Cowles Foundation Paper 806

Journal of Econometrics 51 (1992) 113–150. North-Holland

Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations

In Choi and Peter C.B. Phillips*

Yale University, New Haven, CT 06520, USA

Received January 1990, final version received October 1990

General formulae for the finite sample and asymptotic distributions of the instrumental variable estimators and the Wald statistics in a simultaneous equation model are derived. It is assumed that the coefficient vectors of both endogenous and exogenous variables are only partially identified, even though the order condition for identification is satisfied. This work extends previous results in Phillips (1989) where the coefficient vector of the exogenous variables is partially identified and that of the endogenous variables is totally unidentified. The effect of partial identification on the finite sample and asymptotic distributions of the estimators and the Wald statistics is analyzed by isolating identifiable parts of the coefficient vectors using a rotation of the coordinate system developed in Phillips (1989). The pdf's of the estimators and the Wald statistics are illustrated using simulation and compared with their respective asymptotic distributions.

1. Introduction

Identification of a structural equation in a simultaneous equation system is an important preliminary condition prior to estimation and statistical inference. Standard statistical procedures are almost always based on the assumption that the coefficients of a structural equation are uniquely defined by a priori restrictions on the coefficients of a simultaneous equations model. These restrictions usually arise from economic theory. Conditions for identification have been discussed by various authors [e.g., Fisher (1966), Hsiao (1983), Hausman (1983)]. If an equation is identified, we are usually able to

*Our thanks go to Glena Ames for her skill and effort in keyboarding the manuscript for this paper, to the NSF for partial research support under Grant Number SES 8821180 to the second author, and A. Zellner, the Associate Editor, and the two anonymous referees for their helpful comments. I. Choi is presently at the Ohio State University. He wishes to express his thanks to the second author for his continuing guidance while the research reported in this paper was going on.

0304-4076/92/$03.50

estimate the coefficients consistently and to mount conventional statistical tests relying on asymptotic normal or chi-square distributions.

However, the statistical properties of estimators and tests in cases of identification failure have not received much attention by researchers. We use the term identification failure here to imply that the rank condition for identification in single equation estimation is not fulfilled, even though the order condition is satisfied. Hence studies in identification failure presume that the usual condition that empirical researchers check for in single-equation estimation is met. It is of importance to investigate the statistical properties of estimators and test statistics in case of identification failure, since there are good reasons for suspecting such failures in empirical research [see Phillips (1989) and Sims (1980) for examples]. When such failures occur, the conventional statistical theory does not apply. A finite sample distribution theory of the instrumental variable estimator and the limited information maximum likelihood estimator in case of identification failure was developed in some earlier work by Phillips (1980, 1983, 1984a, 1984b, 1985). In this work it was shown that the exact finite sample densities of the estimators do not carry any information on the coefficients in a structural equation when the true coefficient vector is not identifiable. Moreover, the densities are invariant to changes in the sample size, demonstrating the fact that the uncertainty about the coefficients that is due to lack of identification persists in the limit as the sample size tends to infinity. Recently, a general finite sample and asymptotic distribution theory for the instrumental variable estimator and for Wald test statistics was developed in Phillips (1989). There it is assumed that the whole coefficient vector of the endogenous variables is not identifiable and a general rank condition is given such that the coefficient vector of the exogenous variables is partially identified. By rotating the coordinate system, identified and unidentified parts of the coefficient vector are distinguished. The finite sample distribution theory developed therein shows that only the estimator of identified coefficient vector has a finite sample density that carries any useful information on the true coefficient vector. The asymptotic distribution theory shows that only the estimator of the identified coefficient vector is consistent and has a meaningful limit distribution. These findings remind us how important the necessary and sufficient condition for identification is in terms of the statistical properties of estimators and tests in simultaneous equation models.

This paper is built upon the earlier work in Phillips (1989) and employs a similar approach. The finite sample and asymptotic distribution theory in Phillips (1989) is given for the case where the coefficient vector of the endogenous variables is totally unidentified and that of exogenous variables is partially identified. In the present paper it is assumed that the coefficient vectors of both endogenous and exogenous variables may be partially identified. Hence the current framework is more general than that of Phillips

(1989) and includes total identification, total lack of identification and partial identification of the coefficient vectors of both endogenous and exogenous variables. General formulae for the finite sample and asymptotic densities of the instrumental variable estimators will be given by rotating the coordinate system as in Phillips (1989). This rotation shall show the effect of partial identification on the distributions of the estimators of the whole coefficient vectors in a convenient way. The general formulae provide an economical way of writing down the exact and asymptotic densities of the instrumental variable estimators in various cases of identification and lack of identification and shed light on the effect of identification and lack of identification on the finite sample and asymptotic distribution theory of the estimators. The asymptotic distributions are derived for general martingale difference errors and, as in Phillips (1989), the limit distribution theory is all of the mixture normal class. Limit distributions of Wald test statistics are also derived. These are, in general, not chi-squared and again demonstrate the effect of nonidentifiability.

The plan of this paper is as follows. Section 2 discusses the structure of the current problem and the rotation of the coordinate system that isolates the identifiable components of the coefficient vectors. In section 3, the finite sample and asymptotic distributions of the instrumental variable estimators are derived. Section 4 deals with statistical inference on the whole coefficient vectors. Standard Wald statistics are formulated and are shown to converge in distribution to random variables which are not distributed as chi-square. These results indicate the importance of identification for statistical inference in a simultaneous equations model. Section 5 reports some numerical computations in partially identified models and contains figures of the pdf's of the estimators and the Wald statistics based on simulations. These figures illustrate some of the main properties of econometric estimators and tests in partially identified models. Conclusions are drawn in section 6. All proofs are in the appendix.

A word on notation. We use the symbol '$\Rightarrow$' to signify weak convergence, the symbol '$\equiv$' to signify equality in distribution, and the inequality '$>$' to signify positive definite when applied to matrices. We use $O(n)$ to denote the orthogonal group of $n \times n$ matrices, $V_{k,n}$ to denote the Stiefel manifold $\{H_1$ $(n \times k)$: $H_1'H_1 = I_k\}$. Finally, we use $r(\Pi)$ to signify the rank of the matrix Π and P_Π to signify the orthogonal projection onto the range space of Π with $Q_\Pi = I - P_\Pi$. All the limits are taken as $T \to \infty$, unless specified otherwise.

2. A partially identified structural equation and its estimable functions

We are concerned with a structural equation

$$y_1 = Y_2\beta + Z_1\gamma + u = W\delta + u, \tag{1}$$

where y_1 $(T \times 1)$ and Y_2 $(T \times n)$ denote $n+1$ endogenous variables, Z_1 $(T \times k_1)$ is a matrix of k_1 exogenous variables included in the eq. (1), and u is a random disturbance vector. The reduced form of (1) is written in partitioned format

$$[y_1, Y_2] = [Z_1, Z_2]\begin{bmatrix} \pi_1 & \Pi_1 \\ \pi_2 & \Pi_2 \end{bmatrix} + [v_1, V_2] \tag{2}$$

or

$$Y = Z\Pi + V,$$

where Z_2 $(T \times k_2)$ is a matrix of exogenous variables excluded from (1). It is assumed that $k_2 \geq n$ so that the necessary condition for the identification of (1) is satisfied, and that Z is of full column rank $k = k_1 + k_2$. Eq. (2) is assumed to be in canonical form [see Phillips (1983) for details of the necessary transformations], so that the rows of V are iid$(0, I_m)$, $m = n+1$. We shall require the following distributional assumption for the development of the finite sample theory:

$$V \equiv N_{T,m}(0, I). \tag{C1}$$

In addition, we make the following assumption on the sample second moment matrix of Z:

$$T^{-1}Z'Z = M. \tag{C2}$$

We assume $M = I$. This simplifies the expressions in later sections, but incurs no loss of generality. It will also be convenient in some cases to strengthen (C2) to the following:

$$T^{-1}Z'Z = M + O(T^{-1}) = I + O(T^{-1}). \tag{C2'}$$

We partition M conformally with $Z = [Z_1, Z_2] = [Z_1, Z_3, Z_*]$ as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} = \begin{bmatrix} M_{11} & M_{13} & M_{1*} \\ M_{31} & M_{33} & M_{3*} \\ M_{*1} & M_{*3} & M_{**} \end{bmatrix} \begin{matrix} k_1 \\ k_3 \\ k_* \end{matrix}.$$

The second partition corresponds to the selection of a submatrix of instruments Z_3. The submatrix Z_3 is chosen using the matrix $W' = [I|0]$, i.e., $Z_3 = Z_2W$. Since $M = I$, we have $M_{11} = I$, $M_{22} = I$, $M_{12} = 0$, and $M_{21} = 0$. We also have corresponding results for the second partition.

The identifying relations connecting the parameters of (1) and (2) are

$$\pi_1 - \Pi_1\beta = \gamma, \tag{3}$$

$$\pi_2 - \Pi_2\beta = 0. \tag{4}$$

Eq. (1) is identified iff Π_2 has a full column rank, i.e., $r(\Pi_2) = n \leq k_2$. We call this the fully identified case, following Phillips (1989). If

$$\Pi_2 = 0 \tag{5}$$

and $r(\Pi_2) = 0$, we have what is usually referred to as the leading case in econometric distribution theory [Phillips (1983)]. In this case the parameter vector β is totally unidentified. However, if $\Pi_1 = 0$, for example, the entire coefficient vector γ is identified and is equal to the reduced form subcoefficient vector π_1.

In this paper, we consider the general case where Π_1 and Π_2 are of arbitrary rank. The leading case where $\Pi_2 = 0$ and Π_1 is of arbitrary rank is discussed in earlier work by Phillips (1989). In the leading case the vector β is totally unidentified and only a certain part of γ (i.e., some linear combinations of γ) is identified. In the general case certain components of both β and γ are identified while other components of both vectors are unidentified. The number of components in each category is determined by the ranks of Π_1 and Π_2, which are assumed to be

$$r(\Pi_1) = k_{12} \leq k_1, \tag{C3}$$

$$r(\Pi_2) = n_1 \leq n. \tag{C4}$$

Following the development in Phillips (1989, §2.1) we now rotate coordinates in both the space of the endogenous variables Y_2 and the exogenous variables Z_1 to isolate estimable functions. Define

$$S = [\overset{n_1}{S_1}, \overset{n_2}{S_2}] \in O(n),$$

where S_2 spans the null space of Π_2 and $\Pi_{21} = \Pi_2 S_1$ has a full column rank n_1. Let

$$\beta_1 = S_1'\beta, \quad \beta_2 = S_2'\beta$$

and

$$\Pi_{11} = \Pi_1 S_1, \quad \Pi_{12} = \Pi_1 S_2, \quad \Pi_{21} = \Pi_2 S_1, \quad \Pi_{22} = \Pi_2 S_2 = 0.$$

Then the identifying relations in the new coordinates are

$$\pi_1 - \Pi_{11}\beta_1 - \Pi_{12}\beta_2 = \gamma, \tag{6}$$

$$\pi_2 - \Pi_{21}\beta_1 = 0. \tag{7}$$

In this system we see that β_1 is identifiable and β_2 is totally unidentified. Moreover the structural coefficient vector γ is also unidentified due to the effect of the unidentified coefficient β_2 appearing on the left-hand side of (6).

We now rotate coordinates in eq. (6) to isolate the identifiable part of γ. Again as in Phillips (1989, §2.1) we define an orthogonal matrix:

$$R = [\overset{k_{11}}{R_1} \quad \overset{k_{12}}{R_2}] \in O(k_1),$$

where R_1 is selected to span the null space of Π_1' and $k_1 = k_{11} + k_{12}$. Under R' the equation system (6) becomes

$$R_1'\pi_1 = \gamma_1, \tag{8}$$

$$R_2'\pi_1 - R_2'\Pi_{11}\beta_1 - R_2'\Pi_{12}\beta_2 = \gamma_2, \tag{9}$$

where

$$\gamma_1 = R_1'\gamma \quad \text{and} \quad \gamma_2 = R_2'\gamma.$$

Here γ_1 is identified, while γ_2 is not.

These rotations produce a new structural equation

$$\begin{aligned} y_1 &= Y_2\beta + Z_1\gamma + u \\ &= Y_2SS'\beta + Z_1RR'\gamma + u \\ &= Y_{21}\beta_1 + Y_{22}\beta_2 + Z_{11}\gamma_1 + Z_{12}\gamma_2 + u. \end{aligned} \tag{10}$$

In (10) [which corresponds with eq. (13) in Phillips (1989)] the coefficients (β_1, γ_1) are identified and (β_2, γ_2) are totally unidentified. The original coefficients are recovered from the equations:

$$\beta = S_1\beta_1 + S_2\beta_2, \qquad \gamma = R_1\gamma_1 + R_2\gamma_2.$$

Using these equations, we can find the effect of partial identification on the finite sample and asymptotic distributions of the IV estimators of the entire vectors β and γ.

