DOCUMENT RESUME

ED 173 207 SO 011 765

AUTHOR	Hannan, Michael T.
TITLE	Final Report for Dynamic Models for Causal Analysis of Panel Data. Models for Change in Quantitative Variables, Part III: Estimation from Panel Data. Part II, Chapter 5.
INSTITUTION	Center for Advanced Study in the Behavioral Sciences, Stanford, Calif.; Stanford Univ., Calif.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.; National Science Foundation, Washington, D.C.
PUB DATE	Aug 78
GRANT	BNS-76-22943; NIE-G-76-0082
NOTE	48p.; For related documents, see SO 011 759-772
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Comparative Analysis; *Educational Change; Measurement Techniques; Models; *Research Design; *Research Methodology; Research Problems; *Social Change; Social Science Research; Sociometric Techniques; *Validity

ABSTRACT

This document is part of a series of chapters described in SO 011 759. Addressing the problems of studying change and the change process, the report argues that sociologists should study coupled changes in qualitative and quantitative outcomes (e.g., marital status and earnings). The author presents a model for sociological studies of change in metric variables and focuses on the practical details of using stochastic differential equations (SDEs) in panel analysis. Eleven sections comprise the report. The first considers two broad approaches to empirical studies of SDEs: estimating structural parameters directly from integral equations or using discrete approximations. After arguing for the first approach, the author addresses issues such as autocorrelation of disturbances, unit-specific effects, pooled cross-section and time series estimators, fixed effect estimators, random effects estimators, Monte Carlo studies of small sample properties, and unequally spaced observations. A continuous-time perspective for dealing with problems of analyzing unequally spaced panel data is proposed. The report concludes that researchers can make use of available methods to solve many of the practical problems that arise in applying continuous-time, continuous-state models in sociological research. (Author/KC)

Final Report for

DYNAMIC MODELS FOR CAUSAL ANALYSIS OF PANEL DATA.

MODELS FOR CHANGE IN QUANTITATIVE VARIABLES, PART III: ESTIMATION FROM PANEL DATA*

Michael T. Hannan

April 1978

Revised August 1978

*This report is one in a series from a collaborative project with Nancy Brandon Tuma, funded by National Institute of Education, grant NIE-G-76-0082. While this report was written Hannan was a Fellow at the Center for Advanced Study in the Behavioral Sciences and was supported by National Science Foundation grant BNS76-22943. I wish to thank Nancy Brandon Tuma, Aage B. Sørensen, Glenn Carroll and Barbara Warsavage for comments on an earlier draft. This report is also available as Technical Report #65 of the Laboratory for Social Research, Stanford University.

In Parts I and II (Hannan 1978a,b) we have argued that sociological studies of changes in metric variables should be modeled as stochastic differential equations. We have shown that, at least for linear SDE's, we can obtain explicit solutions for probability densities. Thus there are no obvious impediments to estimating the parameters of such models from conventional panel data. We now turn attention to the practical details of this strategy. But we cannot borrow so directly from the technical literature. While the problem of estimating SDE's has been widely studied,[1] almost all work has considered time series designs. Since we focus on panel analysis, we must modify the usual strategies.

We begin by considering two broad approaches to empirical study of SDE's. One involves estimating structural parameters directly from integral equations; the second uses discrete approximations. We argue for the former strategy and outline the obvious maximum likelihood estimation approach in the panel context. In any realistic application of the methods we propose, disturbances will be autocorrelated. The problem of autocorrelation stands as the major obstacle to sound inference concerning dynamic models. We thus devote considerable attention to this complication, especially in the context of estimation from pooled cross-section and time series designs.

After presenting the large sample estimation theory, we turn to Monte Carlo evidence on the small sample properties of the pooled estimators we use. In particular we contrast the performance of maximum likelihood and generalized least squares estimators.

In the final section we raise a very important practical problem: unequal spacing of observations. Virtually all methodological work on panel analysis assumes that data are equally spaced in time. We suggest that sociologists often obtain data with much less regular spacing. Thus it is important to extend our strategy to such cases. One of the main advantages of using continuous time models is that they permit systematic treatment of unequally spaced data. We illustrate this advantage and show how maximum likelihood estimators may be adapted to handle unequal spacing.

1. Two Approaches

We begin with broad strategic considerations. Analysis of integral-equations for the purpose of obtaining estimates of dynamic parameters as sketched in Hannan (1978a) has obvious appeal. This strategy highlights the connection between the mathematical model and the statistical model. It also permits use of standard estimation techniques. The strategy has at least one drawback that in some circumstances limits its value sharply. In the case of systems of equations, it is difficult to impose constraints on parameters in estimation. Suppose, for example, that theory implies that one of the entries in B, the matrix of parameters of the endogenous portion of the system, is zero. The eigenvalue-eigenvector approach for using data to generate estimates of B does not permit us to use this constraint in estimation. One consequence is non-zero estimates of the parameter "known" to be zero. Moreover, estimates of other parameters will be less than fully efficient. The latter is just an instance of general rule that consistent estimators that ignore constraints have larger variances than other consistent estimators that use them.

This limitation has motivated statisticians and econometricians to seek more efficient estimators. Most attention has focused on so-called "exact discrete approximations" to stochastic differential equations. This strategy replaces the continuous-time model with a discrete-time analogue. The [illegible] is a parameter that, when made sufficiently small, converts the [illegible] to the proper continuous-time model. The advantage of this strateg[illegible] constraints on parameters may be employed routinely. One does not give [illegible]etely on continuous-time modeling since the estimated model is considered an approximation to the continuous-time model.

Exact discrete approximation estimators have been developed and applied to time series data (see Bergstrom 1976) but not, to our knowledge, to panel data. Extensions to panel applications appear not to be trivial. We do not attempt such an extension here but continue to focus on the strategy of using integral equations directly. The most important consideration in this choice concerns the spacing of observations. The "exact discrete approximation" approach requires equally spaced observations. Thus they do not apply to wide classes of sociological research.

So we choose to retain generality at the cost of efficiency in estimation. As long as we use reasonably large samples, the price should not be too high. However, when systems of equations are to be estimated from equally spaced data on small samples, it is worth investigating the alternative approach. We do not pursue this problem here.

2. Single Equation Models

In Part II we treated the following simple extension of the OU process:

$$dY(t) = adt + bY(t)dt + cXdt + \sigma d\beta_t \quad ; \quad b < 0 \tag{1}$$

where β_t is normal Brownian motion. This model may be considered as a stochastic negative feedback or linear partial adjustment model with a single (constant) exogenous variable. As we indicated in the last chapter, this has solution (with initial condition $Y(0) = Y_0$)

$$Y(t) = \frac{a}{b}(e^{b\Delta t}-1) + e^{b\Delta t}Y_0 + \frac{c}{b}(e^{b\Delta t}-1)X + \varepsilon(t) \qquad (2)$$

where

$$\varepsilon(t) \text{ is } N(0, \frac{\sigma^2}{2b}[1-e^{2b\Delta t}]). \qquad (3)$$

We rewrite (2), with obvious substitutions, as

$$Y(t) = a^* + b^*Y_0 + c^*X + \mathcal{E}(t)$$

Suppose that observations on N entities at times 0 and t are available. If all N units follow the same process, we may write

$$Y_i(t) = a^* + b^*Y_i(0) + c^*X_i + \mathcal{E}_i(t) \quad (i = 1,\ldots, N) \qquad (4)$$

As long as the disturbance process $d\beta_{ti}$ is independent from observation to observation, i.e., that the $\mathcal{E}_i(t)$ are independent and identically distributed, this model may be analyzed by ordinary least squares (OLS). In fact, under the generalized Gauss-Markov theorem OLS is a best linear unbiased (BLUE) estimator of a^*, b^*, c^*, and σ^*. That is, among the class of linear unbiased estimators it has smallest variance. Since the disturbances are normally distributed, OLS is identical to maximum likelihood. This fact considerably simplifies the problem of using $\hat{a}^*$, $\hat{b}^*$, $\hat{c}^*$ to estimate the structural parameters a, b, and c.

Maximum likelihood estimators have a strong invariance property: monotonic (i.e., order-preserving) functions of MLE are also MLE. By comparing b^* and b in (2) and (4) we see

$$e^{b\Delta t} = b^*$$
$$\text{or} \quad \frac{\ln b^*}{\Delta t} = b . \qquad (5)$$

Consequently $\delta = \frac{\ln \delta^*}{\Delta t}$ is a maximum likelihood estimator as long as δ^* is ML. This fact does not hold generally for least squares estimators (i.e., when OLS is not ML). Least squares estimators retain consistency under nonlinear transformations but lose asymptotic efficiency.

