Schmidt R A, Bjork R A 1992 New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science* **3**: 207–17.

Smith E, Ford J K, Kozlowski S 1997 Building adaptive expertise: Implications for training design. In: Quinones M A, Ehrenstein A (eds.) *Training for a Rapidly Changing Workplace: Applications of Psychological Research*. APA, Washington, DC

Volpert W, Frommann R, Munzert J 1984 Die Wirkung heuristischer Regeln im Lernprozeß. *Zeitschrift für Arbeitswissenschaft* **38**: 235–40

Von Papstein P, Frese M 1988 Transferring skills from training to the actual work situation: The role of task application knowledge, action styles, and job decision latitude. In: Soloway E, Frye D, Shepard S B (eds.) *Human Factors in Computing Systems, ACM SIGCHI Proceedings, CHI' 88*, pp. 55–60

B. Hesketh and M. Frese

Simultaneous Equation Estimates (Exact and Approximate), Distribution of

A simple example of a system of linear simultaneous equations may consist of production and consumption functions of a nation: $Y = a + bK + cL + \text{error}$, and $C = d + eY + \text{error}$. The variables Y, K, L, and C represent the gross domestic product (GDP), the capital equipment, the labor input, and the consumption, respectively.

These variables are measures of the level of economic activity of a nation. In the production function, Y increases if the inputs K and/or L increase. C increases if Y increases in the consumption equation. Each equation is modeled to explain the variation in the left-hand side 'explained' variable by the variation in the right-hand side 'explanatory' variables. Error terms are added to analyze numerically the effect of the neglected factors from the right-hand side of the equation. These equations are different from the regression equations since the 'explained' variable Y is the 'explanatory' variable in the C equation, and Y and C are simultaneously determined by the two equations. Estimation of unknown coefficients and the properties of estimation methods are not straightforward compared with the ordinary least squares estimator.

In practice, this kind of simultaneous equation system is extended to include more than 100 equations, and regularly updated to measure the economic activities of a nation. It is indispensable to analyze numerically the effect of policy changes and public investments.

In this article, the statistical model and the estimation methods of all the equations are first explained, followed by the estimation methods of a single equation and their asymptotic distributions. Explained next are the exact distributions, the asymptotic expansions, and the higher order efficiency of the estimators.

1. The System of Simultaneous Equations and Identification of the System

We write the structural form of a system consisting of G simultaneous equations as

$$y_i = Y_i \beta_i + Z_i \gamma_i + u_i \equiv (Y_i \ Z_i)\delta_i + u_i,$$
$$i = 1, \ldots, G \quad (1)$$

where y_i and Y_i are 1 and G_i subcolumns in the $T \times G$ matrix of whole endogenous variables $Y = (y_i, Y_i, Y_{ei})$, Z_i consists of K_i subcolumns in the $T \times K$ matrix of whole exogenous variables Z, β_i and γ_i are $G_i \times 1$ and $K_i \times 1$ column vectors of unknown coefficients, $\delta_i = (\beta_i', \gamma_i')'$, and u_i is the $T \times 1$ error term. This system of G equations with T observations is frequently summarized in a simple form

$$YB + Z\Gamma = U \quad (2)$$

the ith column of which is $y_i - Y_i\beta_i - Z_i\gamma_i = u_i$, i.e., Eqn. (1). The ith columns of B and Γ may be denoted as b_i and c_i where $(G - G_i - 1)$ and $(K - K_i)$ elements are zero so that $y_i - Y_i\beta_i - Z_i\gamma_i = Yb_i + Zc_i$. Zero elements are called zero restrictions. It is assumed that each row of U is independently distributed as $N(0, \Sigma)$. The reduced form of Eqn. (2) is

$$Y = Z\Pi + V \quad (3)$$

where the $K \times G$ reduced form coefficient matrix is $\Pi = -\Gamma B^{-1}$, and the $T \times G$ reduced form error term is $V = UB^{-1}$. Each row of V is assumed to be independently distributed as $N(0, \Omega)$, and then $\Sigma = B^{-1}{}'\Omega B^{-1}$.

