

Department of Economics

Working Paper Series

## Second Order Bias of Quasi-MLE for Covariance Structure Models

10-001 | Artem Prokhorov  
Concordia University and CIREQ



UNIVERSITÉ

Concordia

UNIVERSITY

# Second Order Bias of Quasi-MLE for Covariance Structure Models

Artem Prokhorov\*

February 18, 2010

## Abstract

Several recent papers (e.g., Newey et al., 2005; Newey and Smith, 2004; Anatolyev, 2005) derive general expressions for the second-order bias of the GMM estimator and its first-order equivalents such as the EL estimator. Except for some simulation evidence, it is unknown how these compare to the second-order bias of QMLE of covariance structure models. The paper derives the QMLE bias formulas for this general class of models. The bias – identical to the EL second-order bias under normality – depends on the fourth moments of data and remains the same as for EL even for non-normal data so long as the condition for equal asymptotic efficiency of QMLE and GMM derived in Prokhorov (2009) is satisfied.

*JEL Classification:* C13

*Keywords:* (Q)MLE, GMM, EL, Covariance structures.

---

\*Department of Economics, Concordia University, 1455 de Maisonneuve Blvd W, Montreal QC H3G1M8  
Canada; email: artem.prokhorov@concordia.ca

# 1 Motivation and Results

Consider a model formulated in terms of the second moments of the data, i.e. assume that there exists 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  $\mathbb{E}\mathbf{Z} = \mathbf{0}$ ,  $\mathbb{E}\{\|\mathbf{Z}\|^4\} < \infty$  and

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

and expectation is with respect to  $P_{\boldsymbol{\theta}_o}$ .

The measurable real-valued matrix function  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  comes from a structural model such as a factor model, a random effects model, a simultaneous equations model, a conditional expectation model, a LISREL model, etc. The matrix function is such that  $\text{vec}(\boldsymbol{\Sigma})$  is continuous at each  $\boldsymbol{\theta} \in \Theta$ ,  $\text{vec}(\boldsymbol{\Sigma})$  is three times continuously differentiable on a neighborhood of  $\boldsymbol{\theta}_o$  and  $p \leq \frac{1}{2}q(q+1)$ .

For a random sample  $(\mathbf{Z}_1, \dots, \mathbf{Z}_N)$ , where  $\mathbf{Z}_i$  is measured as deviations from the mean, denote

$$\mathbf{S}_i \equiv \mathbf{Z}_i\mathbf{Z}_i' \quad (2)$$

and

$$\mathbf{S} \equiv \frac{1}{N} \sum_{i=1}^N \mathbf{S}_i. \quad (3)$$

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

The fourth moments exist by assumption so, by the central limit theorem,

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

where

$$\Delta(\boldsymbol{\theta}) = \mathbb{V}(\text{vec}(\mathbf{S}_i)) = \mathbb{E}\text{vec}(\mathbf{S}_i)\text{vec}(\mathbf{S}_i)' - \text{vec}(\boldsymbol{\Sigma}(\boldsymbol{\theta}))\text{vec}(\boldsymbol{\Sigma}(\boldsymbol{\theta}))' \quad (4)$$

and  $\text{vec}$  denotes vertical vectorization of a matrix. To save space we will omit the argument of matrix-functions and write  $\boldsymbol{\Sigma}$  instead of  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ ,  $\boldsymbol{\Sigma}_o$  instead of  $\boldsymbol{\Sigma}(\boldsymbol{\theta}_o)$ ,  $\Delta_o$  instead of  $\Delta(\boldsymbol{\theta}_o)$ , etc.

Gaussian (Q)MLE is the traditional estimation method in covariance structure literature (see, e.g., Jöreskog, 1970). It is common to write it as

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

This estimator will be compared with the EL and GMM estimators. The EL estimator is defined as

$$\hat{\boldsymbol{\theta}}_{\text{EL}} = \arg \max_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^N \ln \pi_i, \quad \text{s.t.} \quad \sum_{i=1}^N \pi_i \mathbf{m}(\mathbf{Z}_i; \boldsymbol{\theta}) = 0 \quad \text{and} \quad \sum_{i=1}^N \pi_i = 1,$$

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

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

where

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

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

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

First order asymptotic comparisons of QMLE and GMM are well known (see, e.g., Chamberlain, 1984). Both estimators are consistent under standard assumptions and GMM and its first-order equivalents dominate QMLE in terms of first-order efficiency except under normality as their first-order conditions use optimal weights even under non-normality. It turns out, however, that QMLE of covariance structures preserves its asymptotic efficiency property even under non-normality if certain conditions on higher moments of the data are satisfied (see, e.g., Prokhorov, 2009; Satorra and Neudecker, 1994). In such cases, a comparison of higher order properties is required to rank the estimators. For example, covariance structure QMLE may be preferred to GMM in terms of its second-order bias.

There is a number of simulation-based papers that document significant finite sample biases of the GMM estimator compared with alternative estimators using suboptimal weighting, including the QMLE (see, e.g., Altonji and Segal, 1996; Clark, 1996; Horowitz, 1998).

For instance, Clark (1996) reports simulation results in which Gaussian MLE of covariance structures is unbiased even with non-normal data while the optimal GMM is severely biased.

Aside from the simulation results, second order bias comparison of the QMLE and the GMM first-order equivalents for covariance structure models is not a well studied problem. There are general theoretical results that compare the second order bias of GMM and (Generalized) EL (see Newey and Smith, 2004; Newey et al., 2005; Anatolyev, 2005) and MLE and GMM (see Rilstone et al., 1996) but they are not specialized to covariance structures. Given the QMLE robustness property discussed above, such specialized results would be of value because they may favor the traditional QMLE over the so called asymptotic distribution free covariance structure estimators commonly used in practice (see, e.g., Browne, 1984; Satorra, 1992; Muthen, 1989).

