Journal of Econometrics 42 (1989) 247–265. North-Holland

ON THE EFFICIENT ESTIMATION OF SIMULTANEOUS EQUATIONS WITH COVARIANCE RESTRICTIONS*

Manuel ARELLANO

Institute of Economics and Statistics, Oxford OX1 3UL, England

Received May 1987, final version received June 1988

In a simultaneous equations model with general covariance restrictions, the QML estimator that imposes the covariance restrictions may be less efficient than the unrestricted QML estimator. A sufficient condition on the fourth-order moments is given for the relative inefficiency of the restricted QMLE. The relative inefficiencies of both QML estimators are compared with the optimal minimum distance (MD) estimator and computationally convenient augmented NLLS and QML estimators are proposed that are asymptotically efficient. In the case of separate nonlinear restrictions on the structural covariance matrix, separate estimators of slope and covariance parameters are proposed that are asymptotically equivalent to the joint optimal MD estimator.

1. Introduction

Linear models whose specification includes second-moment restrictions in addition to restrictions on the first moments appear in a variety of contexts. This category includes models that can be represented as a simultaneous equations system with general covariance restrictions. An example of this situation is a dynamic regression from panel data observed over a fixed number of periods [cf. Bhargava and Sargan (1983)] where error components, heteroskedasticity over time, and serial correlation introduce possibly nonlinear restrictions in the serial covariance matrix of the structural errors. In addition, some models specify cross-restrictions linking the structural covariances and the slope parameters, as it is the case in certain expectational and errors-in-variables models. In these problems a popular method of estimation is normal quasi-maximum likelihood (QML). There are both computational and statistical reasons that make QML an attractive choice. It has been shown that QML estimators are consistent and asymptotically normal under fairly general conditions [e.g. Sargan (1975), Gourieroux, Monfort, and Trognon (1984)]. However, this paper argues that when the covariance restrictions are

*This paper is a revised and shortened version of part of Arellano (1986). I wish to thank J.D. Sargan for his help, advice, and encouragement. I am also grateful to an anonymous referee and to an Associate Editor for helpful comments. Responsibility for any errors is mine alone.

not required for identification, so that the unrestricted QMLE is defined, the covariance restricted QMLE and the unrestricted QMLE cannot be ordered and hence enforcing covariance restrictions by QML methods may result in an efficiency loss. We give explicit conditions on the fourth-order moments which are sufficient for the relative inefficiency of the restricted QMLE.

Since the QMLE is asymptotically equivalent to a nonoptimal minimum distance estimator (MDE), our analysis is simplified by conducting it in the MD framework [MD estimators have been considered in detail by Rothenberg (1973) and Chamberlain (1982)]. The asymptotic covariance matrix of the optimal MDE is also given and this provides a lower bound against which the relative efficiency of alternative estimators may be ranked. Using an idea suggested by Hausman, Newey, and Taylor (1987) (HNT henceforth), we find computationally convenient to regard the optimal MDE as a nonlinear least squares estimator (NLLS) in an augmented multivariate regression; on the same lines, an asymptotically equivalent augmented QMLE is also given consideration. Finally, it is shown that in a model where structural covariance parameters are unrelated to slope parameters, separate MDE of covariance parameters based on unrestricted estimates of the *structural* covariance matrix are efficient. This enables us to obtain an efficient separate estimator of slope parameters under nonlinear covariance constraints from an extended 3SLS criterion. The relationship between this estimator and the augmented 3SLS (A3SLS) estimator of HNT for linear covariance constraints is discussed: A3SLS is based on a concentrated GMM criterion while our estimator for nonlinear covariance constraints is based on a conditional GMM criterion. Proofs and standard results employed in the paper are collected in an appendix.

2. The model and the estimators

y_i is an $n \times 1$ vector of endogenous variables and z_i is a $k \times 1$ vector of exogenous variables on which N observations are available. We assume that the first and second conditional moments of y_i are given by

$$\mathrm{E}(y_i|z_i) = \mu + P(\theta)z_i, \tag{1}$$

$$\mathrm{E}(v_i v_i'|z_i) = \Omega(\theta), \tag{2}$$

where

$$v_i = y_i - \mu - P(\theta)z_i = y_i - \Pi z_i^+, \tag{3}$$

and the elements of the $n \times k$ matrix $P(\theta)$ and the $n \times n$ matrix $\Omega(\theta)$ are continuous functions of a $q \times 1$ vector of identified parameters θ; μ is a $n \times 1$ vector of intercepts, $\Pi = (\mu \,\vdots\, P)$, and $z_i^+ = [1 \quad z_i']'$. In addition,

$\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N} z_i^+ z_i^{+\prime} = M^+$ exists and is nonsingular. The errors v_i are assumed to be independent and identically distributed with finite moments up to the fourth order. Letting $m = n(n+1)/2$, we may define

$$\Lambda_3 = \mathrm{E}[v_i v_i' \otimes v_i' | z_i] L',$$

$$\Lambda_4 = L\,\mathrm{E}(v_i v_i' \otimes v_i v_i' | z_i) L',$$

where Λ_3 and Λ_4 are respectively $n \times m$ and $m \times m$ matrices of third- and fourth-order moments, and L is a $m \times n^2$ selection matrix that eliminates from Λ_3 and Λ_4 some of the repeated cross-moments.[1]

When the model has a simultaneous equations representation, then

$$P(\theta) = -B^{-1}(\theta)\Gamma(\theta), \tag{4a}$$

$$\Omega(\theta) = B^{-1}(\theta)\Sigma(\theta)B'^{-1}(\theta), \tag{4b}$$

$$\mu = -B^{-1}(\theta)\gamma, \tag{4c}$$

where $B(\theta)$ is a nonsingular $n \times n$ matrix, $\Sigma(\theta)$ is the structural covariance matrix, and γ is the $n \times 1$ vector of structural constant terms. Similar relationships can be written for higher-order moments. The following one, relating structural and reduced-form fourth-order moments, will be used below. Let $\omega = \nu(\Omega)$, $\sigma = \nu(\Sigma)$, and let F be an $m \times m$ matrix of the form

$$F = F(\theta) = L(B^{-1} \otimes B^{-1})D,$$

so that $\omega = F\sigma$. Then

$$\Lambda_4 = F\Delta_4 F', \tag{4d}$$

where Δ_4 is the $m \times m$ matrix of structural fourth-order moments.