In much substantive work the exogenous variables will not be constant over time. The solution of the differential equation involves terms of the form (see Hannan 1978a):

$$\int_{t_0}^{t} e^{b(s-t_0)} f(s)ds$$

where f(s) is some function of time. Coleman (1968) remarked that common scientific practice when f(s) is unknown is to assume that the exogenous variable(s) change linearly over time. Then the integral equation has the form (see 10.14):

$$Y(t) = a^* + b^*Y_0 + c_1^*X_0 + c_2^*\Delta X(t) + \mathcal{E}(t)$$

which may be analyzed by OLS or ML. Of course, other hypotheses concerning the dynamic behavior of exogenous variables give different "estimation" equations. This matter may not be treated mechanically. After all, the dynamics of the outcome variable must depend on the dynamics of the forcing variables. Unless we specify the latter well, we cannot have much hope of doing a good job with the former. Nonetheless, for simplicity we focus attention on the simplest assumptions concerning f(s).

3. Systems of Equations

Next consider the simple linear system

$$dY_1(t) = a_1dt + b_{11}Y_1(t)dt + b_{12}Y_2(t)dt + \cdots + b_{1K}Y_K(t)dt + c_1Xdt + \sigma_1 d\mathcal{E}_{1t}$$
$$dY_K(t) = a_Kdt + b_{K1}Y_1(t)dt + b_{K2}Y_2(t)dt + \cdots + b_{KK}Y_K(t)dt + c_KXdt + \sigma_K d\mathcal{E}_{Kt} \quad (6)$$

where the β_{kt}, k = 1,...,K, are independent normal Brownian motion processes. We express this system more compactly as

$$d\underline{y}(t) = \underline{a}dt + \underline{B}\underline{y}(t) + \underline{c}x + \underline{\Sigma}d\beta_t \qquad (7)$$

where $\underline{\Sigma} = \sigma\underline{I}$, $\underline{I}$ is a K by K identity matrix. In Part II (Hannan 1978b), we saw that (7) can be integrated with initial condition $\underline{y}(0) = \underline{y}_0$, to yield

$$\underline{y}(t) = \underline{B}^{-1}\underline{a}(e^{\underline{B}\Delta t}-\underline{I}) + e^{\underline{B}\Delta t}\underline{y}_0 + B^{-1}c(e^{\underline{B}\Delta t}-\underline{I})X + \underline{\varepsilon}(t) \qquad (8)$$

where $\underline{I}$ is a K by K identity matrix. Since the disturbances are independent Brownian motions, the $\underline{\varepsilon}(t)$ has a simple structure. Each $\varepsilon_k(t)$ is $N(0, \frac{\sigma_k^2}{2b_k}(1-e^{2b_k\Delta t}))$ and $E[\varepsilon_k(t)\,\varepsilon_j(t)] = 0$ for $k \neq j$. Thus the linear system

$$\underline{y}(t) = \underline{a}^* + \underline{B}^*y_0 + \underline{c}^*x + \underline{\varepsilon}(t) \qquad (9)$$

has independent disturbances. It is, in fact, a recursive system (we do not need "simultaneous equations estimators"). Again OLS and ML are equivalent. They give asymptotically unbiased and efficient estimators of $\underline{a}^*$, $\underline{B}^*$, $\underline{c}^*$, and $\underline{\sigma}^*$.

Therefore we may employ the eigenvalue-eigenvector method of Part I, Section 5 to estimate $\underline{B}$ from $\hat{\underline{B}}^*$. It is then a simple problem in algebra to recover $\hat{\underline{a}}$ and $\hat{\underline{c}}$ from $\hat{\underline{a}}^*$, $\hat{\underline{c}}^*$, and $\hat{\underline{B}}$. We may also, if we wish, calculate standard errors of these parameter vectors from the estimated standard errors of $\hat{\underline{a}}^*$, $\hat{\underline{B}}^*$, $\hat{\underline{c}}^*$, and $\hat{\underline{\sigma}}^*$. All these calculation may be done with any of a set of widely available computer routines for extracting eigenvectors and eigenvalues.

The system case poses only one new inference issue, raised at the outset. The procedure we use does not employ constraints on parameters in $\underline{B}$. It is not, as a result, fully efficient when such constraints are appropriate.

4. Autocorrelation of Disturbances

We have not yet mentioned the practical complication that pervades most discussions of temporal analysis: autocorrelation of disturbances. Wide experience reveals that factors omitted from our models are at least moderately stable over time. As a result disturbances tend to be correlated over time. When disturbances are autocorrelated (i.e., correlated over time), the effects of omitted variables are usually confounded with the effects of the lagged dependent variables(s), $\underline{Y}_0$. This is a standard problem in panel analysis. Unless autocorrelation is handled properly, we will not obtain good estimates of dynamic parameters. The problem is particularly severe with continuous-time models in which all parameter estimates depend on $\hat{\underline{B}}$, since autocorrelation particularly affects estimates of $\underline{B}$ (see Johnston 1972; Hannan and Young 1977).

Although we suspect autocorrelation in practical applications, the model developed so far does not reflect this. Recall that $\underline{\beta}_t$ is a Brownian motion process with independent increments. Thus the increments in successive periods are independent (this is not true in general of Markov processes as we remarked in Part II). So we must modify the model if we are to deal with the autocorrelation problem in a systematic manner. We know of two strategies. The first involves complicating the random forcing function, relaxing the independent increments assumption. The alternative is to introduce individual-specific parameters (effects of stable individual characteristics) into the stochastic differential equation.

Of course, the two strategies may be combined. But it is useful to contrast the substantive interpretations that fit one or the other. The strategy of introducing unit-specific effects fits well those circumstances in which the omitted causal variables are approximately constant over the study period. When the omitted variables change greatly over the study period, the alternative procedure is called for. Our substantive work has relied on the unit-specific effect approach. Let us begin by outlining this approach. Once we have done so, we will be in a better position to clarify the nature of the alternative procedure.

5. Unit-Specific Effects

Suppose that the N units under study change according to the same general process (7) but that each unit has a distinct "constant" rate of change. In the study of individual careers these constants might include physiological characteristics (e.g., energy levels), enduring features of personality, status origins, ethnicity, linguistic styles, etc. In studies of organizations, they would include material infrastructure (e.g., characteristics of physical arrangements), stable features of work technology, long-standing political alliances, cultural attributes of members, etc. For each unit, summarize the effects of all such stable omitted variables as a single quantity, m_{ik} (for the i^{th} unit in the k^{th} equation). In other words, each unit has its own dynamic process (due to the m_{ik}) but the remaining parameters are constrained to be the same for all units. Thus we must consider the system of NK equations:

$$\begin{bmatrix} dY_{11} & \cdots & dY_{1N} \\ \vdots & & \vdots \\ dY_{K1} & \cdots & dY_{KN} \end{bmatrix} = \begin{bmatrix} a_1 & \cdots & a_1 \\ \vdots & & \vdots \\ a_K & \cdots & a_K \end{bmatrix} dt + \begin{bmatrix} m_{11} & \cdots & m_{1N} \\ \vdots & & \vdots \\ m_{K1} & \cdots & m_{KN} \end{bmatrix} dt + \begin{bmatrix} b_{11} & \cdots & b_{1K} \\ \vdots & & \vdots \\ b_{K1} & \cdots & b_{KK} \end{bmatrix}$$

$$\begin{bmatrix} Y_{11} & \cdots & Y_{1N} \\ \vdots & & \vdots \\ Y_{K1} & \cdots & Y_{KN} \end{bmatrix} dt + \begin{bmatrix} c_1 \\ \vdots \\ c_K \end{bmatrix} [X_1, \ldots, X_N]\, dt + \begin{bmatrix} \sigma_1 & & 0 \\ & \sigma & \\ 0 & & \sigma_K \end{bmatrix}$$

$$\begin{bmatrix} d\beta_{11} & \cdots & d\beta_{1N} \\ \vdots & & \vdots \\ d\beta_{K1} & \cdots & d\beta_{KN} \end{bmatrix} \qquad (10)$$

or, equivalently:

$$\underline{dY}(t) = \underline{a}\,\underline{i}'dt + \underline{M}dt + \underline{BY}(t)dt + \underline{cx}'dt + \underline{\Sigma d\beta}_t \qquad (11)$$

where $\underline{i}'$ is an 1 by N vector of ones.

