

On T. W. Anderson's Contributions to the Study of Structural Equation Estimation

Peter C. B. Phillips

Cowles Foundation for Research in Economics, Yale University

1. Introduction

Much of T. W. Anderson's work in statistical theory has had an important influence on econometrics. But it is his work on structural estimation that has had the greatest impact on subsequent developments in the subject. In part this is because his research came at a time in the early postwar years of the late 1940s when econometrics was ripe for new developments; and in part it is because his early papers gave some definitive answers to the estimation and inferential problems that faced empirical researchers who wanted to use the new Keynesian models of macroeconomic behavior. Anderson's work in the 1940s on structural estimation was conducted jointly with Herman Rubin, and their two articles [8, 11] published in *The Annals of Mathematical Statistics* constituted a comprehensive study of single equation estimation under a Gaussian likelihood, covering derivation of the estimates, computational procedures, an asymptotic theory of inference and some aspects of small-sample behavior. Only the latter left major problems for future study.

Anderson began a second phase of research on the topic of single equation structural estimation in the 1970s, more than 20 years after the original investigations. In a series of papers [66, 67, 69, 73, 83, 97, 100, 101, 104, 105] that were published from 1973 to 1985, Anderson and his students and co-workers studied the small-sample properties and performance of single equation structural estimation techniques. These investigations continued the program of research that was begun in the 1940s by providing a corpus of knowledge about the relative small sample performance of competing single equation estimation methods.

Received July 11, 1988.

AMS subject classifications: 62H10, 62H12, 62P20.

Key words and phrases: 2SLS, asymptotic expansions, econometrics, finite sample distributions, limited information maximum likelihood, LIML, LIMLK, maximum likelihood, simultaneous equations in econometrics, simultaneous equation models, single-equation estimators, structural equations, two-stage least squares, TSLS.

2. The Early Work on Limited Information Structural Estimation

Consider the linear structural equation

$$y_1 = Y_2\beta + Z_1\gamma + u, \quad (1)$$

where y_1 ($T \times 1$) and Y_2 ($T \times n$) are an observation vector and observation matrix, respectively, of $n+1$ jointly dependent (or endogenous) variables, Z_1 is a $T \times K_1$ matrix of predetermined (or exogenous) variables and u is a random disturbance vector. The reduced form of (1) is written

$$(y_1, Y_2) = (Z_1, Z_2) \begin{pmatrix} \pi_1 & \Pi_1 \\ \pi_2 & \Pi_2 \end{pmatrix} + (v_1, V_2) \quad (2)$$

or

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

where Z_2 is a $T \times K_2$ matrix of predetermined variables that are known to be excluded from (1). It is assumed that the rows of V are iid($\mathbf{0}, \Omega$) with $\Omega > \mathbf{0}$ (i.e., positive definite) and for the purpose of maximum likelihood estimation the common distribution is taken to be Gaussian. Equations (1) and (2) are connected by the system

$$\pi_1 - \Pi_1\beta = \gamma \quad (3)$$

$$\pi_2 - \Pi_2\beta = \mathbf{0}. \quad (4)$$

These equations are called the identifiability relations because, when $\text{rank}(\Pi_2) = n \leq K_2$, the vector β is uniquely determined by (4) from knowledge of (π_2, Π_2) and γ is then determined by (3).

The first paper [8] by Anderson and Rubin explored the structural estimation of (1) by maximum likelihood. The method became known as limited information maximum likelihood (or LIML), since only the structural information on (1), which here amounts to the exclusion of the variables Z_2 , is accounted for in maximizing the likelihood. It was also known as the Cowles Commission method—see, for example, Malinvaud (1980, pp. 698–702)—since the research was conducted at the Cowles Commission in Chicago. Anderson and Rubin showed that the LIML estimator of β satisfies the generalized eigenvector equation

$$(W - \hat{\lambda}S)\hat{\mathbf{b}} = \mathbf{0}, \quad (5)$$

where $\hat{\mathbf{b}}' = (1, -\hat{\beta}')$ and $\hat{\lambda}$ is the smallest eigenvalue of the matrix $S^{-1}W$, where W and S are the residual moment matrices

$$W = Y'(P_Z - P_{Z_1})Y, \quad S = Y'(I - P_Z)Y \quad (6)$$

and P_A represents the orthogonal projection onto the range space of the matrix A . The estimator $\hat{\beta}$ minimizes the ratio

$$\mathbf{b}'\mathbf{W}\mathbf{b}/\mathbf{b}'\mathbf{S}\mathbf{b} \quad (7)$$

and thus also became known as the least variance ratio (LVR) estimator— see [8, p. 52] and Koopmans and Hood (1953). The corresponding LIML estimator of $\boldsymbol{\gamma}$ is obtained from $\hat{\boldsymbol{\beta}}$ and (3) as

$$\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\pi}}_1 - \hat{\boldsymbol{\Pi}}_1\hat{\boldsymbol{\beta}} = (\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1(\mathbf{y}_1 - \mathbf{Y}_2\hat{\boldsymbol{\beta}})$$

using the maximum likelihood estimates of $\boldsymbol{\pi}_1$ and $\boldsymbol{\Pi}_1$.