The unconstrained least squares estimators of Π and Ω are given by

$$\hat{\Pi} = (\hat{\mu} \,\vdots\, \hat{P}) = (Y'Z^+)(Z^{+\prime}Z^+)^{-1},$$

$$\hat{\Omega} = \left[Y'Y - Y'Z^+(Z^{+\prime}Z^+)^{-1}Z^{+\prime}Y\right]\Big/N,$$

[1] The following conventions are adopted: For any matrix B, vec(B) is obtained by stacking the rows of B. For an $n \times n$ symmetric matrix A, $\nu(A)$ is the $n(n+1)/2$ column vector obtained stacking by rows the lower triangle of A. $\nu(A)$ and vec(A) can be connected by mean of a $n^2 \times n(n+1)/2$ duplication matrix D that maps $\nu(A)$ into vec(A), i.e., $D\nu(A) = \text{vec}(A)$. Furthermore, since $(D'D)$ is nonsingular we also have $\nu(A) = L\,\text{vec}(A)$ with $L = (D'D)^{-1}D'$. The properties of D and L are extensively studied in Magnus and Neudecker (1980).

where

$$Z^{+\prime}Z^{+}=\sum_{i=1}^{N} z_i^{+} z_i^{+\prime},\quad Y'Z^{+}=\sum_{i=1}^{N} y_i z_i^{+\prime},\quad Y'Y=\sum_{i=1}^{N} y_i y_i'.$$

Let $\hat{\pi}=\operatorname{vec}(\hat{\Pi})$ and $\hat{\omega}=\nu(\hat{\Omega})$. The estimators of θ whose properties are discussed below are based on these two statistics. The following lemma states an asymptotic normality result for $\hat{\pi}$ and $\hat{\omega}$.

Lemma. Under the assumptions of our model, $\sqrt{N}(\hat{\pi}-\pi)$ *and* $\sqrt{N}(\hat{\omega}-\omega)$ *have a joint limiting normal distribution with mean zero and covariance matrix with partition given by*

$$V_{\pi\pi}=\Omega\otimes M^{+-1},\quad V_{\pi\omega}=(I\otimes d_1)\Lambda_3,\quad V_{\omega\omega}=\Lambda_4-\omega\omega',$$

where $\pi=\operatorname{vec}(\Pi)$ *and* d_1 *is a* $(k+1)\times 1$ *vector with one in the first position and zero elsewhere.*

Proof. See appendix A.1.

Under normality, $\Lambda_3=0$ and

$$\Lambda_4-\omega\omega'=2L(\Omega\otimes\Omega)L'=A_{\omega\omega},$$

say.[2] The expression for $V_{\pi\omega}$ in the lemma implies that the slope estimators $\hat{P}$ and $\hat{\Omega}$ are asymptotically uncorrelated, and their joint distribution is invariant to departures from symmetry. An explicit expression for $\hat{P}$ is provided by the least squares regression in deviations from sample means

$$\hat{P}=(Y'Z)(Z'Z')^{-1},$$

where $Z'=[z_1\ z_2\ \dots\ z_N]Q$, $Q=I_N-\iota\iota'/N$, and ι is a $N\times 1$ vector of ones. Therefore as a corollary of the lemma we can write

$$\sqrt{N}\begin{pmatrix}\hat{p}-p\\ \hat{\omega}-\omega\end{pmatrix}\xrightarrow{\mathrm{d}} N\left[0,\begin{pmatrix}V_{pp} & 0\\ 0 & V_{\omega\omega}\end{pmatrix}\right],$$

where $\hat{p}=\operatorname{vec}(\hat{P})$, $V_{pp}=\Omega\otimes M^{-1}$, and $M=\operatorname{plim}(Z'Z/N)$. Thus, third-order moments are irrelevant in simultaneous equations with covariance restrictions

[2] That is, the elements of Λ_4 are of the form

$$\lambda_{hjkl}=\omega_{hj}\omega_{kl}+\omega_{hk}\omega_{jl}+\omega_{hl}\omega_{jk},$$

where ω_{hj} is the (h,j) element of Ω.

except when the structural constant terms are restricted. In this paper, we make the simplifying assumption that intercepts are unrestricted[3] and concentrate on the analysis of estimators of θ based on $\hat{p}$ and $\hat{\omega}$. Note that in this case the joint estimation of μ and θ would produce the same estimators of θ, due to the fact that the addition of unrestricted moments does not alter the minimum distance estimator [cf. Chamberlain (1982, p. 45)].

In order to squeeze estimates of the parameters of interest θ from $\hat{p}$ and $\hat{\omega}$ we consider minimizing distance functions of the form

$$s(\theta) = [\hat{p} - p(\theta)]'Q_1[\hat{p} - p(\theta)] + [\hat{\omega} - \omega(\theta)]'Q_2[\hat{\omega} - \omega(\theta)] + 2[\hat{p} - p(\theta)]'Q_3[\hat{\omega} - \omega(\theta)]. \tag{5}$$

Following a general principle (see appendix A.2), the optimal MD estimator $\hat{\theta}_{\mathrm{OMD}}$ uses consistent estimates of V_{pp}^{-1} and $V_{\omega\omega}^{-1}$ as the choice for the weighting matrices Q_1 and Q_2, respectively, and sets $Q_3 = 0$. $\hat{\theta}_{\mathrm{OMD}}$ is asymptotically normal with variance matrix C_{OMD} given by

$$C_{\mathrm{OMD}}^{-1} = H'V_{pp}^{-1}H + K'V_{\omega\omega}^{-1}K, \tag{6}$$

where

$$H = (\partial p/\partial \theta') \quad \text{and} \quad K = (\partial \omega/\partial \theta').^4$$

A distance function similar to (5) is discussed by Rothenberg (1973, p. 81) for the normal case.

If covariance restrictions are not needed for the identification of θ, one can choose to leave Ω unrestricted. The concentrated optimal distance function with respect to Ω is just

$$s^*(\theta) = [\hat{p} - p(\theta)]'V_{pp}^{-1}[\hat{p} - p(\theta)],$$

which defines Malinvaud's (1970) MD estimator $\hat{\theta}_u$. $\hat{\theta}_u$ has asymptotic variance matrix (avm)

$$C_u = \left(H'V_{pp}^{-1}H\right)^{-1}, \tag{7}$$

and it is well known to be asymptotically equivalent to 3SLS, Ω unrestricted QML, and other simultaneous equations estimators [cf. Hendry (1976)]. From

[3]For example, if the model represents a relationship for panel data over n time periods, this implies that time dummies would be included for all periods.

[4]All functions of θ are evaluated at true values unless stated otherwise, except when referring to an estimation criterion in which θ indicates the argument of the function.

expression (6) it is clear that $C_{\mathrm{OMD}}^{-1} - C_u^{-1}$ is positive semidefinite so that $\hat{\theta}_{\mathrm{OMD}}$ is generally efficient relative to $\hat{\theta}_u$.

