6-2022

An Econometrician amongst Statisticians: T. W. Anderson

Peter C. B. Phillips

AN ECONOMETRICIAN AMONGST STATISTICIANS:
T. W. ANDERSON

By

Peter C. B. Phillips

June 2022

COWLES FOUNDATION DISCUSSION PAPER NO. 2333

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

An Econometrician amongst Statisticians: T. W. Anderson *

Peter C. B. Phillips †

Yale University, University of Auckland, Singapore Management University & University of Southampton

June 2, 2022

Abstract

T. W. Anderson did pathbreaking work in econometrics during his remarkable career as an eminent statistician. His primary contributions to econometrics are reviewed here, including his early research on estimation and inference in simultaneous equations models and reduced rank regression. Some of his later works that connect in important ways to econometrics are also briefly covered, including limit theory in explosive autoregression, asymptotic expansions, and exact distribution theory for econometric estimators. The research is considered in the light of its influence on subsequent and ongoing developments in econometrics, notably confidence interval construction under weak instruments and inference in mildly explosive regressions.

Keywords: Asymptotic expansions, Confidence interval construction, Explosive autoregression, LIML, Reduced rank regression, Simultaneous equation models, Weak identification.

AMS 2020 Subject Classifications: 01, 41, 62, 91

Throughout his long career as a statistician Ted Anderson, as he was universally known, had a professional influence that extended beyond statistics into neighboring disciplines where statistical methods were in heavy use and new methods were often needed. For econometrics and psychometrics especially, he forged tools that were suited to the particular models and data being used in those fields. These broad interests began in his undergraduate years at Northwestern where he majored in mathematics, minored in economics, and followed courses in psychology as well as econometrics and statistics. Ted's early decisions as a student played an important role in his subsequent work as a researcher and his continuing links with those

*This paper draws on some material covered in earlier interviews and tributes, including (Phillips, 1986b; DeGroot, 1986; Phillips, 2017) to which readers are referred.

†Support is acknowledged from the NSF under Grant No. SES-18 50860, and a Kelly Fellowship at the University of Auckland.

fields throughout his career.[1] His landmark advanced texts in multivariate analysis and time series (Anderson, 1959a, 1970), together with their subsequent editions, educated generations of students and researchers in those sister disciplines as well as statistics. In addition to his strong pedagogical influence on students and researchers in econometrics through these advanced texts, Ted was a familiar figure and regular keynote speaker at econometric conferences worldwide; and he maintained strong interests in econometric methods until the end of his life, his final published paper appearing in an econometrics journal (Anderson, 2017).

This contribution focuses on Ted Anderson's main contributions to econometrics, which began with his appointment as a research fellow in the Cowles Commission for Research in Economics at Chicago in 1945. Amongst econometricians, the Cowles researchers of the 1940s occupy a special position of seniority because their research opened up promising new fields of study in structural modeling, identification, continuous time processes, explosive time series and causality, all of which continue to resonate in the discipline. Ted Anderson was a leading figure in the Cowles econometrics group during this seminal period.

Limited Information Maximum Likelihood, its origins and longevity

When Ted joined in 1945, the Cowles Commission had commenced a major study developing econometric methodology to address the intrinsic joint dependence of much economic data. The first step in this work involved resolving the problem of identifying the structural parameters that linked the endogenous variables in the equations that represented key variables in the system. In linear systems, identification was resolved by the use of sufficient prior restrictions from economic theory on the variables that entered each equation. Gaussian maximum likelihood, or full information maximum likelihood (FIML) as it became known, was the natural method to use in estimating the parameters of the entire system but it relied on full system identification and presented what were at that time major computational obstacles for practical work. Ted worked with Herman Rubin – another young recruit to Cowles – in undertaking the entirely new task of using maximum likelihood methods at the single equation level, for which the required computations were considerably reduced.[2] Their resulting publications (Anderson and Rubin, 1949, 1950) took a major step forward in structural equation estimation. The methodology they developed liberated empirical researchers to conduct valid estimation and inference about the parameters of a potentially large system of equations on an equation by equation basis with a method that was not substantially more complex than linear regression. This method became known as Limited Information Maximum Likelihood (LIML). Not simply an advance at the time to facilitate computation, LIML proved to be a game changer, delivering properties at the time and since that have made it an attractive approach for both estimation and inference.

