Sparse Factor Analysis Via Post Processing

by

Ju-Young Park

DEPARTMENT OF STATISTICS

GRADUATE SCHOOL
HANKUK UNIVERSITY OF FOREIGN STUDIES

Sparse Factor Analysis Via Post Processing

by

Ju-Young Park

Under the Direction
of

Professor Seokho Lee

A dissertation submitted to the committee of the
Graduate School of Hankuk University of Foreign Studies
in partial fulfillment of the requirements for the degree of
Master of Science

February 2016

Approved by the committees of the Graduate School of Hankuk University of Foreign Studies in partial fulfillment of the requirements for the degree of Master of Science

Thesis Committee : ______________________ __________

______________________ __________

______________________ __________

DEPARTMENT OF STATISTICS

GRADUATE SCHOOL

HANKUK UNIVERSITY OF FOREIGN STUDIES

한국외국어대학교

HANKUK UNIVERSITY OF FOREIGN STUDIES

ABSTRACT

Sparse Factor Analysis Via Post Processing

Factor analysis is a well-known statistical method to express the multivariate data to variance-covariance matrix as coming from a limited number of latent factors. Each factor can be seen as a linear combination of independent variables and each factor loading means the level of its contribution on independent variables. However, the factor model can sometimes be difficult to interpret, because factors are given as linear combinations of all the original variables. For simple interpretation of factor model, an appropriate transformation is used in the complicated factor model. This simple factor structure is made by sparse factor loadings. The most popular representative of the sparse factor loadings method is the varimax rotation method. However, This rotation method fails to create enough sparse factor loadings. In this study, we propose a method and algorithm to make the sparse factor model by using the penalized optimization. The proposed method shows superiority over conventional methods

keyword : factor analysis, varimax rotation method, penalized optimization, principal component analysis.

Contents

Chapter 1 Introduction

Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors (Spearman, 1904). The observed variables, $X_1, X_2, ..., X_p$, are linearly related to a smaller number of unobservable factors, $F_1, F_2, ..., F_k$ $(k \le p)$, and factor loading means the coefficient of common factor and represents how much a factor explains a variable. However, If each variable is given as a linear combination of all the original common factors, it is difficult to interpret the result. In order to make the interpretation simple, a rotation method is generally preferred. Varimax, which was developed by Kaiser (1958), is one of the most popular rotation methods. The rotation method maximizes the sum of the variances of the squared loadings and find the rotation matrix that makes the factor model simpler. However, varimax rotation method cannot always yield sparse loadings factors. We introduce a new sparse factor analysis via post processing. We show that factor analysis can be formulated as a regression -type optimization problem; sparse factor loadings are then obtained by imposing the lasso constraint on the regression coefficients. Hirose and Yamamoto (2013) propose an estimation approach that produce sparse factor loadings in a conventional factor model. They introduce a penalized likelihood procedure that imposes a nonconvex penalty on the factor loadings. However, penalized likelihood method assumes that the underlying distribution for data generation is a normal distribution. If the data does not follow the normal distribution, this model is not appropriate. In this paper, we take an approach to factor analysis, which is similar to the sparse principal component analysis and cames up with sparse factor loadings. In this new approach, any distributional assumption on data is not required. In Section 2, we review the factor analysis method. We review the literature on sparse estimation method in Section 3, including varimax rotation and penalized likelihood method. In Section 4 we provide a new sparse factor analysis which is motivated by using sparse principal component analysis and penalized optimization. Also, we present an estimating algorithm for f itting the model we propose. Section 5 illustrates the performance of the proposed method using Monte Carlo simulations. We conclude the paper with some discussion in Section 6.

Chapter 2 Factor Analysis

Factor analysis is a statistical technique used for summarizing a large number of variables with a smaller number of factors that potentially reflect what sets of variables have in common with one another. Main purpose of the factor analysis is to reduce multivariate data to a smaller set of summary variables. So, with factor analysis, we are able to explore the covariance structure among a set of observable variables and see the underlying relationships between the variables.

2.1 Factor analysis model

Suppose we have a data set of p observed random variables, $x_1, x_2, \ldots, x_p$ with means $\mu_1, \mu_2, \ldots, \mu_p$ and variance-covariance matrix Σ. and these variables are assumed to be generated through the unobserved random variables F_j $(j \in 1, \ldots, k,\ k < p)$ with loading factors l_{ij} $(i \in 1, \ldots, p,\ j \in 1, \ldots, k)$.

