

On Extended Guttman Condition in High Dimensional Factor Analysis



Kentaro Hayashi, Ke-Hai Yuan and Ge (Gabriella) Jiang

Abstract It is well-known that factor analysis and principal component analysis often yield similar estimated loading matrices. Guttman (Psychometrika 21:273–285, 1956) identified a condition under which the two matrices are close to each other at the population level. We discuss the matrix version of the Guttman condition for closeness between the two methods. It can be considered as an extension of the original Guttman condition in the sense that the matrix version involves not only the diagonal elements but also the off-diagonal elements of the inverse matrices of variance-covariances and unique variances. We also discuss some implications of the extended Guttman condition, including how to obtain approximate estimates of the inverse of covariance matrix under high dimensions.

Keywords High dimensions · Principal components · Unique variances

1 Factor Analysis and Principal Component Analysis

Factor analysis (FA) and principal component analysis (PCA) are frequently used multivariate statistical methods for data reduction. In FA (Anderson, 2003; Lawley & Maxwell, 1971), the p -dimensional mean-centered vector of the observed variables y_i , $i = 1, \dots, n$, is linearly related to a m -dimensional vector of latent factors f_i

K. Hayashi (✉)

Department of Psychology, University of Hawaii at Manoa, 2530 Dole Street, Sakamaki C400, Honolulu, HI 96822, USA

e-mail: hayashik@hawaii.edu

K.-H. Yuan

Department of Psychology, University of Notre Dame, Corbett Family Hall, Notre Dame, IN 46556, USA

e-mail: kyuan@nd.edu

G. (Gabriella) Jiang

Department of Educational Psychology, University of Illinois at Urbana-Champaign, 1310 South 6th Street, Campaign, Urbana-Champaign, IL 61820, USA

e-mail: gejiang2@illinois.edu

© Springer Nature Switzerland AG 2019

M. Wiberg et al. (eds.), *Quantitative Psychology*, Springer Proceedings in Mathematics & Statistics 265, https://doi.org/10.1007/978-3-030-01310-3_20

221

via $y_i = \Lambda f_i + \epsilon_i$, where $\Lambda = (\lambda_1, \dots, \lambda_m)$ is a $p \times m$ matrix of factor loadings (with $p > m$), and ϵ_i is a p -dimensional vector of errors. Typically for the orthogonal factor model, the three assumptions are imposed: (i) $f_i \sim N_m(\mathbf{0}, \mathbf{I}_m)$; (ii) $\epsilon_i \sim N_p(\mathbf{0}, \Psi)$, where Ψ is a diagonal matrix with positive elements on the diagonal; (iii) $Cov(f_i, \epsilon_i) = \mathbf{0}$. Then, under these three assumptions, the covariance matrix of y_i is given by $\Sigma = \Lambda \Lambda' + \Psi$. If y_i is standardized, Σ is a correlation matrix.

Let $\Lambda^+ = (\lambda_1^+, \dots, \lambda_m^+)$ be the $p \times m$ matrix whose columns are the standardized eigenvectors corresponding to the first m largest eigenvalues of Σ ; $\Omega = \text{diag}(\omega)$ be the $m \times m$ diagonal matrix whose diagonal elements $\omega = (\omega_1, \dots, \omega_m)'$ are the first m largest eigenvalues of Σ ; and $\Omega^{1/2}$ be the $m \times m$ diagonal matrix whose diagonal elements are the square root of those in Ω . Then principal components (PCs) (c.f., Anderson, 2003) with m elements are obtained as $f_i^* = \Lambda^{+'} y_i$. Clearly, the PCs are uncorrelated with a covariance matrix $\Lambda^{+'} \Sigma \Lambda^+$. When m is properly chosen, there exists $\Sigma \approx \Lambda^+ \Omega \Lambda^{+'} = \Lambda^* \Lambda^{*'}$, where $\Lambda^* = \Lambda^+ \Omega^{1/2}$ is the $p \times m$ matrix of PCA loadings.

