Regularized Common Factor Analysis

Sunho Jung[1] and **Yoshio Takane**[1]
(1) Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC, H3A 1B1, Canada

Abstract

In common factor analysis, estimation of unique variances by unweighted least squares (ULS), generalized least squares (GLS), and maximum likelihood (ML) often leads to improper solutions. To deal with this problem, a new estimation method of the unique variances was proposed by applying the idea of generalized ridge regularization. In regularized common factor analysis (RCFA), unique variances are estimated under three alternative assumptions: 1) unique variances are constant across variables (i.e., proportional to the unit variance), 2) unique variances are proportional to variances of anti-image variables, and 3) unique variances are proportional to Ihara-Kano's non-iterative estimates of unique variances. The constant of proportionality (i.e., the regularization parameter λ) is then estimated by ULS, GLS, or ML. Illustrative examples consisting of Monte Carlo studies and a real data set were given to demonstrate the usefulness of the proposed method.

1. Introduction

Measurement errors are almost ubiquitous in any kinds of measurement in psychological research. In contrast to principal component analysis (PCA), common factor analysis (CFA) explicitly takes into account measurement errors in observed variables. However, to do so requires estimation of variances of measurement errors, and traditionally error variances are equated to unique variances, the variabilities in observed variables which are not shared by other variables. Three methods of estimation, unweighted least squares (ULS), generalized least squares (GLS), and maximum likelihood (ML), are commonly used in estimating unique variances. However, they often yield improper solutions (negative or boundary estimates of unique variances). To avoid the problem, a new estimation method for the unique variances is proposed by applying the idea of generalized ridge regularization.

In ridge regression analysis (Hoerl & Kennard, 1970), estimates of regression coefficients are obtained by adding a small positive value to the variances of predictor variables. This leads to estimates of regression coefficients, which are closer to zero and are on average closer to the true population parameters than the least squares estimates. Presumably, similar shrinkage effects can be obtained by subtracting some positive values from the variances of criterion variables. Since the observed variables are typically regarded as the criterion variables in the CFA model, subtracting unique variances from the diagonals of observed covariance or correlation matrices is expected to have a similar shrinkage effect on estimates of factor loading matrices.

To illustrate, let us look at Tables 1 and 2. They were obtained by applying PCA and ML CFA to 13 variables measuring organizational stereotype from Bergami and Bagozzi's (2000) social identity data. The sample size was 291. Permutation tests indicated two significant components. The varimax rotated component loadings (Table 1) are almost consistently larger than the corresponding common factor loadings (Table 2). The latter are shrunken toward zero, indicating the

Table 1. PCA (Var. = Variable, Comp. Cont. = Component contributions)

Var.	Comp. loadings	
	1	2
1	-.667	-.308
2	-.219	-.711
3	-.196	-.686
4	-.703	-.112
5	-.067	-.625
6	-.481	-.448
7	-.091	-.649
8	-.489	-.424
9	-.757	-.184
10	-.784	-.212
11	-.777	-.098
12	-.746	-.104
13	-.755	-.206
Comp. Cont.	4.43	2.42

Table 2. Maximum likelihood CFA (Var. = Variable, Fac. Cont. = Factor contributions)

Var.	Factor loadings	
	1	2
1	-.606	-.346
2	-.136	-.725
3	-.172	-.589
4	-.612	-.214
5	-.148	-.366
6	-.374	-.506
7	-.159	-.402
8	-.390	-.474
9	-.737	-.211
10	-758	-.262
11	-.739	-.170
12	-.628	-.249
13	-.649	-.339
Fac. Cont.	3.61	2.13

possibility that they are closer to the population values than the former. The factor (component) contributions (sums of squares of loadings) at the bottom of the tables indicate the variances accounted for by respective factors (components), which also indicate the shrunken nature of the CFA loadings relative to the PCA loadings.