The factor model is often expressed as follows:

$$
\begin{aligned}
X_1 &= \mu_1 + l_{11}F_1 + l_{12}F_2 + \ldots + l_{1k}F_k + \epsilon_1 \\
X_2 &= \mu_2 + l_{21}F_1 + l_{22}F_2 + \ldots + l_{2k}F_k + \epsilon_2 \\
&\vdots \\
X_p &= \mu_p + l_{p1}F_1 + l_{p2}F_2 + \ldots + l_{pk}F_k + \epsilon_p
\end{aligned}
$$

μ_i : mean X_i.

l_{ij} : ith variable and jth factor loading.

F_i : ith common factor.

ϵ_i : ith specific factor.

With matrix notations, the set of above equations becomes

$$X-\mu = LF+\epsilon \quad (2.1)$$

If we have n observations, then the dimensions of the above matrices become $X_{p\times n,}$ $L_{p\times k}$, and $F_{k\times n}$. Here, F represents the factors, and L represents the loadings matrix.

In factor model, the below assumptions are often imposed.

Assumption 1. $E(F) = 0_{k\times 1}$, i.e., mean of the common factor is zero.

Assumption 2. $Cov(F) = E(FF^T) = I_{k\times k}$

Assumption 3. $E(\epsilon) = 0_{p\times 1}$, i.e., mean of the specific factor is zero.

Assumption 4. $Cov(\epsilon) = E(\epsilon\epsilon^T) = \Psi_{p\times p} = \begin{pmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{pmatrix}$, i.e., specific factors have no correlation among them and variance of the ith common factor is given as ψ_i.

Assumption 5. $Cov(F, \epsilon) = 0$, i.e., F and ϵ are independent.

These assumptions lender *orthogonal common factor model*. Under a orthogonal common factor model, covariance matrix can be decomposed into a factor loading matrix and a specific matrix (error covariance matrix).

$$\Sigma = Cov(X) = E(X-\mu)(X-\mu)^T = LL^T + \Psi \quad (2.2)$$

2.2 Factor loading estimation method

2.2.1 Maximum likelihood method

In order to use maximum likelihood estimation, the data is assumed to be independently sampled from a multivariate normal distribution with mean vector μ and variance-covariance matrix $\Sigma = LL^T + \Psi$. Suppose that we have a random sample of n observations from $N_p(\mu, LL^T+\Psi)$. Then the log-likelihood function is

$$l(\mu,L,\Psi) = -\frac{np}{2}log2\pi - \frac{n}{2}log|LL^T+\Psi| - \frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)^T(LL^T+\Psi)(X_i-\mu). \quad (2.3)$$

The maximum likelihood estimates, $\hat{L}_{ML}$ and $\hat{\Psi}_{ML}$, are obtained by maximizing (2.3) and, equivalently, solving the below estimating equations:

$$\partial l(\mu,L,\Psi)/\partial L = 0, \quad \partial l(\mu,L,\Psi)/\partial \Psi = 0.$$

2.2.2 Principal factor analysis method

Principal factor analysis concentrates on the decomposition of the variance-covariance matrix. Recall that the variance-covariance matrix can be re-expr essed in the following form:

$$\Sigma = LL^T + \Psi. \quad (2.4)$$

The variance-covariance matrix in (2.4) can be expressed as the sum of p eigenvalues λ_i multiplied by the corresponding eigenvectors e_j.

$$\Sigma_{P\times P} = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T + ... + \lambda_p e_p e_p^T. \quad (2.5)$$

If we ignore the components having small eigenvalues, the variance-covariance matrix is approximated by a set of major principal components. With $p > q$,

$$\begin{aligned}\Sigma_{P\times P} &\approx \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T + ... + \lambda_q e_q e_q^T \\ &= \left(\sqrt{\lambda_1}\, e_1, \sqrt{\lambda_2}\, e_2, ..., \sqrt{\lambda_q}\, e_q\right)\begin{pmatrix}\sqrt{\lambda_1}\, e_1 \\ \sqrt{\lambda_2}\, e_2 \\ \vdots \\ \sqrt{\lambda_q}\, e_q\end{pmatrix} = LL^T \qquad (2.6)\end{aligned}$$

