PART I

Survey papers

CHAPTER 1

Survey of continuous time econometrics

A. R. Bergstrom

1 Introduction

I have, already, published several surveys of the field of continuous time econometrics (1976, 1984a, 1988, 1990). In surveying the field again for this conference, I will emphasize the more recent developments and the importance of continuous time econometrics for dynamic disequilibrium modeling. My survey will cover both the theory of continuous time econometric methods and their empirical applications. It will start, however, with a restatement and discussion of the main arguments for the specification of econometric models in continuous time rather than discrete time.

2 Advantages of continuous time econometric models

I will briefly discuss seven advantages of continuous time econometric models, although the list of advantages could be extended (see e.g. Gandolfo 1993, pp. 2–4).

The first advantage, which is a very general one, is that a continuous time model can take account of the interaction between the variables during the unit observation period. Although the variables in a typical macroeconomic model are measured only at regular discrete intervals (for example, quarterly or annually), we know that they are adjusting at much shorter random intervals as a result of the uncoordinated decisions, at different points of time, of a large number of economic agents. Moreover, we do have information, from economic theory and other sources, concerning the interaction between the variables during the unit observation period. Because of the smallness of the samples with which econometricians must work, it is very important that we should take account of this information in our estimation procedures.

A second advantage, which is closely related to the first one, is that a continuous time model can represent a causal chain or causal system, in which each of the variables responds directly to the stimulus provided by a proper subset only of the other variables in the model, and yet there is interaction between all the variables in the model during the unit observation period. The formulation of econometric models as causal chains was strongly advocated by Herman Wold in his celebrated long-running debate with the Cowles Commission in the 1950s on causal-chain models versus nonrecursive, simultaneous equations models (Wold 1952, 1954, 1956, 1960; Bentzel and Hansen 1954; Strotz and Wold 1960; Strotz 1960). Wold had been brought up in the Stockholm tradition and regarded causal-chain models as more fundamental than nonrecursive simultaneous equations models. Moreover, although the models used in his own work were formulated in discrete time, he recognized the potential importance, for econometrics, of causal-chain models formulated as systems of differential equations (Wold 1956).

From an econometric point of view, an important argument in favor of causal-chain models is that they can take account of a priori information concerning the causal ordering of the variables. When the model is formulated as a system of linear differential equations, most of the coefficients in the system are restricted to zero, and the causal ordering of the variables is represented by the pattern of these zero restrictions. The assumption that the variables can be arranged in a causal chain of this type need not depend on any elaborate or controversial economic theory, but only on our knowledge of the information available to various agents at particular points of time. It is obvious, for example, that aggregate consumer expenditures on a particular day can be influenced by the levels (on that day) of only those variables that are known to consumers – particularly personal income, personal assets, and prices – and not by the levels (on that day) of such variables as exports, imports, and investment. This is very strong information, which can be used to reduce the variances of the parameter estimates. But it can be used efficiently only if the model is formulated in continuous time. For, because of the interaction of all variables during the unit observation period (say a quarter), the conditional expectation of the value of each variable in quarter t, conditional on information up to the end of quarter $t-1$, will be a function of the values in quarter $t-1$ (and, generally, quarters $t-2, t-3, \ldots$) of all the variables in the model, not just a proper subset of them.

A third advantage of continuous time models is that they permit a more accurate representation of the partial adjustment processes in dynamic disequilibrium models than is possible in a discrete time model. Nearly all of the continuous time macroeconometric models that have been developed during the last twenty years, starting with the disequilibrium neoclassical growth model for the United Kingdom by Bergstrom and Wymer

(1976), have been dynamic disequilibrium models formulated according to a common general pattern. A typical equation in such a model can be formally split into two parts. The first part, which can normally be derived from microeconomic equilibrium theory, is an equation relating the partial equilibrium level of the causally dependent variable to a proper subset of the other variables in the model. The second part is a differential equation of first or higher order, representing the adjustment of the causally dependent variable in response to the deviation of its current level from its partial equilibrium level. Such an equation can be obtained by solving a dynamic optimization problem, taking account of adjustment costs of various orders. (See e.g. Bergstrom and Chambers 1990; Hillinger, Reiter, and Wesser 1992.)