Common factor analysis can be viewed as a kind of shrinkage estimation. There is a good theoretical reason for this (Yanai & Takane, 2007). Let X denote an n-subjects by p-variables data matrix, and let $M^{(L)}(\lambda) = I_n + \lambda(XL^{-1}X')^{+}$ represent a generalized ridge metric matrix (Takane & Yanai, 2006), where I_n is the identity matrix of order n, λ is a regularization parameter, L is a *nnd* (non-negative definite) matrix such that $\text{Sp}(L) = \text{Sp}(X')$ (where Sp indicates a range space), and the superscript $^{+}$ indicates the Moore-Penrose inverse. Then, $X'M^{(L)}(\lambda)X = X'X + \lambda L$. When L is set equal to Ψ, the diagonal matrix of unique variances (assuming temporarily that it is known), and the regularization parameter λ is set equal to -1, $X'M^{(\Psi)}(-1)X = C-\Psi$ or $R-\Psi$, where C and R are sample covariance and correlation matrices, respectively, depending on whether X is only columnwise centered or standardized. (In both cases, we assumed that X was further normalized so that $X'X = C$ or $X'X = R$ without dividing $X'X$ by n.) This is the matrix subjected to the eigen analysis to derive a factor loading matrix in CFA (or more precisely, in a special kind of CFA called principal FA). More specifically, eigen value decomposition (EVD) of $R - \Psi$ is obtained in LS FA, and generalized EVD (GEVD) of $C - \Psi$ with respect to Ψ is obtained in

Table 3. Frequencies of improper solutions (out of 100).

Sample size	ULS	GLS	ML
50	39	73	65
100	32	51	54
200	19	34	33
500	10	16	14
1000	4	4	5
50	11	27	22
100	5	6	5
200	0	0	0
500	0	0	0
1000	0	0	0
50	24	37	35
100	16	25	25
200	13	16	22
500	1	1	3
1000	1	1	1

Table 4. A population common factor loadings (F1 and F2) and unique variances (Ψ)

Var.	F1	F2	Ψ
1	.044	.899	.191
2	.807	.061	.345
3	.472	.063	.773
4	.385	.191	.815
5	.652	.207	.533
6	.870	.236	.187
7	.818	.188	.297
8	.635	.193	.559
9	.399	.737	.298
10	.817	.186	.299

GLS and ML FA, while EVD of R is obtained in PCA. Thus, the common factor loadings tend to be shrunk due to a negative value of λ. (Indeed, in GLS and ML FA, generalized eigenvalues of C with respect to Ψ are decremented by 1 to obtain the factor loading matrix.)

In practice, however, Ψ is unknown and is to be estimated in some way. Various methods (e.g., ULS, GSL, and ML) have been developed to estimate this quantity along with the matrix of factor loadings. Most of these methods are iterative, although several non-iterative methods are also available (Hägglund, 1982; Ihara & Kano, 1986; Jennrich, 1987; Kano, 1990). As has been alluded to earlier, these procedures are often susceptible to improper solutions. Furthermore, these procedures can only be asymptotically justified, meaning that they are optimal or nearly optimal for large samples, but are not necessarily so for small samples.

To see how serious the problem of improper solutions in CFA is, let us look at Table 3. This table consists of three parts. The top subtable reports frequencies of improper solutions obtained by applying three estimation methods, ULS, GLS, and ML to 100 data sets generated from a population loadings and unique variances postulated in Table 4. Note that there are only sampling errors in the generated data sets. Solutions are judged as improper as one of unique variance estimates gets smaller than .001 (virtually zero). There are so many cases of improper solutions (these are out of 100); more than 30% of the solutions by GLS and ML are improper even for the sample size of $n = 200$. This number is much less in ULS, but still 19% of the solutions are improper. These frequencies decrease as the sample size increases, but improper solutions can still be obtained for the sample size as large as 1000. We conjectured that this could be due to the peculiarity of the population factor structure. In particular, the second factor is relatively weak. So we modified the population factor structure by adding .5 to the second factor loading of variable 4 in Table 4. (This modified population factor loadings will be again used for a Monte Carlo study.) The middle portions of Table 3 show frequencies of improper solutions for the modified population structure, which are much less than those on the top. However, improper solutions are still observed in small sample sizes. The last portions of the table report frequencies of improper solutions for the modified factor structure, but when the data were contaminated by structural errors. (How the structural errors were built in will be explained later.) These numbers are again very large (although not as large as those in the first situation), indicating the need for further regularization in the estimation of CFA.