The system of equations in (11) has solution (with initial condition $\underline{Y}(t_0) = \underline{Y}_0$):

$$\underline{Y}(t) = \underline{B}^{-1}\underline{a}\,\underline{i}'(e^{\underline{B}\Delta t} - \underline{I}) + \underline{B}^{-1}\underline{M}(e^{\underline{B}\Delta t} - \underline{I}) + \underline{B}^{-1}\underline{c}(e^{\underline{B}\Delta t} - \underline{I})\underline{x}' + \underline{E}(t), \qquad (12)$$

As before we write (12) more compactly as

$$\underline{Y}(t) = \underline{a}^*\underline{i}' + \underline{M}^* + \underline{B}^*\underline{Y}_0 + c^*\underline{x}' + \underline{E}(t) \qquad (13)$$

And this model differs from (9) due only to the presence of $\underline{M}^*$.

Suppose the model in (11) is correct but the analyst ignores the unobservable variables whose effects are summarized in $\underline{M}^*$. That is, he estimates

$$\underline{Y}(t) = \underline{a}^*\underline{i}' + \underline{B}^*\underline{Y}_0 + \underline{c}^*\underline{x}' + \underline{U}(t) \quad (14)$$

where $\underline{U}(t) = \underline{M}^* + \underline{E}(t)$. (15)

It is easy to show that OLS gives biased estimates of $\underline{B}^*$, $\underline{a}^*$, and $\underline{c}^*$. Since the factors in $\underline{M}$ are constant, they affect $\underline{Y}$ at all times including $\underline{Y}(t_0)$. Thus $\underline{M}^*$ must be correlated with $\underline{Y}_0$. Consequently OLS "gives credit" to Y_0 for the effects in $\underline{M}^*$. This gives biased estimates of $\underline{B}^*$, and thus of all the parameters of the model. And this bias is usually substantial as we illustrate below.

This is an instance of the classic autocorrelation problem raised earlier. When the effects in $\underline{M}^*$ are ignored and thereby forced into the disturbance, the latter must become positively autocorrelated. Failure to acknowledge this, i.e., using estimators that assume $\underline{U}(t)$ is uncorrelated with $\underline{Y}_0$, leads to biased estimation. The usual "two-wave" panel does not contain enough information for this autocorrelation problem to be corrected. But as long as the effects of omitted variables are constant ($\underline{M}(t) = \underline{M}$), this autocorrelation problem is easily handled by a change in research design.

6. Pooled Cross-Section and Time Series Estimator

Biometricians (Henderson 1952) and econometricians (Kuh 1959; Balestra and Nerlove 1966) have proposed estimators for such models in a discrete-time framework. Hannan and Freeman (1978) applied similar estimators to a continuous-time model. Before discussing the estimators, we must address a broad methodological issue: whether the unit-specific components are considered _fixed_ or _random_ effects.

As Searle (1970) notes, the fixed effect perspective fits situations in which all the interest attaches to the units under study and no effort will be made to generalize findings to other units. Then the m_{ij} are considered a set of NK parameters to be estimated. When the units studied are chosen to represent some broader class of units (i.e., some population of units), the random effects perspective is appropriate. Then the proper strategy is to model the general distribution of unit-specific effects and to treat those in the sample as instances of the general process generating unit effects. Then interest focuses on the parameters, not of the units, but of the distribution in the population. The distribution of the m_{ij} typically involves far fewer than NK parameters. Usually we assume that the population distribution is normal. Then it is completely specified by two parameters, the mean and variance.

The choice between the fixed and random perspectives is usually discussed in an experimental design context. Consider for example the income maintenance experiment discussed in earlier chapters. We implemented three levels of income support and four tax rates. For example, we use tax rates of 50%, 70%, and 80%. If there were no scientific or policy interest in any other tax rates, a fixed effects model would be appropriate. However, we wish to generalize findings to other tax rates, e.g., 60%; thus we adopt a random effects perspective. But when interest focuses on discrete alternatives, e.g., research on the effectiveness of several qualitatively different organizational design programs, rehabilitation programs, a fixed effects framework may often be more appropriate.

In this chapter we consider effects of unobserved variables. Should these be treated as fixed or random? Since we cannot even enumerate the factors whose effects are summarized in m_{ij} it seems awkward to treat these as fixed effects. One might still argue that the units were chosen because they have some (unmeasured) properties of special scientific interest and that these properties are summarized in m_{ij}. So the choice among the two perspectives appears once again to turn on the question of whether the units were chosen to be representative of some broad class or whether they were selected because they have some very distinctive property. We suspect that most empirical research in the social sciences comes closer to the former than the latter. If so, the random-effects model is more generally applicable. We have focused on this model in our substantive research. Nonetheless we grant that both models have social science utility and we discuss estimators from each perspective.

It suffices to consider only single equation models as we noted above. Suppose we have measurements on the stochastic process $\{Y_i(t)\}$ at times t_0, $t_1, \ldots, t_T$ and assume that the same stochastic differential equation (1) generates all the observations. We specify the following pooled model:

$$Y_{it} = a^* + m_i^* + b^* Y_{i,t-1} + c^* X + \varepsilon_{it} \qquad (i = 1,\ldots,N;\ t = 1,\ldots,T) \tag{16}$$

7. Fixed Effect Estimators

When the m_i ($i = 1,\ldots,N$) are considered fixed parameters, estimation of (16) is simple. As long as $T \geq 3$, we merely add dummy variables for each unit (i.e., variables that are unity for observations in the unit and zero for observations on all other units). Alternatively we may express observations as deviations from unit means (where $Y_{i\cdot} \equiv \sum_t Y_{it}$):

$$Y_{it} - Y_{i\cdot} = a^* + b^*(Y_{i,t-1} - Y_{i\cdot}) + \varepsilon_{it} - \varepsilon_{i\cdot} \tag{17}$$

and apply ordinary least squares. Under the assumptions of the dynamic model, these OLS estimators are again maximum likelihood. And both estimators are asymptotically unbiased and efficient.

Note that the constant exogenous variable has been lost in (17). In this pooled "within-unit" regression, one cannot estimate both unit-parameters and the effects of exogenous variables that do not vary over time. We do not face such a limitation in the random effects model, where c can indeed be estimated.

The $\hat{m}_i^*$ are recovered from estimates $\hat{a}^*$, $\hat{b}^*$ of (17) by straightforward algebraic operations (see Searle (1970)). These are generalizations of the procedures used to recover the intercept in a conventional linear regression of variables taken as deviations from the (grand) mean. So once we have chosen the proper design, the pooled multi-wave model, no new issues arise in estimating the fixed effects model for unit-specific rates of change.

8. Random Effects Estimators

The alternative perspective considers the m_i to be random variables over units but constants over time. The usual specification is that the m_i are independent and identically distribution from a normal distribution with mean zero and variance $\sigma_{m^*}^2$. We further assume

$$E\left[m_i^* X_{i'}\right] = 0 \qquad \text{all } i, i'$$

$$E\left[m_i^* \varepsilon_{i't}\right] = 0 \qquad \text{all } i, i', t$$

Then the m_i^* have the same properties but are transformed from $N(0,\sigma_m^2)$ to $N(0,\sigma_{m^*}^2)$ where $\sigma_{m^*}^2 = \frac{\sigma^2}{2b}(1-e^{2b\Delta t})$.

Since the M_i^* are unobserved random variables, they may be considered a component of the disturbance for purposes of estimation. To emphasize this fact we write the model as

$$Y_{it} = a^* + b^* Y_{i,t-1} + c^* X_i + u^*_{it} \qquad (18)$$

$$u^*_{it} = m^*_i + \varepsilon_{it}$$

Under our assumptions the disturbance, u^*_{it}, has a normal distribution with mean zero and covariance structure:

$$E\, u_{it}\, u_{i't'} = \begin{cases} (\sigma^2_{m^*} + \sigma^2_*) & \text{if } i = i' \text{ and } t = t' \\ \sigma^2_{m^*} & \text{if } i = i',\ t \neq t' \\ 0 & \text{if } i \neq i' \end{cases} \qquad (19)$$

where σ^2_* is $\frac{\sigma^2}{2b}(1-e^{2b\Delta t})$.

