

EFFICIENCY BOUNDS FOR INSTRUMENTAL VARIABLE ESTIMATORS UNDER GROUP-ASYMPTOTICS

Jan van der Ploeg
and
Paul A. Bekker

Department of Economics
University of Groningen

May 1995

The paper derives asymptotic efficiency bounds for estimators of a single linear relation, based on dummy instruments, under asymptotic parameter sequences where the number of instruments is allowed to grow with the number of observations. We assume normality and show that ML-estimators under homo- and heteroscedasticity, do not reach the efficiency bound. It is shown that no uniformly continuously differentiable estimator can reach the bound for all asymptotic parameter sequences considered.

KEYWORDS: Instrumental variable estimators, grouping estimators, natural experiments, pseudo-panels, LIML, K_2 -asymptotics, group-asymptotics, alternative asymptotics, asymptotic efficiency.

We would like to thank Theo Dijkstra and Tom Wansbeek for helpful comments. E-mail: J.van.der.Ploeg@eco.rug.nl; P.A.Bekker@eco.rug.nl

1. Introduction.

Consider the instrumental variable estimation of a single linear relation based on a single instrument. In that case functions of the instrument may also serve as instruments. A rather nonparametric way of defining such instruments is to use group indicator functions that group the observations based on the ordered values of the instrumental variable. The resulting model, with dummy instruments, has been described in Bekker and Van der Ploeg (1994): B&VdP in the sequel.

The model is closely related to various fields of recent econometric research, e.g. natural experiments (Angrist, 1990) and pseudo-panels (Deaton, 1985). In case of homoscedasticity, where the covariance matrices of the observations do not vary between groups, the model is also known as the functional relationship model with replicated observations. The model can also be represented as a simultaneous equations model where the parameters of interest are given by the coefficients of a single equation.

B&VdP describe a variety of estimators for the coefficients of the model. They give asymptotic distributions and confidence intervals based on alternative asymptotic parameter sequences, where the number of instruments, or groups, is allowed to grow as the number of observations increases. Both Bekker (1994) and B&VdP show these alternative asymptotic distributions to be more accurate in their approximations to the finite sample distributions of the estimators compared to large sample approximations. This increased accuracy was found even if the number of instruments is small.

Such parameter sequences are sometimes referred to as large- K_2 asymptotics, which have been studied with respect to the simultaneous equations model and the functional relationship model in Kunitomo (1980, 1981, 1982, 1986, 1987), Morimune and Kunitomo (1980), and Fujikoshi, Morimune, Kunitomo and Taniguchi (1982). However, the number of overidentifying instruments, K_2 , need not be large for the large- K_2 asymptotics to be useful. Therefore, we follow Angrist and Krueger (1995) and Angrist, Imbens and Krueger (1994), who refer to such parameter sequences as group-asymptotics.

The main purpose of this paper is to derive efficiency bounds under group-asymptotics. B&VdP considered the maximum likelihood (ML) estimator under normality, which was not found to be asymptotically efficient relative to other consistent estimators considered in the paper. On the other hand, such relative asymptotic efficiency was found for the homoscedastic case, where LIML is the ML-estimator. This finding is in agreement with Anderson, Kunitomo and Sawa (1982, p. 1025), where it is said that Kunitomo (1980, 1982) and Morimune and Kunitomo (1980) have shown that LIML is efficient under group-asymptotics. A similar result that applies to the case of a single explanatory endogenous variable can be found in Kunitomo (1987, Theorem 3.1). However, by extending the class of estimators considered by Kunitomo (1987), we show for the multivariate case that the LIML-estimator can be improved upon using statistics that arise naturally. That is, LIML is *not* efficient under group-asymptotics.

The paper is organized as follows. Section 2 describes the model, the group-asymptotics, and the asymptotic distribution of the statistics on which we base our class of estimators. Section 3 describes a general approach to derive efficiency bounds that hold under a family of asymptotic parameter sequences. This approach is applied in Section 4 to derive efficiency

bounds for grouping estimators. In Section 5 we describe estimators that reach the efficiency bounds. However, we expect these estimators to be of limited practical value compared to some estimators given in B&VdP. Section 6 uses continuity arguments to indicate that *no* uniformly continuously differentiable estimator can be efficient under all asymptotic parameter sequences.

2. The model, the group-asymptotics, and the statistics

The model we consider and its relation to the literature on instrumental variable estimation, errors-in-variables models, natural experiments, and pseudo-panels has been described in B&VdP. The same holds true for the asymptotic parameter sequences we consider. However, there are some differences. In order to derive asymptotic efficiency bounds, we restrict our analysis to normally distributed observations.

2.1 The model

Consider independent random samples taken from m different $(g + 1)$ -dimensional normal distributions whose expectations satisfy the same linear relation. Let the data consist of scalars y_{ij} and g -vectors x_{ij} , $i = 1, \dots, n_j$ and $j = 1, \dots, m$. The vectors $(y_{ij}, x'_{ij})' = (y_{ij}; x_{ij})$ are independent with normal distributions indexed by j that satisfy $E(y_{ij}) = E(x'_{ij})\delta$, where δ is the parameter vector of interest. Let $E(x_{ij}) = \pi_j$, then

$$\begin{aligned} y_{ij} &= x'_{ij}\delta + \varepsilon_{ij}, & i &= 1, \dots, n_j, \\ x_{ij} &= \pi_j + v_{ij}, & j &= 1, \dots, m, \end{aligned} \quad (1)$$

where $(\varepsilon_{ij}; v_{ij})$ has zero expectation and

$$\text{Var} \{(\varepsilon_{ij}; v_{ij})\} = \Sigma_j = \begin{bmatrix} \sigma_j^2 & \sigma_{j12} \\ \sigma_{j21} & \Sigma_{j22} \end{bmatrix}.$$

For the identification of δ we need the restriction that (π_1, \dots, π_m) has rank g , which implies $m \geq g$.

As noticed by B&VdP, model (1) can be applied in the context of natural experiments (Angrist, 1990) and pseudo-panels (Deaton, 1985). In general, the model can be induced by a single instrument. That is, if a random sample of vectors (y_i, x'_i, z_i) is taken from a population where $y_i = x'_i\delta + \varepsilon_i$ and $E(\varepsilon_i|z_i) = 0$, then z is an instrument and functions of z are also instruments. A rather nonparametric way of defining such functions is to order the observations, such that the elements z_i in the vector z are ordered, and to split up the vector z into m groups. The indicator functions of the groups may thus serve as instruments, which amounts to model (1).

In order to formulate the model in reduced form, let the observations and disturbances of the j -th group be stacked in the matrices (y_j, X_j) and (ε_j, V_j) , resp. Furthermore, let ι_{n_j} be a vector of n_j ones, and let $u_j = \varepsilon_j + V_j\delta$ (we simply drop the statement $j = 1, \dots, m$). Then we have

$$Y_j = (y_j, X_j) = \iota_{n_j}\pi'_j(\delta, I_g) + (u_j, V_j) \quad (2)$$

The rows of (u_j, V_j) are independently normally distributed with zero mean and covariance matrix

$$\text{Var}(u_{ij}; v_{ij}) = \Omega_j = \begin{bmatrix} 1 & \delta' \\ 0 & I_g \end{bmatrix} \Sigma_j \begin{bmatrix} 1 & 0 \\ \delta & I_g \end{bmatrix}.$$

2.2 The group-asymptotics

Asymptotic distributions will be based on parameter sequences where the number of instruments is allowed to grow as the number of observations increases. In Bekker (1994) and B&VdP, the resulting approximations (to the finite sample distributions of estimators of δ) were shown to be more accurate compared to large sample approximations, even if the number of instruments in the actual sample is small.