A fourth advantage, which is closely related to the previous one, is that a continuous time model provides the basis for a more accurate estimation of the distributed lags with which each variable depends on the variables on which it is directly causally dependent. Indeed, the distributed lag relation will have the form of a convolution integral that can be derived from the differential equation representing the partial adjustment relation. In this connection, it is worth recalling the important article of Sims (1971) (see also Geweke 1978), which showed that distributed lag relations satisfied by discrete data that have been generated by a continuous time model can give a very misleading impression of the underlying continuous lag distributions.

A fifth advantage is that procedures for estimating continuous time models make a clear distinction between the treatment of stock variables (e.g., the money supply and the stock of capital), which are measured at points of time, and flow variables (e.g., output and consumption), which are measured as integrals. Standard procedures for the estimation of models formulated in discrete time make no distinction between the treatment of stock and flow variables. Consequently, estimates of the parameters can be seriously biased owing to the specification error resulting from the aggregation over time implicit in the definition of the flow variables.

A sixth advantage is that the form of a continuous time model does not depend on the unit observation period. Discrete time models are much less flexible. Indeed, the form of quite simple discrete time models is dependent on the unit observation period. For example, if the monthly observations of a certain variable satisfy a second-order autoregressive model, then the quarterly observations of the same variables satisfy an autoregressive moving-average model. Since the econometrician can seldom choose the observation period but rather must work with the available data, the dependence of the form of the model on the unit observation period is a serious drawback.

A seventh advantage is that a continuous time model can be used to generate forecasts of the continuous time paths of the variables. Such forecasts can be of considerable value, in spite of the fact that the variables are observable only at discrete intervals of time. For example, a forecast of the continuous time path of the gross domestic product could be used by businessmen for sales forecasting or by the government for policy formulation.

3 Econometric methods for continuous time models

The antecedents and early history of continuous time econometrics were discussed in my historical survey (1988; reproduced in Gandolfo 1993), and I will now review them very briefly.

The problem of estimating the parameters of a continuous time stochastic model from discrete data was first discussed by Bartlett (1946). He pointed out that the economy does not cease to exist between observations, and went on to consider the problem of estimating single first- and second-order differential equations from discrete data. His paper was presented at a Meeting of the Royal Statistical Society and stimulated a lively discussion by M. G. Kendall and other leading statisticians, particularly concerning its potential importance for econometrics. But it seems to have been overlooked by econometricians. A possible reason for this is that, three years earlier, Haavelmo (1943) had introduced his simultaneous equations methodology, which became the dominant econometric methodology for the next thirty years.

The first econometrician to recognize the potential importance of continuous time models in econometrics was Koopmans (1950), who had himself played an important role in the development of the simultaneous equations methodology (see Koopmans, Rubin, and Leipnik 1950). But it was A. W. Phillips (1959) who developed the first detailed algorithm for estimating a continuous time model of sufficient generality to be used in applied work. From an econometric point of view, a limiting feature of the algorithm developed by Phillips was that it could not take account of a priori restrictions on the coefficient matrices of the continuous time model, such as the restriction that certain elements of these matrices should be zero.

I commenced work on continuous time econometrics in the early 1960s, and one of my objectives was to develop methods of estimation that could take account of these a priori restrictions. I decided that it would not be practicable, with the computing resources then available, to obtain asymptotically efficient estimates of the parameters of a continuous time model of ten or more equations, taking account of the exact restrictions on the distribution of the discrete data implied by the continuous time model. I decided, therefore, to follow a simpler approach based on the use of an approximate discrete model in the form of a nonrecursive simultaneous equations model. The idea that a nonrecursive simultaneous equations model is best regarded as some sort of approximation to a causal system had already been put forward by Strotz and Wold (1960).

In Bergstrom (1966) I derived the asymptotic bias of estimates obtained by applying the method of three-stage least squares to such an approximate discrete model. It was assumed, for this purpose, that the continuous time paths of the variables were generated by a closed system of first-order differential equations, with white-noise innovations, and that they were observable at equispaced points of time. The type of approximate discrete model used in this study was extended in subsequent articles by Sargan (1974, 1976) and Wymer (1972) to take account of exogenous variables and flow data. Moreover, a computer program for the estimation of the parameters of a continuous time model by applying the method of full-information maximum likelihood to the approximate discrete model was developed by Wymer (1968), and has been used in the estimation of nearly all of the continuous time models produced during the last twenty years.

