To appear in``The New Palgrave''

S.N.Durlauf and L.E.Blume (Eds.)

C000559

computational methods in econometrics

The computational properties of an econometric method are fundamental determinants of its importance and practical usefulness, in conjunction with the method's statistical properties. Computational methods in econometrics are advanced through successfully combining ideas and methods in econometric theory, computer science, numerical analysis, and applied mathematics. The leading classes of computational methods particularly useful for econometrics are matrix computation, numerical optimization, sorting, numerical approximation and integration, and computer simulation. A computational approach that holds considerable promise for econometrics is parallel computation, either on a single computer with multiple processors, or on separate computers networked in an intranet or over the internet.

1. Introduction

In evaluating the importance and usefulness of particular econometric methods, it is customary to focus on the set of *statistical* properties that a method possesses – for example, unbiasedness, consistency, efficiency, asymptotic normality, and so on. It is crucial to stress, however, that meaningful comparisons cannot be completed without paying attention also to a method's *computational* properties. Indeed the practical value of an econometric method can be assessed only by examining the inevitable interplay between the two classes of properties, since a method with excellent statistical properties may be computationally infeasible and vice versa. Computational methods in econometrics are evolving over time to reflect the current technological boundaries as defined by available computer hardware and software capabilities at a particular period, and hence are inextricably linked with determining what the state of the art is in econometric methodology.

To give a brief illustration, roughly from the late 1950s until the early 1960s we had the 'Stone Age' of econometrics, when the most sophisticated computational instrument was the slide rule, which used two rulers on a logarithmic scale, one sliding into the other, to execute approximate multiplication and division. In this Stone Age, suitably named in honour of Sir Richard Stone, winner of the 1984 Nobel Prize in Economics, the brightest Ph.D. students at the University of Cambridge were toiling for days and days in back rooms using slide rules to calculate ordinary linear regressions, a task which nowadays can be achieved in a split second on modern personal computers.

The classic linear regression problem serves to illustrate the crucial interaction between statistical and computational considerations in comparing competing econometric methods. Given data of size S, with observations on a dependent variable denoted by $S \times 1$ vector y and corresponding observations on k explanatory factors denoted by $S \times k$ matrix X $(k < X)$, the linear plane fitting exercise is defined by Gauss's minimum quadratic distance problem:

$$\hat{\beta} = \arg\min_b (y - Xb)'(y - Xb) \equiv \arg\ \min_b \sum_{s=1}^{S} (y_s - x_s' b)^2 \qquad (1)$$

where x_s' is the sth row of matrix X and b is a $k \times 1$ vector of real numbers defining the regression plane Xb. Under the assumption that X has full column rank k, the solution to this *ordinary least squares* minimization problem

is the linear-in-y expression $\hat{\beta} = (X'X)^{-1}X'y$, which only requires the matrix operations of multiplication and inversion.

Suppose, however, that Gauss had chosen instead as his measure of distance the sum of absolute value of the deviations, and defined instead:

$$\tilde{\beta} = \arg\min_{b} \sum_{s=1}^{S} |y_s - x_s'b|. \tag{2}$$

The vector $\tilde{\beta}$ that solves the second minimization is known as the *least absolute deviations* (LAD) estimator and has no closed-form matrix expression. In fact, calculation of $\tilde{\beta}$ requires highly nonlinear operations for which computationally efficient algorithms were developed only in the 1970s. To give a concrete example, consider the *intercept-only* linear regression model where X is the $S \times 1$ vector of ones. Then the single $\hat{\beta}$ coefficient that solves (1) is the sample mean of y, while $\tilde{\beta}$ that solves (2) is the sample *median* of y. The latter is orders of magnitude more difficult to compute than the former since it involves sorting y and finding the value in the middle, while the former simply adds all elements of y and divides by the sample size. Clearly, it could be quite misleading if $\hat{\beta}$ and $\tilde{\beta}$ where compared solely in terms of statistical properties without any consideration of their substantially different computational requirements.

A second example in a similar vein is the following parametric estimation problem. Suppose a sample of size S is observed on a single variable y. It is believed that each observation y_s is drawn independently from the same uniform distribution on the interval $[\theta,c]$ where the lower value of the support is the single unknown parameter that needs to be estimated, while c is known. Two parametric estimation methods with particularly attractive statistical properties are the generalized method of moments (GMM) and the method of maximum likelihood (MLE). Indeed, for relatively large sample sizes these two methods are comparably attractive in terms of statistical properties, while they differ *drastically* in terms of computational requirements: the GMM solution is $\hat{\theta}_{gmm} = \frac{2}{S}\sum_{s=1}^{S} y_s - c$, thus requiring only the simple calculation of the sample mean $\bar{y}$, while the MLE involves the highly nonlinear operation of finding the minimum of the data vector y, $\hat{\theta}_{mle} = min(y_1, \ldots, y_S)$.