This yields the following estimator for the factor loadings :

$$\hat{l}_{ij} = \hat{e}_{ij}\sqrt{\hat{\lambda}_i}. \quad (2.7)$$

The specific variance Ψ is now obtained by subtracting LL^T from the variance-covariance matrix:

$$\Psi = \Sigma - LL^T. \quad (2.8)$$

The matrix Ψ is a diagonal matrix, so that we generally take the diagonal element of (2.8) in practice. This can be estimated using the following expression:

$$\hat{\psi}_i = s_i^2 - \sum_{i=1}^{m} \lambda_i \hat{e}_{ij}^2$$

Chapter 3 Sparse Estimation Method

3.1 Varimax rotation method

A rotation method is widely used to get simple factor structure. Varimax method is the most popular rotation method (Kaiser, 1958). It finds an orthogonal rotation of data which maximizes the sum of the variances of the squared factor loadings. The resulting factor loadings are often easier to interpret than the conventional factor loadings (this is the reason this method is called varimax), and the variance-covariance structure is unchanged because the orthogonal rotation is used.

Suppose that the initial loading matrix is $\hat{L}$:

$$\hat{L} = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1m} \\ l_{21} & l_{22} & \cdots & l_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \cdots & l_{pm} \end{pmatrix} \qquad (3.1)$$

and the rotated loading matrix is $\hat{L}^*$

$$\hat{L}^* = \hat{L}T = \begin{pmatrix} l_{11}^* & l_{12}^* & \dots & l_{1m}^* \\ l_{21}^* & l_{22}^* & \cdots & l_{2m}^* \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1}^* & l_{p2}^* & \cdots & l_{pm}^* \end{pmatrix}. \quad (3.2)$$

The varimax procedure, as defined below, selects a rotation matrix, T, to achieve the maximum of the below quantity :

$$\max \sum_{j=1}^{m} \left(\frac{1}{p} \sum_{i=1}^{p} (\widehat{l_{ij}^{*2}})^2 - \frac{1}{p} (\sum_{i=1}^{p} \widehat{l_{ij}^{*2}})^2 \right). \quad (3.3)$$

Factor loadings from varimax rotation are often constituted by only a few variables with very high loadings on this factor while the remaining variables have near-zero loadings on this factor. On this account, the varimax rotation brings the loading matrix closer to such simple structure. However, varimax rotation method cannot always have a perfect simple structure and impossib le to make factor loadings to zero exactly.

3.2 Penalized likelihood factor analysis via nonconvex penalties.

Hirosi and Yamamoto (2013) propose an sparse factor analysis method using penalized likelihood maximization approach. They introduce a penalized likelihood procedure that imposes a nonconvex penalty on the factor loadings. A penalized likelihood method maximizes the following penalized log-likelihood function $\ell_\rho(L,\Psi)$:

$$\ell_\rho(L,\Psi) = \ell(L,\Psi) - N\sum_{i=1}^{p}\sum_{j=1}^{m}\rho P(|l_{ij}|), \quad (3.4)$$

where $\rho > 0$ is a regularization parameter and $P(.)$ can be viewed as a penalty function. The regularization parameter ρ controls the amount of shrinkage; the larger the value of ρ, the greater the amount of shrinkage. So this method achieves sparse factor loadings under penalized likelihood procedure. Hirose and Yamamoto (2013) show that the result of sparse loadings superior than varimax rotation.

Chapter 4 Sparse Factor Analysis (SFA)

Hirose and Yamamoto (2013) assumes the strong distribution assumption that the data follows normal distribution. If the normal low is violated, this method does not guarantee a good sparse estimation. To overcome this drawback, we propose a new sparse factor analysis using post processing which is able to produce sparser solutions even though the data is out of normal low.

Our new method, first, conducts a principal factor analysis to yield an initial factor loadings, and then the factor loadings from principal factor method is put into a regularization procedure in order to achieve sparse loadings. Since we use a principal factor method and the regularization procedure is free from the distributional assumption, the normal distribution assumption is not required at all. Therefore, we expect a good sparsity from our method regardless of the distribution which the data actually follows.

4.1 Principal component method in factor analysis

Let the data, X, have p-dimensional independent variables with mean vector $\mu = (\mu_1, ..., \mu_p)$ and variance-covariance matrix Σ. From (2.5) and using the equivalence between the singular value decomposition (SVD) of the data matrix and the eigen structure of the variance-covariance matrix, X is represented as

$$X = ZL^T + E, \quad (4.1)$$

where E is a matrix containing all errors. If we ignore the specific factor in (4.1), the data is approximated by a low rank matrix (Eckart and Young, 1936). For an integer $k \le p$, a matrix $X^{(k)}$, which is called the k-rank approximation of X in the sense that $X \approx X^{(k)}$, is defined as

$$X^{(k)} = Z_1 L_1^T + Z_2 L_2^T + \ldots + Z_k L_k^T, \quad (4.2)$$

where Z_l, L_l $(l = 1, 2, \ldots, k)$ are the lth factor score and factor loading, respectively, matched with the lth singluar value and left singluar vector. In (4.2), the integer k, the rank of matrx $X^{(k)}$, is usually determined by the amount of variation for the data by approximation.

