Department of Economics

Working Paper Series

On Relative Efficiency of Quasi-MLE and GMM Estimators of Covariance Structure Models

08-004 | Artem Prokhorov
Concordia University

Department of Economics, 1455 De Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8
Tel 514–848–2424 # 3900 · Fax 514–848–4536 · econ@alcor.concordia.ca · alcor.concordia.ca/~econ/repec

On relative efficiency of Quasi-MLE and GMM estimators of covariance structure models*

Artem Prokhorov†

May 8, 2008

Abstract

Optimal GMM is known to dominate Gaussian QMLE in terms of asymptotic efficiency (Chamberlain, 1984). I derive a new condition under which QMLE is as efficient as GMM for a general class of covariance structure models. The condition trivially holds for normal data but also identifies non-normal cases for which Gaussian QMLE is efficient.

JEL Classification: C13

Keywords: GMM, (Q)MLE, Covariance structures, LISREL, MIMIC, robustness.

*Helpful comments from Emma Iglesias, Peter Schmidt, Jeffrey Wooldridge and participants of the 2006 Canadian Econometrics Study Group meeting are gratefully acknowledged.

†Department of Economics, Concordia University, 1455 de Maisonneuve Blvd W, Montreal QC H4E2S1 Canada; email: artem.prokhorov@concordia.ca

1 Introduction

Traditionally covariance structure models are estimated by maximum likelihood under the assumption of multivariate normality (see, e.g., Jöreskog, 1970). If the data are not normal, MLE is still consistent. However, the MLE standard errors are wrong and inference may be incorrect. It is common to make inference robust to non-normality by using the "sandwich" form of the variance matrix. The form of the variance matrix for normal quasi-MLE of covariance structures can be found, e.g., in Chamberlain (1984, p. 1295).

However, the Gaussian QMLE is generally inefficient. The optimal generalized method of moments estimator (GMM) makes efficient use of the restrictions on the second moments whether or not the data are in fact normal. It is known to be no worse asymptotically than QMLE (e,g., Chamberlain, 1984).

A trivial case when QMLE is efficient is when the data are in fact normal. The first order conditions of QMLE and GMM are asymptotically identical in this case. But it turns out that QMLE may retain the asymptotic optimality property more generally. The condition I derive in this paper is necessary and sufficient for optimality of QMLE. Thus, this paper is related to the work on asymptotic robustness of covariance structure estimators (e.g., Browne, 1987; Anderson and Amemiya, 1988; Browne and Shapiro, 1988; Anderson, 1989; Mooijaart and Bentler, 1991; Satorra and Neudecker, 1994). However, very few papers consider robustness of the efficiency property. If this kind of robustness is considered, results are stated in terms of the higher-order cumulants (e.g, Mooijaart and Bentler, 1991) or provide conditions that are too weak due to some restriction of the model (Satorra and Neudecker, 1994). The robustness condition derived here is new; it involves the fourth moments of data and applies to a general class of models. With its help, one may easily identify situations in which using the normality assumption does not result in an inefficient estimator. As an example, I show that this is so in problems about the variance of two uncorrelated random variables with the Student-t distribution.

2 Preliminaries

Consider a family of distributions $\{P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p, \Theta \text{ compact}\}$ and a random vector $\mathbf{Z} \in \mathcal{Z} \subset \mathbb{R}^q$ from $P_{\boldsymbol{\theta}_o}, \boldsymbol{\theta}_o \in \Theta$, such that $\mathbb{E}\mathbf{Z} = 0$, $\mathbb{E}\{||\mathbf{Z}||^4\} < \infty$ and

$$\mathbb{E}\left[\mathbf{Z}\mathbf{Z}'\right] = \boldsymbol{\Sigma}(\boldsymbol{\theta}), \text{ if and only if } \boldsymbol{\theta} = \boldsymbol{\theta}_o. \tag{1}$$

Expectation is with respect to $P_{\boldsymbol{\theta}_o}$. The matrix function $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ comes from a structural model, e.g., LISREL, MIMIC, factor analysis, random effects or simultaneous equations model.

