DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

ESTIMATION OF DYNAMIC MODELS WITH ERROR COMPONENTS

T. W. Anderson
Stanford University and
California Institute of Technology

Cheng Hsiao
University of Toronto

Prepared under National Science Foundation
Grants SES79-13976, SES80-07576.

SOCIAL SCIENCE WORKING PAPER 336

August 1980

ESTIMATION OF DYNAMIC MODELS WITH ERROR COMPONENTS*

by

T.W. Anderson and Cheng Hsiao**

1. Introduction

Error components models are widely used in the econometric analysis of cross section and time series data; it is a common practice to assume that the large number of factors which affect the individuals in the sample and the values of the dependent variable observed for them, but which have not been explicitly included as independent variables, may be appropriately summarized by random disturbances. Wallace and Hussain [1969] and Swamy and Arora [1972] have analyzed this type of model when no lagged dependent variables appear as explanatory variables.

Very often we would like to use such a model to study behavioral relationships that are dynamic in character (Balestra and Nerlove [1966]). As it turns out, the problem becomes complicated. Amemiya [1969] and Balestra and Nerlove [1966] have proved the consistency of the maximum likelihood estimator when the length of the time series T tends to infinity within this context. Maddala [1971] has investigated some aspects of the applicability of "covariance techniques." Nerlove [1971] has performed Monte Carlo studies to explore the small-sample properties of various types of

---

*This work was supported by National Science Foundation Grants SES79-13976 and SES 80-07576 at the Institutue for Mathematical Studies in the Social Sciences, Stanford University and Social Sciences and Humanities Research Council of Canada Grant 410-80-0080 at the Institute for Policy Analysis, University of Toronto. This technical report was completed while the first author was a Sherman Fairchild Distinguished Scholar at the California Institute of Technology. The authors are indebted to James Powell for assistance in preparing this paper.

**University of Toronto.

estimates. However, we feel that more attention should be paid to the properties of various estimators in the way in which the time series observations T or the cross sectional units N tends to infinity. In particular more attention should be paid to the typical case where a panel involves a large number of individuals, but only over a short period of time. In this case the relevant limiting distributions have the number of individuals increasing but not the time dimension. Hence, contrary to the dynamic model for a single time series, the assumption about the initial observations plays a much more crucial role in interpreting the model and devising consistent estimates. In this paper we hope to clarify these issues. We shall consider a number of different models arising from different assumptions about the initial observations. The focus will be on (i) the interpretation of the model, and (ii) the asymptotic properties of the estimators under various assumptions.

In Section 2 we consider the interpretations of models under various assumptions and introduce the maximum likelihood and the covariance estimators. In Section 3 we consider the properties of the maximum likelihood and covariance estimators when the initial observations are assumed as fixed constants. Section 4 considers the case of random initial observations with a stationary distribution. Section 5 considers the case of random initial observations with different means, Section 6 with a common mean. Section 7 clarifies the relationship between pseudo and conditional maximum likelihood estimates. Section 8 suggests simple consistent estimators which have the advantage of being independent of the assumption of initial observations. Conclusions are given in Section 9.

2. The Model

A simple model commonly used in the empirical research of a sample consisting of time series observations on a cross section is of the form[1/]

$$y_{it} = \underset{\sim}{z}_{it}\underset{\sim}{\beta} + v_{it} \quad , \quad i = 1,\ldots,N \quad , \quad t = 1,\ldots,T \quad , \tag{2.1}$$

where

$$v_{it} = \alpha_i + u_{it} \quad ,$$

$$E\alpha_i = Eu_{it} = 0 \quad ,$$

$$E\alpha_i u_{jt} = 0 \quad ,$$

$$E(\alpha_i\alpha_j) = \begin{cases} \sigma^2 & \text{if } i = j \quad , \\ 0 & \text{otherwise} \quad , \end{cases} \tag{2.2}$$

$$E(u_{it}u_{js}) = \begin{cases} \lambda\sigma^2 = \sigma_u^2 & \text{if } i = j \ , \quad t = s \quad , \\ 0 & \text{otherwise} \quad , \end{cases}$$

$\underset{\sim}{z}_{it}$ is an $1 \times k$ vector of explanatory variables, $\underset{\sim}{\beta}$ is a $k \times 1$ vector of parameters to be estimated. We are interested in cases where $T \geq 2$, $N \geq 2$. Let

$$\underset{T\times 1}{\underset{\sim}{y}_i} = (y_{i1},\ldots,y_{iT})' \quad ,$$

$$\underset{T\times K}{\underset{\sim}{Z}_i} = (z'_{i1},z'_{i2},\ldots,z'_{iT})' \quad ,$$

$$\underset{T\times 1}{\underset{\sim}{v}_i} = (v_{i1},\ldots,v_{iT})' \quad ,$$

$$\underset{T\times 1}{\underset{\sim}{u}_i} = (u_{i1},\ldots,u_{iT})' \quad ,$$

$$\underset{T\times 1}{\underset{\sim}{e}} = (1,\ldots,1)' \quad .$$

We can rewrite (2.1) as

$$\underset{\sim}{y}_i = \underset{\sim}{Z}_i\underset{\sim}{\beta} + \underset{\sim}{v}_i \quad , \quad i = 1,\ldots,N \quad . \tag{2.3}$$

Premultiplying (2.3) by

$$\underset{\sim}{Q} = \underset{\sim}{I}_T - \frac{1}{T}\underset{\sim}{e}\underset{\sim}{e}' \tag{2.4}$$

we obtain the covariance estimator (CV) of $\underset{\sim}{\beta}$ as

$$\hat{\underset{\sim}{\beta}}_{CV} = [\sum_{i=1}^{N} \underset{\sim}{Z}'_i\underset{\sim}{Q}\underset{\sim}{Z}_i]^{-1} \sum_{i=1}^{N} \underset{\sim}{Z}'_i\underset{\sim}{Q}\underset{\sim}{y}_i \quad . \tag{2.5}$$

If we assume that α_i and u_{it} are normally distributed, we can write down the exact likelihood function of $\underset{\sim}{y}_i$ from the density function of $\underset{\sim}{v}_i$. Maximizing the logarithm of the likelihood function, we obtain the maximum likelihood estimator (MLE) of $\underset{\sim}{\beta}$, σ^2, and λ.

If $\underset{\sim}{Z}_i$ are exogenous, Wallace and Hussain [1969] have proved that the CV is consistent and asymptotically normally distributed when N tends to infinity or T tends to infinity or both. Furthermore, the

CV is asymptotically equivalent to the MLE as long as T tends to infinity (and N is fixed or tends to infinity).

When z_{it} contains endogenous variables, the problem becomes more complicated. Not only may the CV and the MLE be inconsistent, but the interpretation of the model is also not independent of our assumption about the initial conditions. In this paper we shall focus on (1) the interpretation of the model and (2) the asymptotic properties of the estimators for a dynamic model. We shall assume that z_{it} consists of $y_{i,t-1}$ only (namely $k = 1$) because the principle of analysis remains the same, yet the presentation can be greatly simplified. Therefore, instead of (2.1) we shall analyze

(2.6) $$y_{it} = \beta y_{i,t-1} + \alpha_i + u_{it} \quad , \quad i = 1,\ldots,N \quad , \quad t = 1,\ldots,T \ .$$

We also assume that $|\beta| < 1$ and that the mean of the observed variable is known (and taken equal to zero).2/

Based on different assumptions about the initial observations y_{i0} in (2.6) we have essentially three different types of the model. The first type of the model is a conventional one where we assume that y_{i0} are observed fixed constants (Amemiya [1967], Balestra and Nerlove [1966], etc.). This assumption permits a cross-sectional unit starting at some y_{i0} and gradually moving towards a level of $[\alpha_i/(1 - \beta)]$ (Figure 1). To see this, we can rewrite (2.6) in an equivalent form of