The reduced form system (2) can be similarly rewritten in the new coordinate system (following rotation by S) as

$$y_1 = Z_1\pi_1 + Z_2\pi_2 + v_1, \tag{11}$$

$$Y_{21} = Z_1\Pi_{11} + Z_2\Pi_{21} + V_{21}, \tag{12}$$

$$Y_{22} = Z_1\Pi_{12} + V_{22}, \tag{13}$$

where

$$V_{21} = V_2S_1 \quad \text{and} \quad V_{22} = V_2S_2.$$

3. Distribution theory under normality

3.1. Coefficients of the endogenous variables

We shall study the finite sample and asymptotic distributions of instrumental variable (IV) estimators of the structural eq. (1) under the normality assumption (C1). We assume (C3) and (C4), so that Π_1 an Π_2 have arbitrary ranks. [The distributional theory in the case where $\Pi_2 = 0$ and Π_1 is of arbitrary rank was developed in Phillips (1989).] Under (C3) and (C4) and in the transformed coordinate system leading to (10), (β_1, γ_1) is identified while (β_2, γ_2) is totally unidentified. The functional forms of the finite sample distributions of the IV structural coefficient estimators will be derived and, as a corollary under the additional requirement (C2), their asymptotic distributions will be obtained.

The IV estimator of δ in (1) is simply $\hat{\delta} = \text{argmin}_\delta (y - W\delta)P_H(y - W\delta)$, where $H = [Z_1, Z_3]$ is a $T \times (k_1 + k_3)$ matrix of instruments and Z_3 is a submatrix of Z_2 formed by column selection. We require $k_3 \geq n$, so that the order condition of sufficient instruments is satisfied. If $H = [Z_1, Z_2]$, the instrumental variable estimators are equivalent to 2SLS estimators.

Formulae for the subcoefficient vector estimates in the transformed system (10) are easily obtained by stepwise regression as follows:

$$\hat{\beta}_1 = (Y_{21}'EY_{21})^{-1}(Y_{21}'Ey_1),$$

$$\hat{\beta}_2 = (Y_{22}'JY_{22})^{-1}(Y_{22}'Jy_1),$$

$$\hat{\gamma} = R_1'\hat{\gamma} = R_1'(Z_1'Z_1)^{-1}Z_1'y_1 - R_1'(Z_1'Z_1)^{-1}Z_1'[Y_{21}, Y_{22}]\begin{bmatrix}\hat{\beta}_1 \\ \hat{\beta}_2\end{bmatrix},$$

$$\hat{\gamma}_2 = R_2'\hat{\gamma} = R_2'(Z_1'Z_1)^{-1}Z_1'y_1 - R_2'(Z_1'Z_1)^{-1}Z_1'[Y_{21}, Y_{22}]\begin{bmatrix}\hat{\beta}_1 \\ \hat{\beta}_2\end{bmatrix},$$

where

$$E = L - LY_{22}(Y_{22}'LY_{22})^{-1}Y_{22}'L,$$

$$J = L - LY_{21}(Y_{21}'LY_{21})^{-1}Y_{21}'L,$$

$$L = P_H - P_{Z_1}.$$

Of course, E and J are idempotent matrices with ranks $r(E) = \operatorname{tr} L - n_2 = k_3 - n_2$ and $r(J) = \operatorname{tr} L - n_1 = k_3 - n_1$, respectively.

The following theorem gives general expressions for the exact densities of $\hat{\beta}_1$ and $\hat{\beta}_2$.

Theorem 3.1. Under the assumptions (C1), (C3), and (C4),

(a)
$$\hat{\beta}_1 \equiv \int_{V_{k_3-n_2,k_3}} \int_{g \in R^{n_1}} \int_{A>0} N(A(\Theta\Theta')g(\Theta\Theta')\beta_1,\ A(\Theta\Theta'))$$
$$\times \operatorname{pdf}_{\Theta}(A, g)\,\mathrm{d}A\,\mathrm{d}g\,\mathrm{d}\Theta$$
$$= r_1, \quad say,$$

where

$$A|_{\Theta} = A(\Theta\Theta') \equiv W_{n_1}^{-1}(k_3 - n_2 + n_1 + 1, I_{n_1}, T\Pi_{21}'W\Theta\Theta'W'\Pi_{21}),$$

$$g|_{\Theta} = g(\Theta\Theta') \equiv N(T\Pi_{21}'W\Theta\Theta'W'\Pi_{21}', T\Pi_{21}'W\Theta\Theta'W'\Pi_{21}),$$

and

$$W' = \begin{bmatrix} \overset{k_3}{I} & | & \overset{k-k_3}{0} \end{bmatrix} k_3.$$

In the above formula, Θ is a matrix that is distributed uniformly on the Stiefel manifold $V_{k_3-n_2,k_3} = \{\Theta\colon \Theta'\Theta = I_{k_3-n_2}\}$.

(b)
$$\hat{\beta}_2 \equiv \int_{\theta \in R^{k_3 n_1}} \int_{m \in R^{n_1 n_2}} \int_{B>0} N(Bm(\theta)\beta_1, B)$$
$$\times \operatorname{pdf}_{\theta}(B, m)\,\mathrm{d}B\,\mathrm{d}m\,\operatorname{pdf}(\theta)\,\mathrm{d}\theta,$$
$$= r_2, \quad say,$$

where

$$B \equiv W_{n_2}^{-1}(k_3 - n_1 + n_2 + 1, I_{n_2}),$$

$$\theta \equiv N(T^{1/2}W'\Pi_{21}, I),$$

and

$$m(\theta)|_\theta \equiv N(0, T\Pi_{21}'WQ_\theta W'\Pi_{21}).$$

Remarks

(i) We find that both distributions (a) and (b) are mean and covariance matrix mixture normal unless $\beta_1 = 0$. This is an economical way of writing down exact distributions whose series representations are very complicated [these may be deduced from results in Phillips (1980, 1984b)]. Moreover the functional forms given here more easily shed light on the effect of identification and lack of identification on the finite sample distribution theory.

(ii) The density of $\hat{\beta}_2$ is independent of β_2 and carries no information on β_2. This is as we would expect since β_2 is not identified. Interestingly, we find that the density of $\hat{\beta}_2$ is dependent on the identified coefficient β_1. The density of $\hat{\beta}_1$ is also dependent on β_1.

(iii) If Π_2 has full column rank, the whole parameter vector β is identified and the pdf of $\hat{\beta}$ may be expressed as

$$\hat{\beta} \equiv \int_{g \in R^{n^2}} \int_{A>0} N(Ag\beta, A)\mathrm{pdf}(A, g)\, \mathrm{d}A\, \mathrm{d}g = r, \quad \text{say},$$

where

$$A \equiv W_n^{-1}(k_3 - n + 1, I_n, T\Pi_2'WW'\Pi_2),$$

$$g \equiv N(T\Pi_2'WW'\Pi_2, T\Pi_2'WW'\Pi_2).$$

This expression is easily recovered from part (a).

(iv) If $r(\Pi_2) = 0$, then $\Pi_{21} = 0$, and we find from part (b)

$$\hat{\beta} \equiv \int_{B>0} N(0, B)\mathrm{pdf}(B)\, \mathrm{d}B,$$

where

$$B \equiv W_n^{-1}(k_3 - n_1 + n_2 + 1, I_n).$$

This result is consistent with the earlier result in Phillips (1989, theorem 2.1(a)) for the case of totally unidentified β.

(v) The exact densities of $\hat{\beta}_1$ and $\hat{\beta}_2$ undergo some simplification when $\beta_1 = 0$. By setting $\beta_1 = 0$ in (a) and (b), we obtain

$$\hat{\beta}_1 \equiv \int_{V_{k_1 - n_2, k_2}} \int_{A>0} N(0, A(\Theta\Theta'))\mathrm{pdf}(A(\Theta\Theta'))\, \mathrm{d}A\, \mathrm{d}\Theta,$$

$$\hat{\beta}_2 \equiv \int_{B>0} N(0, B)\mathrm{pdf}(B)\, \mathrm{d}B.$$

The latter result was given earlier in Phillips (1988).

(vi) The density of the (totally unidentified) estimator $\hat{\beta}_2$ depends on the matrix $T^{1/2}W'$, as it enters the mean of θ. However, θ occurs in the distribution of $\hat{\beta}_2$ only in terms of the projection operator Q_θ. It is clear, therefore, that the distribution of $\hat{\beta}_2$ depends explicitly only on W', which is finite. Nevertheless, the dependence on W' and the noncentrality of the distribution of θ make the problem rather different from that studied in Phillips (1988, 1989).

Under assumption (C2), we may develop an asymptotic distribution theory of $\hat{\beta}_1$ and $\hat{\beta}_2$ as follows:

Corollary 3.1. Under (C1), (C2′), (C3), and (C4),

(a) $$\sqrt{T}(\hat{\beta}_1-\beta_1)$$

$$\Rightarrow \int_{V_{k_3-n_2,k_3}} N\left(0, \left(\Pi_{21}'M_{32\cdot1}'M_{33\cdot1}^{-1/2}\Theta\Theta'M_{33\cdot1}^{-1/2}M_{32\cdot1}\Pi_{21}\right)^{-1}\right)\mathrm{d}\Theta$$

$$= \int_{V_{k_3-n_2,k_3}} N\left(0, (\Pi_{21}'\Theta\Theta'\Pi_{21})^{-1}\right)\mathrm{d}\Theta = \bar{r}_1, \quad say,$$

(b) $$\hat{\beta}_2 \Rightarrow \int_{B>0} N(0,B)\,\mathrm{pdf}(B)\,\mathrm{d}B$$

$$= \bar{r}_2, \quad say,$$

where

$$M_{33.1} = M_{33} - M_{31}M_{11}^{-1}M_{13} = I,$$

$$M_{32.1} = M_{32} - M_{31}M_{11}^{-1}M_{12} = I,$$

and

$$B \equiv W_{n_2}^{-1}(k_3 - n_1 + n_2 + 1, I_{n_2}).$$

Remarks

(i) The estimator $\hat{\beta}_1$ is consistent to β_1, as we would expect for the identified subcoefficient β_1. On the other hand, $\hat{\beta}_2$ converges in law to a nondegenerate distribution so that the uncertainty about β_2 that results from the lack of identification persists in the limit. This is analogous to the cases explored earlier in Phillips (1988, 1989).

(ii) Note that the limiting distribution of $\hat{\beta}_1$ is a covariance matrix mixture normal. Hence the conventional asymptotic theory for identified coefficient estimators does not apply here. This is because the lack of identifiability of β_2 affects the limiting distribution of $\hat{\beta}_1$ by producing an additional variability that is manifested in the covariance matrix mixing variate.

(iii) If the entire parameter vector β is identified, we have directly from (a)

$$\sqrt{T}(\hat{\beta}-\beta) \Rightarrow N\left(0, (\Pi_2' M_{32\cdot 1}' M_{33\cdot 1}^{-1} M_{32\cdot 1}\Pi_2)^{-1}\right) = N\left(0, (\Pi_2'\Pi_2)^{-1}\right),$$

corresponding to traditional asymptotic theory.

(iv) If $r(\Pi_2)=0$, then

$$\hat{\beta} \Rightarrow \int_{B>0} N(0,B)\,\mathrm{pdf}(B)\,\mathrm{d}B,$$

so that the usual leading case result applies in the limit.

(v) Note that the limiting distribution of $\hat{\beta}_2$ is different from its finite sample distribution. Again, this is also different from the leading case where the finite sample distribution is invariant to T.

(vi) Since $\hat{\beta} = S_1\hat{\beta}_1 + S_2\hat{\beta}_2$, we find that $\hat{\beta} \Rightarrow S_2\bar{r}_2$. Thus, the effects of the lack of identifiability of β_2 are manifested in the original coordinates in the nondegeneracy of $\hat{\beta}$.

(vii) Corollary 3.1 continues to hold if the rows of V form a sequence of stationary, ergodic martingale differences with covariance matrix I, as is discussed in Phillips (1989). Thus the asymptotic results hold for a much wider class of errors.

3.2. Coefficients of the exogenous variables

The estimates of the coefficients of exogenous variables have the following finite sample distributions. These depend on the joint pdf of $\hat{\beta}_1$ and $\hat{\beta}_2$.