There are some specialized theoretical and simulation-based results on asymptotic bias of MLE and GMM for certain classes of covariance structure models, such as factor models and structural equation models (see, e.g., Ogasawara, 2004, 2005). They suggest that QMLE of some parameters in these types of models possesses an asymptotic robustness property in the sense that its standard errors and first order asymptotic biases do not change under deviations from normality. Example 2 of this paper contains a similar result – it shows that when the QMLE standard errors are robust to deviations from normality, QMLE biases coincide with those of EL. This result is of independent interest – it describes the circumstance in which QMLE is clearly preferred to GMM.

Besides comparisons with other estimators, a specialized expression of QMLE bias for covariance structures permits construction of a bias-corrected QMLE. I do not pursue this point further in this paper although, given the derived bias expression, this is a straightforward exercise. Finally, the stochastic expansion I use allows for the QMLE bias to be expressed in terms of higher order moments of the distribution. This expression is simpler than the one in terms of cumulants, which is available in the literature (see, e.g., Rothenberg, 1984; McCullagh, 1987; Ogasawara, 2005).

Higher order stochastic expansions are based on the Taylor approximation of the first-order

conditions at the true value. The expansions have the following form

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) = \boldsymbol{\mu} + N^{-\frac{1}{2}}\boldsymbol{\tau} + O_p(N^{-1}), \quad (6)$$

where  $\boldsymbol{\mu}$  and  $\boldsymbol{\tau}$  are  $O_p(1)$  random vectors. Since QMLE and the GMM first-order equivalents are  $\sqrt{N}$  consistent, their first-order bias, which can be obtained by taking the expectation of the first term, is zero. Similarly, the first-order variances can be obtained as the expectation of the outer product of the first term. The second-order bias is based on the expectation of the first two terms in (6). Alternatively, the second-order bias can be obtained using the Edgeworth approximation to the distribution as in Rothenberg (1984) and McCullagh (1987).

General expressions for  $\boldsymbol{\mu}$  and  $\boldsymbol{\tau}$  of extremum and minimum distance estimators with many examples can be found in Newey and Smith (2004); Rilstone et al. (1996); Newey et al. (2005); Ullah (2004). For instance, Newey and Smith (2004) in Lemma A4 of Appendix provide a general form of  $\boldsymbol{\mu}$  and  $\boldsymbol{\tau}$  for m-estimators. Derivation of higher order stochastic expansions involves higher order derivatives of the objective functions. Rilstone et al. (1996) use a recursive definition of derivatives. Here, I follow Newey et al. (2005) and Newey and Smith (2004) in using the traditional definition because I do not go to orders higher than two and because I wish to compare the QMLE bias to the GMM and EL bias expressions they derive.

## 2 Proofs and Discussion

Let  $\mathbf{G}(\boldsymbol{\theta})$  denote the *Jacobian* matrix of the  $\frac{1}{2}q(q+1)$  distinct second-order moments in (4) and let  $\mathbf{D}$  denote the *duplication* matrix that transforms *vech* into *vec* (see Magnus and Neudecker, 1988, p. 49). Also, define the Moore-Penrose inverse of  $\mathbf{D}$ ,  $\mathbf{D}^+ = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ . Note that

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

The following lemma is used in derivation of the main results of the paper. It is well known and thus given without proof (see, e.g., Chamberlain, 1984).

**Lemma 2.1** *The first order condition for  $\hat{\boldsymbol{\theta}}_{\text{QMLE}}$  is*

$$\mathbf{G}'\mathbf{D}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1}\mathbf{D}[\text{vec}(\mathbf{S}) - \text{vec}(\boldsymbol{\Sigma})] = 0. \quad (7)$$

In proofs of the main results I follow Newey and Smith (2004) and use an alternative way of writing the first order condition, which circumvents the need to operate with the inverse. Define  $\boldsymbol{\lambda} = -\mathbf{D}'[\boldsymbol{\Sigma}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}(\boldsymbol{\theta})]^{-1}\mathbf{D}\mathbf{m}_N(\boldsymbol{\theta})$ , where  $\mathbf{m}_N(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N \mathbf{m}(\mathbf{Z}_i; \boldsymbol{\theta}) = \text{vech}(\mathbf{S}) - \text{vech}(\boldsymbol{\Sigma})$ . Then the QMLE of  $\boldsymbol{\theta}$  also solves the following equation

$$\mathbf{s}_N(\boldsymbol{\beta}) \equiv \frac{1}{N} \sum_{i=1}^N \mathbf{s}_i(\boldsymbol{\beta}) = 0,$$

where

$$\mathbf{s}_i(\boldsymbol{\beta}) = - \begin{bmatrix} \mathbf{G}'\boldsymbol{\lambda} \\ \mathbf{m}(\mathbf{Z}_i; \boldsymbol{\theta}) + [\mathbf{D}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1}\mathbf{D}]^{-1}\boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \mathbf{G}'\boldsymbol{\lambda} \\ \mathbf{m}(\mathbf{Z}_i; \boldsymbol{\theta}) + [\mathbf{D}^+(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}^+]\boldsymbol{\lambda} \end{bmatrix}.$$

So the QMLE of  $\boldsymbol{\theta}$  is identical to the upper part of the  $(p+q^2)$ -vector  $\boldsymbol{\beta} = (\boldsymbol{\theta}', \boldsymbol{\lambda}')'$  that solves this equation.