The definition $\Pi = -\Gamma B^{-1}$, or $-\Pi B = \Gamma$ is the key to identify structural coefficients. Coefficients in β_i are identified if they can be uniquely determined by the equation $-\Pi b_i = c_i$ given Π. This equation is reduced to $-\Pi_0(1, \beta_i')' = 0$ denoting the $(K - K_i) \times (1 + G_i)$ submatrix in Π as Π_0. (Rows and columns are selected according to the zero elements in c_i and non-zero elements in b_i, respectively.) Given Π_0, this includes $(K - K_i)$ linear equations and G_i unknowns, and β_i is solvable if rank $(\Pi_0) = G_i$. This means $(K - K_i)$ must be at least G_i, or $L = K - K_i - G_i$ must be at least 0. If $L = 0$, β_i is uniquely determined. For the positive L, there are L linearly dependent rows in Π_0 since only G_i rows are necessary to determine β_i uniquely. Once β_i is determined, γ_i is determined by other K_i equations through $-\Pi b_i = c_i$. L is called the number of the degrees of overidentifiability of the ith equation.

Structural coefficients are not uniquely estimable if they are not identified (see *Statistical Identification and Estimability*).

2. Estimation Methods of the Whole System and the Asymptotic Distribution

The full information maximum likelihood (FIML) estimator of all nonzero structural coefficients δ_i, $i = 1,\ldots, G$, follows from Eqn. (3). Since it is in a linear regression form, the likelihood function can first be minimized with respect to Ω. Once Ω is replaced by the first-order condition, the likelihood function is concentrated where only B and Γ are unknown. The concentrated likelihood function is proportional to

$$\ln|\Omega_R|, \quad \Omega_R = (Y - Z\Pi_R)'(Y - Z\Pi_R)/T,$$
$$\Pi_R = -\Gamma B^{-1}, \quad (4)$$

and all zero restrictions are included in B and Γ matrices. In the FIML estimation, it is necessary to minimize $|\Omega_R|$ with respect to all non-zero structural coefficients.

The FIML estimator is consistent, and the asymptotic distribution is derived by the central limit theorem. Stacking δ_i, $i = 1,\ldots, G$ in a column vector δ, the FIML estimator $\hat{\delta}$ asymptotically approaches $N(0, -I^{-1})$ as follows:

$$\sqrt{T}(\hat{\delta} - \delta) \xrightarrow{D} N(0, -I^{-1}), \quad I = \lim_{T \to \infty} \frac{1}{T} E\left(\frac{\partial^2 \ln|\Omega_R|}{\partial\delta\,\partial\delta'}\right). \quad (5)$$

I is the limit of the average of the information matrix, i.e., $-I^{-1}$ is the asymptotic Cramer–Rao lower bound. Then the FIML estimator is the best among consistent and asymptotically normal (BCAN) estimators.

The right-hand side endogenous variable Y_i in (1) is defined by a set of G_i columns in (3) such as $Y_i = Z\Pi_i + V_i$. By the definition of V, Y_i or, equivalently, V_i is correlated with u_i since columns in U are correlated with each other. The least squares estimator applied to (1) is inconsistent because of the correlation between Y_i and u_i. Since Z is assumed to be not correlated with U in the limit, Z is used as K instruments in the instrumental variable method estimator. Premultiplying Z' to (1), it follows that

$$Z'y_i = (Z'Y_i, Z'Z_i)\delta_i + Z'u_i$$
$$= (0,\ldots, 0, Z'Y_i, Z'Z_i, 0,\ldots, 0)\delta + u_i^*,$$
$$i = 1,\ldots, G \quad (6)$$