Theorem 3.2. Under (C1), (C2), and (C3),

(a) $$\hat{\gamma}_1 \equiv \int_{R^{n_1+n_2}} N(\gamma_1, T^{-1}(1+r^{*\prime}r^*))\,\mathrm{pdf}(r^*)\,\mathrm{d}r^* = s_1, \quad \textit{say},$$

where $r^* = [r_1, r_2]'$.

(b) $$\hat{\gamma}_2 \equiv \int_{R^{n_1+n_2}} N(R_2'\pi_1 - R_2'\Pi_1 S r^*, T^{-1}(1+r^{*\prime}r^*))\,\mathrm{pdf}(r^*)\,\mathrm{d}r^* = s_2, \quad \textit{say}.$$

Remarks

(i) Both $\hat{\gamma}_1$ and $\hat{\gamma}_2$ have mixture normal distributions since they are clearly normal conditional on the estimates of the endogenous variable coefficients.

(ii) When $\Pi_2 = 0$, $\hat{\gamma}_1$ and $\hat{\gamma}_2$ have distributions identical to those derived in Phillips (1989).

(iii) When $\Pi_1 = 0$, the whole parameter vector γ is identified. Its pdf is easily recovered from part (a) of Theorem 3.2 as follows:

$$\hat{\gamma} \equiv \int_{R^{n_1+n_2}} N(\gamma, T^{-1}(1 + r^{*\prime}r^*))\mathrm{pdf}(r^*)\,\mathrm{d}r^*.$$

If, in addition, $\Pi_2 = 0$, we find from the above that

$$\hat{\gamma} \equiv \int_{R^n} N(\gamma, T^{-1}(1 + \bar{r}'\bar{r}))\mathrm{pdf}(\bar{r})\,\mathrm{d}\bar{r},$$

where

$$\bar{r} \equiv \int_{B>0} N(0, B)\mathrm{pdf}(B)\,\mathrm{d}B, \qquad B \equiv W_{n_2}^{-1}(k_3 - n_1 + n_2 + 1, I_n).$$

The exact formula of the pdf of $\hat{\gamma}$ in this case was derived by Phillips (1984) by a rather different method.

(iv) If Π_1 has full row rank, γ is totally unidentified. In this case, the pdf of $\hat{\gamma}$ is expressed as

$$\hat{\gamma} \equiv \int_{R^{n_1+n_2}} N(\pi_1 - \Pi_{11}r_1 - \Pi_{12}r_2, T^{-1}(1 + r^{*\prime}r^*))\mathrm{pdf}(r^*)\,\mathrm{d}r^*.$$

(v) Note that the density of $\hat{\gamma}_1$ contains γ_1 in its pdf as an argument. From eq. (9) we have $R_2'\pi_1 = \gamma_2 + R_2'\Pi_1 S\beta$, which yields $R_2'\pi_1 - R_2'\Pi_1 S r^* = \gamma_2 - R_2'\Pi_1 S(r^* - \beta)$. Thus we obtain

$$\hat{\gamma}_2 \equiv \int_{R^{n_1+n_2}} N(\gamma_2 - R_2'\Pi_1 S(r^* - \beta), T^{-1}(1 + r^{*\prime}r^*))\mathrm{pdf}(r^*)\,\mathrm{d}r^*.$$

This expression shows that the distribution of $\hat{\gamma}_2$ depends on γ_2. This is in contrast to the case of $\hat{\beta}_2$ where there is no comparable parameterization for $\hat{\beta}_2$ which will allow β_2 to enter the pdf of $\hat{\beta}_2$. This expression reduces to the one in (iii) when $\Pi_1 = 0$.

(vi) The densities of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ both depend on the sample size T. Thus we can expect that both will have asymptotic distributions that are different from the finite sample distributions given here.

The asymptotic distributions of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are given as follows:

Corollary 3.2. Under (C1), (C2'), (C3), and (C4),

(a) $$\sqrt{T}(\hat{\gamma}_1 - \gamma_1) \Rightarrow \int_{R^{n_2}} N(0, T^{-1}(1 + \bar{\beta}_1'\bar{\beta}_1 + \bar{r}_2'\bar{r}_2))\mathrm{pdf}(\bar{r}_2)\,\mathrm{d}\bar{r}_2$$
$$= \bar{s}_1, \quad say.$$

(b) $$\hat{\gamma}_2 \Rightarrow R_2'\pi_1 - R_2'\Pi_{11}\bar{\beta}_1 - R_2'\Pi_{12}\bar{r}_2 = \bar{s}_2, \quad say.$$

Remarks

(i) The estimator $\hat{\gamma}_1$ is consistent to γ_1 as expected for the identified subcoefficient vector γ_1. However, $\hat{\gamma}_2$ converges in law to a nondegenerate distribution due to the lack of identification. This is analogous to our former results on the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$.

(ii) As is the case with the estimator $\hat{\beta}_1$, the limiting distribution of $\hat{\gamma}_1$ is a covariance matrix mixture normal. Here again, we find that the conventional asymptotic theory for identified coefficient estimators does not apply.

(iii) When $\Pi_2 = 0$, the earlier results obtained in Phillips (1989) can be readily recovered from our general formulae for the asymptotic distributions.

(iv) When $\Pi_1 = 0$, the whole parameter vector γ is identified irrespective of the rank of the reduced form coefficient matrix Π_2. The limiting distribution of $\hat{\gamma}$ might be expressed as

$$\sqrt{T}(\hat{\gamma} - \gamma) \Rightarrow \int_{R^{n_2}} N(0, T^{-1}(1 + \bar{\beta}_1'\bar{\beta}_1 + \bar{r}_2'\bar{r}_2))\mathrm{pdf}(\bar{r}_2)\,\mathrm{d}\bar{r}_2.$$

Moreover, if $\Pi_2 = 0$, we find that

$$\sqrt{T}(\hat{\gamma} - \gamma) \Rightarrow \int_{R^n} N(0, T^{-1}(1 + \bar{r}'\bar{r}))\mathrm{pdf}(\bar{r}),$$

where

$$\bar{r} \equiv \int_{B>0} N(0, B)\mathrm{pdf}(B)\,\mathrm{d}B, \qquad B \equiv W_n^{-1}(k_3 - n_1 + n_2 + 1, I_n).$$

Clearly, $\hat{\gamma}$ is consistent to γ, manifesting its identifiability. However, its limiting distribution is a covariance matrix mixture normal due to the lack of identifiability of the endogenous parameter vector β.

(v) If Π_1 has a full row rank, the whole parameter vector γ is totally unidentified. As can be readily found from part (b), $\hat{\gamma}$ converges in law to a nondegenerate distribution as shown below:

$$\hat{\gamma} \Rightarrow \pi_1 - \Pi_{11}\beta_1 - \Pi_{12}\bar{r}_2.$$

(vi) The limiting distribution of $\hat{\gamma}_2$ is different from its finite sample distribution. This is consistent with the leading case discussed in Phillips (1989).

(vii) Since $\hat{\gamma} = R_1\hat{\gamma}_1 + R_2\hat{\gamma}_2$, we find that $\hat{\gamma} \Rightarrow R_2\bar{s}_2$. The nondegeneracy of $\hat{\gamma}$ is due to the lack of identifiability of the coefficient vector β.

(viii) As is the case with Corollary 3.1, Corollary 3.2 holds if the rows of V form a sequence of stationary, ergodic martingale differences with covariance matrix I.

4. Statistical tests on the coefficients

In this section, we shall consider the problem of testing hypotheses on the coefficients of the endogenous and exogenous regressors. We shall formulate Wald statistics for the hypotheses

$$\mathrm{H}_\beta: \quad H_1\beta = h_1,$$

where H_1 is $p_1 \times n$ of rank p_1 $(\leq n)$ and

$$\mathrm{H}_\gamma: \quad H_2\gamma = h_2,$$

where H_2 is $p_2 \times n$ of rank p_2 $(\leq n)$. The error variance estimator for the Wald statistics is defined by

$$\begin{aligned}\hat{\sigma}^2 &= T^{-1}\left(y_1 - W_1\hat{\delta}\right)'\left(y_1 - W_1\hat{\delta}\right) \\ &= T^{-1}\left(y_1 - Y_{21}\hat{\beta}_1 - Y_{22}\hat{\beta}_2\right)' Q_{Z_1}\left(y_1 - Y_{21}\hat{\beta}_1 - Y_{22}\hat{\beta}_2\right).\end{aligned}$$

The Wald statistics for H_β and H_γ are, respectively,

$$W_\beta = \left(H_1\hat{\beta} - h_1\right)'\left\{H_1\left[Y_2'(P_H - P_{Z_1})Y_2\right]^{-1}H_1'\right\}^{-1}\left(H_1\hat{\beta} - h_1\right)/\hat{\sigma}^2$$

and

$$W_\gamma = (H_2\hat{\gamma} - h_2)'\left[H_2(Z_1'QZ_1)^{-1}H_2'\right]^{-1}(H_2\hat{\gamma} - h_2)/\hat{\sigma}^2,$$

where

$$Q = P_H - P_H Y_2(Y_2'P_H Y_2)^{-1}Y_2'P_H.$$

The following lemmas will be employed in deriving the asymptotic distributions of W_β and W_γ. As in the previous section, we assume that (C2'), (C3), and (C4) hold. Note also that the following lemmas hold under the assumption that the rows of V form a sequence of stationary, ergodic martingale differences with covariance matrix I. This replaces condition (C1).

Lemma 4.1. $\hat{\sigma}^2 \Rightarrow 1 + \beta_1'\beta_1 + \bar{r}_2'\bar{r}_2$.

Lemma 4.2. $[Y_2'(P_H - P_{Z_1})Y_2]^{-1} \Rightarrow S_2\bar{l}_{22}S_2'$, *where*

$$\bar{l}_{22} \equiv \bar{B} + \bar{B}\left[\int_{V_{k_3-n_2,k_3}} W_{n_2}\left(n_1, \bar{l}_{11}^{1/2}\Pi_{21}'\Pi_{21}\bar{l}_{11}^{1/2}\right)\mathrm{d}\Theta\right]\bar{B},$$

$$\bar{B} \equiv W_{n_2}^{-1}(n_2 + k_3 + 1, I_{n_2}),$$

and

$$\bar{l}_{11} \equiv \{\Pi_{21}'\Theta\Theta'\Pi_{21}\}^{-1}.$$

Lemma 4.3. $Z_1'QZ_1 \Rightarrow \Pi_{12}^{+\prime}\xi[Q_F - Q_F\lambda\bar{f}_{11}\lambda'Q_F]\xi\Pi_{12}^{+}$

$$\equiv \Pi_{12}^{+\prime}W_{n_2}(k_1 + k_3 - n, I)\Pi_{12}^{+},$$

where Π_{12}^{+} *is the Moore–Penrose inverse of* Π_{12}, $\xi \equiv N(0, I_{(k_1+k_3)n_2})$,

$$F = \begin{bmatrix} 0 \\ \Pi_{12} \end{bmatrix}, \qquad \lambda = \begin{bmatrix} \Pi_{21} \\ \Pi_{11} \end{bmatrix},$$

and

$$\bar{f}_{11} = (\lambda Q_F\lambda')^{-1}.$$

Remarks

(i) Lemma 4.1 shows that the standard error of regression converges weakly to a random variable due to the lack of identifiability.

(ii) If the coefficient vector β is totally unidentified, we find that $T^{-1}Y'Q_{Z_1}Y \to_p I$ and that $\hat{\sigma}^2 \Rightarrow 1 + \bar{r}'\bar{r}$, where $\bar{r} \equiv \int_{B>0} N(0, B)\mathrm{pdf}(B)\,\mathrm{d}B$ and $\bar{B} \equiv W_n^{-1}(k_3 - n_1 + n_2 + 1)$. This is consistent with the result obtained by Phillips (1989).

(iii) When $\beta = 0$, we find from Lemma 4.1 that $\hat{\sigma}^2 \Rightarrow 1 + \bar{r}_2'\bar{r}_2$. Thus under the null H_0: $\beta = 0$, the nondegeneracy of $\hat{\sigma}^2$ persists in the limit as well.

(iv) If the system is totally identified, we would have $[Y_2'(P_H - P_{Z_1})Y_2]^{-1} = O_p(T^{-1})$. Lemma 4.2 shows that $[Y_2'(P_H - P_{Z_1})Y_2]^{-1} = O_p(1)$ in contrast. The difference arises since the coefficient vector β is only partially identified.

(v) If the coefficient vector β were totally unidentified, we would have $[Y_2'(P_H - P_{Z_1})Y_2]^{-1} \Rightarrow W_n^{-1}(n + k_3 + 1, I)$ as was obtained by Phillips (1989).