4.2 Factor loadings under penalization

Factor loadings from principal factor method are not sufficient sparse in most cases. So, in this section, we study a sparse procedure for factor loading. To this end, we provide a sparse factor loadings method under penalized optimization.

4.2.1 Sparse factor loadings of procedure

The factor loadings from principal factor method can be expressed as the low rank approximation of the data matrix. Here, the best rank-one approximation with the fist factor can be described as the below residual matrix:

$$X_1 = X - \sum_{l=2}^{k} X_l, \quad (4.3)$$

with $X_l = Z_l L_l^T$ for $l = 1, 2, ..., k$. Then, (4.3) formular can be represented as the following :

$$X_1 = X - \sum_{l=2}^{k} X_l = Z_1 L_1^T. \quad (4.4)$$

In the above, L_1 and Z_1 are

$$L_1^* = L_1/|L_1|, \quad Z_1 = X_1 L_1^* (L_1^{*T} L_1^*)^{-1}, \qquad (4.5)$$

where L_1 is the first principal factor loading from the matrix $X - Z_{(-1)} L_{(-1)}^T$, and $Z_{(-1)}$ and $L_{(-1)}$ are first factor score and loadings for a new data matrix which is subtracted matrix X_1. Now, in order to get a sparse L_1 solution, we use lasso regression framework:

$$\widehat{L_1} = \arg min_{L_1} \|X_1 - Z_1 L_1\|_F^2 + \lambda \|L_1\|. \qquad (4.6)$$

In the (4.6), the sparse factor loadings is considered as the coefficient vector of lasso regression without an intercept. The tuning parameter λ controls the amount of shrinkage; the larger the value λ, the further the factor loadings towards zero. The same scheme is repeated through $l = 2, ..., k$: the second sparse factor analysis can calculate lasso regression method with the newly updated residual matrix. the residual matrix is given as

$$X_2 = X - (X_1 + X_3 + .. + X_k) \qquad (4.7)$$

and second sparse factor loading is obtained, in the similar way, by

$$\widehat{L_2} = \mathrm{arg}min_{L_2} \|X_2 - Z_2 L_2\|_F^2 + \lambda \|L_2\|. \quad (4.8)$$

In this sparse factor method, we sequentially produce the sparse factor loadings. The advantage of this method is each sparse factor loading can be calculated without being affected by other factor loadings and this method is faster than method of calculate at once sparse factor loadings.

4.3 Algorithm

As described earlier, sparse factor loading vectors are obtained one by one in the sequel. In this section, we generalize this and introduces sparse estimating algorithm for the ith factor loading.

Algorithm 1 K-fold CV Tuning Parameter Selection

Step1. Initialize : Apply the principal factor analysis to X and obtain the loadings and factor scores.

(a) $X_i = X - Z_{(-i)} L_{(-i)}^T$

(b) Z_i Given L_i : $Z_i = X_i L_i^* (L_i^{*T} L_i^*)^{-1}$ where $L_i^* = L_i / \|L_i\|$

Step2 . Update :

(a) $L_i^{new} = \mathrm{arg}min_{L_i^*} \|X_i - Z_i L_i^*\|_F^2 + \lambda \|L_i^*\|$

(b) $\hat{L}_i = L_i^{new} / \|L_i^{new}\|, \quad Z_i^{new} = X_i \hat{L}_i (\widehat{L_i^T} \hat{L}_i)^{-1}$

Step3. Repeat Step2. replacing Z_i and L_i by Z_i^{new} and L_i^{new} until convergence.

Description of Algorithm 1 is as follows. In Step1, the initial value for the i th factor loading is obtained from principal factor method and this factor loading is normalized. And then, the corresponding initial factor score is calculated by projection on the normalized factor loading. In Step2, we used the penalized optimization for spare factor loadings. We propose to use lasso penalty. From this penalized optimization, we obtain a sparse factor loading solution and then normalize it to be a unity in length. The factor score is updated by the projection of the residual data on the normalized factor loading. This algorithm is repeated until convergence.

4.4 Tuning parameter λ selection

Amount of the sparsity in factor loading is determined according to the choice of tuning parameter λ value. In this section, we propose to use k-fold cross validation technique to choose the optimal λ.

