ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 10 — 1988

A Note on Instrumental Variables and Maximum Likelihood Estimation Procedures

Alberto HOLLY, Jan R. MAGNUS *

ABSTRACT. – This paper presents two alternative formulations of the instrumental variables (IV) procedure. One of the formulations, termed the "completed model", serves to show, in a general way, that the IV estimator of a vector of regression parameter is asymptotically as efficient as the maximum likelihood estimator, under the hypothesis of normality. The expressions for the information matrix and its inverse are obtained. The calculation of this inverse requires some new results of matrix algebra that are demonstrated in an appendix.

Note sur les procédures d'estimation par variables instrumentales et par le maximum de vraisemblance

RÉSUMÉ. – Cet article présente deux formulations alternatives de la procédure d'estimation par variables instrumentales. L'une des formulations, appelée ici « modèle complété » permet de montrer que, de façon générale, l'estimateur de variables instrumentales traditionnel est asymptotiquement aussi efficace que l'estimateur du maximum de vraisemblance sous l'hypothèse de normalité. L'expression de la matrice d'information et celle de son inverse sont obtenues. Le calcul de cette inverse nécessite l'emploi de résultats matriciels nouveaux qui sont démontrés en annexe de cet article.

* A. HOLLY : Département d'Économétrie et d'Économie Politique, Université de Lausanne, École des HEC, 1015-Lausanne, Switzerland; J. R. MAGNUS : Department of Economics, The London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom. We would like to thank Alain TROGNON for clarifying our confusion concerning Proposition 2 in an earlier version of this paper. We would also like to thank Alain MONFORT and two anonymous referees for their constructive comments.

1 Introduction

The instrumental variables (IV) procedure is traditionnally formulated by assuming the existence of instruments which are asymptotically uncorrelated with the disturbance term and correlated with the explanatory variables. In this paper, we present an alternative formulation in terms of a "completed model". The equivalence of these two formulations does not seem to be known in the econometric literature.

By using the "completed model" framework, this paper also shows that the IV estimator of a vector of regression parameters is asymptotically as efficient as the maximum likelihood (ML) estimator. This result is well known in the context of simultaneous equations, but we show that it holds in a rather more general setting.

The two alternative formulations are presented in sections 2 and 3, and the result about asymptotic efficiency of IV is explained in section 4. Four appendices, of independent interest, accompany the paper. In Appendix 1 some new results on the duplication matrix are presented; in Appendix 2 we derive the asymptotic information matrix; in Appendix 3 we prove Proposition 3; and in Appendix 4 we apply the results of Appendices 1-3 and obtain the complete inverse of the asymptotic information matrix.

2 Two Alternative Formulations of IV: the Geometric Approach

In the standard formulation of IV, the regularity conditions are written in terms of probability limits. In this section we shall adopt an alternative formulation due to Monfort [1978] based on geometric properties of the Hilbert space $\mathscr{L}^2$ of square integrable variables defined in some probability space $(\Omega, \mathscr{A}, P)$. The basic reason for this choice is that it will enable us to show an equivalence between two alternative formulations of IV which does not depend on the observations but rather on model formulation. In the following section we shall indicate the main difference between the geometric and the probability limit approaches with regard to these two alternative formulations of IV.

Let y be an observable random variable belonging to $\mathscr{L}^2$, z an $n \times 1$ observable random vector whose components belong to $\mathscr{L}^2$, x an observable $K \times 1$ random vector whose components, called instrumental variables, also belong to $\mathscr{L}^2$.

The random variable y and the random vectors z and x are supposed to be linked by the equations [1]

$$y = z'\delta_0 + u \tag{1}$$

$$E(xu) = 0 \tag{2}$$

where δ_0 is an unknown $n \times 1$ vector and u is defined by (1).

Since some components of x and z may be the same, we shall write z and x as

$$z = (y_1', x_1')'$$

$$x = (x_1', x_2')'$$

where y_1 is an $N \times 1$ vector, x_1 is a $K_1 \times 1$ vector (so that $n = N + K_1$), and x_2 is a $K_2 \times 1$ vector with $K = K_1 + K_2$. We also write

$$\delta_0 = (\alpha_0', \beta_0')'$$

where α_0 is an $N \times 1$ vector and β_0 is a $K_1 \times 1$ vector.

It should be noted at this stage that, according to Theorem 2 in MONFORT [1978], equation (1) is identifiable if and only if rank $E(xz') = N + K_1$. For this reason, we shall adopt the following set-up for the first formulation of IV:

$$\text{(IV 1)} \qquad \begin{cases} y = y_1' \alpha_0 + x_1' \beta_0 + u \\ x = (x_1', x_2')' \\ E(xu) = 0 \\ \text{rank } E[x(y_1', x_1')] = N + K_1 \end{cases}$$

Now, by analogy with the formulation of simultaneous equations models, it is natural to consider a "completed model" where one adds to equation (1) a multivariate regression type equation for y_1. More specifically, if we adopt the following set-up for the second formulation of IV,

$$\text{(IV 2)} \qquad \begin{cases} y = y_1' \alpha_0 + x_1' \beta_0 + u \\ y_1 = \Pi_{10}' x_1 + \Pi_{20}' x_2 + v \\ x = (x_1', x_2')' \\ E(xu) = 0 \\ E(xv') = 0 \\ \text{rank}(\Pi_{20}) = N, \end{cases}$$

then the following result emerges.

PROPOSITION 1: Assuming that $E(xx')$ is nonsingular, (IV 1) and (IV 2) are equivalent.