(vi) Lemma 4.3 is obtained in the same way as Lemma 2.7 of Phillips (1989). The result can be sharpened as follows:

$$\begin{aligned} Z_1'QZ_1 &\Rightarrow \Pi_{12}^{+\prime} W_{n_2}(k_1 + k_3 - n, I)\Pi_{12}^{+} \\ &\equiv \Pi_{12}^{+\prime} PP'\Pi_{12}^{+} \\ &= \bar{P}\bar{P}'. \end{aligned}$$

Columns of $\bar{P}$ are now iid $N(0, (\Pi_{12}\Pi_{12}')^{+})$. That is, $\bar{P}$ has a singular matrix normal distribution. Note that Lemma 2.7 of Phillips (1989) uses an additional rotation to obtain the required result on the assumptions that Π_1 has full row rank and that $\Pi_2 = 0$.

Theorem 4.4. If the rows of V form a sequence of stationary, ergodic martingale differences with covariance matrix I and if (C2′), (C3), and (C4) apply, then under the null

$$\text{(a)} \qquad W_\beta \Rightarrow \frac{(\bar{r}_2 - \beta_2)'S_2'H_1'\{H_1 S_2 \hat{l}_{22} S_2' H_1'\}^{-1} H_1 S_2(\bar{r}_2 - \beta_2)}{1 + \beta_1'\beta_1 + \bar{r}_2'\bar{r}_2},$$

$$\text{(b)} \qquad W_\gamma \Rightarrow \frac{(\bar{s}_2 - \gamma_2)'R_2'H_2'\left\{H_2\{\Pi_{12}^{+\prime} W_{n_2}(k_1 + k_3 - n, I)\Pi_{12}^{+}\}^{-1} H_2'\right\}^{-1} \times H_2 R_2(\bar{s}_2 - \gamma_2)}{1 + \beta_1'\beta_1 + \bar{r}_2'\bar{r}_2}.$$

Remarks

(i) Theorem 4.4 shows that the limiting distributions of the Wald statistics are not chi-squared. Here the conventional theory for hypothesis tests in simultaneous equations system does not apply. Under the alternative hypotheses H_β: $H_1\beta \neq h_1$ and H_γ: $H_2\gamma \neq h_2$, $[Y_2'(P_H - P_{Z_1})Y_2]^{-1}$, $Z_1'QZ_1$, $H_1\hat{\beta} - h_1$, and $H_2\hat{\gamma} - h_2$ have the same order of magnitude $O_p(1)$, so that the Wald statistics do not diverge under the alternatives as T goes to infinity and the tests are also inconsistent. Of course, this is a consequence of the lack of identification.

(ii) When $\Pi_2 = 0$, our result on W_β reduces to that obtained by Phillips (1989). W_γ has the same limiting distribution as in Theorem 4.1 in this case.

(iii) The limiting distributions of W_β and W_γ when β and γ are totally unidentified were derived by Phillips (1989).

We also study the case in which the coefficient vector γ is fully identified as in Phillips (1989). In this case, we have $\Pi_1 = 0$, so that $Y_{22} = V_{22}$. Hence it follows that $Z_1'QZ_1 = O_p(T)$ as below:

Lemma 4.5. If $\Pi_1 = 0$, *then*

$$T^{-1}Z_1'QZ_1 \Rightarrow \rho\Phi\left[I - \Phi'\Psi(\Psi'\Phi\Phi'\Psi)^{-1}\Psi'\Phi\right]\Phi'\rho',$$

where

$$\Psi = \begin{bmatrix} \Pi_{21} \\ 0 \end{bmatrix}, \quad \rho = [0, I],$$

and Φ *is distributed uniformly on the Stiefel manifold* $V_{n_2, k_1+k_3} = \{\Phi: \Phi'\Phi = I_{n_2}\}$.

Theorem 4.6. If $\Pi_1 = 0$, *then under the same conditions as in Theorem 4.4,*

$$W_\gamma \Rightarrow \frac{\bar{s}'H_2'\left\{H_2\left\{\rho\Phi\left[I - \Phi'\Psi(\Psi'\Phi\Phi'\Psi)^{-1}\Psi'\Phi\right]\Phi'\rho'\right\}^{-1}H_2\right\}^{-1}H_2\bar{s}}{1 + \beta_1'\beta_1 + \bar{r}_2'\bar{r}_2}.$$

Here

$$\bar{s} = \int_{R^{n_2}} N\left(0, T^{-1}\left(1 + \beta_1'\beta_1 + \bar{r}_2'\bar{r}_2\right)\right)\mathrm{pdf}(\bar{r}_2)\,\mathrm{d}\bar{r}_2.$$

Remarks

(i) When $\Pi_2 = 0$, Lemma 4.4 yields $T^{-1}Z_1'QZ_1 \Rightarrow \rho\Phi\Phi'\rho'$. This can be sharpened further to give rise to a result equivalent to Lemma 2.9 of Phillips (1989).

(ii) Under the alternative H_γ: $H_2\gamma \neq h_2$, $\sqrt{T}(H_2\hat{\gamma} - h_2)$ diverges, and so does the statistic W_γ. The test is therefore consistent. However, the limiting distribution of W_γ is not chi-squared due to the nonidentifiability of β.

(iii) When $\Pi_2 = 0$, Theorem 4.6 reduces to the result discussed in Theorem 2.10 of Phillips (1989).

5. Illustration of probability density functions

We illustrate the finite sample and asymptotic distributions of the two-stage least squares (2SLS) estimator in a single equation of a simultaneous equation system. We consider the structural equation (1), the reduced form equations of which are

$$y_1 = Z_1\pi_1 + Z_2\pi_2 + v_1,$$

$$Y_{21} = Z_1\Pi_{11} + Z_2\Pi_{21} + V_{11},$$

$$Y_{22} = Z_1\Pi_{12} + Z_2\Pi_{22} + V_{12}.$$

It is assumed that an appropriate rotation is performed to partition Y_2, Π_1, Π_2, and V_2 as in section 2. The endogenous variables, y_1, Y_{21}, and Y_{22} are $T\times 1$ vectors. The exogenous variables Z_1 and Z_2 are $T\times 1$ and $T\times 2$ vectors, respectively. The values of parameters in the reduced form equations are set to be $\pi_1 = 2$, $\pi_2' = [0,0]$, $\Pi_{11} = 9$, $\Pi_{21}' = [1,5]$, $\Pi_{12} = 2.5$, and $\Pi_{22}' = [0,0]$. Notice that the value of β_1 is zero by construction. The identifying relations show that β_1 is identified and that β_2 and γ are not identified in this experimental format. For the purpose of simulation, 4,000 iterations are made to generate normal variates for v_1, V_{21}, and V_{22}. For Z_1 and Z_2, random numbers from a uniform distribution are generated. Since the variables Z_1 and Z_2 are exogenous, they are fixed throughout the iterations. In each iteration, we calculate the 2SLS estimates of β_1, β_2, and γ using the formulae in section 3. These are used to plot the finite sample and asymptotic distributions. Since the parameter β_1 is identified, its estimator $\hat{\beta}_1$ has a distribution which is not invariant to the sample size, as is discussed in Phillips (1988). At a given sample size, the finite sample distribution is plotted by using $\hat{\beta}_1$, while the asymptotic distribution is charted by the rescaled values of $\hat{\beta}_1$, i.e., $T^{1/2}\hat{\beta}_1$. The 2SLS estimator of β_2 is known to be invariant to the sample size and, as discussed in Phillips (1988), has a standard Cauchy distribution. The parameter γ is unidentified, but its estimator has a distribution that varies with the sample size.

Fig. 1 illustrates the finite sample and asymptotic pdf of the identified coefficient estimator $\hat{\beta}_1$. Sample sizes $T = 30, 80, 200$ are used. As discussed in section 4, both the pdf's are scale mixture normal. The identifiability of β_1 is manifested by the fact that the pdf's are centered on the true value of β_1 and concentrate as T increases. In fig. 2, the pdf of $T^{1/2}\hat{\beta}_1$ at $T = 80$ is plotted together with normal pdf's. The normal pdf's are charted by a normal density function formula with the variance $\Pi_{21}' M_{22\cdot 1}^{1/2}\Theta\Theta' M_{22\cdot 1}^{1/2}\Pi_{21}$ conditioned on certain Θ as in Corollary 3.1. Note that $M_{22\cdot 1} = M_{22} - M_{21}M_{11}^{-1}M_{12}$. The pdf's of normal 1 and normal 2 are generated by setting $\Theta' = [\cos(\pi/2), \sin(\pi/2)]$ and $\Theta' = [\cos(0), \sin(0)]$, respectively. The tail of the pdf of $T^{1/2}\hat{\beta}_1$ is in between those of normal 1 and normal 2, reflecting the scale mixture normality of $T^{1/2}\hat{\beta}_1$. Fig. 3 shows the pdf's of $\hat{\beta}_2$ and the standard

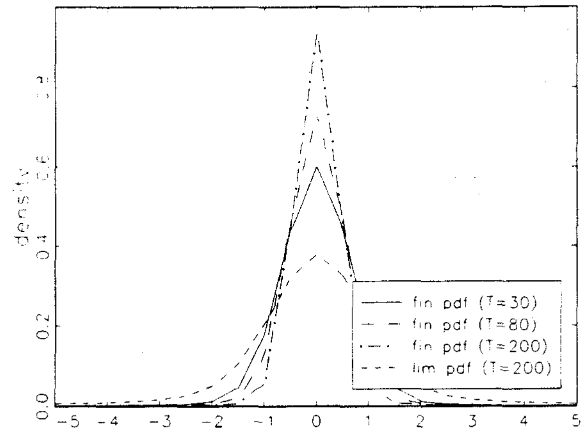

Fig. 1. Pdf's of beta 1.

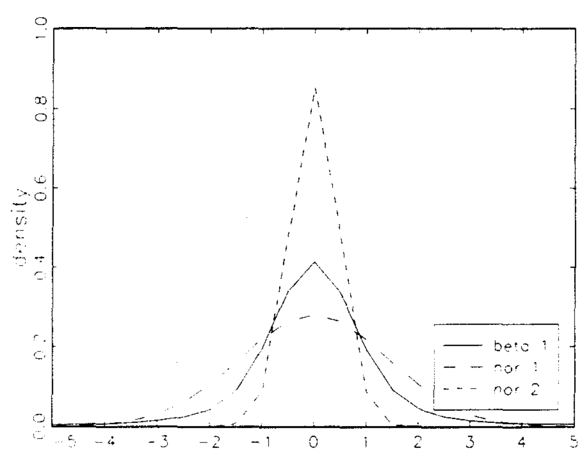

Fig. 2. Pdf's of beta 1 and normals.

Cauchy. The sample size here is $T = 80$. We find that both pdf's have similar shapes.

The pdf's of $\hat{\gamma}$ at $T = 30, 80$ are plotted in fig. 4. The distributions are shown to have very large variances. This is what we would expect from the fact that $\hat{\beta}_2$ has a Cauchy distribution. Both the pdf's show only a slight difference in shape.

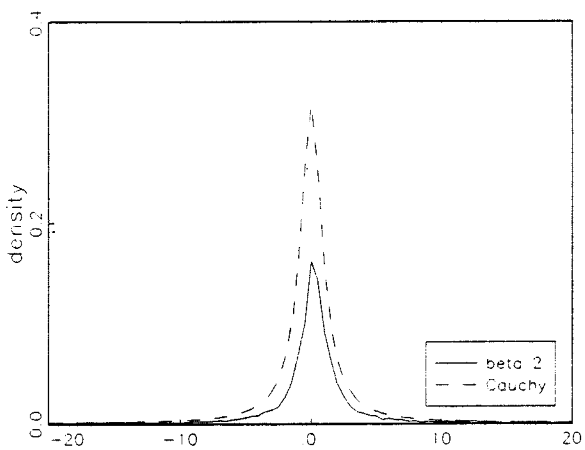

Fig. 3. Pdf's of beta 2 and Cauchy.

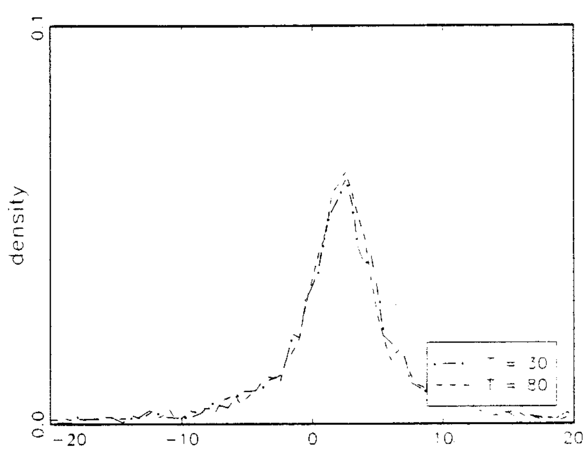

Fig. 4. Pdf's of unidentified gamma.

