First and Second Order Asymptotics of Covariance Structure Models*

Artem Prokhorov†

May 28, 2008

Abstract

This paper derives some new first and second order asymptotic properties of well known $\sqrt{N}$-consistent estimators for covariance structure models. Generally, optimal GMM is known to dominate Gaussian QMLE in terms of first order asymptotic efficiency. There are however nontrivial conditions under which Gaussian QMLE preserves its asymptotic optimality property even if the distribution is non-Gaussian. I derive such conditions for a general class of covariance structure models and provide an example when they hold. The conditions can be stated as restrictions on the fourth order moments of the distribution. They trivially hold for normal data but also identify non-normal cases for which Gaussian QMLE is asymptotically efficient. This result supports the much criticized use of traditional Gaussian QMLE in a range of econometric applications that employ such covariance structure models as linear structural relationship (LISREL) models, multiple indicators multiple causes (MIMIC) models, factor analysis and random effect models.

In recent papers, Newey et al (2003, 2004) derived and compared the second order bias of the Emprirical Likelihood estimator and its first order equivalents, such as GMM and Exponential Tilting estimators, for covariance structure models. The GMM estimator has a second order bias that contains more terms than that of the EL estimator and other

*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

estimators of this family. The extra bias terms come from the estimation of the optimal weighting matrix and the derivative matrix that are both parts of the GMM first order conditions. It is unknown how these biases compare to the second order bias of the Gaussian QMLE. I derive the Gaussian QMLE second order bias in a form that allows such a comparison and show that, under normality, the expressions for the EL and QMLE bias are identical. There are several advantages to having a formal proof of these results. First, an explicit form of the QMLE bias is helpful in making comparisons with GMM and EL-type estimators under other distributions than normal. Second, it can be used to explain the finite sample performance of Gaussian QMLE and to construct its bias-corrected version. Finally, I use a higher order stochastic expansion and the bias expression I obtain involves higher order moments of the distribution rather than the cumulants and is thus relatively simple.

JEL Classification: C13

Keywords: GMM, (Q)MLE, EL, Covariance structures, LISREL, MIMIC, efficiency, second-order bias.

1 Introduction

This paper considers estimation of covariance structure models, i.e. models formulated in terms of the second moments of the data. One situation when such models arise is when there are unobserved variables whose presence in the model introduces a particular pattern of correlation between observed variables (e.g., linear structural relationship (LISREL) models, multiple indicators multiple causes (MIMIC) models, factor analysis and random effect models).

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.

This result is given in Section 3. Section 2 describes the general model and the estimators. Section 4 is devoted to derivation of the second order bias of QMLE and its comparison with the empirical likelihood estimator.

Intuitively, one may argue that the comparison is obvious without even looking at the first order conditions the two estimators solve. The bias should be identical. If the true distribution of the data is discrete then MLE and EL are identical estimators. Furthermore, bias terms usually do not depend on discreteness, only on existence of certain moments. So if the assumed distribution (normal, in the case of Gaussian QMLE) turns out to be correct, we should expect the same bias.

There are several advantages to having a formal proof of this intuition. First, an explicit form of the QMLE bias is helpful on its own right. It may be used in making comparisons with

EL under other distributions than normal. Second, it is useful for explaining the finite sample performance of QMLE and for constructing its bias-corrected version. Finally, we use a higher order stochastic expansion and the bias expression we obtain involves higher moments of the distribution rather than cumulants and is thus relatively simple.

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:

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

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

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

The EL estimator is

$$\hat{\boldsymbol{\theta}}_{\text{EL}} = \arg\max_{\theta\in\Theta}\sum_{i=1}^{N}\ln\pi_i$$

subject to

$$\sum_{i=1}^{N}\pi_i\mathbf{m}(\mathbf{z}_i;\boldsymbol{\theta}) = 0$$

and

$$\sum_{i=1}^{N}\pi_i = 1.$$

A GMM estimator is based on the moment condition

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

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

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

It is a standard result that, under certain regularity conditions, the normal QMLE, the optimal GMM and the EL 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), and Owen (2001).