Such problematic situations as above call for an estimation procedure that can avoid improper solutions. A new estimation method (called regularized CFA, or RCFA for short) for unique variances is proposed to meet this demand. More specifically, it is assumed that the unique variances are proportional to some tentative (mostly non-iterative) estimates of "unique variances", and an optimization is done only with respect to the constant of proportionality. Given a tentative estimate of unique variances obtained non-iteratively, the computation of RCFA reduces to choosing the best value of the constant of proportionality (i.e., the regularization parameter λ) in such a way that a discrepancy function associated with ML, GLS, or ULS is minimized.

Three kinds of regularization scheme are considered in this paper. The simplest case is to set $\Psi = \lambda I$ (e.g., Anderson, 1984). This is based on the assumption that unique variances are constant across variables. This should work reasonably well when the unique variances are relatively uniform across manifest variables. The second alternative is to set $\Psi = \lambda\,\mathrm{diag}(S^{-1})^{-1}$ (e.g., Yanai & Mukherjee, 1987), where S represents C for ML and GLS, and R for LS. In this scheme, unique variances are assumed to be proportional to the variances of anti-image variables. This is similar to image factor analysis proposed by Jöreskog (1969), but the optimization is much simpler in our case because in Jöreskogs case, λ is a function of the model, while here λ is a function of data. The third alternative is to assume that Ψ is proportional to Ihara and Kano's non-iterative estimates of unique variances (Ihara & Kano, 1986; Kano, 1990). (See also Cudeck (1991) for an account of the best possible subset selection.) We can derive a variety of RCFA by all possible combinations of the three types of regularization scheme and the three estimation criteria (ULS, GLS, and ML). We may regard those varieties as one-parameter (λ) family of ULS, GLS, and ML, while traditional iterative methods as full ULS, GLS, and ML.

2. Design of the study

A comparative study between RCFA and three widely used estimation methods is conducted in three parts. The first two parts concern Monte Carlo studies to demonstrate the usefulness of RCFA. The population covariance matrix is assumed known from which a large number of data sets are generated. (We use the modified population factor structure.) The difference between those two experiments is that the first study considers data sets containing only sampling error, while the second both sampling error and structural error. In the third part, RCFA is applied to the same 13 variables from Bergami and Bagozzi's data as introduced earlier to demonstrate how well it works in the analysis of empirical data.

2.1. Sampling error

One hundred replicated data sets of varying sample sizes (50, 100, 200, 500, and 1000) were generated using the hypothesized population obtained by modifying the factor structure in Table 4 in the way described above. The first experiment is designed to apply RCFA and the conventional estimation methods to these data sets containing only sampling error.

2.2. Combining sampling error and structural error

We assume that there are many other influences (on covariances among observed variables) from outside the area of interest which an experimenter has a little or no control over. For example, an item score may be inadvertently influenced by the item location (i.e., context effect) in a test in computerized adaptive testing (CAT), in which difficulty of the item may depend on where it appears to some extent. This concept was formulated as minor factors in the context of CFA model (Tucker, Koopman, & Linn, 1969). Minor factors may be viewed as structural error in the sense that the CFA model may not exactly account for the population covariance matrix. Tucker el al. (1969) proposed a procedure to generate sample correlation matrices containing the effect of minor factors, which have some influence on measured variables. Following their procedure, structural error is explicitly built in the population covariance matrix for the present study. To incorporate such an error, minor factor loadings are computed on 50 factors. The proportion of the variance in a measured variable accounted for by the minor factors is set to 0.075 (for a good fit) equally across 10 variables in the population. The total contribution of minor factors to the variance is distributed over a series of minor factors in such a way that the contribution of each subsequent minor factor is smaller than the preceding one. The unique variance of the measured variable is then set in such a way that the total variance of the measured variable is unity. As before both RCFA and the three full versions of ULS, GLS, and ML estimation methods were applied to the generated data sets.

2.3. An empirical data set

The 13 variables from Bergami and Bagozzi's (2000) social identity study measure organizational stereotype. Participants ($n = 291$) recorded their responses on how much each of characteristics such as "innovative", "dynamic", "democratic", and so on describes their company, using 5-point scales (i.e., 1 = not at all, and 5 = very much).