where the $K \times 1$ transformed right-hand side variables $Z'Y_i$ is not correlated with u_i^* in the limit. Stacking all G transformed equations in a column form, the G equations are summarized as $w = X\delta + u^*$ where w and u^* stack $Z'y_i$ and u_i^*, $i = 1,\ldots, G$, respectively, and are $GK \times 1$. The covariance between u_i^* and u_j^* is $\sigma_{ij}(Z'Z)$ which is the ith row and jth column sub-block in the

covariance matrix of u^*. (The whole covariance matrix can be written as $\Sigma \otimes (Z'Z)$ where $\otimes$ signifies the Kroneker product.) Once Σ is estimated consistently (by the 2SLS method explained in the next section), δ is efficiently estimated by the generalized least squares method

$$\hat{\delta}_{3SLS} = \{X'[\hat{\Sigma}^{-1} \otimes (Z'Z)^{-1}]X\}^{-1}\{X'[\hat{\Sigma}^{-1} \otimes (Z'Z)^{-1}]w\}. \quad (7)$$

This is the three-stage least squares (3SLS) estimator by Zellner and Theil (1962). The assumption of the normal distribution error is not required in this estimation. The 3SLS estimator is consistent and is BCAN since it has the same asymptotic distribution as the FIML estimator.

3. Estimation Methods of a Single Equation and the Asymptotic Distribution

An alternative way of estimating structural coefficients is to pick up one structural equation, the ith, in a G-equation system, and estimate δ_i neglecting zero restrictions in other equations. Because of this, the other $(G-1)$ structural equations can be rewritten equivalently as $(G-1)$ reduced form equations. The limited information maximum likelihood (LIML) estimator by Anderson and Rubin (1949) applies the FIML method to a $(1 + G_i)$-equation system consisting of (1) and $Y_i = Z\Pi_i + V_i$. This means the first column and the second G_i columns of B are $(1, -\beta_i')'$ and $(0, I)'$, respectively, and the first column and the second G_i columns of Γ are $(\gamma_i', 0')'$ and Π_i, respectively. (If we denote Y as (y_i, Y_i, Y_{ei}), Y_{ei} is weakly exogenous in estimating (1) or, equivalently, the structural and reduced form parameters of (y_i, Y_i) given Y_{ei} are variation-free from the reduced form coefficient of Y_{ei}. Then Y_{ei} is omitted from (2), (3), (4), and (8).) Using these limited information B and Γ matrices, we minimize $|\hat{\Omega}_R|$ with respect to δ_i and Π_i. Defining $P_F = F(F'F)^{-1}F'$ for any full column matrix F, $G = (y_i, Y_i)'(P_Z - P_{Z_i})(y_i, Y_i)$, and $C = (y_i, Y_i)'(I - P_Z)(y_i, Y_i)$, it turns out that β_i is estimated by minimizing the least variance ratio

$$\lambda(\beta_i) = \frac{(1, -\beta_i')G(1, -\beta_i')'}{(1, -\beta_i')C(1, -\beta_i')'}. \quad (8)$$

γ_i is estimated by the least squares method applied to (1) replacing β_i with the estimator.

The two stage least squares (2SLS) estimator is the generalized least squares estimator applied to Eqn. (6) using $Z'Z$ as the weight matrix. (See Eqn. (10), where k is set to 1.) The assumption of the normal distribution of the error term is not required in this estimation. Both LIML and 2SLS estimators are consistent, and the large sample distribution is

$$\sqrt{T}(\hat{\delta}_i - \delta_i) \xrightarrow{D} N(0, -I^{-1}(\delta_i, \delta_i)) \quad (9)$$

where I is calculated similarly to Eqn. (5) but Ω_R is defined using the limited information B and Γ matrices, and only the diagonal submatrix of I^{-1} which corresponds to δ_i is used. (Partial derivatives are calculated with respect to δ_i and columns in Π_i.) This asymptotic distribution is a particular case of Eqn. (5). Both estimators are consistent and BCAN under the zero restrictions imposed on Eqn. (1).