Define

$$\begin{aligned} \mathbf{M}_j &= \frac{\partial^2 \mathbf{s}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\beta}_j}, \quad \text{where } \boldsymbol{\beta}_j \text{ is the } j\text{-th element of } \boldsymbol{\beta}, \\ \mathbf{R} &= \{\mathbf{G}'[\mathbf{D}^+(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}^+]^{-1}\mathbf{G}\}^{-1} = [\mathbf{G}'\mathbf{D}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1}\mathbf{D}\mathbf{G}]^{-1}, \\ \mathbf{Q} &= \mathbf{R}\mathbf{G}'[\mathbf{D}^+(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}^+]^{-1} = \mathbf{R}\mathbf{G}'\mathbf{D}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1}\mathbf{D}, \\ \mathbf{P} &= \mathbf{D}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1}\mathbf{D}(\mathbb{I} - \mathbf{G}\mathbf{Q}). \end{aligned} \tag{8}$$

Note that  $\mathbf{M}_j$  does not depend on  $i$  because derivatives of  $\mathbf{m}_i$  are not random. As before, I use subscript  $o$  to denote matrices evaluated at  $\boldsymbol{\beta}_o = (\boldsymbol{\theta}'_o, \mathbf{0}')'$ .

**Theorem 2.1** *The estimator  $\hat{\boldsymbol{\beta}}_{\text{QMLE}}$  satisfies (6) with*

$$\begin{aligned} \boldsymbol{\mu} &= \begin{bmatrix} \mathbf{Q}_o \\ \mathbf{P}_o \end{bmatrix} \frac{1}{\sqrt{N}} \sum_{i=1}^N [\text{vech}(\mathbf{S}_i) - \text{vech}(\boldsymbol{\Sigma}_o)], \\ \boldsymbol{\tau} &= 1/2 \begin{bmatrix} -\mathbf{R}_o & \mathbf{Q}_o \\ \mathbf{Q}'_o & \mathbf{P}_o \end{bmatrix} \sum_{j=1}^{p+q^2} \boldsymbol{\mu}_j \mathbf{M}_{j_o} \boldsymbol{\mu}, \end{aligned} \tag{9}$$

where  $\boldsymbol{\mu}_j$  is the  $j$ -th element of  $\boldsymbol{\mu}$ .

**Proof.** Let  $\bar{\mathbf{M}}(\boldsymbol{\beta}) = \frac{1}{N} \sum_{i=1}^N \frac{\partial \mathbf{s}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}$ ,  $\mathbf{M}(\boldsymbol{\beta}) = \mathbb{E} \frac{\partial \mathbf{s}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}$ ,  $\bar{\mathbf{M}}_j(\boldsymbol{\beta}) = \frac{1}{N} \sum_{i=1}^N \frac{\partial^2 \mathbf{s}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\beta}_j}$  and  $\bar{\boldsymbol{\beta}}$  between  $\hat{\boldsymbol{\beta}}$  and  $\boldsymbol{\beta}_o$ . Note that because  $\frac{\partial \mathbf{s}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}$  is non-random,  $\bar{\mathbf{M}}(\boldsymbol{\beta}) = \mathbf{M}(\boldsymbol{\beta})$ . By the second-

order Taylor expansion of (7) around  $\beta_o$ , we have

$$\begin{aligned}
\mathbf{s}_N(\hat{\beta}) &= \mathbf{0} \\
&= \mathbf{s}_N(\beta_o) + \bar{\mathbf{M}}(\beta_o)(\hat{\beta} - \beta_o) + \frac{1}{2} \sum_{j=1}^{p+q^2} (\hat{\beta}_j - \beta_{oj}) \bar{\mathbf{M}}_j(\bar{\beta})(\hat{\beta} - \beta_o) \\
&= \mathbf{s}_N(\beta_o) + \mathbf{M}(\beta_o)(\hat{\beta} - \beta_o) + [\bar{\mathbf{M}}(\beta_o) - \mathbf{M}(\beta_o)](\hat{\beta} - \beta_o) \\
&\quad + \frac{1}{2} \sum_{j=1}^{p+q^2} (\hat{\beta}_j - \beta_{oj}) \mathbf{M}_j(\beta_o)(\hat{\beta} - \beta_o) + \\
&\quad + \frac{1}{2} \sum_{j=1}^{p+q^2} (\hat{\beta}_j - \beta_{oj}) [\bar{\mathbf{M}}_j(\bar{\beta}) - \mathbf{M}_j(\beta_o)](\hat{\beta} - \beta_o).
\end{aligned}$$

Since  $\bar{\mathbf{M}}(\beta_o) = \mathbf{M}(\beta_o)$ , the third term in the last equation is zero. Also note that the last term is  $O_p(N^{-3/2})$ .