3 First Order 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 3.1 *Under regularity conditions, the first order conditions for* $\hat{\boldsymbol{\theta}}_{\mathrm{QMLE}}$ *and* $\hat{\boldsymbol{\theta}}_{\mathrm{GMM}}$ *are, respectively,*

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

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

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 3.2 *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,*

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

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

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

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

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

This proves both (i) and (ii). □

Theorem 3.1 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 Second Order Analysis

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

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) = \boldsymbol{\mu} + N^{-\frac{1}{2}}\boldsymbol{\tau} + O_p(N^{-1}), \tag{10}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\tau}$ are $O_p(1)$ random vectors.

Since QMLE and EL 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 first term. The second order bias is based on the expectation of the first two terms in (10). 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, e.g., in Rilstone et al. (1996); Ullah (2004). Specialized expressions for the GMM and (generalized) EL can be found in Newey et al. (2005) and Newey and Smith (2004).

Derivation of higher order stochastic expansions involves higher order derivatives of the objective functions. Rilstone et al. (1996) use a recursive definition of derivatives which is useful in general settings. In our derivation we follow Newey and Smith (2004) in using the usual definition because we do not go to the order higher than two and because we wish to compare the QMLE bias to the EL bias expression they derive.

In derivations of results in this section, we will use an alternative way of writing the first order condition (6)-(7), which circumvents the need to operate with the inverse. Define $\boldsymbol{\lambda} = -[\boldsymbol{\Sigma}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}(\boldsymbol{\theta})]^{-1}\mathbf{D}\,\mathbf{m}_N(\boldsymbol{\theta})$, where $\mathbf{m}_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{m}(\mathbf{Z}_i;\boldsymbol{\theta}) = vech(\mathbf{S}) - vech(\boldsymbol{\Sigma})$. Then the QMLE first order condition can be written as

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

where

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

and we now have a $p+q^2$-vector of parameters $\boldsymbol{\beta} = (\boldsymbol{\theta}', \boldsymbol{\lambda}')'$. A similar representation was used by Newey and Smith (2004) for the EL first order condition.

Define

$$\begin{aligned} \mathbf{M}_j &= \frac{\partial^2 \mathbf{s}_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}_j}, \quad \text{where } \boldsymbol{\beta}_j \text{ is the } j\text{-th element of } \boldsymbol{\beta}, \qquad (11) \\ \mathbf{R} &= [\mathbf{G}'\mathbf{D}'(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})^{-1}\mathbf{D}\mathbf{G}]^{-1}, \\ \mathbf{Q} &= \mathbf{R}\mathbf{G}'\mathbf{D}'(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})^{-1}, \\ \mathbf{P} &= (\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})^{-1} - (\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})^{-1}\mathbf{D}\mathbf{G}\mathbf{Q}. \end{aligned}$$

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

Theorem 4.1 *The estimator* $\hat{\boldsymbol{\beta}}_{\mathrm{QMLE}}$ *satisfies (10) with*

$$
\begin{aligned}
\boldsymbol{\mu} &= \begin{bmatrix} \mathbf{Q}_o \\ \mathbf{P}_o \end{bmatrix} \mathbf{D} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [vech(\mathbf{S}_i) - vech(\boldsymbol{\Sigma}_o)], \qquad (12) \\
\boldsymbol{\tau} &= 1/2 \begin{bmatrix} -\mathbf{R}_o & \mathbf{Q}_o \\ \mathbf{Q}_o' & \mathbf{P}_o \end{bmatrix} \sum_{j=1}^{p+q^2} \boldsymbol{\mu}_j \mathbf{M}_{jo} \boldsymbol{\mu},
\end{aligned}
$$

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 $\bar{\boldsymbol{\beta}}$ be 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 (6) around $\boldsymbol{\beta}_o$, we have

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

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

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

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o &= -[\mathbf{M}(\boldsymbol{\beta}_o)]^{-1} \mathbf{s}_N(\boldsymbol{\beta}_o) \\
&\quad - \frac{1}{2}[\mathbf{M}(\boldsymbol{\beta}_o)]^{-1} \sum_{j=1}^{p+q^2} (\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{oj}) \mathbf{M}_j(\boldsymbol{\beta}_o)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) \\
&\quad + O_p(N^{-3/2}). \qquad (13)
\end{aligned}
$$