In [8] Anderson and Rubin derived the LIML estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$, outlined a procedure for their computation and derived the likelihood ratio test

$$l = (1 + \hat{\lambda})^{-T/2} \quad (8)$$

of the hypothesis that the structural equation (1) is overidentified [that is, there are more than enough exclusion restrictions on (1) to identify its coefficients]. They also found small-sample confidence regions for the structural coefficient vector $\boldsymbol{\beta}$ under Gaussian assumptions and gave an approximate small-sample F test of the identifiability hypothesis which could be used in place of (8).

The companion paper [11] was concerned with asymptotics. They showed the consistency of LIML, found its limit distribution to be normal, gave formulas for the asymptotic covariance matrix, and demonstrated that the likelihood-ratio test statistic (8) gives rise to an asymptotic χ^2 test in the usual way, namely,

$$-2 \ln l = T \ln(1 + \hat{\lambda}) \xrightarrow{D} \chi^2_{K_2 - n},$$

where the degrees of freedom $K_2 - n$ represent the degree of overidentification of the structural equation (1). These are the main results of [11] and they have been heavily used in subsequent research.

Some other aspects of [11] are, however, less well known:

(i) Stable autoregressive specifications in (1) and (2) were permitted, so that some components of the predetermined variables could be lagged dependent variables.

(ii) Nonlinearities were permitted in the other structural equations of the system, leading to a nonlinear reduced form in place of (2).

(iii) In determining the limit behavior of the LIML estimator $\hat{\boldsymbol{\beta}}$, Anderson and Rubin showed that in large samples $\hat{\boldsymbol{\beta}}$ is effectively determined by the equation

$$\mathbf{W}\hat{\boldsymbol{\beta}} = \mathbf{0}$$

in place of (5). This leads directly to the expression

$$\tilde{\boldsymbol{\beta}} = [\mathbf{Y}'_2(\mathbf{P}_Z - \mathbf{P}_{Z_1})\mathbf{Y}_2]^{-1}[\mathbf{Y}'_2(\mathbf{P}_Z - \mathbf{P}_{Z_1})\mathbf{y}_1] \quad (9)$$

for what is now called the two-stage least squares (2SLS) estimator of $\boldsymbol{\beta}$. While Anderson and Rubin did not give the formula (9) explicitly, it is clear that they were aware of its

existence: Indeed, as Anderson himself explained it in a recent interview published in *Econometric Theory* (Phillips, 1986):

Now, in proving the asymptotic normal distribution of the estimators in the second of the two papers with Rubin, the first step in proving this was to show that the contribution of the smallest root times the estimate of the covariance matrix drops out. So in doing this we reduced the estimator to what's now called the two stage least squares estimator. So the two stage least squares estimator is actually in that second paper.

Of course, the 2SLS estimator $\tilde{\beta}$ was subsequently discovered independently by Basman (1957) and Theil (1958). It was further shown by Theil (1958) that $\hat{\beta}$ and $\tilde{\beta}$ are both members of a general class of estimators known as the k -class.

3. Later Work on Small-Sample Properties

The second phase of Anderson's research on single equation structural estimation began in 1973 with the publication of [66]. This was a joint article with T. Sawa and it gave exact finite-sample distributions and asymptotic expansions (both under Gaussian assumptions) for the k -class estimator of β in (1) and for the special case of two variables ($n = 1$). Finite-sample investigations of this type were already underway, and the field had been active in econometrics since the early 1960s, with contributions by Nagar (1959), Basman (1961), Bergstrom (1962), Richardson (1968), Sawa (1969), and Sargan and Mikhail (1971). Somewhat earlier than [66] Mariano and Sawa (1972) had found the exact density of the LIML estimator $\hat{\beta}$, again for the special case of $n = 1$. It was not until a decade after the publication of [66] that the exact finite-sample distribution of the LIML was found for the general case of $n+1$ endogenous variables (Phillips, 1984, 1985). The exact distribution of the 2SLS estimator $\tilde{\beta}$ in the general case was found in Phillips (1980a). In one sense, this completed the research program begun in [8] and [11].