Fig. 5 and fig. 6 display the empirical pdf's of the Wald statistics under the null hypotheses on β and on γ at $T = 30, 80$ against the chi-square distribution with the degrees of freedom 2. Here

$$\pi_1' = [2,0], \quad \pi_2' = [0,0], \quad \Pi_1 = \begin{bmatrix} 9 & 1 \\ 0 & 0 \end{bmatrix}, \quad \Pi_2 = \begin{bmatrix} 1 & 0 \\ 5 & 0 \end{bmatrix},$$

hence β and γ are only partially identified. When β is partially identified, fig.

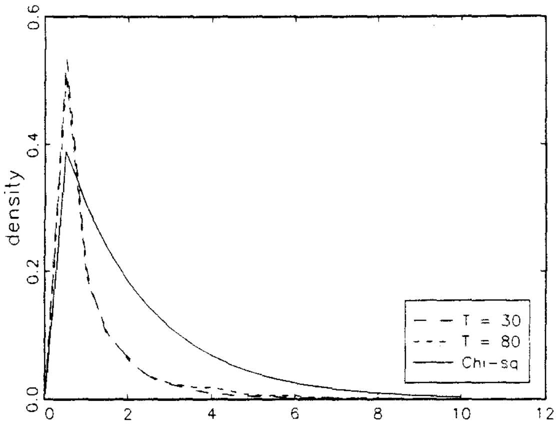

Fig. 5. Pdf's of the Wald statistics.

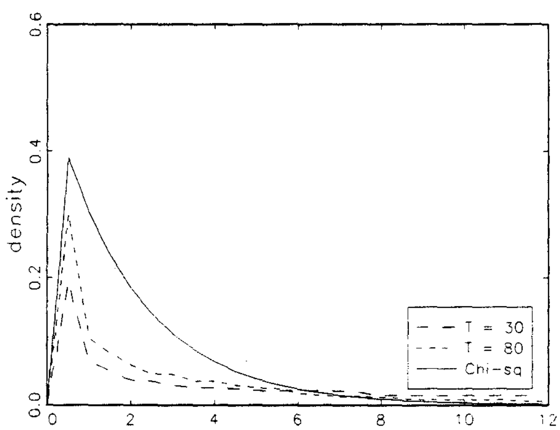

Fig. 6. Pdf's of the Wald statistics.

5 shows that the null distributions of the Wald statistics have high peaks around the origin and that their tails are thinner than that of the chi-square distribution. Contrastingly, when γ is partially identified, the null distributions of the Wald statistics display more dispersion than the chi-square variates, as we see fig. 6. Obviously we find that using a nominal chi-square in hypothesis testing in case of identification failure results in size distortion.

6. Conclusion

General formulae for the finite sample and asymptotic distributions of the instrumental variable estimators and Wald test statistics in a simultaneous equation model are derived under the assumption of partial identification of the coefficient vectors of both the endogenous and exogenous variables. In order to isolate identifiable components from the coefficient vectors, we employed a simple rotation of the coordinate system. Both identified and unidentified components estimates can then be studied separately and the consequences for estimation and inference in the original coordinate system follow directly.

Only estimators of identified coefficient vector have finite sample distributions that carry information about the true coefficient vectors. The estimators of the unidentified coefficient vectors are independent of the true coefficients. But they are not invariant to the sample size. All the finite sample distributions are expressed as mixed normal. The formulae derived are general enough to include totally identified, totally unidentified and general partially identified systems. Asymptotic analysis informs us that only the estimators of the identified coefficient vector have meaningful limit distributions. However, these are not normal, demonstrating that the effects of the nonidentifiability of other components persist asymptotically. Only the estimators of the identified coefficient vector are consistent of course. The limit distributions of Wald statistics for the whole coefficient vectors are not chi-square, as we would expect from the lack of identification. As a matter of fact, these tests are inconsistent in the general case of partial identification. Simulation results show the empirical pdf's of the estimators of both the identified and unidentified coefficients and of the Wald statistics under the null against the respective asymptotic distributions. The difference in the properties of the estimators and of the test statistics in the two polar cases is vividly shown in the figures. These results indicate the importance of identification in simultaneous equations theory and highlight the consequences for estimation and for statistical inference of identification failure. In empirical research where order conditions for identification are all that are checked, these consequences should always be borne in mind.

Appendix

Proof of Theorem 3.1

(a) The coefficient estimate of β_1 is written as

$$\hat{\beta}_1 = (Y_{21}'EY_{21})^{-1}(Y_{21}'Ey_1),$$

where $E = L - LY_{22}(Y_{22}'LY_{22})^{-1}Y_{22}'L$ and $L = P_{II} - P_{Z_1}$. Factorize the idempotent matrix L as $L = D_1D_1'$, where $D_1 = Q_{Z_1}Z_3(Z_3'Q_{Z_1}Z_3)^{-1/2}$ [cf. eq. (33) of Phillips (1989)]. Using this factorization, we write E as

$$E = D_1D_1' - D_1D_1'V_{22}(V_{22}'D_1D_1'V_{22})^{-1}V_{22}'D_1D_1'.$$

It follows from (C1) that $\delta = D_1'V_{22} \equiv N(0, I)$. Hence E can be rewritten as follows:

$$\begin{aligned} E &= D_1D_1' - D_1\delta(\delta'\delta)^{-1}\delta'D_1' \\ &= D_1\left[I - \delta(\delta'\delta)^{-1}\delta'\right]D_1' \\ &= Q_{Z_1}Z_3\left(Z_3'Q_{Z_1}Z_3\right)^{-1/2}\Theta\Theta'\left(Z_3'Q_{Z_1}Z_3\right)^{-1/2}Z_3'Q_{Z_1}, \end{aligned}$$

where Θ is distributed uniformly on the Stiefel manifold

$$V_{k_3-n_2,k_3} = \left\{\Theta : \Theta'\Theta = I_{k_3-n_2}\right\}.$$

We find trivially that

$$\Theta'\left(Z_3'Q_{Z_1}Z_3\right)^{-1/2}Z_3'Q_{Z_1}Y_{21}|_{V_{22}} \equiv N\left(\Theta'\Lambda_T\Pi_{21}, I_{(k_3-n_2)n_1}\right), \qquad (A.1)$$

where $\Lambda_T = (Z_3'Q_{Z_1}Z_3)^{-1/2}(Z_3'Q_{Z_1}Z_2) = T^{1/2}W'$ under (C2). Standard multivariate theory then yields the (conditional) Wishart distribution:

$$\begin{aligned} Y_{21}'EY_{21}|_{V_{22}} &= W_{n_1}\left(k_3 - n_2, I_{n_1}, \Pi_{21}'\Lambda_T'\Theta\Theta'\Lambda_T\Pi_{21}\right) \\ &= W_{n_1}\left(k_3 - n_2, I_{n_1}, T\Pi_{21}'W\Theta\Theta'W'\Pi_{21}\right) \end{aligned}$$

and (conditional) inverted Wishart

$$\begin{aligned} A(\Theta\Theta')|_{\Theta} &= \left(Y_{21}'EY_{21}\right)^{-1}|_{V_{22}} \\ &\equiv W_{n_1}^{-1}\left(k_3 - n_2 + n_1 + 1, I_{n_1}, \Pi_{21}'\Lambda_T'\Theta\Theta'\Lambda_T\Pi_{21}\right) \\ &= W_{n_1}^{-1}\left(k_3 - n_2 + n_1 + 1, I_{n_1}, T\Pi_{21}'W\Theta\Theta'W'\Pi_{21}\right). \end{aligned} \qquad (A.2)$$

Next

$$Y_{21}'Ey_1\Big|_{\substack{V_{21} \\ V_{22}}} \equiv N\left(Y_{21}'EZ_2\Pi_{21}\beta_1, Y_{21}'EY_{21}\right). \qquad (A.3)$$

Set $g(\Theta\Theta') = Y_{21}' E Z_2 \Pi_{21}$, and note that $g(\Theta\Theta')$ is normally distributed conditional on V_{22} (or equivalently on Θ) as follows:

$$\begin{aligned} g(\Theta\Theta')|_{\Theta} &\equiv N(\Pi_{21}' Z_2 E Z_2 \Pi_{21}, \Pi_{21}' Z_2' E Z_2 \Pi_{21}) \\ &= N(\Pi_{21}' \Lambda_T' \Theta\Theta' \Lambda_T \Pi_{21}, \Pi_{21}' \Lambda_T' \Theta\Theta' \Lambda_T \Pi_{21}) \\ &= N(T\Pi_{21}' W \Theta\Theta' W' \Pi_{21}, T\Pi_{21}' W\Theta\Theta' W' \Pi_{21}). \end{aligned}$$

Combining (A.2) and (A.3), we have

$$\begin{aligned} \hat{\beta}\Big|_{\substack{V_{21}\\V_{22}}} &= (Y_{21}' E Y_{21})^{-1} (Y_{21}' E y_1) \\ &\equiv N(A(\Theta\Theta') g(\Theta\Theta') \beta_1, A(\Theta\Theta')). \end{aligned}$$

Next, integrating with respect to A and g conditional on Θ, we find

$$\begin{aligned} \hat{\beta}_1 |_{V_{22}} = \int_{g \in R^{n_2}} \int_{A>0} & N(A(\Theta\Theta') g(\Theta\Theta') \beta_1, A(\Theta\Theta')) \\ & \times \mathrm{pdf}_{\Theta}(A, g) \, \mathrm{d}A \, \mathrm{d}G, \end{aligned}$$

where $\mathrm{pdf}_{\Theta}(A, g)$ denotes the joint pdf of A and g conditional on Θ. Finally, letting V_{22} go free and integrating over $\Theta \in V_{k_3 - n_2, k_3}$, we obtain

$$\begin{aligned} \hat{\beta}_1 = \int_{V_{k_3-n_2,k_3}} \int_{g \in R^{n_2}} \int_{A>0} & N(A(\Theta\Theta') g(\Theta\Theta') \beta_1, A(\Theta\Theta')) \\ & \times \mathrm{pdf}_{\Theta}(A, g) \, \mathrm{d}A \, \mathrm{d}G \, \mathrm{d}\Theta. \end{aligned}$$

(b) The coefficient estimate of β_2 is

$$\hat{\beta}_2 = (Y_{22}' J Y_{22})^{-1} (Y_{22}' J y_1),$$

where $J = L - LY_{21}(Y_{21}' L Y_{21})^{-1} Y_{21}' L$. L is defined as in (a) and factorized in the same manner. J is rewritten such that

$$\begin{aligned} J &= D_1 D_1' - D_1 D_1'(Z_2 \Pi_{21} + V_{21})((Z_2 \Pi_{21} + V_{21})' D_1 D_1' (Z_2 \Pi_{21} + V_{21}))^{-1} \\ &\quad \times (Z_2 \Pi_{21} + V_{21})' D_1 D_1' \\ &= D_1 \left[I - \Theta(\Theta'\Theta)^{-1} \Theta' \right] D_1', \end{aligned}$$

where

$$\Theta = D_1'(Z_2\Pi_{21} + V_{21}) = \Lambda_T\Pi_{21} + D_1'V_{21} \equiv N(T^{1/2}W'\Pi_{21}, I) \qquad (A.4)$$

and $\Lambda_T = (Z_3'Q_{Z_1}Z_3)^{-1/2}(Z_3'Q_{Z_1}Z_2) = T^{1/2}W'$. Proceeding in the same way as (a) we find that

$$J = Q_{Z_1}Z_3(Z_3'Q_{Z_1}Z_3)^{-1/2}\Omega\Omega'(Z_3'Q_{Z_1}Z_3)^{-1/2}Z_3'Q_{Z_1},$$

where $\Omega = \Omega(\theta)$ is distributed on the Stiefel manifold $V = \{\Omega; \Omega'\Omega = I_{k_3-n_1}\}$ with a nonuniform distribution induced by that of θ in (A.4). It is easily deduced that

$$\Omega'(Z_3'Q_{Z_1}Z_3)^{-1/2}Z_3Q_{Z_1}Y_{22}|_{V_{21}} \equiv N(0, I_{(k_3-n_1)n_2}).$$