Algorithm 2 K-fold CV Tuning Parameter Selection

Step1. Randomly divide the rows of X into K roughly equal-sized groups, denoted by $X_{1,} X_2, \ldots, X_k$;

Step2. For each tuning parameter λ_j $j \in \{0,1,\ldots,r\}$, do the following :

(a) For $k=1,2,\ldots,K$, let X^{-k} be the data matrix X leaving out X^k.

Apply Algorithm 1 on X^{-k} to drive the factor loading $L^{-k}(\lambda_j)$.

Then factor score Z^k obtain the project X^k onto $L^{-k}(\lambda_j)$ as $Z^k(\lambda_j) = X^k L^{-k}(\lambda_j)$;

(b) Calculate the K-fold CV estimate defined as

$$CV(\lambda_j) = \sum_{k=1}^{K} \frac{\sum_{i=1}^{n_k} \sum_{l=1}^{p} \left\{ x_{il}^k - l_i^k(\lambda_j) z_l^{-k}(\lambda_j) \right\}^2}{n_k r}$$

where n_k is the number of observations in X^k, and l_i^k and z_l^{-k} are respectively the ith and lth elements of l^k and z^{-k} ;

Step3. Select the degree of sparsity as $\lambda_0 = \arg min_j CV(\lambda_j)$

In practice, K is usually chosen to be 5 or 10 for computational efficiency. We use $K=5$ in our simulation studies.

Chapter 5 Simulation

In the simulation study, we used the following true factor loadings (Hirosi and Yamamoto, 2013):

$$L^{True} = \begin{pmatrix} 0.95 & 0.90 & 0.85 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.80 & 0.75 & 0.70 \end{pmatrix}^T. \quad (5.1)$$

We generated the simulation data that possess the true factor loadings, i.e,

$$X = ZL^T + E \quad (5.2)$$

The data sets were generated over various distributions for Z and E: the combinations of normal distribution , t-distribution and Gamma distribution. We consider 6 scenarios.

Scenario 1 Hirosi and Yamamoto (2013) :

$$X_{n\times 6} \sim N_6(0, LL^T + \Psi) \quad \text{where} \quad \Psi \sim diag(I - LL^T)$$

The Scenario 1 data generated Hirose and Yamamoto paper simulation method.

Scenario 2 $N(0,1)$ for factor scores and $N(0,1)$ for random noises

$$X_{n\times 6} = ZL^T + N_6(0, 0.1) \quad \text{where} \quad Z \sim N_2(0,1)$$

Scenario 3 $N(0,1)$ for factor scores and $t(df)$ for random noises

$$X_{n\times 6} = ZL^T + t(df) \quad \text{where} \quad Z \sim N_2(0,1)$$

Scenario 4 $t(df)$ for factor scores and $N(0,1)$ for random noises

$$X_{n\times 6} = ZL^T + N(0,0.1) \quad \text{where} \quad Z \sim t(df)$$

Scenario 5 $t(df)$ for factor scores and $t(df)$ for random noises

$$X_{n\times 6} = ZL^T + t(df) \quad \text{where} \quad Z \sim t(df)$$

Scenario 6 $N(0,1)$ for factor scores and $Gamma(\alpha,\beta)$ for random noises

$$X_{n\times 6} = \Gamma(\alpha,\beta) \quad \text{where} \quad Z \sim N_2(0,1)$$

We used the degree of freedom 3 for t-distribution. For the random error having t-distribution, we scaled them to have 0.1 standard deviation. For gamma distribution in Scenrio 6, parameter α and β were set to be the mean ZL^T and 0.1 standard deviation. We consider the sample size $N = 50, 100, 200$ and 400 in the simulation study.

5.1 Evaluating performance

We used mean squared error (MSE) and true positive/negative rate for evaluating methods we considered. MSE measures the closeness between the true factor loadings and their estimates. And true positive/negative rate indicates how well true nonzero/zero loadings the estimated factor loadings represent the true ones.

5.1.1 MSE_{Λ} (mean squared error)

MSE is calculated based on the squared Frobenius norm of the difference of matrices L^{true} and $\hat{L}$:

$$MSE_L = \frac{1}{1000pm} \sum_{s=1}^{1000} \| L - \hat{L}^{(s)} \|^2$$

5.1.2 True Positive Rate and True Negative Rate

True positive/negative rate (TPR/TNR respectively) is defined as follows:

TPR = number of estimated positive loadings / number of true positive loadings.

