EFFICIENT ESTIMATION OF DYNAMIC ERROR COMPONENTS MODELS WITH PANEL DATA

by

Lung-Fei Lee

Discussion Paper No. 79-118, October 1979

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, Minnesota 55455

EFFICIENT ESTIMATION OF DYNAMIC ERROR COMPONENTS MODELS WITH PANEL DATA

Lung-Fei Lee*

1. Introduction

In Balestra and Nerlove [1966], error components models are introduced as useful models for pooling cross sections of time series data. Error components regression models have subsequently been analyzed in Wallace and Hussain [1969], Maddala [1971], Nerlove [1971] among others. Dynamic error components models have also been introduced in Balestra and Nerlove [1967]. Difficulties in estimating the dynamic error components models were pointed out in Balestra and Nerlove [1967], Nerlove [1971] and Trognon [1978]. For dynamic models with exogenous variables, the instrumental variables approach was introduced in Balestra and Nerlove. In time series models without exogenous variables, conditional maximum likelihood procedures are suggested. However, these procedures are far from satisfactory. The instrumental variables estimates are consistent but not efficient. For panel data with large units of cross sections but short time periods, conditional maximum likelihood procedures are inconsistent.

In this article, we will consider efficient estimation of the dynamic error components model with and without exogenous variables.

---

*This research is supported by Grant SOC-78-07304 from the National Science Foundation.

We consider panel data with a large number of cross-sectional units but short time periods; specifically, the number of time periods T are assumed to be fixed and greater than three.

2. AR(1) process and error components model - Conditional and Unconditional MLE.

The following first order stationary autoregressive process with the error components structure, as introduced in Balestra and Nerlove [1967], can be regarded as the simplest dynamic model,

$$(2.1)\qquad \begin{aligned} y_{it} &= \rho y_{it-1} + \varepsilon_{it} && i=1,\ldots,N \quad |\rho| < 1 \\ &= \rho y_{it-1} + u_i + w_{it} && t=1,\ldots,T \end{aligned}$$

where $u_i \sim \text{iid } N(0, \sigma_u^2)$, $w_{it} \sim \text{iid } N(0, \sigma_w^2)$ and $E(u_i w_{jt}) = 0$ for all i, j and t. The error components structure introduced in this model allows for heterogeneity in the different units i.

Balestra and Nerlove considered a generalized least squares estimation procedure (GLS) applied to (2.1) with the y_{i1}'s as initial values. This procedure is a conditional maximum likelihood procedure conditional on y_{i1}. It is well known in the time series literature (see e.g. Pierce [1971]), that as T goes to infinity, conditional maximum likelihood estimates (MLE) are consistent and asymptotically efficient. However this will not hold for our model (2.1) as T is fixed and the y_{i1}'s are generated by the same stochastic process. The inconsistency of the conditional MLE is shown as follows:

Without loss of generality, assume that σ_u^2 and σ_w^2 are known. Let $y_i = (y_{i2},\ldots,y_{iT})'$ and $x_i = (y_{i1},\ldots,y_{iT-1})'$. The conditional MLE $\hat{\rho}_G$ of ρ_G is

$$(2.2)\qquad \hat{\rho}_G = [\Sigma_{i=1}^N x_i'(I_{T_1} - c\ell_{T_1}\ell_{T_1}')x_i]^{-1}\Sigma_{i=1}^N x_i'(I_{T_1} - c\ell_{T_1}\ell_{T_1}')y_i$$

where $T_1 = T-1$, $c = \dfrac{\sigma_u^2}{T_1\sigma_u^2 + \sigma_w^2}$, I_{T_1} is a $T_1 \times T_1$ identity matrix and ℓ_{T_1} is a $T_1 \times 1$ vector with unit components. Denote $x_{i.}' = (x_{i2},\ldots,x_{iT})$. It follows

$$\hat{\rho}_G - \rho_G = [\frac{1}{N}\Sigma_{i=1}^N \Sigma_{t=2}^T x_{it}' x_{it} - T_1^2 c \frac{1}{N} \Sigma_{i=1}^N \bar{x}_{i.}' \bar{x}_{i.}]^{-1} (\frac{1}{N}\Sigma_{i=1}^N \Sigma_{t=2}^T x_{it}' \varepsilon_{it}$$