---

1. We adopt a notation analogous to the one commonly used in the context of simultaneous equations models.

Proof: Let

$$\Pi_0^* = \begin{pmatrix} \Pi_{10} & I_{K_1} \\ \Pi_{20} & 0 \end{pmatrix} \tag{3}$$

and note that

$$\operatorname{rank}(\Pi_0^*) = \operatorname{rank}(\Pi_{20}) + K_1.$$

Now assume (IV 2). Then

$$x(y_1', x_1') = xx' \Pi_0^* + (xv', 0),$$

so that

$$E\, x(y_1', x_1') = (E\, xx') \Pi_0^*,$$

and hence

$$\operatorname{rank}[E\, x(y_1', x_1')] = \operatorname{rank}(\Pi_0^*) = N + K_1.$$

Next assume (IV 1). We can always project y_1 orthogonaly on $\mathcal{M}(x)$, the closed subspace of $\mathcal{L}^2$ spanned by the components of x. Let $\Pi_{10}' x_1 + \Pi_{20}' x_2$ be this projection. Defining v as $y_1 - \Pi_{10}' x_1 - \Pi_{20}' x_2$ we have, by construction, $E(xv') = 0$. In addition,

$$\operatorname{rank}[E\, x(y_1', x_1')] = \operatorname{rank}[(E\, xx') \Pi_0^*] = \operatorname{rank}(\Pi_0^*) = N + K_1$$

and hence

$$\operatorname{rank}(\Pi_{20}) = N. \quad \square$$

3 The Probability Limit Approach

Suppose now that we have T observations on y, y_1, x_1, x_2 which we write as y_t, $y_{t\,1}$, $x_{t\,1}$, $x_{t\,2}$ $(t = 1, \ldots, T)$.

Define

$$\begin{aligned} u &= (u_1, \ldots, u_t, \ldots, u_T)' \\ y &= (y_1, \ldots, y_t, \ldots, y_T)' \\ Y &= (y_{11}, \ldots, y_{t\,1}, \ldots, y_{T\,1})' \\ X_1 &= (x_{11}, \ldots, x_{t\,1}, \ldots, x_{T\,1})' \\ X_2 &= (x_{12}, \ldots, x_{t\,2}, \ldots, x_{T\,2})' \\ V &= (v_{11}, \ldots, v_{t\,1}, \ldots, v_{T\,1})' \\ X &= (X_1, X_2) \end{aligned}$$

$$Z=(Y, X_1)$$

$$\Pi_0=(\Pi'_{10}, \Pi'_{20})'.$$

The T observations on the "completed model" may be written compactly as

$$\begin{cases} y=Z\,\delta_0+u \\ Y=X\,\Pi_0+V. \end{cases}$$

Let us now show the difference between the geometric approach to IV and the standard set-up (using probability limits).

The standard IV framework is

$$\text{(IV 1)}' \quad \begin{cases} y=Y\,\alpha_0+X_1\,\beta_0+u \\ Z=(Y, X_1) \\ X=(X_1, X_2) \\ \operatorname{plim}(X'u/T)=\mathcal{M}_{Xu}=0 \\ \operatorname{plim}(X'Z/T)=\mathcal{M}_{XZ} \quad \text{is a } K\times(N+K_1) \text{ matrix of rank } N+K_1. \end{cases}$$

Now, consider, similarly to (IV 2), the following formulation:

$$\text{(IV 2)}' \quad \begin{cases} y=Y\,\alpha_0+X_1\,\beta_0+u \\ Y=X_1\,\Pi_{10}+X_2\,\Pi_{20}+V \\ X=(X_1, X_2) \\ \operatorname{plim}(X'u/T)=\mathcal{M}_{Xu}=0 \\ \operatorname{plim}(X'V/T)=\mathcal{M}_{XV}=0 \\ \operatorname{rank}(\Pi_{20})=N. \end{cases}$$

Then we can prove the following result as the counterpart to Proposition 1.

PROPOSITION 2: Assuming that $\operatorname{plim}(X'X/T)=\mathcal{M}_{XX}$ is nonsingular, (IV 1)′ and (IV 2)′ are equivalent.

Proof: Define Π_0^* as in (3). Using a similar argument as in the proof of Proposition 1, it is easy to show that (IV 2)′ implies (IV 1)′.

Now assume (IV 1)′. Partition $\mathcal{M}_{XZ}$ as $\mathcal{M}_{XZ}=(\mathcal{M}_{XY}, \mathcal{M}_{XX_1})$ and define Π_0 and V as

$$\Pi_0=\mathcal{M}_{XX}^{-1}\,\mathcal{M}_{XY} \qquad \text{and} \qquad V=Y-X\,\Pi_0.$$

It is easy to see that

$$\operatorname{plim}(X'V/T)=0.$$

This implies that

$$\operatorname{plim}(X'Z/T)=\mathcal{M}_{XX}\,\Pi_0^*.$$

It follows that

$$\operatorname{rank}(\Pi_0^*) = \operatorname{rank} \operatorname{plim}(X'Z/T) = N + K_1,$$

and hence

$$\operatorname{rank}(\Pi_{20}) = N. \quad \square$$

4 Asymptotic Efficiency of IV

The previous two sections dealt exclusively with model formulation and not with estimation. Let us now consider the estimation problem and, in particular, let us compare the asymptotic efficiency of the IV estimator of δ_0 in model (IV 1) or (IV 1)′ with the asymptotic efficiency of the ML estimator of δ_0 in the "completed model" (IV 2) or (IV 2)′.

To estimate the parameters of (IV 2) we shall assume that the random vectors $(u_t, v_t', x_t')'$ $(t = 1, \ldots, T)$ are independent and identically distributed as a normal distribution. Moreover, the conditions $E(x_t u_t) = 0$ and $E(x_t v_t') = 0$ imply that x_t is stochastically independent from (u_t, v_t'). It is easy to verify, under these conditions, that the ML estimators of the parameters of the "completed model" are also the ML estimators of the likelihood function of (y, Y) conditional on X. In other words, the exogeneity property of $\{x_1, \ldots, x_t, \ldots, x_T\}$ implied by the above conditions permit X to be treated as fixed for purposes of inference about the parameters of the "completed model". This explains why the result stated below holds for (IV 2) as well as for (IV 2)′ where the x_t's may not be random.

It is useful to note that under the normality assumption a weak law of large numbers applies and that we can identify $E(xu)$, $E(xz')$ and $E(xx')$ with $\mathscr{M}_{Xu}$, $\mathscr{M}_{XZ}$ and $\mathscr{M}_{XX}$ respectively. We have the following general result.