If we arrange observations in (18) so that the first T are from unit 1, the next T from unit 2, etc. The variance-covariance matrix of disturbances has the simple block diagonal form:

$$E(\underline{u}\ \underline{u}') = V \begin{bmatrix} S & & 0 \\ & S & \\ & & \ddots \\ 0 & & & S \end{bmatrix} \qquad (20)$$

with $V = (\sigma^2_{m^*} + \sigma^2_*)$ and each block has the structure: (21)

$$S = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & & \\ \vdots & & \ddots & \\ \rho & \cdots & & 1 \end{bmatrix} \qquad (22)$$

and $\rho = \sigma^2_{m^*} / (\sigma^2_{m^*} + \sigma^2_*)$ (23)

Note that p is the proportion of "error variance" that is unit-specific. That is, it may be considered a measure of the importance of the unit-specific effects relative to the Brownian motion noise process. The parameter p is called the autocorrelation coefficient for the unit-specific effects model. The simple model we are considering holds that units are homogenous in the sense that ρ is constant over units.

Before considering estimators, we consider the systems case. For simplicity we continue to focus on the case where $\underline{x}$ is constant. The model may be written:

$$\underline{y}_t = \underline{Q}_t \underline{\gamma} + \underline{u}_t \tag{24}$$

where

$$\underline{y}_t = (y_{11}, y_{12}, \ldots, y_{1T}, \ldots, y_{N1}, \ldots, y_{NT})'$$

$$\underline{y}_{t-1} = (y_{10}, y_{11}, \ldots, y_{1,T-1}, \ldots y_{N0}, \ldots, y_{N_{T-1}})'$$

$$\underline{x}_1 = (x_{11}, x_{11}, \ldots, x_{11}, \ldots, x_{1N}, \ldots, x_{1N})'$$

$$\vdots$$

$$x_{Jt} = (x_{J1}, x_{J1}, \ldots, x_{JI}, \ldots, x_{JN}, \ldots, x_{JN})'$$

$$\underline{Q}_t = (\underline{i}, \underline{y}_{t-1}\ \underline{x}_1, \ldots, x_J) \quad \text{(where i is an NT x 1 vectors of ones)}$$

$$\underline{u}_t = (u^*_{11}, u^*_{12}, \ldots, u^*_{1T}, \ldots, u^*_{NI}, \ldots, u^*_{NT})'$$

and

$$\underline{\gamma} = (a^*, b^*, c^*_1, \ldots, c_J^*)'$$

At this point it is natural to search for a consistent estimator which avoids the problem in the disturbances. The existence of such an estimator is suggested by the fact that we can transform (24) in such a way as to produce "well-behaved" disturbances. What we need is to find a matrix $\underline{\Omega}$ which when applied to (24) yields

$$\underline{\Omega}^{-1/2}\underline{y}_t = \underline{\Omega}^{-1/2}\underline{Q}_t'\underline{\gamma} + \underline{\Omega}^{-1/2}\underline{u}_t \tag{25}$$

$$E[\underline{\Omega}^{-1/2}\underline{u}_t\underline{u}_t'\underline{\Omega}^{-1/2}] = \sigma^2 I \tag{26}$$

Nothing in the causal structure has been changed and we can apply ordinary least squares to (25). Because of (26) OLS applied to the transformed model is now a consistent and asymptotically efficient estimator.

The procedure suggested in (25) is an application of the widely useful generalized least squares (GLS) approach to estimation. The application of GLS to pooled models is commonly advocated in the econometric and biometric literatures (Nerlove, 1971; Searle, 1971).

Since we will make continued reference to the GLS estimator we need a somewhat more formal representation. The GLS estimators is defined as

$$\underline{\gamma}_{GLS} = (\underline{Q}_t'\underline{\Omega}^{-1}\underline{Q}_t)^{-1}\underline{Q}_t)^{-1}\underline{Q}_t'\underline{\Omega}^{-1}y_t \qquad (26)$$

where

$$\Omega^{-1} = \begin{bmatrix} \underline{S}^{-1} & 0 & \cdots & 0 \\ 0 & \underline{S}^{-1} & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \underline{S}^{-1} \end{bmatrix}$$

and (cf. Hannan and Young, 1974)

$$\underline{A}^{-1} = (1/\eta)(\underline{I}_T - \underline{1}\underline{1}'/\underline{T}) + (1/\xi)(\underline{1}\underline{1}'/T) \qquad (27)$$

where $\eta = (1 - \rho) + T\rho$, and $\underline{1}$ is a $(T \times 1)$ vector of ones.

The form of GLS transformation (27) can be intuitively motivated as follows. The peculiar feature of pooled models is the use of both cross-sectional (between-unit) and longitudinal (within unit) variation to estimate causal parameters. The richness of the data presents an implicit choice, how to weight one type of variation relative to another. Generalized least squares uses ρ to weight the two types of information. To see this, consider the case where $\rho = 0$. Then $\underline{S}^{-1/2} = \underline{I}_T$ and observations are transformed in (25) by an identity transformation. GLS reduces to OLS where cross-sectional and time series variation are weighted proportionately to N and T (see Maddala, 1971). At the other extreme, when $\rho = 1$, $\underline{S}^{-1} = \underline{11}'/t^2$. This transformation averages observations over time for each unit. The result is a regression on grouped observations where all of the weight is placed on cross-sectional variation. In cases where ρ takes on a value $0 \leq \rho \leq 1$, GLS weights time series variation inversely to ρ. Such a weighting seems appropriate since ρ measures redundancy in the time series. The more redundancy, the lower the weight attached to longitudinal variation.

So far we have treated ρ as known a priori. But we know of no realistic cases where sociological researchers have prior knowledge of the value of ρ. Thus we consider methods of estimating ρ and properties of generalized least squares estimators that use estimates of ρ.

The most widely used procedure for estimating ρ uses the fixed-effects estimator discussed in the previous section. The results of LSC can be used to calculate $\hat{\rho}$ as follows. To estimate β we need an estimate of σ^2_{m*}. Nerlove (1971) suggests

$$\sigma^2_m = \sum_{i=1}^{N} (\hat{m}^*_i + \sum_{i=1}^{N} \hat{m}^*_i / N)^2 / N \quad (28)$$

An obvious estimator of $\sigma^2_{m^*} + \sigma^2_*$ is the sum of squared residuals from the LSDV regression divided by NT. Then

$$\hat{\rho} = \hat{\sigma}^2_{m^*}/(\hat{\sigma}^2_{m^*} + \hat{\sigma}^2_*) \quad (29)$$

Nerlove chose $\hat{\rho}$ in (12.29) over a maximum likelihood estimate to avoid negative values of $\hat{\rho}$ (which are implausible in most applications). Unfortunately the estimator in (12.29) is upwardly biased (at least in small samples) with the magnitude of the bias inversely related to ρ.

Recall that GLS requires consistent estimates of ρ. The bias in $\hat{\rho}$ does not, however, appear to unduly damage the resulting GLS estimators (Amemiya, 1967). We study this issue further below. To acknowledge the fact that we are using estimates of ρ rather than the true values, it is more precise to refer to this estimator

$$\hat{\gamma}_{MGLS} = (Q_t' \hat{\Omega}^{-1} Q_t')^{-1} Q_t' \hat{\Omega}^{-1} y_t$$

as modified generalized least squares (MGLS).

This estimator is consistent and asymptotically efficient even though it uses biased estimates of ρ. All that is required for these large sample properties is that $\hat{\rho}$ be a consistent estimator of ρ (Aitken 1934). Empirical researchers are often more concerned with the behavior of estimators in small or moderate size samples. And, the bias in $\hat{\rho}$ may be damaging in such samples. We report results on small sample properties below.

Finally, we may form maximum likelihood estimators for the random effects model. Since the u^*_{it} are joint normally distributed, this amounts to a standard ML regression problem. Estimates of a^*, b^*, $c^*_1, \ldots, c^*_J$, σ^2, and ρ may be found by maximizing the log likelihood function

$$L = -\frac{NT}{2} \log 2\pi - 1/2 \log |\Omega| - 1/2\ u'\ \Omega^{-1} u \ . \qquad (30)$$

Since both σ^2 and ρ must be non-negative, we may maximize (30) subject to this constraint.

The MLE (both unconstrained and constrained) have the good large sample properties (consistency, efficiency) discussed in earlier applications. However, unlike cases discussed to this point, ML is not identical for this case with the best least squares estimator, MGLS. There are three reasons why the two estimators will differ. First, least squares and ML estimates of variance components differ. Second, MGLS is a two round procedure while ML estimates all parameters simultaneously. Finally, as there is no closed-form solution to (30) MLE are found by iteration. Thus the numerical values of MLE depend as well on the shape of the likelihood function and the quality of the iterative procedure.