TNR = number of estimated negative loadings / number of true negative loadings.

5.2 Result

We summarize MSE, TPR, and TNR in Tables 1~6 over all scenario. In these tables, we used abbreviated terms, HY for Hirosi and Yamamoto (2013), SFA for sparse factor analysis. Depending on the initial values for loading matrix, SFA method is denoted differently. SFA vari stands for SFA method when the initial factor loadings are rotated by varimax method, and SFA w/o stands for SFA method when the initial factor loadings are not rotated.

Table1. Scenario 1 data is constructed normal distribution.
In this data, MSE, TPR and TNR results for each models show below table1.

		$Varimax$	HY	$SFA\ w/o$	$SFA\ vari$
		$N=50$			
MSE_L		0.046	0.054	0.21	0.21
TPR	$loadings\,1$	1.00	0.75	0.97	0.99
	$loadings\,2$	1.00	0.75	0.97	0.98
TNR	$loadings\,1$	0.00	0.50	0.65	0.67
	$loadings\,2$	0.00	0.25	0.36	0.36
		$N=100$			
MSE_L		0.022	0.024	0.162	0.160
TPR	$loadings\,1$	1.00	0.70	0.97	1.00
	$loadings\,2$	1.00	0.72	0.97	1.00
TNR	$loadings\,1$	0.00	0.26	0.65	0.71
	$loadings\,2$	0.00	0.25	0.34	0.43
		$N=200$			
MSE_L		0.012	0.011	0.113	0.111
TPR	$loadings\,1$	1.00	0.73	1.00	1.00
	$loadings\,2$	1.00	0.71	1.00	1.00
TNR	$loadings\,1$	0.00	0.31	0.73	0.74
	$loadings\,2$	0.00	0.29	0.51	0.51
		$N=400$			
MSE_L		0.007	0.005	0.077	0.076
TPR	$loadings\,1$	1.00	0.73	1.00	1.00
	$loadings\,2$	1.00	0.71	1.00	1.00
TNR	$loadings\,1$	0.00	0.31	0.76	0.74
	$loadings\,2$	0.00	0.29	0.56	0.56

Table2. Scenario 2 data is constructed normal-distribution. This is same situation with scenario 1. Specific factor Ψ and factor score Z of data are calculated normal distribution. In this data, MSE, TPR and TNR results for each models show below table2.

		Varimax	*HY*	*SFA w/o*	*SFA vari*
		$N=50$			
MSE_L		0.162	0.003	0.007	0.011
TPR	*loadings* 1	1.00	0.81	1.00	1.00
	loadings 2	1.00	0.89	1.00	1.00
TNR	*loadings* 1	0.00	0.11	0.79	0.77
	loadings 2	0.00	0.17	0.64	0.63
		$N=100$			
MSE_L		0.261	0.002	0.003	0.003
TPR	*loadings* 1	1.00	0.81	1.00	1.00
	loadings 2	1.00	0.87	1.00	1.00
TNR	*loadings* 1	0.00	0.12	0.78	0.76
	loadings 2	0.00	0.17	0.61	0.61
		$N=200$			
MSE_L		0.378	0.001	0.002	0.002
TPR	*loadings* 1	1.00	0.82	1.00	1.00
	loadings 2	1.00	0.87	1.00	1.00
TNR	*loadings* 1	0.00	0.12	0.79	0.78
	loadings 2	0.00	0.18	0.63	0.62
		$N=400$			
MSE_L		0.486	0.001	0.001	0.001
TPR	*loadings* 1	1.00	0.82	1.00	1.00
	loadings 2	1.00	0.87	1.00	1.00
TNR	*loadings* 1	0.00	0.12	0.77	0.76
	loadings 2	0.00	0.19	0.62	0.63

Table3. Scenario 3 data is constructed t-distribution. Specifically, specific factor Ψ of data is t-distribution with degree 3. On the other hand, Factor scores of data Z is standard normal distribution. In this data, MSE, TPR and TNR results for each models show below table3.