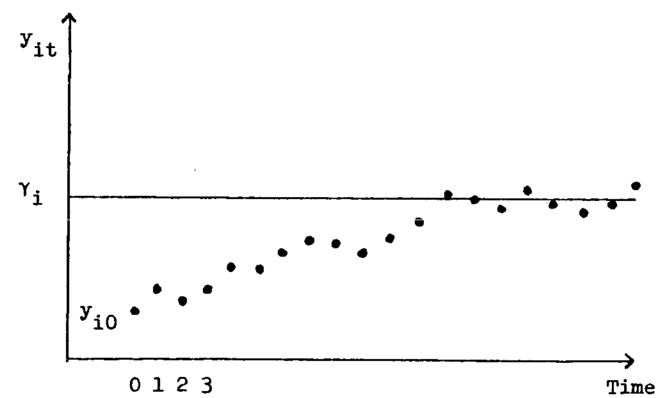
(2.7) $$(y_{it} - \gamma_i) = \beta(y_{i,t-1} - \gamma_i) + u_{it} \quad ,$$

where

(2.8) $$\alpha_i = (1 - \beta)\gamma_i \quad , \quad E\gamma_i = 0 \quad , \quad \mathrm{Var}\ (\gamma_i) = \sigma_\gamma^2 = \frac{\sigma^2}{(1 - \beta)^2} \ .$$

Then the assumption of fixed initial conditions implies that a cross-sectional unit may start from an arbitrary initial position, and gradually drift towards its respective level γ_i according to a probability law. The individual level γ_i is a random draw from a population with mean zero and variance σ_γ^2. This is a reasonable model, but there might be a question of treating y_{i0} as fixed if the decision of when to start sampling is arbitrary, in particular, independent of the value of y_{i0}.

Figure 1



Since α_i represents effects not taken into account explicitly, it may be unrealistic to assume y_{i0} as a fixed constant because α_i is then distributed independently of the starting value y_{i0}. The omitted effects are not brought into the model at time 0, but affect the process at time 1 and later and determine the eventual level.

We may re-write (2.6) as

$$(2.9) \qquad y_{it} = w_{it} + \gamma_i \quad , \quad t = 0,1,\ldots,T \quad ,$$

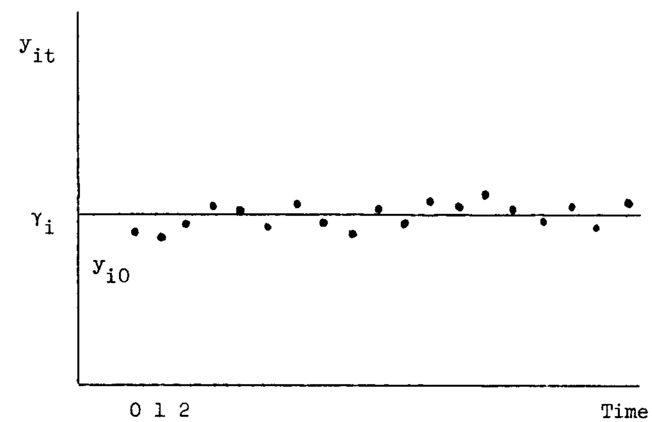
$$(2.10) \qquad w_{it} = \beta w_{i,t-1} + u_{it} \quad , \quad t = 1,\ldots,T \quad ,$$

where w_{it}, γ_i, and u_{it} are unobservable. It is natural in (2.6) to assume that the starting observable value and the level are independent. In the form (2.9) and (2.10) it is natural to assume that level γ_i and the unobservable process $\{w_{it}\}$ are independent; then the starting value y_{i0} is correlated with the level γ_i. If we allow correlation between y_{i0} and γ_i, the two models are equivalent via (2.7) and (2.8).

In the model (2.9) and (2.10), alternative standard assumptions about w_{i0} are (a) stationary with variance $\lambda\sigma^2/(1 - \beta^2)$, (b) random with arbitrary variance $(\lambda/\eta)\sigma^2$, and (c) w_{i0} fixed constants. We may express the initial conditions in three different ways. In case (a) y_{i0} is considered to have the marginal normal distribution

determined by the stationary process; that is, y_{i0} is viewed to form part of a stationary process as any other y_{it} (Figure 2). In case (b) the starting value y_{i0} is a random draw from a population which may

Figure 2



not have the same marginal distribution as at later periods. In case (c) it is similar to the first model where an individual may start at some value y_{i0} and move towards a level of γ_i, except that in this case the individual equilibrium level γ_i affects y_{i0}.

The third model we consider is that the initial observations are random with common mean but uncorrelated with the time disturbances We may assume that

$$(2.11) \qquad y_{i0} = c + \varepsilon_i \quad , \quad i = 1,\ldots,N \quad .$$

Then we may say that ε_i represents the effect of initial individual endowments (corrected for the mean). Depending on the assumption about ε_i, the impact of initial endowments will be different as time goes by. For instance, if we assume ε_i to be random with mean zero and variance σ_ε^2 and to be independent of α_i and u_{it}, its impact gradually diminishes and eventually vanishes. The model is somewhat like the first model in which the starting value and the level γ_i are independent, except that now the starting observable value is not a fixed constant but a random draw from a population with mean c and variance σ_ε^2

If we want to assume that the initial endowment affects the level, we may let

(2.12) $$y_{i0} = c + \alpha_i \quad , \quad i = 1,\ldots,N \quad .$$

Then, as time goes by, the effect of initial endowments cumulates and eventually reaches a level of $[\alpha_i/(1 - \beta)]$.

3. Fixed Initial Observations

In this section we assume that the initial observations y_{i0} are fixed constants and observable such that

(3.1) $$\lim_{N\to\infty} \frac{\sum_{i=1}^{N} y_{i0}^2}{N}$$

exists. Then

(3.2) $$\operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} [\sum_{t=1}^{T} y_{i,t-1}^2 - T\bar{y}_{i,-1}^2] \quad ,$$

exist and is nonzero. We define

(3.3) $$\bar{y}_i = \frac{\sum_{t=1}^{T} y_{it}}{T} \quad , \quad \bar{y}_{i,-1} = \frac{\sum_{t=1}^{T} y_{i,t-1}}{T} \quad .$$

We first consider the property of the CV of β. The CV for (2.6) is obtained by solving the following normal equation.

(3.4) $$\sum_{i=1}^{N} [\frac{1}{T} \sum_{t=1}^{T} y_{i,t-1}^2 - \bar{y}_{i,-1}^2]\hat{\beta}_{CV}$$

$$= \sum_{i=1}^{N} [\frac{1}{T} \sum_{t=1}^{T} y_{it}y_{i,t-1} - \bar{y}_i\bar{y}_{i,-1}] \quad .$$

We note that $Q\underset{\sim}{v}_i = Q\underset{\sim}{u}_i$. Equation (3.4) is equivalent to applying the least squares estimation of β to the model

(3.5) $$y_{i,t} = \beta y_{i,t-1} + a_i + u_{it} \; .$$

Thus, when T tends to infinity (regardless of whether N is fixed or tends to infinity) we can prove the consistency and asymptotic normality of the covariance estimator in exactly the same form as that of Anderson ([1971], ch. 5, Section 5.5, pp. 200-203). The variance of the limiting distribution of $\sqrt{NT}(\hat{\beta}_{CV} - \beta)$ is

$$\lambda\sigma^2\{\operatorname*{plim}_{T\to\infty} \frac{1}{NT}[\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}^2 - T\bar{y}_i^2]\}^{-1} = 1 - \beta^2 \quad . \tag{3.6}$$

On the other hand, when T is fixed but N tends to infinity the CV is inconsistent. This can be seen by noting that

$$\hat{\beta}_{CV} = \beta + \frac{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}u_{it} - \frac{1}{N}\sum_{i=1}^{N}\bar{y}_{i,-1}\bar{u}_i}{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}^2 - \frac{1}{N}\sum_{i=1}^{N}\bar{y}_{i,-1}^2} \tag{3.7}$$

where $\bar{u}_i = \sum_{t=1}^{T} u_{it}/T$. By a law of large numbers

$$\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}u_{it} = 0 \quad , \tag{3.8}$$

$$\begin{aligned} \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \bar{y}_{i,-1}\bar{u}_i &= \frac{\lambda\sigma^2}{T^2}[(T-1) + (T-2)\beta + (T-3)\beta^2 + \ldots + \beta^{T-2}] \\ &= \frac{\lambda\sigma^2}{T}\,\frac{T - 1 - T\beta + \beta^T}{(1-\beta)^2} \quad , \end{aligned} \tag{3.9}$$

which is not equal to zero.