2 Closeness Conditions Between Factor Analysis and Principal Component Analysis

It has been well-known that FA and PCA often yield approximately the same results, especially their estimated loading matrices $\hat{\Lambda}$ and $\hat{\Lambda}^*$, respectively (e.g., Velicer & Jackson, 1990). Conditions under which the two matrices are close to each other are of substantial interest. At the population level, two such conditions identified by Guttman (1956) and Schneeweiss (1997) are among the most well-known.

2.1 Guttman Condition

Consider the factor analysis model $\Sigma = \Lambda \Lambda' + \Psi$, where Ψ is a diagonal unique variance matrix, with $(\Sigma^{-1})_{jj} = \sigma^{jj}$ and $(\Psi)_{jj} = \psi_{jj}$, $j = 1, \dots, p$. Let m be the number of common factors, Guttman (1956; See also Theorem 1 of Krijnen, 2006) has shown that if $m/p \rightarrow 0$ as $p \rightarrow \infty$, then $\psi_{jj} \sigma^{jj} \rightarrow 1$ for almost all j . Here, “for almost all j ” means $\lim_{p \rightarrow \infty} \#\{j : \psi_{jj} \sigma^{jj} < 1\} / p = 0$. That is, the number of j that satisfies $\psi_{jj} \sigma^{jj} < 1$ is ignorable as p goes to infinity.

2.2 Schneeweiss Condition

The closeness condition between the loading matrix from FA and that from PCA by Schneeweiss and Mathes (1995) and Schneeweiss (1997) is $ev_m(\Lambda' \Psi^{-1} \Lambda) \rightarrow$

∞ , where $ev_k(A)$ is the k -th largest eigenvalue of a square matrix A . Obviously, $ev_m(\Lambda'\Psi^{-1}\Lambda)$ is the smallest eigenvalue of $\Lambda'\Psi^{-1}\Lambda$.

Related with the Schneeweiss condition, Bentler (1976) parameterized the correlation structure of the factor model as $\Psi^{-1/2}\Sigma\Psi^{-1/2} = \Psi^{-1/2}\Lambda\Lambda'\Psi^{-1/2} + I_p$ and showed that, under this parameterization, a necessary condition for $ev_m(\Lambda'\Psi^{-1}\Lambda) = ev_m(\Psi^{-1/2}\Lambda\Lambda'\Psi^{-1/2}) \rightarrow \infty$ is that as p increases, the sum of squared loadings on each factor has to go to infinity ($\lambda'_k\lambda_k \rightarrow \infty, k = 1, \dots, m$, as $p \rightarrow \infty$).

2.3 Relationship Between Guttman and Schneeweiss Conditions

The relationship between Guttman and Schneeweiss conditions is summarized in Table 1. Schneeweiss condition ($ev_m(\Lambda'\Psi^{-1}\Lambda) \rightarrow \infty$) is sufficient for Guttman condition ($m/p \rightarrow 0$ as $p \rightarrow \infty$) (Krijnen, 2006, Theorem 3). What we would like is for the converse ($m/p \rightarrow 0$ as $p \rightarrow \infty \Rightarrow ev_m(\Lambda'\Psi^{-1}\Lambda) \rightarrow \infty$) to hold in practical applications, as to be discussed in the next section.

First, the condition of $m/p \rightarrow 0$ as $p \rightarrow \infty$ is sufficient for $\psi_{jj}\sigma^{jj} \rightarrow 1$ for almost all j (Guttman, 1956; Krijnen, 2006, Theorem 1). Also, $\psi_{jj}\sigma^{jj} \rightarrow 1$ for all j implies $ev_m(\Lambda'\Psi^{-1}\Lambda) \rightarrow \infty$ (Krijnen, 2006, Theorem 4). Here, “ $\psi_{jj}\sigma^{jj} \rightarrow 1$ for all j ” is slightly stronger than “ $\psi_{jj}\sigma^{jj} \rightarrow 1$ for almost all j .” However, in practice, it seems reasonable to assume that the number of loadings on every factor increases with p proportionally, as stated in Bentler (1976). Then the condition of $m/p \rightarrow 0$ as $p \rightarrow \infty$ becomes equivalent to $ev_m(\Lambda'\Psi^{-1}\Lambda) \rightarrow \infty$. That is, Guttman and Schneeweiss conditions become interchangeable.