LIML concentrates on the estimation of a single structural equation $y_1 = Y_2\beta + X_1\gamma + u$

[1] Some of Ted's personal reminiscences on his training and career are given in his published interviews in *Econometric Theory* and *Statistical Science*, vide Phillips (1986b) and DeGroot (1986).

[2] Ted describes his early work at the Cowles Commission in Anderson (2015).

within a complete system of n observations of $m+1$ endogenous variables written in observation form as $Y = [y_1, Y_2]$ with reduced form $Y = [X_1, X_2]\Pi + [v_1, V_2] = X\Pi + V$, where $X = [X_1, X_2]$ is an $n \times K (= K_1 + K_2)$ matrix of exogenous variables of which the K_1 variables X_1 are included in the structural equation of interest and X_2 variables are excluded, which assures identification of the parameters in the structural equation when $K_2 \geq m$. The error matrix V has n rows that are assumed to be independent and normally distributed. The LIML estimator $(\hat{\beta}', \hat{\gamma}')'$ is the Gaussian maximum likelihood estimator of $(\beta', \gamma')'$ in this complete system for which the structural equations for Y_2 are unspecified and instead given only in reduced form. The LIML estimator $\hat{\beta}$ is obtained by solving the generalized eigenvector equation $(W - \hat{\lambda}S)\hat{b} = 0$ for $\hat{b}' = (1, -\hat{\beta}')$ and $\hat{\lambda}$, the smallest eigenvalue of the matrix $S^{-1}W$, where W and S are the residual moment matrices $W = Y'(P_X - P_{X_1})Y$ and $S = Y'(I - P_X)Y$ where P_A is the orthogonal projection onto the range of the matrix A. The LIML vector $\hat{\beta}$ minimizes the ratio $b'Wb/b'Sb$ and so became known as the least variance ratio estimator. When only a few endogenous variables are present in the structural equation, the computations required to find $\hat{\beta}$ were within reach of a hand calculator at the time, thereby leading to its early practical implementation even in large simultaneous equation systems.[3]

In addition to the development of the LIML procedure, Anderson and Rubin (1949) provided a likelihood ratio test based on the statistic $\ell = (1 + \hat{\lambda})^{-n/2}$ of the hypothesis that the structural equation was overidentified, viz., that $K_2 > m$, ensuring that there were more than enough exclusion restrictions in the equation to identify its structural coefficient m-vector β. The second paper (Anderson and Rubin, 1950) was concerned with asymptotic theory. The authors proved the consistency of LIML, established its limiting normal distribution, gave formulae for the asymptotic variance matrix and demonstrated that the likelihood ratio statistic delivered the usual asymptotic χ^2 test, so that $-2\log(\ell) = n\log(1 + \hat{\lambda}) \rightsquigarrow \chi^2_{K_2 - m}$ as $n \to \infty$. The treatment in the paper was general enough to include stable autoregressive specifications in the formulation of the structural system so that predetermined variables were permitted; and nonlinearities were allowed in the other structural equations so that the reduced form could be nonlinear. Further, in determining the limit distribution of LIML they demonstrated that in large samples the estimator $\hat{\beta}$ is essentially the same as the two stage least squares (2SLS) estimator $\tilde{\beta} = [Y_2'QY_2]^{-1}[Y_2'Qy_1]$ where $Q = P_Z - P_{Z_1}$. This anticipated the later discovery of 2SLS (Basmann, 1957), the asymptotic equivalence of the two estimators, and the emergence of more general instrumental variable (IV) procedures in econometrics (Sargan, 1958).

LIML is known to work very well in practice and there has been much subsequent investigation of its properties. First, extensive simulations have confirmed that the LIML estimator $\hat{\beta}$ is a better general purpose estimator than the 2SLS estimator $\tilde{\beta}$ in finite samples: specifically, LIML is better centered about the true value with a nearly symmetric distribution, close to median unbiased, and better approximated by its asymptotic distribution. These properties

[3] In one of the first large scale empirical applications of LIML estimation Bergstrom (1955) used the method to estimate the parameters of a 19 equation macroeconomic model of the demand and supply for New Zealand's exports.

apply even though it is known that LIML has no finite integer moments and heavier tails than 2SLS in general with an exact distribution whose leading term is Cauchy, a property that holds also for FIML, vide (Phillips, 1984, 1986a). Second, the differences that favor the LIML estimator are most striking when the degree of equation overidentification is large and when there is a high correlation between the endogenous regressor and the structural equation error. Anderson (1982) provides a detailed account of this work. These conclusions match the findings of more recent research in which it has been discovered that LIML clearly dominates IV estimation procedures such as 2SLS in high dimensional cases where the degree of overidentification $K_2 = K_{2,n} \to \infty$ as $n \to \infty$.

