finite sample theory in econometrics

Economic models, which provide relationships between economic variables, are useful in making scientific predictions and policy evaluations. Well-known examples include classical linear regression models, where the explanatory variables are assumed to be non-stochastic (fixed) and the errors are normally distributed, and non-classical models, where these assumptions are violated. These non-classical models are frequently used in empirical work, and they include the simultaneous equations model, models with serial correlation and heteroscedasticity, limited dependent-variables models, panel and spatial models, non-linear models, and models with non-normal errors.

Based on sample data, econometric methods provide techniques of estimation and hypothesis testing related to these and other models. The commonly used estimators are the least squares (LS) or the generalized LS (GLS), the maximum likelihood (ML), the generalized method of moments (GMM), the empirical likelihood (EL) and the quantiles. The hypothesis-testing procedures used are Wald's (W), Rao's score (RS) and the likelihood ratio (LR) methods. Since these are based on sample information, the statistical properties (unbiasedness, consistency, efficiency, distributions) of these procedures are of great interest for both small and large samples. This has led to the development of asymptotic theory (large sample) econometrics (White, 2001) and finite sample econometrics (Ullah, 2004).

The large sample theory properties may not imply finite sample behaviour of econometric estimators and test statistics, and they can give misleading results for small or even moderately large samples. As an example, consider a regression model

$$y_i = x_i\beta + u_i \qquad i = 1,2\ldots,n,$$

where y_i is a univariate response, x_i is a univariate fixed regressor, β is an unknown parameter to be estimated, and u_i is an additive error assumed to be independently and identically distributed (i.i.d.) with mean zero and variance σ^2. Let b_1, and $b_2 = (1-1/n)b_1$ be two estimators of β, where b_1 is the LS estimator. Then, the asymptotic distributions of b_1 and b_2 are

$$\sqrt{n}(b_1 - \beta) \sim N(0, \sigma^2/m_{xx}),$$
$$\sqrt{n}(b_2 - \beta) \sim N(0, \sigma^2/m_{xx}),$$

where $m_{xx} = \Sigma_{i=1}^{n} \frac{x_i^2}{n}$ as n tends to ∞.

Thus, asymptotically, both estimators are unbiased, and they have the same variances and distributions. But these results do not hold for finite samples (small or moderately large), since in this case $Eb_1 = \beta$, $Eb_2 = \beta(1-1/n)$, $V(b_1) = \sigma^2/\Sigma_{i=1}^{n} x_i^2$, $V(b_2) = (1-1/n)^2 V(b_1)$, that is, while b_1 is unbiased, b_2 is biased and their variances are different. Further, the distributions of b_1 and b_2 are generally not known but, if we assume normality of errors, then both b_1 and b_2 are normally distributed.

Fisher (1921; 1922) and then the work of Cramér (1946) laid the foundations of statistical finite sample theory on the exact distributions and moments which are valid for any sample size. This exact theory on distributions and moments was brought into econometrics by the seminal work of Haavelmo (1947) and Anderson and Rubin (1949) on the exact confidence regions of structural coefficients, Hurwicz (1950) on the exact LS bias in an autoregressive model, Basmann (1961) and Phillips (1983) on the exact density and moments of the estimators in the structural model, and Ullah (2004) on the exact moments. However, these exact results are often very complicated for drawing meaningful inferences since they are expressed in terms of multivariate integrals or complex infinite series. Also, the results are not derivable for non-classical models, especially for non-linear models or models with non-normal errors.