$$- T_1^2 c \frac{1}{N}\Sigma_{i=1}^N \bar{x}_{i.}' \bar{\varepsilon}_{i.})$$

where $\bar{x}_{i.} = \frac{1}{T_1}\Sigma_{t=2}^T x_{it}$ and $\bar{\varepsilon}_{i.} = \frac{1}{T_1}\Sigma_{t=2}^T \varepsilon_{it}$ are the sample means of x_{it} and ε_{it} respectively. If $\hat{\rho}_G$ were consistent, $\frac{1}{N}\Sigma_{i=1}^N x_{it}'\varepsilon_{it}$ and $\frac{1}{N}\Sigma_{i=1}^N \bar{x}_{i.}' \bar{\varepsilon}_{i.}$ would converge to zero in probability. Since

$$\begin{aligned} x_{it} &= y_{it-1} \\ &= u_i/(1-\rho) + w_{it-1}/(1-\rho L) \end{aligned}$$

where L, $Lw_{it} = w_{it-1}$, is the left shift operator. Then by the law of large numbers,

$$\begin{aligned} \underset{N\to\infty}{\text{plim}} \frac{1}{N} \Sigma_{i=1}^N \bar{x}_{i.}' \bar{\varepsilon}_{i.} &= \underset{N\to\infty}{\text{plim}} \frac{1}{N} \Sigma_{i=1}^N (\frac{u_i}{1-\rho} + \frac{1}{T_1} \Sigma_{t=2}^N \frac{w_{it-1}}{1-\rho L})(u_i + \frac{1}{T_1} \Sigma_{t=2}^T w_{it}) \\ &= \underset{N\to\infty}{\text{plim}} \frac{1}{N} \Sigma_{i=1}^N \frac{u_i^2}{1-\rho} + \underset{N\to\infty}{\text{plim}} \frac{1}{N} \Sigma_{i=1}^N (\frac{1}{T_1} \Sigma_{t=2}^T \frac{w_{it-1}}{1-\rho L}) \\ &\quad \cdot (\frac{1}{T_1} \Sigma_{t=2}^T w_{it}) \\ &= \frac{\sigma_u^2}{1-\rho} + \underset{N\to\infty}{\text{plim}} \frac{1}{N} \Sigma_{i=1}^N (\frac{1}{T_1} \Sigma_{t=2}^T w_{it})(\Sigma_{s=2}^{T-1} \rho_s w_{is}) \\ &= \frac{\sigma_u^2}{1-\rho} + \frac{1}{T_1} \Sigma_{s=1}^{T-1} \rho_s \sigma_w^2 \end{aligned}$$

where $\rho_s = \frac{1}{T_1}\Sigma_{t=0}^{T-s-1}\rho^t = \frac{1}{T_1(1-\rho)}(1-\rho^{T-s})$. Since

$$\frac{1}{T_1}\Sigma_{s=2}^{T-1}\rho_s = \frac{1}{T_1^2(1-\rho)}\left[T_1-1-\frac{\rho}{(1-\rho)}(1-\rho^{T_1-1})\right]$$

$$= \frac{1}{T_1^2(1-\rho)^2}(T_1-1-T_1\rho+\rho^{T_1}) .$$

Hence

$$\underset{N\to\infty}{\text{plim}}\frac{1}{N}\Sigma_{i=1}^N\bar{x}_{i\cdot}'\bar{\varepsilon}_{i\cdot} = \frac{\sigma_u^2}{1-\rho}+\frac{\sigma_w^2(T_1-1-T_1\rho+\rho^{T_1})}{T_1^2(1-\rho)^2} .$$

Evidently,

$$\underset{N\to\infty}{\text{plim}}\frac{1}{N}\Sigma_{i=1}^N\Sigma_{t=2}^T x_{it}\varepsilon_{it} = \underset{N\to\infty}{\text{plim}}\frac{1}{N}\Sigma_{i=1}^N\Sigma_{t=2}^T y_{it-1}(u_i+w_{it})$$

$$= \underset{N\to\infty}{\text{plim}}\frac{T_1}{N}\Sigma_{i=1}^N\left(\frac{u_i}{1-\rho}+\frac{1}{T_1}\Sigma_{t=2}^T\frac{w_{it-1}}{1-\rho L}\right)u_i$$

$$= \frac{T_1\sigma_u^2}{1-\rho}$$

It follows that

$$\underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\Sigma_{i=1}^N\Sigma_{t=2}^T x_{it}'\varepsilon_{it} - T_1^2 c\,\frac{1}{N}\Sigma_{i=1}^N\bar{x}_{i\cdot}'\bar{\varepsilon}_{i\cdot}\right)$$