Third, the dominance of LIML is especially marked when the instruments provided by the excluded exogenous variables are weak. The analysis of weak instrumentation has received considerable attention following work on unidentified models (Phillips, 1989) and on weakly identified systems (Staiger and Stock, 1997). A primary import of this research is that under certain general conditions weak instrument asymptotics reproduce a version of the exact finite sample theory under Gaussian distributional assumptions, thereby invalidating the usual asymptotic theory of estimation and inference. In such cases, of course, both LIML and 2SLS are inconsistent. However, the situation changes when the number of instruments $K_2 = K_{2,n} \to \infty$ as $n \to \infty$. Research by Chao and Swanson (2005) initiated the analysis of structural regression estimation with an asymptotically infinite number of weak instruments, showing that in such conditions, LIML recovers consistency but that 2SLS remains inconsistent unless bias correction measures are employed. Subsequent research has pursued this line of investigation to generalized method of moments (GMM) estimation with many moment conditions, which includes a nonlinear model extension of LIML to the continuously updated GMM and the analysis of its properties (Han and Phillips, 2006; Newey and Windmeijer, 2009).

All of these features of LIML and particularly its advantages under conditions of weak instrumentation with large numbers of instruments have contributed to its longevity as a method of estimation in simultaneous equations systems. But there was a separate finite sample contribution in Anderson and Rubin (1949) that was concerned with inference that has had the most significant subsequent influence both in theoretical work and empirical application.

The Anderson-Rubin method and its unexpectedly long aftermath

In addition to developing the LIML estimation procedure, Anderson and Rubin (1949) developed a finite sample theory of interval estimation and a statistic for testing the null hypothesis $\mathcal{H}_0 : \beta = \beta_0$. The procedures apply when the confidence region or null hypothesis concerns the full vector β of endogenous variable coefficients in the structural equation. The statistic took the form of the ratio

$$\mathrm{AR}(\beta_0) = \frac{(y_1 - Y_2\beta_0)'(P_X - P_{X_1})(y_1 - Y_2\beta_0)/K_2}{(y_1 - Y_2\beta_0)'Q_X(y_1 - Y_2\beta_0)/(n-K)}, \quad Q_X = I_n - P_X \tag{1}$$

of standardized quadratic forms in the structural equation difference $y^* := y_1 - Y_2\beta_0$. When $\mathcal{H}_0$ is true $y^* := X_1\gamma + u$ and when $\mathcal{H}_0$ is false y^* is a linear function of both X_1 and X_2. So $\mathcal{H}_0$ is tested by a conventional F test of the hypothesis that the coefficient vector of X_2 is zero in the regression of y^* on X_1 and X_2, leading directly to (1). When the structural equation errors $u \sim_d \mathcal{N}(0, \sigma^2 I_n)$, the test statistic $\mathrm{AR}(\beta_0) = \frac{u'(P_X - P_{X_1})u/K_2}{u'Q_X u/(n-K)}$ has an exact $F_{K_2,n-K}$ distribution with respective numerator and denominator degrees of freedom K_2 and $n-K$. When $\beta \neq \beta_0$, the distribution is noncentral $F_{K_2,n-K}$. $\mathrm{AR}(\beta_0)$ may be viewed as a simple structural equation extension of the original Fisher regression F test. The asymptotic distribution as $n \to \infty$ is $1/K_2$ times a $\chi^2_{K_2}$ distribution. The AR test statistic is pivotal, asymptotically pivotal, and consistent; and valid confidence regions are constructed in the usual manner. These properties are appealing and have ensured a long reaching influence in applications.