Assume that  $\bar{\mathbf{M}}(\beta_o)$  is not singular. Then,

$$\begin{aligned}
\hat{\beta} - \beta_o &= -[\mathbf{M}(\beta_o)]^{-1} \mathbf{s}_N(\beta_o) \\
&\quad - \frac{1}{2} [\mathbf{M}(\beta_o)]^{-1} \sum_{j=1}^{p+q^2} (\hat{\beta}_j - \beta_{oj}) \mathbf{M}_j(\beta_o)(\hat{\beta} - \beta_o) \\
&\quad + O_p(N^{-3/2}).
\end{aligned} \tag{10}$$

But  $\mathbf{M}(\beta_o) = - \begin{bmatrix} \mathbf{0} & \mathbf{G}'_o \\ \mathbf{G}_o & \mathbf{D}^+(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o) \mathbf{D}^{+'} \end{bmatrix}$ ,  $\mathbf{s}_N(\beta_o) = - \begin{bmatrix} \mathbf{0} \\ \mathbf{m}_N(\boldsymbol{\theta}_o) \end{bmatrix}$  and the second term is  $O_p(N^{-1})$ . It follows that

$$\begin{aligned}
\hat{\beta} - \beta_o &= \frac{1}{\sqrt{N}} \begin{bmatrix} \mathbf{Q}_o \\ \mathbf{P}_o \end{bmatrix} \frac{1}{\sqrt{N}} \sum_{i=1}^N [\text{vech}(\mathbf{S}_i) - \text{vech}(\boldsymbol{\Sigma}_o)] + O_p(N^{-1}) \\
&= \frac{1}{\sqrt{N}} \boldsymbol{\mu} + O_p(N^{-1}).
\end{aligned} \tag{11}$$

Substituting (11) into (10), multiplying by  $\sqrt{N}$  and collecting terms of the same order yields the result.  $\square$



Note that  $\mathbb{E}\boldsymbol{\mu} = 0$  and the first order variance of  $\hat{\boldsymbol{\beta}}_{\text{QMLE}}$  based on (9) can be written as

$$\begin{aligned}\mathbb{E}\boldsymbol{\mu}\boldsymbol{\mu}' &= \begin{bmatrix} \mathbf{Q}_o \\ \mathbf{P}_o \end{bmatrix} \mathbb{E}[\mathbf{m}(\mathbf{Z}_i, \boldsymbol{\theta}_o)\mathbf{m}(\mathbf{Z}_i, \boldsymbol{\theta}_o)'] \begin{bmatrix} \mathbf{Q}_o \\ \mathbf{P}_o \end{bmatrix}' \\ &= \begin{bmatrix} \mathbf{Q}_o\mathbf{C}_o\mathbf{Q}_o' & \mathbf{Q}_o\mathbf{C}_o\mathbf{P}_o' \\ \mathbf{P}_o\mathbf{C}_o\mathbf{Q}_o' & \mathbf{P}_o\mathbf{C}_o\mathbf{P}_o' \end{bmatrix},\end{aligned}\tag{12}$$

where  $\mathbf{C} = \mathbf{D}^+\boldsymbol{\Delta}\mathbf{D}^{+'}$  and I have used the fact that  $\mathbb{E}[\mathbf{m}(\mathbf{Z}_i, \boldsymbol{\theta})\mathbf{m}(\mathbf{Z}_i, \boldsymbol{\theta})'] = \mathbf{D}^+\boldsymbol{\Delta}_o\mathbf{D}^{+'}$ . The upper left  $p \times p$  block of (12) is the traditional expression for the asymptotic variance of  $\hat{\boldsymbol{\theta}}_{\text{QMLE}}$  (see, e.g., Chamberlain, 1984).

Let  $\mathbb{B}$  denote the second order bias of the relevant estimator. Using (6), the bias can be written in terms of the expected value of  $\boldsymbol{\tau}$  as

$$\mathbb{B} = \mathbb{E}\boldsymbol{\tau}/N.$$

Thus, an explicit form of the QMLE bias contains  $\mathbb{E}\boldsymbol{\mu}_j\mathbf{M}_{j_o}\boldsymbol{\mu}$ ,  $j = 1, \dots, p + q^2$ . But  $\mathbf{M}_{j_o}$  can be written as

$$\begin{aligned}\mathbf{M}_{j_o} &= -\frac{\partial^2}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}_j} \left[ \begin{array}{c} \mathbf{G}'\boldsymbol{\lambda} \\ \mathbf{m}(\mathbf{Z}_i, \boldsymbol{\theta}_o) + [\mathbf{D}^+(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}^{+'}]\boldsymbol{\lambda} \end{array} \right] \Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o, \boldsymbol{\lambda}=0} \\ &= \begin{cases} -\begin{bmatrix} \mathbf{0} & \mathbf{G}_o^{j'} \\ \mathbf{G}_o^j & \boldsymbol{\Omega}_o^j \end{bmatrix}, & j = 1, \dots, p \\ -\begin{bmatrix} \mathbf{G}_{j-p,o} & \mathbf{0} \\ \boldsymbol{\Omega}_{j-p,o} & \mathbf{0} \end{bmatrix}, & j = p + 1, \dots, p + q^2 \end{cases}\end{aligned}$$

where  $\mathbf{G}_o^j = \frac{\partial}{\partial\boldsymbol{\theta}_j}\mathbf{G}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o}$ ,  $\mathbf{G}_{j-p,o} = \frac{\partial}{\partial\boldsymbol{\theta}}[\mathbf{G}'\mathbf{e}_{j-p}]\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o}$ ,  $\boldsymbol{\Omega}_o^j = \frac{\partial}{\partial\boldsymbol{\theta}_j}[\mathbf{D}^+(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}^{+'}]\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o}$ ,  $\boldsymbol{\Omega}_{j-p,o} = \frac{\partial}{\partial\boldsymbol{\theta}'}[\mathbf{D}^+(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}^{+'}\mathbf{e}_{j-p}]\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o}$ , and  $\mathbf{e}_{j-p}$  is a  $q^2$ -vector of zeros with the  $(j-p)$ -th element equal to 1. Therefore  $\mathbf{M}_j$  is non-random and we can write

$$\mathbb{E}\boldsymbol{\mu}_j\mathbf{M}_{j_o}\boldsymbol{\mu} = \begin{cases} -\begin{bmatrix} \mathbf{0} & \mathbf{G}_o^{j'} \\ \mathbf{G}_o^j & \boldsymbol{\Omega}_o^j \end{bmatrix} \mathbb{E}\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{e}_j, & j = 1, \dots, p \\ -\begin{bmatrix} \mathbf{G}_{j-p,o} & \mathbf{0} \\ \boldsymbol{\Omega}_{j-p,o} & \mathbf{0} \end{bmatrix} \mathbb{E}\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{e}_j, & j = p + 1, \dots, p + q^2, \end{cases}\tag{13}$$

where  $\mathbf{e}_k$  is a  $p + q^2$ -vector of zeros with the  $k$ -th element equal to 1. Substituting (12) into (13) and simplifying yields the result of the following theorem.