The CV of β is also the MLE under the assumption that α_i are fixed constants and u_{it} are normally distributed. In this paper, however, we assume that α_i are random. When α_i and u_{it} are normally distributed, we can write down the logarithm of the likelihood function as

$$\begin{aligned} \log L = &- \frac{NT}{2}\log 2\pi - \frac{NT}{2}\log\sigma^2 - \frac{N}{2}\log|\underset{\sim}{\Lambda}| \\ &- \frac{1}{2\sigma^2}\sum_{i=1}^{N}(\underset{\sim}{y}_i - \underset{\sim}{y}_{i,-1}\beta)'\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{y}_i - \underset{\sim}{y}_{i,-1}\beta) \quad , \end{aligned} \tag{3.10}$$

where

$$\underset{T\times 1}{\underset{\sim}{y}_{i,-1}} = (y_{i0}, y_{i1}, \ldots, y_{i,T-1})' \quad ,$$

$$\underset{T\times T}{\underset{\sim}{\Lambda}} = \begin{pmatrix} 1+\lambda & 1 & 1 & \ldots & 1 \\ 1 & 1+\lambda & 1 & \ldots & 1 \\ & 1 & 1+\lambda & \ldots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 1 & \ldots & 1+\lambda \end{pmatrix} = \lambda \underset{\sim}{I}_T + \underset{\sim}{e}\underset{\sim}{e}' \quad , \tag{3.11}$$

$$|\Lambda| = \lambda^{T-1}(T + \lambda)$$

$$\Lambda^{-1} = \frac{1}{\lambda(T + \lambda)}\begin{pmatrix} \lambda + (T - 1) & -1 & \cdots & -1 \\ -1 & \lambda + (T - 1) & \cdots & -1 \\ -1 & -1 & \cdots & -1 \\ \vdots & \vdots & & \vdots \\ -1 & -1 & \cdots & \lambda + (T - 1) \end{pmatrix}$$

$$= \frac{1}{\lambda} I_T - \frac{1}{\lambda(T + \lambda)} ee' \ .$$

The derivative equations for the MLE are:

(3.12)
$$\frac{\partial \log L}{\partial \beta} = \frac{1}{\sigma^2\lambda(\lambda + T)}\Big[(\lambda + T) \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \beta y_{i,t-1})y_{i,t-1} - \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \beta y_{i,t-1})\sum_{t=1}^{T} y_{i,t-1}\Big] = 0 \ ,$$

(3.13)
$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{NT}{2\sigma^2} + \frac{1}{2\sigma^4\lambda(\lambda + T)}\Big\{(\lambda + T)\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \beta y_{i,t-1})^2 - \sum_{i=1}^{N}\Big(\sum_{t=1}^{T} y_{it} - \beta \sum_{t=1}^{T} y_{i,t-1}\Big)^2\Big\} = 0 \ ,$$

(3.14)
$$\frac{\partial \log L}{\partial \lambda} = -\frac{N(T - 1)}{2\lambda} - \frac{N}{2(\lambda + T)} + \frac{1}{2\sigma^2\lambda^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \beta y_{i,t-1})^2 - \frac{2\lambda + T}{2\sigma^2\lambda^2(\lambda + T)^2}\sum_{i=1}^{N}\Big(\sum_{t=1}^{T} y_{it} - \beta\cdot\sum_{t=1}^{T} y_{i,t-1}\Big)^2 = 0 \ .$$

If there is a unique solution to these three equations with that $\sigma^2 > 0$, $0 < \lambda < \infty$, the solution is the MLE. Maddala [1971] has shown that the boundary solution of $\lambda = 0$ (that is, $\sigma_u^2 = 0$) cannot occur, but the boundary solution of $\sigma^2 = 0$ (that is, $\lambda = \infty$) will occur when $\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \beta y_{i,t-1})^2 \geq T^2\sum_{i=1}^{N}(\bar{y}_i - \beta\bar{y}_{i,-1})^2$. The solution, then, does not satisfy all of the derivative equation. Nerlove [1971] obtained such solutions in a simulation study. However, the probability of a boundary solution tends to 0 as $T\to\infty$ or as $N\to\infty$.

Amemiya [1967] and Balestra and Nerlove [1966] have shown that the the MLE is consistent and asymptotically normally distributed when T tends to infinity (and N is fixed or tends to infinity). When T is fixed and N tends to infinity, the MLE is consistent and asymptotically normally distributed under the assumption that y_{i0} are fixed constants.

To show that the MLE is consistent when N tends to infinity and T is fixed, we let $\theta = (\beta,\sigma^2,\lambda)'$. Noting that $E(u_{it}y_{i,t+j}) = \beta^j\lambda\sigma^2$ for $j \geq 0$, and $E(\alpha_i y_{it}) = (1 + \beta + \ldots + \beta^{t-1})\sigma^2$ for $t \geq 2$ and $E\alpha_i y_{i1} = \sigma^2$, we can show that $[1/(NT)](\partial \log L/\partial\theta)$ converges in probability to 0 at the true θ when T is fixed and N tends to infinity.

Furthermore, $[1/(NT)](\partial^2 \log L/\partial\underset{\sim}{\theta}\partial\underset{\sim}{\theta}')$ around its true value converges in probability to

$$(3.15)\quad -\begin{pmatrix} \frac{1}{\lambda\sigma^2}\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}[\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1}^2-\frac{T}{\lambda+T}\bar{y}_{i,-1}^2] & 0 & 0 \\ 0 & \frac{1}{2\sigma^4} & \frac{\lambda+(T-1)}{2\sigma^2\lambda(\lambda+T)} \\ 0 & \frac{\lambda+(T-1)}{2\sigma^2\lambda(\lambda+T)} & \frac{[\lambda^2+2\lambda(T-1)+T(T-1)]}{2\lambda^2(\lambda+T)^2} \end{pmatrix}$$

which is negative definite as long as $T \geq 2$. By Amemiya's [1973] Lemmas 3 and 4, we know that there is a consistent root for the MLE. We can also show the asymptotic normality of the MLE by an argument similar to that of Anderson [1978].

The solution of the MLE is complicated. An iterative procedure such as Newton-Raphson type will have to be used. However, if T tends to infinity, the CV is not only consistent, but is also asymptotically equivalent to the MLE. [See (3.6) and (3.15)].

It is interesting to note that when N is fixed and T tends to infinity, it is not possible to obtain separate consistent estimates of σ^2 and λ (Amemiya [1969]). Yet if T is fixed, as long as it is greater than one, we can get separate consistent estimates of σ^2 and λ when N tends to infinity. When T is one, we have N independent random variables $v_{i1} = \alpha_i + u_{i1}$ with variance $\sigma^2(1+\lambda)$. It is not possible to distinguish λ and σ^2. However, the MLE of β

in this case is consistent and it is the same as the least squares estimate.

The incidental parameters problem does not arise under the assumption that y_{i0}'s are fixed because they are observed. The individual component α_i only gives rise to a special form of the covariance matrix of $\underset{\sim}{v}_i$. The consistency of the MLE is a consequence of the fact that we are maximizing the likelihood function of N random vectors $\underset{\sim}{y}_i$ which are independently normally distributed.