The procedures have two main shortcomings: (i) they apply only to the full vector of endogenous variable coefficients; and (ii) the degrees of freedom of the limit distribution equal the degree of overidentification, so when the number of instruments strongly exceeds the number of structural coefficients, which is common in empirical work, the test has low power. These issues have attracted considerable attention in the recent years. An important step forward in resolving (ii) was made by Kleibergen (2002) who suggested a slight modification to the projection geometry of the AR statistic retaining its asymptotic pivotal properties but with a limiting χ^2_m distribution that has degrees of freedom equal to the number of structural parameters. This statistic therefore has an asymptotic distribution with a minimal number of degrees of freedom and is largely unhampered by poor power performance that may be induced by points of underidentification or large numbers of instruments. Research on (i) concerned with the development of subvector tests has proved more demanding but is of greater importance in empirical work. Robust inference, particularly weak-instrument robust inference, is a hard problem because the structural coefficients not under test become nuisance parameters that present an obstacle to pivotal limit theory. Typically it is difficult to control size without rendering the test too conservative (Guggenberger et al., 2012) but the use of data-dependent critical values that adapt to the strength of the instruments has proved promising in controlling size and raising test power (Guggenberger et al., 2019).

In addition to testing, Anderson and Rubin (1949) suggested inverting the test statistic to obtain confidence regions for β_0. This is a procedure that is now widely employed throughout econometrics. It is particularly useful in the case of tests that are robust to weak identification, a situation where there is low correlation between the endogenous regressor variables (viz., Y_2 in the above system) and the instruments used for Y_2 in the regression (viz. X_2). Problems of weak instrumentation arise frequently in empirical economic applications (Andrews et al., 2019) and typically lead to failure in the bootstrap, which treats low correlations as non-zero even when the correlations are zero asymptotically. The AR confidence regions also suffer difficulties, as they may be unbounded or disconnected, reflecting the fact that the instruments may supply very limited information about the structural coefficients. To mitigate these difficulties several alternative approaches have been suggested and analyzed in the econometric literature including conditional likelihood ratio tests (Moreira, 2003), partially restricted reduced form methods Phillips and Gao (2017), Lagrange multiplier tests and modifications to the AR test (Kleibergen, 2002). Overviews of some of these procedures and performance evaluations are given in Mikusheva (2010); Andrews et al. (2019). Much of this recent work and ongoing progress in the 21st century speaks to the long-reaching influence of the many contributions in the two original articles by Anderson and Rubin (1949, 1950) that were written under the auspices of the Cowles Commission more than seven decades ago.

Reduced rank regression and its later link to cointegration

In his doctoral dissertation on the noncentral Wishart distribution at Princeton (Anderson, 1945) Ted considered applications to the problems of estimation and inference in the presence of linear restrictions on regression coefficients. A full treatment was later provided in Anderson (1951). This work studied multivariate regressions of the form $y_t = A_1 x_{1t} + A_2 x_{2t} + u_t$ where the errors $u_t \sim_a iid\mathcal{N}(0, \Sigma)$ and the $m \times k (= k_1 + k_2)$ partitioned coefficient matrix $A = [A_1, A_2]$ has an $m \times k_2$ submatrix A_2 of deficient rank r with an outer product form $A_2 = \alpha\beta'$ where α is $m \times r$, β is $r \times k_2$ and both α and β are of full rank. If $\alpha_\perp$ is an $m \times (m - r)$ orthogonal complement of α then $\alpha'_\perp y_t = \alpha'_\perp A x_t + u_t = \alpha'_\perp A_1 x_{1t} + \alpha'_\perp u_t$, revealing the restricted system. Gaussian maximum likelihood estimation gives rise to a canonical correlation problem solved by finding the latent roots $\{\hat{\lambda}_i\}_{i=1}^m$ and eigenvectors $\{\hat{v}_i\}_{i=1}^m$ of the determinantal equation $|\lambda S_{yy} - S_{yx} S_{xx}^{-1} S_{xy}| = 0$, where $S_{ab} = \frac{1}{n} \sum_{t=1}^n a_t b_t'$, leading to estimates of the restricted regression coefficients and a likelihood ratio test of the restrictions. The framework and the asymptotic theory associated was applicable under general conditions to stationary vector autoregressions (VARs).