A popular alternative to enforce the covariance restrictions is the Ω-restricted QML estimator, $\hat{\theta}_{\mathrm{QML}}$, that minimizes

$$l(\theta) = \log\det \Omega(\theta) + \operatorname{tr}\left(\Omega^{-1}(\theta)\left[\hat{\Omega} + (\hat{P} - P(\theta))\left(\frac{Z'Z}{N}\right)(\hat{P} - P(\theta))'\right]\right), \tag{8}$$

where $l(\theta)$ is (minus) the quasi-log-likelihood concentrated with respect to the vector of constant terms. The $\hat{\theta}_{\mathrm{QML}}$ has been shown to be asymptotically equivalent to the MD estimator that uses V_{pp}^{-1}, $A_{\omega\omega}^{-1}$, and 0 as the choice for Q_1, Q_2, and Q_3, respectively [cf. Chamberlain (1982)].[5] The optimal MD estimator is asymptotically equivalent to the (quasi) MLE when the errors are normal but in general, using again results from the appendix A.2, the avm of $\hat{\theta}_{\mathrm{QML}}$ can be obtained as

$$C_{\mathrm{QML}} = \left(C_u^{-1} - K'A_{\omega\omega}^{-1}K\right)^{-1}\left(C_u^{-1} + K'A_{\omega\omega}^{-1}V_{\omega\omega}A_{\omega\omega}^{-1}K\right) \times \left(C_u^{-1} + K'A_{\omega\omega}^{-1}K\right)^{-1}.$$

Thus $\hat{\theta}_{\mathrm{QML}}$ is inefficient relative to $\hat{\theta}_{\mathrm{OMD}}$ unless $V_{\omega\omega} = A_{\omega\omega}$. More worryingly, Ω-restricted QMLE and Ω-unrestricted QMLE cannot be ranked in terms of asymptotic efficiency, so that it is not clear whether any efficiency gain will ensue from enforcing the covariance constraints by minimizing $l(\theta)$. This can be readily seen by rewriting C_{QML} as[6]

$$C_{\mathrm{QML}} = C_u + C_u K'(A_{\omega\omega} + KC_uK')^{-1}KC_u + C_uK'(A_{\omega\omega} + KC_uK')^{-1}\kappa_4(A_{\omega\omega} + KC_uK')^{-1}KC_u, \tag{9}$$

[5] The inverse of $A_{\omega\omega}$ is given by

$$A_{\omega\omega}^{-1} = \tfrac{1}{2}D'(\Omega^{-1} \otimes \Omega^{-1})D,$$

[cf. Richard (1975)].

[6] Eq. (9) is obtained by using the 'matrix inversion lemma'. We use

$$\left(C_u^{-1} + K'A_{\omega\omega}^{-1}K\right)^{-1} = C_u - C_uK'(A_{\omega\omega} + KC_uK')^{-1}KC_u$$

and

$$A_{\omega\omega}^{-1}K\left(C_u^{-1} + K'A_{\omega\omega}^{-1}K\right)^{-1} = (A_{\omega\omega} + KC_uK')^{-1}KC_u.$$

where κ_4 is the $m \times m$ matrix of fourth-order cumulants given by

$$\kappa_4 = \Lambda_4 - A_{\omega\omega} - \omega\omega'.$$

That is, if $\kappa_4 = 0$, as it is the case under normality, then $C_u - C_{\mathrm{QML}} > 0$, but in general the direction of this inequality is uncertain as it will depend on the relative size of the two last terms on the RHS of (9). We state this result in the following proposition.

Proposition. $\hat{\theta}_{\mathrm{QML}}$ *is efficient relative to* $\hat{\theta}_u$ *if and only if*

$$J\left[\left(A_{\omega\omega} + KC_uK'\right) - \kappa_4\right]J' > 0,$$

where

$$J = C_uK'\left(A_{\omega\omega} + KC_uK'\right)^{-1}.$$

Therefore a sufficient condition is

$$\Lambda_4 < 4L(\Omega \otimes \Omega)L' + \omega\omega' + KC_uK'.$$

As a simple illustration of the previous discussion let us consider the problem of estimating α and β from the model

$$y_{1i} = \gamma + \alpha y_{2i} + \beta z_{1i} + u_{1i},$$

$$y_{2i} = \mu_0 + \mu_1 z_{1i} + \mu_2 z_{2i} + u_{2i},$$

$$\begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix} \sim \mathrm{iid}\left[0, \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}\right],$$

where $Z = (z_1\, z_2)Q$ is a $N \times 2$ nonstochastic matrix in deviations from sample means such that M is positive definite in the notation above. Notice that we assume the random vectors $\{(u_{1i}u_{2i}), i = 1, \ldots, N\}$ to be independently and identically distributed with a possibly nonnormal bivariate distribution, so that although u_{1i} and u_{2i} are assumed to be uncorrelated they are not necessarily independent. Letting $\delta = (\alpha\ \beta)'$ and $X = (y_1\ y_2)$, the QML estimator of δ that leaves (σ_{ij}) unrestricted is the IV estimator

$$\hat{\delta}_{\mathrm{IV}} = (Z'X)^{-1}Z'y_1,$$

while the QML estimator that takes into account the restriction $\sigma_{12} = 0$ is the

OLS estimator

$$\hat{\delta}_{\mathrm{OLS}} = (X'X)^{-1}X'y_1.$$

In general, $\hat{\delta}_{\mathrm{IV}}$ and $\hat{\delta}_{\mathrm{OLS}}$, although both consistent, cannot be ordered and neither of the two is optimal. It can be shown that $\mathrm{avm}(\hat{\delta}_{\mathrm{OLS}}) < \mathrm{avm}(\hat{\delta}_{\mathrm{IV}})$ if and only if [7]

$$\mathrm{E}\left(u_{1i}^2 u_{2i}^2\right) < \left(2 + \sigma_{11} m^{11}/\mu_1^2\right)\sigma_{11}\sigma_{22},$$

where m^{11} is the (1,1) element of M^{-1}. Clearly, in this example fourth-order independence of u_{1i} and u_{2i} suffices for the previous condition to hold and OLS to be asymptotically efficient.

As another illustration, the relative efficiency of FIML with respect to 3SLS when the structural covariance matrix is diagonal was first shown by Rothenberg and Leenders (1964). Since 3SLS is asymptotically equivalent to the QMLE that leaves Ω unrestricted, the proposition delimits the scope of this result when the errors are not necessarily normally distributed.