		$Varimax$	HY	$SFA\ w/o$	$SFA\ vari$
$N=50$					
MSE_L		0.303	0.003	0.007	0.011
TPR	$loadings\,1$	1.00	0.83	1.00	1.00
	$loadings\,2$	1.00	0.90	1.00	1.00
TNR	$loadings\,1$	0.00	0.11	0.78	0.76
	$loadings\,2$	0.00	0.19	0.64	0.59
$N=100$					
MSE_L		0.434	0.002	0.003	0.003
TPR	$loadings\,1$	1.00	0.83	1.00	1.00
	$loadings\,2$	1.00	0.90	1.00	1.00
TNR	$loadings\,1$	0.00	0.11	0.79	0.78
	$loadings\,2$	0.00	0.18	0.64	0.62
$N=200$					
MSE_L		0.5627	0.001	0.002	0.002
TPR	$loadings\,1$	1.00	0.82	1.00	1.00
	$loadings\,2$	1.00	0.88	1.00	1.00
TNR	$loadings\,1$	0.00	0.11	0.78	0.78
	$loadings\,2$	0.00	0.18	0.64	0.64
$N=400$					
MSE_L		0.667	0.001	0.001	0.001
TPR	$loadings\,1$	1.00	0.82	1.00	1.00
	$loadings\,2$	1.00	0.86	1.00	1.00
TNR	$loadings\,1$	0.00	0.14	0.79	0.77
	$loadings\,2$	0.00	0.17	0.64	0.62

Table4. Scenario 4 data is composed t-distribution. On the contrary to table2, Specific factor Ψ of data is standard normal distribution and Factor scores of data Z is t-distribution with degree 3. In this data, MSE, TPR and TNR results for each models show below table4.

		$Varimax$	HY	$SFA\ w/o$	$SFA\ vari$
		$N=50$			
MSE_L		0.067	0.005	0.014	0.017
TPR	$loadings\,1$	1.00	0.82	0.99	0.99
	$loadings\,2$	1.00	0.87	1.00	0.99
TNR	$loadings\,1$	0.00	0.13	0.77	0.77
	$loadings\,2$	0.00	0.20	0.60	0.62
		$N=100$			
MSE_L		0.123	0.002	0.007	0.004
TPR	$loadings\,1$	1.00	0.80	1.00	0.99
	$loadings\,2$	1.00	0.86	1.00	0.99
TNR	$loadings\,1$	0.00	0.14	0.79	0.79
	$loadings\,2$	0.00	0.18	0.66	0.65
		$N=200$			
MSE_L		0.198	0.001	0.003	0.002
TPR	$loadings\,1$	1.00	0.82	1.00	1.00
	$loadings\,2$	1.00	0.87	1.00	1.00
TNR	$loadings\,1$	0.00	0.15	0.83	0.82
	$loadings\,2$	0.00	0.20	0.67	0.66
		$N=400$			
MSE_L		0.276	0.001	0.001	0.001
TPR	$loadings\,1$	1.00	0.81	1.00	1.00
	$loadings\,2$	1.00	0.86	1.00	1.00
TNR	$loadings\,1$	0.00	0.14	0.78	0.80
	$loadings\,2$	0.00	0.19	0.65	0.66

Table5. Scenario 5 data is t-distribution. In the scenario 5 data, Both factor Ψ and factor scores Z of data are t-distribution with degree 3. In this data, MSE, TPR and TNR results for each models show below table5.

		$Varimax$	HY	$SFA\ w/o$	$SFA\ vari$
		$N=50$			
MSE_L		0.160	0.005	0.014	0.015
TPR	$loadings\,1$	1.00	0.82	0.99	0.99
	$loadings\,2$	1.00	0.89	1.00	0.99
TNR	$loadings\,1$	0.00	0.12	0.78	0.77
	$loadings\,2$	0.00	0.19	0.62	0.62
		$N=100$			
MSE_L		0.260	0.002	0.001	0.001
TPR	$loadings\,1$	1.00	0.81	0.99	1.00
	$loadings\,2$	1.00	0.87	1.00	1.00
TNR	$loadings\,1$	0.00	0.13	0.79	0.79
	$loadings\,2$	0.00	0.19	0.64	0.65
		$N=200$			
MSE_L		0.014	0.006	0.003	0.001
TPR	$loadings\,1$	1.00	0.82	1.00	1.00
	$loadings\,2$	1.00	0.86	1.00	1.00
TNR	$loadings\,1$	0.00	0.14	0.80	0.82
	$loadings\,2$	0.00	0.19	0.66	0.68
		$N=400$			
MSE_L		0.473	0.001	0.001	0.001
TPR	$loadings\,1$	1.00	0.82	1.00	1.00
	$loadings\,2$	1.00	0.86	1.00	1.00
TNR	$loadings\,1$	0.00	0.15	0.80	0.80
	$loadings\,2$	0.00	0.20	0.63	0.66

Table6. Scenario 5 data menas *Gamma*-distribution. This data is constructed that factor Ψ of data is *Gamma*-distribution and The parameters α,β are three. In this data, MSE, TPR and TNR results for each models show below table6.