Subsequent work by Johansen (1988) nearly four decades later showed that the same framework could be used to analyze cointegrated VAR systems of nonstationary time series. In this case the model is a simple VAR written in what is known in the econometrics literature as the 'error correction' form $\Delta z_t = A z_{t-1} + \sum_{j=1}^p \Delta z_{t-j} + u_t$, which characterizes the adjustment process towards a 'long run equilibrium' relation. That relationship is embodied in the leading $m \times m$ coefficient matrix $A = \alpha\beta'$ written in outer product form with reduced rank $r < m$ so that the r relations $\beta' z_{t-1}$ are the stationary errors about equilibrium, giving a 'cointegrating' linkage among the nonstationary variables z_{t-1} in the previous time period to which the system variables dynamically adjust by correcting these stationary errors. The generating mechanism of this system involves two components: (i) the $m - r$ dimensional unit root system $\alpha'_\perp \Delta z_t = \sum_{j=1}^p \alpha'_\perp \Delta z_{t-j} + \alpha'_\perp u_t$; and (ii) a complementary r dimensional system $\beta' z_t = (I_r + \beta'\alpha)\beta' z_{t-1} + \sum_{j=1}^p \beta' \Delta z_{t-j} + \beta' u_t$ that is stationary when the leading coefficient matrix $R = I_r + \beta'\alpha$ has stable roots. In (ii) the $r \times m$ matrix β' describes the cointegrating space of the system giving the linear linkages among the system variables that reduce the nonstationary components of the multivariate times series z_t to a vector of stationary time series $\beta' z_t$. Some further algebraic manipulations with (i) and (ii) lead to the partial sum and moving average representation $z_t = C \sum_{s=1}^t u_s + \alpha(\beta'\alpha)^{-1} R(L) \beta' u_t + C z_0$, where $C = \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \alpha'_\perp$,

$\beta_\perp$ is an $m \times (m-r)$ matrix orthogonal complement of β and $R(L) = \sum_{i=0}^{\infty} R^i L^i$. This system reveals the nonstationary component $C\sum_{s=1}^{t} u_s$ (with reduced rank coefficient matrix C) and the stationary component $\alpha(\beta'\alpha)^{-1}R(L)\beta' u_t$ (with reduced rank moving average operator $\alpha(\beta'\alpha)^{-1}R(L)$) of the data z_t and is called the Granger-Johansen representation (Johansen, 1995). In view of the nonstationary elements in the generating mechanism of z_t, the asymptotic theory of estimation and inference for this cointegrated system required functional central limit theory, weak convergence and stochastic integral methods that had been developed for multivariate regressions involving time series with some unit roots (so-called integrated processes) in (Phillips, 1986c; Phillips and Durlauf, 1986; Phillips, 1988; Chan and Wei, 1988).

In economics the cointegrating transform matrix $\beta' z_t$ provides a time series representation of the concept of a long run equilibrium linking the variables of z_t. This strong connection with economic theory coupled with the methods for deriving asymptotic theory for nonstationary regressions opened up a wide arena of new research in time series econometrics. The resulting methodology has spawned a vast literature of applications throughout economics and more widely across the social and business sciences. Anderson (2002) revisited the reduced rank regression model, related some of the subsequent developments in econometrics to his earlier framework in Anderson (1951), and compared the limit theory of the reduced rank estimator with that of least squares. An interesting further connection to Ted's research is that the LIML estimator in the simulataneous equations model may itself be derived as a reduced rank regression estimator when taking into account the full system including the reduced form and attendant restrictions associated with the structural equation of interest.

Explosive autoregression, invariance principles and asset bubble detection

Anderson (1959b) broke new ground in the analysis of stationary autoregression, showing that only a finite variance (not a higher order moment condition) was required for the usual asymptotic distribution theory for the least squares (LS) estimator. In the same article, building on the earlier findings of White (1958), Ted studied the explosive autoregression $x_t = \theta x_{t-1} + u_t$ with $u_t \sim_{iid} (0, \sigma^2)$ and $\theta > 1$ showing, under normally distributed innovations and a zero initial condition $x_0 = 0$, that: (i) the asymptotic distribution of the LS estimator $\hat{\theta} = \sum_{t=1}^{n} x_t x_{t-1} / \sum_{t=1}^{n} x_{t-1}^2$ is Cauchy; (ii) the self normalized and centred statistic $(\sum_{t=1}^{n} x_{t-1}^2)^{1/2}(\hat{\theta} - \theta) \rightsquigarrow \mathcal{N}(0, \sigma^2)$ as $n \to \infty$; and (iii) in the general case where the innovations are not normal the asymptotic distribution of $\hat{\theta}$ depends on the distribution of the innovations and the initial condition. Importantly, in all these cases, no central limit theory applies. Ted's paper also considered the explosive VAR model $x_t = Ax_{t-1} + u_t$ with $u_t \sim_{iid} \mathcal{N}(0, \Sigma)$, Σ positive definite, and all latent roots λ_i of the $m \times m$ coefficient matrix explosive with $\lambda_i > 1$. The results for this VAR model mirrored those of the scalar case provided the explosive roots were all distinct. Again, no invariance principle applied. In addition, Ted noted that when there were some common explosive roots with $\lambda_i = \lambda_j$ for some $i \neq j$ or when there were stable roots, as well as explosive roots, these results no longer applied and that this case 'would be much more involved'.