4. Random Initial Observations with a Stationary Distribution

In this and the next section we shall consider the second model [(2.9) and (2.10)] in which the initial observations y_{i0} are treated as random and correlated with α_i. We first consider case (a) where w_{i0} is assumed to be normally distributed with mean zero and variance $[\lambda\sigma^2/(1-\beta^2)]$. Then y_{i0} will be normally distributed with mean zero and variance $[\lambda\sigma^2/(1-\beta^2) + \sigma^2/(1-\beta)^2]$, and $E(\alpha_i y_{i0}) = \sigma^2/(1-\beta)$. The joint density of $(y_{i0}, y_{i1}, \ldots, y_{iT})$ is

$$(4.1)\quad f_i(y_{i0},\ldots,y_{iT}) = (2\pi)^{-\frac{(T+1)}{2}}(\sigma^2)^{-\frac{(T+1)}{2}}|\underset{\sim}{\Omega}|^{-\frac{1}{2}}$$

$$\cdot \exp\{-\frac{1}{2\sigma^2}[y_{i0}, y_{i1}-\beta y_{i0},\ldots,y_{iT}-\beta y_{i,T-1}]\underset{\sim}{\Omega}^{-1}$$

$$\cdot [y_{i0}, y_{i1}-\beta y_{i0},\ldots,y_{iT}-\beta y_{i,T-1}]'\} \quad ,$$

where

$$\underset{(T+1)\times(T+1)}{\Omega} = \lambda \begin{pmatrix} \frac{1}{1-\beta^2} & & & \\ & 1 & & 0 \\ & & \ddots & \\ & 0 & & 1 \end{pmatrix} + \begin{pmatrix} \frac{1}{1-\beta} \\ 1 \\ \vdots \\ 1 \end{pmatrix} \left(\frac{1}{1-\beta}, 1, \ldots, 1\right) ,$$

(4.2) $$|\Omega| = \frac{\lambda^T}{1-\beta^2}\left[\lambda + T + \frac{1+\beta}{1-\beta}\right] ,$$

$$\Omega^{-1} = \frac{1}{\lambda}\left\{ \begin{pmatrix} 1-\beta^2 & & & \\ & 1 & & 0 \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix} - \left(\lambda + T + \frac{1+\beta}{1-\beta}\right)^{-1} \begin{pmatrix} 1+\beta \\ 1 \\ \vdots \\ 1 \end{pmatrix} (1+\beta, 1, \ldots, 1) \right\} .$$

It might be noticed that the density of $y_{i1},\ldots,y_{iT}$ in Section 3 is <u>not</u> obtained from (4.1) as the conditional density given y_{i0}.

The derivatives of the logarithm of $\prod_{i=1}^{N} f_i(y_{i0},\ldots,y_{iT})$ with respect to β, σ^2 and λ are:

(4.3) $$\frac{\partial \log L}{\partial \beta} = -\frac{N\beta}{1-\beta^2} - \frac{N}{2(1-\beta)} + \frac{N(\lambda+T-1)}{2[2+(\lambda+T-1)(1-\beta)]}$$

$$+ \frac{\beta}{\lambda\sigma^2}\sum_{i=1}^{N} y_{i0}^2 + \frac{1}{\lambda\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \beta y_{i,t-1})y_{i,t-1}$$

$$- \frac{1}{\lambda\sigma^2}\cdot\frac{1}{[2+(\lambda+T-1)(1-\beta)]^2}\sum_{i=1}^{N}[(1+\beta)y_{i0}$$

$$+ \sum_{t=1}^{T}(y_{it} - \beta y_{i,t-1})]^2 - \frac{1}{\lambda\sigma^2}\cdot\frac{(1-\beta)}{[2+(\lambda+T-1)(1-\beta)]}$$

$$\sum_{i=1}^{N}[(1+\beta)y_{i0} + \sum_{t=1}^{T}(y_{it} - \beta y_{i,t-1})][\sum_{t=1}^{T-1} y_{it}] ,$$

(4.4) $$\frac{\partial \log L}{\partial \sigma^2} = -\frac{N(T+1)}{2\sigma^2} + \frac{1}{2\lambda\sigma^4}\sum_{i=1}^{N}\{(1-\beta^2)y_{i0}^2 + \sum_{t=1}^{T}(y_{it} - \beta y_{i,t-1})^2$$

$$- \frac{(1-\beta)}{2+(\lambda+T-1)(1-\beta)}[(1+\beta)y_{i0}$$

$$+ \sum_{t=1}^{T}(y_{it} - \beta y_{i,t-1})]^2\} ,$$

$$(4.5)\quad \frac{\partial \log L}{\partial \lambda} = -\frac{NT}{2\lambda} - \frac{N(1-\beta)}{2[2+(\lambda+T-1)(1-\beta)]} + \frac{1}{2\lambda^2\sigma^2}\sum_{i=1}^{N}\{(1-\beta^2)y_{i0}^2 + \sum_{t=1}^{T}(y_{it}-\beta y_{i,t-1})^2\} - \frac{[2+(2\lambda+T-1)(1-\beta)](1-\beta)}{2\sigma^2\{\lambda[2+(\lambda+T-1)(1-\beta)]\}^2} \cdot \sum_{i=1}^{N}[(1+\beta)y_{i0} + \sum_{t=1}^{T}(y_{it}-\beta y_{i,t-1})]^2 ,$$

When N is fixed and T tends to infinity, the consistency of the MLE follows from the usual argument that the asymptotic theory is the same for both y_{i0} fixed and y_{i0} random with a stationary distribution (e.g., see Anderson [1971], ch. 5). The consistency of the MLE when N tends to infinity and T is fixed follows from the fact that the MLE is obtained by maximizing the joint density of N independently normally distributed random vectors $(y_{i0},y_{i1},\ldots,y_{iT})$. Again, there is no incidental parameters problem in this case.

There are similar conditions for the occurrence of boundary solutions. The boundary solution of $\lambda = 0$ (or $\sigma_u^2 = 0$) cannot occur, but the boundary solution that $\sigma^2 = 0$ will occur if $(\underset{\sim}{e}'\underset{\sim}{\Phi}^{-1}\underset{\sim}{e})\sum_{i=1}^{N}\underset{\sim}{y}_i'\underset{\sim}{\Phi}^{-1}\underset{\sim}{y}_i \geq T \cdot \sum_{i=1}^{N}(\underset{\sim}{y}_i'\underset{\sim}{\Phi}^{-1}\underset{\sim}{e})^2$, where

$$\underset{\sim}{\Phi} = \begin{pmatrix} 1 & \beta & \cdots & \beta^{t-1} \\ \beta & 1 & \cdots & \beta^{t-2} \\ \vdots & \vdots & & \\ \beta^{t-1} & \beta^{t-2} & \cdots & 1 \end{pmatrix}$$

(Lee [1979]).

In this model the CV will have the same consistency properties as in the first model. As $T \to \infty$, the last term in (3.7) converges in probability to 0. As $N \to \infty$, the last term in (3.7) converges in probability to the negative of (3.9), which is not 0. Thus the CV is consistent as $T \to \infty$ and is inconsistent as $N \to \infty$.

Although the consistency of the CV is not independent of the way T or N goes to infinity, the MLE in this case is consistent and independent of the way T or N goes to infinity. However (4.3)-(4.5) do not have a simple solution and a complicated iterative scheme will have to be used. Unfortunately, the assumption that w_{i0} are random with arbitrary variance $(\lambda/\eta)\sigma^2$ (case (b)) does not simplify the computation either. One can see this by noting that the joint density $f_i(y_{i0},\ldots,y_{iT})$ is still of the form (4.1) with the definition of $\underset{\sim}{\Omega}$ (4.2) replaced by

$$(4.6)\quad \underset{\sim}{\Omega} = \lambda\begin{pmatrix} \frac{1}{\eta} & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix} + \begin{pmatrix} \frac{1}{1-\beta} \\ 1 \\ \vdots \\ 1 \end{pmatrix}\left(\frac{1}{1-\beta},1,\ldots,1\right) ,$$

$$|\Omega| = \eta^{-1}\lambda^{T}[\lambda + T + \frac{\eta}{(1-\beta)^2}] \quad ,$$

$$\Omega^{-1} = \frac{1}{\lambda}\left\{\begin{pmatrix} \eta & & & \\ & 1 & & 0 \\ & & \cdot & \\ 0 & & & \cdot \\ & & & 1 \end{pmatrix} - [\lambda + T + \frac{\eta}{(1-\beta)^2}]^{-1}\begin{pmatrix} \frac{\eta}{1-\beta} \\ 1 \\ \cdot \\ \cdot \\ 1 \end{pmatrix}(\frac{\eta}{1-\beta}, 1, \ldots, 1)\right\}.$$

The MLE is still consistent if either T tends to infinity and N is fixed or N tends to infinity and T is fixed, but the computation is again complicated.