Let the number of observations in the j th group be a fixed proportion, w_j , of the total number of observations: $n_{jn} = w_j n$. As $n \rightarrow \infty$, each group is split up into an increasing number of k_{jn} subgroups with f_{jn} observations each. The $n_{jn} = k_{jn} f_{jn}$ observations in the j -th group are required to have the same covariance matrix Ω_j . So the number of parameters in the m matrices Ω_j remains fixed. Thus we have a sequence of samples satisfying

$$Y_j = Z_{jn} A'_{jn} (\delta, I_g) + (u_j, V_j), \quad (3)$$

where $Z_{jn} = I_{k_{jn}} \otimes I_{f_{jn}}$ and the A_{jn} are $(g \times k_{jn})$ matrices.

In the actual sample we have $n_{jn} = n_j$, $k_{jn} = 1$ and $A_{jn} = \pi_j$. For the asymptotic parameter sequences we require, as $n \rightarrow \infty$,

$$\begin{aligned} \frac{k_{jn}}{n_{jn}} &\rightarrow \frac{\alpha_j}{w_j m} < 1, \\ \frac{A_{jn} Z'_{jn} Z_{jn} A'_{jn}}{n_{jn}} &= \frac{A_{jn} A'_{jn}}{k_{jn}} \rightarrow \Pi_j \Pi'_j, \end{aligned} \quad (4)$$

where $\Pi_j : g \times r_j$, $r_j \leq g$.

If $k_{jn} = 1$ is fixed then $\alpha_j = 0$, which corresponds to large sample asymptotics. If $\alpha_j \neq 0$, then the total number of instruments grows with the total number of observations such that

$$\frac{\sum_{j=1}^m k_{jn}}{n} \rightarrow \bar{\alpha} = \frac{1}{m} \sum_{j=1}^m \alpha_j.$$

For an actual approximation we would use $\Pi_j = \pi_j$ (cf. Bekker, 1994). However, to make comparisons with the literature, we use formulation (4), where $\sum_{j=1}^m w_j \Pi_j \Pi'_j$ is nonsingular for identification purposes. The sequences (4) are slightly more general than those considered by B&VdP, where $\alpha_j = \bar{\alpha}$ and $\Pi_j = \pi_j$. On the other hand we assume normality in this paper.

2.3 The statistics

The statistics on which we base our class of estimators can be computed using group mean vectors and group covariance matrices. Define the projection matrix $P_{Z_{jn}} = Z_{jn} (Z'_{jn} Z_{jn})^{-1} Z_{jn}$

and let ¹

$$\begin{aligned}\bar{S}_j &= \frac{1}{n} Y_j' P_{Z_{jn}} Y_j \\ S_j^\perp &= \frac{1}{n} Y_j' (I_{n_{jn}} - P_{Z_{jn}}) Y_j.\end{aligned}\tag{5}$$

Notice the order of these matrices, $(g + 1) \times (g + 1)$, remains fixed in the sequence defined by (4). We consider estimators of δ that are functions of

$$s_j = (\bar{s}_j; s_j^\perp) = (\text{vec}(\bar{S}_j); \text{vec}(S_j^\perp)).$$

According to B&VdP we have

$$n^{1/2}(s_j - E(s_j)) \overset{A}{\rightsquigarrow} N(0, \lim_{n \rightarrow \infty} n \text{Var}(s_j)).$$

If we restrict our sequences to cases where the difference between the left-hand-sides of (4) and their limits is of order $o(n^{-1/2})$, we also have

$$n^{1/2}(s_j - \psi_j) \overset{A}{\rightsquigarrow} N(0, V_j),\tag{6}$$

where

$$\psi_j = \begin{pmatrix} \bar{\psi}_j \\ \psi_j^\perp \end{pmatrix} = \begin{pmatrix} \text{plim}(\bar{s}_j) \\ \text{plim}(s_j^\perp) \end{pmatrix} = \begin{pmatrix} \lim_{n \rightarrow \infty} E(\bar{s}_j) \\ \lim_{n \rightarrow \infty} E(s_j^\perp) \end{pmatrix},$$

$$V_j = \begin{bmatrix} \bar{V}_j & 0 \\ 0 & V_j^\perp \end{bmatrix} = \begin{bmatrix} \lim_{n \rightarrow \infty} n \text{Var}(\bar{s}_j) & 0 \\ 0 & \lim_{n \rightarrow \infty} n \text{Var}(s_j^\perp) \end{bmatrix}.$$

The block-diagonality of V_j is due to the normality of the observations.

Using the properties of noncentral Wishart distributions (Muirhead, 1982, Problem 10.8; Eaton, 1983, Theorem 8.13), we find

¹ In the notation of B&VdP we have $\bar{S}_j = M_j$ and $S_j^\perp = N_j - M_j$.

$$\begin{aligned}
\bar{\psi}_j &= \text{vec} \{w_j(\delta, I_g)' \Pi_j \Pi_j'(\delta, I_g) + \frac{\alpha_j}{m} \Omega_j\} \\
&= R'^{-1} \text{vec} \{w_j(0, I_g)' \Pi_j \Pi_j'(0, I_g) + \frac{\alpha_j}{m} \Sigma_j\}, \\
\psi_j^\perp &= \text{vec} \{(w_j - \frac{\alpha_j}{m}) \Omega_j\} \\
&= R'^{-1} \text{vec} \{(w_j - \frac{\alpha_j}{m}) \Sigma_j\}, \\
\bar{V}_j &= H[\Omega_j \otimes \{\frac{\alpha_j}{2m} \Omega_j + w_j(\delta, I_g)' \Pi_j \Pi_j'(\delta, I_g)\}] H \\
&= H R'^{-1} [\Sigma_j \otimes \{\frac{\alpha_j}{2m} \Sigma_j + w_j(0, I_g)' \Pi_j \Pi_j'(0, I_g)\}] R^{-1} H, \\
V_j^\perp &= H\{\Omega_j \otimes \frac{1}{2}(w_j - \frac{\alpha_j}{m}) \Omega_j\} H \\
&= H R'^{-1} \{\Sigma_j \otimes \frac{1}{2}(w_j - \frac{\alpha_j}{m}) \Sigma_j\} R^{-1} H,
\end{aligned}$$

with

$$\begin{aligned}
R &= \begin{pmatrix} 1 & 0 \\ -\delta & I_g \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ -\delta & I_g \end{pmatrix}, \\
H &= I_{(g+1)^2} + K,
\end{aligned}$$

and K is the $(g+1)^2 \times (g+1)^2$ permutation matrix, defined by $K \text{vec}(M) = \text{vec}(M')$ for any $(g+1) \times (g+1)$ matrix M . Notice that H and R commute: $HR = RH$.

Stacking all statistics, we define

$$\begin{aligned}
s &= (s_1; \dots; s_m), \\
\psi &= (\psi_1; \dots; \psi_m), \\
V &= \begin{bmatrix} V_1 & & 0 \\ & \ddots & \\ 0 & & V_m \end{bmatrix}.
\end{aligned}$$

Due to the independency of the observations we find

$$n^{1/2}(s - \psi) \stackrel{A}{\sim} N(0, V). \tag{7}$$

The asymptotic distribution of s depends on the parameters of the model and on the particular sequence that is followed. That is, the asymptotic distribution depends on the value of the vector

$$\alpha = (\alpha_1; \dots; \alpha_m).$$

B&VdP considered $\alpha_j = \bar{\alpha}$ and described functions of s as estimators of δ that are consistent for any $0 \leq \bar{\alpha} < w_j m$. The ML-estimator is one of them. It is asymptotically efficient under large sample asymptotics: $\bar{\alpha} = 0$. However, for $\bar{\alpha} > 0$, it was not found to be asymptotically efficient relative to the other estimators considered.

We also consider the homoscedastic case, where the matrices Ω_j do not vary between groups. In that case ML is given by the LIML-estimator, which is a function of the statistics $\sum_{j=1}^m s_j^\perp$ and $\sum_{j=1}^m \bar{s}_j$. Under homoscedasticity, LIML is consistent for any $0 \leq \bar{\alpha} < w_j m$ and B&VdP found it to be asymptotically efficient relative to the other estimators considered in that paper. This finding is in agreement with statements made in the literature (Anderson, Kunitomo and Sawa, 1982, p.1025) that say that LIML is both consistent and efficient under group asymptotics. However, here we show that if $\alpha \neq 0$, then LIML is not asymptotically efficient within the class of estimators that are functions of s .