This is also an unconditional distribution. Hence we find that

$$Y_{22}'JY_{22} \equiv W_{n_2}(k_3 - n_1, I_{n_2})$$

and that

$$B = (Y_{22}'JY_{22})^{-1} \equiv W_{n_2}^{-1}(k_3 - n_1 + n_2 + 1, I_{n_2}). \qquad (A.5)$$

Next we consider the conditional distribution

$$Y_{22}'Jy_1\Big|_{\substack{V_{21}\\V_{22}}} \equiv N(Y_{22}'JZ_2\Pi_{21}\beta_1, Y_{22}'JY_{22})\Big|_{\substack{V_{21}\\V_{22}}}. \qquad (A.6)$$

Observe that $m(\theta)|_\theta = Y_{22}'JZ_2\Pi_{21}|_{V_{21}}$ is distributed as

$$\begin{aligned} m(\theta)|_\theta &\equiv N(0, \Pi_{21}'Z_2'JZ_2\Pi_{21}) \\ &= N(0, \Pi_{21}'\Lambda_T'\Omega\Omega'\Lambda_T\Pi_{21}) \\ &= N(0, T\Pi_{21}'W'\Omega\Omega'W'\Pi_{21}). \end{aligned}$$

From (A.5) and (A.6) we have

$$\hat{\beta}_2\Big|_{\substack{V_{21}\\V_{22}}} \equiv (Y_{22}'JY_{22})(Y_{22}'Jy_1)\Big|_{\substack{V_{21}\\V_{22}}} = N(Bm(\theta)\beta_1, B).$$

In the manner of (a), we then have

$$\hat{\beta}_2|_{V_{21}} = \int_{m\in R^{n_1 n_2}} \int_{B>0} N(Bm(\theta)\beta_1, B)\,\mathrm{pdf}_\theta(B,m)\,\mathrm{d}B\,\mathrm{d}m,$$

and finally

$$\hat{\beta}_2 \equiv \int_{\theta\in R^{k_2 n_1}} \int_{m\in R^{n_1 n_2}} \int_{B>0} N(Bm(\theta)\beta_1, B)$$

$$\times\,\mathrm{pdf}_\theta(B,m)\,\mathrm{d}B\,\mathrm{d}m\,\mathrm{pdf}(\theta)\,\mathrm{d}\theta,$$

where $\mathrm{pdf}_\theta(B,m)$ denotes the joint probability density of B and m conditional on θ.

Proof of Corollary 3.1

(a) Write

$$Y_{21}'EY_{21} = (\Pi_{21}'Z_2' + V_{21}')D_1\left[I - \delta(\delta'\delta)^{-1}\delta'\right]D_1'(Z_2\Pi_{21} + V_{21}).$$

Now characterize the asymptotic behavior of each term on the right-hand side. First,

$$\begin{aligned}
T^{-1/2}D_1'Z_2\Pi_{21} &= T^{-1/2}\left(Z_3'Q_{Z_1}Z_3\right)^{-1/2}\left(Z_3'Q_{Z_1}Z_2\right)\Pi_{21} \\
&= T^{-1/2}\left(Z_3'Z_3 - Z_3'Z_1(Z_1'Z_1)^{-1}Z_1'Z_3\right)^{-1/2} \\
&\quad \times\left(Z_3'Z_2 - Z_3'Z_1(Z_1'Z_1)^{-1}Z_1'Z_2\right)\Pi_{21} \\
&\to \left(M_{33} - M_{31}M_{11}^{-1}M_{13}\right)^{-1/2}\left(M_{32} - M_{31}M_{11}^{-1}M_{12}\right)\Pi_{21} \\
&= M_{33\cdot 1}^{-1/2}M_{32\cdot 1}\Pi_{21} \\
&= \Pi_{21}.
\end{aligned}$$

Next,

$$\delta = D_1'V_{22} \Rightarrow N(0, I_{k_3 n_2}), \qquad D_1'V_{21} \Rightarrow N(0, I_{k_3 n_2}).$$

The last two results follow directly from Phillips (1989, lemma 2.3). Thus we obtain the limit theory:

$$T^{-1}Y_{21}'EY_{21} \Rightarrow \Pi_{21}'M_{32\cdot 1}'M_{33\cdot 1}^{-1/2}\Theta\Theta'M_{33\cdot 1}^{-1/2}M_{32\cdot 1}\Pi_{21}.$$

Similarly, we write

$$Y_{21}'Ey_1 = (\Pi_{21}'Z_2' + V_{21}')D_1\left[I - \delta(\delta'\delta)^{-1}\delta'\right]D_1'(Z_2\pi_2 + v_1).$$

Since $D_1'v_1 \Rightarrow N(0, I_{k_3})$, we find that

$$T^{-1}Y_{21}'EZ_2\pi_2 \Rightarrow \Pi_{21}'M_{32\cdot 1}'M_{33\cdot 1}^{-1/2}\Theta\Theta'M_{33\cdot 1}^{-1/2}M_{32\cdot 1}\pi_2$$

$$= \Pi_{21}'M_{32\cdot 1}'M_{33\cdot 1}^{-1/2}\Theta\Theta'M_{33\cdot 1}^{-1/2}M_{32\cdot 1}\Pi_{21}\beta_1$$

and

$$T^{-1/2}Y_{21}'Ev_1 \Rightarrow \int_{V_{k_3-n_2,k_3}} N(0, \Pi_{21}'M_{32\cdot 1}'M_{33\cdot 1}^{-1/2}\Theta\Theta'M_{33\cdot 1}^{-1/2}M_{32\cdot 1}\Pi_{21})\,\mathrm{d}\Theta.$$

Rewrite $\hat{\beta}_1$ as

$$\hat{\beta}_1 = (Y_{21}'EY_{21})^{-1}(Y_{21}'Ey_1)$$

$$= (Y_{21}'EY_{21})^{-1}(Y_{21}'EZ_2\Pi_{21}\beta_1) + (Y_{21}'EY_{21})^{-1}(Y_{21}'EV_1)$$

$$= \left(T^{-1}Y_{21}'EY_{21}\right)^{-1}\left(T^{-1}Y_{21}'EZ_2\Pi_{12}\beta_1\right)$$

$$+ T^{-1/2}\left(T^{-1}Y_{21}'EY_{21}\right)^{-1}\left(T^{-1/2}Y_{21}'Ev_1\right).$$

Thus,

$$\sqrt{T}\,(\hat{\beta}_1 - \beta_1)|_{V_{22}} \Rightarrow N\left(0, \left(\Pi_{21}'M_{32\cdot 1}'M_{33\cdot 1}^{-1/2}\Theta\Theta'M_{33\cdot 1}^{-1/2}M_{32\cdot 1}\Pi_{21}\right)^{-1}\right),$$

and removing the conditioning we obtain

$$\sqrt{T}\,(\hat{\beta}_1 - \beta_1) \Rightarrow \int_{V_{k_3-n_2,k_3}} N\left(0, (\Pi_{21}'M_{32\cdot 1}'M_{33\cdot 1}^{-1/2}\Theta\Theta'M_{33\cdot 1}^{-1/2}M_{32\cdot 1}\Pi_{21})^{-1}\right)$$

$$= \int_{V_{k_3-n_2,k_3}} N(0, (\Pi_{21}'\Theta\Theta'\Pi_{21})^{-1})\,\mathrm{d}\Theta.$$

(b) Write

$$Y_{22}'Jy_1\Big|_{\substack{V_{21}\\ V_{22}}} = Y_{22}'J(Z_2\pi_2 + v_1)$$

$$\equiv Y_{22}'JZ_2\pi_2 + N(0, Y_{22}'JY_{22}).$$

Here

$$Y_{22}'JZ_2\pi_2 = V_{22}'D_1\left[I - \theta(\theta'\theta)^{-1}\theta'\right]D_1'Z_2\pi_2.$$

Since

$$T^{-1/2}\theta = T^{-1/2}\Lambda_T\Pi_{21} + T^{-1/2}D_1'V_{21} \underset{p}{\rightarrow} W'\Pi_{21} = \bar{\theta}, \quad \text{say},$$

and $D_1'V_{22} \Rightarrow N(0, I_{k_3 n_2})$, it follows that $Q_\theta \rightarrow_p Q_{\bar{\theta}}$ and $T^{-1/2}Q_\theta D_1'Z_2\pi_2 = T^{-1/2}Q_\theta D_1'Z_2\Pi_{21}\beta \rightarrow_p 0$. Thus

$$T^{-1/2}Y_{22}'JZ_2\pi_2 \underset{p}{\rightarrow} 0. .$$

Moreover, since $T^{-1}Z'Z = I + O(T^{-1})$, we have

$$D_1'Z_2 = \left(Z_3'Q_{Z_1}Z_3\right)^{-1/2}\left(Z_3'Q_{Z_1}Z_2\right) = T^{1/2}W' + O(T^{-1/2}).$$

Next note that $Q_\theta D_1'Z_2 = (Q_{\bar{\theta}} + O_p(T^{-1}))(T^{1/2}W + O(T^{-1/2})) = o_p(1)$, since $Q_{\bar{\theta}}W = 0$. Thus $Y_{22}'JZ_2\pi_2 \rightarrow_p 0$ also. Now consider

$$\begin{aligned} Y_{22}'JY_{22} &\Rightarrow V_{22}'D_1'\left[I - \theta(\theta'\theta)^{-1}\theta'\right]D_1'V_{22} \\ &\Rightarrow N(0, I)Q_{\bar{\theta}}N(0, I) \\ &= N(0, I)\Delta\Delta'N(0, I), \end{aligned}$$

where $\Delta'\Delta = I$. Hence $Y_{22}'JY_{22} \Rightarrow B \equiv W_{n_2}(k_3 - n_1, I_{n_2})$. Next decompose $\hat{\beta}_2$ as

$$\begin{aligned} \hat{\beta}_2 &= (Y_{22}'JY_{22})^{-1}(Y_{22}'JZ_2\pi_2) + (Y_{22}'JY_{22})^{-1}(Y_{22}'Jv_1) \\ &= (Y_{22}'JY_{22})^{-1}(Y_{22}'Jv_1) + o_p(1), \end{aligned}$$

and since

$$(Y_{22}'JY_{22})^{-1}Y_{22}'Jv_1 \Rightarrow \int_{B>0} N(0, B)\,\mathrm{pdf}(B)\,\mathrm{d}B,$$

the stated limit law follows directly.

Proof of Theorem 3.2

(a) $\hat{\gamma}_1$ is written as

$$\begin{aligned}\hat{\gamma}_1 &= R_1'\hat{\gamma}\\ &= R_1'(Z_1'Z_1)^{-1}Z_1'y_1 - R_1'(Z_1'Z_1)^{-1}Z_1'[Y_{21}Y_{22}]\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix}\\ &= R_1'\pi_1 + R_1'(Z_1'Z_1)^{-1}Z_1'Z_2\pi_2 - R_1'(Z_1'Z_1)^{-1}Z_1'Z_2\Pi_{21}\hat{\beta}_1\\ &\quad + R_1'(Z_1'Z_1)^{-1}Z_1'V\begin{bmatrix}1\\ -\hat{\beta}_1\\ -\hat{\beta}_2\end{bmatrix}.\end{aligned}$$

$Z_1'V$ is independent of $\hat{\beta}_1 = r_1$ and $\hat{\beta}_2 = r_2$, since $Z_1'E = Z_1'J = 0$. Thus, we have the conditional distribution

$$\begin{aligned}\hat{\gamma}_1|_{r_2}^{r_1} &\equiv N\big(\gamma_1 + R_1'(Z_1'Z_1)^{-1}Z_1'Z_2\pi_2 - R_1'(Z_1'Z_1)^{-1}Z_1'Z_2\Pi_{21}r_1,\\ &\qquad (1 + r_1'r_1 + r_2'r_2)R_1'(Z_1'Z_1)^{-1}R_1\big)\\ &= N\big(\gamma_1, T^{-1}(1 + r^{*\prime}r^*)\big),\end{aligned}$$

so that

$$\hat{\gamma}_1 \equiv \int_{R^{n_1+n_2}} N\big(\gamma_1, T^{-1}(1 + r^{*\prime}r^*)\big)\mathrm{pdf}(r^*)\,\mathrm{d}r^*.$$

(b) We write $\hat{\gamma}_2$ as

$$\begin{aligned}\hat{\gamma}_2 &= R_2'\hat{\gamma}\\ &= R_2'(Z_1'Z_1)^{-1}Z_1'y_1 - R_2'(Z_1'Z_1)^{-1}Z_1'[Y_{21} \quad Y_{22}]\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix}\\ &= R_2'\pi_1 - R_2'(Z_1'Z_1)^{-1}Z_1'Z_2\pi_2 - R_2'\Pi_{11}\hat{\beta}_1\\ &\quad - R_2'(Z_1'Z_1)^{-1}Z_1'Z_2\Pi_{21}\hat{\beta}_1 - R_2'\Pi_{21}\hat{\beta}_2\\ &\quad + R_2'(Z_1'Z_1)^{-1}Z_1'V\begin{bmatrix}1\\ -\hat{\beta}_1\\ -\hat{\beta}_2\end{bmatrix}.\end{aligned}$$

In the same way as part (a), we obtain

$$\hat{\gamma}_2\Big|_{\substack{r_1\\ r_2}} \equiv N\big(R_2'\pi_1 + R_2'(Z_1'Z_1)^{-1}Z_2'Z_2\pi_2 - R_2'\Pi_{11}r_1$$

$$-R_2'(Z_1'Z_1)^{-1}Z_1'Z_2\Pi_{21}r_1 - R_2'\Pi_{12}r_2,$$

$$(1 + r_1'r_1 + r_2'r_2)R'(Z_1'Z_1)^{-1}R_2\big)$$

$$= N\big(R_2'\pi_1 - R_2'\Pi_{11}r_1 - R_2'\Pi_{12}r_2, T^{-1}(1 + r_1'r_1 + r_2'r_2)\big)$$

$$= N\big(R_2'\pi_1 - R_2'\Pi_1 Sr^*, T^{-1}(1 + r^{*\prime}r^*)\big),$$

since

$$[\Pi_{11} \quad \Pi_{12}]\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \Pi_1 Sr^*.$$

Hence, the result follows as in part (a).