**Theorem 2.2** *The second order bias of  $\hat{\beta}_{\text{QMLE}}$  can be written as follows*

$$\begin{aligned} \mathbb{B}_{\text{QMLE}} = & -\frac{1}{2N} \begin{bmatrix} -\mathbf{R}_o & \mathbf{Q}_o \\ \mathbf{Q}'_o & \mathbf{P}_o \end{bmatrix} \left\{ \sum_{j=1}^p \begin{bmatrix} \mathbf{0} & \mathbf{G}_o^{j'} \\ \mathbf{G}_o^j & \mathbf{\Omega}_o^j \end{bmatrix} \begin{bmatrix} \mathbf{Q}_o \mathbb{C}_o \mathbf{Q}'_o \\ \mathbf{P}_o \mathbb{C}_o \mathbf{Q}'_o \end{bmatrix} \mathbf{e}_j \right. \\ & \left. + \sum_{j=p+1}^{p+q^2} \begin{bmatrix} \mathbf{G}_{j-p,o} \\ \mathbf{\Omega}_{j-p,o} \end{bmatrix} \mathbf{Q}_o \mathbb{C}_o \mathbf{P}'_o \mathbf{e}_{j-p} \right\}, \end{aligned} \quad (14)$$

where  $\mathbb{C}_o = \mathbf{D}^+ \mathbf{\Delta}_o \mathbf{D}^{+'}$  and  $\mathbf{e}_k$  is the zero vector of relevant dimension in which the  $k$ -th element is 1.

Newey and Smith's (2004, Theorems 4.1 and 4.6) second-order bias for the GMM and the EL estimators of  $\theta_o$  can be written as follows

$$\mathbb{B}_{\text{EL}} = -\frac{1}{2N} \mathbf{Q}_o^{\text{EL}} \sum_{j=1}^p \mathbf{G}_o^j \mathbf{R}_o^{\text{EL}} \mathbf{e}_j \quad (15)$$

$$\mathbb{B}_{\text{GMM}} = \mathbb{B}_{\text{EL}} + \frac{1}{N} \mathbf{Q}_o^{\text{EL}} \mathbb{E}[\mathbf{m}(\mathbf{Z}_i, \theta) \mathbf{m}(\mathbf{Z}_i, \theta)' \mathbf{P}_o^{\text{EL}} \mathbf{m}(\mathbf{Z}_i, \theta)] \quad (16)$$

where

$$\begin{aligned} \mathbf{Q}^{\text{EL}} &= \mathbf{R}^{\text{EL}} \mathbf{G}' [\mathbb{E} \mathbf{m}(\mathbf{Z}_i, \theta) \mathbf{m}(\mathbf{Z}_i, \theta)']^{-1} \\ \mathbf{R}^{\text{EL}} &= (\mathbf{G}' [\mathbb{E} \mathbf{m}(\mathbf{Z}_i, \theta) \mathbf{m}(\mathbf{Z}_i, \theta)']^{-1} \mathbf{G})^{-1} \\ \mathbf{P}^{\text{EL}} &= [\mathbb{E} \mathbf{m}(\mathbf{Z}_i, \theta) \mathbf{m}(\mathbf{Z}_i, \theta)']^{-1} (\mathbb{I} - \mathbf{G} \mathbf{Q}^{\text{EL}}) \end{aligned} \quad (17)$$

It is not clear how these compare to  $\mathbb{B}_{\text{QMLE}}$  in general. The examples that follow show several cases when such comparisons are possible. Example 1 shows the obvious point that if the data are normally distributed, the upper block of  $\mathbb{B}_{\text{QMLE}}$  is equal to  $\mathbb{B}_{\text{EL}}$ , while the extra term in  $\mathbb{B}_{\text{GMM}}$  is generally non-zero. This is expected but the derivation of this result is useful because of what follows in Example 2. In Example 2, I deviate from normality and instead consider the condition of equal (first-order) asymptotic efficiency of QMLE and GMM (EL) derived in Prokhorov (2009). It turns out that when the asymptotic efficiency property of QMLE is robust to deviations from normality, that is when Prokhorov's (2009) condition

holds, QMLE's asymptotic bias is also robust to such deviations, that is, QMLE's and EL's biases are identical. Finally, Example 3 shows a situation when the size of QMLE bias is smaller than that of GMM.

**Example 1 – multivariate normality.** In order to show that  $\mathbb{B}_{\text{QMLE}}(\boldsymbol{\theta}_o) = \mathbb{B}_{\text{EL}}$  under normality, recall that for the multivariate normal distribution, the fourth moments can be expressed in terms of the second moments as follows (see, e.g., Magnus and Neudecker, 1988, p. 253)

$$\Delta_o = (\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)(\mathbb{I}_{q^2} + \mathbf{K}_{q^2}) = (\mathbb{I}_{q^2} + \mathbf{K}_{q^2})(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o), \quad (18)$$

where  $\mathbb{I}_k$  is the identity matrix of dimension  $k$ ,  $\mathbf{K}_{m^2}$  is the *commutation matrix*, i.e. such an  $m^2 \times m^2$ -matrix that  $\mathbf{K}_{m^2} \text{vec}(\mathbf{A}) = \text{vec}(\mathbf{A}')$ , for any  $m \times m$  matrix  $\mathbf{A}$ .