3. Results

We shall first present results from the Monte Carlo studies and then the analysis of a real data set. Since we may not assume that unique variances are uniform across all the manifest variables, and since GLS has the same asymptotic properties as ML, we present the Monte Carlo results for one-parameter ULS and ML methods only under the anti-image and Ihara-Kano's estimator (the I-K estimator for short). The quality of estimators is evaluated in terms of how close they are on average to the population parameters. Mean square error (MSE) is calculated to measure the average discrepancy, which is defined by $\frac{1}{N}\sum_{i=1}^{N} \mathrm{SS}(\hat{\theta}_i - \theta)$, where $\hat{\theta}$ is the vector of parameter estimates obtained from sample i, θ is the vector of population parameters, and N indicates the number of Monte Carlo samples. Here, θ could be a population covariance matrix, while $\hat{\theta}_i$ a reproduced covariance matrix derived from the estimates of factor loadings and unique variances. The MSE is then normalized by dividing it by $\mathrm{SS}(\theta)$.

Table 5. Normalized MSEs for full ULS and one-parameter family of ULS (sampling error only). (n = sample size and Freq. = frequencies of improper solutions.)

ULS	n	Freq.	Covariance matrix	Factor loadings	Unique variances
Full	50	11	.077	.042	.061
	100	5	.042	.022	.030
	200	0	.023	.011	.016
	500	0	.008	.004	.006
	1000	0	.004	.002	.003
Anti-image	50	0	.078	.043	.076
	100	0	.042	.022	.048
	200	0	.023	.012	.036
	500	0	.008	.005	.030
	1000	0	.004	.003	.027
I-K estimator	50	4	.077	.043	.074
	100	3	.042	.021	.037
	200	0	.023	.011	.018
	500	0	.008	.004	.007
	1000	0	.004	.002	.004

3.1. Sampling error only

Table 5 gives the results of fitting the CFA model by full ULS and one-parameter ULS under the two assumptions (anti-image and the I-K estimator) to 100 samples generated from the population for sampling-error-only condition. As the sample size increases, MSE decreases monotonically. One-parameter ULS under the anti-image assumption yields reasonably good results compared to full ULS in all cases of reproduced covariance and factor loading matrices. This method seems to work well in approximating the population unique variances, particularly when sample size is small. Note, in particular, that no improper solutions are obtained with this procedure in all sample sizes. One-parameter ULS with the I-K estimator shows good performance for approximating the population covariance and factor loading matrices. It yields improper solutions for small sample sizes less frequently than the full ULS. Perhaps for this reason, there are relatively larger MSEs for unique variances for small sample sizes. When the sample size is large (more than 200), no improper solutions were found, and little difference between the full ULS and one-parameter ULS with the I-K estimator was observed in estimating unique variances.

Table 6 shows MSEs and frequencies of improper solutions obtained by the full ML and one-parameter ML for the sampling-error-only condition. One-parameter ML methods under both assumptions yield almost the same MSEs relative to the full ML in all cases for reproduced covariance and factor loading matrices. The one-parameter ML estimation method under the anti-image assumption works well in approximating unique variances for sample size as small as 50, while one-parameter ML under the I-K estimator works well from sample size 100 to 1000. One-parameter ML under the I-K estimator yields more improper solutions than the full ULS, but much less frequently than the full ML for the sample size of 50.

Table 6. Normalized MSEs for full ML and one-parameter family of ML (sampling error only) (n = sample size and Freq. = frequencies of improper solutions.)

ML	n	Freq.	Covariance matrix	Factor loadings	Unique variances
Full	50	22	.077	.042	.068
	100	5	.042	.022	.031
	200	0	.023	.011	.016
	500	0	.007	.004	.006
	1000	0	.004	.002	.003
Anti-image	50	0	.078	.042	.076
	100	0	.042	.022	.050
	200	0	.024	.012	.039
	500	0	.008	.005	.031
	1000	0	.005	.003	.029
I-K estimator	50	6	.080	.048	.116
	100	3	.043	.025	.040
	200	1	.023	.011	.018
	500	0	.008	.004	.007
	1000	0	.004	.002	.004