The k-class estimator $\hat{\delta}_i(k)$ unifies both LIML and 2SLS estimators (Theil 1961). It is

$$\hat{\delta}_i(k) = \begin{pmatrix} Y_i'P_zY_i - (k-1)Y_i(I-P_z)Y_i & Y_i'Z_i \\ Z_i'Y_i & Z_i'Z_i \end{pmatrix}^{-1}$$
$$\times \begin{pmatrix} Y_i'P_z - (k-1)Y_i'(I-P_z) \\ Z_i' \end{pmatrix} y_i. \quad (10)$$

This is the least squares estimator, the 2SLS estimator, and the LIML estimator when k is 0, 1, and $1+\lambda$, respectively. There are two important properties in the k-class estimator. It is consistent if $p\lim_{T\to\infty} k = 1$, and is BCAN if $p\lim_{T\to\infty} \sqrt{T}(k-1) = 0$. If k satisfies these conditions, the k-class estimator is consistent and BCAN even when $Y_i'(I-P_z)Y_i$ and $Y_i'(I-P_z)y_i$ are replaced with any matrix and a vector of order $O_P(T)$.

4. Exact Distributions of the Single-equation Estimators

Several early studies compared the bias and mean squared errors of OLS, LIML, 2SLS, FIML, and 3SLS estimators by the Monte Carlo simulations since all but OLS estimators are consistent and are indistinguishable. The OLS estimator was often found as reliable as other consistent estimators. Later, the studies went on to t ratios, and the real defect of OLS estimators was found: the deviation from the standard normal distribution is worse than any other simultaneous equation methods. See Cragg (1967) for related papers.

Drawing general qualitative comparisons from simulations is difficult since simulations require setting values of all population parameters. Simulation studies on the small sample properties led to the derivation of the exact distributions, which were expected to permit the drawing of general comparisons without depending on the particular parameter values.

If $n \times 1$ column vectors x_t, $t = 1,\ldots,T$ are independently distributed $N(m_t, \Omega)$, the density function of $\Sigma_{t-1,T} x_t x_t'$ is the non-central Wishart matrix denoted as $W_n(T, \Omega, M)$ where the non-centrality parameter is $M = \Sigma_{t-1,T} m_t m_t'$ (Anderson 1958, Chap. 13). The study of the exact distribution of the single-equation estimators started from the fact that $G = (G_{kl})$ and $C = (C_{kl})$, k, $l = 1,\ldots,1+G_i$ in Eqn. (8)

are the noncentral Wishart matrix $W_{1+G_i}(K-K_i, \Omega, M)$ with the noncentrality parameter matrix $M = \Pi_R'Z'(P_z - P_{z_i})Z\Pi_R$, and the central Wishart matrix $W_{1+G_i}(T-K, \Omega, 0)$, respectively. The 2SLS, OLS, and LIML estimators of β_i are $G_{22}^{-1}G_{21}$, $(G_{22}+C_{22})^{-1}(G_{21}+C_{21})$, and $(G_{22}-\lambda C_{22})^{-1}(G_{21}-\lambda C_{21})$, respectively, and λ is the minimum root in the polynomial equation $|G - \lambda C| = 0$. Since all estimators are functions of elements in the G and C matrices, their distributions can be characterized by degrees of freedom of the two Wishart matrices, M and Ω matrices.

In deriving the exact density functions, the 2SLS and OLS estimators can be treated in a similar way. In the 2SLS estimator, the joint density of G is transformed into the joint density of $G_{22}^{-1}G_{21}$, $G_{11}-G_{12}G_{22}^{-1}G_{21}$, and G_{22}. Integrating out $G_{11}-G_{12}G_{22}^{-1}G_{21}$ and G_{22} results in the joint density of $G_{22}^{-1}G_{21}$. The resulting density function includes infinite terms, and zonal polynomials when G_i is greater than one. A pedagogical derivation is found in Press (1982, Chap. 5). In the LIML estimator, the joint density of G and C is transformed into that of characteristic roots and vectors. Since $\hat{\beta}$ is rewritten as a ratio of elements in a characteristic vector, the density function is derived by integrating out unnecessary random variables from the joint density function. However, the analytical operations are not easy when there are many endogenous variables. See Phillips (1983, 1985) to find comprehensive reviews on the 2SLS and LIML estimators, respectively.