Table 1 Relationships among conditions and results

Condition(s)	Result	Source
1. $m/p \rightarrow 0$ as $p \rightarrow \infty$	$\psi_{jj}\sigma^{jj} \rightarrow 1$ for almost all j	Guttman (1956), Krijnen (2006, Thm 1)
2. $\psi_{jj}\sigma^{jj} \rightarrow 1$ for almost all j ; $ev_m(\Lambda'\Psi^{-1}\Lambda) > c > 0$	$m/p \rightarrow 0$ as $p \rightarrow \infty$	Krijnen (2006, Thm 2)
3. $ev_m(\Lambda'\Psi^{-1}\Lambda) \rightarrow \infty$	$m/p \rightarrow 0$ as $p \rightarrow \infty$	Krijnen (2006, Thm 3), Hayashi and Bentler (2000, Obs 8b)
4. $ev_m(\Lambda'\Psi^{-1}\Lambda) \rightarrow \infty$	$\psi_{jj}\sigma^{jj} \rightarrow 1$ for all j	Krijnen (2006, Thm 4), Hayashi and Bentler (2000, after Obs 7)
5. $\psi_{jj}\sigma^{jj} \rightarrow 1$ for all j	$ev_m(\Lambda'\Psi^{-1}\Lambda) \rightarrow \infty$	Krijnen (2006, Thm 4)
6. $ev_k(\Lambda'\Psi^{-1}\Lambda) \rightarrow \infty, k = 1, \dots, r$; $ev_k(\Lambda'\Psi^{-1}\Lambda) < C < \infty, k = r + 1, \dots, m$; $p \rightarrow \infty, m$ fixed	$\psi_{jj}\sigma^{jj} \rightarrow 1$ for almost all j	Krijnen (2006, Thm 5)
7. $ev_m(\Lambda'\Psi^{-1}\Lambda) \rightarrow \infty$	$\Psi^{-1} - \Sigma^{-1} \rightarrow \mathbf{0}$	

Notes (i) 2 is a partial converse of 1; (ii) 5 is the converse of 4; (iii) 7 is the matrix version of 4

3 Extended Guttman Condition

By far the most important consequence of the Schneeweiss condition is that, when $ev_m(\Lambda'\Psi^{-1}\Lambda) \rightarrow \infty$, the second term in the right-hand side of the Sherman-Morrison-Woodbury formula (see, e.g., Chap. 16 of Harville, 1997):

$$\Sigma^{-1} = \Psi^{-1} - \Psi^{-1}\Lambda(I_m + \Lambda'\Psi^{-1}\Lambda)^{-1}\Lambda'\Psi^{-1} \quad (1)$$

vanishes, so that

$$\Psi^{-1} - \Sigma^{-1} \rightarrow \mathbf{0} \quad \text{as} \quad ev_m(\Lambda'\Psi^{-1}\Lambda) \rightarrow \infty \quad (2)$$

As we noted in the previous section, the condition of $m/p \rightarrow 0$ as $p \rightarrow \infty$ can be equivalent to $ev_m(\Lambda'\Psi^{-1}\Lambda) \rightarrow \infty$ in practical applications. Therefore, we have $\Psi^{-1} - \Sigma^{-1} \rightarrow \mathbf{0}$ under high dimensions with a large p . We call $\Psi^{-1} - \Sigma^{-1} \rightarrow \mathbf{0}$ the extended Guttman condition. It is an extension of the original Guttman condition in the sense that $\psi_{jj}\sigma^{jj} \rightarrow 1$ can be expressed as $\psi_{jj}^{-1} - \sigma^{jj} \rightarrow 0$, as long as ψ_{jj} is bounded above ($\psi_{jj} \leq \psi_{\text{sup}} < \infty$).