Proposition 3: Consider (IV 2) and assume that $E(xx')$ is nonsingular [or consider (IV 2)′ and assume that $\mathscr{M}_{XX}$ is nonsingular]. Assume further that the T rows of (u, V) are independent and identically distributed as $\mathscr{N}(0, \Psi_0)$ where Ψ_0 is a positive definite $(K_1+1) \times (K_1+1)$ matrix partitioned as

$$\Psi_0 = \begin{pmatrix} \sigma_0^2 & \theta_0' \\ \theta_0 & \Omega_0 \end{pmatrix}.$$

Let $\hat{\delta}_{MLE}$ denote the ML estimator of δ_0. Then [2]

$$\mathscr{A}\mathscr{D}[T^{1/2}(\hat{\delta}_{MLE}-\delta_0)]=\mathscr{N}(0, \sigma_0^2 G_0^{-1})$$

with

$G_0 = \text{plim}\,[T^{-1} Z' X (X'X)^{-1} X' Z] = \mathscr{M}_{ZX} \mathscr{M}_{XX}^{-1} \mathscr{M}_{XZ}$.

Proof: See Appendix 3. □

As an immediate consequence of Proposition 3 we have

PROPOSITION 4: Let $\hat{\delta}_{IV}$ be the IV estimator of δ_0. Under the assumptions of Proposition 3, $\hat{\delta}_{IV}$ is asymptotically efficient.

Proof: This result follows from the fact that $\sigma_0^2 G_0^{-1}$ is also the covariance matrix of the asymptotic distribution of $T^{1/2}(\hat{\delta}_{IV}-\delta_0)$. □

The reader familiar with the econometric literature on simultaneous equation models will recognise that the "completed model" (IV 2)′ is often referred to as the "incomplete system" (see, for example, GODFREY and WICKENS [1982]). It is well known that the ML estimator of δ_0 in the "incomplete system" and the Limited Information Maximum Likelihood Estimator (LIML) are identical. The results stated in Proposition 3 and 4 are not new in the context of simultaneous equation models. They show, however, that they are valid for the general IV procedure traditionally formulated as (IV 1) or (IV 1)′.

The reader will note that in order to obtain the expression for G_0 we need the entire information matrix of the "completed model". This matrix is derived in Appendix 2 [3]. The information matrix then needs to be inverted and this requires new properties of the duplication matrix which are given in Appendix 1. Only part of the inverted matrix is needed in the proof of Proposition 3 (see Appendix 3); the complete inverse, which gives the asymptotic variance and covariance matrices of all parameters in the "completed model", is derived in Appendix 4 and will prove useful in other contexts as well.

2. The abreviation $\mathscr{A}\mathscr{D}$ is used to mean the asymptotic distribution of a random vector.
3. This matrix corresponds, with different notation (and a correction of sign), to the information matrix derived by SMITH [1983], [1985] in the context of LIML interpreted as FIML on the « incomplete system ».

Some New Properties of the Duplication Matrix

For any square matrix A, we denote by $v(\mathrm{A})$ the vector that is obtained from vec A by eliminating all supradiagonal elements of A. For example, if A is a 3×3 matrix,

$$\mathrm{vec\ A}=(a_{11}\quad a_{21}\quad a_{31}\quad a_{12}\quad a_{22}\quad a_{32}\quad a_{13}\quad a_{23}\quad a_{33})'$$

and

$$v(\mathrm{A})=(a_{11}\quad a_{21}\quad a_{31}\quad a_{22}\quad a_{32}\quad a_{33})'.$$

Thus, if A is an $n\times n$ matrix, then $v(\mathrm{A})$ is a $\frac{1}{2}\,n(n+1)\times 1$ vector. The unique $n^2\times\frac{1}{2}\,n(n+1)$ matrix which transforms, for any *symmetric* $n\times n$ matrix A, $v(\mathrm{A})$ into vec A is called the *duplication matrix* and is denoted D_n. Thus,

(4) $$\mathrm{D}_n\,v(\mathrm{A})=\mathrm{vec\ A}\qquad(\mathrm{A}=\mathrm{A}').$$

The duplication matrix was studied *inter alia* by MAGNUS and NEUDECKER [1980]. A brief review is given in MAGNUS and NEUDECKER [1986] and a complete treatment in Magnus [1988]. We recall the following properties for an arbitrary $n\times n$ matrix A (not necessarily symmetric) and an $n\times 1$ vector b:

(5) $$\mathrm{D}_n^+=(\mathrm{D}_n'\mathrm{D}_n)^{-1}\mathrm{D}_n'$$

(6) $$\mathrm{D}_n'\,\mathrm{vec\ A}=v(\mathrm{A}+\mathrm{A}'-\mathrm{dg\,A})$$

(7) $$\mathrm{D}_n\mathrm{D}_n^+(b\otimes\mathrm{A})=\frac{1}{2}(b\otimes\mathrm{A}+\mathrm{A}\otimes b)$$

(8) $$\mathrm{D}_n\mathrm{D}_n^+(\mathrm{A}\otimes\mathrm{A})\,\mathrm{D}_n^{+\prime}=(\mathrm{A}\otimes\mathrm{A})\,\mathrm{D}_n^{+\prime}$$

where dg A denotes the diagonal matrix containing the diagonal elements of A. Also, if A is nonsingular,

(9) $$(\mathrm{D}_n'(\mathrm{A}^{-1}\otimes\mathrm{A}^{-1})\,\mathrm{D}_n)^{-1}=\mathrm{D}_n^+(\mathrm{A}\otimes\mathrm{A})\,\mathrm{D}_n^{+\prime}.$$

Properties (5)-(9) can be found in MAGNUS [1988] as Theorem 4.1 (ii), Theorem 4.7 (iv), Exercice 4.5 [use Theorem 3.11 (v) and 4.2 (iii)] and Theorem 4.11 (ii).

We shall now present two new lemmas and one corollary which describe the relationship between D_{n+1} and D_n; these results are applied in Appendix 4.