It was somewhat fruitless to derive exact distributions because these include nuisance parameters and infinite terms. It was difficult to draw general conclusions on the qualitative properties of the estimators from the numerical evaluations of these distributions. See Anderson et al. (1982).

Qualitative properties of the estimators followed from the exact moments of estimators. Kinal (1980) proved that the (fixed) k-class estimator in a multiple endogenous variables case has moments up to $(T-K_i-G_i)$ if $0 \leqslant k < 1$, and up to L if $k = 1$. Mariano and Sawa (1972) proved that, in the $G_i = 1$ case, the mean and variance of the LIML estimator do not exit. (In Monte Carlo simulations, the bias of LIML estimators was often found to be smaller than that of others even though the exact mean is infinite. This showed the clear limitation of the simulation methods.)

5. Asymptotic Expansions of the Distributions of the Single-equation Estimators

Asymptotic expansion of the distribution was introduced as an analytical tool which is more accurate than the asymptotic distribution but is less complicated than the exact distributions. For instance, the

t ratio statistic, say X, which is commonly used in econometrics, has the density function $f(x) = c \cdot [1 + (x^2/m)]^{-(1+m)/2}$ where c is a constant and m is the degree of freedom under conditions including the normally distributed error terms. Since the mean and variance of X is 0 and $m/(m-2)$, the standardized statistic $Z = \sqrt{(m-2)/m} \cdot X$ has the density $f(z) = c' \cdot [1 + (z^2/(m-2))]^{-(1+m)/2}$ where c' is a new constant. This density function is expanded to the third-order term as

$$f(z) = \phi(z)\{1 + \frac{1}{4m}(z^4 - 6z^2 + 3)\} + o(m^{-1}) \quad (11)$$

where $\phi(z)$ is the standard normal density function, and the constant is adjusted so that the area under the curve is one. (Since the t distribution is symmetric around the origin, the $O(1/\sqrt{m})$ term does not appear in the right-hand side of the equation.) Rewriting in terms of X, the asymptotic expansion of the t statistic with m degrees of freedom is

$$f(x) = \phi(x) + \frac{1}{4m}(x^4 - 2x^2 - 1)\phi(x) + o(m^{-1}) \quad (12)$$

The first term in the right-hand side is the $n(0, 1)$ density function. The second term in the right-hand side converges to zero if the limit is taken with respect to m. This second term gives deviation of $f(x)$ from $\phi(x)$, and is called the third-order term. For finite m, the third-order term is expected to improve the standard normal approximation. The numerical evaluation of this expansion is easy.

The asymptotic expansions in the simultaneous equation estimators are long and include nuisance parameter matrices such as q below. See Phillips (1983) for a review and Phillips (1977) for the validity of expansion. The asymptotic expansion does not require the assumption of the normal distribution of the error term. Fujikoshi et al. (1982) gave the expansion of the joint density of the estimators of δ_i. In their study, the bias of the estimators is calculated from the asymptotic expansions as

$$AM(\sqrt{T}(\hat{\delta}_i - \delta_i)) = \frac{1}{\sqrt{T}}(dL - 1)\Gamma^{-1}q + o(T^{-1}),$$

$$q = \frac{1}{\sigma^2}\begin{pmatrix} \varpi_{21} - \Omega_{22}\beta_i \\ 0 \end{pmatrix}, \quad (13)$$

where $\hat{\delta}$ is the estimator, d is 1 and 0 fo1r 2SLS and LIML estimators, respectively, and the Ω matrix is partitioned into submatrices conformable with the partitions of G and C matrices. $AM(\cdot)$ stands for the mean operator, but uses the asymptotic expansion for the density function. (Recall that the exact mean does

not exist in the LIML estimator.) It is possible to compare the two estimators in terms of the calculated bias. For example, the bias of the 2SLS estimator is 0 when the degree of overidentifiability L is 1.