The absence of an invariance principle in explosive autoregression asymptotics presented a barrier to the use of these models in practical work. This obstacle was removed in Phillips and Magdalinos (2007) where it was shown that martingale central limit theory (MGCLT) delivers invariance principle results for an explosive autoregression whose coefficient $\theta = 1 + c/n^{\gamma} > 1$, where $\gamma \in (0, 1)$ and $c > 0$ are rate and scale constants. The time series is then 'mildly explosive' in the sense that θ is near unity on the explosive side but more distant from unity than local to unity departures for which $\theta = 1 + c/n$. (Phillips, 1987; Chan and Wei, 1987). Use of the MGCLT gives precisely the same limit theory for this mildly integrated case as that of the purely explosive case and holds also for weakly dependent innovations u_t. These results opened up a large literature in econometrics concerned with the detection of financial and real estate bubbles, where mildly explosive behavior in asset prices is a characteristic of asset prices during the expansive phase of a bubble in which prices depart from economic fundamentals (Phillips and Yu, 2011; Phillips et al., 2015). The methods that have been developed allow for real time detection of emergent asset bubbles and are used by banks and monetary authorities to assess ongoing financial market conditions.

Dynamic panel regression

In the early 1980's two jointly authored papers (Anderson and Hsiao, 1981, 1982) by Ted and his former student Cheng Hsiao opened up new possibilities in dynamic panel modeling methodology. By this time such models had evolved into a considerable subfield of applied research in economics. Longitudinal datasets were increasingly available and in many cases the panels were short and wide, with a cross section sample size N that vastly dominated the time series sample size n, which often numbered in the single digits. Regressions in these dynamic panels were subject when n was small to very considerable autoregressive estimation bias problems that translated into inconsistencies when the asymptotics were driven solely as $N \to \infty$ (Nickell, 1981). The Anderson-Hsiao papers had an immediate and lasting impact demonstrating that, with use of a simple choice of a suitable lagged endogenous instrumental variable, serious bias problems in stationary dynamic panel models could be eliminated. This idea became the foundation stone of what quickly grew into a cathedral of methodology based on generalized method of moment estimation of dynamic panels, creating in the process a sub-discipline in itself with its own graduate courses and textbooks, complete with tentacles of application that stretched across many different applied sciences. It was quickly realized that the Anderson-Hsiao estimators were inefficient because they relied on a single lagged instrument. Research extending their approach and progressive searches for a complete set of instruments led to high dimensional instrumental variable and GMM approaches that were asymptotically efficient under certain conditions. Subsequent research on the asymptotic theory of these general approaches included nonstationary data cases (Hahn and Kuersteiner, 2002; Phillips and Sul, 2007) and has provided a foundation for statistical inference in applied work. Publication of the influential advanced text by Hsiao (2022), first in 1986 and now in its third edition, provided an accessible general reference to these methods and the wide scope of empirical applications they encompassed.

Finite sample theory

During the 1970s Ted renewed his analytic research on structural equation estimation, focusing on the small sample properties of LIML and related estimators such as 2SLS. This line of research was initiated a few years earlier by Sargan and Mikhail (1971) who developed an asymptotic expansion of the instrumental variable estimator in a general single structural equation. In a series of papers published over the period 1973-1985 and aided by his students and former students, Ted derived the exact distributions and asymptotic expansions of some of these single equation structural estimation techniques in various special cases under Gaussian assumptions. This work commenced with the asymptotic expansion of the LIML estimator (Anderson, 1974) in an equation with two endogenous variables and continued with several comparative studies involving both exact theory and expansions for similar special cases (Anderson, 1976, 1977, 1982; Anderson et al., 1982; Morimune, 1983). The exact finite sample distribution of LIML was obtained in the general case of a structural equation with $m + 1$ endogenous variables in Phillips (1985). Since LIML, 2SLS, and IV estimators all depend on quadratic functions of the data, the exact distribution of each of these estimators can be viewed as the distribution of a nonlinear function of a matrix that has a noncentral Wishart distribution. This formulation of the distribution theory links the problem closely to Ted's original thesis research at Princeton that was published in Anderson (1946). The behavior of these exact distributions and asymptotic expansions (and hence that of the corresponding estimators) therefore all depends on the noncentrality parameter matrix, which is itself a quadratic form in the reduced form parameter matrix and is thereby affected by the strenth of the instruments.