For a random sample $(\mathbf{Z}_1, \ldots, \mathbf{Z}_N)$, denote

$$\mathbf{S}_i \equiv \mathbf{Z}_i\mathbf{Z}_i'$$

and

$$\mathbf{S} \equiv \frac{1}{N}\sum_{i=1}^{N}\mathbf{S}_i.$$

The problem is to estimate $\boldsymbol{\theta}_o$ given $(\mathbf{Z}_1, \ldots, \mathbf{Z}_N)$.

Since we assumed existence of the fourth moments, $\mathbf{S}$ satisfies the central limit theorem:

$$\sqrt{N}(vec(\mathbf{S}) - vec(\boldsymbol{\Sigma}(\boldsymbol{\theta}_o))) \rightarrow N(\mathbf{0}, \Delta(\boldsymbol{\theta}_o)),$$

where

$$\Delta(\boldsymbol{\theta}) = \mathbb{V}(vec(\mathbf{S}_i)) = \mathbb{E}vec(\mathbf{S_i})vec(\mathbf{S_i})' - vec(\boldsymbol{\Sigma}(\boldsymbol{\theta}))vec(\boldsymbol{\Sigma}(\boldsymbol{\theta}))' \tag{2}$$

and *vec* denotes vertical vectorization. To save space we will omit the argument of matrix-functions.

It is well known (see, e.g., Magnus and Neudecker, 1988, p. 253) that the multivariate normal distribution satisfies

$$\Delta_o = (\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)(\mathbb{I} + \mathbf{K}) = (\mathbb{I} + \mathbf{K})(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o), \tag{3}$$

where $\otimes$ is the Kronecker product, $\mathbb{I}$ is the identity matrix, $\mathbf{K}$ is the *commutation matrix*, such that $\mathbf{K}\, vec(\mathbf{A}) = vec(\mathbf{A}')$, for any square matrix $\mathbf{A}$. Thus the fourth moments of the multivariate normal distribution are expressed in terms of the second moments.

The normal QML estimator is

$$\hat{\boldsymbol{\theta}}_{\text{QMLE}} = \arg\min_{\theta \in \Theta} \{\log |\boldsymbol{\Sigma}| + \text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1})\}.$$

The optimal GMM estimator of $\boldsymbol{\theta}_o$ is based on the moment conditions

$$\mathbb{E}[\mathbf{m}(\mathbf{Z}_i; \boldsymbol{\theta}_o)] = \mathbf{0}, \tag{4}$$

where $\mathbf{m}(\mathbf{Z}_i; \boldsymbol{\theta}) = vech(\mathbf{S}_i) - vech(\boldsymbol{\Sigma})$ and *vech* denotes vertical vectorization of the lower triangle of a matrix.

The optimal GMM estimator is

$$\hat{\boldsymbol{\theta}}_{\text{GMM}} = \arg\min_{\theta \in \Theta} \{\mathbf{m}_N(\boldsymbol{\theta})'\mathbf{W}\mathbf{m}_N(\boldsymbol{\theta})\},$$

where

$$\begin{aligned} \mathbf{m}_N(\boldsymbol{\theta}) &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{m}(\mathbf{z}_i; \boldsymbol{\theta}) \\ &= vech(\mathbf{S}) - vech(\boldsymbol{\Sigma}), \end{aligned}$$

and the asymptotically optimal weighting matrix is the inverse of the asymptotic variance matrix of the moment functions:

$$\mathbf{W}_o = \{\mathbb{E}[\mathbf{m}(\boldsymbol{Z}_i; \boldsymbol{\theta}_o)\mathbf{m}(\boldsymbol{Z}_i; \boldsymbol{\theta}_o)']\}^{-1}. \tag{5}$$