LEMMA 5: Let

$$A_1 = \begin{pmatrix} \alpha & a' \\ a & A \end{pmatrix}, \qquad B_1 = \begin{pmatrix} \beta & b' \\ b & B \end{pmatrix},$$

where A and B are symmetric $n \times n$ matrices, a and b are $n \times 1$ vectors and α and β are scalars. Then

$$\text{(10)} \quad D'_{n+1}(A_1 \otimes B_1) D_{n+1} = \begin{pmatrix} \alpha\beta & \alpha b' + \beta a' & (a' \otimes b') D_n \\ \alpha b + \beta a & \alpha B + \beta A + ab' + ba' & (a' \otimes B + b' \otimes A) D_n \\ D'_n (a \otimes b) & D'_n (a \otimes B + b \otimes A) & D'_n (A \otimes B) D_n \end{pmatrix}$$

and

$$\text{(11)} \quad D^+_{n+1}(A_1 \otimes B_1) D^{+\prime}_{n+1} = \begin{pmatrix} \alpha\beta & \frac{1}{2}(\alpha b' + \beta a') & (a' \otimes b') D^{+\prime}_n \\ \frac{1}{2}(\alpha b + \beta a) & \frac{1}{4}(\alpha B + \beta A + ab' + ba') & \frac{1}{2}(a' \otimes B + b' \otimes A) D^{+\prime}_n \\ D^+_n (a \otimes b) & \frac{1}{2} D^+_n (a \otimes B + b \otimes A) & D^+_n (A \otimes B) D^{+\prime}_n \end{pmatrix}$$

Proof: Let X_1 be an arbitrary symmetric $(n+1) \times (n+1)$ matrix partitioned conformably with A_1 and B_1 as

$$X_1 = \begin{pmatrix} \xi & x' \\ x & X \end{pmatrix}.$$

Then,

$$\text{(12)} \quad \operatorname{tr} A_1 X_1 B_1 X_1 = (\operatorname{vec} X_1)'(A_1 \otimes B_1)(\operatorname{vec} X_1) = (v(X_1))' D'_{n+1}(A_1 \otimes B_1) D_{n+1} v(X_1)$$

and also

$$\text{(13)} \quad \operatorname{tr} A_1 X_1 B_1 X_1 = \alpha\beta\xi^2 + 2\xi(\alpha b' x + \beta a' x) + \alpha x' B x + \beta x' A x + 2(a' x)(b' x) + 2\xi a' X b + 2(x' BX a + x' AX b) + \operatorname{tr} AXBX$$
$$= \begin{pmatrix} \xi \\ x \\ v(X) \end{pmatrix}' \times \begin{pmatrix} \alpha\beta & \alpha b' + \beta a' & (a' \otimes b') D_n \\ \alpha b + \beta a & \alpha B + \beta A + ab' + ba' & (a' \otimes B + b' \otimes A) D_n \\ D'_n (a \otimes b) & D'_n (a \otimes B + b \otimes A) & D'_n (A \otimes B) D_n \end{pmatrix} \begin{pmatrix} \xi \\ x \\ v(X) \end{pmatrix}.$$

The first result now follows from (12) and (13), the symmetry of all matrices concerned and the fact that

$$(v(X_1))' = (\xi, x', (v(X))').$$

By letting $A_1 = B_1 = I_{n+1}$, we obtain as a special case of (10):

$$D'_{n+1} D_{n+1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2I_n & 0 \\ 0 & 0 & D'_n D_n \end{pmatrix} \tag{14}$$

and hence

$$(D'_{n+1} D_{n+1})^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} I_n & 0 \\ 0 & 0 & (D'_n D_n)^{-1} \end{pmatrix}. \tag{15}$$

Pre-and postmultiplying (10) by (15) yields (11) using (5). □

LEMMA 6: Let

$$A_1 = \begin{pmatrix} \alpha & b' \\ a & A \end{pmatrix},$$

where A is an $n \times n$ matrix (not necessarily symmetric), a and b are $n \times 1$ vectors and α is a scalar. Then

$$D'_{n+1} \operatorname{vec} A_1 = \begin{pmatrix} \alpha \\ a+b \\ D'_n \operatorname{vec} A \end{pmatrix}, \qquad D^+_{n+1} \operatorname{vec} A_1 = \begin{pmatrix} \alpha \\ \frac{1}{2}(a+b) \\ D^+_n \operatorname{vec} A \end{pmatrix}.$$

Proof: We have, using (6),

$$D'_{n+1} \operatorname{vec} A_1 = v(A_1 + A'_1 - \operatorname{dg} A_1)$$

$$= \begin{pmatrix} \alpha \\ a \\ v(A) \end{pmatrix} + \begin{pmatrix} \alpha \\ b \\ v(A') \end{pmatrix} - \begin{pmatrix} \alpha \\ 0 \\ v(\operatorname{dg} A) \end{pmatrix} = \begin{pmatrix} \alpha \\ a+b \\ D'_n \operatorname{vec} A \end{pmatrix}.$$

Also, using (15),

$$D^+_{n+1} \operatorname{vec} A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} I_n & 0 \\ 0 & 0 & (D'_n D_n)^{-1} \end{pmatrix} \begin{pmatrix} \alpha \\ a+b \\ D'_n \operatorname{vec} A \end{pmatrix}$$

$$= \begin{pmatrix} \alpha \\ \frac{1}{2}(a+b) \\ D^+_n \operatorname{vec} A \end{pmatrix}.$$

COROLLARY 7: Let A be an $n \times p$ matrix and b a $p \times 1$ vector. Then

$$D'_{n+1}\begin{pmatrix} b' \\ A \\ 0_1 \end{pmatrix}=\begin{pmatrix} b' \\ A \\ 0_2 \end{pmatrix}, \qquad D^+_{n+1}\begin{pmatrix} b' \\ A \\ 0_1 \end{pmatrix}=\begin{pmatrix} b' \\ \frac{1}{2}A \\ 0_2 \end{pmatrix},$$

where 0_1 and 0_2 denote null matrices of orders $n(n+1)\times p$ and $\frac{1}{2}n(n+1)\times p$ respectively.