All of these studies just cited are essentially mathematical exercises. But they are all motivated by the important concern of learning more about the small-sample behavior of competing structural equation estimators such as $\hat{\beta}$ and $\tilde{\beta}$. To satisfy this goal Anderson and his co-workers in the series of papers cited in the Introduction developed asymptotic expansions of the distributions of various estimators and performed numerical calculations to calibrate their accuracy. As far as possible, the expansions were used to shed light on properties such as distributional location and concentration about the true coefficient value. The main conclusion to emerge from this body of work is that, at least in the two variable case, $\tilde{\beta}$ is a better general purpose estimator of β than $\hat{\beta}$ in finite samples. This conclusion is supported by evidence which indicates that the distribution of $\tilde{\beta}$ is well located about β , nearly symmetric and is well approximated by the asymptotic distribution. On the other hand, the distribution of $\hat{\beta}$ is poorly located, skewed and less well approximated by asymptotic theory.

Since the 2SLS estimator $\tilde{\beta}$ inherits many of the features (including the general distributional form) of the ordinary least squares (OLS) estimator of β in (1), the poor performance of this estimator is not so surprising. Inevitably, the behavior of $\tilde{\beta}$ in finite samples is contaminated by properties of OLS, which is known to be asymptotically biased. What is more surprising and what has so far not been well explained is the remarkably good performance of LIML. After all, maximum likelihood is well known

to produce estimators that are seriously biased and poorly approximated by asymptotic results in other settings. In fact, if LIML is applied to (1) when \mathbf{Y}_2 is an integrated process it has been shown (Phillips, 1988) that even the limiting distribution of $\hat{\beta}$ is biased and skewed. The good performance of LIML in the present setting should therefore be explicable solely in terms of the characteristics of the model (1) and (2). Three factors seem important. The first is that $\hat{\beta}$, unlike $\tilde{\beta}$, is invariant to the normalization of the structural equation (1). This is a point that was subsequently emphasized by Anderson in the interview (Phillips, 1986, p. 258). It has also been investigated by Hillier (1988). The second factor is that, in the present case, neglecting the remaining equations of the system results only in a potential efficiency loss for LIML. That this is not always the case for LIML procedures is demonstrated in Phillips (1988). Finally, it can be argued that $\hat{\beta}$ in minimizing the ratio (7) mimics a property of the true coefficient β , namely, that $T^{-1}\mathbf{b}'\mathbf{W}\mathbf{b}$ and $T^{-1}\mathbf{b}'\mathbf{S}\mathbf{b}$ are both unbiased estimators of the equation error variance in (1), for which value the ratio (7) achieves its minimum of unity.

A second topic in this area that is now ripe for further research is structural estimation in time series models. In particular, the finite sample properties of structural estimators such as LIML and 2SLS need to be explored in systems where the predetermined variables include lagged dependent variables. Some steps in this direction have been taken using Edgeworth expansions to shed light on the finite-sample distributions (see Phillips, 1980b). But the area is wide open for further development, and the need for more evidence is substantial since most empirical structural models fall into the dynamic not the static model category.

The above remarks are intended to show that the research agenda inspired by the original papers of Anderson and Rubin [8, 11] is still active. Structural models have always been a central field of econometric research and in that field the contributions of T. W. Anderson will continue to occupy a position of special significance.

References to Publications by T. W. Anderson

- Anderson, T. W. (1974). An asymptotic expansion of the distribution of the limited information maximum likelihood estimate of a coefficient in a simultaneous equation system. *J. Amer. Statist. Assoc.* **69** 565–573; **71** 1010 (1976). (Paper [67] in these *Collected Papers*, pp. 995–1003.)
- Anderson, T. W. (1976). Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics (with discussion). *J. Roy. Statist. Soc. Ser. B* **38** 1–36. (Paper [69] in these *Collected Papers*, pp. 1027–1062.)
- Anderson, T. W. (1977). Asymptotic expansions of the distributions of estimates in simultaneous equations for alternative parameter sequences. *Econometrica* **45** 509–518. (Paper [73] in these *Collected Papers*, pp. 1087–1096.)
- Anderson, T. W. (1982). Some recent developments on the distributions of single-equation estimators. *Advances in Econometrics* (W. Hildenbrand, ed.), Cambridge University Press, 109–122. (Paper [101] in these *Collected Papers*, pp. 1409–1422.)
- Anderson, T. W., Kunitomo, N., and Sawa, T. (1982). Evaluation of the distribution function of the limited information maximum likelihood estimator. *Econometrica* **50** 1009–1027. (Paper [100] in these *Collected Papers*, pp. 1389–1407.)
- Anderson, T. W., Kunitomo, N., and Sawa, T. (1983). Comparison of the densities of the TSLS and LIMLK estimators for simultaneous equations. *Global Econometrics: Essays in Honor of*