$\mathbf{W}$ in (5) and $\boldsymbol{\Delta}$ in (2) are connected through the duplication matrix (see, e.g., Magnus and Neudecker, 1988, p. 49). The *duplication* matrix $\mathbf{D}$ is such that $\mathbf{D}\, vech(\mathbf{A}) = vec(\mathbf{A})$. $\mathbf{D}$ transforms *vech* into *vec*, while the Moore-Penrose inverse of $\mathbf{D}$, $\mathbf{D}^+ = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$, transforms *vec* into *vech*. We will use four properties of $\mathbf{D}$ and $\mathbf{D}^+$:

(i) $\mathbf{D}^+\mathbf{D} = \mathbb{I}$;

(ii) $\mathbf{K}\,\mathbf{D} = \mathbf{D}$, where $\mathbf{K}$ is the commutation matrix defined above;

(iii) $\mathbf{D}\,\mathbf{D}^+ = \frac{1}{2}(\mathbb{I} + \mathbf{K})$;

(iv) $(\mathbb{I} + \mathbf{K})\,\mathbf{D} = 2\,\mathbf{D}$ and $\mathbf{D}^+\,(\mathbb{I} + \mathbf{K}) = 2\,\mathbf{D}^+$.

Thus, $\boldsymbol{\Delta} = \mathbb{V}[vec(\mathbf{S}_i)] = \mathbb{V}[\mathbf{D}\, vech(\mathbf{S}_i)] = \mathbf{D}\mathbb{V}[vech(\mathbf{S}_i)]\mathbf{D}'$. But $\mathbb{V}[vech(\mathbf{S}_i)] = \mathbb{E}[\mathbf{m}(\boldsymbol{Z}_i;\boldsymbol{\theta})\mathbf{m}(\boldsymbol{Z}_i;\boldsymbol{\theta})']$. So

$$\mathbf{W}_o = [\mathbf{D}^+\boldsymbol{\Delta}_o\mathbf{D}^{+'}]^{-1}.$$

It is a standard result that, under certain regularity conditions, the normal QMLE and the optimal GMM estimators of $\boldsymbol{\theta}_o$ are consistent and asymptotically normal. See Chamberlain (1984, p. 1289), Newey and McFadden (1994, Theorems 2.6 and 3.4).

3 Asymptotic Analysis

Let $\mathbf{G}(\boldsymbol{\theta})$ denote the *Jacobian* matrix of the moment functions in (4). Then

$$\mathbf{G} \equiv \mathbf{G}(\boldsymbol{\theta}) = \frac{\partial \mathbf{m}(\mathbf{z}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = -\frac{\partial vech(\boldsymbol{\Sigma})}{\partial \boldsymbol{\theta}'}.$$

The following lemmas are used in derivation of the main result of the paper; they are well known and thus given without proof (see, e.g., Chamberlain, 1984; Hansen, 1982).

Lemma 1 *Under regularity conditions, the first order conditions for* $\hat{\boldsymbol{\theta}}_{\text{QMLE}}$ *and* $\hat{\boldsymbol{\theta}}_{\text{GMM}}$ *are, respectively,*

$$\mathbf{G}'\mathbf{D}'(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})^{-1}\mathbf{D}\,[vech(\boldsymbol{S}) - vech(\boldsymbol{\Sigma})] = 0 \tag{6}$$

$$\mathbf{G}'\mathbf{W}^{-1}[vech(\boldsymbol{S}) - vech(\boldsymbol{\Sigma})] = 0. \tag{7}$$

It is clear from (6)-(7) that the only thing that distinguishes the two estimators is the way in which the empirical moments $\mathbf{m}_N(\boldsymbol{\theta})$ are weighted. One way to compare the first order variances of GMM and normal QMLE is to note that $\hat{\boldsymbol{\theta}}_{\text{QMLE}}$ comes from the GMM problem that employs a suboptimal weighting matrix $\mathbf{G}'\mathbf{D}'(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})^{-1}\mathbf{D}$ and is therefore inferior to $\hat{\boldsymbol{\theta}}_{\text{GMM}}$ in terms of first-order relative efficiency unless the weighting matrices are the same. However, this argument cannot be used to derive our equal efficiency condition.