It is now known that the exact distributions and their findings in terms of the relative performance of the estimation procedures are relevant under much more general asymptotic conditions that do not rely on Gaussian innovations when the instrumental variables are only weakly identifying. The reason for this widened generality is that, whereas the innovations themselves are not Gaussian, standardized linear combinations of them do satisfy martingale central limit theorems, which implies that in the limit as the sample size $n \to \infty$, the estimators and test statistics typically behave as if the data were Gaussian. So in the limit the same functional dependence on quadratic forms of a non-central Wishart distribution applies, thereby enriching the implications of the findings.

Econometric Legacy

The broad sweep of Ted's research has influenced many different areas of econometrics, from structural equation estimation and inference, to commonly used time series and dynamic panel methods, from asymptotic theory to finite sample theory, from linear reduced rank regression to cointegration, and from Gaussian limit theory to the surprising simplicity of matrix Cauchy limit distributions in explosive autoregressions. Perhaps Ted's primary and most enduring legacy to econometrics lies in his seminal contributions to simultaneous equations estimation, the development of the LIML estimator and the Anderson-Rubin inferential procedure that has found such a wide range of applications in modern partially identified systems. Ted himself extolled the virtues of LIML and its associated procedures in Anderson (2015). It is a fitting memorial to his intellectual contributions to econometrics that his last published paper (Anderson, 2017) delivered further desirable properties of the LIML likelihood ratio test in terms of its best invariance and admissability, thereby closing the circle of his research career back to his original work in the Cowles Commission.

References

Anderson, T. W. (1945): "The non-central Wishart distribution and its application to problems of multivariate statistics," *Ph.D Thesis, Princeton University Library.*

——— (1946): "The non-central Wishart distribution and certain problems of multivariate statistics," *The Annals of Mathematical Statistics*, 409–431.

——— (1951): "Estimating linear restrictions on regression coefficients for multivariate normal distributions," *The Annals of Mathematical Statistics*, 327–351.

——— (1959a): *An introduction to multivariate statistical analysis.*

——— (1959b): "On asymptotic distributions of estimates of parameters of stochastic difference equations," *The Annals of Mathematical Statistics*, 676–687.

——— (1970): *The statistical analysis of time series*, Wiley New York.

——— (1974): "An asymptotic expansion of the distribution of the limited information maximum likelihood estimate of a coefficient in a simultaneous equation system," *Journal of the American Statistical Association*, 69, 565–573.

——— (1976): "Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometrics," *Journal of the Royal Statistical Society: Series B (Methodological)*, 38, 1–20.

——— (1977): "Asymptotic expansions of the distributions of estimates in simultaneous equations for alternative parameter sequences," *Econometrica*, 509–518.

——— (1982): "Some recent developments on the distributions of single-equation estimators," in *Advances in Econometrics*, ed. by W. Hildebrand, Cambridge University Press, 1409–1422.

——— (2002): "Reduced rank regression in cointegrated models," *Journal of Econometrics*, 106, 203–216.

——— (2015): "My Reminiscences of Trygve Haavelmo at the Cowles Commission," *Econometric Theory*, 31, 180–192.

——— (2017): "Admissible Significance Tests in Simultaneous Equation Models," *Econometric Theory*, 33, 534–550.

Anderson, T. W. and C. Hsiao (1981): "Estimation of dynamic models with error components," *Journal of the American statistical Association*, 76, 598–606.

——— (1982): "Formulation and estimation of dynamic models using panel data," *Journal of Econometrics*, 18, 47–82.

Anderson, T. W., N. Kunitomo, and T. Sawa (1982): "Evaluation of the distribution function of the limited information maximum likelihood estimator," *Econometrica*, 1009–1027.

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

——— (1950): "The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations," *The Annals of Mathematical Statistics*, 570–582.

Andrews, I., J. H. Stock, and L. Sun (2019): "Weak instruments in instrumental variables regression: Theory and practice," *Annual Review of Economics*, 11, 727–753.

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

Bergstrom, A. R. (1955): "An econometric study of supply and demand for New Zealand's exports," *Econometrica*, 258–276.

