation:

- L_m: m-dimensional second-order cone ($\{x \in \mathbb{R}^m : x_1 \geq \sqrt{x_2^2 + \ldots + x_m^2}\}$)
- R_m^+: Positive orthant in $\mathbb{R}^m$
- $X \succeq 0$: Indicates that the symmetric matrix X is positive semidefinite
- $svec(X)$: Vectorization operator for symmetric matrix X
- $svec(X) = (X_{11}, X_{21}\sqrt{2}, X_{22} \ldots, X_{n1}\sqrt{2} \ldots, X_{n,n-1}\sqrt{2}, X_{nn})$

Solution Approach

SDPT3 can solve problems of the form:

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{n_s} \mathrm{tr}(C_j^s X_j^s) + \sum_{i=1}^{n_q} (c_i^q)^T x_i^q + (c^l)^T x^l \\ \text{subject to} \quad & \sum_{j=1}^{n_s} (A_j^s)^T \mathrm{svec}(X_j^s) + \sum_{i=1}^{n_q} (A_i^q)^T \mathbf{x}_i^q + (A^l)^T \mathbf{x}^l = b \\ & X_j^s \succeq 0, \quad \forall j \\ & \mathbf{x}_i^q \in L^{q_i}, \quad \forall i \\ & \mathbf{x}^l \in R_+^{n_l}, \end{aligned}$$

where C_j^s is a symmetric matrix, c_i^q is a q_i-dimensional vector, c^l is an n_l-dimensional vector, and the dimensions of matrices A and vector b are clear from the context.

Similarly to equation (8), the objective function of expression (5) can be rewritten as

$$\mathrm{tr}\{(B-\hat{B}^{LS})^T X^T X(B-\hat{B}^{LS})\} = \mathrm{tr}\{C^T C\}$$

up to a constant free of B where $C = \Lambda^{1/2}Q(B-\hat{B}^{LS})$ and $Q\Lambda Q^T$ is the eigenvalue decomposition of $X^T X$.
By the definition of the second-order cone, expression (5) can be equivalently written as

$$\min_{M,C,B} M \quad \text{such that} \quad \left[M, C_{11}, \ldots, C_{1q}, C_{21}, \ldots, C_{pq}\right] \in L_{pq+1},$$

$$\sum_{i=1}^{q} \sigma_i(B) \leq t,\ C = \Lambda^{1/2}Q(B-\hat{B}^{LS}).$$

Then, utilizing $\Sigma\sigma_i(B) = \Sigma\sigma_i(QB) \leq t$ is equivalent to

$$\sum_{i=1}^{min(p,q)} \mu_i(A) \leq t,$$

where $\mu_i(A)$ is the ith eigenvalue of A and

$$A = \begin{bmatrix} 0 & (QB)^T \\ QB & 0 \end{bmatrix}.$$

Together with formula (4.2.2) of Ben-Tal and Nemirovski (2001), page 147, this constraint is also equivalent to

$$qs + \mathrm{tr}(Z) \leq t,$$

$$Z - \begin{bmatrix} 0 & (\Lambda^{-1/2}C + Q\hat{B}^{LS})^T \\ \Lambda^{-1/2}C + Q\hat{B}^{LS} & 0 \end{bmatrix} + s\mathbb{I} \succeq 0,$$

$$Z \succeq 0.$$

Then, introducing variables to turn inequalities into equalities, expression (5) becomes equivalent to

$$\min_{M,C,s,Z_1,Z_2}(M) \quad \text{subject to}$$

$$q(s_1 - s_2) + \mathrm{tr}(Z_1) + s_3 = t,$$

$$Z_2 - Z_1 + \begin{pmatrix} 0 & (\Lambda^{-1/2}C)^T \\ (\Lambda^{-1/2}C) & 0 \end{pmatrix} - (s_1 - s_2)\mathbb{I} = \begin{pmatrix} 0 & -(Q\hat{B}^{LS})^T \\ -(Q\hat{B}^{LS}) & 0 \end{pmatrix},$$

$$Z_1, Z_2 \succeq 0,$$

$$(M, C_{11}, \ldots, C_{1q}, C_{21}, \ldots, C_{pq})^T \in L^{pq+1},$$

$$s \in \mathbb{R}^3_+.$$

This equivalent form is readily computable using SDPT3.

Thank you !

Zirui Pan
Haole Tian

School of Mathematics,
University of Science and Technology of China