Proof: Let β_i be the ith component of b and let a_i be the ith column of A $(i=1, \ldots, p)$. Define the $(n+1)\times(n+1)$ matrice

$$C_i=\begin{pmatrix} \beta_i & 0' \\ a_i & 0 \end{pmatrix} \qquad (i=1, \ldots, p).$$

Then, using Lemma 6,

$$\text{vec } C_i=\begin{pmatrix} \beta_i \\ a_i \\ 0 \end{pmatrix}, \qquad D'_{n+1} \text{ vec } C_i=\begin{pmatrix} \beta_i \\ a_i \\ 0 \end{pmatrix}, \qquad D^+_{n+1} \text{ vec } C_i=\begin{pmatrix} \beta_i \\ \frac{1}{2}a_i \\ 0 \end{pmatrix}$$

for $i=1, \ldots, p$, and the result follows from the fact that

$$\begin{pmatrix} b' \\ A \\ 0_1 \end{pmatrix}=(\text{vec } C_1, \text{vec } C_2, \ldots, \text{vec } C_p). \quad \square$$

The Asymptotic Information Matrix

We consider the model

(16) $$y = Z\delta_0 + u_0$$

(17) $$Y = X\Pi_0 + V_0$$

where δ_0 and Π_0 denote the true (but unknown) values of the parameter vector δ and parameter matrix Π to be estimated. We partition

$$X = (X_1, X_2), \qquad Z = (Y, X_1),$$
$$\Pi = (\Pi_1', \Pi_2')', \qquad \delta = (\alpha', \beta')'.$$

The orders of X_1 and Y are $T \times K_1$ and $T \times N$ respectively. We assume that

(i) X_1 and X_2 are nonrandom "fixed" matrices such that $X'X/T$ tends to a positive definite matrix $\mathscr{M}_{XX}$ as $T \to \infty$;

(ii) $\operatorname{plim}(X' u_0/T) = 0$, $\operatorname{plim}(X' V_0/T) = 0$;

(iii) Π_{20} has full column-rank; and

(iv) the T rows of (u_0, V_0) are i. i. d. $N(0, \Psi_0)$, where Ψ_0 (positive definite) is the true value of

(18) $$\Psi = \begin{pmatrix} \sigma^2 & \theta' \\ \theta & \Omega \end{pmatrix}.$$

Defining $u = u(\delta) = y - Z\delta$, $V = V(\Pi) = Y - X\Pi$, and $W = (u, V)$, we obtain the log-likelihood function

(19) $$\mathscr{L} = \mathscr{L}(\delta, \pi, \psi) = \text{Const.} - (T/2)\log|\Psi| - \frac{1}{2}\operatorname{tr} W\Psi^{-1}W',$$

where $\pi = \operatorname{vec}\Pi$ and $\psi = v(\Psi)$, and its first and second differentials

(20) $$\begin{cases} d\mathscr{L} = -(T/2)\operatorname{tr}\Psi^{-1}d\Psi + \frac{1}{2}\operatorname{tr}W\Psi^{-1}(d\Psi)\Psi^{-1}W' \\ \qquad -\operatorname{tr}W\Psi^{-1}(dW)' \\ d^2\mathscr{L} = (T/2)\operatorname{tr}\Psi^{-1}(d\Psi)\Psi^{-1}d\Psi \\ \qquad -\operatorname{tr}W\Psi^{-1}(d\Psi)\Psi^{-1}(d\Psi)\Psi^{-1}W' \\ \qquad +2\operatorname{tr}W\Psi^{-1}(d\Psi)\Psi^{-1}(dW)' - \operatorname{tr}(dW)\Psi^{-1}(dW)', \end{cases}$$

using the (obvious) facts that both Ψ and W are *linear* in the parameters, so that $d^2\Psi = 0$ and $d^2W = 0$. We clearly have

$$\operatorname{plim} W_0' W_0/T = \Psi_0,$$

where $W_0 = (u_0, V_0)$, and also, since $dW = -(Z\,d\delta, X\,d\Pi)$,

$$\text{plim}\,(dW)'\,W_0/T = -\begin{pmatrix} (d\delta)'\,E'\,\Psi_0 \\ 0 \end{pmatrix}$$

and

$$\text{plim}\,(dW)'\,(dW)/T = \begin{pmatrix} (d\delta)'\,\mathcal{M}_{ZZ}\,d\delta & (d\delta)'\,\mathcal{M}_{ZX}\,(d\Pi) \\ (d\Pi)'\,\mathcal{M}_{XZ}\,d\delta & (d\Pi)'\,\mathcal{M}_{XX}\,(d\Pi) \end{pmatrix}$$

where E denotes the $(N+1)\times(N+K_1)$ matrix

(21) $$E = \begin{pmatrix} 0 & 0 \\ I_N & 0 \end{pmatrix}$$

and $\mathcal{M}_{ZZ} = \text{plim}\,(1/T)\,Z'Z$, $\mathcal{M}_{XZ} = \text{plim}\,(1/T)\,X'Z$, and $\mathcal{M}_{ZX} = \text{plim}\,(1/T)\,Z'X$. Hence, writing the inverse of Ψ as

(22) $$\Psi^{-1} = \frac{1}{\eta^2}\begin{pmatrix} 1 & -\theta'\Omega^{-1} \\ -\Omega^{-1}\theta & \eta^2\Omega^{-1} + \Omega^{-1}\theta\theta'\Omega^{-1} \end{pmatrix},$$

where $\eta^2 = \sigma^2 - \theta'\Omega^{-1}\theta$, and letting e denote the $(N+1)\times 1$ vector $(1, 0, \ldots, 0)'$, we obtain

$$\begin{aligned}
-\text{plim}\,(1/T)\,d^2\,\mathcal{L}(\delta_0, \pi_0, \psi_0) &= (1/2)\,\text{tr}\,\Psi_0^{-1}(d\Psi)\,\Psi_0^{-1}(d\Psi) \\
&\quad + 2\,(d\delta)'\,E'\,(d\Psi)\,\Psi_0^{-1}\,e \\
&\quad + (1/\eta_0^2)\,(d\delta)'\,\mathcal{M}_{ZZ}\,d\delta \\
&\quad - (2/\eta_0^2)\,\theta_0'\,\Omega_0^{-1}\,(d\Pi)'\,\mathcal{M}_{XZ}\,d\delta \\
&\quad + (1/\eta_0^2)\,\text{tr}\,(\eta_0^2\,\Omega_0^{-1} \\
&\quad + \Omega_0^{-1}\,\theta_0\,\theta_0'\,\Omega_0^{-1})\,(d\Pi)'\,\mathcal{M}_{XX}\,d\Pi \\
&= (1/2)\,(dv(\Psi))'\,D_{N+1}'\,(\Psi_0^{-1}\otimes\Psi_0^{-1})\,D_{N+1}\,dv(\Psi) \\
&\quad + 2\,(d\delta)'\,(e'\,\Psi_0^{-1}\otimes E')\,D_{N+1}\,dv(\Psi) \\
&\quad + (1/\eta_0^2)\,(d\delta)'\,\mathcal{M}_{ZZ}\,d\delta \\
&\quad - (2/\eta_0^2)\,(d\delta)'\,(\theta_0'\,\Omega_0^{-1}\otimes\mathcal{M}_{ZX})\,d\,\text{vec}\,\Pi \\
&\quad + (1/\eta_0^2)\,(d\,\text{vec}\,\Pi)'\,[(\eta_0^2\,\Omega_0^{-1} + \Omega_0^{-1}\,\theta_0\,\theta_0'\,\Omega_0^{-1}) \\
&\quad \otimes\mathcal{M}_{XX}]\,d\,\text{vec}\,\Pi,
\end{aligned}$$

using the defining property (4) of the duplication matrix. The asymptotic information matrix is therefore