3. Computing efficient estimators from an augmented multivariate regression

HNT have suggested to augment the original structural equation system by equations involving a linearization of the covariance restrictions around an initial consistent estimator of θ, and to estimate the resulting system by joint 3SLS (augmented 3SLS or A3SLS). This method can be used when there are no cross-restrictions linking slope and covariance coefficients in the structural form and the structural covariance constraints are linear. A3SLS is particularly attractive when the covariance restrictions are not required for identification, since in this case any IV estimator can be used to linearize the restrictions and the A3SLS can be computed without iteration, provided the restrictions in B and Γ are linear. In section 4, we propose an estimator of the structural slope coefficients under nonlinear covariance restrictions which is efficient in the absence of constraints linking slopes and covariances, and we also discuss the relationship of this estimator to the A3SLS of HNT. In this section, we rewrite the 'reduced form' model (1)–(3) in the form of an augmented multivariate regression with unrestricted covariance matrix, and consider conventional (unrestricted covariance) estimators of the augmented system. The nonlinear least squares estimator (NLLS) of the augmented system is the optimal MD estimator of the previous section and the QML estimator is an asymptotically equivalent alternative. It may be useful to regard efficient

[7] For a detailed discussion of this example, see Arellano (1986).

estimators as augmented NLLS or QML of multivariate regressions because computer packages for estimators of this type are often available.

Let us consider the augmented multivariate regression

$$y_i = \mu + P(\theta) z_i + v_i, \tag{10a}$$

$$w_i = \omega(\theta) + \varepsilon_i, \tag{10b}$$

where $w_i = \nu(v_i v_i')$, and the ε_i are $m \times 1$ vectors of errors such that $\mathrm{E}(\varepsilon_i | z_i) = 0$ and

$$\operatorname{var}\left[\begin{pmatrix} v_i \\ \varepsilon_i \end{pmatrix} \middle| z_i \right] = \Omega^* = \begin{bmatrix} \Omega & \Lambda_3 \\ \Lambda_3' & \Lambda_4 - \omega\omega' \end{bmatrix}. \tag{11}$$

The unrestricted least squares estimator of ω is given by

$$\tilde{\omega} = N^{-1} \sum_{i=1}^{N} \nu(v_i v_i').$$

Since $\hat{\omega}$ is asymptotically equivalent to $\tilde{\omega}$, the w_i can be replaced by the observable variables $\hat{w}_i = \nu(\hat{v}_i \hat{v}_i')$, where $\hat{v}_i = y_i - \hat{\mu} - \hat{P} z_i$ with no loss of asymptotic information about θ. In effect, straightforward algebra reveals that the optimal MD criterion function of the form given in (5) satisfies

$$\begin{aligned}
&[\hat{p} - p(\theta)]' \left(\hat{\Omega}^{-1} \otimes \frac{Z'Z}{N} \right) [\hat{p} - p(\theta)] \\
&+ [\hat{\omega} - \omega(\theta)]' \left(\hat{\Lambda}_4 - \hat{\omega}\hat{\omega}' \right)^{-1} [\hat{\omega} - \omega(\theta)] \\
&= \frac{1}{N} \sum_{i=1}^{N} \Big\{ (y_i^* - P(\theta) z_i^*)' \hat{\Omega}^{-1} (y_i^* - P(\theta) z_i^*) \\
&\quad + (\nu[\hat{v}_i^* \hat{v}_i^{*\prime} - \Omega(\theta)])' \left(\hat{\Lambda}_4 - \hat{\omega}\hat{\omega}' \right)^{-1} \nu[\hat{v}_i^* \hat{v}_i^{*\prime} - \Omega(\theta)] \Big\} - (n+m),
\end{aligned} \tag{12}$$

where y_i^*, z_i^*, and $\hat{v}_i^*$ are sample mean deviations of y_i, z_i, and $\hat{v}_i$, respectively, and

$$\hat{\Lambda}_4 = N^{-1} L \sum_{i=1}^{N} (\hat{v}_i^* \hat{v}_i^{*\prime} \otimes \hat{v}_i^* \hat{v}_i^{*\prime}) L'.$$

Therefore $\hat{\theta}_{\text{OMD}}$ is the NLLS estimator of θ in the set of equations

$$y_i^* = P(\theta) z_i^* + v_i^*,$$

$$\nu(\hat{v}_i^* \hat{v}_i^{*\prime}) = \omega(\theta) + \hat{\varepsilon}_i^*,$$

which uses the inverse of $\operatorname{diag}(\hat{\Omega}, \hat{\Lambda}_4 - \hat{\omega}\hat{\omega}')$ as the norm. If we wish to estimate efficiently the constant terms together with θ, we can use NLLS in (10a) and (10b) after replacing w_i by $\hat{w}_i$ and using an estimator of Ω^* as the norm.

On the other hand, we can use the normal log-likelihood as an alternative distance function. Letting $V' = (v_1 \dots v_N)$ and $\hat{E}' = (\hat{\varepsilon}_1 \dots \hat{\varepsilon}_N)$ with $\hat{\varepsilon}_i = \hat{w}_i - \omega(\theta)$, the quasi-log-likelihood function, apart from an irrelevant constant, is given by

$$l_a = -\frac{N}{2} \log \det \Omega^* - \tfrac{1}{2} \operatorname{tr}\left[\Omega^{*-1} \begin{pmatrix} V'V & V'\hat{E} \\ \hat{E}'V & \hat{E}'\hat{E} \end{pmatrix} \right],$$

and concentrating Ω^* out of l_a, we obtain

$$l_a^* = -\frac{N}{2} \log \det \left(\frac{V'V}{N} \right) - \frac{N}{2} \log \det \frac{1}{N} \left(\hat{E}'\hat{E} - \hat{E}'V(V'V)^{-1} V'\hat{E} \right), \tag{13}$$

where

$$\left(\frac{V'V}{N} \right) = \hat{\Omega} + [\hat{\Pi} - \Pi(\mu, \theta)] \left(\frac{Z^{+\prime} Z^{+}}{N} \right) [\hat{\Pi} - \Pi(\mu, \theta)]', \tag{13a}$$

$$\left(\frac{\hat{E}'\hat{E}}{N} \right) = (\hat{\Lambda}_4 - \hat{\omega}\hat{\omega}') + [\hat{\omega} - \omega(\theta)][\hat{\omega} - \omega(\theta)]', \tag{13b}$$

$$\left(\frac{V'\hat{E}}{N} \right) = \hat{\Lambda}_3 + [\hat{\Pi} - \Pi(\mu, \theta)] N^{-1} \sum_{i=1}^{N} z_i^{+} [\hat{w}_i - \omega(\theta)]'. \tag{13c}$$

$\hat{\Lambda}_3$ and $\hat{\Lambda}_4$ are sample counterparts of Λ_3 and Λ_4, respectively, based on unrestricted least squares residuals $\hat{v}_i$.

The minimizer of l_a^*, $\hat{\theta}_{\text{AQML}}$, is asymptotically equivalent to $\hat{\theta}_{\text{OMD}}$. Augmented NLLS and QML estimators can be computed without modification when the covariance restrictions are required in order to identify the structural parameters, because the reduced form residuals can always be constructed. However, in some cases the restrictions in the reduced form may be unnecessarily complicated if a simpler augmented equation system can be specified in terms of the structural form. Some of these cases are discussed in the next section.