Using this fact along with the properties of  $\mathbf{D}^+$  (see, e.g., Magnus and Neudecker, 1988, p. 49), it is easy to show that

$$\begin{aligned} \mathbf{Q}_o \mathbf{C}_o \mathbf{Q}'_o &= 2\mathbf{R}_o, \\ \mathbf{Q}_o \mathbf{C}_o \mathbf{P}'_o &= \mathbf{0}. \end{aligned}$$

Note that this makes the QMLE variance matrix (12) block-diagonal just like its EL counterpart (see, e.g., Qin and Lawless, 1994, Theorem 1).

We can now use these simplifications to rewrite (14) as follows

$$\begin{aligned} \mathbb{B}_{\text{QMLE}} &= -\frac{1}{2N} \begin{bmatrix} -\mathbf{R}_o & \mathbf{Q}_o \\ \mathbf{Q}'_o & \mathbf{P}_o \end{bmatrix} \left\{ \sum_{j=1}^p \begin{bmatrix} \mathbf{0} & \mathbf{G}_o^{j'} \\ \mathbf{G}_o^j & \boldsymbol{\Omega}_o^j \end{bmatrix} \begin{bmatrix} 2\mathbf{R}_o \\ \mathbf{0} \end{bmatrix} \mathbf{e}_j \right\} \\ &= -\frac{1}{2N} \begin{bmatrix} -\mathbf{R}_o & \mathbf{Q}_o \\ \mathbf{Q}'_o & \mathbf{P}_o \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ 2 \sum_{j=1}^p \mathbf{G}_o^j \mathbf{R}_o \mathbf{e}_j \end{bmatrix} \\ &= -\frac{1}{N} \begin{bmatrix} \mathbf{Q}_o \sum_{j=1}^p \mathbf{G}_o^j \mathbf{R}_o \mathbf{e}_j \\ \mathbf{P}_o \sum_{j=1}^p \mathbf{G}_o^j \mathbf{R}_o \mathbf{e}_j \end{bmatrix}. \quad (19) \end{aligned}$$

The upper block of (20) does now look similar to (15) but not identical. The difference is that the expression for  $\mathbb{B}_{\text{QMLE}}$  contains  $\mathbf{D}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1} \mathbf{D}$ , while  $\mathbb{B}_{\text{EL}}$  contains  $\frac{1}{2} \mathbb{E}[\mathbf{m}(\mathbf{Z}_i, \boldsymbol{\theta}) \mathbf{m}(\mathbf{Z}_i, \boldsymbol{\theta})'] = \frac{1}{2} \mathbf{D}^+ \boldsymbol{\Delta}_o \mathbf{D}^{+'}$ . But for the normal distribution,  $\mathbf{D}^+ \boldsymbol{\Delta}_o \mathbf{D}^{+'} = 2\mathbf{D}^+(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}^{+'}$ . If we further

note that  $[\mathbf{D}^{+'}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}^+]^{-1} = \mathbf{D}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1}\mathbf{D}$ , then

$$\begin{aligned}
\mathbf{R}^{\text{EL}} &= \{\mathbf{G}'[2\mathbf{D}^+(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}^{+'}]^{-1}\mathbf{G}\}^{-1} \\
&= 2[\mathbf{G}'\mathbf{D}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1}\mathbf{D}\mathbf{G}]^{-1} \\
&= 2\mathbf{R}, \\
\mathbf{Q}^{\text{EL}} &= \mathbf{R}^{\text{EL}}\mathbf{G}'[2\mathbf{D}^+(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}^{+'}]^{-1} \\
&= \mathbf{R}\mathbf{G}'\mathbf{D}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})^{-1}\mathbf{D} \\
&= \mathbf{Q},
\end{aligned}$$

which confirms that the bias expressions are identical.

Finally, the second term of  $\mathbb{B}_{\text{GMM}}$  contains the third moments of  $\mathbf{m}_i$ , i.e. the sixth moments of  $\mathbf{Z}_i$ . This term is generally non-zero.

**Example 2 – equal asymptotic variance of QMLE and EL.** Prokhorov (2009) shows that the asymptotic variance of QMLE of covariance structures and GMM and its first order equivalents, including EL, is identical under the following condition. Let  $\mathbb{C}_o \equiv \mathbf{D}^+\boldsymbol{\Delta}_o\mathbf{D}^{+'}$  and  $\mathbb{A}_o \equiv \mathbf{D}'(\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o)^{-1}\mathbf{D}$ . Then equal asymptotic efficiency occurs if and only if  $\mathbf{G}_o$  is in the column space of  $\mathbb{C}_o\mathbb{A}_o\mathbf{G}_o$ , i.e., for some  $\frac{q(q+1)}{2} \times \frac{q(q+1)}{2}$  matrix  $\mathbb{D}$ ,  $\mathbf{G}_o = \mathbb{C}_o\mathbb{A}_o\mathbf{G}\mathbb{D}$ . Clearly, in Example 1 this condition holds with  $\mathbb{D} = 2$  because for normal data  $\mathbb{C}_o = 2\mathbb{A}_o$ . Given that the asymptotic weighting matrix used in GMM and its first-order equivalents is  $\mathbf{G}'\mathbb{C}_o^{-1}$ , the condition  $\mathbf{G}'_o\mathbb{C}_o^{-1} = \mathbb{D}'\mathbf{G}'_o\mathbb{A}_o$  basically means that the equations solved by QMLE are asymptotically first-order equivalent to the equations solved by GMM and EL up to a linear transformation – this is why the estimators are equally (first-order) efficient. It turns out that under this condition,  $\mathbb{B}_{\text{QMLE}}(\boldsymbol{\theta}_o) = \mathbb{B}_{\text{EL}}$  so the robustness property of QMLE carries over to its first-order bias.