$$= \frac{T_1\sigma_u^2}{1-\rho}\,\frac{\sigma_w^2}{T_1\sigma_u^2+\sigma_w^2} - \frac{c\sigma_w^2(T_1-1-T_1\rho+\rho^{T_1})}{(1-\rho)^2}$$

$$= \frac{\sigma_u^2\sigma_w^2(1-\rho^{T_1})}{(1-\rho)^2(T_1\sigma_u^2+\sigma_w^2)}$$

which is positive. Since $\Sigma_{i=1}^N x_i'(I_{T_1} - c\ell_{T_1}\ell_{T_1}')x_i > 0$, $\hat{\rho}_G > \rho$ i.e., the conditional MLE is biased upward. The inconsistency of the conditional maximum likelihood procedure points out the importance of the initial values in estimating models with short time panel data.[1]

---

[1] The inconsistency of the conditional MLE has been pointed out in Chamberlain [1978] for $T = 2$. For discrete panel data, similar problems also occur; see Chamberlain [1978] and Heckman [1979].

In contrast to the conditional MLE approach, the unconditional MLE is consistent, asymptotic normal and efficient if the sample y_i (T multivariate random variables) $i=1,\ldots,N$ are i.i.d. (see Rao [1973]) and if the model is identifiable. Evidently this model is identifiable only if $T \geq 3$; consequently, the appropriate method of estimation is the unconditional MLE applied to (2.1). Equation (2.1) implies

$$(2.1)' \qquad y_{it} = \frac{u_i}{1-\rho} + \frac{w_{it}}{1-\rho L}$$

$$= u_i^* + w_{it}^*$$

where $u_i^* = \frac{u_i}{1-\rho} \sim N(0, \frac{\sigma_u^2}{(1-\rho)^2})$ and $w_{it}^* = \rho w_{it-1}^* + w_{it}$ is an AR(1) process. Let $\sigma_{u*}^2 = \frac{\sigma_u^2}{(1-\rho)^2}$, $\sigma_{w*}^2 = \frac{\sigma_w^2}{(1-\rho^2)}$ be the variances of u_i^* and w_{it}^* respectively, and let $\Omega = \sigma_{u*}^2 \ell_T \ell_T' + \sigma_{w*}^2 V$ be the variance matrix of y_i where

$$V = \begin{pmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ \rho^{T-1} & \cdots & \cdots & 1 \end{pmatrix}$$

The log likelihood function is

$$(2.3) \qquad L(\rho, \sigma_{u*}^2, \sigma_{w*}^2) = -\frac{NT}{2} \ln(2\Pi) - \frac{N}{2} \ln |\Omega| - \frac{1}{2} \Sigma_{i=1}^N y_i' \Omega^{-1} y_i$$

This exact likelihood function is relatively more complicated than the conditional one; however, a simple algorithm exists for the derivation of the unconditional MLE. To estimate (2.3), it is convenient to reparameterize the model. Let $\sigma_R^2 = \sigma_{u*}^2 / \sigma_{w*}^2$ be ratio of the two variances. As

$$|\Omega| = |\sigma_{w*}^2(\sigma_R^2 \ell_T \ell_T' + V)|$$

$$= \sigma_{w*}^{2T}|V| \; |1 + \sigma_R^2 \ell_T' V^{-1} \ell_T|$$

and

$$\Omega^{-1} = \frac{1}{\sigma_{w*}^2} (V^{-1} - V^{-1}\ell_T(\frac{1}{\sigma_R^2} + \ell_T' V^{-1}\ell_T)^{-1}\ell_T' V^{-1}) ,$$

the log likelihood function (2.3) becomes

$$(2.3)' \quad L(\rho, \sigma_{w*}^2, \sigma_R^2) = -\frac{NT}{2}\ell n(2\Pi) - \frac{NT}{2}\ell n\, \sigma_{w*}^2 - \frac{N}{2}\ell n|V| - \frac{N}{2}\ell n|1$$

$$+ \sigma_R^2 \ell_T' V^{-1}\ell_T| - \frac{1}{2\sigma_{w*}^2}[\Sigma_{i=1}^N y_i' V^{-1} y_i$$

$$- (\frac{1}{\sigma_R^2} + \ell' V^{-1}\ell_T)^{-1}\Sigma_{i=1}^N (y_i' V^{-1}\ell_T)^2]$$

$$= -\frac{NT}{2}\ell n(2\Pi) - \frac{NT}{2}\ell n\sigma_{w*}^2 - \frac{N}{2}\ell n|V| -$$

$$- \frac{N}{2}\ell n|1 + \sigma_R^2 \ell_T' V^{-1}\ell_T| - \frac{1}{2\,\sigma_{w*}^2}[\Sigma_{i=1}^N y_i' V^{-1} y_i$$

$$- \Sigma_{i=1}^N y_i' V^{-1}\ell_T \ell_T' V^{-1} y_i / \ell_T' V^{-1}\ell_T$$

$$+ (1 + \sigma_R^2 \ell_T' V^{-1}\ell_T)^{-1}\Sigma_{i=1}^N y_i' V^{-1}\ell_T \ell_T' V^{-1} y_i / \ell_T' V^{-1}\ell_T]$$