4. Separate restrictions on the structural covariance matrix

For the purpose of the discussion below, it is convenient to represent an optimal MD criterion function of the type given in (5) as an extended 3SLS criterion function. If the covariance restrictions are not needed for identification, we can obtain consistent estimates of the structural covariance and fourth-order moment matrices, $\tilde{\Sigma}$ and $\tilde{\Delta}_4$, respectively. Replacing $\hat{\omega}$ by $\tilde{\omega} = \nu(V'V/N)$ and using as our 'estimates' of V_{pp} and $V_{\omega\omega}$,

$$\tilde{V}_{pp} = B^{-1}\tilde{\Sigma}B'^{-1} \otimes (Z'Z/N)^{-1},$$

$$\tilde{V}_{\omega\omega} = F(\tilde{\Delta}_4 - \tilde{\sigma}\tilde{\sigma}')F',$$

where $\tilde{\sigma} = \nu(\tilde{\Sigma})$, and V, B, and F are regarded as functions of θ, we obtain a new criterion function given by

$$\begin{aligned} s^+(\theta) &= [\hat{p} - p(\theta)]'\tilde{V}_{pp}^{-1}[\hat{p} - p(\theta)] + [\tilde{\omega} - \omega(\theta)]'\tilde{V}_{\omega\omega}^{-1}[\tilde{\omega} - \omega(\theta)] \\ &= N^{-1}[\mathrm{vec}(U')]'\left[\tilde{\Sigma}^{-1} \otimes Z(Z'Z)^{-1}Z'\right]\mathrm{vec}(U') \\ &\quad + \left[\nu\left(\Sigma - \frac{U'U}{N}\right)\right]'\left(\tilde{\Delta}_4 - \tilde{\sigma}\tilde{\sigma}'\right)^{-1}\nu\left(\Sigma - \frac{U'U}{N}\right), \end{aligned} \tag{14}$$

where U' is the $n \times N$ matrix of structural errors in deviations from sample means whose ith column is $u_i = By_i^* + \Gamma z_i^*$. Note that the first term of $s^+(\theta)$ is a standard 3SLS criterion. Moreover, note that $s^+(\theta)$ is also an optimal generalized method-of-moments (GMM) criterion of the form

$$s^+(\theta) = \left(\frac{1}{N}\sum_{i=1}^{N}\zeta_i\right)'\hat{\psi}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\zeta_i\right),$$

where

$$\zeta_i = \begin{pmatrix} u_i \otimes z_i^* \\ L(u_i \otimes u_i) - \sigma \end{pmatrix}$$

and

$$\hat{\psi} = \mathrm{diag}\left(\tilde{\Sigma} \otimes Z'Z/N, \tilde{\Delta}_4 - \tilde{\sigma}\tilde{\sigma}\right).$$

Let us suppose we have $B = B(\theta_1)$, $\Gamma = \Gamma(\theta_1)$, and $\Sigma = \Sigma(\theta_2)$, where $(\theta_1 \; \theta_2)$ is a partition of θ and the parameter space takes the form $\Theta_1 \times \Theta_2$. There are several models which conform to this pattern, including the classical simultaneous equations system with zero covariance restrictions and certain models for panel data. Interest centers in general on the estimation of the slope coefficients θ_1, although the testing of the restrictions in $\Sigma(\theta_2)$ may also be of importance. We assume that the restrictions in Σ are not required for the identification of θ_1. In these models Ω still depends on both θ_1 and θ_2, so an efficiency gain may be expected by using an optimal joint criterion to estimate θ_1. Let

$$s^{+}(\theta_1, \theta_2) = s_1^{+}(\theta_1) + s_2^{+}(\theta_1, \theta_2),$$

where the two terms of the RHS correspond to those in (14), and let θ_2 be $q_2 \times 1$. If the restrictions in Σ are linear, so that $\sigma = G\theta_2$ or $S\sigma = 0$, where G and S are $m \times q_2$ and $(m - q_2) \times m$ matrices of known constants, respectively, such that $SG = 0$, θ_2 can be concentrated out of $s^{+}(\theta_1, \theta_2)$ thus obtaining the HNT GMM criterion function

$$s^{++}(\theta_1) = s_1^{+}(\theta_1) + [S\nu(U'U/N)]'\left[S\left(\tilde{\Delta}_4 - \tilde{\sigma}\tilde{\sigma}'\right)S'\right]^{-1} S\nu(U'U/N). \tag{15}$$

The A3SLS estimator of Hausman, Newey, and Taylor minimizes a criterion similar to $s^{++}(\theta_1)$ where $S\nu(U'U/N)$ has been replaced by its first-order Taylor expansion about a consistent estimate of θ_1.

On the other hand, in the absence of cross-restrictions linking B and Γ to Σ, it is possible to define alternative MD estimators of θ_2 based on unrestricted efficient estimates of Σ. This section shows that the estimators of θ_2 obtained in this way are asymptotically efficient and therefore can be inserted in the joint distance function, thus obtaining a simplified conditional criterion for θ_1 without loss of efficiency, which can still be used with nonlinear covariance restrictions. Note that if a consistent but inefficient estimate of θ_2 is inserted in the joint criterion function, the resulting conditional estimate of θ_1 will be inefficient, since the joint Hessian matrix is not block-diagonal between θ_1 and θ_2 (see appendix A.4 for details).

If Σ is left unrestricted, Ω is also unrestricted, hence the unconstrained MD estimator of Σ is just a transformation of $\hat{\theta}_{1u}$ and $\hat{\Omega}$:

$$\hat{\Sigma} = B\left(\hat{\theta}_{1u}\right)\hat{\Omega}B'\left(\hat{\theta}_{1u}\right), \tag{16}$$

where $\hat{\theta}_{1u}$ is $\hat{\theta}_u$ in the notation of section 2.[8] In appendix A.3 it is shown that the avm of $\hat{\sigma} = \nu(\hat{\Sigma})$ is given by

$$V_{\sigma\sigma} = \Delta_4 - \sigma\sigma' + F^{-1}K_1\left(H_1'V_{pp}^{-1}H_1\right)^{-1}K_1'F'^{-1}, \tag{17}$$

and that the avm of the joint optimal MD estimator of θ_1 and θ_2 given in (6) can be partitioned as

$$\operatorname{avm}^{-1}\left(\hat{\theta}_{1,\mathrm{OMD}}\right)$$

$$= H_1'V_{pp}^{-1}H_1 + K_1'V_{\omega\omega}^{-1}\left(V_{\omega\omega} - K_2\left(K_2'V_{\omega\omega}^{-1}K_2\right)^{-1}K_2'\right)V_{\omega\omega}^{-1}K_1 \tag{18a}$$

and

$$\operatorname{avm}^{-1}\left(\hat{\theta}_{2,\mathrm{OMD}}\right) = G'V_{\sigma\sigma}^{-1}G, \tag{18b}$$

where

$$H_1 = \left(\partial p/\partial\theta_1'\right), \quad K_j = \left(\partial\omega/\partial\theta_j'\right), \quad G = \left(\partial\sigma/\partial\theta_2'\right), \qquad j = 1,2.$$