Chan, N. H. and C. Z. Wei (1987): "Asymptotic inference for nearly nonstationary AR(1) processes," *The Annals of Statistics*, 1050–1063.

——— (1988): "Limiting distributions of least squares estimates of unstable autoregressive processes," *The Annals of Statistics*, 367–401.

Chao, J. C. and N. R. Swanson (2005): "Consistent estimation with a large number of weak instruments," *Econometrica*, 73, 1673–1692.

DeGroot, M. H. (1986): "A conversation with T W Anderson," *Statistical Science*, 1, 97–105.

Guggenberger, P., F. Kleibergen, and S. Mavroeidis (2019): "A more powerful subvector Anderson Rubin test in linear instrumental variables regression," *Quantitative Economics*, 10, 487–526.

Guggenberger, P., F. Kleibergen, S. Mavroeidis, and L. Chen (2012): "On the asymptotic sizes of subset Anderson–Rubin and Lagrange multiplier tests in linear instrumental variables regression," *Econometrica*, 80, 2649–2666.

Hahn, J. and G. Kuersteiner (2002): "Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large," *Econometrica*, 70, 1639–1657.

Han, C. and P. C. B. Phillips (2006): "GMM with many moment conditions," *Econometrica*, 74, 147–192.

Hsiao, C. (2022): *Analysis of panel data*, Cambridge University Press.

Johansen, S. (1988): "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

——— (1995): *Likelihood-based inference in cointegrated vector autoregressive models*, Oxford University Press.

Kleibergen, F. (2002): "Pivotal statistics for testing structural parameters in instrumental variables regression," *Econometrica*, 70, 1781–1803.

Mikusheva, A. (2010): "Robust confidence sets in the presence of weak instruments," *Journal of Econometrics*, 157, 236–247.

Moreira, M. J. (2003): "A conditional likelihood ratio test for structural models," *Econometrica*, 71, 1027–1048.

Morimune, K. (1983): "Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size," *Econometrica*, 821–841.

Newey, W. K. and F. Windmeijer (2009): "Generalized method of moments with many weak moment conditions," *Econometrica*, 77, 687–719.

Nickell, S. (1981): "Biases in dynamic models with fixed effects," *Econometrica*, 1417–1426.

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

——— (1985): "The exact distribution of LIML: II," *International Economic Review*, 26, 21–36.

——— (1986a): "The distribution of FIML in the leading case," *International Economic Review*, 239–243.

——— (1986b): "The ET Interview: Professor T. W. Anderson," *Econometric Theory*, 2, 249–288.

——— (1986c): "Understanding spurious regressions in econometrics," *Journal of Econometrics*, 33, 311–340.

——— (1987): "Time series regression with a unit root," *Econometrica*, 55, 277–301.

——— (1988): "Multiple regression with integrated time series," *Contemporary Mathematics*, 80, 79–106.

——— (1989): "Partially identified econometric models," *Econometric Theory*, 5, 181–240.

——— (2017): "Tribute to T. W. Anderson," *Econometric Theory*, 33, 529–533.

Phillips, P. C. B. and S. N. Durlauf (1986): "Multiple time series regression with integrated processes," *The Review of Economic Studies*, 53, 473–495.

Phillips, P. C. B. and W. Y. Gao (2017): "Structural inference from reduced forms with many instruments," *Journal of Econometrics*, 199, 96–116.

Phillips, P. C. B. and T. Magdalinos (2007): "Limit theory for moderate deviations from a unit root," *Journal of Econometrics*, 136, 115–130.

Phillips, P. C. B., S. Shi, and J. Yu (2015): "Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500," *International Economic Review*, 56, 1043–1078.

Phillips, P. C. B. and D. Sul (2007): "Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence," *Journal of Econometrics*, 137, 162–188.

Phillips, P. C. B. and J. Yu (2011): "Dating the timeline of financial bubbles during the subprime crisis," *Quantitative Economics*, 2, 455–491.

Sargan, J. D. (1958): "The estimation of economic relationships using instrumental variables," *Econometrica*, 393–415.

Sargan, J. D. and W. M. Mikhail (1971): "A general approximation to the distribution of instrumental variables estimates," *Econometrica*, 131–169.

Staiger, D. and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

White, J. S. (1958): "The limiting distribution of the serial correlation coefficient in the explosive case," *The Annals of Mathematical Statistics*, 1188–1197.