But $\mathbf{M}(\boldsymbol{\beta}_o) = -\begin{bmatrix} \mathbf{0} & \mathbf{G_o'D'} \\ \mathbf{DG}_o & \boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o \end{bmatrix}$, $\mathbf{s}_N(\boldsymbol{\beta}_o) = -\begin{bmatrix} \mathbf{0} \\ \mathbf{D}\,\mathbf{m}_N(\boldsymbol{\theta}_o) \end{bmatrix}$ and the second term is $O_p(N^{-1})$. We thus have

$$\begin{aligned} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o &= \frac{1}{\sqrt{N}} \begin{bmatrix} \mathbf{Q}_o \\ \mathbf{P}_o \end{bmatrix} \mathbf{D} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [vech(\mathbf{S}_i) - vech(\boldsymbol{\Sigma}_o)] + O_p(N^{-1}) \\ &= \frac{1}{\sqrt{N}} \boldsymbol{\mu} + O_p(N^{-1}). \end{aligned} \tag{14}$$

Substituting (21) into (20), multiplying by $\sqrt{N}$ and collecting terms of the same order yields the result. □

Note that $\mathbb{E}\boldsymbol{\mu} = 0$ and the first order variance of $\hat{\boldsymbol{\beta}}_{\text{QMLE}}$ based on (12) can be written as

$$\begin{aligned} \mathbb{E}\boldsymbol{\mu}\boldsymbol{\mu}' &= \begin{bmatrix} \mathbf{Q}_o \\ \mathbf{P}_o \end{bmatrix} \mathbf{D}\,\mathbb{E}[\mathbf{m}(\mathbf{Z}_i, \boldsymbol{\theta}_o)\mathbf{m}(\mathbf{Z}_i, \boldsymbol{\theta}_o)'] \mathbf{D}' \begin{bmatrix} \mathbf{Q}_o \\ \mathbf{P}_o \end{bmatrix}' \\ &= \begin{bmatrix} \mathbf{Q}_o \boldsymbol{\Delta}_o \mathbf{Q}_o' & \mathbf{Q}_o \boldsymbol{\Delta}_o \mathbf{P}_o' \\ \mathbf{P}_o \boldsymbol{\Delta}_o \mathbf{Q}_o' & \mathbf{P}_o \boldsymbol{\Delta}_o \mathbf{P}_o' \end{bmatrix}, \end{aligned} \tag{15}$$

where the upper left $p \times p$ block of (15) 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 (10), 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 $\mathbb{E}\boldsymbol{\mu}_j \mathbf{M}_{jo} \boldsymbol{\mu}$, $j = 1, \dots, p + q^2$. But $\mathbf{M}_{jo}$ can be written as

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

where $\mathbf{G}_o^j = \frac{\partial}{\partial \boldsymbol{\theta}_j}\mathbf{G}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o}$, $\mathbf{G}_{j-p,o} = \frac{\partial}{\partial \boldsymbol{\theta}}[\mathbf{G}'\mathbf{D}'\mathbf{e}_{j-p}]\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o}$, $\boldsymbol{\Omega}_o^j = \frac{\partial}{\partial \boldsymbol{\theta}_j}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o}$, $\boldsymbol{\Omega}_{j-p,o} = \frac{\partial}{\partial \boldsymbol{\theta}}[(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\mathbf{e}_{j-p}]\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o}$, 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

$$\mathbb{E}\boldsymbol{\mu}_j\mathbf{M}_{jo}\boldsymbol{\mu} = \begin{cases} -\begin{bmatrix} \mathbf{0} & \mathbf{G}_o^{j\prime}\mathbf{D}' \\ \mathbf{D}\mathbf{G}_o^j & \boldsymbol{\Omega}_o^j \end{bmatrix}\mathbb{E}\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{e}_j, & j = 1,\ldots,p \\ -\begin{bmatrix} \mathbf{G}_{j-p,o} & \mathbf{0} \\ \boldsymbol{\Omega}_{j-p,o} & \mathbf{0} \end{bmatrix}\mathbb{E}\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{e}_j, & j = p+1,\ldots,p+q^2, \end{cases} \tag{16}$$

where $\mathbf{e}_k$ is a $p+q^2$-vector of zeros with the k-th element equal to 1. Substituting (15) into (16) and simplifying yields the result of the following theorem.