3. A general approach

In order to describe efficiency bounds that hold under a variety of parameter sequences we first give, in the first subsection, a brief discussion of the main idea of our approach. The second subsection formally derives lower bounds that will be used in the next section to derive efficiency bounds for grouping estimators, i.e. estimators of δ that are functions of s as defined in Section 2.

3.1 A heuristic presentation of the main idea

We consider sequences of vectors of statistics s , of fixed order l , say, such that as $n \rightarrow \infty$

$$n^{1/2}(s - \psi(\delta; \tau)) \overset{A}{\rightsquigarrow} N(0, V(\delta, \tau)). \quad (8)$$

Here the probability limits ψ and the asymptotic covariance matrices V depend on a $g \times 1$ parameter vector of interest δ and a vector of nuisance parameters τ_1 contained in the $h \times 1$ vector $\tau = (\tau_1; \tau_2)$. The vector τ_2 indicates which sequence has been followed.

In order to derive an estimator $\hat{\delta}(s)$ that is consistent and asymptotically efficient, under all asymptotic parameter sequences indicated by τ_2 , we may treat the elements of τ_2 as if they were parameters just as the elements of δ and τ_1 . We can then use the same arguments that lead to the consistency and asymptotic efficiency of the minimum chi-square estimator

$$(\hat{\delta}; \hat{\tau}) = \arg \min_{(\delta; \tau)} (s - \psi(\delta; \tau))' \hat{V}^{-1} (s - \psi(\delta; \tau)), \quad (9)$$

where \hat{V} is an estimator of V that is consistent for all parameter sequences indicated by τ_2 . As a result $\hat{\delta}$ will be consistent and asymptotically efficient for all parameter sequences indicated by τ_2 .

An efficiency bound is now given by the asymptotic covariance matrix of $n^{1/2}\hat{\delta}$:

$$(I_g, 0)(\Delta' V^{-1} \Delta)^{-1} (I_g; 0), \quad (10)$$

where Δ is an $l \times (g + h)$ matrix:

$$\Delta = (\Delta_1, \Delta_2) = \left(\frac{\partial \psi}{\partial \delta'}, \frac{\partial \psi}{\partial \tau'} \right). \quad (11)$$

However, the rather informally derived lower bound (10) becomes a very complicated function of the parameters of the model when applied to grouping estimators. This makes it very hard to give an analytical comparison between (10) and the asymptotic covariance matrices of grouping estimators as described in B&VdP. The next subsection gives a more formal derivation of another representation of the efficiency bound.

3.2 A general formulation of efficiency bounds

Let $(\delta; \tau) \in \Theta$, which is an open subset of R^{g+h} . Furthermore, let $\psi : \Theta \rightarrow \mathbf{R}^l$ be differentiable. We assume the vector of statistics s has an asymptotic distribution as indicated in (8) for any $(\delta; \tau) \in \Theta$. However, we do not assume $V(\delta; \tau)$ to be nonsingular.

Consider the class of all differentiable consistent estimators $f(s)$, so that $\text{plim } f(s) = \delta$ and $f(\psi(\delta; \tau)) = \delta$. Define the $g \times l$ matrix F as

$$F(\delta; \tau) = \frac{\partial f(\psi)}{\partial \psi'}. \quad (12)$$

By the delta-method we find

$$n^{1/2}(f(s) - \delta) \overset{A}{\sim} N(0, FVF'). \quad (13)$$

In order to find an asymptotically efficient estimator within this class of differentiable consistent estimators we notice that

$$\frac{\partial f}{\partial (\delta', \tau')} = F\Delta = (I_g, 0), \quad (14)$$

where Δ is defined in (11). A lower bound for FVF' can now be found using the following algebraic theorem, where $+$ denotes the Moore-Penrose inverse.

Theorem 3.1 *Let $F\Delta = F(\Delta_1, \Delta_2) = (I_g, 0)$ and $VV^+\Delta = \Delta$, then $FVF' \geq V_L$, where*

$$V_L = (I_g, 0)(\Delta'V^+\Delta)^+(I_g; 0); \quad (15)$$

this lower bound is reached for

$$F' = V^+\Delta(\Delta'V^+\Delta)^+(I_g; 0).$$

If, furthermore, U is a matrix of appropriate order such that

$$U'\Delta_2 = 0, \quad (16)$$

$$\text{rank}(\Delta_1, VU, \Delta_2) = \text{rank}(VU, \Delta_2),$$

then

$$V_L = (\Delta_1'U(U'VU)^+U'\Delta_1)^{-1}. \quad (17)$$

The proof is given in Appendix 1.

In the next section, and Appendix 2, we will use the representation (17) to derive efficiency bounds for grouping estimators.

4. Efficiency bounds for grouping estimators

Returning to model (1) and the asymptotic distribution of the statistics s as given in (6) and (7), let $\hat{\delta}(s)$ be differentiable such that $\text{plim } \hat{\delta}(s) = \delta$ for all parameter sequences (4). That is, $\hat{\delta}(s)$ is consistent for all $0 \leq \alpha_j < w_j m$.

In the following theorems we assume the matrices Ω_j to be nonsingular. For such cases the theorems give efficiency bounds, $\text{Avar}\{n^{1/2}\hat{\delta}(s)\} \geq V_L$, under heteroscedasticity, where the matrices Ω_j vary between groups, and homoscedasticity, $\Omega_j = \Omega$, resp. Let

$$\phi_j = \sigma_{j21}/\sigma_j,$$

where we drop the index j in case of homoscedasticity.

Theorem 4.1 *Under heteroscedasticity, where the matrices Ω_j are nonsingular, the efficiency bound for differentiable consistent estimators $\hat{\delta}(s)$ is given by*

$$\begin{aligned} V_L &= \left[\sum_{j=1}^m \left\{ A_j (A_j + B_j)^+ A_j \right\} \right]^{-1}, \\ A_j &= \sigma_j^{-2} w_j \Pi_j \Pi_j', \\ B_j &= \frac{\sigma_j^{-2} w_j \alpha_j}{m w_j - \alpha_j} (\Sigma_{j22} - \phi_j \phi_j'). \end{aligned} \tag{18}$$

The proof is given in Appendix 2.

Theorem 4.2 Under homoscedasticity, where $\Omega_j = \Omega$ is nonsingular, the efficiency bound for differentiable consistent estimators $\hat{\delta}(s)$ is given by

$$\begin{aligned}
V_L &= \sigma^2 (J' A (A + B)^+ A J)^{-1}, \\
J &= \iota_m \otimes I_g, \\
A &= \begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{bmatrix}, \quad A_j = w_j \Pi_j \Pi_j', \\
B &= \left(\frac{1}{m} \text{Diag}(\alpha) + \frac{1}{m^2(1-\bar{\alpha})} \alpha \alpha' \right) \otimes (\Sigma_{22} - \phi \phi').
\end{aligned} \tag{19}$$

The proof is given in Appendix 2.

Notice that the nonsingularity of Ω_j implies the nonsingularity of both Σ_j and $\Sigma_{j22} - \phi_j \phi_j'$. Consequently, if $\alpha_j > 0$ the Moore-Penrose inverses in Theorems 4.1 and 4.2 can be replaced by regular inverses.