It will be useful, at this stage, to return to the simpler case discussed in my 1966 article, and consider the relation between the approximate discrete model introduced in that paper and the exact discrete model satisfied by the data. The assumed continuous time model is a first-order system that can be written as

$$dx(t) = A(\theta)x(t)\,dt + \zeta(dt), \tag{1}$$

where $\{x(t), \infty < t < \infty\}$ is an n-dimensional continuous time random process, A is an $n \times n$ matrix whose elements are known functions of a p vector θ of unknown parameters ($p \leq n^2$), and $\zeta(dt)$ is a vector of white-noise innovations (see Bergstrom 1984a for a precise interpretation of this system). A sequence of equispaced observations $x(0), x(1), x(2), \ldots$ generated by the system (1) satisfies the exact discrete model

$$x(t) = F(\theta)x(t-1) + \epsilon_t, \tag{2}$$

where

$$F = e^A = I + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \cdots \tag{3}$$

and

$$E(\epsilon_s \epsilon_t') = 0, \quad s \neq t. \tag{4}$$

By integrating the system (1) over the interval $(t-1, t)$ and making a trapezoidal approximation to the unobservable integral $\int_{t-1}^{t} x(r)\,dr$, we obtain the approximate discrete model

$$x(t)-x(t-1)=\tfrac{1}{2}A\{x(t)+x(t-1)\}+u_t, \tag{5}$$

which is a simultaneous equations model.

The degree of accuracy of the approximate discrete model can be most easily assessed by considering its reduced form

$$x(t)=\Pi x(t-1)+v_t, \tag{6}$$

where

$$\begin{aligned}\Pi &= [I-\tfrac{1}{2}A]^{-1}[I+\tfrac{1}{2}A] \\ &= I+A+\tfrac{1}{2}A^2+\tfrac{1}{4}A^3+\cdots.\end{aligned} \tag{7}$$

Comparing (3) and (7), we see that the difference between F and Π depends only on A^3 and higher powers of A. The model (1) implies that the elements of A have the same dimension as the unit of time, which we are here identifying with the unit observation period. Our comparison of F and Π thus implies that the error in the model (6), regarded as an approximation to the exact discrete model (2), is of the same order as the cube of the unit observation period (as the latter tends to zero).

If the only restrictions on A are that certain elements be zero, in which case θ is just the vector of unknown elements of A, then θ can be estimated by applying standard simultaneous equations estimation procedures to the approximate discrete model (5). Numerical examples considered in my 1966 article showed that the asymptotic bias of such estimates is quite small under realistic assumptions about the speeds of adjustment and the unit observation period. Moreover, it was later shown by Sargan (1974, 1976) that the asymptotic bias of such estimates is of the order of the square of the unit observation period as the latter tends to zero.

In spite of these asymptotic results, an important Monte Carlo study by P. C. B. Phillips (1972) showed that, with finite samples, spectacular gains in efficiency can be obtained by using the exact discrete model (2) rather than the approximate discrete model. Assuming a first-order three-equation model with observations at equispaced points of time, he obtained parameter estimates first by applying three-stage least squares to the approximate discrete model and then by applying the minimum distance estimation procedure of Malinvaud (1980, chap. 9) to the exact discrete model. He found that use of the exact discrete model led to a more than 50-percent reduction in the root-mean-square errors of the estimates of most of the parameters of his model, and that this was mainly due to a reduction in the variance.

The results obtained by Phillips suggested that we should, ultimately, develop methods of estimation based on the exact discrete model. But the use of such methods would not have been practicable with the computing resources available in the 1970s, except for the first-order models with no flow variables. When the variables are a mixture of stocks and flows, then the exact discrete model corresponding to even a first-order system is a vector autoregressive moving-average model with complicated restrictions between the autoregresive and moving-average coefficient matrices. When estimating continuous time models from flow data, the normal procedure (until very recently) has been first to transform the data using an autoregressive transformation introduced by Bergstrom and Wymer (1976) and then to use the transformed data to estimate the approximate discrete model (1) or, in some cases, the discrete model (5). But estimates obtained in this way are not asymptotically efficient, even if the model (5) is used.