The following distance function defines an estimator of θ_2 based on $\hat{\Sigma}$:

$$d(\theta_2) = \left[\hat{\sigma} - \sigma(\theta_2)\right]'\hat{V}_{\sigma\sigma}^{-1}\left[\hat{\sigma} - \sigma(\theta_2)\right], \tag{19}$$

where $\hat{V}_{\sigma\sigma}$ is a consistent estimator of $V_{\sigma\sigma}$. Let $\bar{\bar{\theta}}_2$ be the minimizer of $d(\theta_2)$. Following the general principle in appendix A.2, avm($\bar{\bar{\theta}}_2$) turns out to be the same as avm($\hat{\theta}_{2,\mathrm{OMD}}$) and therefore $\bar{\bar{\theta}}_2$ is asymptotically efficient. This is not surprising given the fact that $\bar{\bar{\theta}}_2$ is based on an efficient estimator of the structural covariance matrix Σ in the absence of restrictions, and that there are no cross-restrictions linking slope and covariance coefficients. Incidentally, remark that replacing $\hat{V}_{\sigma\sigma}$ in (19) by a consistent estimate of

$$A_{\sigma\sigma} = 2L(\Sigma \otimes \Sigma)L' + F^{-1}K_1\left(H_1'V_{pp}^{-1}H_1\right)^{-1}K_1'F'^{-1}$$

defines a nonoptimal estimator of θ_2 that is asymptotically equivalent to the QMLE.

The previous result suggests to consider an estimating criterion for θ_1 of the same form as $s^+(\theta_1, \theta_2)$ but in which $\bar{\bar{\theta}}_2$ replaces θ_2:

$$r(\theta_1) = s_1^+(\theta_1) + \left[\nu\left(\bar{\bar{\Sigma}} - U'U/N\right]'\left(\tilde{\Delta}_4 - \tilde{\sigma}\tilde{\sigma}'\right)^{-1}\nu\left(\bar{\bar{\Sigma}} - U'U/N\right), \tag{20}$$

[8]An alternative asymptotically equivalent estimator of Σ is the unrestricted QMLE

$$\hat{\Sigma}_{\mathrm{QML}} = A\left(\hat{\theta}_{1u}\right)\left(X'X/N\right)A'\left(\hat{\theta}_{1u}\right),$$

where $A(\theta) = [B(\theta) \vdots \Gamma(\theta)]$ and $X = (QY \vdots Z)$. In the discussion below, $\hat{\Sigma}_{\mathrm{QML}}$ could be used in place of $\hat{\Sigma}$. $\hat{\theta}_{1u}$ is also indistinctly the unrestricted MD, 3SLS, or unrestricted QML estimator of θ_1.

where $\bar{\bar{\Sigma}} = \Sigma(\bar{\bar{\theta}}_2)$. Since $\bar{\bar{\theta}}_2$ is efficient, the minimizer of $r(\theta_1)$, $\tilde{\theta}_1$, is asymptotically equivalent to $\hat{\theta}_{1,\mathrm{OMD}}$. $\tilde{\theta}_1$ is easier to compute than $\hat{\theta}_{1,\mathrm{OMD}}$ because it does not require the simultaneous calculation of estimates for θ_2.

Finally, since Σ-restricted estimates of θ_1 are inconsistent when invalid covariance constraints are enforced, it is important to test for the validity of the restrictions. It is certainly possible to use Hausman specification tests based on the difference $(\hat{\theta}_{1u} - \tilde{\theta}_1)$, but since $Nd(\bar{\bar{\theta}}_2) \xrightarrow{d} \chi^2(m - q_2)$ [cf. Chamberlain (1982)] we can use minimum chi-square tests of the covariance constraints without having to calculate $\tilde{\theta}_1$. This is particularly useful when the precise form of the covariance restrictions is uncertain and several sets of restrictions should be tested.

5. Conclusions

This paper has compared the relative inefficiencies of the QML estimator that imposes covariance restrictions in a simultaneous equations model and the unrestricted QML estimator, with respect to the optimal minimum distance estimator.

It is found that the QMLE that makes use of the covariance restrictions may be less efficient than the unrestricted QMLE. This is the case if the fourth-order moments of the errors are large enough relative to the variances and covariances. A sufficient condition is given for the relative inefficiency of the restricted QMLE. Distributions with long tails are common in practice due to the presence of extreme observations in the sample, and they lead to large standardized fourth-order moments.

Computationally convenient augmented NLLS and QML estimators are proposed which are asymptotically equivalent to the optimal MD estimator. In the case of separate nonlinear restrictions on the structural covariance matrix, it is shown that there are available separate estimators of slope and covariance parameters that are asymptotically equivalent to the joint optimal MD estimator.

Appendix

A.1. Proof of the lemma

Let $V = Y - Z^+\Pi'$. Then $\hat{\Pi} - \Pi = V'Z^+(Z^{+\prime}Z^+)^{-1}$ and

$$\hat{\Omega} = \frac{V'V}{N} - \frac{V'Z^+}{N}\left(\frac{Z^{+\prime}Z^+}{N}\right)^{-1}\frac{Z^{+\prime}V}{N} = \frac{V'V}{N} + o_p(N^{-1/2}).$$

These expressions can be written as

$$\sqrt{N}\,(\hat{\pi}-\pi)=\left[I_n\otimes\left(Z^{+\prime}Z^{+}/N\right)^{-1}\right]N^{-1/2}\sum_{i=1}^{N}\left(v_i\otimes z_i^{+}\right)$$

and

$$\sqrt{N}\,(\hat{\omega}-\omega)=N^{-1/2}\sum_{i=1}^{N}\left[L(v_i\otimes v_i)-\omega\right]+\mathrm{o}_p(1).$$