Further, the mean of the squared errors was calculated from the asymptotic expansions and used to compare estimators. It was proved that the mean squared error of the 2SLS estimator is smaller than that of the LIML estimator when L is less than or equal to 6. Historically, this kind of comparison of 'approximate' mean squared errors goes back to the 'Nagar expansion' by Nagar (1959) and the 'small disturbance expansion' by Kadane (1971). These qualitative comparisons gave researchers some guidance about the choice of estimators.

It was interesting to examine the accuracy of the approximations calculated by the asymptotic expansions of distributions. If the asymptotic expansions were accurate, calculation of the exact distributions could be avoided, and properties of estimators could be found easily from the approximations. However, the approximations were found not to be accurate enough to replace the exact distributions. There are many cases where the asymptotic expansion is accurate when the asymptotic distribution, the first term in the expansion, is already accurate; the asymptotic expansion is inaccurate when the asymptotic distribution is inaccurate. In particular, the asymptotic expansions are inaccurate when

(a) The value of asymptotic variance is small;
(b) The value of L is large; or
(c) The structural error term is highly correlated with the right-hand side endogenous variable.

It is noted, first, $Z'Z/T$ is assumed to converge a nonsingular fixed matrix in the asymptotic theory. Second, for an attempt to improve the accuracy of asymptotic distributions by incorporating large L values, see Morimune (1983). Third, the asymptotic expansions cannot trace closely the skewed exact distributions that happen particularly when the correlation is high.

6. Higher-order Efficiency of the System and the Single-equation Estimators

The asymptotic expansions of distributions can be used in comparing the probabilities of concentration of estimators about the true parameter values. One estimator is more desirable than another if its probability is greater than that of the other. This measure was used in comparing the single-equation estimators, and some qualitative results were derived.

Furthermore, the third-order efficiency criterion was brought in the comparisons. This criterion requires that estimators be adjusted to have the same asymptotic bias as in Eqn. (13). Then the adjusted estimators are compared, and the maximum likelihood

estimator is proved most efficient. It has the highest concentration about the true parameter values in terms of the asymptotic expansion of distribution to the third-order $O(1/T)$ terms. (Akahira and Takeuchi (1981), for example.) The adjusted maximum likelihood estimator has the smallest mean-squared error at the same time since the difference among estimators is found only in the mean-squared errors.

In whole system estimation, the FIML estimator is third-order efficient. The 3SLS estimator is less efficient than the FIML estimator in terms of the asymptotic probability of concentration once the bias of the two estimators is adjusted to be the same. Morimune and Sakata (1993) derived a simple adjustment of the 3SLS estimator so that the adjusted estimator has the same asymptotic expansion as the FIML estimator to the third-order $O(1/T)$ terms. This estimator is explained by modifying Eqn. (7). In Eqn. (6), Y_i is replaced by $\hat{Y}_i \equiv Z\,\hat{\Pi}_i = Z(Z'Z)^{-1} Z'Y_i$ so that the X matrix consists of $Z'\hat{Y}_i$ and $Z'Z_i$. In the modified estimator, we estimate Σ and Π by the first round 3SLS estimator and replace $\hat{Y}_i$ in X by $\tilde{Y}_i \equiv Z\tilde{\Pi}_i$ where $\tilde{\Pi}_i$ consists of proper subcolumns in $\tilde{\Pi} = -\hat{\Gamma}\hat{B}^{-1}$. The new X matrix is denoted as $\tilde{X}$. Finally, the new estimator is

$$\tilde{\delta}_{\text{M3SLS}} = \{\tilde{X}'[\tilde{\Sigma}^{-1} \otimes (Z'Z)^{-1}]X\}^{-1}\{\tilde{X}'[\tilde{\Sigma}^{-1} \otimes (Z'Z)^{-1}]w\}. \quad (14)$$

This estimator has the same asymptotic expansion as the FIML estimator to the third-order terms and is third-order efficient.