Note that there exists a similar identity for the FA model:

$$\Psi^{-1} - \Psi^{-1}\Lambda(\Lambda'\Psi^{-1}\Lambda)^{-1}\Lambda'\Psi^{-1} = \Sigma^{-1} - \Sigma^{-1}\Lambda(\Lambda'\Sigma^{-1}\Lambda)^{-1}\Lambda'\Sigma^{-1} \quad (3)$$

(see, e.g., Hayashi & Bentler, 2001). Clearly, as $ev_m(\Lambda'\Psi^{-1}\Lambda) \rightarrow \infty$, not only the second term on the left-hand side of Eq. (3) but the second term on the right-hand side of Eq. (3) vanishes.

As we have just seen, the extended Guttman condition is a direct consequence of the Schneeweiss condition. Because $\Psi^{-1}\Lambda(I_m + \Lambda'\Psi^{-1}\Lambda)^{-1}\Lambda'\Psi^{-1} < \Psi^{-1}\Lambda(\Lambda'\Psi^{-1}\Lambda)^{-1}\Lambda'\Psi^{-1}$ and $I_m + \Lambda'\Psi^{-1}\Lambda$ is only slightly larger than $\Lambda'\Psi^{-1}\Lambda$ when $\Lambda'\Psi^{-1}\Lambda$ is large in the sense that $ev_m(I_m + \Lambda'\Psi^{-1}\Lambda) = ev_m(\Lambda'\Psi^{-1}\Lambda) + 1$, the speed of convergence in $\Psi^{-1} - \Sigma^{-1} \rightarrow \mathbf{0}$ is approximately at the rate of the reciprocal of smallest eigenvalues of $\Lambda'\Psi^{-1}\Lambda$, that is, of $1/ev_m(\Lambda'\Psi^{-1}\Lambda)$.

4 Approximation of the Inverse of the Covariance Matrix

An important point to note here is that the original Guttman condition of $\psi_{jj}\sigma^{jj} \rightarrow 1$ (for almost all j) has to do with only the diagonal elements of Ψ (or Ψ^{-1}) and Σ^{-1} , while $\Psi^{-1} - \Sigma^{-1} \rightarrow \mathbf{0}$ involves both the diagonal and the off-diagonal elements of the matrices. It justifies the interchangeability of Σ^{-1} and Ψ^{-1} as $ev_m(\Lambda'\Psi^{-1}\Lambda) \rightarrow \infty$, or assuming that the number of loadings on every factor increases with p proportionally, as $m/p \rightarrow 0$ with $p \rightarrow \infty$. The important implication is that all the off-diagonal elements of Σ^{-1} approach zero in the limit. Thus, it is a result of spar-

sity of the off-diagonal elements of the inverted covariance (correlation) matrix in high dimensions.

One of the obvious advantages of being able to approximate Σ^{-1} by Ψ^{-1} in high dimensions is that the matrix of unique variances Ψ is a diagonal matrix and thus it can be inverted only with p operations. Note that, in general, the inversion of a p –dimensional square matrix requires operations of order $O(p^3)$ (see, e.g., Pourahmadi, 2013, p. 121).

Consequently, the single most important application of the extended Guttman condition is to approximate the inverse of the covariance matrix Σ^{-1} by Ψ^{-1} under high dimensions. This implication is very important because Σ^{-1} is involved in the quadratic form for the log likelihood function of the multivariate normal distribution. Even if Σ is positive definite so that Σ^{-1} exists in the population, the inverse S^{-1} of the sample covariance matrix S does not exist under high dimensions when $p > n$. When S^{-1} does not exist, we cannot estimate Ψ^{-1} under the FA model using the generalized least squares (GLS) or the maximum likelihood (ML) method, without resorting to certain regularization method(s), either. Thus, a natural choice would be to employ the unweighted least square (ULS) estimation method that minimizes the fit function of $F_{ULS}(S, \Sigma) = tr\{(S - \Sigma)^2\}$, which does not require to compute S^{-1} or the estimate of Σ^{-1} . Note that $1 - 1/s^{jj}$, a common initial value for the j -th communality cannot be used because it requires the computation of S^{-1} . Then, we can use the value of 1 as the initial communality estimates. In this case, the initial solution is identical to PCA.