Theorem 4.2 *The second order bias of $\hat{\boldsymbol{\beta}}_{\mathrm{QMLE}}$ can be written as follows*

$$\begin{aligned} \mathbb{B}_{\mathrm{QMLE}} &= -\frac{1}{2N}\begin{bmatrix} -\mathbf{R}_o & \mathbf{Q}_o \\ \mathbf{Q}_o' & \mathbf{P}_o \end{bmatrix}\left\{\sum_{j=1}^{p}\begin{bmatrix} \mathbf{0} & \mathbf{G}_o^{j\prime}\mathbf{D}' \\ \mathbf{D}\mathbf{G}_o^j & \boldsymbol{\Omega}_o^j \end{bmatrix}\begin{bmatrix} \mathbf{Q}_o\boldsymbol{\Delta}_o\mathbf{Q}_o' \\ \mathbf{P}_o\boldsymbol{\Delta}_o\mathbf{Q}_o' \end{bmatrix}\mathbf{e}_j \right. \\ &\quad \left. + \sum_{j=p+1}^{p+q^2}\begin{bmatrix} \mathbf{G}_{j-p,o} \\ \boldsymbol{\Omega}_{j-p,o} \end{bmatrix}\mathbf{Q}_o\boldsymbol{\Delta}_o\mathbf{P}_o'\mathbf{e}_{j-p}\right\}, \end{aligned} \tag{17}$$

where $\mathbf{e}_k$ is the zero vector of relevant dimension in which the k-th element is 1.

McCullagh (1987) and Linton (1997) give expressions for the second order bias of QMLE in terms of cumulants; our higher-moment representation is simpler and it enables comparison with the second order biases derived in Newey and Smith (2004).

Newey and Smith's (2004, Theorems 4.1 and 4.6) second order bias for the EL estimator of $\boldsymbol{\theta}_o$ is, in our notation,

$$\mathbb{B}_{\mathrm{EL}} = -\frac{1}{2N}\mathbf{Q}_o^{\mathrm{EL}}\sum_{j=1}^{p}\mathbf{G}_o^j\mathbf{R}_o^{\mathrm{EL}}\mathbf{e}_j, \tag{18}$$

where

$$\begin{aligned} \mathbf{Q}^{\mathrm{EL}} &= \mathbf{R}^{\mathrm{EL}}\mathbf{G}'[\mathbb{E}\mathbf{m}(\mathbf{Z}_i,\boldsymbol{\theta})\mathbf{m}(\mathbf{Z}_i,\boldsymbol{\theta})'], \\ \mathbf{R}^{\mathrm{EL}} &= (\mathbf{G}'[\mathbb{E}\mathbf{m}(\mathbf{Z}_i,\boldsymbol{\theta})\mathbf{m}(\mathbf{Z}_i,\boldsymbol{\theta})']^{-1}\mathbf{G})^{-1}. \end{aligned}$$

It is not clear how this compares to $\mathbb{B}_{\mathrm{QMLE}}(\boldsymbol{\theta})$ in general. However, when $\mathbf{Z}$ is multivariate normal, it is relatively easy to show that the upper block of $\mathbb{B}_{\mathrm{QMLE}}$ is equal to (18).

In order to show this final result we make use of two results mentioned before. One result is about stating fourth order moments of the normal distribution in terms of the second order moments, the other result is about properties of the duplication matrix.

Using the results, it is easy to show that

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

Note that this makes the QMLE variance matrix (15) block diagonal just like its EL counterpart (see, e.g., Qin and Lawless, 1994, Theorem 1).

We can now use these simplifications to rewrite (17) as follows

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

The upper block of (19) does now look similar to (18) but not identical. The difference is that the expression for $\mathbb{B}_{\mathrm{QMLE}}$ contains $\mathbf{D}'(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})^{-1}\mathbf{D}$, while $\mathbb{B}_{\mathrm{EL}}$ contains $\frac{1}{2}\mathbb{E}[\mathbf{m}(\mathbf{Z}_i,\boldsymbol{\theta})\mathbf{m}(\mathbf{Z}_i,\boldsymbol{\theta})']$.