Letting $\xi_{1i}=v_i\otimes z_i^{+}$ and $\xi_{2i}=L(v_i\otimes v_i)-\omega$, it follows that $\mathrm{E}(\xi_{1i}\xi_{1i}')=\Omega\otimes z_i^{+}z_i^{+\prime}$, $\mathrm{E}(\xi_{1i}\xi_{2i}')=(I_n\otimes z_i^{+})\Lambda_3$, and $\mathrm{E}(\xi_{2i}\xi_{2i}')=\Lambda_4-\omega\omega'$. Since $\xi_i=(\xi_{1i}'\,\xi_{2i}')'$ are independently distributed random vectors, the Liapunov central limit theorem implies that $N^{-1/2}\sum_{i=1}^{N}\xi_i$ is asymptotically normal with mean zero and covariance matrix given by

$$\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}\mathrm{E}(\xi_i\xi_i')=\begin{pmatrix}\Omega\otimes M^{+} & (I_n\otimes m^{+})\Lambda_3\\ \Lambda_3'(I_n\otimes m^{+\prime}) & \Lambda_4-\omega\omega'\end{pmatrix}.$$

Finally, using the Cramer linear transformation theorem

$$\sqrt{N}\begin{pmatrix}\hat{\pi}-\pi\\ \hat{\omega}-\omega\end{pmatrix}\xrightarrow{\mathrm{d}}N\left[0,\begin{pmatrix}\Omega\otimes M^{+-1} & \left(I_n\otimes M^{+-1}m^{+}\right)\Lambda_3\\ \Lambda_3'\left(I_n\otimes m^{+\prime}M^{+-1}\right) & \Lambda_4-\omega\omega'\end{pmatrix}\right],$$

where

$$m^{+}=\operatorname{plim}N^{-1}\sum_{i=1}^{N}z_i^{+}.$$

$M^{+-1}m^{+}$ is in fact a vector with one in the first position and zero elsewhere; this can be easily seen noting that $M^{+-1}m^{+}=\operatorname{plim}(Z^{+\prime}Z^{+})^{-1}Z^{+\prime}\iota$, where ι is a $N\times 1$ vector of ones. But when $Z^{+}=(\iota\,\vdots\,Z_1)$, since $(Z^{+\prime}Z^{+})^{-1}Z^{+\prime}(\iota\,\vdots\,Z_1)=I_{k+1}$, we have $(Z^{+\prime}Z^{+})^{-1}Z^{+\prime}\iota=(1\,0\ldots0)'\doteq d_1$.

A.2. The limiting distribution of the MD estimator

Let $\hat{p}$ be an unconstrained estimator of the $s\times 1$ coefficient vector p such that

$$\operatorname*{plim}_{N\to\infty}\hat{p}=p, \tag{A.1}$$

$$\sqrt{N}\,(\hat{p}-p)\xrightarrow{\mathrm{d}}N(0,V_p). \tag{A.2}$$

Assume that p depends on a set of constraint parameters δ, $p = p(\delta)$. We further assume that $p(\delta^*) = p(\delta)$ for some δ^* in the parameter space implies that $\delta^* = \delta$, and that $p(\delta)$ has continuous second partial derivatives in a neighbourhood of δ. It is also assumed that $D = \partial p(\delta)/\partial \delta'$ has full column rank.

Let $\tilde{\delta}$ be the minimiser of the distance function

$$s(\delta) = [\hat{p} - p(\delta)]'\hat{Q}[\hat{p} - p(\delta)], \tag{A.3}$$

where $\hat{Q}$ is an $s \times s$ matrix such that $\operatorname{plim} \hat{Q} = Q$ exists and is positive definite. (A.1) and our identification assumptions ensure the consistency of $\tilde{\delta}$ for δ. By the definition of $\tilde{\delta}$, $\partial s(\tilde{\delta})/\partial \delta = 0$ so that using a first-order expansion about δ it is straightforward to show that

$$\sqrt{N}(\tilde{\delta} - \delta) \xrightarrow{d} N(0, V_\delta),$$

where

$$V_\delta = (D'QD)^{-1}(D'QV_pQD)(D'QD)^{-1}. \tag{A.4}$$

Clearly, an optimal choice for Q is V_p^{-1}, in which case the avm of $\tilde{\delta}$ reduces to

$$V_\delta = \left(D'V_p^{-1}D\right)^{-1}. \tag{A.5}$$

A.3. The avm of the optimal MDE of slope and covariance parameters under separate structural covariance restrictions

Let us consider a partition in C_{OMD}^{-1} with blocks given by

$$C^{11} = H_1'V_{pp}^{-1}H_1 + K_1'V_{\omega\omega}^{-1}K_1,$$

$$C^{12} = K_1'V_{\omega\omega}^{-1}K_2,$$

$$C^{22} = K_2'V_{\omega\omega}^{-1}K_2.$$

Using formulae for partitioned inverses

$$\begin{aligned} &\mathrm{avm}^{-1}(\hat{\theta}_{1,\mathrm{OMD}}) \\ &= C^{11} - C^{12}(C^{22})^{-1}C^{21} \\ &= H_1'V_{pp}^{-1}H_1 + K_1'V_{\omega\omega}^{-1}\left[V_{\omega\omega} - K_2\left(K_2'V_{\omega\omega}^{-1}K_2\right)^{-1}K_2'\right]V_{\omega\omega}^{-1}K_1. \end{aligned} \tag{A.6}$$

Note that $K_2 = FG$ and $V_{\omega\omega} = F(\Delta_4 - \sigma\sigma')F'$. If Σ is left unrestricted $\theta_2 = \sigma$ and $G = I$ so that

$$C^{22} = \left(\Delta_4 - \sigma\sigma'\right)^{-1},$$

$$C^{21} = \left(\Delta_4 - \sigma\sigma'\right)^{-1} F^{-1} K_1,$$

and the second term of the RHS of (A.6) vanishes.

On the other hand,

$$\begin{aligned} &\operatorname{avm}\left(\hat{\theta}_{2,\mathrm{OMD}}\right) \\ &= \left(C^{22}\right)^{-1} + \left(C^{22}\right)^{-1} C^{21}\left(C^{11} - C^{12}\left(C^{22}\right)^{-1} C^{21}\right)^{-1} C^{12}\left(C^{22}\right)^{-1}, \end{aligned}$$

thus

$$\operatorname{avm}(\hat{\sigma}) = V_{\sigma\sigma} = \left(\Delta_4 - \sigma\sigma'\right) + F^{-1} K_1\left(H_1' V_{pp}^{-1} H_1\right)^{-1} K_1' F'^{-1}. \tag{A.7}$$

Finally, since it is also true that

$$\begin{aligned} &\operatorname{avm}^{-1}\left(\hat{\theta}_{2,\mathrm{OMD}}\right) \\ &= C^{22} - C^{21}\left(C^{11}\right)^{-1} C^{12} \\ &= K_2'\left[V_{\omega\omega}^{-1} - V_{\omega\omega}^{-1} K_1\left(H_1' V_{pp}^{-1} H_1 + K_1' V_{\omega\omega}^{-1} K_1\right)^{-1} K_1' V_{\omega\omega}^{-1}\right] K_2, \end{aligned}$$

using the matrix inversion lemma

$$= K_2'\left[V_{\omega\omega} + K_1\left(H_1' V_{pp}^{-1} H_1\right)^{-1} K_1'\right]^{-1} K_2 = G' V_{\sigma\sigma}^{-1} G.$$