Thus there are two major alternative approaches to the estimation of dynamic parameters in models with random unit-specific effects: maximum likelihood and generalized least squares. There is actually a third estimator that might be considered. The fixed-effects LSC estimator is also consistent and asymptotically consistent for the random-effects model (Amemiya 1967). Of the three, ML is preferred in large samples for reasons discussed earlier. It retains minimum variance properties under the non-linear transformations required to go from integral to differential equations. But what about smaller samples?

Throughout the discussion we have relied on large-sample theory. As we mentioned earlier, it is important for empirical researchers to obtain some information about the behavior of such estimators in small and moderate sized samples. Two issues are important here: we want to compare the efficiency of the various consistent estimators in finite samples, and we also want to compare the performance of the consistent estimators with those of inconsistent estimators (OLS for example) which may have smaller mean squared error in small samples (cf. Hurd, 1972). We have not yet seen analytical results on these issues. So we consider the results of Monte Carlo experiments on the small sample properties of the various estimators.

9. Monte Carlo Studies of Small Sample Properties

We summarize results from two simulations that used the same structure. The two studies partially overlap but also study some different estimators. We concentrate here on the similar cases so as to give an overall comparison of all the estimators under consideration. For more details see Hannan and Young (1974, 1977) and Tuma and Young (1976).

Data Generation. Both studies generated data that fits the following model:

$$Y_{it} = \gamma_1 Y_{i,t-1} + \gamma_2 X_{it} + u_{it} \tag{31}$$

$$u_{ik} = m_i + \varepsilon_{it}$$

where the components of u_{it} have the properties stated in Section 8. The exogenous variable has the structure:

$$X_{it} = 0.1t + 0.5\,X_{i,t-1} + w_{it}$$

where the w_{it} are independent normal variables. In these respects the simulations followed Nerlove's (1971) procedure. However, they differed from Nerlove's in four respects: First, we have chosen the number of individuals N as fifty and the number of time periods T as five, whereas Nerlove chose twenty-five and ten, respectively. We chose the former values of N and T because they are representative of many available data sets. Second, we have generated pseudo-random variates by Marsaglia's rectangle-wedge-tail algorithm, recommended as best by Knuth (1969), rather than the method described by Nerlove (1971). Third, we have studied somewhat different combinations of parameter values. In each combination we set $a^* = 0.0$ and $\sigma = 1.0$. We selected five values for ρ: 0.0, 0.25, 0.50, 0.75, and 0.90. To examine the dependence of estimator quality on the relative strength of the effects of the lagged endogenous and exogenous variables, we chose three combinations of b and c: $(b,c) = (0.3,1.0)$, $(0.8,1.0)$, and $(0.8,0.5)$. Thus, we examined a total of fifteen combinations of parameter values. Fourth, for each combination of parameter values we generated 100 sets of data, where Nerlove generated 50. The additional data sets give increased confidence about the properties of estimators.

Estimators: We study the behavior of the following estimators:

(1) Ordinary least squares (OLS). A consistent estimator only when $\rho = 0$.

(2) Least squares with constants (LSC), the fixed-effects estimator. Consistent and asymptotically efficient.

(3) "True" generalized least squares (GLS) using known (true) values of ρ. A minimum variance consistent estimator.

(4) Modified generalized least squares (MGLS) with $\hat{\rho}$ calculated as in (29) from an LSC first stage estimator. Consistent and asymptotically efficient.

(5) Maximum likelihood constrained (MLC) with $\sigma \geq 0$ and $0 \leq \rho \leq 1$. Asymptotically unbiased and efficient.

(6) Maximum likelihood unconstrained (MLU): asymptotically unbiased but inefficient relative to MLC.

An initial set of parameter estimates must be provided to find the ML estimates in both methods (5) and (6). We compared the performance of two types of starting values for five different parameter combinations (a total of 500 data sets) using unconstrained ML: the LSDV estimates and the true values used to generate the data. The two types of initial estimates produced nearly identical final estimates for the four combinations in which $\rho > 0$. For $\rho = 0$ the two sets of parameter estimates differed in only a handful of cases, and by a negligible amount. Therefore, because of the cost involved in obtaining the LSDV estimates, we used the true parameter values as starting estimates in all remaining ML estimations. We report only the results obtained from using this latter type of initial estimates.

Whereas Nerlove (1971) used the Fletcher-Powell algorithm (1963) programmed by Wells (1967) to maximize L, Tuma and Young (1976) used the Gill-Murray algorithm (1972) programmed by Wright (1975). Both algorithms are iterative procedures and are based on modified steepest descent methods of function minimization. Gill, Murray and their coworkers (1972a, 1972b) have shown that the Gill-Murray algorithm converges more rapidly and more reliably than the Fletcher-Powell algorithm. However, when both converge, they report that the two algorithms give extremely similar estimates of the function optimum for a variety of functions.

Tuma and Young's (1976) treatment of contraints on parameter values for

σ^2 and ρ departed markedly from Nerlove's (1971). Nerlove constrained σ^2 to be positive by maximizing L with respect to σ rather than σ^2. He imposed a nonnegativity constraint on ρ by equating it with $\sin^2\theta$ and maximizing L with respect to θ rather than ρ. As Nerlove acknowledges, this method of applying constraints causes L to have multiple maxima with respect to θ since $\sin\theta$ is a periodic function. Murray (1972) warns against employment of trigonometric constraints. Such a procedure can increase the nonlinearity of the function being maximized and cause the matrix of second derivatives (which must be negative definite at the maximum of the likelihood function) to become singular or ill-conditioned.

The Gill-Murray algorithm used by Tuma and Young (1976) for ML estimation utilizes a projection method of optimization that permits any feasible equality or inequality constraints to be imposed on parameter values. For a detailed discussion of this constrained optimization procedure, see Gill and Murray (1972). This method does not increase the nonlinearity of the function being optimized or the number of local maxima.

To our knowledge there is no previous evidence indicating the magnitude of the effects of constraining $\hat{\rho}$ and $\hat{\sigma}^2$ on ML parameter estimates for the model we have simulated. Thus, we do not know whether the mean squared errors (MSE's) of the constrained estimates of ρ and σ^2 will be appreciably smaller than the unconstrained versions. Further, we do not know the effects of constraining $\hat{\rho}$ and $\hat{\sigma}^2$ on the quality of the estimates of γ_1 and γ_2. Finally, it is important to learn whether the poor performance of the ML method in Nerlove (1971) results from the small-sample properties of ML estimation of this model or from the implementation of parameter constraints.

Results: Before looking at mean squared error and bias of estimators we comment on the effectiveness and practicality of the maximum likelihood procedure used. This issue has heightened importance in the present context as Nerlove (1971) in a very influential paper reports that MLE failed to converge on most occasions and thus did not stand as serious practical alternatives to MGLS. Tuma and Young (1976) that implementation of the ML methods was both successful and practical. Not only did ML estimation converge to a solution for every data set, but also the time required for this was short. On the average the ML solution was found in four to ten iterations, depending on the particular combination of parameter values. The MLC and MLU methods required nearly identical numbers of iterations to converge. For both methods several more iterations were usually needed for high values of ρ, especially when ($\gamma_1 = 0.8, \gamma_2 = 0.5$). These higher numbers of iterations occur together with poor quality of the ML estimates of β_1, σ^2 and ρ, as described more fully later in this section. It is helpful to know which parameter combinations led to activation of constraints. Obviously for the cases in which no constraints were activated, the MLC and MLU estimates are identical. The constraints that σ^2 be positive and that ρ be less than or equal to one were never brought into play (cf. Nerlove 1971). However, the constraint that ρ be nonnegative was activated in about sixty percent of the cases in which $\rho = 0$ or $\gamma_1 = 0.8$, $\gamma_2 = 1.0$; and $\rho = 0.90$. Thus the quality of MLC and MLU estimators is unlikely to differ except for these combinations of parameters.

We begin our assessment[2] of the quality of the stimators by comparing overall mean squared errors averaged over the five choices of ρ. These are reported in Table 1. Overall the OLS and LSC estimators

Table 1 about here

are inferior as we expected. The simulation results agree with the large sample theory in that OLS has largest MSE in each case. Moreover, these two estimators are notably poor in estimating γ_1, and, as we have noted repeatedly, this failure has serious consequences in analysis of continuous-time models. On the basis of these results and the further evidence in Hannan and Young (1977) we advise against use of OLS and LSC for random-effects models. Henceforth we direct attention only to the MGLS and ML estimators.