Lemma 2 *Let* $\mathbb{V}$ *denote the asymptotic variance matrix of the relevant estimator, i.e.* $\mathbb{V} =$

$Avar[N^{-\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)]$. *Then, under regularity conditions,*

$$\begin{aligned}
\mathbb{V}_{\mathrm{QMLE}} &= [\mathbf{G}_o'\mathbf{D}'(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)^{-1}\mathbf{D}\mathbf{G}_o]^{-1} \\
&\quad \times \mathbf{G}_o'\mathbf{D}'(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)^{-1}\boldsymbol{\Delta}_o(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)^{-1}\mathbf{D}\mathbf{G}_o \qquad (8)\\
&\quad \times [\mathbf{G}_o'\mathbf{D}'(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)^{-1}\mathbf{D}\mathbf{G}_o]^{-1}, \\
\mathbb{V}_{\mathrm{GMM}} &= [\mathbf{G}_o'(\mathbf{D}^+\boldsymbol{\Delta}_o\mathbf{D}^{+'})^{-1}\mathbf{G}_o]^{-1}. \qquad (9)
\end{aligned}$$

If the data are multivariate normal then the two variance matrices are the same. On using properties of the duplication matrix and equation (3), the following simplifications apply:

$$\begin{aligned}
\mathbf{D}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1}\boldsymbol{\Delta}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1}\mathbf{D} &= \mathbf{D}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1}(\mathbb{I} + \mathbf{K})\mathbf{D} \\
&= 2\,\mathbf{D}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1}\mathbf{D},
\end{aligned}$$

$$\begin{aligned}
\mathbf{D}^+\boldsymbol{\Delta}\mathbf{D}^{+'} &= \mathbf{D}^+(\mathbb{I} + \mathbf{K})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}^{+'} \\
&= 2\,\mathbf{D}^+(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}^{+'}.
\end{aligned}$$

But $[\mathbf{D}^{+'}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}^+]^{-1}$ is equal to $\mathbf{D}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1}\mathbf{D}$ because

$$\begin{aligned}
\mathbf{D}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1}\mathbf{D}\,\mathbf{D}^+(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}^{+'} &= \frac{1}{2}\,\mathbf{D}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1}(\mathbb{I} + \mathbf{K})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}^{+'} \\
&= \frac{1}{2}\,\mathbf{D}'(\mathbb{I} + \mathbf{K})\mathbf{D}^{+'} \\
&= \mathbb{I}.
\end{aligned}$$

It is not immediately clear from the form of (8)-(9) that QMLE is dominated by GMM and under what condition they are equally efficient. The main result is stated in the next theorem.

Theorem 1 *Under the regularity conditions,* $\hat{\boldsymbol{\theta}}_{\mathrm{GMM}}$ *is no less asymptotically efficient than* $\hat{\boldsymbol{\theta}}_{\mathrm{QMLE}}$. *Equal efficiency occurs under the following equivalent conditions:*

(i) $\mathbf{G}_o$ *is in the column space of* $\mathbf{D}^+\boldsymbol{\Delta}_o(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)^{-1}\mathbf{D}\mathbf{G}_o$*;*

(ii) There exists a $\frac{q(q+1)}{2} \times \frac{q(q+1)}{2}$ *matrix* $\mathbb{D}$ *such that*

$$\mathbf{G}_o = \mathbf{D}^+\boldsymbol{\Delta}_o(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)^{-1}\mathbf{D}\mathbf{G}_o\mathbb{D}.$$