Denote $\phi = 1 + \sigma_R^2 \ell_T' V^{-1}\ell_T$. The MLE of $(\hat{\sigma}_{w*}^2, \hat{\sigma}_R^2, \hat{\rho})$ can be derived by solving the first order conditions. The first order conditions $\frac{\partial L}{\partial \sigma_{w*}^2} = 0$ and $\frac{\partial L}{\partial \phi} = 0$ imply respectively the following relationships,

$$(2.4)\qquad \sigma_{w*}(\rho) = \frac{1}{NT}\,(\Sigma_{i=1}^{N} y_i'V^{-1}y_i - \Sigma_{i=1}^{N} y_i'V^{-1}\ell_T\ell_T'V^{-1}y_i/\ell_T'V^{-1}\ell_T$$

$$+ \phi^{-1}\Sigma_{i=1}^{N} y_i'V^{-1}\ell_T\ell_T'V^{-1}y_i/\ell_T'V^{-1}\ell_T)$$

$$(2.5)\qquad \phi(\rho) = \frac{1}{N\sigma_{w*}^2}\;\Sigma_{i=1}^{N} y_i'V^{-1}\ell_T\ell_T'V^{-1}y_i/\ell_T'V^{-1}\ell_T$$

The estimates σ_{w*}^2 and ϕ can be solved explicitly from (2.4) and (2.5) in terms of ρ as

$$(2.6)\qquad \hat{\sigma}_{w*}^2(\rho) = \frac{1}{N(T-1)}\,(\Sigma_{i=1}^{N} y_i'V^{-1}y_i - \Sigma_{i=1}^{N} y_i'V^{-1}\ell_T\ell_T'V^{-1}y\,/\ell_T'V^{-1}\ell_T)$$

and

$$(2.7)\qquad \hat{\phi}(\rho) = (T-1)\;\Sigma_{i=1}^{N} y_i'V^{-1}\ell_T\ell_T'V^{-1}y_i/(\Sigma_{i=1}^{N} y_i'V^{-1}y_i\,\ell_T'V^{-1}\ell_T$$

$$- \Sigma_{i=1}^{N} y_i'V^{-1}\ell_T\ell_T'V^{-1}y_i)$$

Substituting equations (2.6) and (2.7) into (2.3)', we have the following concentrated likelihood function

$$(2.8)\qquad L(\rho) = k - \frac{N}{2}\ln|V| + \frac{NT}{2}\,\ln(\ell_T'V^{-1}\ell_T) - \frac{N}{2}\,\ln\Sigma_{i=1}^{N}(y_i'V^{-1}\ell_T)^2$$

$$- \frac{N(T-1)}{2}\,\ln(\ell_T'V^{-1}\ell_T\cdot\Sigma_{i=1}^{N} y_i'V^{-1}y_i - \Sigma_{i=1}^{N} y_i'V^{-1}\ell_T\ell_T'V^{-1}y_i)$$

where $k = -\frac{NT}{2}\,\ln(2\Pi) + \frac{NT}{2}\,\ln(T-1)N - \frac{N}{2}\,\ln(T-1) - \frac{NT}{2}$.

As $|V| = (1-\rho^2)^{T-1}$, $\ell_T'V^{-1}\ell_T = (T-(T-2)\rho)/(1+\rho)$, $\ell_T'V^{-1}y_i = (1-\rho)T\bar{y}_{i.}$ $+ \rho(y_{i1} + y_{iT})$ and $y_i'V^{-1}y_i = y_{i1}^2 + \Sigma_{t=2}^{T}(y_{it}-y_{it-1})^2/(1-\rho^2)$