In the following section we discuss in turn the leading classes of methods that are of particular importance in modern econometrics, while Section 3 introduces the concept of parallel processing and describes its current value and future promise in aiding dramatically econometric computation.

2. Computational methods important for econometrics

The advancement of computational methods for econometrics relies on understanding the interplay between the disciplines of econometric theory, computer science, numerical analysis, and applied mathematics. In the five subsections below we discuss the leading classes of computational methods that have proven of great value to modern econometrics.

2.1 Matrix computation and specialized languages

To start with the fundamental econometric framework of linear regression, the *sine qua non* of econometric computation is the ability to program and perform efficiently matrix operations. To this end, specialized matrix computer languages have been developed which include Gauss and Matlab.

Fundamental estimators of the linear regression coefficient vector β, like the OLS $(X'X)^{-1}X'y$ and its generalized least squares (GLS) variant $(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$, are leading examples of the usefulness of such matrix languages, where the $S \times S$ matrix Ω is a positive definite, symmetric variance-covariance matrix of the disturbance vector $\varepsilon \equiv y - X\beta$. Matrix operations are useful even for nonlinear econometric methods discussed below, since a generally useful approach is to apply linearization approximations through the use of differentiation and Taylor's expansions.

In implementing econometric methods that involve matrix operations, special attention needs to be paid to the dimensionality of the various matrices, as well as to any special properties a matrix may posses, which can affect very substantially the feasibility and performance of the computational method to be adopted. Looking at the OLS and GLS formulae, we see three different matrices that require inversion: $X'X$, Ω, and $X'\Omega^{-1}X$. The first and the third are of dimension $k \times k$, while the second is $S \times S$. Since the number of regressors k is typically considerably smaller than the sample size S, the inversion of these matrices can involve vastly different burden in terms of total number of computer operations required as well as memory locations necessary for holding the information during those calculations. (For example, in panel data settings where multiple observations are observed in different time-periods for a cross-section of economic agents, it is not uncommon to have total sample sizes of 300,000 or more.) To this end, econometric analysts have focused on importing from numerical analysis matrix algorithms that are particularly efficient in handling sparse as opposed to dense matrices. By their very nature, sparse matrices exhibit a very high degree of compressibility and concomitantly lower memory requirements. See Drud (1977) for the use of sparse matrix techniques in econometrics. A matrix is called sparse if it is primarily populated by zeros, for example, the variance-covariance matrix of a disturbance vector following the moving-average-of-order-1 model:

$$\Omega_{ma1} = \sigma^2 \begin{pmatrix} 1 & \frac{\lambda}{1+\lambda^2} & 0 & \cdots & 0 \\ \frac{\lambda}{1+\lambda^2} & 1 & \frac{\lambda}{1+\lambda^2} & \ddots & \vdots \\ 0 & \frac{\lambda}{1+\lambda^2} & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 & \frac{\lambda}{1+\lambda^2} \\ 0 & \cdots & 0 & \frac{\lambda}{1+\lambda^2} & 1 \end{pmatrix}.$$

In contrast, a stationary autoregressive disturbance of order 1 has a dense variance-covariance matrix:

$$\Omega_{ar1} = \sigma^2 \begin{pmatrix} 1 & \gamma & \gamma^2 & \cdots & \gamma^{S-1} \\ \gamma & 1 & \gamma & \ddots & \vdots \\ \gamma^2 & \gamma & \ddots & \ddots & \gamma^2 \\ \vdots & \ddots & \ddots & 1 & \gamma \\ \gamma^{S-1} & \cdots & \gamma^2 & \gamma & 1 \end{pmatrix}.$$

Other matrix algebra methods especially important in econometrics are the Cholesky factorization (see Golub,1969) of a positive definite matrix A into the product $A = R'R$ where R is an upper-triangular matrix, and the singular value decomposition that allows the calculation of pseudo-inverse of any

matrix B which may be non-square, and if square, not positive definite (see Belsley, 1974).