The relative quality of the ML and MGLS estimates varies according to the size of the ratio of γ_1, the coefficient of the lagged endogenous variable, to γ_2, the coefficient of the exogenous variable. We find that ML is superior when the effect of the lagged endogenous variable is small in comparison to the effect of the exogenous variable, while MGLS is best when the opposite is true. As we report below, the dependence of the relative quality of the ML and MGLS estimates of regression coefficients on the relative effects of γ_1 and γ_2 becomes even more apparent when the simulation results are not aggregated over values of ρ.

We now turn our attention to a more detailed examination of the performance of the MLU and MLC estimates, contrasted to each other and to the best of the least squares methods, MGLS. We use the measure:

$$\% \text{ bias}(\hat{\theta}) = 100\% * \text{bias}(\hat{\theta})/\theta.$$

For both γ_1 and γ_2 the % biases of the MLC and MLU estimates are very similar across all parameter combinations (see Tables 2 and 3 respectively).

Tables 2 and 3 about here

Both the ML and MGLS methods display consistently low % biases in $\hat{\gamma}_2$ across all parameter combinations. However, both methods of estimation produce widely varying % biases in $\hat{\gamma}_1$. For each combination of γ_1 and γ_2 the % bias in ML estimates of γ_1 tends to become worse as ρ increases. However, for the first two combinations of γ_1 and γ_2 there is a downturn in the % bias for very high values of ρ. On the other hand, the MGLS estimates of γ_1 are downwardly biased for low values of ρ but the % bias increases monotonically as ρ increases, approaching a negligible % bias for $\rho = 0.9$.

Of course, the MSE's reported earlier also depend on the variances of estimators. However, there are only slight differences between the two ML and MGLS in variances. For both types of estimators the variance falls off sharply as ρ increases. As this is the only interesting pattern in the variances we do not report the actual figures (see Tuma and Young 1976, Tables 4 and 5).

Finally we look at estimates of ρ. In both MLU and MLC estimates of ρ the biases are usually negative and very similar (see Table 4).

Table 4 about here

The magnitude of the bias in $\hat{\rho}$ is somewhat smaller for MLC than for MLU when ($\gamma_1 = 0.8$, $\gamma_2 = 0.5$, $0.25 \leq \rho \leq 0.75$) but slightly larger for MLC than for MLU when $\rho = 0.0$. The size of the bias in ML estimates of ρ tends to increase as ρ increases; however, for two of the three combinations of the regression coefficients, there is a downturn in the bias in $\hat{\rho}$ as the value of ρ becomes very large.

The ML and MGLS methods perform optimally at opposite ends of ρ continuum. Whereas ML estimates of ρ are almost always downwardly biased, the MGLS estimates of ρ are almost always upwardly biased. And, as we found in our examination of the % biases of $\hat{\gamma}_1$ the performance of the ML method tends to be best when MGLS is at its worst, and vice versa. Thus, we find that while the bias in ML estimates of ρ is greatest for high values of ρ and least for low values of ρ, just the opposite is true for MGLS. The MGLS estimates of ρ are most biased when ρ is near zero and least biased when ρ is near unity.

Nonetheless, the ML and MGLS methods have two obvious similarities: (1) there is an inverse relationship between % bias in $\hat{\gamma}_1$ and bias in $\hat{\rho}$, and (2) absolute values of the biases in $\hat{\gamma}_1$ and $\hat{\rho}$ are positively associated. These similarities are curious becuase the ML and MGLS methods have opposite signs to the biases of their estimates of ρ and of γ_1. Though the two methods differ dramatically in their tendencies to attribute stability in the dependent variable to serial correlation of residuals for individual units rather than to inertia in the dependent variable, for both methods there are compensating effects. That is, for both methods error in one direction in estimating the strength of serial correlation of residuals is accompanied by error in the opposite direction in estimating the strength of the lagged endogenous variable.

We conclude that both ML and MGLS perform relatively well with panel data of the size usually available to sociologists ($N = 50$, $T = 5$). They clearly outperform OLS and LSDV. It appears that MGLS does best when γ_1 is small. This implies that MGLS has best small sample properties when systems under study adjust rapidly relative to the time scale chosen (or, under the alternative interpretation, have strong negative feedback). On the other hand, MLE appear preferable for systems that adjust more slowly. In light of previous work on these issues, perhaps the most important conclusion is that both ML and MGLS are practical and appear to have good small sample properties.

We also provide at least a partial answer to the question: Should natural constraints on parameters be imposed? Tuma and Young (1976) find, as did Nerlove (1971), that in practice only the nonnegativity constraint on ρ is at issue because other natural constraints are never violated. These results show that in terms of the mean squared error of $\hat{\gamma}_1$ and $\hat{\gamma}_2$, ML estimation with constraints on ρ has a slight advantage over that without constraints. Clearly constrained ML estimation gives more reasonable estimates of ρ because it prevents $\hat{\rho}$ from having a negative value, which is contrary to the assumptions of the model. In addition, the constrained ML estimates of the regression coefficients always have a smaller variance than the unconstrained ones, and this usually compensates for occasionally larger biases in the constrained estimates. Still, the differences between the constrained and unconstrained ML estimates are never large--and always negligible for those parameter combinations in which ML estimates are superior in quality to MGLS estimates. Consequently, this research provides no evidence that omitting constraints on ρ will seriously damage the quality of ML estimates of regression coefficients in the model.

10. Unequally Spaced Observations

To this point, we have assumed equally spaced panel observations. As long as waves in the panel are repeated with constant period for all units, several approaches to estimation have merit. We saw that two broad strategies have been proposed. Within each strategy, several estimators have good properties. But once we venture beyond this standard design to consider unequal spacing, we face greatly limited alternatives. In fact only one strategy and one estimator appear feasible: maximum likelihood applied to integral equations.

Two classes of designs may yield unequally spaced data. The first is the conventional multi-wave panel where the length of lags between waves varies but is the same for all units. In field research such variability in the timing of waves may arise from the vagaries of flows of research funds, problems of entry into sites, renewed interest in some earlier panel, etc. For example, Meyer's (1975) three-wave panel of finance agencies has a three-year lag between the first two waves and a six-year lag between the second and the third. The widely analyzed Sewell (see, for example, Hauser and Sewell 1975) panel of Wisconsin high school seniors was interviewed in 1957, 1964, and 1976. Exactly the same sorts of problems arise in archival research since official sources often release data at intermittent intervals. Moreover, researchers using secondary sources must often depend on the timing of several scholars or groups of scholars. They are this often confronted with unequally spaced data.

The second, perhaps more important, problem concerns timing that varies from unit to unit. This problem may also arise for the reasons discussed above. Some individuals may be "lost" to a panel and only recovered at some later time. However, there is a more systematic reason why the timing of observations may vary among units. Panel observations may vary among units. Panel observations may be linked to events that are generated by a stochastic process. Sometimes this is done within a retrospective design. For example, the Parnes (1975) "mature woman panel" contains work histories at marriage, at first birth, etc. Since different women have different timing of events, the panel will have extreme unequal spacing.

We argue that sociologists ought to study coupled changes in qualitative and quantitative outcomes (e.g. marital status and earnings). One fruitful approach to such systems involves studying changes in quantitative variables over periods that begin and end with events (changes in state or qualitative variables). We surmise that, if we are to make progress on the important class of problems that involve coupled changes in quantity and quality, we must solve the problem of analyzing unequal spaced panel data.[3]

The first type of unequal spacing is usually dealt with by analyzing pairs of waves separately. But this is an unsatisfactory solution in many instances since it obviates the possibility of adjusting for unit-specific disturbances. In most sociological applications, such a failure makes a real difference in estimates of the parameters of the underlying continuous-time model from different lags. Should we be tempted as a consequence to treat the data as generated by several discrete-time processes with different lags, there is another problem. As we showed above, there is no metric available to compare results from different lags when the process is viewed as discrete in time. Thus the analyst cannot draw sound inferences about stability or change in the process. He has not one but two or more processes. Thus there is a tremendous loss of generality. So the analyst with panel data with the simpler form of unequal spacing faces two unhappy alternatives: report different estimates of one process (with the fear that the differences reflect only autocorrelation bias) or report estimates of several discrete-time processes for the same substantive problem (where the lag structures are determined by the peculiarities of the research design).

We have not seen any analysis of data of the more extreme type of unequal spacing that pays attention to these methodological problems. Moreover, we have not yet found any systematic treatment of the general problem. So, despite its obvious practical importance and its possible substantive importance, the issue of how to estimate models from unequally spaced data has received surprisingly little attention. We attribute this lacuna to the common preoccupation with discrete-time models in the social sciences. We now show that shifting to a continuous-time perspective suggests solutions to unequal spacing problems.