The lower bounds V_L in Theorems 4.1 and 4.2 hold for estimators $\hat{\delta}(s)$ that are consistent for all $0 \leq \alpha_j < w_j m$. However, if we restrict this set to subsets where either $\alpha = \alpha^0$, or $\alpha_j = \bar{\alpha}$, we do not find lowerbounds that are smaller than V_L as given in theorems 4.1 and 4.2, i.e. the lowerbounds remain valid in these cases. In particular, in case of large sample asymptotics, where $\alpha = 0$, we find under heteroscedasticity,

$$V_L = \left(\sum_{j=1}^m \sigma_j^{-2} w_j \Pi_j \Pi_j' \right)^{-1};$$

under homoscedasticity σ_j^2 can be replaced by σ^2 . Comparing these results with the asymptotic covariance matrices of the ML-estimators, as described in B&VdP, we find the expected result that these estimators are indeed efficient if $\alpha = 0$.

However, if we consider parameter sequences where $\alpha \neq 0$, we find a different result. Under heteroscedasticity V_L can be rewritten as follows. Let

$$J = \iota_m \otimes I_g, \quad A = \begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{bmatrix}, \quad B = \begin{bmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_m \end{bmatrix},$$

then

$$V_L = [J'A(A+B)^+AJ]^{-1} \quad (20)$$

Now consider the projection matrix

$$P = (A+B)^{1/2}J(J'(A+B)J)^{-1}J'(A+B)^{1/2}.$$

We find

$$\begin{aligned} V_L &\leq [J'A(A+B)^{+1/2}P(A+B)^{+1/2}AJ]^{-1} \\ &= (J'AJ)^{-1}J'(A+B)J(J'AJ)^{-1}, \\ J'AJ &= \sum_{j=1}^m \sigma_j^{-2} w_j \Pi_j \Pi_j', \\ J'(A+B)J &= \sum_{j=1}^m \left\{ \sigma_j^{-2} w_j \Pi_j \Pi_j' + \frac{\sigma_j^{-2} w_j \alpha_j}{w_j m - \alpha_j} (\Sigma_{j22} - \phi_j \phi_j') \right\}. \end{aligned} \quad (21)$$

For $\alpha_j = \bar{\alpha}$ the right-hand-side of (21) is equal to the asymptotic covariance matrix of the ML-estimator as given in B&VdP.

Under homoscedasticity we find, in a similar way,

$$\begin{aligned} V_L &\leq \sigma^2 (J'AJ)^{-1} J'(A+B) J (J'AJ)^{-1}, \\ J'AJ &= \sum_{j=1}^m w_j \Pi_j \Pi_j', \\ J'(A+B)J &= \left(\sum_{j=1}^m w_j \Pi_j \Pi_j' \right) + \frac{\bar{\alpha}}{1 - \bar{\alpha}} (\Sigma_{22} - \phi \phi'). \end{aligned} \quad (22)$$

Now the right-hand-side of (22) is equal to the asymptotic covariance matrix of the LIML-estimator as given in B&VdP.

So the ML-estimators do not reach the asymptotic efficiency bounds as given in Theorems 4.1 and 4.2. In particular for the homoscedastic case this result contradicts earlier statements made in the literature (Anderson, Kunitomo and Sawa (1982, p. 1025) and Kunitomo (1987)), where it is said that LIML is asymptotically efficient under group-asymptotics. The question is how these earlier results relate to our result that says that ML is inefficient under group-asymptotics.

The problem is easily resolved. We consider the asymptotic performance of the LIML-estimator relative to the estimators that are functions of the statistics \bar{s}_j and s_j^\perp , j, \dots, m ,

whereas Kunitomo's (1987) efficiency result holds only relative to estimators that are functions of $\sum_{j=1}^m \bar{s}_j$ and $\sum_{j=1}^m s_j^\perp$. If we only consider this subset of estimators then, indeed, LIML is efficient. That is, we may simply apply Theorem 4.2 with the number of groups m equal to one and $A_1 = \sum_{j=1}^m w_j \Pi_j \Pi_j'$. In that case the lower bound V_L is equal to the asymptotic covariance matrix of the LIML-estimator. However, if we consider the larger set of estimators based on \bar{s}_j and \bar{s}_j^\perp , we find a smaller lower bound V_L . In fact, in the next section we consider an estimator that reaches this lower bound V_L . So LIML is inefficient under group-asymptotics.

5. Efficient grouping estimators

In section 3 we already indicated, (2), how an efficient estimator might be formulated. Here we consider a more simple asymptotically efficient estimator. Let

$$\begin{aligned} a &= (a_1; \dots; a_m), \quad a_j = (\bar{a}_j; a_j^\perp), \\ \bar{a}_j &= \bar{S}_j(1; -\delta) - \frac{\alpha_j}{m} \gamma_j, \\ a_j^\perp &= S_j^\perp(1; -\delta) - (w_j - \frac{\alpha_j}{m}) \gamma_j. \end{aligned}$$

Then, for $\gamma_j = \Omega_j(1; -\delta)$, we find $\text{plim } a = 0$. Due to the asymptotic normality of s , (6) and (7), we find

$$n^{1/2} a \overset{A}{\rightsquigarrow} N(0, W),$$

$$W = \begin{bmatrix} W_1 & & 0 \\ & \ddots & \\ 0 & & W_m \end{bmatrix}, \quad W_j = \begin{bmatrix} \bar{W}_j & 0 \\ 0 & W_j^\perp \end{bmatrix},$$

$$\bar{W}_j = w_j \sigma_j^2(\delta, I_g)' \Pi_j \Pi_j'(\delta, I_g) + \frac{\alpha_j}{m} (\sigma_j^2 \Omega_j + \Omega_j \begin{pmatrix} 1 \\ -\delta \end{pmatrix} \begin{pmatrix} 1 \\ -\delta \end{pmatrix}' \Omega_j),$$

$$W_j^\perp = (w_j - \frac{\alpha_j}{m}) (\sigma_j^2 \Omega_j + \Omega_j \begin{pmatrix} 1 \\ -\delta \end{pmatrix} \begin{pmatrix} 1 \\ -\delta \end{pmatrix}' \Omega_j),$$

where the expressions for \bar{W}_j and W_j^\perp can be verified by Lemma 11.1 in B&VdP.

In the following theorems we again assume the matrices Ω_j to be nonsingular. Furthermore, we restrict the parameter sequences (4) to cases where $\alpha_j > 0$. In that case the matrices \bar{W}_j and W_j^\perp are invertible and we have the following result.

Theorem 5.1 *Under heteroscedasticity, where the matrices Ω_j are nonsingular, and for parameter sequences with $\alpha_j > 0$, let $\gamma = (\gamma_1; \dots; \gamma_m)$ and define*

$$(\hat{\delta}; \hat{\gamma}; \hat{\alpha}) = \underset{(\delta; \gamma; \alpha)}{\text{arg min}} a'(\delta; \gamma; \alpha) \hat{W}^{-1} a(\delta; \gamma; \alpha),$$

where \hat{W} is a consistent estimator of W . Then

$$n^{1/2}(\hat{\delta} - \delta) \overset{A}{\rightsquigarrow} N(0, V_L),$$

where V_L is given in theorem 4.1.

The proof is given in Appendix 3.

Theorem 5.2 *Under homoscedasticity, where $\Omega_j = \Omega$ is nonsingular, and for parameter sequences where $\alpha_j > 0$, let $\gamma_j = \gamma$ and define*

$$(\hat{\delta}; \hat{\gamma}; \hat{\alpha}) = \arg \min_{(\delta; \gamma; \alpha)} a'(\delta; \gamma; \alpha) \hat{W}^{-1} a(\delta; \gamma; \alpha),$$

where \hat{W} is a consistent estimator of W . Then

$$n^{1/2}(\hat{\delta} - \delta) \overset{A}{\sim} N(0, V_L),$$

where V_L is given in theorem 4.2.

The proof is given in Appendix 3.