(23) $$I = \begin{pmatrix} I_{\delta\delta} & I_{\delta\pi} & I_{\delta\psi} \\ I_{\pi\delta} & I_{\pi\pi} & 0 \\ I_{\psi\delta} & 0 & I_{\psi\psi} \end{pmatrix},$$

where

(24) $$I_{\delta\delta} = (1/\eta_0^2)\,\mathcal{M}_{ZZ}$$

(25) $$I_{\pi\pi} = (1/\eta_0^2)\,(\eta_0^2\,\Omega_0^{-1} + \Omega_0^{-1}\,\theta_0\,\theta_0'\,\Omega_0^{-1})\otimes\mathcal{M}_{XX}$$

(26) $$I_{\psi\psi} = (1/2)\, D'_{N+1} (\Psi_0^{-1} \otimes \Psi_0^{-1})\, D_{N+1}$$

(27) $$I_{\delta\pi} = -(1/\eta_0^2)\, \theta'_0\, \Omega_0^{-1} \otimes \mathcal{M}_{ZX}$$

(28) $$I_{\delta\psi} = (e'\, \Psi_0^{-1} \otimes E')\, D_{N+1}.$$

We notice that the asymptotic information matrix given above corresponds to SMITH [1983, equation (5)] or SMITH [1985, equation (3)] except that $I_{\delta\psi}$ has the wrong sign in Smith's formulae. We see this immediately if we write

$$I_{\delta\psi} = \begin{pmatrix} I_{\alpha\psi} \\ I_{\beta\psi} \end{pmatrix}$$

and note from (28) that

(29) $$I_{\alpha\psi} = [e'\, \Psi_0^{-1} \otimes (0,\ I_N)]\, D_{N+1}, \qquad I_{\beta\psi} = 0.$$

The Asymptotic Covariance Matrix of $\hat{\delta}$

The asymptotic covariance matrix is the inverse of the asymptotic information matrix. We have from (23)

$$I^{-1} = \begin{pmatrix} I^{\delta\delta} & I^{\delta\pi} & I^{\delta\psi} \\ I^{\pi\delta} & I^{\pi\pi} & I^{\pi\psi} \\ I^{\psi\delta} & I^{\psi\pi} & I^{\psi\psi} \end{pmatrix} \tag{30}$$

with

$$I^{\delta\delta} = (I_{\delta\delta} - I_{\delta\pi} I_{\pi\pi}^{-1} I_{\pi\delta} - I_{\delta\psi} I_{\psi\psi}^{-1} I_{\psi\delta})^{-1} \tag{31}$$

$$I^{\delta\pi} = -I^{\delta\delta} I_{\delta\pi} I_{\pi\pi}^{-1} \tag{32}$$

$$I^{\delta\psi} = -I^{\delta\delta} I_{\delta\psi} I_{\psi\psi}^{-1} \tag{33}$$

$$I^{\pi\pi} = I_{\pi\pi}^{-1} + I_{\pi\pi}^{-1} I_{\pi\delta} I^{\delta\delta} I_{\delta\pi} I_{\pi\pi}^{-1} \tag{34}$$

$$I^{\pi\psi} = I_{\pi\pi}^{-1} I_{\pi\delta} I^{\delta\delta} I_{\delta\psi} I_{\psi\psi}^{-1} \tag{35}$$

$$I^{\psi\psi} = I_{\psi\psi}^{-1} + I_{\psi\psi}^{-1} I_{\psi\delta} I^{\delta\delta} I_{\delta\psi} I_{\psi\psi}^{-1}. \tag{36}$$

The asymptotic covariance matrix of $\hat{\delta}$ is thus given by (31). In order to evaluate this expression we need some intermediary results:

$$I_{\pi\pi}^{-1} = (\Omega_0 - (1/\sigma_0^2)\,\theta_0\,\theta_0') \otimes \mathcal{M}_{XX}^{-1}$$
$$I_{\delta\pi} I_{\pi\pi}^{-1} = -(1/\sigma_0^2)\,\theta_0' \otimes \mathcal{M}_{ZX} \mathcal{M}_{XX}^{-1}$$
$$I_{\delta\pi} I_{\pi\pi}^{-1} I_{\pi\delta} = \{(1/\eta_0^2) - (1/\sigma_0^2)\}\, \mathcal{M}_{ZX} \mathcal{M}_{XX}^{-1} \mathcal{M}_{XZ}$$

and also, using (9), (8) and (7),

$$I_{\psi\psi}^{-1} = 2\, D_{N+1}^{+} (\Psi_0 \otimes \Psi_0)\, D_{N+1}^{+\prime}$$
$$I_{\delta\psi} I_{\psi\psi}^{-1} = 2\,(e' \otimes E'\,\Psi_0)\, D_{N+1}^{+\prime}$$
$$I_{\delta\psi} I_{\psi\psi}^{-1} I_{\psi\delta} = (e'\,\Psi_0^{-1}\, e)\, E'\,\Psi_0\, E = (1/\eta_0^2)\, E'\,\Psi_0\, E,$$

since $E'e = 0$. Thus

$$\begin{aligned} I^{\delta\delta} &= \{(1/\eta_0^2)\,[\operatorname{plim}(1/T)\, Z'(I_T - X(X'X)^{-1}X')Z - E'\Psi_0 E] \\ &\quad + (1/\sigma_0^2) \operatorname{plim}(1/T)\, Z'X(X'X)^{-1}X'Z\}^{-1} \\ &= \sigma_0^2 (\mathcal{M}_{ZX} \mathcal{M}_{XX}^{-1} \mathcal{M}_{XZ})^{-1}. \end{aligned} \tag{37}$$

This gives the asymptotic covariance matrix of $\hat{\delta}$ and proves Proposition 3.