It is important to note that on occasion a brilliant theoretical development can simplify enormously the computational burden of econometric methods that, though possessing attractive statistical properties, were thought to be infeasible with existing computation technology in the absence of the theoretical development. A case in point is the GLS/MLE estimator for the one-factor random effects model proposed by Balestra and Nerlove (1966), which is of great importance in the analysis of linear panel data models. The standard formulation gives rise to the GLS formula requiring the inversion of an equi-correlated variance covariance matrix Ω of dimension $S \times S$, where S is of the order of the product of the number of available observations in the cross-section dimension times the number available in the time dimension. For modern panel data-sets, this can exceed 300,000, thus making the calculation of Ω^{-1} infeasible even on today's super-computers, let alone with the slide rules available in 1966. Fuller and Battese (1973), however, showed that the equi-correlated nature of the one-factor random effects model made calculation of the GLS estimator equivalent to an OLS problem, where the dependent variable $\tilde{y}$ and the regressors $\tilde{X}$ are simple linear combinations of the original data y_{it}, $x_{1it}, \ldots,$ x_{kit} and its time averages $\bar{y}_{i.}, \bar{x}_{1i.}, \ldots, \bar{x}_{ki.}$ defined by $\bar{y}_{i.} \equiv \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $\tilde{y}_{it} \equiv y_{it} - \lambda \bar{y}_{i.}$, and analogously for the regressor variables. This realization allowed the calculation of the GLS estimator without the need for inverting the usually problematically large Ω matrix.

Another important case where a theoretical development in methodology led to a dramatic lowering of the computational burden and hence allowed the calculation of models that would otherwise have had to wait perhaps for decades for sufficient advancements in computer technology is the simulation-based inference for Limited Dependent Variable models, associated with the name of Daniel McFadden (1989). See Section 2.5 below, MCFADDEN, DANIEL and SIMULATION-BASED ESTIMATION.

2.2 Optimization

Many econometric estimators with attractive statistical properties require the optimization of a (generally) nonlinear function of the form:

$$q \equiv \arg\max_{\theta} F(\theta; data) \tag{3}$$

over a vector of unknown parameters θ of dimension p, typically considerably larger than 1. Examples are: the method of maximum likelihood, minimum-distance (OLS, LAD, GMM), and other extremum estimators. (The need to optimize functions numerically is also important for certain problems in computational economics, for example, the problem of optimal control.) Algorithms for optimizing functions of many variables are a key component in the collection of tools for econometric computation. The suitability of a certain algorithm to a specific optimization econometric problem depends on the following classification:

1. *Algorithms that require the calculation of first and possibly second derivatives* Versus *algorithms that do not*. Clearly, if the function to be optimized is not twice continuously differentiable (as is the case with LAD) or even discontinuous (as is the case with the maximum score estimator for the semiparametric analysis of the binary response model – see Manski, 1975), algorithms that require differentiability will not be suitable. The

leading example of an algorithm not relying on derivatives is the nonlinear simplex method of Nelder and Meade (1965).

2. *Local* Versus *global algorithms*. Optimization algorithms of the first type (for example, Gauss-Newton, Newton-Raphson, and Berndt et al. (1974)) search for an optimum in the vicinity of the starting values fed into the algorithm. This strategy may not necessarily lead to a global optimum over the full set of parameter space. This is of particular importance if the function to be optimized has multiple local optima, where typically the estimator with the desirable statistical properties corresponds to locating the overall optimum of the function. In such cases, global optimization algorithms (for example, simulated annealing and genetic optimization algorithm) should be employed instead.

Special methods are necessary for constrained optimization, where a function must be maximized or minimized subject to a set of equality or inequality constraints. These problems, in general considerably more demanding than unconstrained optimization, can be handled through three main alternative approaches: interior, exterior, and re-parameterization methods.

Comprehensive reviews of optimization methods in econometrics can be found in Goldfeld and Quandt (1972), Quandt (1983), and Dennis and Schnabel (1984). These studies also discuss the related issue of the numerical approximation of derivatives and illustrate the fundamental link in terms of computation between optimization and the problem of solving linear and nonlinear equations. For similar methods used in economics, see NUMERICAL OPTIMIZATION METHODS IN ECONOMICS and NONLINEAR PROGRAMMING.

2.3 Sorting

Of special importance for computing the class of estimators known as robust or semiparametric methods is the ability to sort data rapidly and computationally efficiently. Such a need arises in the calculation of order statistics, for example, the sample median and sample minimum required by the first two estimation examples given above. The leading sorting algorithms, bubble-, heap-, and quick-sort, have fundamentally different properties in terms of computation speed and memory requirements, in general depending on how close to being sorted the original data series happens to be. For a practical review of the leading sorting algorithms, see Press et al. (2001, ch. 8).