The estimators defined in theorems 5.1 and 5.2 serve interesting theoretical purposes: they reach the asymptotic efficiency bound V_L . However, the assumptions that lead to the invertibility of W are quite restrictive. If a matrix Ω_j is singular, or $\alpha_j = 0$, the estimators are not well-defined. If Ω_j is close to singularity, or α_j close to zero, the matrix W will be close to singularity and the inversion \hat{W}^{-1} may lead to all kinds of difficulties. No such difficulties are encountered for the ML-estimators, which are consistent over the full parameter space. However, they are efficient only for $\alpha = 0$. The problem whether or not there exist estimators that are asymptotically efficient over the full parameter space and for all sequences (4) with $0 \leq \alpha_j < w_j m$ is further discussed in the next section.

6. Efficiency on the full parameter space

Consider estimators $\hat{\delta}(s)$ that are consistent for δ over the full parameter space, without excluding cases where Ω_j is singular or $\alpha_j = 0$. Examples are given in B&VdP. If $\hat{\delta}$ is a uniformly continuously differentiable function of s , $\hat{\delta}$ will have an asymptotic normal distribution,

$$n^{1/2}(\hat{\delta} - \delta) \overset{A}{\sim} N(0, V),$$

where V is a continuous function in the interior of the parameter space and has a continuous extension to the parameters $(\delta; \pi; \omega; \alpha)$, as defined in Appendix 2.

If $\hat{\delta}$ is efficient for points $(\delta; \pi; \omega; \alpha)$ in the interior of the parameter space, where Ω_j is nonsingular and $\alpha_j > 0$, then, according to theorems 4.1 and 4.2, $V = V_L$ for such interior points. For both the homoscedastic case and the heteroscedastic case V_L can be represented (cf. (19) and (20)) for such interior points as

$$V_L = (J'A(A + B)^{-1}AJ)^{-1}.$$

Due to the continuity of the inverse, this is, indeed, a continuous function of the parameter points on the interior of the parameter space.

As V is a continuous function, we may derive from the equality $V = V_L$, which holds on the interior, the value of V on the boundary where B is singular: $B = B^*$, say. That is, we may consider a path from the interior to the boundary such that $B = B^* + \lambda C$, with $\lambda > 0$, $\lambda \rightarrow 0$ and $C > 0$. It follows from Lemma 6.1 in Appendix 4 that in that case

$$\lim_{\lambda \rightarrow 0} A(A + B^* + \lambda C)^{-1}A = A(A + B^*)^+A.$$

Consequently, if V is continuous and $V = V_L$ on the interior, then V should take the form

$$V = (J'A(A + B)^+AJ)^{-1}, \quad (23)$$

on the full parameter space.

However, it follows from Lemma 6.2 in Appendix 4 that V as given in (23) is not a continuous function of the elements of B , which is a *contradiction*, since V is continuous. Consequently, there is no uniformly continuously differentiable consistent estimator of δ , based on the statistics s , that is efficient under group asymptotics over the full parameter space.

On the other hand if we consider a single asymptotic parameter sequence where $\alpha = \alpha^0 \neq 0$, say, then we find, as is shown in sections 4 and 5, that the sequence of ML estimators is not *Best Asymptotically Normal*. Here the incidental parameters in the matrices A_{jn} , (3), do not affect the consistency of ML, but instead affect its asymptotic efficiency.

Appendix 1
Proof of Theorem 3.1

Projecting $V^{1/2}F'$ on $V^{+1/2}\Delta$, we find

$$\begin{aligned} FVF' &\geq FV^{1/2}(V^{+1/2}\Delta(\Delta'V^+\Delta)^+\Delta'V^{+1/2})V^{1/2}F' \\ &= (I_g, 0)(\Delta'V^+\Delta)^+(I_g; 0), \end{aligned}$$

which is (15).

To prove (17) we notice that (16) implies the existence of matrices A and B such that

$$\Delta = (VU, \Delta_2) \begin{bmatrix} A & 0 \\ B & I_h \end{bmatrix},$$

and

$$\Delta'V^+\Delta = (A, 0)'U'VU(A, 0) + (B, I_h)'\Delta_2'V^+\Delta_2(B, I_h).$$

As $F\Delta = (I_g, 0)$, we have $FVUA = I_g$. So $V^{1/2}UA$ must have full column rank and $A'U'VUA$ must be nonsingular. Using general results on Schur-complements, as collected in Ouellette (1981, theorem 4.6 (iii)), we find

$$V_L = (A'U'VUA)^{-1}.$$

As $U'\Delta_1 = U'VUA$, we find the result as given in (17). □

APPENDIX 2
Proof of theorems 4.1 and 4.2

First we proof the theorems for cases where $\alpha_j > 0$. Then, at the end of the proofs, it is indicated that the same arguments can be used for cases where $\alpha_j = \bar{\alpha}$, or $\alpha = \alpha^0$, in particular $\alpha_j = 0$. For the regular cases, where $\alpha_j > 0$ and $\Omega_j > 0$, a vector x satisfies $x' \bar{V}_j = 0$, or $x' V_j^\perp = 0$, if and only if $x' H = 0$. So the singularity of V is due only to the symmetry of the statistics \bar{S}_j and S_j^\perp . As a result we have $V V^+ \Delta = H H^+ \Delta = \Delta$, which we need in order to apply Theorem 3.1.

We use the following notation:

$$\begin{aligned} \Delta &= ((\Delta_{11}, \Delta_{12}); \dots; (\Delta_{m1}, \Delta_{m2})), \\ (\Delta_{j1}, \Delta_{j2}) &= \begin{bmatrix} \bar{\Delta}_{j1} & \bar{\Delta}_{j2} \\ \Delta_{j1}^\perp & \Delta_{j2}^\perp \end{bmatrix} = \left(\frac{\partial(\bar{\psi}_j; \psi_j^\perp)}{\partial \delta'}, \frac{\partial(\bar{\psi}_j; \psi_j^\perp)}{\partial(\pi', \omega', \alpha')} \right), \\ \pi &= \text{vec}(\Pi_1; \dots; \Pi_m), \\ \omega &= \text{vech}(\Omega), \\ \alpha &= (\alpha_1; \dots; \alpha_m), \end{aligned}$$

where $\text{vech}(\Omega)$ is a stacking of the diagonal and subdiagonal elements of Ω . We also use

$$T = \frac{\partial \text{vec}(\Omega)}{\partial \text{vech}'(\Omega)}.$$

Notice that $\text{vec}(\Omega) = R'^{-1} \text{vec}(\Sigma)$.

In each proof we give the matrices Δ_{j1} and Δ_{j2} , which were found by application of the rules of differential matrix calculus (e.g. Balestra, 1976). Then matrices $U_j = (\bar{U}_j; U_j^\perp)$ are given, which build the matrix $U = (U_1; \dots; U_m)$, such that $U' \Delta_2 = \sum_{j=1}^m U_j' \Delta_{j2} = 0$. Next we give the matrices $V_j U_j = (\bar{V}_j \bar{U}_j; V_j^\perp U_j^\perp)$ so that $V U = (V_1 U_1; \dots; V_m U_m)$. Finally, we show that if a vector x satisfies both $x' V U = 0$ and $x' \Delta_2 = 0$, it also satisfies $x' \Delta_1 = 0$. In that case the matrix U satisfies condition (16) of Theorem 1 so that the asymptotic efficiency bound is given by (17).