The Complete Asymptotic Covariance Matrix

Having determined $I^{\delta\delta}$, we now proceed to obtain expressions for the other blocks of the asymptotic covariance matrix I^{-1} given in (30). We define

$$G_0 = \mathcal{M}_{ZX}\mathcal{M}_{XX}^{-1}\mathcal{M}_{XZ} = \text{plim}(1/T)\,Z'X(X'X)^{-1}X'Z,$$

so that $I^{\delta\delta} = \sigma_0^2 G_0^{-1}$, and partition

$$G_0^{-1} = \begin{pmatrix} G_0^{YY} & G_0^{YX_1} \\ G_0^{X_1 Y} & G_0^{X_1 X_1} \end{pmatrix} = (G_0^{ZY} \,\vdots\, G_0^{ZX_1}).$$

Then,

$$\begin{aligned}
I^{\delta\pi} &= \theta_0' \otimes G_0^{-1}\mathcal{M}_{ZX}\mathcal{M}_{XX}^{-1},\\
I^{\delta\psi} &= -2\sigma_0^2(e'\otimes G_0^{-1}E'\Psi_0)D_{N+1}^{+\prime}\\
&= -2\sigma_0^2(G_0^{-1}E'\Psi_0 \,\vdots\, 0)D_{N+1}^{+\prime}\\
&= -2\sigma_0^2(G_0^{ZY}\theta_0 \,\vdots\, G_0^{ZY}\Omega_0 \,\vdots\, 0)D_{N+1}^{+\prime}\\
&= -2\sigma_0^2\left(G_0^{ZY}\theta_0 \,\vdots\, \frac{1}{2}G_0^{ZY}\Omega_0 \,\vdots\, 0\right)
\end{aligned}$$

by Corollary 7,

$$\begin{aligned}
I^{\pi\pi} &= \Omega_0 \otimes \mathcal{M}_{XX}^{-1}\\
&\qquad -(1/\sigma_0^2)\,\theta_0\theta_0' \otimes (\mathcal{M}_{XX}^{-1} - \mathcal{M}_{XX}^{-1}\mathcal{M}_{XZ}G_0^{-1}\mathcal{M}_{ZX}\mathcal{M}_{XX}^{-1}),\\
I^{\pi\psi} &= -I_{\pi\pi}^{-1}I_{\pi\delta}I^{\delta\psi}\\
&= -2(\theta_0\otimes\mathcal{M}_{XX}^{-1}\mathcal{M}_{XZ})\left(G_0^{ZY}\theta_0 \,\vdots\, \frac{1}{2}G_0^{ZY}\Omega_0 \,\vdots\, 0\right)\\
&= -\theta_0\otimes\{2\mathcal{M}_{XX}^{-1}\mathcal{M}_{XZ}G_0^{ZY}\theta_0 \,\vdots\, \mathcal{M}_{XX}^{-1}\mathcal{M}_{XZ}G_0^{ZY}\Omega_0 \,\vdots\, 0\},\\
I^{\psi\psi} &= 2D_{N+1}^{+}(\Psi_0\otimes\Psi_0)D_{N+1}^{+\prime}\\
&\qquad +4\sigma_0^2 D_{N+1}^{+}(e\otimes\Psi_0 E)G_0^{-1}(e'\otimes E'\Psi_0)D_{N+1}^{+\prime}\\
&= 2D_{N+1}^{+}(\Psi_0\otimes\Psi_0)D_{N+1}^{+\prime}\\
&\qquad +4\sigma_0^2 D_{N+1}^{+}(ee'\otimes\Psi_0 EG_0^{-1}E'\Psi_0)D_{N+1}^{+\prime}\\
&= 2\begin{pmatrix} \sigma_0^4 & \sigma_0^2\theta_0' & (\theta_0'\otimes\theta_0')D_N^{+\prime} \\ \sigma_0^2\theta_0 & \frac{1}{2}(\sigma_0^2\Omega_0+\theta_0\theta_0') & (\theta_0'\otimes\Omega_0)D_N^{+\prime} \\ D_N^{+}(\theta_0\otimes\theta_0) & D_N^{+}(\theta_0\otimes\Omega_0) & D_N^{+}(\Omega_0\otimes\Omega_0)D_N^{+\prime} \end{pmatrix}
\end{aligned}$$

$$
+\sigma_0^2 \begin{pmatrix} 4\theta_0' G_0^{YY} \theta_0 & 2\theta_0' G_0^{YY} \Omega_0 & 0 \\ 2\Omega_0 G_0^{YY} \theta_0 & \Omega_0 G_0^{YY} \Omega_0 & 0 \\ 0 & 0 & 0 \end{pmatrix}
$$

using Lemma 5 and the fact that

$$
\Psi_0 E G_0^{-1} E' \Psi_0 = \begin{pmatrix} \theta_0' G_0^{YY} \theta_0 & \theta_0' G_0^{YY} \Omega_0 \\ \Omega_0 G_0^{YY} \theta_0 & \Omega_0 G_0^{YY} \Omega_0 \end{pmatrix}.
$$

Corresponding to the parametrization in terms of $(\delta, \text{vec}\, \Pi, \sigma^2, \theta, v(\Omega))$ we now write I^{-1} as

$$
I^{-1} = \begin{pmatrix} I^{\delta\delta} & I^{\delta\pi} & I^{\delta\sigma^2} & I^{\delta\theta} & 0 \\ I^{\pi\delta} & I^{\pi\pi} & I^{\pi\sigma^2} & I^{\pi\theta} & 0 \\ I^{\sigma^2\delta} & I^{\sigma^2\pi} & I^{\sigma^2\sigma^2} & I^{\sigma^2\theta} & I^{\sigma^2\omega} \\ I^{\theta\delta} & I^{\theta\pi} & I^{\theta\sigma^2} & I^{\theta\theta} & I^{\theta\omega} \\ 0 & 0 & I^{\omega\sigma^2} & I^{\omega\theta} & I^{\omega\omega} \end{pmatrix}
$$