Proof. $\mathbb{V}_{\text{QMLE}} - \mathbb{V}_{\text{GMM}}$ is positive semidefinite (PSD) if and only if $\mathbb{V}_{\text{GMM}}^{-1} - \mathbb{V}_{\text{QMLE}}^{-1}$ is PSD. Denote $\mathbf{D}^{+}\boldsymbol{\Delta}_o\mathbf{D}^{+'}$ by $\mathbb{C}$ and $\mathbf{D}'(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)^{-1}\mathbf{D}$ by $\mathbb{A}$. We have

$$\begin{aligned}
\mathbb{V}_{\text{GMM}}^{-1} - \mathbb{V}_{\text{QMLE}}^{-1} &= \mathbf{G}_o'\mathbb{C}^{-1}\mathbf{G}_o - \mathbf{G}_o'\mathbb{A}\mathbf{G}_o[\mathbf{G}_o'\mathbb{A}\mathbb{C}\mathbb{A}\mathbf{G}_o]^{-1}\mathbf{G}_o'\mathbb{A}\mathbf{G}_o \\
&= \mathbf{G}_o'\mathbb{C}^{-\frac{1}{2}}[\mathbb{I} - \mathbb{C}^{\frac{1}{2}}\mathbb{A}\mathbf{G}_o[\mathbf{G}_o'\mathbb{A}\mathbb{C}^{\frac{1}{2}}\mathbb{C}^{\frac{1}{2}}\mathbb{A}\mathbf{G}_o]^{-1}\mathbf{G}_o'\mathbb{A}\mathbb{C}^{\frac{1}{2}}]\mathbb{C}^{-\frac{1}{2}}\mathbf{G}_o.
\end{aligned}$$

This is PSD because the middle part is the idempotent projection matrix onto $\mathbb{C}^{1/2}\mathbb{A}\mathbf{G}_o$. This proves the first part of the theorem.

The difference is zero if and only if $\mathbb{C}^{-1/2}\mathbf{G}_o$ is in the column space spanned by $\mathbb{C}^{1/2}\mathbb{A}\mathbf{G}_o$, or equivalently, $\mathbf{G}_o$ is in the column space of $\mathbb{C}\mathbb{A}\mathbf{G}_o$. Note that

$$\begin{aligned}
\mathbb{C}\mathbb{A}\mathbf{G}_o &= \mathbf{D}^{+}\boldsymbol{\Delta}_o\mathbf{D}^{+'}\mathbf{D}'(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)^{-1}\mathbf{D}\mathbf{G}_o \\
&= \mathbf{D}^{+}\boldsymbol{\Delta}_o\frac{1}{2}(\mathbb{I} + \mathbf{K})(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)^{-1}\mathbf{D}\mathbf{G}_o \\
&= \mathbf{D}^{+}\boldsymbol{\Delta}_o(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)^{-1}\frac{1}{2}(\mathbb{I} + \mathbf{K})\mathbf{D}\mathbf{G}_o \\
&= \mathbf{D}^{+}\boldsymbol{\Delta}_o(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)^{-1}\frac{1}{2}2\mathbf{D}\mathbf{G}_o \\
&= \mathbf{D}^{+}\boldsymbol{\Delta}_o(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)^{-1}\mathbf{D}\mathbf{G}_o.
\end{aligned}$$

This proves both (i) and (ii). □

Theorem 1 is novel in that it states the first order efficiency properties of QMLE and GMM explicitly in terms of the fourth moments of $\mathbf{Z}$ in $\boldsymbol{\Delta}$.

Not surprisingly, the conditions of the theorem hold for the multivariate normal distribution. Using (3), we have

$$\mathbf{D}^{+}\boldsymbol{\Delta}_o(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)^{-1}\mathbf{D}\mathbf{G}_o = \mathbf{D}^{+}(\mathbb{I} - \mathbf{K})\mathbf{D}\mathbf{G}_o = 2\,\mathbf{D}^{+}\mathbf{D}\mathbf{G}_o = 2\,\mathbf{G}_o.$$

So condition (ii) trivially holds. However, there may exist other distributions that satisfy the equal efficiency condition. The following example uses a Student-t distribution with ν degrees of freedom ($\nu > 4$) to show that the condition holds.