Proof of Theorem 4.1 We find

$$(A1) \quad \Delta_{j1} = \begin{pmatrix} \bar{\Delta}_{j1} \\ \Delta_{j1}^\perp \end{pmatrix} = \begin{pmatrix} w_j H R'^{-1} \left(e_1 \otimes \begin{pmatrix} 0 \\ I_s \end{pmatrix} \right) \Pi_j \Pi_j' \\ 0 \end{pmatrix},$$

where e_1 is the first column of I_{g+1} ;

$$(A2) \quad \Delta_{j2} = \begin{pmatrix} \bar{\Delta}_{j2} \\ \Delta_{j2}^\perp \end{pmatrix} =$$

$$\begin{bmatrix} w_j e_j' \otimes [HR'^{-1} \left\{ \begin{pmatrix} 0 \\ I_g \end{pmatrix} \otimes \begin{pmatrix} 0 \\ I_g \end{pmatrix} \right\} (\Pi_j \otimes I_g)] & \frac{\alpha_j}{m} e_j' \otimes T & \frac{1}{m} e_j' \otimes \text{vec}(\Omega_j) \\ 0 & (w_j - \frac{\alpha_j}{m}) e_j' \otimes T & -\frac{1}{m} e_j' \otimes \text{vec}(\Omega_j) \end{bmatrix},$$

where e_j is the j th column of I_m . Now let

$$(A3) \quad U_j = \begin{pmatrix} \bar{U}_j \\ U_j^\perp \end{pmatrix} = \begin{bmatrix} e_j' \\ \alpha_j e_j' \\ \alpha_j - m w_j \end{bmatrix} \otimes \left[HR \left\{ e_1 \otimes \begin{bmatrix} \frac{-\sigma_{j12}}{\sigma_j^2} \\ I_g \end{bmatrix} \right\} \right].$$

It can be easily verified that $U' \Delta_2 = 0$. Furthermore, we find

$$\bar{V}_j \bar{U}_j = e_j' \otimes \left[2HR'^{-1} \left\{ \Sigma_j e_1 \otimes \begin{pmatrix} 0 \\ I_g \end{pmatrix} \right\} \left(\frac{\alpha_j}{m} (\Sigma_{j22} - \phi_j \phi_j') + w_j \Pi_j \Pi_j' \right) \right],$$

$$V_j^\perp U_j^\perp = \frac{\alpha_j e_j'}{\alpha_j - m w_j} \otimes \left[2(w_j - \frac{\alpha_j}{m}) HR'^{-1} \left\{ \Sigma_j e_1 \otimes \begin{pmatrix} 0 \\ I_g \end{pmatrix} \right\} (\Sigma_{j22} - \phi_j \phi_j') \right].$$

Consider vectors $x = (x_1; \dots; x_m)$, $x_j = (\bar{x}_j; x_j^\perp)$, such that $x' \Delta_2 = 0$ and $x' V U = 0$. We will prove that x also satisfies $x' \Delta_1 = 0$. For cases where \bar{x}_j and x_j^\perp are vectorizations of skew-symmetric matrices, so that $\bar{x}_j' H = x_j^{\perp'} H = 0$, the proof is trivial. As any square matrix can be written as a sum of a symmetric and a skew-symmetric matrix, we only need consider cases where \bar{x}_j and x_j^\perp are vectorizations of symmetric matrices. Let

$$\bar{x}_j = R \text{vec} \begin{bmatrix} a_j & b_j' \\ b_j & C_j \end{bmatrix},$$

(A4)

$$x_j^\perp = R \text{vec} \begin{bmatrix} \tilde{a}_j & \tilde{b}_j' \\ \tilde{b}_j & \tilde{C}_j \end{bmatrix},$$

where C_j and \tilde{C}_j are symmetric $g \times g$ matrices. As $x' \Delta_2 = 0$ implies

$$\bar{x}'_j H R'^{-1} \left\{ \begin{pmatrix} 0 \\ I_g \end{pmatrix} \otimes \begin{pmatrix} 0 \\ I_g \end{pmatrix} \right\} (\Pi_j \otimes I_g) = 0,$$

$$\left(\frac{\alpha_j}{m} \bar{x}'_j + \left(w_j - \frac{\alpha_j}{m} \right) x_j^{\perp'} \right) T = 0,$$

we find

$$(A5) \quad C_j \Pi_j = 0,$$

$$\frac{\alpha_j}{m} \begin{bmatrix} a_j & b'_j \\ b_j & C_j \end{bmatrix} + \left(w_j - \frac{\alpha_j}{m} \right) \begin{bmatrix} \tilde{a}_j & \tilde{b}'_j \\ \tilde{b}_j & \tilde{C}_j \end{bmatrix} = 0.$$

For the implications of $x' V U = 0$ we consider

$$\bar{x}'_j \bar{V}_j \bar{U}_j = e'_j \otimes 4(\sigma_j^2 b'_j + \sigma_{j12} C_j) \left(\frac{\alpha_j}{m} (\Sigma_{j22} - \phi_j \phi'_j) + w_j \Pi_j \Pi'_j \right),$$

$$x_j^{\perp'} V_j^{\perp} U_j^{\perp} = \frac{\alpha_j e'_j}{\alpha_j - m w_j} \otimes 4 \left(w_j - \frac{\alpha_j}{m} \right) (\sigma_j^2 \tilde{b}'_j + \sigma_{j12} \tilde{C}_j) (\Sigma_{j22} - \phi_j \phi'_j).$$

So,

$$\begin{aligned} x' V U (e_i \otimes I_g) &= \left(\sum_{j=1}^m x'_j V_j U_j \right) (e_i \otimes I_g) \\ &= 4(\sigma_i^2 b'_i + \sigma_{i12} C_i) \left(\frac{\alpha_i}{m} (\Sigma_{i22} - \phi_i \phi'_i) + w_i \Pi_i \Pi'_i \right) + \\ &\quad \frac{4\alpha_i}{\alpha_i - m w_i} \left(w_i - \frac{\alpha_i}{m} \right) (\sigma_i^2 \tilde{b}'_i + \sigma_{i12} \tilde{C}_i) (\Sigma_{i22} - \phi_i \phi'_i) \\ &= 4\sigma_i^2 w_i b'_i \Pi_i \Pi'_i + \frac{4\alpha_i}{m} (\sigma_i^2 b'_i + \sigma_{i12} C_i) (\Sigma_{i22} - \phi_i \phi'_i) - \\ &\quad \frac{4\alpha_i}{\alpha_i - m w_i} \frac{\alpha_i}{m} (\sigma_i^2 b'_i + \sigma_{i12} C_i) (\Sigma_{i22} - \phi_i \phi'_i), \end{aligned}$$

where the latter equality is due to (A5). So, if $x' V U = 0$, then

$$\begin{aligned}
& x' V U (e_i \otimes I_g) (\sigma_i^2 b_i + C_i \sigma_{i21}) \\
&= 4\sigma_i^4 w_i b_i' \Pi_i \Pi_i' b_i + \\
& \quad \frac{4\alpha_i}{m} (\sigma_i^2 b_i' + \sigma_{i12} C_i) (\Sigma_{i22} - \phi_i \phi_i') (\sigma_i^2 b_i + C_i \sigma_{i21}) + \\
& \quad \frac{4\alpha_i^2}{m(mw_i - \alpha_i)} (\sigma_i^2 b_i' + \sigma_{i12} C_i) (\Sigma_{i22} - \phi_i \phi_i') (\sigma_i^2 b_i + C_i \sigma_{i21}) = 0.
\end{aligned}$$

In the expressions above, the three terms are nonnegative so $b_i' \Pi_i \Pi_i' b_i = 0$, which holds if and only if $b_i' \Pi_i = 0$. As

$$\bar{x}_j' \bar{\Delta}_{j1} = 2w_j b_j' \Pi_j \Pi_j',$$

we find, indeed, that $x' \Delta_2 = 0$ and $x' V U = 0$ imply $x' \Delta_1 = 0$.

As a result the efficiency bound V_L is given by (17), where the submatrices of Δ and U are given in (A1) and (A3). This amounts to the bound (18) given in Theorem 4.1.