		$Varimax$	HY	$SFA\ w/o$	$SFA\ vari$
$N=50$					
MSE_L		0.157	0.004	0.007	0.031
TPR	$loadings\,1$	1.00	0.81	1.00	1.00
	$loadings\,2$	1.00	0.89	1.00	1.00
TNR	$loadings\,1$	0.00	0.11	0.78	0.78
	$loadings\,2$	0.00	0.17	0.62	0.62
$N=100$					
MSE_L		0.262	0.002	0.003	0.003
TPR	$loadings\,1$	1.00	0.80	1.00	1.00
	$loadings\,2$	1.00	0.87	1.00	1.00
TNR	$loadings\,1$	0.00	0.12	0.78	0.78
	$loadings\,2$	0.00	0.17	0.62	0.63
$N=200$					
MSE_L		0.379	0.001	0.002	0.002
TPR	$loadings\,1$	1.00	0.82	1.00	1.00
	$loadings\,2$	1.00	0.87	1.00	1.00
TNR	$loadings\,1$	0.00	0.13	0.77	0.79
	$loadings\,2$	0.00	0.18	0.64	0.64
$N=400$					
MSE_L		0.484	0.001	0.001	0.001
TPR	$loadings\,1$	1.00	0.82	1.00	1.00
	$loadings\,2$	1.00	0.88	1.00	1.00
TNR	$loadings\,1$	0.00	0.14	0.79	0.77
	$loadings\,2$	0.00	0.19	0.66	0.63

In Tables 1 and 2 of the case that factor scores and random errors come from the normal distribution, HY show the best performance in MSE sense. SFA methods (w/o and vari), on the other hand, is the best in TPR and TNR perspective. In the other words, SFA w/o and vari methods are the most excellent performer in the sparse solutions. Tables 3~6 summarize the result in case of non-normal distributions. In these cases, although HY shows the best result in MSE, the difference between HY and SFA becomes marginal as the sample size increases. As in the normal case, SFA outperforms HY in terms of TPR and TNR.

Chapter 6 Conclusion

In the study, we propose a new sparse factor analysis method by post-processing the principal factor loadings under a sparse procedure and provide its effective algorithm. This new method can be a good alternative to rotation method, including varimax, which is not able to create truly sparse solution. We demonstrated the performance of our method using Monte Carlo simulation studies. In simulation studies, we compared our methods with the existing sparse factor analysis methodology. We observed that our method is very competitive to the existing method in terms of reconstruction errors (measured by MSE), especially when the distribution is not normal. However, our method always outperforms the existing method when the signal detection is of most concern (measured by TPR and TNR). This empirical evidence demonstrates that our method is the best when non-normal distribution is involved.

Reference

Hirosi, K. and Yamamoto, M. (2013) Sparse estimation via nonconcave penalized likelihood in factor analysis model, *Statistics and Computing*, **25**, 863-875.

Kaiser, H. F. (1958) The varimax criterion for analytic rotation in factor analysis, *Psychometrika*, **23**, 187-200.

국문초록

사후프로세싱을 통한 희소인자분석

인자분석(factor analysis)은 다변량자료를 공분산 구조를 이용하여 관측되지 않는 몇개의 인자로 표현하는 방법이다. 각 인자들은 자료의 변수들의 선형결합으로 표현되며, 각 인자들의 계수인 인자적재값(factor loadings)은 인자에 대한 변수의 기여도를 의미한다. 그러나 인자들의 수가 많을 경우 어떤 인자가 변수에 기여하는지 파악하기 쉽지 않다. 용이한 해석을 위해 적절한 변환을 통하여 단순한 구조를 만들 수 있다. 이러한 단순한 구조는 인자적재값을 희소(sparse)하게 만들어서 얻을 수 있다. 그 중 가장 대표적으로 사용되는 방법으로는 인자의 축을 회전 시키는 회전변환 방법 중 베리멕스 회전(varimax rotation)방법이 있다. 그러나 이러한 베리멕스 회전변환은 인자적재값들을 충분히 희소하게 만들지 못한다. 본 연구에서는 벌점최적화(penalized optimization)를 이용하여 인자적재값들을 희소하게 만드는 방법 및 알고리즘을 제안한다. 제안한 방법의 우수성을 컴퓨터 모의실험 상에서 기존의 방법과 비교를 통해 확인하였다.

주요용어 : 인자분석, 베리멕스 회전변환, 벌점최적화, 주성분분석