Another major development took place through the pioneering work of Nagar (1959) on obtaining the approximate moments of the k-class estimators in simultaneous equations. This was followed by Sargan (1975) and Phillips (1980), who rigorously developed the theory and applications of the Edgeworth expansions to derive the approximate distribution functions of econometric estimators. (The idea of the Edgeworth expansions originates from the fundamental work of Edgeworth, 1896.) The approximate distributions and moments provide results which can tell us how much we lose by using asymptotic results and how far we are from the exact results if they are known. Most of the contributions, however, were confined to the analytical derivation of the moments and distributions in the simultaneous equations model and the dynamic first-order autoregressive (AR (1)) model, but with i.i.d. normal observations. These also included the finite sample results using the Monte Carlo methodology (Hendry, 1984) and advances in bootstrapping (resampling) procedures (see Efron, 1979; Hall, 1992). The analytical and bootstrap results for non-classical models, especially those that are non-linear with non-normal and non-i.i.d.

observations, remain a challenging task for future development in this area of research. For the approximate analytical results some development has begun to take place (Rilstone, Srivastava and Ullah, 1996) with a non-i.i.d. extension in Ullah (2004). This provides results which can be used to evaluate the approximate bias and mean-squared error of a class of estimators (ML, LS, GMM) for linear and non-linear models with normal or non-normal errors, and the observations can be i.i.d. or non-i.i.d. In the same spirit Newey and Smith (2004) develop the properties of generalized empirical likelihood estimators. Similarly, there are developments in the bootstrapping procedures for studying the properties of the GMM and extremum estimators in various econometric models with i.i.d. as well as dependent and non-stationary observations (see Horowitz, 2001).

The progress in finite sample econometrics has indeed been ongoing. The developments described provide analytical and simulation-based procedures for finite sample analysis of econometric models. In the broad sense, the frontier of this research area has moved on. With the advances in computer technology this subject will further develop in both the analytical and the bootstrapping domains.

Aman Ullah

See also bootstrap; econometrics; Edgeworth expansion; simultaneous equations models

Bibliography

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

Basmann, R. 1961. Note on the exact finite sample frequency functions of generalized classical linear estimators in two leading overidentified cases. *Journal of the American Statistical Association* 56, 619–36.

Cramér, H. 1946. *Mathematical Methods of Statistics Princeton*. Princeton: Princeton University Press.

Edgeworth, F. 1896. The asymmetrical probability curve. *Philosophical Magazine* 41, 90–9.

Efron, B. 1979. Bootstrap methods: another look at the jacknife. *Annals of Statistics* 7, 1–26.

Hall, P. 1992. *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Haavelmo, T. 1947. Methods of measuring the marginal propensity to consume. *Journal of the American Statistical Association* 42, 105–22.

Hendry, D. 1984. The Monte Carlo experimentation in econometrics. In *Handbook of Econometrics*, vol. 2, ed. M. Intriligator and Z. Griliches. Amsterdam: North-Holland.

Horowitz, J. 2001. The bootstrap in econometrics. In *Handbook of Econometrics*, vol. 5, ed. J. Heckman and E. Leamer. Amsterdam: North-Holland.

Hurwicz, L. 1950. Least square bias in time series. In *Statistical Inference in Dynamic Economic Models*, ed. T. Koopmans. New York: Wiley.

Fisher, R. 1921. On the probable error of a coefficient of correlation deduced from a small sample. *Metron* 1, 1–32.

Fisher, R. 1922. The goodness of fit of regression formulae and the distribution of regression coefficients. *Journal of the Royal Statistical Society* 85, 597–612.

Nagar, A. 1959. The bias and moments matrix of the general k-class estimators of the parameters in structural equations. *Econometrica* 27, 575–95.

Newey, W. and Smith, R. 2004. Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72, 219–55.

Phillips, P. 1980. Finite sample theory and the distribution of alternative estimators of the marginal propensity to consume. *Review of Economic Studies* 47, 183–224.

Phillips, P. 1983. Exact small sample theory in simultaneous equations models. In *Handbook of Econometrics*, vol. 1, ed. M. Intriligator and Z. Griliches. Amsterdam: North-Holland.

Rilstone, P., Srivatsava, V. and Ullah, A. 1996. The second order bias and MSE of nonlinear estimators. *Journal of Econometrics* 75, 239–395.

Sargan, J. 1975. Gram-Charlier approximations applied to t ratios of k-class estimators. *Econometrica* 43, 326–46.

Ullah, A. 2004. *Finite Sample Econometrics*. New York: Oxford University Press.

White, H. 2001. *Asymptotic Theory for Econometricians*. New York: Academic Press.