Notice that if $\alpha_j = \bar{\alpha}$, the last m columns of Δ_2 reduce to a single column: ψ is differentiated with respect to the scalar $\bar{\alpha}$. In that case both U and (A5) are not affected, so the proof is the same; which also holds true if the vector α is fixed, $\alpha = \alpha^0$, and ψ is not differentiated with respect to α . One may verify that if $\alpha_j = 0$, the columns of Δ are still located in the space spanned by the columns of V , so $V V^+ \Delta = \Delta$. \square

Proof of theorem 4.2 We find Δ_{j1} equal to (A1) and

$$(A6) \quad \Delta_{j2} = \begin{bmatrix} w_j e_j' \otimes [H R'^{-1} \left\{ \begin{pmatrix} 0 \\ I_g \end{pmatrix} \otimes \begin{pmatrix} 0 \\ I_g \end{pmatrix} \right\} (\Pi_j \otimes I_g)] & \frac{\alpha_j}{m} T & \frac{1}{m} e_j' \otimes \text{vec}(\Omega) \\ 0 & (w_j - \frac{\alpha_j}{m}) T & -\frac{1}{m} e_j' \otimes \text{vec}(\Omega) \end{bmatrix}.$$

Let

$$(A7) \quad U_j = \begin{bmatrix} e_j' \\ \frac{\alpha_j}{m(\bar{\alpha}-1)} \end{bmatrix} \otimes [H R \{e_1 \otimes \begin{bmatrix} \frac{-\sigma_{j2}}{\sigma_j^2} \\ I_g \end{bmatrix} \}].$$

Then it can be easily verified that $U' \Delta_2 = 0$ and

$$\bar{V}_j \bar{U}_j = e'_j \otimes \left[2HRR'^{-1} \{ \Sigma e_1 \otimes \begin{pmatrix} 0 \\ I_g \end{pmatrix} \} \left(\frac{\alpha_j}{m} (\Sigma_{22} - \phi\phi') + w_j \Pi_j \Pi'_j \right) \right],$$

$$V_j^\perp U_j^\perp = \frac{\alpha'}{m(\bar{\alpha} - 1)} \otimes \left[2(w_j - \frac{\alpha_j}{m}) HRR'^{-1} \{ \Sigma e_1 \otimes \begin{pmatrix} 0 \\ I_g \end{pmatrix} \} (\Sigma_{22} - \phi\phi') \right].$$

Define vectors $x = (x_1; \dots; x_m)$, $x_j = (\bar{x}_j; x_j^\perp)$, analogous to (A4). Then $x' \Delta_2 = 0$ implies, slightly different from (A5):

$$(A8) \quad C_j \Pi_j = 0,$$

$$\sum_{j=1}^m \left\{ \frac{\alpha_j}{m} \begin{bmatrix} a_j & b'_j \\ b_j & C_j \end{bmatrix} + (w_j - \frac{\alpha_j}{m}) \begin{bmatrix} \tilde{a}_j & \tilde{b}'_j \\ \tilde{b}_j & \tilde{C}_j \end{bmatrix} \right\} = 0.$$

For the implications of $x' VU = 0$ consider

$$\bar{x}'_j \bar{V}_j \bar{U}_j = e'_j \otimes 4(\sigma^2 b'_j + \sigma_{12} C_j) \left(\frac{\alpha_j}{m} (\Sigma_{22} - \phi\phi') + w_j \Pi_j \Pi'_j \right),$$

$$x_j^{\perp'} V_j^\perp U_j^\perp = \frac{\alpha'}{m(\bar{\alpha} - 1)} \otimes 4(w_j - \frac{\alpha_j}{m}) (\sigma^2 \tilde{b}'_j + \sigma_{12} \tilde{C}_j) (\Sigma_{22} - \phi\phi').$$

So,

$$\begin{aligned} x' VU (e_i \otimes I_g) &= \left(\sum_{j=1}^m x'_j V_j U_j \right) (e_i \otimes I_g) \\ &= 4(\sigma^2 b'_i + \sigma_{12} C_i) \left(\frac{\alpha_i}{m} (\Sigma_{22} - \phi\phi') + w_i \Pi_i \Pi'_i \right) + \\ &\quad \frac{4\alpha_i}{m(\bar{\alpha} - 1)} \sum_{j=1}^m (w_j - \frac{\alpha_j}{m}) (\sigma^2 \tilde{b}'_j + \sigma_{12} \tilde{C}_j) (\Sigma_{22} - \phi\phi') \\ &= 4\sigma^2 w_i b'_i \Pi_i \Pi'_i + \frac{4\alpha_i}{m} (\sigma^2 b'_i + \sigma_{12} C_i) (\Sigma_{22} - \phi\phi') - \\ &\quad \frac{4\alpha_i}{m(\bar{\alpha} - 1)} \sum_{j=1}^m \frac{\alpha_j}{m} (\sigma^2 b'_j + \sigma_{12} C_j) (\Sigma_{22} - \phi\phi'), \end{aligned}$$

where the latter equality follows from (A8). So, if $x' VU = 0$, then

$$\begin{aligned}
& \sum_{i=1}^m x' V U (e_i \otimes I_g) (\sigma^2 b_i + C_i \sigma_{21}) \\
&= 4\sigma^4 \sum_{i=1}^m w_i b_i' \Pi_i \Pi_i' b_i + \\
& \quad 4 \sum_{i=1}^m \frac{\alpha_i}{m} (\sigma^2 b_i' + \sigma_{12} C_i) (\Sigma_{22} - \phi \phi') (\sigma^2 b_i + C_i \sigma_{21}) + \\
& \quad \frac{4}{1 - \bar{\alpha}} \sum_{i=1}^m \sum_{j=1}^m \frac{\alpha_j}{m} (\sigma^2 b_j' + \sigma_{12} C_j) (\Sigma_{22} - \phi \phi') (\sigma^2 b_i + C_i \sigma_{21}) \frac{\alpha_i}{m} = 0.
\end{aligned}$$

As all three terms are nonnegative, we find $\sum_{i=1}^m b_i' \Pi_i \Pi_i' b_i = 0$, which implies $b_i' \Pi_i = 0$. As

$$\bar{x}_j' \bar{\Delta}_{j1} = 2w_j b_j' \Pi_j \Pi_j',$$

we find that $x' \Delta_2 = 0$ and $x' V U = 0$ imply $x' \Delta_1 = 0$. So V_L is given by (17), where the submatrices of Δ_1 and U are given by (A1) and (A7), which amounts to the bound (19) given in Theorem 4.2.

The remarks made at the end of the proof of Theorem 2 with respect to special cases where $\alpha_j = \bar{\alpha}$, $\alpha = \alpha^0$, or $\alpha_j = 0$, also apply in the homoscedastic case. \square

Appendix 3
Proofs of Theorems 5.1 and 5.2

In both proofs we use the fact, known from minimum chi-square estimation,

$$n^{1/2} \left(\hat{\delta}; \hat{\gamma}; \hat{\alpha} \right) - (\delta; \gamma; \alpha) \stackrel{A}{\sim} N(0, (\Delta' W^{-1} \Delta)^{-1}),$$

where $\gamma_j = \Omega_j(1; -\delta)$ and $\Delta = (\Delta_1, \Delta_2)$;

$$\Delta_1 = \text{plim} \left(\frac{\partial a}{\partial \delta'} \right), \quad \Delta_2 = \text{plim} \left(\frac{\partial a}{\partial (\gamma', \alpha')} \right).$$

So the asymptotic distribution of $n^{1/2} \hat{\delta}$ is given by

$$n^{1/2} (\hat{\delta} - \delta) \stackrel{A}{\sim} N(0, (I_g, 0) (\Delta' W^{-1} \Delta)^{-1} (I_g, 0)).$$

Now let U be a matrix such that $U' \Delta_2 = 0$, $U' \Delta_1$ has full column rank and $U' W U$ is nonsingular, then

$$\begin{aligned} \Delta' W^{-1} \Delta &\geq \Delta' W^{-1/2} (W^{1/2} U (U' W U)^{-1} U' W^{1/2}) W^{-1/2} \Delta \\ &= \begin{bmatrix} \Delta_1' U (U' W U)^{-1} U' \Delta_1 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

Using results on the positive semi-definiteness of partitioned matrices (Bekker, 1988, Theorem 3), we find

$$(A9) \quad (I_g, 0) (\Delta' W^{-1} \Delta)^{-1} (I_g, 0) \leq (\Delta_1' U (U' W U)^{-1} U' \Delta_1)^{-1}.$$