The LIML and 2SLS estimators are simple cases of the FIML and 3SLS estimators, respectively. Then the modified 2SLS estimator which follows from Eqn. (14) has the same asymptotic expansion as the LIML estimator to the third-order term. The LIML estimator and the modified 2SLS estimator are third-order efficient in the single-equation estimation.

7. Conclusion

Laurence Klein received the 1980 Nobel prize for the creation of a macroeconometric model which is an empirical form of a simultaneous equation system, and for the application to the analysis of economic fluctuations and economic policies. The macroeconometric model became a standard tool to analyze the economies and policies of nations. Trygve Haavelmo received the 1989 Nobel prize for his contribution in the analysis of simultaneous structures. Haavelmo, together with other researchers at the Cowles Commission for Research in Economics, then at the University of Chicago, became the founders of simultaneous equation analysis in econometrics. Part of their research is collected in Hood and Koopmans

(1953). Studies on the exact and approximate distributions of estimators came after the research conducted at the Cowles group, and helped to make econometrics rigorous.

Access to computers was the main concern when econometric model-building started spreading all over the world in the 1970s. Since then, computer facilities surrounding econometric model-building have changed greatly. Bulky mainframe computer have been replaced by personal computers. Computer programs were written individually, mostly in Fortran, and were used for regression analyses in model estimation as well as for simulation studies in econometrics theory. The packaged least squares programs replaced the individually written programs later in model estimation. They run on personal computers and have facilitated greatly the conducting of empirical studies.

See also: Simultaneous Equation Estimation: Overview

Bibliography

Akahira M, Takeuchi K 1981 *Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency.* Springer, New York

Anderson T W 1958 *An Introduction to Multivariate Statistical Analysis.* Wiley, New York

Anderson T W, Rubin H 1949 Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* **20**: 46–63

Anderson T W, Kunitomo N, Sawa T 1982 Evaluation of the distribution function of the limited information maximum likelihood estimator. *Econometrica* **50**: 1009–27

Basmann R L 1957 A generalized classical method of linear estimation of coefficients in a structural equation. *Econometrica* **25**: 77–83

Cragg J G 1967 On the relative small sample properties of several structural equation estimators. *Econometrica* **35**: 89–110

Fujikoshi Y, Morimune K, Kunitomo N, Taniguchi M 1982 Asymptotic expansions of the distributions of the estimates of coefficients in a simultaneous equation system. *Journal of Econometrics* **18**: 191–205

Hood W C, Koopmans T C 1953 *Studies in Econometric Methods.* Wiley, New York

Kadane J B 1971 Comparison of k-class estimates when the disturbance is small. *Econometrica* **39**: 723–37

Kinal T W 1980 The existence of moments of k-class estimators. *Econometrica* **48**: 241–9

Mariano R S, Sawa T 1972 The exact finite sample distribution of the limited information maximum likelihood estimator in the case of two included endogenous variables. *Journal of the American Statistical Association* **67**: 159–63

Morimune K 1983 Approximate distributions of k-class estimators when the degree of over-identifiability is large compared with the sample size. *Econometrica* **51**: 821–41

Morimune K, Sakata S 1993 Modified three stage-least squares estimator which is third-order efficient. *Journal of Econometrics* **57**: 257–76

Nagar A L 1959 The bias and moment matrix of the general kappa-class estimators of the parameters in simultaneous equations. *Econometrica* **27**: 575–95

Phillips P C B 1977 The general theorem in the theory of asymptotic expansions as approximation to the finite sample distributions of econometric estimators. *Econometrica* **45**: 1517–34

Phillips P C B 1983 Exact Phillips. In: Griliches Z, Intriligator M D (eds.) *Handbook of Econometrics.* North-Holland, Amsterdam, Vol. 1, Chap. 8

Phillips P C B 1985 The exact distribution of LIML: 2. *International Economic Reviews* **25**(1): 249–61

Press S J 1982 *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference* [original edn. 1972, 2nd edn.] Robert E Krieger, Malabar, FL