First, note that, similar to Example 1, this condition implies that  $\mathbf{Q}_o\mathbb{C}_o\mathbf{P}'_o = \mathbf{0}$  and

$\mathbf{Q}_o \mathbf{C}_o \mathbf{Q}' = \mathbb{D}^{-1} \mathbf{R}$ , assuming  $\mathbb{D}$  is not singular. To see this, write omitting the subscript

$$\begin{aligned} \mathbf{QCP}' &= \mathbf{R}[\mathbf{G}'\mathbf{A}\mathbf{C}\mathbf{A} - \mathbf{G}'\mathbf{A}\mathbf{C}\mathbf{A}\mathbf{G}'(\mathbf{G}'\mathbf{A}\mathbf{G})^{-1}\mathbf{A}] \\ &= \mathbf{R}[\mathbb{D}'^{-1}\mathbf{G}'\mathbf{A} - \mathbb{D}'^{-1}\mathbf{G}'\mathbf{A}\mathbf{G}'(\mathbf{G}'\mathbf{A}\mathbf{G})^{-1}\mathbf{A}] = \mathbf{0} \\ \mathbf{QCQ}' &= \mathbf{R}\mathbf{G}'\mathbf{A}\mathbf{C}\mathbf{A}\mathbf{G}\mathbf{R} \\ &= \mathbf{R}\mathbf{G}'\mathbf{A}'\mathbf{G}\mathbb{D}^{-1}\mathbf{R} \\ &= \mathbb{D}^{-1}\mathbf{R}. \end{aligned}$$

Then,  $\mathbb{B}_{\text{QMLE}}$  simplifies as follows:

$$\begin{aligned} \mathbb{B}_{\text{QMLE}} &= -\frac{1}{2N} \begin{bmatrix} -\mathbf{R}_o & \mathbf{Q}_o \\ \mathbf{Q}'_o & \mathbf{P}_o \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \sum_{j=1}^p \mathbf{G}_o^j \mathbf{Q}_o \mathbf{C}_o \mathbf{Q}'_o \mathbf{e}_j \end{bmatrix} \\ &= -\frac{1}{2N} \begin{bmatrix} \mathbf{Q}_o \sum_{j=1}^p \mathbf{G}_o^j \mathbb{D}^{-1} \mathbf{R}_o \mathbf{e}_j \\ \mathbf{P}_o \sum_{j=1}^p \mathbf{G}_o^j \mathbb{D}^{-1} \mathbf{R}_o \mathbf{e}_j \end{bmatrix}. \end{aligned} \quad (20)$$

Finally, to see that  $\mathbb{B}_{\text{QMLE}}(\boldsymbol{\theta}) = \mathbb{B}_{\text{EL}}$ , note that since  $\mathbf{G}'\mathbf{C}^{-1} = \mathbb{D}'\mathbf{G}'\mathbf{A}$ , then  $\mathbf{R}^{\text{EL}} = (\mathbf{G}'\mathbf{C}^{-1}\mathbf{G})^{-1} = \mathbf{R}\mathbb{D}'^{-1}$  and  $\mathbf{Q}^{\text{EL}} = \mathbf{R}^{\text{EL}}\mathbf{G}'\mathbf{C}^{-1} = \mathbf{R}\mathbb{D}'^{-1}\mathbb{D}'\mathbf{G}'\mathbf{A} = \mathbf{Q}$ . The asymptotic bias of QMLE remains equal to that of EL even when the distribution is non-normal so long as the condition of equal first-order efficiency holds.

**Example 3 – Student-t covariates.** Let  $\mathbf{Z}_i$  be correlated realizations from bivariate Student-t distribution with degrees of freedom  $\nu = 7$ . The parameter of interest is the covariance  $\rho$  from

$$\mathbb{V}(\mathbf{Z}_i) = \begin{pmatrix} \frac{7}{5} & \rho \\ \rho & \frac{7}{5} \end{pmatrix},$$

which, given  $\nu$ , takes values in  $[-\frac{\nu}{\nu-2}, \frac{\nu}{\nu-2}] = [-\frac{7}{5}, \frac{7}{5}]$ . This is a situation for which the general comparison from Example 2 does not apply. However, using the expressions for biases of

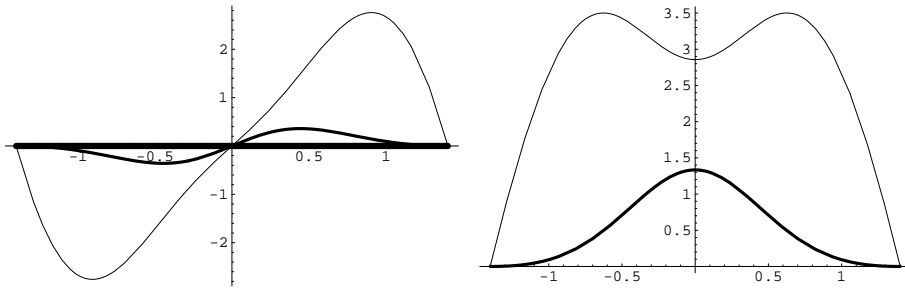


Figure 1: Left panel: second-order bias times sample size for EL (thick), QMLE (thinner), and GMM (thin) in Example 3; Right panel: relative biases, i.e. second-order bias times sample size over true parameter value, for QMLE (thick) and GMM (thin) in Example 3.