Proof of Corollary 3.2

(a) Using $\pi_2 - \Pi_{21}\hat{\beta}_1 = \pi_2 - \Pi_{21}\beta_1 - \Pi_{21}(\hat{\beta}_1 - \beta_1) = -\Pi_{21}(\hat{\beta}_1 - \beta_1)$, we deduce from Corollary 3.1(a) and Theorem 3.2(a) that

$$\sqrt{T}(\hat{\gamma}_1 - \gamma_1) \sim -R_1'M_{11}^{-1}M_{12}\Pi_{21}\sqrt{T}(\hat{\beta}_1 - \beta_1)$$

$$+R_1M_{11}^{-1}(T^{-1/2}Z_1'V)\begin{pmatrix} 1 \\ -\hat{\beta}_1 \\ -\hat{\beta}_2 \end{pmatrix}$$

$$\Rightarrow \int_{R^{n_2}} N\big(0, T^{-1}(1 + \beta_1'\beta_1 + \bar{r}_2'\bar{r}_2)\big)\mathrm{pdf}(\bar{r}_2)\,\mathrm{d}\bar{r}_2.$$

(b) This is easily deduced from Theorem 3.2(b).

Proof of Lemma 4.1

$$\hat{\sigma}^2 = T^{-1}\big(y_1 - Y_{21}\hat{\beta}_1 - Y_{22}\hat{\beta}_2\big)'Q_{Z_1}\big(y_1 - Y_{21}\hat{\beta}_1 - Y_{22}\hat{\beta}_2\big)$$

$$= \big(1, -\hat{\beta}_1, -\hat{\beta}_2\big)\big(T^{-1}Y'Q_{Z_1}Y\big)\begin{pmatrix} 1 \\ -\hat{\beta}_1 \\ -\hat{\beta}_2 \end{pmatrix}.$$

We consider the middle term $T^{-1}Y'Q_{Z_1}Y$ and write $Q_{Z_1} = I - D_2D_2'$ where $D_2 = Z_1(Z_1'Z_1)^{-1/2}$. Then we have

$$Y'Q_{Z_1}Y = \begin{bmatrix} (Z_2\pi_2 + V)' \\ (Z_1\Pi_{21} + V_{21})' \\ V_{22}' \end{bmatrix} (I - D_2D_2') \times (Z_2\pi_2 + V_1 \quad Z_2\Pi_{21} + V_{21} \quad V_{22}).$$

Since we know that $D_2'V_1 \Rightarrow N(0, I_{k_1})$, $D_2'V_{21} \Rightarrow N(0, I_{k_1n_1})$, $D_2'V_{22} \Rightarrow N(0, I_{k_2n_2})$, $T^{-1}D_2'Z_2 \to_p 0$, and $T^{-1}Z_2'D_2D_2'Z_2 \to M_{21}M_{11}^{-1}M_{12}$, we obtain the following results:

$$T^{-1}(Z_2\pi_2 + v_1)'(Z_2\pi_2 + v_1) \underset{p}{\to} \pi_2'M_{22}\pi_2 + 1 = \pi_2'\pi_2 + 1,$$

$$T^{-1}(Z_2\pi_2 + v_1)'(Z_2\Pi_{21} + V_{21}) \underset{p}{\to} \pi_2'M_{22}\Pi_{21} = \pi_2'\Pi_{21},$$

$$T^{-1}(Z_2\pi_2 + v_1)'V_{22} \underset{p}{\to} 0,$$

$$T^{-1}(Z_2\Pi_{21} + V_{21})'(Z_2\Pi_{21} + V_{21}) \underset{p}{\to} \Pi_{21}'M_{22}\Pi_{21} + I_{n_1} = \Pi_{21}'\Pi_{21} + I_{n_1},$$

$$T^{-1}(Z_2\Pi_{21} + V_{21})'V_{22} \underset{p}{\to} 0,$$

$$T^{-1}V_{22}'V_{22} \underset{p}{\to} I_{n_2},$$

$$T^{-1}(Z_2\pi_2 + v_1)'D_2D_2'(Z_2\pi_2 + v_1) \underset{p}{\to} \pi_2'M_{21}M_{11}^{-1}M_{12}\pi_2 = 0,$$

$$T^{-1}(Z_2\pi_2 + v_1)'D_2D_2'(Z_2\Pi_{21} + V_{21}) \underset{p}{\to} \pi_2'M_{21}M_{11}^{-1}M_{12}\Pi_{21} = 0,$$

$$T^{-1}(Z_2\pi_2 + v_1)'D_2D_2'V_{22} \underset{p}{\to} 0,$$

$$T^{-1}(Z_2\Pi_{21} + V_{21})'D_2D_2'(Z_2\Pi_{21} + V_{21}) \underset{p}{\to} \Pi_{21}'M_{21}M_{11}^{-1}M_{12}\Pi_{21} = 0,$$

$$T^{-1}(Z_2\Pi_{21} + V_{21})'D_2D_2'V_{22} \underset{p}{\to} 0,$$

$$T^{-1}V_2'D_2D_2'V_{22} \underset{p}{\to} 0.$$

Hence,

$$T^{-1}Y'Q_{Z_1}Y \underset{p}{\rightarrow} \begin{bmatrix} \pi_2' M_{22}\pi_2 + 1 & \pi_2' M_{22}\Pi_{21} & 0 \\ \Pi_{21}' M_{22}\pi_2 & \Pi_{21}' M_{22}\Pi_{21} + I_{n_1} & 0 \\ 0 & 0 & I_{n_2} \end{bmatrix}$$

$$- \begin{bmatrix} \pi_2' M_{21}M_{11}^{-1}M_{12}\pi_2 & \pi_2' M_{21}M_{11}^{-1}M_{12}\Pi_{21} & 0 \\ \Pi_{21}' M_{21}M_{11}^{-1}M_{12}\pi_2 & \Pi_{21}' M_{21}M_{11}^{-1}M_{12}\Pi_{21} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \pi_2'\pi_2 + 1 & \pi_2'\Pi_{21} & 0 \\ \Pi_{21}'\pi_2 & \Pi_{21}'\Pi_{21} + I_{n_1} & 0 \\ 0 & 0 & I_{n_2} \end{bmatrix}.$$

Using Corollary 3.1, we now obtain

$$\hat{\sigma}^2 = \left(1, -\hat{\beta}_1', -\hat{\beta}_2'\right)\left(T^{-1}Y'Q_{Z_1}Y\right)\begin{pmatrix} 1 \\ -\hat{\beta}_1 \\ -\hat{\beta}_2 \end{pmatrix}$$

$$\Rightarrow (1, -\beta_1', -\bar{r}_2') \begin{bmatrix} \pi_2'\pi_2 + 1 & \pi_2'\Pi_{21} & 0 \\ \Pi_{21}'\pi_2 & \Pi_{21}'\Pi_{21} + I_{n_1} & 0 \\ 0 & 0 & I_{n_2} \end{bmatrix} \begin{pmatrix} 1 \\ -\beta_1 \\ -\bar{r}_2 \end{pmatrix}$$

$$= 1 + \beta_1'\beta_1 + \bar{r}_2'\bar{r}_2.$$

Proof of Lemma 4.2

Rotating Y_2 by the matrix

$$S = [\overset{n_1}{S_1}, \overset{n_2}{S_2}] \in O(n),$$

we have

$$S'Y_2'(P_H - P_{Z_1})Y_2S = \begin{bmatrix} Y_{21}' \\ Y_{22}' \end{bmatrix}(P_H - P_{Z_1})[Y_{21} \quad Y_{22}]$$

$$= \begin{bmatrix} Y_{21}'D_1D_1'Y_{21} & Y_{21}'D_1D_1'Y_{22} \\ Y_{22}'D_1D_1'Y_{21} & Y_{22}'D_1D_1'Y_{22} \end{bmatrix},$$

where $D_1 = Q_{Z_1}Z_3(Z_3'Q_{Z_1}Z_3)^{-1/2}$. Write the partitioned inverse of $S'Y_2'(P_H - P_{Z_1})Y_2S$ in the form

$$S'\left(Y_2'(P_H - P_{Z_1})Y_2\right)^{-1}S = \begin{bmatrix} l_{11} & l_{21}' \\ l_{21} & l_{22} \end{bmatrix},$$

with

$$l_{11} = \left[Y_{21}'D_1D_1'Y_{21} - Y_{21}'D_1D_1'Y_{22}(Y_{22}'D_1D_1'Y_{22})^{-1}Y_{22}'D_1D_1'Y_{21}\right]^{-1},$$

$$l_{21} = -(Y_{22}'D_1D_1'Y_{22})^{-1}Y_{22}'D_1D_1'Y_{22}l_{11},$$

$$l_{22} = (Y_{22}'D_1D_1'Y_{22})^{-1} + (Y_{22}'D_1D_1'Y_{22})^{-1}Y_{22}'D_1D_1'Y_{22}l_{11}$$

$$\times Y_{21}'D_1D_1'Y_{22}(Y_{22}'D_1D_1'Y_{22})^{-1}.$$

Now we study the asymptotic behavior of $Y_{21}'D_1D_1'Y_{21}$, $Y_{22}'D_1D_1'Y_{22}$, and $Y_{21}'D_1D_1'Y_{22}$. Write $Y_{21}'D_1D_1'Y_{21}$ as follows:

$$\begin{aligned} Y_{21}'D_1D_1'Y_{21} &= (Z_2\Pi_{21} + V_{21})'D_1D_1'(Z_2\Pi_{21} + V_{21}) \\ &= \Pi_{21}'Z_2'D_1D_1'Z_2\Pi_{21} + \Pi_{21}'Z_2D_1D_1'V_{21} \\ &\quad + V_{21}'D_1D_1'Z_2\pi_2 + V_{21}'D_1D_1'V_{21}. \end{aligned}$$

Since

$$T^{-1/2}D_1'Z_2 \to M_{33\cdot1}^{-1/2}M_{32\cdot1} = I$$

and

$$D_1'V_{21} \Rightarrow N(0, I_{K_3n_1}),$$

we have

$$T^{-1}Y_{21}'D_1D_1'Y_{21} \underset{p}{\to} \Pi_{21}'M_{32\cdot1}'M_{33\cdot1}^{-1}M_{32\cdot1}\Pi_{21} = \Pi_{21}'\Pi_{21}. \tag{A.7}$$

Next, we find

$$Y_{22}'D_1D_1'Y_{22} = V_{22}'D_1D_1'V_{22} \Rightarrow W_{n_2}(k_3, I_{n_2}). \tag{A.8}$$

Lastly, writing

$$Y_{21}'D_1D_1'Y_{22} = (Z_2\Pi_{21} + V_{21})'D_1D_1'V_{22}$$
$$= \Pi_{21}'Z_2'D_1D_1'V_{22} + V_{21}'D_1D_1'V_{22},$$

we obtain

$$T^{-1/2}Y_{21}'D_1D_1'Y_{22} \Rightarrow \Pi_{21}'M_{32\cdot 1}'M_{33\cdot 1}^{-1/2}N(0, I_{k_3n_2})$$
$$= \Pi_{21}'N(0, I_{k_3n_2}). \tag{A.9}$$