Theil H 1961 *Economic Forecasts and Policy*, 2nd edn. North-Holland, New York, pp. 231–2, 334–6

Zellner A, Theil H 1962 Three-stage least squares: simultaneous estimation of simultaneous equations. *Econometrica* **30**: 54–78

K. Morimune

Simultaneous Equation Estimation: Overview

Simultaneous equations are important tools for understanding behavior when two or more variables are determined by the interaction of two or more relationships in such a way that causation is joint rather than unidirectional. Such situations abound in economics, but also occur elsewhere. Haavelmo (1943, 1944) began their modern econometric treatment.

The simplest economic example is the interaction of buyers and sellers in a competitive market which jointly determines the quantity sold and price. Another important example is the interaction of workers, consumers, investors, firms, and government in determining the economy's output, employment, price level, and interest rates, as in macroeconometric forecasting models. Even when one is interested only in a single equation, it often is best interpreted as one of a system of simultaneous equations.

Strotz and Wold (1960) argued that in principle every economic action is a response to a previous action of someone else, but even they agreed that simultaneous equations are useful when the data are yearly, quarterly, or monthly, because these periods are much longer than the typical market response time.

This article discusses the essentials of simultaneous equation estimation using a simple linear example.

1. A Simple Supply–Demand Example

Let q and p stand for the quantity sold and price of a good, y and w for the income and wealth of buyers (assumed independently determined), u_1 and u_2 for unobservable random shocks, and Greek letters for

unknown constant parameters. Suppose that market equilibrium, in which the price is such that suppliers want to sell the same quantity that demanders what to buy, is described by these linear equations:

$$\text{supply:} \quad q = \gamma_1 + \beta_1 p + u_1, \quad \beta_1 > 0 \tag{1}$$

$$\text{demand:} \quad q = \gamma_2 + \delta_2 y + \varepsilon_2 w + \beta_2 p + u_2, \quad \beta_2 < 0 \tag{2}$$

Neither equation alone can determine either p or q, because each equation contains both: these are *simultaneous equations*. Solve them for p and q thus:

$$p = \pi_{11} + \pi_{12} y + \pi_{13} w + (u_2 - u_1)/\Delta \tag{3}$$

$$q = \pi_{21} + \pi_{22} y + \pi_{23} w + (\beta_1 u_2 - \beta_2 u_1)/\Delta \tag{4}$$

where

$$\pi_{11} = (\gamma_2 - \gamma_1)/\Delta, \quad \pi_{12} = \delta_2/\Delta, \quad \pi_{13} = \varepsilon_2/\Delta \tag{5}$$

$$\pi_{21} = (\beta_1 \gamma_2 - \beta_2 \gamma_1)/\Delta, \quad \pi_{22} = \beta_1 \delta_2/\Delta,$$

$$\pi_{23} = \beta_1 \varepsilon_2/\Delta \tag{6}$$

and

$$\Delta = \beta_1 - \beta_2 \tag{7}$$

2. Types of Variables: Structural and Reduced Form Equations

The variables p and q which are to be explained by the model are *endogenous*. Equations (1) and (2) are structural equations. Each of them describes the behavior of one of the building blocks of the model, and (as is typical) contains more than one endogenous variable. Equations (3) and (4) are reduced form equations. Each of them contains just one endogenous variable, and determines its value as a function of parameters, shocks, and explanatory variables (here y and w).

The explanatory variables are predetermined if the shocks for any given period are independent of the explanatory variables for that period and all previous periods; they are *exogenous* if the shocks for each period are independent of the explanatory variables for every period. Thus all exogenous variables are predetermined, but not conversely. For example, the value of an endogenous variable from a previous period cannot be exogenous, but it is predetermined if the shocks are serially independent.

3. The Need for Estimation of Parameters

One job of a reduced form equation is to tell how an endogenous variable responds to a change in any predetermined or exogenous variable. Typically, no single structural equation can do this. Another job of

14106