5. Random Initial Observations with Different Means

Here we assume in model (2.9) and (2.10) (case (c)) that w_{i0} are fixed constants. Then at $t = 0$ $\gamma_i = y_{i0} - w_{i0}$. The density of $(y_{i0}, y_{i1}, \ldots, y_{it})$ is 3/

(5.1) $$f_i(y_{i0}, y_{i1}, \ldots, y_{iT}) = f_i(y_{i1}, \ldots, y_{iT} | y_{i0}) f_i(y_{i0})$$

$$= (\frac{1}{\sqrt{2\pi}\,\sigma_u})^T \exp\left\{-\frac{1}{2\sigma_u^2}\sum_{t=1}^{T}[(y_{it} - y_{i0} + w_{i0}) - \beta(y_{i,t-1} - y_{i0} + w_{i0})]^2\right\}$$

$$\cdot \frac{1}{\sqrt{2\pi}\,\sigma_\gamma}\exp\left\{-\frac{1}{2\sigma_\gamma^2}(y_{i0} - w_{i0})^2\right\} .$$

For given y_{it}'s it is a function of w_{i0}'s, β, σ_u^2 and σ_γ^2. If we let $w_{i0} = y_{i0}$, then the second factor is $1/(\sqrt{2\pi}\,\sigma_\gamma)$. As $\sigma_\gamma \to 0$, this approaches ∞ and the likelihood function approaches ∞. The likelihood function does not have a maximum and, hence, the MLE does not exist.

However, if we focus our attention on the interior solutions we may take the partial deviation of the logarithm of the joint likelihood function $\prod_{i=1}^{N} f_i(y_{i0}, y_{i1}, \ldots, y_{iT})$,

(5.2) $$\log L = -\frac{NT}{2}\log 2\pi - \frac{NT}{2}\log\sigma_u^2 - \frac{1}{2\sigma_u^2}\sum_{i=1}^{N}\sum_{t=1}^{T}[(y_{it} - y_{i0} + w_{i0}) - \beta(y_{i,t-1} - y_{i0} + w_{i0})]^2 - \frac{N}{2}\log 2\pi - \frac{N}{2}\log\sigma_\gamma^2 - \frac{1}{2\sigma_\gamma^2}\sum_{i=1}^{N}(y_{i0} - w_{i0})^2 ,$$

with respect to w_{i0}, β, σ_u^2 and σ_γ. These are

(5.3) $$\frac{\partial \log L}{\partial w_{i0}} = -\frac{1}{\sigma_u^2}\sum_{t=1}^{T}[(y_{it} - y_{i0} + w_{i0}) - \beta(y_{i,t-1} - y_{i0} + w_{i0})] \cdot (1 - \beta) + \frac{1}{\sigma_\gamma^2}(y_{i0} - w_{i0}) \quad , \quad i = 1, \ldots, N ,$$

(5.4) $$\frac{\partial \log L}{\partial \beta} = \frac{1}{\sigma_u^2}\sum_{i=1}^{N}\sum_{t=1}^{T}[(y_{it} - y_{i0} + w_{i0}) - \beta(y_{i,t-1} - y_{i0} + w_{i0})] \cdot (y_{i,t-1} - y_{i0} + w_{i0}) ,$$

$$\frac{\partial \log L}{\partial \sigma_u^2} = -\frac{NT}{2\sigma_u^2} + \frac{1}{2\sigma_u^4} \sum_{i=1}^{N} \sum_{t=1}^{T} [(y_{it} - y_{i0} + w_{i0}) - \beta(y_{i,t-1} - y_{i0} + w_{i0})]^2 \tag{5.5}$$

$$\frac{\partial \log L}{\partial \sigma_\gamma^2} = -\frac{N}{2\sigma_\gamma^2} + \frac{1}{2\sigma_\gamma^4} \sum_{i=1}^{N} (y_{i0} - w_{i0})^2 \ . \tag{5.6}$$

Setting these equal to zero, we obtain

$$[T(1-\hat\beta)^2\hat\sigma_\gamma^2 + \hat\sigma_u^2]\hat w_{i0} = \hat\sigma_u^2 y_{i0} - \hat\sigma_\gamma^2(1-\hat\beta)\sum_{t=1}^{T}[(y_{it} - y_{i0}) - \hat\beta(y_{i,t-1} - y_{i0})] \quad , \ i = 1,\ldots,N \ , \tag{5.7}$$

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - y_{i0} + \hat w_{i0})(y_{i,t-1} - y_{i0} + \hat w_{i0}) = \hat\beta \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - y_{i0} + \hat w_{i0})^2 \ , \tag{5.8}$$

$$\hat\sigma_u^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}[(y_{it} - y_{i0} + \hat w_{i0}) - \hat\beta(y_{i,t-1} - y_{i0} + \hat w_{i0})]^2}{NT} \ , \tag{5.9}$$

$$\hat\sigma_\gamma^2 = \frac{\sum_{i=1}^{N}(y_{i0} - \hat w_{i0})^2}{N} \ . \tag{5.10}$$

We want to show first that the interior solution is asymptotically equivalent to the CV as $T \to \infty$. The coefficient of $\hat\beta$ on the right-hand side of (5.8) divided by NT can be written

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[(y_{i,t-1} - \bar y_{i,-1}) + (\bar y_{i,-1} - y_{i0} + \hat w_{i0})]^2 = \frac{1}{N}\sum_{i=1}^{N}[\frac{1}{T}\sum_{t=1}^{T}(y_{i,t-1} - \bar y_{i,-1})^2 + (\bar y_{i,-1} - y_{i0} + \hat w_{i0})^2] \ . \tag{5.11}$$

From (5.7) we find

$$\hat w_{i0} - y_{i0} = -\frac{\hat\sigma_\gamma^2(1-\hat\beta)T}{T(1-\hat\beta)^2\hat\sigma_\gamma^2 + \hat\sigma_u^2}(\bar y_i - \hat\beta \bar y_{i,-1}) \ . \tag{5.12}$$

Then we see that the second term on the right-hand side of (5.11) converges in probability to 0. Similarly the left-hand side of (5.8) divided by NT can be written

$$\frac{1}{N}\sum_{i=1}^{N}[\frac{1}{T}\sum_{t=1}^{T}(y_{i,t} - \bar y_i)(y_{i,t-1} - \bar y_{i,-1}) + (\bar y_i - y_{i0} + \hat w_{i0})(\bar y_{i,-1} - y_{i0} + \hat w_{i0})] \ , \tag{5.13}$$

and the second term converges in probability to 0. Then as $T \to \infty$, (5.8) is equivalent to (3.4). The solution is consistent as $T \to \infty$.

Now let us consider the case of $N \to \infty$. If we substitute (5.12) into (5.8), (5.9) and (5.10), we obtain three polynomial equations in $\hat{\beta}$, $\hat{\sigma}_u^2$, and $\hat{\sigma}_\gamma^2$. There may be several different solutions. These give stationary points of the likelihood function; perhaps some are relative maxima. But none can give an absolute maximum. Neither does any of these relative maxima yield a consistent root. To show this, we assume that $(1/N) \sum_{i=1}^{N} w_{i0}^2$ converges to a finite constant. We solve (5.7) and put it into (5.8). Then we subtract the right-hand side from the left-hand side. If $\hat{\beta}$ is consistent, we can replace it by β and find the probability limit of this difference divided by NT. Its probability limit is not equal to zero; it is equal to

$$(5.14) \qquad \frac{1}{T(1-\beta)^2\hat{\sigma}_\gamma^2 + \hat{\sigma}_u^2}[(1-\beta)^2\hat{\sigma}_\gamma^2]\{-\frac{\sigma_u^2}{T}[(T-1) + (T-2)\beta + \ldots + \beta^{T-2}]\}$$

$$+ \frac{1}{[T(1-\beta)^2\hat{\sigma}_\gamma^2 + \hat{\sigma}_u^2]^2}(1-\beta)\hat{\sigma}_u^2[\hat{\sigma}_u^2\sigma_\gamma^2 - \sigma_u^2\hat{\sigma}_\gamma^2] \; .$$

This contradiction shows that $\hat{\beta}$ is not consistent.

The analysis of the behavior of the CV proceeds as in the two previous sections. The CV is consistent as $T \to \infty$ and is inconsistent as $N \to \infty$.

It may be of interest to note that when $T = 1$ this is similar to the classical problem of incidental parameters.