3.2. Sampling error and structural error combined

Table 7 shows MSEs by the full ML and one-parameter ML in the combined-errors condition. The one-parameter ML method has almost the same MSEs as the full ML for reproduced covariance and factor loading matrices. Due to the structural error, MSEs of unique variances by both one-parameter ML and the full ML are consistently larger than those in the sampling-error-only case. It is shown that the one-parameter ML method with the anti-image scheme works well in approximating the population unique variances for small sample sizes (e.g., $n = 50$ and 100), while the ML method with the I-K scheme for moderate to large sample sizes (more than 200). In the combined-errors situation, the one-parameter ML method under the anti-image assumption is likely to provide a good (even a better for the sample size of 50) approximation to the true unique variances for small sample sizes with no risk of improper solutions. (We do not present any results from the ULS method for this condition because essentially the same patterns of MSEs as found by the ML method were found in the ULS method.) Since the structural error has been incorporated in the CFA model, the true values of unique variances were set closer to zero than in the sampling-error-only case. This might cause improper solutions more frequently under the I-K assumption.

3.3. An empirical data set

We compare the results of the one-parameter ML method with the results from the full ML method in Table 2. Table 8 reports estimated factor loadings and unique variances obtained by the one-parameter ML method under the anti-image assumption. Table 9 is essentially a replica of Table 2 with the estimates of unique variances by the full ML method appended as the last column. It is clear that the derived factor loadings in Table 8 are quite similar to those estimated by the full ML method in Table 9. We may conclude that this proposed procedure performs reasonably well in a real data set.

Table 7. Normalized MSEs for full ML and one-parameter family of ML (sampling error and structural error) (n = sample size and Freq. = frequencies of improper solutions.)

ML	n	Freq.	Covariance matrix	Factor loadings	Unique variances
Full	50	35	.076	.043	.119
	100	25	.040	.023	.067
	200	22	.018	.011	.057
	500	3	.010	.006	.051
	1000	1	.004	.004	.049
Anti-image	50	0	.077	.043	.102
	100	0	.041	.023	.073
	200	0	.018	.011	.066
	500	0	.010	.006	.061
	1000	0	.005	.003	.059
I-K estimator	50	17	.077	.044	.126
	100	13	.040	.023	.076
	200	7	.018	.011	.058
	500	1	.010	.006	.051
	1000	1	.004	.003	.050

4. Concluding remarks

Common factor analysis (CFA) aims to discover a simple pattern of factor loadings which can account for structural relationships among observed variables. It is important that it explicitly models measurement errors in observed variables, so commonly observed in research in psychology and other social sciences. Loadings on common factors are shrunken toward zero (relative to PCA loadings) to take into account the measurement errors. However, the conventional estimation methods for unique variances often yield improper solutions. Computational remedies have been proposed to prevent the problem (Harman & Fukuda, 1966; Jennrich, 1986; Jöreskog & Goldberger, 1972). Several strategies have been used in practical settings to deal with the problem in ML CFA (Jöreskog, 1967). Each strategy (or a modification in the computational routine) might be an alternative, but not a resolution. Using the traditional methods, CFA always suffers from a potential risk of improper solutions.

To avoid the problem of improper solutions, we proposed a new approach of estimation called regularized CFA (RCFA) by applying the idea of generalized ridge regularization. It is assumed that the unique variances are proportional to a diagonal matrix of tentative estimates of unique variances obtained non-iteratively. RCFA reduces the number of parameters in estimating unique variances, so that it could dramatically decrease the frequency of improper solutions.

A variety of RCFA can be derived by combining three kinds of regularization scheme and three widely used estimation criteria. We developed one-parameter family of ULS, GLS, and ML, which are much simpler and efficient computationally, are much more stable numerically, and have no risk of improper solutions unless tentative estimates of "unique variances" are improper to begin with (as are possible in the case of the I-K estimator). The numerical results from the Monte Carlo studies and a real data set are encouraging. Particularly when the sample size is small, one-parameter ULS and ML estimation methods under anti-image assumption are likely to produce estimates of unique variances which are good