10. MLE for Unequally Spaced Observations

The simplest case concerns the linear stochastic differential equation with no unit-specific components:

$$dY(t) = a\,dt + bY(t)dt + cX(t)dt + \sigma d\beta_t$$

where β_t is a normal Brownian motion. (1) has solution (subject to initial conditions $Y(t_0) = Y_0$, $X(t_0) = X_0$ and assuming $X(t)$ changes linearly over Δt)

$$Y_i(t) = \frac{a}{b}(e^{b\Delta t_i} - 1) + e^{b\Delta t_i}Y_{i0} + \frac{c}{b}(e^{b\Delta t_i}-1)X_{i0} + \frac{c}{b}\left(\frac{e^{b\Delta t_{i-1}}}{b\Delta t_i} - 1\right)\Delta X(t) + \sigma\int_0^{t_i} e^{b(t-s)}d\beta_s \tag{33}$$

Because Δt_i varies, each unit has its own set of parameters in the integrated form. Let us rewrite (33) as

$$Y_i(t_i) = a_i^* + b_i^* Y_{i0} + c_i^* X_{i0} + d_i^* \Delta X_i(t_i) + \varepsilon_i^* \tag{34}$$

We know that ε_i^* is $N(0, \sigma_i^2)$ where

$$\sigma_i^2 = \frac{\sigma^2}{2b}(1-e^{-2b\Delta t_i})$$

Consequently we may write the likelihood function

$$\mathcal{L}(a,b,c,\sigma;\ \text{data}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left[\frac{\varepsilon_i^2}{2\sigma_i^2}\right] \quad (35)$$

or

$$\log \mathcal{L} = L = \sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi\sigma_i}} + \sum_{i=1}^{N} \frac{\varepsilon_i^2}{2\sigma_i^2} \quad (36)$$

where σ_i is defined in (35). But under the model,

$$\varepsilon_i^2 = [Y_i(t) - a_i^* - b_i^* Y_{i0} - c_i^* X_{i0} - d_i^* \Delta X(t_i)]^2 \quad (37)$$

and, see (32), a_i^*, b_i^*, c_i^* and d_i^* are explicit functions of the dynamic parameters of interest.

Since the Δt_i are observed data, this likelihood may be maximized with respect to a, b, c, and σ. This requires writing out first and second derivatives of (36) with respect to these parameters and using these expressions in one of the standard iterative routines. We have adapted the Gill-Murray algorithm used to estimate rates, for this purpose. We will illustrate the procedure below.

Suppose one has reason to believe that each unit changes in response to unobserved constant factors as discussed above. Then the model is

$$dY_i(t) = a\,dt + bY(t)dt + cX(t)dt + m_i dt + \sigma d\beta_t \quad (38)$$

and subject to the same conditions stated above, has solution

$$Y_i(t_i) = \frac{a}{b}(e^{b\Delta t_i} - 1) + e^{b\Delta t_i} Y_{i0} + \frac{c}{b}(e^{b\Delta t_i} - 1)X_{i0} + \frac{c}{b}\left[\frac{e^{b\Delta t_i} - 1}{b\Delta t_i} - 1\right]\Delta X_i(t_i)$$

$$+ \frac{m_i}{b}(e^{b\Delta t_i} - 1) + \sigma \int_0^{t_i} e^{b(t_i - s)} d\beta_t \quad (39)$$

or $\quad Y_i(t_i) = a_i^* + b_i^* Y_{i0} + c_i^* X_{i0} + d_i^* \Delta X_i(t_i) + u_i(t_i) \quad (40)$

where $u_i(t) = \frac{m_i}{b}(e^{b\Delta t_i} - 1) + \sigma \int_0^{t_i} e^{b(t_i - s)} d\beta_t$

$$= m_i^* + \varepsilon_i(t_i) \quad (41)$$

One cannot identify σ_m^2 from only two waves of observations. However, when three or more waves are available on each unit, all the dynamic parameters may be identified.

Let us consider the general case where the number of observations varies from unit to unit. We denote the number of observations on unit i by T_i and let $\Delta t_{ij} = t_{i,j+1} - t_{ij}$ where t_{ij} is the jth observation on the i^{th} unit. Then we write a pooled model as follows:

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1T_1} \\ Y_{21} \\ \vdots \\ Y_{NT_N} \end{bmatrix} = \frac{a}{b} \begin{bmatrix} e^{b\Delta t_{11}}-1 \\ e^{b\Delta t_{12}}-1 \\ \vdots \\ \vdots \\ e^{b\Delta t_{NT_N}}-1 \end{bmatrix} + \begin{bmatrix} e^{b\Delta_{11}}Y_{10} \\ e^{b\Delta_{12}}Y_{11} \\ \vdots \\ e^{b\Delta_{1T_1}}Y_{1,T_1-1} \\ \vdots \\ e^{b\Delta_{NT_N}}Y_{N,T_N-1} \end{bmatrix} + \frac{c}{b} \begin{bmatrix} (e^{b\Delta_{11}}-1)X_{10} \\ \vdots \\ \vdots \\ (e^{b\Delta_{N,T_N}}-1)X_{N,T_N-1} \end{bmatrix}$$

$$+ \frac{c}{b^2} \begin{bmatrix} \left[\frac{e^{b\Delta_{11}}-1}{\Delta t_{11}} - 1\right]\Delta X_{11} \\ \vdots \\ \vdots \\ \left(\frac{e^{b\Delta t_{10,T_N}}-1}{\Delta t_{N,T_N}} - 1\right)\Delta X_{N,T_N} \end{bmatrix} + \begin{bmatrix} u_{11} \\ \vdots \\ \vdots \\ u_{N,T_N} \end{bmatrix} \tag{42}$$

Where

$$\begin{bmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{N_1 T_N} \end{bmatrix} = \begin{bmatrix} \frac{m_1}{b}\,(e^{b\Delta t_{11}}-1) \\ \frac{m_1}{b}\,(e^{b\Delta t_{12}}-1) \\ \vdots \\ \frac{m_N}{b}\,(e^{b\Delta t_{N,T_N}}-1) \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{N,T_N} \end{bmatrix}$$

And, "disturbance" vector of the integrated form is $N(\underline{0}, \underline{\Sigma})$ with

$$\underline{\Sigma} = \begin{bmatrix} S_1 & & & \\ & S_2 & & 0 \\ & & \ddots & \\ & 0 & & S_N \end{bmatrix}$$

Where $\underline{\Sigma}_i$ is a T_i by T_i matrix with

$$\underline{\Sigma}_i = \begin{bmatrix} \frac{\sigma^2(1-e^{-2b\Delta t_1})}{2b} & \sigma_m^2(e^{b\Delta t_{i1}}-1) & \cdots & \sigma_m^2(e^{b\Delta t_{i1}}-1)(e^{b\Delta t_{i1}}-1) \\ & \frac{\sigma^2(1-e^{-2b\Delta t_2})}{2b} & & \\ & & \ddots & \\ \sigma_m^2(e^{b\Delta t_{iT_i}}-1)(e^{b\Delta t_{i1}}-1) & \cdots & & \frac{\sigma^2(1-e^{-2b\Delta t_1})}{2b} \end{bmatrix}$$

Now $\underline{\Sigma}$ may be expressed as a function of observable variables Y_{-1}, X_{-1}, ΔX as above. Since $\underline{\Sigma}$ is normal, we may write an explicit likelihood function and estimate parameters by standard iterative algorithms. Rather extensive programming is required, however, to make this scheme operational. Our research group is currently conducting this work in preparation for application of these methods to sociological data.

11. Conclusions

The thrust of this report has been to show that we can use available methods to solve many of the practical problems that arise in apply continuous-time, continuous-state models in sociological research. In part we have shown that sociologists have begun to estimate implicitly (by use of normal theory assumptions with deterministic models) linear change models driven by Brownian motion. Only a slight change in perspective is required for the usual estimates to be transformed into estimates of a simple probabilistic model for change in quantitative variables.

We have devoted considerable attention to the likely problem of autocorrelation. We suggest that a combination of Brownian motion disturbances and unit-specific permanent effects may apply meaningfully to a variety of sociological analyses. If so, modest extensions of available estimators for pooled cross-section and time series data may be used profitably. We showed that both generalized least squares and maximum likelihood estimators have good properties for sample sizes typically used by sociologists.

Finally, we illustrated one of the major advantages of continuous-time modeling of social processes: the ability to handle data collected with unequal spacing in time. The maximum likelihood estimators we discuss may be extended to this case in a straightforward, though tedious, way.