During the last decade there have been enormous developments in computing technology, and the latest generation of supercomputers is capable of speeds of several billion floating-point operations per second. Concurrent with these developments in computing technology has been an immense amount of work on the development of algorithms for the computation of asymptotically efficient estimates of the structural parameters of continuous time models. These algorithms are applicable to mixed stock and flow data and take account of the exact restrictions on the distribution of the data implied by the continuous time model. They are also applicable to models formulated as systems of second- or higher-order differential equations. Such models permit a much richer dynamic specification than the predominantly first-order models developed during the last twenty years.

The most general program that has been developed for the computation of asymptotically efficient estimates of the parameters of a continuous time model is that of Nowman (1992). This program is applicable to a second-order system with mixed stock and flow data and yields the exact Gaussian estimates (which are the exact maximum likelihood estimates if the innovations are Brownian motion) when either the model is closed or the exogenous variables are quadratic functions of time.

I will now describe, in some detail, the algorithm implemented by Nowman's program and developed in a series of articles published during the 1980s (see Bergstrom 1983, 1985, 1986, 1989). For various extensions and limiting cases, see Agbeyegbe (1984, 1987, 1988), Chambers (1991a), and Nowman (1991). The basic model to which the algorithm is applicable can be written in the form

$$\begin{aligned} d[Dx(t)] = [A_1(\theta)Dx(t) + A_2(\theta)x(t) + B(\theta)z(t)]\,dt \\ + \zeta(dt), \quad t \geq 0, \\ x(0) = y_1, \quad Dx(0) = y_2, \end{aligned} \tag{8}$$

where $x(t)$ is an n-dimensional vector of endogenous variables, $z(t)$ is an m-dimensional vector of exogenous variables, θ is a p-dimensional vector of unknown structural parameters ($p \leq 2n^2+nm$), A_1, A_2, B are matrices whose elements are known functions of θ, y_1 and y_2 are nonrandom vectors, D is the mean-square differential operator, and $\zeta(dt)$ is a vector of white-noise innovations. The model (8) can be made more general by including the first and second derivatives of $z(t)$; the computer program of Nowman (1992) is directly applicable to the more general model using interpolation formulas for $Dz(t)$ and $D^2z(t)$ given in Nowman (1991). However, for simplicity of exposition, I will confine the discussion to the model (8).

We assume that both the endogenous and exogenous variables are a mixture of stocks and flows, and (without loss of generality) that they are ordered so that the vectors $x(t)$, $z(t)$, y_1, and y_2 can be partitioned as

$$x(t) = \begin{bmatrix} x^s(t) \\ x^f(t) \end{bmatrix}, \qquad z(t) = \begin{bmatrix} z^s(t) \\ z^f(t) \end{bmatrix},$$

$$y_1 = \begin{bmatrix} y_1^s \\ y_1^f \end{bmatrix}, \qquad y_2 = \begin{bmatrix} y_2^s \\ y_2^f \end{bmatrix},$$

where $x^s(t)$ and $z^s(t)$ are vectors of stock variables (which are observable at points of time) while $x^f(t)$ and $z^f(t)$ are vectors of flow variables (which are observable as integrals over the period $(t-1, t)$).

We define the observable vectors $\bar{x}_t$ and $\bar{z}_t$ by

$$\bar{x}_t = \begin{bmatrix} x^s(t) - x^s(t-1) \\ \int_{t-1}^{t} x^f(r)\,dr \end{bmatrix}, \qquad \bar{z}_t = \begin{bmatrix} \frac{1}{2}\{z^s(t) + z^s(t-1)\} \\ \int_{t-1}^{t} z^f(r)\,dr \end{bmatrix}.$$

The sample then comprises the set of vectors

$$\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_T, \bar{z}_1, \bar{z}_2, \ldots, \bar{z}_T$$

together with the initial stock vector $y_1^s = x^s(0)$, which is the only part of the initial state vector that is observable.