We proceed by showing that, regardless of the distribution, we can write $\mathbb{E}[\mathbf{m}(\mathbf{Z}_i,\boldsymbol{\theta})\mathbf{m}(\mathbf{Z}_i,\boldsymbol{\theta})']$ as $[\mathbf{D}^{+}\boldsymbol{\Delta}_o\mathbf{D}^{+'}]$. This is because $\boldsymbol{\Delta} = \mathbb{V}[vec(\mathbf{S}_i)] = \mathbb{V}[\mathbf{D}\,vech(\mathbf{S}_i)] = \mathbf{D}\mathbb{V}[vech(\mathbf{S}_i)]\mathbf{D}'$. But $\mathbb{V}[vech(\mathbf{S}_i)] = \mathbb{E}[\mathbf{m}(\mathbf{Z}_i,\boldsymbol{\theta})\mathbf{m}(\mathbf{Z}_i,\boldsymbol{\theta})']$. Now for the normal distribution, the two facts above imply that $[\mathbf{D}^{+}\boldsymbol{\Delta}_o\mathbf{D}^{+'}]$ can be written as $2\mathbf{D}^{+}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\mathbf{D}^{+'}$.

We further note that $[\mathbf{D}^{+'}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\mathbf{D}^{+}]^{-1}$ is equal to $\mathbf{D}'(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})^{-1}\mathbf{D}$ as discussed in Section

3. We can therefore write

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

which confirms that the bias expressions are identical.

Note the relationship between the proof and the intuition mentioned above: the proof establishes equality of certain moments in the EL bias expression with those of QMLE, while the intuition stresses that the EL bias terms do not depend on discreteness and so the equivalence of discrete MLE and EL should carry over to continuous distributions.

5 Concluding Remarks

The paper derives new properties of three known consistent estimators – GMM, MLE, EL. Specifically, it is argued that the Gaussian MLE preserves its asymptotic efficiency property even if the data are non-normal, provided certain conditions hold. The conditions are 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.

In this paper we compared Gaussian QMLE to optimal GMM in terms of the first order asymptotics 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.

The main second order asymptotic result of the paper is the expression for the MLE bias. It may be useful in many settings including finite sample studies, but we use it here to show that it is identical to the bias expression for the EL estimator under normality.

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.

Linton, O. (1997): "An asymptotic expansion in the GARCH(1,1) model," *Econometric Theory*, 13, 558.

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.

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.

Newey, 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.

Newey, W. K. and R. J. Smith (2004): "Higher order properties of GMM and Generalized Empirical Likelihood estimators," *Econometrica*, 72, 219–255.

Owen, A. B. (2001): *Empirical likelihood*, Monographs on statistics and applied probability; 92, Boca Raton, Fla. : Chapman and Hall.

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 statictics," in *Handbook of Econometrics*, ed. by Z. Griliches and M. D. Intriligator, vol. II, 881–935.

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.

A Proofs

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 $\bar{\boldsymbol{\beta}}$ be 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 (6) around $\boldsymbol{\beta}_o$, we have

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

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

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

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o &= -[\mathbf{M}(\boldsymbol{\beta}_o)]^{-1}\mathbf{s}_N(\boldsymbol{\beta}_o) \\
&\quad - \frac{1}{2}[\mathbf{M}(\boldsymbol{\beta}_o)]^{-1}\sum_{j=1}^{p+q^2}(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{oj})\mathbf{M}_j(\boldsymbol{\beta}_o)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) \\
&\quad + O_p(N^{-3/2}). \qquad (20)
\end{aligned}
$$

But $\mathbf{M}(\boldsymbol{\beta}_o) = -\begin{bmatrix} \mathbf{0} & \mathbf{G_o'D'} \\ \mathbf{DG}_o & \boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o \end{bmatrix}$, $\mathbf{s}_N(\boldsymbol{\beta}_o) = -\begin{bmatrix} \mathbf{0} \\ \mathbf{D}\,\mathbf{m}_N(\boldsymbol{\theta}_o) \end{bmatrix}$ and the second term is $O_p(N^{-1})$. We thus have

$$\begin{aligned} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o &= \frac{1}{\sqrt{N}} \begin{bmatrix} \mathbf{Q}_o \\ \mathbf{P}_o \end{bmatrix} \mathbf{D} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [vech(\mathbf{S}_i) - vech(\boldsymbol{\Sigma}_o)] + O_p(N^{-1}) \\ &= \frac{1}{\sqrt{N}} \boldsymbol{\mu} + O_p(N^{-1}). \end{aligned} \tag{21}$$

Substituting (21) into (20), multiplying by $\sqrt{N}$ and collecting terms of the same order yields the result.