6. Random Initial Observations with a Common Mean

In this section we first consider the model (2.11) where the initial endowment ε_i does not affect the level γ_i and the disturbances

u_{it}. Then the joint likelihood function of $(y_{i0},\ldots,y_{iT})$ is

$$(6.1) \qquad f_i(y_{i0},\ldots,y_{iT}) = f_i(y_{i1},\ldots,y_{iT}|y_{i0})f_i(y_{i0})$$

$$= (2\pi\sigma^2)^{-\frac{T}{2}} |\underset{\sim}{\Lambda}|^{-\frac{1}{2}} \exp\{-\frac{1}{2\sigma^2}(\underset{\sim}{y}_i - \underset{\sim}{y}_{i,-1}\beta)'\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{y}_i - \underset{\sim}{y}_{i,-1}\beta)\}$$

$$\cdot (2\pi\sigma_\varepsilon^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}(y_{i0} - c)^2\right\} .$$

Therefore, the MLE of β, λ, and σ^2 is identical to the MLE under the assumption that y_{i0} are fixed constants, except now that in addition to estimating these parameters, we also estimate c and σ_ε^2 by

$$(6.2) \qquad \hat{c} = \frac{\sum_{i=1}^{N} y_{i0}}{N} \; ,$$

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{N} (y_{i0} - \hat{c})^2}{N} \; .$$

On the other hand, if we assume (2.12), the joint likelihood function will be

$$(6.3) \qquad f_i(y_{i0},\ldots,y_{iT}) = (2\pi)^{-\frac{T+1}{2}} (\sigma^2)^{-\frac{T+1}{2}} |\tilde{\underset{\sim}{\Omega}}|^{-\frac{1}{2}} \exp\{-\frac{1}{2\sigma^2}$$

$$\cdot (y_{i0} - c,\; y_{i1} - \beta y_{i0},\ldots,y_{iT} - \beta y_{i,T-1})\tilde{\underset{\sim}{\Omega}}^{-1}$$

$$\cdot\,(y_{i0} - c, \ldots, y_{iT} - \beta y_{i,T-1})'$$

$$= f_i(y_{i1}, \ldots, y_{iT} | y_{i0}) f_i(y_{i0})$$

$$= (2\pi)^{-\frac{T}{2}} (\sigma^2\lambda)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2\lambda\sigma^2} \sum_{t=1}^{T} [(y_{it} - y_{i0}) - \beta y_{i,t-1} + c]^2\right\} \cdot (2\pi)^{-\frac{1}{2}} (\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{(y_{i0} - c)^2}{2\sigma^2}\right\} ,$$

where

$$\underset{(T+1)\times(T+1)}{\tilde{\Omega}} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1+\lambda & 1 & \cdots & 1 \\ 1 & 1 & 1+\lambda & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \\ 1 & 1 & 1 & & 1+\lambda \end{pmatrix} .$$

The MLE of $\underset{\sim}{\delta} = (\beta, c, \sigma^2, \lambda)'$ is obtained by taking the partial derivatives of the logarithm of $\prod_{i=1}^{N} f_i(y_{i0}, \ldots, y_{iT})$ and setting them equal to zero:

$$(6.5) \qquad \frac{\partial \log L}{\partial \beta} = \frac{1}{\lambda\sigma^2} \sum_{i=1}^{N} \sum_{t=1}^{T} [(y_{it} - y_{i0}) - \beta y_{i,t-1} + c] y_{i,t-1} = 0 ,$$

$$(6.6) \qquad \frac{\partial \log L}{\partial c} = -\frac{1}{\lambda\sigma^2} \sum_{i=1}^{N} \sum_{t=1}^{T} [(y_{it} - y_{i0}) - \beta y_{i,t-1} + c] + \frac{1}{\sigma^2} \sum_{i=1}^{N} (y_{i0} - c) = 0 ,$$

$$(6.7) \qquad \frac{\partial \log L}{\partial \sigma^2} = -\frac{N(T+1)}{2\sigma^2} + \frac{1}{2\lambda\sigma^2} \sum_{i=1}^{N} \sum_{t=1}^{T} [(y_{it} - y_{i0}) - \beta y_{i,t-1} + c]^2 + \frac{1}{2\sigma^4} \sum_{i=1}^{N} (y_{i0} - c)^2 = 0 ,$$

$$(6.8) \qquad \frac{\partial \log L}{\partial \lambda} = -\frac{NT}{2\lambda} + \frac{1}{2\lambda^2\sigma^2} \sum_{i=1}^{N} \sum_{t=1}^{T} [(y_{it} - y_{i0}) - \beta y_{i,t-1} + c]^2 = 0 .$$

Contrary to previous cases, the solution to the derivative equations (6.5) - (6.8) is always an interior one and there is no boundary value problem.

It is easy to show that $[1/(N(T+1))] \cdot [(\partial \log L)/\partial \underset{\sim}{\delta}]$ converges in probability to $\underset{\sim}{0}$ at the true value and $[1/(N(T+1))] \cdot [(\partial^2 \log L)/\partial \underset{\sim}{\delta} \partial \underset{\sim}{\delta}']$ converges in probability to a negative definite matrix when either T tends to infinity and N is fixed or N tends to infinity and T is fixed or both. Therefore, the MLE is consistent in either case.

We note that conditional on y_{i0} in (6.3) we can maximize $L^* = \prod_{i=1}^{N} f_i(y_{i1}, \ldots, y_{iT} | y_{i0})$ with respect to β, c, and $\lambda\sigma^2$. This conditional MLE is consistent when either T or N or both tend to infinity, and asymptotically normally distributed. Of course, when T is fixed and N tends to infinity the unconditional MLE of β and c is more efficient than the conditional MLE (in the sense of having smaller variance-covariance matrix). But it is computationally more tedious than the conditional MLE. Taking partial derivatives of $\log L^*$, we have

$$(6.9)\qquad \frac{\partial \log L^*}{\partial \beta} = \frac{1}{\lambda\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}[(y_{it} - y_{i0}) - \beta y_{i,t-1} + c]y_{i,t-1} = 0 \quad ,$$

$$(6.10)\qquad \frac{\partial \log L^*}{\partial c} = -\frac{1}{\lambda\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}[(y_{it} - y_{i0}) - \beta y_{i,t-1} + c] = 0 \quad ,$$

$$(6.11)\qquad \frac{\partial \log L^*}{\partial(\lambda\sigma^2)} = -\frac{NT}{2(\lambda\sigma^2)} + \frac{1}{2(\lambda\sigma^2)^2}\sum_{i=1}^{N}\sum_{t=1}^{T}[(y_{it} - y_{i0}) - \beta y_{i,t-1} + c]^2 = 0 \quad .$$

Equation (6.9)-(6.11) are nothing but the least squares regression of $(y_{it} - y_{i0})$ on $y_{i,t-1}$ and a constant term.4/ This solution can either be used as a consistent estimate or be used to start the iterative procedure to obtain unconditional MLE.

The property of the CV is the same as in other cases. It is consistent when T tends to infinity and inconsistent when T is fixed and N tends to infinity.

7. Pseudo and Conditional Maximum Likelihood Estimators

The purpose of using a covariance estimator is to eliminate the individual effect α_i. This can be done by premultiplying $\underset{\sim}{v}_i$ by the $(T - 1) \times T$ transformation matrix

$$(7.1)\qquad \underset{(T-1)\times T}{\underset{\sim}{D}} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}$$

Then

$$7.2)\qquad \underset{\sim}{D}\underset{\sim}{v}_i = \underset{\sim}{D}\underset{\sim}{u}_i = \underset{\sim}{D}\begin{pmatrix} u_{i1} \\ \cdot \\ \cdot \\ u_{iT} \end{pmatrix} \quad ,$$

is normally distributed with mean $\underset{\sim}{0}$ and covariance matrix

$$7.3)\qquad \sigma_u^2 \underset{\sim}{D}\underset{\sim}{\Lambda}\underset{\sim}{D}' = \sigma_u^2\begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix} \quad .$$

From (2.6) we have

$$(7.4)\qquad y_{it} - y_{i,t-1} = \beta(y_{i,t-1} - y_{i,t-2}) + u_{it} - u_{i,t-1}$$

However, from this we cannot obtain MLE's as claimed by some people. This can be seen by noting that although $\underset{\sim}{D}\underset{\sim}{u}_i$ has a properly defined density function, $\underset{\sim}{D}\underset{\sim}{y}_i$ does not. Even under the assumption that y_{i0} are fixed, y_{i1} are still random. Thus, $\underset{\sim}{D}\underset{\sim}{y}_i$ leaves the density of $(y_{i1} - y_{i0})$ undefined.