FOOTNOTES

[1]Such studies are most common in the engineering literature, where the term "filtering" is used instead of "estimation"--see Jazwinski (1970) for a review.

[2]The tables that follow contain excerpts from various tables in Hannan and Young (1977) and Tuma and Young (1976). Both reports contain considerable additional detail.

[3]In many cases, both quantitative and qualitative outcomes will be measured in the same interviews. Singer and Spilerman (1976) have demonstrated that panel studies of discrete (i.e., qualitative) stochastic processes should not use a constant lag between waves, but should be irregularly spaced. If this advice is followed, the quantitative records from such interviews will have the structure we discuss here.

REFERENCES

Amemiya, T.

1967 "A note on the estimation of Balestra-Nerlove models." Technical Report No. 4, Institute for Mathematical Studies in the Social Sciences, Stanford University.

Balestra, P. and M. Nerlove

1966 "Pooling cross-section and time series data in the estimation of a dynamic model: the demand for natural gas." Econometrica 34: 585-612.

Bergstrom, A. R. (ed.)

1976 Statistical Inference in Continuous-Time Economic Models. Amsterdam: North-Holland.

Coleman, J. S.

1968 "The mathematical study of change." Pp. 428-78 in H. M. Blalock, Jr. and A. Blalock (eds.), Methodology in Social Research. New York: McGraw-Hill.

Fletcher, P. and M. J. D. Powell

1963 "A rapidly convergent descent method for minimization." Computer Journal 6: 163-68.

Gill, P. F. and W. Murray

1972a "The implementation of two modified Newton algorithms for unconstrained optimization." National Physical Laboratory Report NAC 24.

1972b "Two methods for the solution of linearly constrained and unconstrained optimization problems." National Physical Laboratory Report NAC 25.

Hannan, M. T.

1978a "Models for change in quantitative variables, part I: deterministic models." Technical Report No. 63, Laboratory for Social Research, Stanford University.

1978b "Models for change in quantitative variables, part II: stochastic models." Technical Report No. 64, Laboratory for Social Research, Stanford University.

Hannan, M. T. and J. Freeman

1978 "Internal politics of growth and decline." Pp. 177-99 in M. Meyer et al. (eds.), Environment and Organization. San Francisco: Jossey-Bass.

Hannan, M. T. and A. A. Young

1974 "Estimation of pooled cross-section and time series models: preliminary Monte Carlo results." Paper presented at Conference on Policy Research in Education: Methods and Implications, University of Wisconsin, Madison.

1977 "Estimation in panel models: results on pooling cross-sections and time series." Pp. 52-83 in D. Heise (ed.), Sociological Methodology 1977. San Francisco: Jossey-Bass.

Hauser, R. M. and W. Sewell

1975 Education, Occupation, and Earnings. New York: Academic Press.

Henderson, C. R.

1952 "Specific and combining ability." Pp. 352-70 in J. W. Gowens (ed.), Heterosis. Ames, Iowa: Iowa State College Press.

Hurd, M. D.

1972 "Small sample estimation of a structural equation with auto-correlated errors." Journal of the American Statistical Association 67: 567-73.

Johnston, J.

1972 Econometric Methods, 2nd edition. New York: McGraw-Hill.

Knuth, D. E.

1969 The Art of Computer Programming, Vol. 2. Cambridge, Mass.: Addison-Wiley.

Kuh, F.

1959 "The validity of cross-sectionally estimated behavior equations in time series applications." Econometrica 27: 197-214.

Meyer, M.

1975 "Leadership and organizational structure." American Journal of Sociology 81: 514-41.

Murrary, W. (ed.)

1972 Numerical Methods for Unconstrained Optimization. New York: Academic Press.

Nerlove, M.

1971 "Further evidence on the estimation of dynamic economic relations from a time series of cross-sections." Econometrica 39: 359-82.

Parnes, H. S.

1975 "The national longitudinal surveys: new vistas for labor market research." American Economic Review 65: 244-49.

Searle, S. R.

1971 "Topics in variance components estimation." Biometrics 27: 1-76.

Singer, B. and S. Spilerman

1976 "The representation of social processes by Markov models." American Journal of Sociology 82: 1-54.

Tuma, N. B. and A. A. Young

1976 "Constrained and unconstrained maximum likelihood estimation of a variance components model of cross-sections pooled over time." Technical Report No. 60, Laboratory for Social Research, Stanford University.

Wells, M.

1967 "Function minimization." Algorithm 251 in Collected Algorithms from CACM. New York: Association for Computing Machinery.

Wright, M.

1975 "LCMNA: a set of FORTRAN subroutines for minimization of a set of non-linear functions subject to linear inequality or equality constraints." Stanford Computation Center Documentation, Stanford University.

Table 1

Mean Squared Error of Estimates

(Cases averaged over all values of ρ; each entry based on 500 sets of estimates)

	$\hat{\gamma}_1$	$\hat{\gamma}_2$
$b^* = 0.3$, $c^* = 1.0$		
OLS	6.449	.348
LSC	.822	.239
MGLS	.226	.194
MLC	.169	.182
MLU	.175*	.182
$b^* = 0.8$, $c^* = 1.0$		
OLS	1.592	.341
LSC	.748	.228
MGLS	.146	.198
MLC	.720	.199
MLU	.722	.199
$b^* = 0.8$, $c^* = 0.5$		
OLS	2.420	.228
LSC	3.865	.220
MGLS	.925	.194
MLC	2.352	.208
MLU	2.415	.218

* All entries in this table have been multiplied by 10^2.

Table 2

Percent Bias in γ_1

(Each entry based on 100 sets of estimates)

ρ = :	0.0	0.25	0.5	0.75	0.9
$\gamma_1 = 0.3,\ \gamma_2 = 1.0$					
MLU	0.0%*	3.7	3.9	2.3	0.9
MLC	-1.2	3.7	3.9	2.3	0.9
MGLS	-22.4	-12.2	-5.9	-1.5	-0.1
$\gamma_1 = 0.8,\ \gamma_2 = 1.0$					
MLU	0.2%	9.3	13.6	14.2	6.4
MLC	-0.8	9.2	13.6	14.2	6.4
MGLS	-7.3	-2.4	.3	1.9	1.7
$\gamma_1 = 0.8,\ \gamma_2 =$ [illegible].5					
MLU	-0.4	16.7	21.4	23.4	23.9
MLC	-1.5	15.8	21.0	23.3	23.9
MGLS	-19.0	-12.4	-7.7	-1.8	1.9

*All entries in this table have been rounded off to the nearest tenth of a percent.

Table 3

Percent Bias in γ_2

Percent Bias

(Each entry based on 100 sets of estimates)

ρ = :	0.0	0.25	0.5	0.75	0.9
$\gamma_1 = 0.3$, $\gamma_2 = 1.0$					
MLU	-0.0%*	-0.0	0.1	0.1	-0.0
MLC	0.1	-0.0	0.1	0.0	-0.0
MGLS	-0.3	-0.4	-0.3	-0.2	-0.1
$\gamma_1 = 0.8$, $\gamma_2 = 1.0$					
MLU	0.0	-1.6	-1.6	-0.4	0.4
MLC	0.1	-1.6	-1.6	-0.4	0.4
MGLS	-0.1	-0.3	-0.2	0.0	0.0
$\gamma_1 = 0.8$, $\gamma_2 = 0.5$					
MLU	0.0	-2.5	-3.1	-2.8	-1.5
MLC	0.2	-2.0	-2.6	-2.6	-1.5
MGLS	-0.9	-1.4	-1.1	-0.5	0.0

*All entries in this table have been rounded off to the nearest tenth of a percent.

Table 4

Bias of $\hat{\rho}$

(Each entry based on 100 sets of estimates)

ρ = :	0.0	0.25	0.5	0.75	0.9
$\gamma_1 = 0.3$, $\gamma_2 = 1.0$					
MLU	-.006	-.040	-.048	-.025	-.008
MLC	.017	-.040	-.048	-.025	-.008
MGLS	.254	.215	.145	.064	.021
$\gamma_1 = 0.8$, $\gamma_2 = 1.0$					
MLU	-.007	-.169	-.345	-.398	-.101
MLC	.016	-.166	-.344	-.398	-.101
MGLS	.325	.320	.219	.092	.027
$\gamma_1 = 0.8$, $\gamma_2 = 0.5$					
MLU	-.005	-.273	-.519	-.728	-.766
MLC	.017	-.242	-.491	-.718	-.766
MGLS	.445	.477	.340	.156	.047