Consider the problem of estimating the common variance $\theta_o = \frac{\nu_o}{\nu_o - 2}$ $(0 < \theta_o < 2)$ of two uncorrelated random variables with zero mean:

$$\mathbf{\Sigma} = \theta \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{G} = -\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{D}^{+} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{D}^{+}\mathbf{\Delta}\mathbf{D}^{+'} = \frac{\theta^2}{2-\theta} \begin{bmatrix} 1+\theta & 0 & 1+\theta \\ 0 & 1 & 0 \\ 1+\theta & 0 & 1+\theta \end{bmatrix}, \quad \mathbf{D}'(\mathbf{\Sigma}_o \otimes \mathbf{\Sigma}_o)^{-1}\mathbf{D} = \frac{1}{\theta^2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{D}^{+}\mathbf{\Delta}(\mathbf{\Sigma} \otimes \mathbf{\Sigma})^{-1}\mathbf{D}\mathbf{G} = -\frac{2\theta}{2-\theta} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

The condition of the theorem clearly holds with $\mathbb{D} = \frac{2-\theta_o}{2\theta_o}$. Normal QMLE is efficient. In fact, $\mathbb{V}_{\text{QMLE}} = \mathbb{V}_{\text{GMM}} = \frac{\theta_o^3}{2-\theta_o}$.

4 Concluding Remarks

The paper derives a new condition under which Gaussian QMLE of a general class of covariance structure models preserves its asymptotic optimality property under deviations from the normal distribution. The condition is problem specific so it is hard to say how easy it is to violate it. But it is easy to use as illustrated by the Student-t example.

I compared Gaussian QMLE to optimal GMM but, of course, the same result holds for asymptotic equivalents of optimal GMM such as the empirical likelihood and exponential tilting estimators because their asymptotic variance is identical to GMM.

References

Anderson, T. W. (1989): "Linear latent variable models and covariance structures," *Journal of Econometrics*, 41, 91–119.

ANDERSON, T. W. AND Y. AMEMIYA (1988): "The Asymptotic Normal Distribution of Estimators in Factor Analysis under General Conditions," *The Annals of Statistics*, 16, 759–771.

BROWNE, M. W. (1987): "Robustness of statistical inference in factor analysis and related models," *Biometrika*, 74, 375–384.

BROWNE, M. W. AND A. SHAPIRO (1988): "Robustness of normal theory methods in the analysis of linear latent variable models," *British Journal of Mathematical and Statistical Psychology*, 41, 193–208.

CHAMBERLAIN, G. (1984): "Panel data," in *Handbook of Econometrics*, ed. by Z. Griliches and M. D. Intriligator, vol. II, 1248–1313.

HANSEN, L. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029–1054.

JÖRESKOG, K. G. (1970): "A general method for analysis of covariance structures," *Biometrika*, 57, 239–251.

MAGNUS, J. R. AND H. NEUDECKER (1988): *Matrix differential calculus with applications in statistics and econometrics*, Wiley Series in Probability and Statistics, Chichester: John Wiley and Sons Ltd.

MOOIJAART, A. AND P. BENTLER (1991): "Robustness of normal theory statistics in structural equation models," *Statistica Neerlandica*, 45, 159–171.

NEWEY, W. AND D. MCFADDEN (1994): "Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, ed. by R. Engle and D. McFadden, vol. IV, 2113–2241.

SATORRA, A. AND H. NEUDECKER (1994): "On the Asymptotic Optimality of Alternative Minimum-Distance Estimators in Linear Latent-Variable Models," *Econometric Theory*, 10, 867–883.