Using (A.7), (A.8), and (A.9), we obtain the following results:

$$Tl_{11} = \left[T^{-1}Y_{21}'D_1D_1'Y_{21} - T^{-1/2}Y_{21}'D_1D_1'Y_{22}(Y_{22}'D_1D_1'Y_{22})^{-1} \times T^{-1/2}Y_{22}'D_1D_1'Y_{22}\right]^{-1}$$
$$\Rightarrow \left\{\Pi_{21}'\left[I - N(0, I_{k_3n_2})\left(W_{n_2}(k_3, I_{n_2})\right)^{-1}N(0, I_{k_3n_2})\Pi_{21}\right\}.$$

Or, equivalently,

$$Tl_{11} \Rightarrow \left\{\Pi_{21}'\left[I - \delta(\delta'\delta)^{-1}\delta'\right]\Pi_{21}\right\}^{-1}$$
$$= \{\Pi_{21}'\Theta\Theta'\Pi_{21}\}^{-1}$$
$$= \bar{l}_{11}, \quad \text{say}. \tag{A.10}$$

Here δ and Θ are as in the proof of Theorem 3.1(a).

$$T^{1/2}l_{21} = -(Y_{22}'D_1D_1'Y_{22})^{-1}T^{-1/2}Y_{22}'D_1D_1'Y_{21}Tl_{11}$$
$$\to -\left[W_{n_2}(K_3, I_{n_2})^{-1}N(0, I_{K_3n_2})\Pi_{21}\bar{l}_{11}\right.$$
$$\equiv \int_{\bar{B}>0} N(0, \bar{B}^2 \otimes \Pi_{21}')\bar{l}_{11}\,\mathrm{pdf}(\bar{B})\,\mathrm{d}\bar{B}$$
$$= \int_{V_{k_3-n_2,k_3}} \int_{\bar{B}>0} N\left(0, \bar{B}^2 \otimes \bar{l}_{11}\Pi_{21}'\Pi_{21}\bar{l}_{11}\right)\mathrm{pdf}(\bar{B},\Theta)\,\mathrm{d}\bar{B}\,\mathrm{d}\Theta, \tag{A.11}$$

where $\bar{B} \equiv W_{n_2}^{-1}(n_2 + k_3 + 1, I_{n_2})$ and pdf$(\bar{B}, \Theta)$ denotes the joint pdf of $\bar{B}$ and Θ.

$$\begin{aligned} l_{22} &= (Y_{22}' D_1 D_1' Y_{22})^{-1} + (Y_{22}' D_1 D_1' Y_{22})^{-1} T^{-1/2} Y_{22}' D_1 D_1' Y_{21} T l_{11} \\ &\Rightarrow \bar{B} + \bar{B} N(0, I_{k_3 n_2}) \Pi_{21} \bar{l}_{11} \Pi_{21}' N(0, I_{k_3 n_2}) \bar{B} \\ &= \bar{B} + \bar{B}\left[\int_{V_{k_3 - n_2, k_3}} W_{n_2}\left(n_1, \bar{l}_{11}^{1/2} \Pi_{21}' \Pi_{21} \bar{l}_{11}^{1/2}\right) \mathrm{d}\Theta \right] \bar{B} \\ &= \bar{l}_{22}, \quad \text{say.} \end{aligned} \tag{A.12}$$

We deduce from (A.10), (A.11), and (A.12) that

$$\begin{aligned} \left[Y_2'(P_H - P_{Z_1}) Y_2 \right]^{-1} &\Rightarrow S \begin{bmatrix} 0 & 0 \\ 0 & \bar{l}_{22} \end{bmatrix} S' \\ &= S_2 \bar{l}_{22} S_2'. \end{aligned}$$

Proof of Lemma 4.3

This is based on Lemma 2.7 of Phillips (1989). Let $P_H = D^* D^{*\prime}$, $D^{*\prime} Y_{21} = \bar{Y}_{21}$, $D^{*\prime} Y_{22} = \bar{Y}_{22}$, and $D^{*\prime} V_{22} = \bar{V}_{22}$. Then $Z_1' Q Z_1$, with $Q = P_H - P_H Y_2 (Y_2' P_H Y_2)^{-1} Y_2' P_H$, may be decomposed as follows [see eq. (A.6) of Phillips (1989)]:

$$\begin{aligned} Z_1' Q Z_1 &= Z_1' D^* \left[I - D^{*\prime} Y_2 (Y_2' D^* D^{*\prime} Y_2)^{-1} Y_2' D^* \right] D^{*\prime} Z_1 \\ &= Z_1' D^* \left[Q_{\bar{Y}_{22}} - Q_{\bar{Y}_{22}} \bar{Y}_{21} f_{11} \bar{Y}_{21} Q_{\bar{Y}_{22}} \right] D^{*\prime} Z_1, \end{aligned}$$

where $f_{11} = (\bar{Y}_{11}' Q_{\bar{Y}_{22}} \bar{Y}_{21})^{-1}$.

Now proceeding as in the proof of Lemma 2.7 of Phillips (1989), we have

$$\begin{aligned} T^{-1/2} \bar{Y}_{22} &= T^{-1/2} D^{*\prime} Z_1 \Pi_{12} + T^{-1/2} D^{*\prime} V_{22} \\ &= F_T + T^{-1/2} D^{*\prime} V_{22}, \quad \text{say,} \end{aligned}$$

and

$$F_T' Q_{\bar{Y}_{22}} F_T = T^{-1} \bar{V}_{22} Q_{F_T} \bar{V}_{22} + O_p(T^{-3/2}).$$

Now write $T^{-1/2}D^{*\prime}Z_1 = F_T\Pi_{12}^{+} + O_p(T^{-1/2})$, where $\Pi_{12}^{+} = (\Pi_{12}'\Pi_{12})^{+}\Pi_{12}'$ is the Moore–Penrose inverse of Π_{12}. We have

$$\begin{aligned} Z_1'QZ_1 &= T\Big\{\Pi_{12}^{+\prime}F_T'\Big[Q_{\bar{Y}_{22}} - Q_{\bar{Y}_{22}}\bar{Y}_{21}f_{11}\bar{Y}_{21}'Q_{\bar{Y}_{22}}\Big]F_T\Pi_{12}^{+}\Big\} \\ &= T\Pi_{12}^{+\prime}\Big\{T^{-1}\bar{V}_{22}'Q_{F_T}\bar{V}_{22} - T^{-1}\bar{V}_{22}'Q_{F_T}Y_{21}f_{11}\bar{Y}_{21}'Q_{F_T}\bar{V}_{22} \\ &\quad + O_p(T^{-3/2})\Big\}\Pi_{12}^{+} \\ &= \Pi_{12}^{+\prime}\bar{V}_{22}'\Big[Q_{F_T} - Q_{F_T}\bar{Y}_{21}f_{11}\bar{Y}_{21}'Q_{F_T}\Big]\bar{V}_{22}\Pi_{12}^{+} + O_p(T^{-1/2}) \\ &= \Pi_{12}^{+\prime}\bar{V}_{22}'\Big[Q_{F_T} - Q_{F_T}T^{-1/2}\bar{Y}_{21}Tf_{11}T^{-1/2}\bar{Y}_{21}'Q_{F_T}\Big]\bar{V}_{22}\Pi_{12}^{+} + o_p(1). \end{aligned}$$

Writing $\bar{V}_{22} \equiv N(0, I_{(k_1+k_3)n_2}) = \xi$, we find

$$\begin{aligned} Z_1'QZ_1 &\Rightarrow \Pi_{12}^{+\prime}\xi'\big[Q_F - Q_F\lambda\bar{f}_{11}\lambda'Q_F\big]\xi\Pi_{12}^{+} \\ &\equiv \Pi_{12}^{+\prime}\big(W_{n_2}(k_1+k_3-n, I)\big)\Pi_{12}^{+}, \end{aligned}$$

where

$$F = \begin{bmatrix} 0 \\ M_{11}^{1/2}\Pi_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ \Pi_{12} \end{bmatrix},$$

$$\lambda = \begin{bmatrix} M_{33\cdot1}^{-1/2}M_{32\cdot1}\Pi_{21} \\ M_{11}^{-1/2}\Pi_{11} + M_{11}^{-1/2}M_{12}\Pi_{21} \end{bmatrix} = \begin{bmatrix} \Pi_{21} \\ \Pi_{11} \end{bmatrix},$$

$$\bar{f}_{11} = (\lambda Q_F\lambda')^{-1}.$$

The last line is obtained as in Lemma 2.7 of Phillips (1989) using the fact that $\operatorname{tr}\{Q_F - Q_F\lambda\bar{f}_{11}\lambda'W_F\} = k_1 + k_3 - n$.

Proof of Theorem 4.4

(a) Since $\hat{\beta} = S_1\hat{\beta}_1 + S_2\hat{\beta}_2$, we find under the null that $H_1\hat{\beta} - h_1 = H_1S_1(\hat{\beta}_1 - \beta_1) + H_1S_2(\hat{\beta}_2 - \beta_2)$. Using Corollary 3.1, we obtain

$$H_1\hat{\beta} - h_1 \Rightarrow H_1S_2(\bar{r}_2 - \beta_2).$$

Combining this with Lemma 4.1 and Lemma 4.2 yields the result as required.

(b) Under the null $H_2\hat{\gamma} - h_2 = H_2R_1(\hat{\gamma}_1 - \gamma_1) + H_2R_2(\hat{\gamma}_2 - \gamma_2)$. Appealing to Corollary 3.2 gives

$$H_2\hat{\gamma} - h_2 \Rightarrow H_2R_2(\bar{s}_2 - \gamma_2).$$

Now applying Lemma 4.1 and Lemma 4.3, we obtain the result directly.

Proof of Lemma 4.5

Observe that

$$T^{-1/2}\bar{Y}_{21} \underset{p}{\to} \Psi \quad \text{and} \quad \bar{Y}_{22} \Rightarrow N(0, I) = \xi, \quad \text{say},$$

where

$$\Psi = \begin{bmatrix} M_{33\cdot 1}^{1/2}M_{32\cdot 1}\Pi_{21} \\ M_{11}^{1/2}M_{12}\Pi_{21} \end{bmatrix} = \begin{bmatrix} \Pi_{21} \\ 0 \end{bmatrix}.$$

Proceeding as in the proof of Lemma 4.3, we find that

$$T^{-1}Z_1'QZ_1 \Rightarrow \rho\left[Q_\xi - Q_\xi\Psi(\Psi'Q_\xi\Psi)^{-1}\Psi'Q_\xi\right]\rho'$$

$$= \rho\Phi\left[I - \Phi'\Psi(\Psi'\Phi\Phi'\Psi)^{-1}\Psi'\Phi\right]\Phi'\rho',$$

as required.

Proof of Theorem 4.6

This is an easy consequence of Corollary 3.2 and Lemma 4.4.

References

Fisher, F.M., 1966, The identification problem in econometrics (McGraw-Hill, New York, NY).

Hausman, J.A., 1983, Specification and estimation of simultaneous equation models, Ch. 7 in: M.D. Intriligator and Z. Griliches, eds., Handbook of econometrics (North-Holland, Amsterdam).

Hsiao, C., 1983, Identification, Ch. 4 in: M.D. Intriligator and Z. Griliches, eds., Handbook of econometrics (North-Holland, Amsterdam).

Muirhead, R., 1982, Aspects of multivariate statistical theory (Wiley, New York, NY).

Phillips, P.C.B., 1980, The exact finite sample density of instrumental variable estimators in an equation with $n + 1$ endogenous variables, Econometrica 48, 861–868.

Phillips, P.C.B., 1983, Exact small sample theory in the simultaneous equations model, Ch. 8 in: M.D. Intriligator and Z. Griliches, eds., Handbook of econometrics (North-Holland, Amsterdam).

Phillips, P.C.B., 1984a, The exact distribution of LIML: I, International Economic Review 25, 249–261.

Phillips, P.C.B., 1984b, The exact distribution of exogenous variable coefficient estimators, Journal of Econometrics 26, 387–398.

Phillips, P.C.B., 1985, The exact distribution of LIML: II, International Economic Review 26, 21–36.

Phillips, P.C.B., 1988, Structural estimation under partial identification (problem and solution), Econometric Theory 4, 172–173.

Phillips, P.C.B., 1989, Partially identified econometric models, Econometric Theory 5, 181–240.

Sims, C.A., 1980, Macroeconomics and reality, Econometrica 48, 1–43.