The remainder of the initial state vector, that is, the vector defined by

$$y_1 = \begin{bmatrix} y_1^f \\ y_2^f \end{bmatrix},$$

is unobservable and must be treated as part of the parameter vector to be estimated. We can also parameterize the covariance matrix of the white-noise innovation vector by writing it in the form

$$E\{\zeta(dt)\zeta'(dt)\} = dt\Sigma(\mu),$$

where Σ is a positive definite matrix whose elements are known functions of a parameter vector μ. The complete vector of parameters to be estimated is then $[\theta, \mu, y']$.

In order to derive the Gaussian likelihood function, we make use of a discrete time dynamic model which is exactly satisfied by the data when the unobservable continuous time paths of the exogenous variables are polynomials in t of degree not exceeding 2 and which is also a very good approximation under much more general conditions. The assumption that the exogenous variables are quadratic functions of time was first used in the derivation of estimates of the parameters of continuous time models by Phillips (1974, 1976). He dealt only with first-order systems, however, and assumed that all variables are observable at equispaced points of time; that is, there are no flow variables.

The exact discrete model derived from the second-order system (8) is, in its most basic form, the following set of equations (9)–(11):

$$\bar{x}_1 = G_{11}y_1 + G_{12}y_2 + E_{11}\bar{z}_1 + E_{12}\bar{z}_2 + E_{13}\bar{z}_3 + \eta_1, \tag{9}$$

$$\bar{x}_2 = C_{11}\bar{x}_1 + G_{21}y_1 + G_{22}y_2 + E_{21}\bar{z}_1 + E_{22}\bar{z}_2 + E_{23}\bar{z}_3 + \eta_2, \tag{10}$$

$$\bar{x}_t = F_1\bar{x}_{t-1} + F_2\bar{x}_{t-2} + E_0\bar{z}_t + E_1\bar{z}_{t-1} + E_2\bar{z}_{t-2} + \eta_t, \quad t = 3, \dots, T, \tag{11}$$

where

$$\begin{aligned}
E(\eta_1\eta_1') &= \Omega_{11}, & E(\eta_2\eta_2') &= \Omega_{22},\\
E(\eta_2\eta_1') &= \Omega_{21}, & E(\eta_3\eta_1') &= \Omega_{31},\\
E(\eta_3\eta_2') &= \Omega_{32}, & E(\eta_4\eta_2') &= \Omega_{42},\\
E(\eta_t\eta_t') &= \Omega_0, \quad t = 3, \dots, T,\\
E(\eta_t\eta_{t-1}') &= \Omega_1, \quad t = 4, \dots, T,\\
E(\eta_t\eta_{t-2}') &= \Omega_2, \quad t = 5, \dots, T,\\
E(\eta_t\eta_{t-r}') &= 0, \quad r > 2, \ t = 3, \dots, T.
\end{aligned}$$

The matrices of coefficients in (9)–(11) are all functions of θ and whose precise forms are given in Bergstrom (1986); the matrices Ω_{11}, Ω_{22}, Ω_{21}, Ω_{31}, Ω_{32}, Ω_{42}, Ω_0, Ω_1, and Ω_2 are all functions of $[\theta, \mu]$ and whose precise forms are also given in Bergstrom (1986). Moreover, (11) is exactly satisfied for any interval $[t'-3, t']$ over which the elements of $z(t)$ are quadratic functions of t, and the coefficient matrices E_0, E_1, E_2 depend only on θ and not on the parameters of the quadratic functions defining $z(t)$. For this reason, the model represented by equations (9)–(11) is a very good approximation even if the behavior of the exogenous variable varies greatly (as between different parts of the sample period), provided that the paths are sufficiently smooth.

From (9)–(11) and the listed conditions on the covariance matrix of the vector $[\eta_1', \eta_2', \dots, \eta_T']'$, we can derive (see Bergstrom 1990, chap. 7) the exact discrete model in the form of a generalized VARMAX model, which is a very convenient basis for estimation, testing, and forecasting. The generalized VARMAX model is given by equations (12)–(14), in which