Substituting (7.4) into the density of $\underset{\sim}{D}\underset{\sim}{u}_1,\ldots,\underset{\sim}{D}\underset{\sim}{u}_N$ and deriving the estimators by maximizing this quantity with respect to β, λ and σ^2 yields estimators that are not consistent when N tends to infinity and T is fixed. We show the inconsistency of these pseudo maximum likelihood estimators by considering the case where y_{i0} are fixed and $T = 3$. Then

$$(7.5) \qquad \underset{2\times 3}{\underset{\sim}{D}\underset{\sim}{u}_i} = \begin{pmatrix} u_{i2} - u_{i1} \\ u_{i3} - u_{i2} \end{pmatrix}$$

is normally distributed with mean zero and variance-covariance matrix

$$(7.6) \qquad \sigma_u^2 \underset{\sim}{D}\underset{\sim}{A}\underset{\sim}{D}' = \sigma_u^2 \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

The determinant of $\underset{\sim}{D}\underset{\sim}{A}\underset{\sim}{D}'$ is now equal to 3 and the inverse of it is

$$\frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

If we consider (7.4), $t = 2,3$ as a transformation from $(u_{i2} - u_{i1})$ and $(u_{i3} - u_{i2})$ to y_{i2}, and y_{i3}, $i=1,\ldots N$, the Jacobian of the transformation is equal to one. Thus, the logarithm of the pseudo likelihood function is equal to

$$(7.7) \qquad \text{Log } \tilde{L} = \text{constant} - N \log \sigma_u^2 - \frac{1}{3\sigma_u^2}\sum_{i=1}^{N}\{[(y_{i2} - y_{i1}) - \beta(y_{i1} - y_{i0})]^2$$
$$+ [(y_{i3} - y_{i2}) - \beta(y_{i2} - y_{i1})]^2$$
$$+ [(y_{i2} - y_{i1}) - \beta(y_{i1} - y_{i0})][(y_{i3} - y_{i2}) - \beta(y_{i2} - y_{i1})]\}$$

Taking partial derivatives of (7.7) with respect to β and solving for β, we have

$$(7.8) \qquad \hat{\beta}_{PML} = \frac{\sum_{i=1}^{N}[2(y_{i2} - y_{i1})(y_{i1} - y_{i0}) + 2(y_{i3} - y_{i2})(t_{i2} - y_{i1}) + (y_{i2} - y_{i1})^2 + (y_{i3} - y_{i2})(y_{i1} - y_{i0})]}{\sum_{i=1}^{N}[2(y_{i1} - y_{i0})^2 + 2(y_{i2} - y_{i1})^2 + (y_{i2} - y_{i1})(y_{i1} - y_{i0}) + (y_{i2} - y_{i1})(y_{i1} - y_{i0})]} .$$

The probability limit of this pseydo MLE $\hat{\beta}_{PML}$ is equal to

$$(7.9) \qquad \underset{N\to\infty}{\text{plim}}\ \hat{\beta}_{PML} = \beta - \frac{2 - \beta}{\text{constant}} ,$$

Hence, it is inconsistent.

On the other hand, the maximization of the joint density of $(y_{i2},\ldots,y_{iT})$ conditional on y_{i1} over i does yield a consistent estimator. This follows from the fact that conditional on y_{i1} (and y_{i0} fixed) we are maximizing $(T - 1)$ - component independently distributed random vectors, the i-th having density

$$f_i(y_{i2},\ldots,y_{iT}|y_{i1}) = (2\pi)^{-\frac{T-1}{2}}(\sigma^2)^{-\frac{T-1}{2}}|\underset{\sim}{\Lambda}^*|^{-\frac{1}{2}} \tag{7.10}$$

$$\cdot \exp\{-\frac{1}{2\sigma^2}[y_{i2} - \beta y_{i1} - \frac{1}{1+\lambda}(y_{i1} - \beta y_{i0}),$$

$$y_{i3} - \beta y_{i2} - \frac{1}{1+\lambda}(y_{i1} - \beta y_{i0}),$$

$$\ldots, y_{iT} - \beta y_{i,T-1} - \frac{1}{1+\lambda}(y_{i1} - \beta y_{i0})]\underset{\sim}{\Lambda}^{*-1}$$

$$[y_{i2} - \beta y_{i1} - \frac{1}{1+\lambda}(y_{i1} - y_{i0}) \quad ,$$

$$\ldots, y_{iT} - \beta y_{i,T-1} - \frac{1}{1+\lambda}(y_{i1} - \beta y_{i0})]'\} \quad , \quad i = 1,\ldots,N,$$

where

$$\underset{(T-1)\times(T-1)}{\underset{\sim}{\Lambda}^*} = (\lambda I_{T-1} + \underset{\sim}{e}_{T-1}\underset{\sim}{e}'_{T-1}) - \frac{1}{1+\lambda}\underset{\sim}{e}_{T-1}\underset{\sim}{e}_{T-1} \ . \tag{7.11}$$

We illustrate the argument of consistency by considering the case that $T = 2$. Then the conditional density of y_{i2} given y_{i0} and y_{i1} is normal with mean $\delta_0 y_{i0} + \delta_1 y_{i1}$ and variance τ^2, where

$$\delta_0 = -\frac{\beta}{1+\lambda} \tag{7.12}$$

$$\delta_0 = \beta + \frac{1}{1+\lambda}$$

$$\tau^2 = \frac{\lambda(2+\lambda)\sigma^2}{1+\lambda} \ .$$

The MLE's of δ_o, δ_1, τ^2 are (strongly) consistent if

$$\left[\sum_{i=1}^{N}\begin{pmatrix} y_{i0} \\ y_{i1}\end{pmatrix}(y_{i0} \quad y_{i1})\right]^{-1} \to \underset{\sim}{0} \ . \tag{7.13}$$

(Anderson and Taylor [1979]). Since the transformation $(\delta_0,\delta_1,\tau^2)$ and (β,λ,σ^2) is one-to-one (in the proper region $(\lambda > 0, \sigma^2 > 0)$), the MLE's of β, λ, σ^2 are (strongly) consistent if (7.13) is fulfilled.

If y_{i1} is a random draw from $y_{i1} = \beta y_{i0} + \alpha_i + u_{it}$, then

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix} y_{i0} \\ y_{i1}\end{bmatrix}(y_{i0} \quad y_{i1}) \tag{7.14}$$

$$= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} y_{i0}^2 \begin{pmatrix} 1 \\ \beta \end{pmatrix}(1 \quad \beta) + \begin{pmatrix} 0 & 0 \\ 0 & (1+\lambda)\sigma^2 \end{pmatrix} \quad ,$$

which is positive definite. Therefore, the conditional MLE is consistent when T is fixed and N tends to infinity. It is the pseudo MLE which is inconsistent. We suspect it is this confusion about the proper form of the conditional likelihood function which caused the confusion about the consistency of the conditional MLE (e.g., Chamberlein [1979], Lee [1979]).

8. Simple Consistent Estimates

Although the maximization of (7.4) does not yield consistent estimates, it does suggest some simple consistent estimators. From (7.4) we know that we may use either $y_{i,t-2}$ or $(y_{i,t-2} - y_{i,t-3})$ as instruments and estimate β by

$$(8.1)\quad \hat{\beta}_{IV} = \frac{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{it} - y_{i,t-1})(y_{i,t-2} - y_{i,t-3})}{\sum_{i=1}^{N}\sum_{t=3}^{N}(y_{i,t-1} - y_{i,t-2})(y_{i,t-2} - y_{i,t-3})}$$

or

$$(8.2)\quad \tilde{\beta}_{IV} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}(y_{it} - y_{i,t-1})y_{i,t-2}}{\sum_{i=1}^{N}\sum_{t=2}^{T}(y_{i,t-1} - y_{i,t-2})y_{i,t-2}}$$

Both (8.1) and (8.2) are consistent when N tends to infinity or T tends to infinity or both.