with

$$
\begin{aligned}
I^{\delta\delta} &= \sigma_0^2 G_0^{-1} \\
I^{\delta\pi} &= \theta_0' \otimes G_0^{-1} \mathcal{M}_{ZX} \mathcal{M}_{XX}^{-1} \\
I^{\delta\sigma^2} &= -2\sigma_0^2 G_0^{ZY} \theta_0 \\
I^{\delta\theta} &= -\sigma_0^2 G_0^{ZY} \Omega_0 \\
I^{\pi\pi} &= \Omega_0 \otimes \mathcal{M}_{XX}^{-1} - (1/\sigma_0^2)\, \theta_0 \theta_0' \\
&\quad \otimes (\mathcal{M}_{XX}^{-1} - \mathcal{M}_{XX}^{-1} \mathcal{M}_{XZ} G_0^{-1} \mathcal{M}_{ZX} \mathcal{M}_{XX}^{-1}) \\
I^{\pi\sigma^2} &= -2\theta_0 \otimes \mathcal{M}_{XX}^{-1} \mathcal{M}_{XZ} G_0^{ZY} \theta_0 \\
I^{\pi\theta} &= -\theta_0 \otimes \mathcal{M}_{XX}^{-1} \mathcal{M}_{XZ} G_0^{ZY} \Omega_0 \\
I^{\sigma^2\sigma^2} &= 2\sigma_0^4 + 4\sigma_0^2 \theta_0' G_0^{YY} \theta_0 \\
I^{\sigma^2\theta} &= 2\sigma_0^2 (\theta_0' + \theta_0' G_0^{YY} \Omega_0) \\
I^{\sigma^2\omega} &= 2(\theta_0' \otimes \theta_0') D_N^{+\prime} \\
I^{\theta\theta} &= \sigma_0^2 (\Omega_0 + \Omega_0 G_0^{YY} \Omega_0) + \theta_0 \theta_0' \\
I^{\theta\omega} &= 2(\theta_0' \otimes \Omega_0) D_N^{+\prime} \\
I^{\omega\omega} &= 2 D_N^{+} (\Omega_0 \otimes \Omega_0) D_N^{+\prime}.
\end{aligned}
$$

The remaining blocks of I^{-1} follow from its symmetry.

Remarks

1. Without difficulty we can show that

$$
\mathcal{M}_{XX}^{-1} \mathcal{M}_{XZ} = \begin{pmatrix} \Pi_{10} & I_{K_1} \\ \Pi_{20} & 0 \end{pmatrix}.
$$

2. Since the matrix $\mathcal{M}_{XX}^{-1} - \mathcal{M}_{XX}^{-1} \mathcal{M}_{XZ} G_0^{-1} \mathcal{M}_{ZX} \mathcal{M}_{XX}^{-1}$ is positive semidefinite, we have

$$\{\Omega_0 - (1/\sigma_0^2)\,\theta_0\,\theta_0'\} \otimes \mathcal{M}_{XX}^{-1} \leqslant var(\text{vec}\,\hat{\Pi}) \leqslant \Omega_0 \otimes \mathcal{M}_{XX}^{-1}$$

Where $B \leqslant A$ means that $A - B$ is positive semidefinite.

3. Clearly, $\hat{\Omega}$ is asymptotically uncorrelated with $\hat{\delta}$ and $\hat{\Pi}$.

4. The following three expressions for G_0^{YY} are equivalent:

$$G_0^{YY} = [\text{plim}\,(1/T)\,Y'X\,(X'X)^{-1}X'Y - \text{plim}\,(1/T)\,Y'X_1\,(X_1'X_1)^{-1}X_1'Y]^{-1}$$

$$G_0^{YY} = [\Pi_0'\,\{\lim\,(1/T)\,(X'X - (X'X_1\,(X_1'X_1)^{-1}X_1'X)\}\,\Pi_0]^{-1}$$

$$G_0^{YY} = [\Pi_{20}'\,\{\lim\,(1/T)\,X_2'(I - X_1\,(X_1'X_1)^{-1}X_1')\,X_2\}\,\Pi_{20}]^{-1}.$$

5. From the expression of I^{-1} we can, of course, derive the asymptotic variance or covariance matrices of subvectors or submatrices, e. g., the asymptotic covariance matrix between $\hat{\alpha}$ and $\hat{\beta}$, or the asymptotic variance matrix of vec $\hat{\Pi}_1$. An interesting case was considered by SMITH [1985] who partitions $\theta = (\theta_1', \theta_2')'$ and examines the Wald test for the hypothesis $H_0 : \theta_{20} = 0$. For this he needs the asymptotic variance matrix of $\hat{\theta}_2$ which we obtain as the appropriate submatrix of

$$as.\,var\,\hat{\theta} = \sigma_0^2\,(\Omega_0 + \Omega_0\,G_0^{YY}\,\Omega_0) + \theta_0\,\theta_0'.$$

SMITH [1985, equation (5)] gives the asymptotic variance of $\hat{\theta}_2$ *only under* H_0 (where $\theta_{20} = 0$). This does not, of course, effect the validity of his Wald test, but it does explain the difference between his result and ours.

• References

GODFREY, L. G. and WICKENS, M. R. (1982). – "A Simple Derivation of the Limited Information Maximum Likelihood Estimator", *Economics Letters*, 10, pp. 277-283.

MAGNUS, J. R. (1988). – *Linear Structures*, Charles Griffin, London and Oxford University Press, New York.

MAGNUS, J. R. and NEUDECKER, H. (1980). – "The Elimination Matrix: Some Lemmas and Applications", *SIAM Journal on Algebraic and Discrete Methods*, 1, pp. 422-449.

MAGNUS, J. R. and NEUDECKER, H. (1986). – "Symmetry, 0-1 Matrices and Jacobians: a Review", *Econometric Theory*, 2, pp. 157-190.

MONFORT, A. (1978). – "First-Order Identification in Linear Models", *Journal of Econometrics*, 7, pp. 333-350.

SMITH, R. J. (1983). – "On the Classical Nature of the Wu-Hausman Statistics for the Independence of Stochastic Regressors and Disturbance", *Economics Letters*, 11, pp. 357-364.

SMITH, R. J. (1985). – "Wald Tests for the Independence of Stochastic Variables and Disturbance of a Single Linear Stochastic Simultaneous Equation", *Economics Letters*, 17, pp. 87-90.