Alternatively, when p is huge, we can employ the following “approximate” FA model with equal unique variances (e.g., Hayashi & Bentler, 2000), using standardized variables, that is, applying to the correlation matrix:

$$\Sigma \approx \Lambda^* \Lambda^{*'} + kI_p, \tag{4}$$

with a positive constant k . Note that this model is also called the probabilistic PCA in statistics (Tipping & Bishop, 1999). Use of the FA model with equal unique variances seems reasonable, because the eigenvectors of $(\Sigma - kI_p)$ are the same as the eigenvectors of Σ , and the eigenvalues of $(\Sigma - kI_p)$ are smaller than the eigenvalues of Σ by only the constant of k . Thus, the FA model with equal unique variances is considered as a variant of the PCA, and, the loading matrices between the FA and the PCA approach the same limit values as $ev_m(\Lambda' \Psi^{-1} \Lambda) \rightarrow \infty$, or they become essentially equivalent, under high dimensions.

In Eq. (4), let $\Psi^* = kI_p$, then $\Psi^{*-1} = k^{-1}I_p$. Thus, we can use Ψ^{*-1} as quick and fast approximation for Ψ^{-1} . The natural estimator of k is the MLE for k given Λ^* (Tipping & Bishop, 1999):

$$\hat{k} = \frac{1}{p - m} \sum_{j=m+1}^p ev_j(S). \tag{5}$$

However, a more practical method seems as follows: Once estimating $\Psi^* = kI_p$ by $\hat{\Psi}^* = \hat{k}I_p$, we can compute loadings $\hat{\Lambda}^*$ using the eigenvalues and eigenvectors of $(S - \hat{k}I_p)$ and find the estimates of Ψ^* as $\hat{\Psi}^* = \text{diag}(S - \hat{\Lambda}^* \hat{\Lambda}^{*\prime})$. Note that $\hat{\Psi}^*$ is no longer a constant times the identity matrix. Now, invoke the estimator version of the extended Guttman condition $\hat{\Psi}^{*-1} - \hat{\Sigma}^{-1} \approx \mathbf{0}$ to find the approximate estimator $\hat{\Sigma}^{-1}$ of Σ^{-1} .

5 Illustration

The compound symmetry correlation structure is expressed as $\Sigma = (1 - \rho)I_p + \rho \mathbf{1}_p \mathbf{1}_p'$ with a common correlation ρ , $0 < \rho < 1$. Obviously, it is a one-factor model with the vector of factor loadings $\lambda_1 = \sqrt{\rho} \mathbf{1}_p$ and the diagonal matrix unique variances $\Psi = (1 - \rho)I_p$. Because the first eigenvalue and the corresponding standardized eigenvector of $\Sigma = (1 - \rho)I_p + \rho \mathbf{1}_p \mathbf{1}_p'$ are $\omega_1 = 1 + (p - 1)\rho$ and $\lambda_1^+ = (1/\sqrt{p})\mathbf{1}_p$, respectively, the first PC loading vector is

$$\lambda_1^* = \lambda_1^+ \sqrt{\omega_1} = (1/\sqrt{p})\sqrt{1 + (p - 1)\rho} \cdot \mathbf{1}_p = \sqrt{1/p + (1 - 1/p)\rho} \cdot \mathbf{1}_p, \quad (6)$$

which approaches the vector of factor loadings $\lambda_1 = \sqrt{\rho} \mathbf{1}_p$ with $m/p = 1/p \rightarrow 0$ and $p \rightarrow \infty$. The remaining $p - 1$ eigenvalues are $\omega_2 = \dots = \omega_p = 1 - \rho$. Thus, obviously, the constant k in the FA model with equal unique variances is $k = 1 - \rho$. Note that the Schneeweiss condition also holds