2.4. Numerical approximation and integration

Numerical approximation is necessary for any mathematical function that does not have a closed form solution, for example, exponential, natural logarithm, and error functions. See Abramowitz and Stegun (1964) for an exhaustive study of mathematical functions and their efficient approximation. Judd (1996) focuses on numerical approximation methods particularly useful in economics and econometrics.

Numerical integration, also known as numerical quadrature, is a related approximation problem that is crucial to modern econometrics. There are two key fields of econometrics where integrals without a closed form must be evaluated numerically. The first is Bayesian inference where moments of

posterior densities need to be evaluated, which take the form of high-dimensional integrals. See, inter alia, Zellner, Bauwens and VanDijk (1988). The second main class is classical inference in limited dependent variable (LDV) models; for example, Hajivassiliou and Ruud (1994). See Geweke (1996) for an exhaustive review of numerical integration methods in computational economics and econometrics, and Davis and Rabinowitz (1984) for earlier results.

It is important to highlight a crucial difference between the numerical integration problems in Bayesian inference and those in classical inference for LDV models, which makes various integration-by-simulation algorithms be useful to one field and not the other: in the Bayesian case, typically a single or a few high-dimensional integrals have to be evaluated accurately. In contrast, in the classical LDV inference case, quite frequently hundreds of thousands of such integrals need to be approximated.

2.5 Computer simulation

The need for efficient generation of pseudo-random numbers with good statistical properties on a computer appears very routinely in econometrics. Leading examples include:

- Statistical methods based on resampling, primarily the 'jackknife' and the 'bootstrap', as introduced by Efron (1982). These methods have proven of special value in improving the small sample properties of certain econometric estimators and test procedures, for example in reducing estimation bias. They are also used to approximate the small sample variance of estimators for which no closed form expressions can be derived.
- Evaluation of econometric estimators through Monte Carlo experiments, where hypothetical data-sets with certain characteristics are simulated repeatedly and the econometric estimators under study are calculated for each set. This allows the calculation of empirical (simulated) properties of the estimators, either to compare to theoretical mathematical calculations or because the latter are intractable.
- Calculation of frequency probabilities of possible outcomes in large-scale decision trees, for which the outcome probabilities are impossible to characterize theoretically.
- Sensitivity analyses and what-if studies, where an econometric model is 'run' on a computer under different scenarios of policy measures.
- Simulation-based Bayesian and classical inference, where integrals are approximated through computer simulation (known as Monte Carlo integration). Particularly important methods in this context are the following: frequency simulation; importance sampling; and Markov chain Monte Carlo methods (the leading exponents being Gibbs resampling and the Metropolis/Hastings algorithm). A related class of methods, known as variance-reduction simulation techniques, includes control variates and antithetics. See Geweke (1988) and Hajivassiliou, McFadden and Ruud (1996) for reviews. See also SIMULATION-BASED ESTIMATION.

3. Parallel computation

Parallel processing, where a computation task is broken up and distributed across different computers, is a technique that can afford huge savings in

terms of total time required for solving particularly difficult econometric problems. For example, the simulation-based estimators mentioned in the previous section exhibit the potential of significant computational benefits by calculating them on computers with massively parallel architectures, because the necessary calculations can be organized in essentially an independent pattern. An example of such a computer is the Connection Machine CM-5 at the National Center for Supercomputing Applications in Illinois with 1,024 identical processors in a multiple-instruction/multiple-data (MIMDI) configuration. The benefits of such a parallel architecture on the problem of solving an econometric optimization classical estimator not involving simulation can also be substantial, since such estimators involve the evaluation of contributions to the criterion (for example, likelihood) function in the case of independently and identically distributed (i.i.d.) observations. Since typical applications in modern applied econometrics using cross-sectional and longitudinal data sets involve several thousands of i.i.d. observations, the potential benefits of parallel calculations of such estimators should be obvious. The benefits of a massively parallel computer architecture become even more pronounced in the case of simulation-based estimators. See Nagurney (1996) for a discussion of parallel computation in econometrics.

An alternative approach for parallel computation that does not involve a single computer with many processors has been developed recently and offers considerable promise for computational econometrics. Through the use of specialized computer languages, many separate computers are harnessed together over an organization's intranet or even over the internet, and an econometric computation task is distributed across them. The benefits of this approach depend critically on the relative burden of the overhead of communicating across the individual computers when organizing the splitting of the tasks and then collecting and processing the separate partial results. Such distributed parallel computation has the exciting potential of affording formidable super-computing powers to econometric researchers with only modest computer hardware.