QMLE, EL and GMM, it is possible to show the following:<sup>1</sup>

$$\begin{aligned}\mathbb{B}_{\text{EL}} &= 0 \\ N\mathbb{B}_{\text{QMLE}}(\rho) &= \frac{196\rho(49 - 25\rho^2)^2}{3(49 + 25\rho^2)^3} \\ N\mathbb{B}_{\text{GMM}} &= \frac{140\rho(2401 + 4900\rho^2 - 3125\rho^4)}{(343 + 125\rho^2)^2}\end{aligned}$$

Both QMLE and GMM are (second-order) biased in this example and their biases are not the same in general. Figure 1 shows how the size of the biases changes over  $\rho$ . The left panel plots  $\mathbb{B}_{\text{EL}}$  (thick line),  $N\mathbb{B}_{\text{QMLE}}(\rho)$  (thinner line), and  $N\mathbb{B}_{\text{GMM}}$  (thin line), while the right panel plots the relative biases, i.e.  $\frac{N\mathbb{B}_{\text{QMLE}}(\rho)}{\rho}$  (thick line) and  $\frac{N\mathbb{B}_{\text{GMM}}}{\rho}$  (thin line). Several important observations can be made from Figure 1. First, the GMM bias is much more severe than the QMLE bias for all value of  $\rho$  except  $\rho = 0, |\rho| = \frac{7}{5}$ , for which the three biases are equal to zero. The direction of the bias corresponds to the sign of correlation between the covariates. It is zero if the covariates are uncorrelated or if their correlation is one. Second, it is perhaps surprising how small is the QMLE bias compared to GMM. The relative bias of QMLE is several times smaller than GMM and the relative bias of QMLE is larger the closer  $\rho$  is to zero but vanishes as correlation grows.

<sup>1</sup>A mathStatica code deriving these expressions is available at:

[http://alcor.concordia.ca/~aprokhor/papers/student\\_t\\_theta\\_BIAS.nb](http://alcor.concordia.ca/~aprokhor/papers/student_t_theta_BIAS.nb)

## References

- ALTONJI, J. G. AND L. M. SEGAL (1996): “Small-sample bias in GMM estimation of covariance structures,” *Journal of Business and Economic Statistics*, 14, 353–366.
- ANATOLYEV, S. (2005): “GMM, GEL, Serial Correlation, and Asymptotic Bias,” *Econometrica*, 73, 983–1002.
- BROWNE, M. W. (1984): “Asymptotically distribution-free methods for the analysis of covariance structures,” *British Journal of Mathematical and Statistical Psychology*, 37, 62–83.
- CHAMBERLAIN, G. (1984): “Panel data,” in *Handbook of Econometrics*, ed. by Z. Griliches and M. D. Intriligator, vol. II, 1248–1313.
- CLARK, T. E. (1996): “Small-Sample Properties of Estimators of Nonlinear Models of Covariance Structure,” *Journal of Business and Economic Statistics*, 14, 367–73.
- HOROWITZ, J. L. (1998): “Bootstrap methods for covariance structures,” *The Journal of Human Resources*, 33, 39–61.
- 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.
- MCCULLAGH, P. (1987): *Tensor Methods in Statistics*, Monographs on Statistics and Applied Probability, London: Chapman and Hall.
- MUTHEN, B. (1989): “Multiple group structural modelling with nonnormal continuous variables,” *British Journal of Mathematical and Statistical Psychology*, 42, 55–62.
- NEWHEY, W. K., J. S. RAMALHO, AND R. J. SMITH (2005): “Asymptotic Bias for GMM and GEL Estimators with Estimated Nuisance Parameters,” in *Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg*, ed. by D. Andrews and J. Stock, Cambridge University Press, 245–281.
- NEWHEY, W. K. AND R. J. SMITH (2004): “Higher order properties of GMM and Generalized Empirical Likelihood estimators,” *Econometrica*, 72, 219–255.
- OGASAWARA, H. (2004): “Asymptotic biases in exploratory factor analysis and structural equation modeling,” *Psychometrika*, 69, 235–256.
- (2005): “Asymptotic robustness of the asymptotic biases in structural equation modeling,” *Computational Statistics and Data Analysis*, 49, 771–783.

- PROKHOROV, A. (2009): “On relative efficiency of quasi-MLE and GMM estimators of covariance structure models,” *Economics Letters*, 102, 4–6.
- QIN, J. AND J. LAWLESS (1994): “Empirical likelihood and general estimating equations,” *The Annals of Statistics*, 22, 300–325.
- RILSTONE, P., V. SRIVASTAVA, AND A. ULLAH (1996): “The second-order bias and mean squared error of nonlinear estimators,” *Journal of Econometrics*, 75, 369–395.
- ROTHENBERG, T. (1984): “Approximating the distributions of econometric estimators and test statistics,” in *Handbook of Econometrics*, ed. by Z. Griliches and M. D. Intriligator, vol. II, 881–935.
- SATORRA, A. (1992): “Asymptotic robust inferences in the analysis of mean and covariance structures,” in *Sociological Methodology*, ed. by P. Marsden, Oxford, UK: Blackwell, 249–278.
- 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.
- ULLAH, A. (2004): *Finite Sample Econometrics*, Advanced Texts in Econometrics, Oxford University Press.