$$\lambda_1' \Psi^{-1} \lambda_1 = (\sqrt{\rho} \mathbf{1}_p)' \{ (1/(1 - \rho)) I_p \} (\sqrt{\rho} \mathbf{1}_p) = p \cdot \rho / (1 - \rho) \rightarrow \infty \quad (7)$$

with $m/p = 1/p \rightarrow 0$ as $p \rightarrow \infty$. The inverse of the correlation matrix is:

$$\begin{aligned} \Sigma^{-1} &= \Psi^{-1} - \Psi^{-1} \lambda_1 (1 + \lambda_1' \Psi^{-1} \lambda_1)^{-1} \lambda_1' \Psi^{-1} \\ &= \left(\frac{1}{1 - \rho}\right) I_p - \left(\frac{1}{1 - \rho}\right) I_p \cdot (\sqrt{\rho} \mathbf{1}_p) \cdot \left(1 + \frac{\rho}{1 - \rho} \cdot p\right)^{-1} \cdot (\sqrt{\rho} \mathbf{1}_p)' \cdot \left(\frac{1}{1 - \rho}\right) I_p \\ &= \left(\frac{1}{1 - \rho}\right) I_p - \left(\frac{\rho}{1 - \rho}\right) \left(\frac{1}{(1 - \rho) + \rho \cdot p}\right) (\mathbf{1}_p \mathbf{1}_p') \rightarrow \left(\frac{1}{1 - \rho}\right) I_p = \Psi^{-1} \end{aligned} \quad (8)$$

with $m/p = 1/p \rightarrow 0$ as $p \rightarrow \infty$.

For example, it is quite easy to show that if $\rho = 0.5$, then for $p = 10$, the diagonal elements of the inverse of the compound symmetry correlation structure are $2 - 1/5.5 = 1.818$ and the off-diagonal elements are $-1/5.5 = -0.182$. At $p = 100$, the diagonal and the off-diagonal elements become $2 - 1/50.5 = 1.980$ and $-1/50.5 = -0.0198$, respectively. Furthermore, at $p = 1000$, the diagonal and the off-diagonal elements become $2 - 1/500.5 = 1.998$ and $-1/500.5 = -0.001998$. Again, we see the off-diagonal elements of Σ^{-1} approaching 0 as p increases. Also, the diagonal elements

of Σ^{-1} approach 2, which are the value of the inverse of the unique variances in the FA model.

6 Discussion

We discussed the matrix version of the Guttman condition for closeness between FA and PCA. It can be considered as an extended Guttman condition in the sense that the matrix version involves not only the diagonal elements but also the off-diagonal elements of the matrices Σ^{-1} and Ψ^{-1} . Because Ψ^{-1} is a diagonal matrix, the extended Guttman condition implies that the off-diagonal elements of Σ^{-1} approach zero as the dimension increases. We showed how the phenomenon happens with the compound symmetry example in the Illustration section. We also discussed some implications of the extended Guttman condition, which include the ease of inverting Ψ compared against inverting Σ . Because the ULS estimation method does not involve any inversion of either the sample covariance matrix S or the estimated model implied population covariance matrix $\hat{\Sigma}$, the ULS should be the estimation of choice when sample size n is smaller than the number of variables p . Furthermore, we proposed a simple method to approximate Σ^{-1} by Ψ^{-1} using the FA model with equal unique variances, or equivalently, the probabilistic PCA model.

Some other implications of the extended Guttman condition (especially with respect to algorithms) are as follows: First of all, suppose we add the $(p + 1)$ th variable at the end of already existing p variables. Then, while the values of σ^{jj} , $j = 1, \dots, p$, can change, ψ_{jj}^{-1} , $j = 1, \dots, p$, remain unchanged. Thus, with the extended Guttman condition, only one additional element needs to be computed.