- Lawrence R. Klein (F. G. Adams and B. G. Hickman, eds.), MIT Press, Cambridge, Mass., 103–124. (Paper [105] in these *Collected Papers*, pp. 1465–1486.)
- Anderson, T. W., Morimune, K. A., and Sawa, T. (1983). The numerical values of some key parameters in econometric models. *J. Econometrics* **21** 229–243. (Paper [104] in these *Collected Papers*, pp. 1449–1463.)
- Anderson, T. W. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Statist.* **20** 46–63. (Paper [8] in these *Collected Papers*, pp. 99–116.)
- Anderson, T. W. and Rubin, H. (1950). The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Statist.* **21** 570–582. (Paper [11] in these *Collected Papers*, pp. 153–165.)
- Anderson, T. W. and Sawa, T. (1973). Distributions of estimates of coefficients of a single equation in a simultaneous system and their asymptotic expansions. *Econometrica* **41** 683–714. (Paper [66] in these *Collected Papers*, pp. 963–994.)
- Anderson, T. W. and Sawa, T. (1979). Evaluation of the distribution function of the two-stage least squares estimate. *Econometrica*, **47** 163–182. (Paper [83] in these *Collected Papers*, pp. 1177–1196.)
- Anderson, T. W. and Sawa, T. (1982). Exact and approximate distributions of the maximum likelihood estimator of a slope coefficient. *J. Roy. Statist. Soc. Ser. B* **44** 52–62. (Paper [97] in these *Collected Papers*, pp. 1331–1341.)
- Phillips, P. C. B. (1986). The ET interview: Professor T. W. Anderson interviewed by Peter C. B. Phillips. *Econometric Theory* **2** 249–288.

Other References

- Basmann, R. L. (1957). A generalized classical method of linear estimation of coefficients in a structural equation. *Econometrica* **25** 77–83.
- Basmann, R. L. (1961). Note on the exact finite sample frequency functions of generalized classical linear estimators in two leading overidentified cases. *J. Amer. Statist. Assoc.* **56** 619–636.
- Bergstrom, A. R. (1962). The exact sampling distributions of least squares and maximum likelihood estimators of the marginal propensity to consume. *Econometrica* **30** 480–490.
- Hillier, G. H. (1988). On the interpretation of exact results for structural equation estimators. Discussion Paper No. 6/88, Monash University.
- Koopmans, T. C. and Hood, W. (1953). The estimation of simultaneous linear economic relationships. *Studies in Econometric Method* (T. C. Koopmans and W. Hood, eds.), Yale University Press, New Haven, Connecticut, 112–199.
- Malinvaud, E. (1980). *Statistical Methods of Econometrics*. North-Holland, Amsterdam.
- Mariano, R. S. and Sawa, T. (1972). The exact finite-sample distribution of the limited information maximum likelihood estimator in the case of two included endogenous variables. *J. Amer. Statist. Assoc.* **67** 159–163.
- Nagar, A. L. (1959). The bias and moment matrix of the general k -class estimators of the parameters in structural equations. *Econometrica* **27** 575–595.
- Phillips, P. C. B. (1980a). The exact finite sample density of instrumental variable estimators in an equation with $n+1$ endogenous variables. *Econometrica* **48** 861–878.
- Phillips, P. C. B. (1980b). Finite sample theory and the distributions of alternative estimators of the marginal propensity to consume. *Rev. Econom. Stud.* **47** 183–224.
- Phillips, P. C. B. (1984). The exact distribution of LIML: I. *Internat. Econom. Rev.* **25** 249–261.

- Phillips, P. C. B. (1985). The exact distribution of LIML: II. *Internat. Econom. Rev.* **26** 21–36.
- Phillips, P. C. B. (1988). Optimal inference in cointegrated systems. Cowles Foundation Discussion Paper No. 866, Yale University. To appear in *Econometrica*.
- Richardson, D. H. (1968). The exact distribution of a structural coefficient estimator. *J. Amer. Statist. Assoc.* **62** 1214–1226.
- Sargan, J. D. and Mikhail, M. W. (1971). A general approximation to the distribution of instrumental variable estimates. *Econometrica* **39** 131–169.
- Sawa, T. (1969). The exact finite sampling distribution of ordinary least squares and two stage least squares estimator. *J. Amer. Statist. Assoc.* **64** 923–936.
- Theil, H. (1958). *Economic Forecasts and Policy*. North-Holland, Amsterdam.