(2.8)'
$$L(\rho) = k - \frac{N(T-1)}{2}\,\ell n(1-\rho^2) + \frac{NT}{2}\,\ell n(\frac{T-(T-2)\rho}{1+\rho})$$
$$- \frac{N}{2}\,\ell n \Sigma_{i=1}^{N}((1-\rho)T\bar{y}_{i.} + \rho(y_{i1} + y_{iT}))^2$$
$$- \frac{N(T-1)}{2}\,\ell n[\frac{T-(T-2)\rho}{1+\rho}\quad \Sigma_{i=1}^{N}(y_{i1}^2 + \frac{\Sigma_{t=2}^{T}(y_{it}-y_{it-1})^2}{(1-\rho^2)})$$
$$- \Sigma_{i=1}^{N}((1-\rho)T\bar{y}_{i.} + \rho(y_{i1} + y_{iT})^2)]$$

Thus the MLE $\hat{\rho}$ is formed by searching for the maximum value of $L(\rho)$ on (-1, 1). One can divide the admissible range (-1, 1) by M equidistant points ρ_i, i=1,...,M. For each ρ_i, evaluate $L(\rho_i)$ and choose as the estimator of ρ, say $\hat{\rho}$, that maximizes $L(\rho_i)$. For a given finite sample, there is a necessary condition for the existence of the ML estimates $\hat{\sigma}_{w*}^2$, $\hat{\sigma}_R^2$ and $\hat{\rho}$. This condition follows from equation (2.7). Equation (2.7) implies

(2.9)
$$\sigma_R^2 = (\phi-1)/\ell_T{}'V^{-1}\ell_T$$
$$= [T\Sigma_{i=1}^{N}y_i'V^{-1}\ell_T\ell_T{}'V^{-1}y_i - (\ell_T{}'V^{-1}\ell_T)\Sigma_{i=1}^{N}y_i'V^{-1}y_i]/d^2$$

where $d^2 = (\ell_T{}'V^{-1}\ell_T)(\ell_T{}'V^{-1}\ell_T\Sigma_{i=1}^{N}y_i'V^{-1}y_i - \Sigma_{i=1}^{N}y_i'V^{-1}\ell_T\ell_T{}'V^{-1}y_i)$ which is positive with probability one since $\ell_T{}'V^{-1}\ell_T \cdot y_i'V^{-1}y_i > y_i'V^{-1}\ell_T\ell_T{}'V^{-1}y_i$ by the Schwartz inequality. The necessary condition is $g_n(\rho) > 0$ for some $\rho \varepsilon (-1, 1)$ where

(2.10)
$$g_N(\rho) = T\Sigma_{i=1}^{N}y_i{}'V^{-1}\ell_T\ell_T{}'V^{-1}y_i - (\ell_T{}'V^{-1}\ell_T)\Sigma_{i=1}^{N}y_i'V^{-1}y_i \quad .$$

If this condition is not satisfied, a positive estimate of $\sigma^2_{w_R}$ does not exist. Thus for a given sample, one needs to search for the maximum value of $L(\rho)$ over the region $\{\rho \in (-1, 1) | g_N(\rho) > 0\}$. It is easy to check that by the strong law of large numbers, $\underset{N\to\infty}{\text{plim}}\ g_N(\rho) > 0$ for all $\rho \in (-1,1)$ when $T > 1$. Once the MLE $\hat{\rho}$ is determined, the MLE of $\hat{\sigma}^2_R$ and $\hat{\sigma}^2_{w*}$ follow from (2.9) and (2.6). Under the assumption that the true parameter $\theta = (\sigma^2_{w*}, \sigma^2_R, \rho)'$ is an interior point of the parameter space $\Theta \subseteq R^3$ which is compact, the MLE $\hat{\theta}$ is consistent, asymptotic normal and asymptotic efficient (see Rao [1973], Jennrich [1969]). As shown in Magnus [1978],

$$\sqrt{N}\ (\hat{\theta}-\theta) \overset{D}{\to} M(0,\ 2\Phi_\theta^{-1})$$

in distribution, where

$$(2.11) \qquad \Phi_\theta = \frac{\partial(\text{vec}\,\Omega^{-1})'}{\partial\theta}\ (\Omega\otimes\Omega)\ \frac{\partial\text{vec}\,\Omega^{-1}}{\partial\theta'}$$