Another implication is on the ridge estimator, which is among the methods to deal with singularity of S or the estimator of its covariance matrix by introducing some small bias term (see e.g., Yuan & Chan, 2008, 2016). Warton (2008, Theorem 1) showed that the ridge estimator of the covariance (correlation) matrix $\hat{\Sigma}_\eta = \eta \hat{\Sigma} + (1 - \eta) \mathbf{I}_p$ (with the tuning parameter η) is the maximum penalized likelihood estimator with the penalty term proportional to $-tr(\Sigma^{-1})$. Unfortunately, as the dimension p increases (or the ratio p/n increases), it becomes more difficult to obtain the inverse of the covariance matrix. Therefore, in high dimensions, it is not practical to express the ridge estimator of the covariance matrix in the form of the maximum penalized likelihood with the penalty term involving $-tr(\Sigma^{-1})$. This naturally leads to employing an “approximate” maximum penalized likelihood with the penalty term approximately proportional to $-tr(\Psi^{-1})$ in place of the penalty term proportional to $-tr(\Sigma^{-1})$, assuming the factor analysis model, when the dimension p is large.

We are aware that, perhaps except approximations of the inverse of covariance matrix, the majority of implications that we discussed in this article may be of limited practical utility. For example, because the original Guttman condition, the Schneeweiss condition, and the extended Guttman condition are all conditions for closeness between FA and PCA, we can simply employ PCA as an approximation to FA when the conditions hold. Also, we did not discuss regularized FA with L1

regularization here, which in itself is a very interesting topic. Yet, we think the implications we discussed are still of theoretical interest that should continue to be studied. The compound symmetry example used in the Illustration is probably only an approximation to the real world. We will need to do an extensive simulation to come up with some empirical guidelines regarding how to best apply the theoretical results in practice.

Acknowledgements The authors are thankful to Dr. Dylan Molenaar's comments. Ke-Hai Yuan's work was supported by the National Science Foundation under Grant No. SES-1461355.

References

- Anderson, T. W. (2003). *An introduction to multivariate statistical analysis* (3rd ed.). New York: Wiley.
- Bentler, P. M. (1976). Multistructure statistical model applied to factor analysis. *Multivariate Behavioral Research, 11*, 3–15.
- Guttman, L. (1956). Best possible systematic estimates of communalities. *Psychometrika, 21*, 273–285.
- Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. New York: Springer.
- Hayashi, K., & Bentler, P. M. (2000). On the relations among regular, equal unique variances and image factor analysis. *Psychometrika, 65*, 59–72.
- Hayashi, K., & Bentler, P. M. (2001). The asymptotic covariance matrix of maximum-likelihood estimates in factor analysis: The case of nearly singular matrix of estimates of unique variances. *Linear Algebra and its Applications, 321*, 153–173.
- Krijnen, W. P. (2006). Convergence of estimates of unique variances in factor analysis, based on the inverse sample covariance matrix. *Psychometrika, 71*, 193–199.
- Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method* (2nd ed.). New York: American Elsevier.
- Pourahmadi, M. (2013). *High-dimensional covariance estimation*. New York: Wiley.
- Schneeweiss, H. (1997). Factors and principal components in the near spherical case. *Multivariate Behavioral Research, 32*, 375–401.
- Schneeweiss, H., & Mathes, H. (1995). Factor analysis and principal components. *Journal of Multivariate Analysis, 55*, 105–124.
- Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B, 61*, 611–622.
- Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research, 25*, 1–28.
- Warton, D. I. (2008). Penalized normal likelihood and ridge regularization of correlation and covariance matrices. *Journal of the American Statistical Association, 103*, 340–349.
- Yuan, K.-H., & Chan, W. (2008). Structural equation modeling with near singular covariance matrices. *Computational Statistics & Data Analysis, 52*, 4842–4858.
- Yuan, K.-H., & Chan, W. (2016). Structural equation modeling with unknown population distributions: Ridge generalized least squares. *Structural Equation Modeling, 23*, 163–179.