Vassilis A. Hajivassiliou

See also

< xref = xyyyyyy > longitudinal data analysis;
< xref = xyyyyyy > McFadden, Daniel;
< xref = xyyyyyy > nonlinear programming;
< xref = xyyyyyy > numerical optimization methods in economics;
< xref = xyyyyyy > robust estimators in econometrics;
< xref = xyyyyyy > simulation-based estimation.

Bibliography

Abramowitz, M. and Stegun, I. 1964. *Handbook of Mathematical Functions*. Washington, DC: National Bureau of Standards.

Balestra, P. and Nerlove, M. 1966. Pooling cross-section and time-series data in the estimation of a dynamic model. *Econometrica* 34, 585–612.

Belsley, D. 1974. Estimation of system of simultaneous equations and computational specifications of GREMLIN. *Annals of Economic and Social Measurement* 3, 551–614.
Berndt, E.K., Hall, B.H., Hall, R.E. and Hausman, J.A. 1974. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement* 3, 653–66.
Davis, P.J. and Rabinovitz, P. 1984. *Methods of Numerical Integration*. New York: Academic Press.
Dennis, J.E. and Schnabel, R.B. 1984. *Unconstrained optimization and Nonlinear Equations*. Englewood Cliffs, NJ: Prentice-Hall.
Drud, A. 1977. An optimization code for nonlinear econometric models based on sparse matrix techniques and reduced grades. *Annals of Economic and Social Measurement* 6, 563–80.
Efron, B. 1982. *The Jackknife, the Bootstrap, and Other Resampling Plans*. CBMS-NSF Monographs No. 38.Philadelphia: SIAM.
Fuller, W.A. and Battese, G.E. 1973. Transformations for estimation of linear models with nested-error structure. *Journal of the American Statistical Association* 68, 626–32.
Geweke, J. 1988. Antithetic acceleration of Monte Carlo integration in Bayesian inference. *Journal of Econometrics* 38, 73–90.
Geweke, J. 1996. Monte Carlo simulation and numerical integration. In *Handbook of Computational Economics*, vol. 1, ed. H. Amman, D. Kendrik and J. Rust. Amsterdam: North-Holland.
Golub, G.H. 1969. Matrix decompositions and statistical calculations. In *Statistical Computation*, ed. R.C. Milton and J.A. Milder. New York: Academic Press.
Goldfeld, S. and Quandt, R. 1972. *Nonlinear Methods in Econometrics*. Amsterdam: North-Holland.
Hajivassiliou, V.A. and Ruud, P.A. 1994. Classical estimation methods using simulation. In *Handbook of Econometrics*, vol. 4, ed. R. Engle and D. McFadden. Amsterdam: North-Holland.
Hajivassiliou, V.A., McFadden, D.L. and Ruud, P.A. 1996. Simulation of multivariate normal rectangle probabilities and derivatives: theoretical and computational results. *Journal of Econometrics* 72(1, 2), 85–134.
Judd, K. 1996. Approximation, perturbation, and projection methods in economic analysis. In *Handbook of Computational Economics*, vol. 1, ed. H. Amman, D. Kendrik and J. Rust. Amsterdam: North-Holland.
Manski, C. 1975. Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics* 3, 205–28.
McFadden, D. 1989. A method of simulated moments for estimation of multinomial discrete response models. *Econometrica* 57, 995–1026.
Nagurney, A. 1996. Parallel computation. In *Handbook of Computational Economics*, vol.1, ed. H. Amman, D. Kendrik and J. Rust. Amsterdam: North-Holland.
Nelder, J.A. and Meade, R. 1965. A simplex method for function minimization. *Computer Journal* 7, 308–13.
Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. 2001. *Numerical Recipes in Fortran 77: The Art of Scientific Computing*. Cambridge: Cambridge University Press.
Quandt, R. 1983. Computational problems and methods. In *Handbook of Econometrics*, vol. 1, ed. Z. Griliches and M. Intriligator. Amsterdam: North-Holland.
Zellner, A., Bauwens, L. and VanDijk, H. 1988. Bayesian specification analysis and estimation of simultaneous equation models using Monte Carlo methods. *Journal of Econometrics* 38, 73–90.

Index terms

Bayesian inference
bootstrap
classical inference
computational methods
generalized least squares
generalized method of moments
importance sampling simulation

jackknife
least absolute deviations
maximum likelihood
numerical integration
optimal control
ordinary least squares
random effects models
simulation-based estimation
Stone, J. R. N.
Markov chain Monte Carlo methods
parallel computation