Estimator (8.2) has the advantage over (8.1) in the sense that the minimum time period required is two, while (8.1) requires $T \geq 3$. However, (8.1) and (8.2) have different asumptotic variances. Under the assumption that y_{i0} random with a stationary distribution (Section 4) the asymptotic variance of (8.1) is

$$(8.3)\quad \text{asy. var}\,[\sqrt{N}(\hat{\beta}_{IV} - \beta)] = \frac{4}{T-2}\cdot\frac{(1-\beta^2)(1-\beta)}{(1+\beta^2)^2}\ ,$$

the asymptotic variance of (8.2) is

$$(8.4)\quad \text{asy. var}\,[\sqrt{N}(\tilde{\beta}_{IV} - \beta)] = \frac{2(1+\beta)}{T-1}\left[\frac{1}{1-\beta} + \frac{1+\beta}{(1-\beta)^2}\frac{1}{\lambda}\right]\ .$$

Therefore, (8.1) is preferred to (8.2) if

$$(8.5)\quad \frac{2(T-1)}{T-2} < \frac{(1+\beta^2)^2}{(1-\beta)^4}\left[(1-\beta) + (1+\beta)\frac{1}{\lambda}\right]\ .$$

Without knowledge of β and λ, there is little to choose between these two estimators. However, it appears that (8.5) is more likely to be satisfied if β is positive. Thus, as a rough rule of thumb, we may want to use (8.1) if there is prior belief that successive observations are positively correlated and use (8.2) if successive observations are negatively correlated.

As we have seen, different assumptions about the initial observations do not affect the consistency of the MLE's when T tends to infinity. However, a typical panel usually involves a large number of individuals, but only over a short period of time. As it turns out, the properties of the MLE depend crucially on the assumption of the initial conditions. Different assumptions about the initial conditions call for different methods to obtain the MLE. Mistaking one case for the other in general will not lead to asymptotically equivalent formulas. Consequently, the misused estimator may be inconsistent. Unfortunately, usually we have little information to rely upon in making a correct choice of the initial conditions. Estimator (8.1) or (8.2) has the advantage that it is consistent independent of what the initial conditions are. Thus, the instrumental variable method, although is less efficient, does have its merit. Furthermore, if we know the correct choice of the initial conditions, we can always use the instrumental variable estimates as the initial value to start the iterative process to obtain the more efficient MLE.

9. Conclusions

In this paper we have studied the problems of estimating a dynamic model with error components in panel data when either the number of time point T or the number of cross-sectional unit N tends to infinity. We examined several models arising from different assumptions about the initial conditions. We attempted an interpretation and studied the properties of the MLE's and covariance estimators for each of these models. The main conclusions may be summarized in the following table.[5/]

As we can see from Table, the MLE is consistent when T tends to infinity no matter what are the assumptions about the initial conditions. When T is fixed and N tends to infinity the consistency of the MLE will depend on the assumptions about the initial conditions.

On the other hand the covariance estimators always use the same estimation method no matter what the initial conditions are. When T tends to infinity it is always consistent. When T is fixed it is always inconsistent no matter how large N is and no matter what are our initial conditions. Because the justification of using the covariance estimator for a dynamic model mainly rests on the asymptotic properties as the length of series T tends to infinity and the typical panel has a large number of individuals observed over a short period, it appears that the case for the use of the covariance estimator is not favorable.

Although we favor the use of the MLE because its desirable asymptotic properties (with T or N or both tend to infinity) in most circumstances, the computations of the MLE's are complicated. In the special case where the individual effects may be viewed as the effect of the initial

observation or initial endowment (corrected for the mean) and affect the individual equilibrium level (Section 6), the conditional MLE becomes very simple. We only need to modify the dependent variable as the actual subtracting the initial observation and apply the least squares regression to the transformed model.

It should be noted that the method of obtaining the MLE is different under different assumptions about the initial conditions. Mistaking one case for the other will not give us a consistent estimator no matter how large N is. A simple instrumental variable method was therefore suggested in Section 8. Although it is less efficient, it does have the advantage that it is consistent independent of what the initial conditions are.

Table 1

Interpretations and Statistical Properties of the MLE's
and CV's for Models Under Different Assumptions about the Initial Observations

		y_{i0} fixed		y_{i0} random					
				with a stationary distribution		with different means		with a common mean	
Interpretation of the Model		A cross-sectional unit starts from an arbitrary initial position and gradually drift towards its mean or no individual effects at the initial period but shows up at all later periods		All cross-sectional observation are random realizations of a stochastic process with same distribution but different levels and the initial observation is no different from any other observations		A cross-sectional unit may start at some position and gradually move toward its equilibrium level. But the individual equilibrium level affects the starting value		The starting value corrected for the mean may be viewed as the initial endowment. Depends on the assumption, the initial endowment may or may not affect the equilibrium level	
		$T \to \infty$, N fixed	T fixed, $N \to \infty$	$T \to \infty$, N fixed	T fixed, $N \to \infty$	$T \to \infty$, N fixed	T fixed, $N \to \infty$	$T \to \infty$, N fixed	T fixed, $N \to \infty$
Statistical properties of	MLE	Consistent	Consistent	Consistent	Consistent	Consistent	Inconsistent	Consistent	Consistent
	CV	Consistent	Inconsistent	Consistent	Inconsistent	Consistent	Inconsistent	Consistent	Inconsistent

Footnotes

1/ We assume no correlation between the unobserved effects and the observed explanatory variables; this assumption is unlike that of Mundlak [1978]. As will be discussed later, we essentially follow a different interpretation of the model from that of Mundlak [1978].

2/ The stationarity assumption may be relaxed when T is fixed and N tends to infinity (e.g., see Anderson [1978]). We keep this assumption for simplicity of exposition and because it allows us to provide a unified approach towards various assumptions about the initial conditions to be discussed later.

3/ Note that we use σ_u^2 in place of $\lambda\sigma^2$ and σ_γ^2 in place of $\sigma^2/(1-\beta)^2$ in this section for ease of exposition.

4/ Note that if the original model contains an intercept term, the conditional MLE can only provide a consistent estimate of c subtracting the intercept. Neither can the conditional MLE distinguish λ and σ^2. The unconditional MLE can distinguish the intercept, c, λ and σ^2.

5/ In Table 1, the MLE for y_{i0} random with different means should be interpreted as the interior solution. See Section 5.

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.

Ameyiya, T. [1973], "Regression Analysis When the Dependent Variable Is Truncated Normal," Econometrica, 41, 997-1016.

Anderson, T.W. [1971], The Statistical Analysis of Time Series, New York: John Wiley and Sons, Inc.

Anderson, T.W. [1978], "Repeated Measurments on Autoregressive Processes," Journal of the American Statistical Association, 73, 371-378.

Anderson, T.W. and J.B. Taylor [1979], "Strong Consistency of Least Squares Estimates in Dynamic Models," Annals of Statistics, 7, 484-489.

Balestra, P. and M. Nerlove [1966], "Pooling Cross-Section and Time Series Date in the Estimation of a Dynamic Model: The Demand for Natural Gas," Econometrica, 34, 585-612.

Chamberlain, G. [1979], "Heterogeneity, Omitted Variable Bias, and Duration Dependence," Discussion Paper No. 691, Harvard Institute of Economic Research, Harvard University.

Lee, L.F. [1979], "Efficient Estimation of Dynamic Error Components Models with Panel Data", Discussion paper No. 79-118 Center, for Economic Research, University of Minnesota.

Maddala, G.S. [1971], "The Use of Variance Components Models in Pooling Cross-Section and Time Series Data," Econometrica, 39, 341-358.

Mundlak, Y. [1978], "On the Pooling of Time Series and Cross Section Data," Econometrica, 36, 69-84.

Nerlove, M. [1971], "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections," Econometrica, 39, 359-382.

Swamy, P.A.V.B. and S.S. Arora [1972], "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models," Econometrica, 40, 261-276.

Wallace, T.D. and A. Hussain [1969], "The Use of Error Components Models in Combining Cross Section with Time Series Data," Econometrica, 37, 55-72.