A.4. Efficient criterion functions and two-stage estimation

Let $s_N(\theta)$, $\theta \in \Theta$, be an estimator criterion subject to usual regularity and identification conditions [i.e., $\operatorname{plim} N^{-1} s_N(\theta) = s_\infty(\theta)$ uniformly in θ, and $s_\infty(\theta)$ attains a global minimum at $\bar{\theta}$, the true value of θ]. We also assume:

(i) $\quad \operatorname{plim}_{N \to \infty}\left(-N^{-1} \partial^2 s_N(\bar{\theta}) / \partial\theta\,\partial\theta'\right) = A$, a positive definite matrix,

(ii) $\quad N^{-1/2} \partial s_N(\bar{\theta}) / \partial\theta \xrightarrow{\mathrm{d}} N(0, B)$.

$s_N(\theta)$ is said to be an efficient criterion if $A = B$, in which case

$$\sqrt{N}\,(\hat{\theta} - \bar{\theta}) \xrightarrow{\mathrm{d}} N(0, A^{-1}),$$

where $\hat{\theta}$ is the minimizer of $s_N(\theta)$.

Let us consider a partition $\hat{\theta} = (\hat{\theta}_1' \; \hat{\theta}_2')'$ and accordingly $A = \{A_{ij}\}$, $A^{-1} = \{A^{ij}\}$, $i, j = 1, 2$. Suppose that an alternative consistent estimator of $\bar{\theta}_2$ is available, $\tilde{\theta}_2$ say. Then we can consider an alternative criterion for θ_1 alone defined as

$$c_N(\theta_1) = s_N(\theta_1, \tilde{\theta}_2).$$

Let $\tilde{\theta}_1$ be the minimizer of $c_N(\theta_1)$. We wish to investigate the conditions under which $\tilde{\theta}_1$ is asymptotically equivalent to $\hat{\theta}_1$. A first-order Taylor expansion of $\partial c_N(\tilde{\theta}_1)/\partial\theta_1$ about $\bar{\theta}_1$ and $\bar{\theta}_2$ gives

$$\begin{aligned} 0 = N^{-1/2}\frac{\partial c_N(\tilde{\theta}_1)}{\partial\theta_1} &= N^{-1/2}\frac{\partial s_N(\tilde{\theta}_1, \tilde{\theta}_2)}{\partial\theta_1} \\ &= N^{-1/2}\frac{\partial \bar{s}_N}{\partial\theta_1} + \left(N^{-1}\frac{\partial^2 \bar{s}_N}{\partial\theta_1\,\partial\theta_1'}\right)N^{1/2}(\tilde{\theta}_1 - \bar{\theta}_1) \\ &\quad + \left(N^{-1}\frac{\partial^2 \bar{s}_N}{\partial\theta_1\,\partial\theta_2'}\right)N^{1/2}(\tilde{\theta}_2 - \bar{\theta}_2) + \mathrm{o}_p(1), \end{aligned}$$

where $\bar{s}_N = s_N(\bar{\theta}_1, \bar{\theta}_2)$. Also

$$A_{11}\sqrt{N}\,(\tilde{\theta}_1 - \bar{\theta}_1) + A_{12}\sqrt{N}\,(\tilde{\theta}_2 - \bar{\theta}_2) = N^{-1/2}(\partial\bar{s}_N/\partial\theta_1) + \mathrm{o}_p(1).$$

However, using a similar expansion for $\partial s_N(\hat{\theta})/\partial\theta_1$ about $\bar{\theta}$, we can write

$$\begin{aligned} A_{11}\sqrt{N}\,(\tilde{\theta}_1 - \bar{\theta}_1) + A_{12}\sqrt{N}\,(\tilde{\theta}_2 - \bar{\theta}_1) &= A_{11}\sqrt{N}\,(\hat{\theta}_1 - \bar{\theta}_1) \\ &\quad + A_{12}\sqrt{N}\,(\hat{\theta}_2 - \bar{\theta}_2) + \mathrm{o}_p(1), \end{aligned}$$

so that

$$\sqrt{N}\,(\tilde{\theta}_1 - \hat{\theta}_1) = -A_{11}^{-1}A_{12}\sqrt{N}\,(\tilde{\theta}_2 - \hat{\theta}_2) + \mathrm{o}_p(1).$$

Therefore if $A_{12} = 0$, $\tilde{\theta}_1$ is asymptotically equivalent to $\hat{\theta}_1$ for any consistent estimator $\tilde{\theta}_2$. Otherwise, we require that $\tilde{\theta}_2$ is asymptotically equivalent to $\hat{\theta}_2$ for $\operatorname{plim}\sqrt{N}\,(\tilde{\theta}_1 - \hat{\theta}_1) = 0$ to hold.

Note that this result is still valid for nonefficient criteria. That is, when $A \neq B$.

References

Arellano, M., 1986, Estimating simultaneous equations with covariance restrictions, Paper presented at the European Meeting of the Econometric Society, Budapest.

Bhargava, A. and J.D. Sargan, 1983, Estimating dynamic random effects models from panel data covering short time periods, Econometrica 51, 1635–1659.

Chamberlain, G., 1982, Multivariate regression models for panel data, Journal of Econometrics 18, 5–46.

Gourieroux, C., A. Monfort, and A. Trognon, 1984, Pseudo maximum likelihood methods: Theory, Econometrica 52, 681–700.

Hausman, J., W. Newey, and W. Taylor, 1987, Efficient estimation and identification of simultaneous equation models with covariance restrictions, Econometrica 55, 849–874.

Hendry, D.F., 1976, The structure of simultaneous equations estimators, Journal of Econometrics 4, 51–88.

Magnus, J.R. and H. Neudecker, 1980, The elimination matrix: Some lemmas and applications, SIAM Journal of Algebraic and Discrete Methods 1, 422–449.

Malinvaud, E., 1970, Statistical methods of econometrics (North-Holland, Amsterdam).

Richard, J.-F., 1975, A note on the information matrix of the multivariate normal distribution, Journal of Econometrics 3, 57–60.

Rothenberg, T.J., 1973, Efficient estimation with a priori information, Cowles Foundation monograph no. 23 (Yale University Press, New Haven, CT).

Rothenberg, T.J. and C.T. Leenders, 1964, Efficient estimation of simultaneous equation systems, Econometrica 32, 57–76.

Sargan, J.D., 1975, The identification and estimation of sets of simultaneous stochastic equations, London School of Economics econometrics programme discussion paper no. A.1.