Table 8. One-parameter ML CFA: Anti-image assumption

	Factor loadings		
Var.	1	2	$\hat{\Psi}$
1	-.615	-.336	.497
2	-.170	-.648	.616
3	-.176	-.577	.657
4	-.617	-.207	.583
5	-.131	-.394	.756
6	-.374	-.520	.576
7	-.140	-.440	.741
8	-.387	-.496	.573
9	-.722	-.224	.444
10	-743	-.267	.404
11	-.731	-.169	.459
12	-.645	-.237	.513
13	-.663	-.324	.453
Fac. Cont.	3.608	2.084	

Table 9. Full Maximum Likelihood CFA

	Factor loadings		
Var.	1	2	$\hat{\Psi}$
1	-.606	-.346	.513
2	-.136	-.725	.457
3	-.172	-.589	.624
4	-.612	-.214	.580
5	-.148	-.366	.763
6	-.374	-.506	.603
7	-.159	-.402	.800
8	-.390	-.474	.623
9	-.737	-.211	.413
10	-758	-.262	.357
11	-.739	-.170	.425
12	-.628	-.249	.544
13	-.649	-.339	.464
Fac. Cont.	3.609	2.133	

approximations to the population unique variances. For moderate to large sample sizes, one-parameter ULS and ML methods under the I-K estimator perform reasonably well with a few exceptional cases of improper solutions.

Estimation of unique variances in CFA is generally difficult when the sample size is small. In such cases, the sample covariance matrix may be singular (or near-singular) and cannot be inverted, which causes all sorts of algebraic and numerical problems. To resolve the singularity problem, Yuan (2007) recently proposed to use a regularized covariance matrix which is obtained by adding a small positive value to the diagonal elements of the sample covariance matrix. The full ML estimation method is then used for the modified covariance matrix. This procedure may not only be able to deal with the singularity problem, but may also be effective in reducing the frequency of improper solutions. The comparison of our procedure with this method would undoubtedly be interesting. The regularized covariance matrix might be useful in medical imaging analysis (e.g., fMRI), where the number of subjects is much smaller than the number of variables.

5. Acknowledgement

The work reported in this paper was supported by Grant A6294 to the second author from the Natural Sciences and Engineering Research Council of Canada.

References

Anderson, T. W. (1984). Estimating linear statistical relationships. *The Annals of Statistics, 12*, 1–45.

Bergami, M., & Bagozzi, R. P. (2000). Self-categorization, affective commitment and group self-esteem as distinct aspects of social identity in the organization. *British Journal of Social Psychology, 39*, 555–577.

Cudeck, R. (1991). Noniterative factor anlaysis estimators, with algorithms for subset and instrumental variable selection. *Journal of Educational Statistics, 16*, 35–52.

Hägglund, G. (1982). Factor analysis by instrumental variables methods. *Psychometrika, 47*, 209–222.

Harman, H. H., & Fukuda, Y. (1966). Resolution of the heywood case in the minres solution. *Psychometrika, 31*, 563–571.

Hoerl, A. F., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*, 55–67.

Ihara, M., & Kano, Y. (1986). A new estimator of the uniqueness in factor analysis. *Psychometrika, 51*, 563–566.

Jennrich, R. I. (1987). Tableau algorithms for factor analysis by instrumental variable methods. *Psychometrika, 52*, 469–476.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32*, 443–482.

Jöreskog, K. G. (1969). Efficient estimation in image factor analysis. *Psychometrika, 34*, 51–75.

Jöreskog, K. G., & Goldberger, A. S. (1972). Factor analysis by generalized least squares. *Psychometrika, 37*, 243–260.

Kano, Y. (1990). Noniterative estimation and the choice of the number of factors in exploratory factor analysis. *Psychometrika, 55*, 277–291.

Takane, Y., & Yanai, H. (2006) On ridge operators. Submitted for publication.

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika, 34*, 421–459.

Yanai, H., & Mukherjee, B. N. (1987). A generalized method of image analysis from an intercorrelation matrix which may be singular. *Psychometrika, 52*, 555–564.

Yanai, H., & Takane, T. (2007). Matrix methods and their applications to factor analysis. Handbook of Latent Variable and Related Models. Amsterdam: Elsevier.

Yuan, K.-H., & Chen, W. (2007). Sturctural equation modeling with near singular covariance matrices. Paper presented at IMPS-07, Tokyo.