For both the hetero- and the homoscedastic case we give such matrices U for which the right-hand-side of (A9) is equal to V_L as given in Theorems 2 and 3, respectively. As V_L is a lower bound, it follows that (A9) is an equality, which proves Theorems 5.1 and 5.2.

proof of Theorem 5.1: We find

$$\Delta_{j1} = \text{plim} \left(\frac{\partial a_j}{\partial \delta'} \right) = \begin{bmatrix} -w_j(\delta, I_g)' \Pi_j \Pi_j' - \frac{\alpha_j}{m} \Omega_j(0; I_g) \\ -(w_j - \frac{\alpha_j}{m}) \Omega_j(0; I_g) \end{bmatrix},$$

(A10)

$$\Delta_{j2} = \text{plim} \left(\frac{\partial a_j}{\partial (\gamma', \alpha')} \right) = \begin{bmatrix} -e_j' \otimes \frac{\alpha_j}{m} I_{g+1} & -e_j' \otimes \frac{1}{m} \gamma_j \\ -e_j' \otimes (w_j - \frac{\alpha_j}{m}) I_{g+1} & e_j' \otimes \frac{1}{m} \gamma_j \end{bmatrix}.$$

Let $U = (U_1; \dots; U_m)$,

$$U_j = \begin{bmatrix} e_j' \\ \alpha_j e_j' \\ \alpha_j - m w_j \end{bmatrix} \otimes [\Omega_j^{-1}(\delta, I_g)'],$$

then it can be easily verified that $U' \Delta_2 = 0$ and

$$(\Delta_1' U (U' W U)^{-1} U' \Delta_1)^{-1} = V_L,$$

where V_L equals (18) as given in Theorem 3.1. □

Proof of Theorem 5.2: We find Δ_{j1} equal to (A10), with Ω_j replaced by Ω , and

$$\Delta_{j2} = \begin{bmatrix} -\frac{\alpha_j}{m} I_{g+1} & -e_j' \otimes \frac{1}{m} \gamma \\ -(w_j - \frac{\alpha_j}{m}) I_{g+1} & e_j' \otimes \frac{1}{m} \gamma \end{bmatrix}.$$

Let $U = (U_1; \dots; U_m)$,

$$U_j = \begin{bmatrix} e_j' \\ \alpha_j' \\ m(\alpha_j - 1) \end{bmatrix} \otimes [\Omega^{-1}(\delta, I_g)'],$$

then it can be easily verified, if $\gamma = \Omega(1; -\delta)$, that $U' \Delta_2 = 0$ and

$$(\Delta_1' U (U' W U)^{-1} U' \Delta_1)^{-1} = V_L,$$

where V_L equals (19) as given in Theorem 3.2. □

Appendix 4

Lemma 6.1 *Let $A \geq 0$, $C > 0$, $\lambda > 0$, $x = AA^+x$, then*

$$\lim_{\lambda \rightarrow 0} x'(A + \lambda C)^{-1}x = x'A^+x.$$

PROOF. As

$$\begin{bmatrix} x'A^+x & x' \\ x & A + \lambda C \end{bmatrix} = \begin{bmatrix} x'A^+ \\ I \end{bmatrix} A(A^+x, I) + \lambda \begin{pmatrix} 0 \\ I \end{pmatrix} C(0, I) \geq 0,$$

we find (Bekker, 1988, Theorem 1) that

$$x'(A + \lambda C)^{-1}x \leq x'A^+x.$$

As the left-hand-side is increasing as $\lambda \rightarrow 0$, we find that its limit exists:

$$(A11) \quad \lim_{\lambda \rightarrow 0} x'(A + \lambda C)^{-1}x = q \leq x'A^+x.$$

Furthermore

$$\begin{bmatrix} x'(A + \lambda C)^{-1}x & x' \\ x & A + \lambda C \end{bmatrix} \geq 0.$$

So the limit of this matrix is positive semi-definite:

$$\begin{bmatrix} q & x' \\ x & A \end{bmatrix} \geq 0.$$

Hence

$$(A12) \quad q \geq x'A^+x.$$

Together (A11) and (A12) imply the result in Lemma 6.1. □

For the application in Section 6 notice that if $x = Ay$, for some vector y , so that $AA^+x = x$, and $0 \leq A \leq A + B^*$, then also $x = (A + B^*)(A + B^*)^+x$.

Lemma 6.2 *Let x and y be vectors such that $\text{rank}(x, y) = 2$. Let $y \rightarrow x$, then*

$$\lim_{y \rightarrow x} x'(xx' + yy')^+x = 1 \neq x'(2xx')^+x = 1/2.$$

PROOF. If $\text{rank}(x, y) = 2$, we find (Ouellette, 1981, Theorem 4.3)

$$\text{rank} \begin{bmatrix} 1 & x' \\ x & xx' + yy' \end{bmatrix} =$$

$$\text{rank}(xx' + yy') + \text{rank}(1 - x'(xx' + yy')^+x) = 2.$$

So, $x'(xx' + yy')^+x = 1$, for any $y \neq x \neq 0$. □

References

- Anderson, T.W., N. Kunitomo and T. Sawa (1982): Evaluation of the distribution function of the limited-information maximum likelihood estimator, *Econometrica* 50, 1009-1027.
- Angrist, J.D. (1990): Lifetime earnings and the Vietnam era draft lottery: Evidence from social security administrative records, *The American Economic Review* 80, 313-336.
- Angrist, J.D., G.W. Imbens and A. Krueger (1994): Jackknife instrumental variable estimation, working-paper (MIT, Harvard, Princeton, NBER).
- Angrist, J.D., and A. Krueger (1995): Split sample instrumental variables estimates of the return to schooling, Forthcoming in *Journal of Business and Economic Statistics*.
- Balestra, P. (1976): *La dérivation matricielle*, Collection de l'IME, No. 12. Sirey, Paris.
- Bekker, P.A. (1988): The positive semidefiniteness of partitioned matrices, *Linear Algebra Appl.* 111, 261-278.
- Bekker, P.A. (1994): Alternative approximations to the distributions of instrumental variable estimators, *Econometrica* 62, 657-681.
- Bekker, P.A., and J. van der Ploeg (1994): Grouping Estimators, University of Groningen, *SOM Research Report* 94309.
- Deaton, A. (1985): Panel data from time series of cross-sections, *Journal of Econometrics* 30, 109-126.
- Eaton, M.L. (1983): *Multivariate statistics*, John Wiley & Sons.
- Fujikoshi, Y., K. Morimune, N. Kunitomo and M. Taniguchi (1982): Asymptotic expansions of the distributions of the estimates of coefficients in a simultaneous equation system, *Journal of Econometrics* 18, 191-205.
- Kunitomo, N. (1980): Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations, *JASA* 75, 693-700.
- Kunitomo, N. (1981): Asymptotic optimality of the limited information maximum likelihood estimator in large econometric models, *Economic Studies Quarterly* 32, 247-266.
- Kunitomo, N. (1982): Asymptotic efficiency and higher order efficiency of the limited information maximum likelihood estimator in large econometric models: *Technical Report* No. 365, Institute for Mathematical Studies in the Social Sciences, Stanford University.
- Kunitomo, N. (1986): Comparing some modified maximum likelihood estimators of a slope coefficient in a linear functional relationship, *J. Japan Statist. Soc.* 16, 173-185.
- Kunitomo, N. (1987): A third order optimum property of the ML estimator in a linear functional relationship and simultaneous equation system in econometrics, *Annals of the Institute of Statistical Mathematics* 39, 575-591.
- Morimune, K., and N. Kunitomo (1980): Improving the maximum likelihood estimate in linear functional relationships for alternative parameter sequences, *JASA* 73, 867-871.
- Muirhead, R.J. (1982): *Aspects of multivariate statistical theory*, John Wiley & Sons.

Ouellette, D.V. (1981): Schur complements and statistics, *Linear Algebra Appl.* 36, 187-295.