Since $(\Phi_\theta)_{ij} = \text{tr}(\Omega^{-1}\frac{\partial\Omega}{\partial\theta}\Omega^{-1}\frac{\partial\Omega}{\partial\theta_j})$, $\theta' = (\theta_1, \theta_2, \theta_3) = (\sigma^2_{w*}, \sigma^2_R, \rho)$,

the expressions of elements of Φ_θ can be simplified as

$$(2.12) \qquad \text{tr}(\Omega^{-1}\frac{\partial\Omega}{\partial\sigma^2_{w*}}\ \Omega^{-1}\frac{\partial\Omega}{\partial\sigma^2_{w*}}) = \frac{T}{\sigma^4_\varepsilon}$$

$$(2.13) \qquad \text{tr}(\Omega^{-1}\frac{\partial\Omega}{\partial\sigma^2_{w*}}\ \Omega^{-1}\frac{\partial\Omega}{\partial\sigma^2_R}) = \frac{1}{\sigma^2_{w*}\sigma^2_R}\ \ell_T'V^{-1}\ell_T$$

$$(2.14) \qquad \text{tr}(\Omega^{-1}\frac{\partial\Omega}{\partial\sigma^2_R}\ \Omega^{-1}\frac{\partial\Omega}{\partial\sigma^2_R}) = \frac{1}{\sigma^4_R}(\ell_T'V^{-1}\ell_T)^2$$

(2.15) $$\mathrm{tr}(\Omega^{-1}\frac{\partial\Omega}{\partial\sigma^2_{w*}}\Omega^{-1}\frac{\partial\Omega}{\partial\rho}) = \frac{1}{\sigma^2_{w*}}[(\frac{1}{\sigma^2_R}+\ell_T'V^{-1}\ell_T)^{-1}(\frac{2\rho}{1-\rho^2}\ell_T'V^{-1}\ell_T + \frac{1}{1-\rho^2}\ell_T'D\ell_T) + \frac{2\rho(1-T)}{1-\rho^2}]$$

(2.16) $$\mathrm{tr}(\Omega^{-1}\frac{\partial\Omega}{\partial\sigma^2_R}\Omega^{-1}\frac{\partial\Omega}{\partial\rho}) = \frac{2(T-1)(1-\rho)}{(1+\sigma^2_R\ell_T'V^{-1}\ell_T)^2}$$

(2.17) $$\mathrm{tr}(\Omega^{-1}\frac{\partial\Omega}{\partial\rho}\Omega^{-1}\frac{\partial\Omega}{\partial\rho}) = \frac{2(T-1)(1+2\rho^2)}{1-\rho^2} - \frac{2\ell_T'\frac{\partial V^{-1}}{\partial\rho}V\frac{\partial V^{-1}}{\partial\rho}\ell_T}{(\frac{1}{\sigma^2_R}+\ell_T'V^{-1}\ell_T)} + \frac{(\ell_T'\frac{\partial V^{-1}}{\partial\rho}\ell_T)^2}{(\frac{1}{\sigma^2_R}+\ell_T'V^{-1}\ell_T)^2}$$

where $\ell_T'D\ell_T = -2(T-1) + 2(T-2)\rho$. The detail derivations of the above expressions are straightforward but tedious and are omitted here.

Since there is a one-one correspondence between $(\sigma^2_{w*}, \sigma^2_R, \rho)$ and $(\sigma^2_w, \sigma^2_u, \rho)$, the MLE of $(\sigma^2_w, \sigma^2_u, \rho)$ can be derived as

(2.18) $$\hat\sigma^2_w = (1-\hat\rho^2)\hat\sigma^2_{w*} .$$

(2.19) $$\hat\sigma^2_u = (1-\hat\rho)^2\hat\sigma^2_{w*}\hat\sigma^2_R .$$

By a Taylor expansion, one can show that

$$\sqrt{N}\left(\begin{pmatrix}\hat\sigma^2_w\\ \hat\sigma^2_u\\ \hat\rho\end{pmatrix} - \begin{pmatrix}\sigma^2_w\\ \sigma^2_u\\ \rho\end{pmatrix}\right) \xrightarrow{D} N(0,\ 2P\Phi_\theta^{-1}P')$$

where $$P = \begin{bmatrix} 1-\rho^2 & 0 & -2\sigma^2_{w*}\rho \\ (1-\rho)^2\sigma^2_R & (1-\rho)^2\sigma^2_{w*} & -2(1-\rho)\sigma^2_{w*}\sigma^2_R \\ 0 & 0 & 1 \end{bmatrix} .$$

3. Dynamic Models with Time Invariant Exogenous Variables

In many econometric models, exogenous variables are presented. In this section, we consider the following dynamic model as in (2.1) with the inclusion of exogenous variables:

$$(3.1) \qquad y_{it} = \rho y_{it-1} + x_{it}\beta + u_i + w_{it} \qquad \begin{array}{l} i=1,\ldots,N \\ t=1,\ldots,T \end{array}$$

where x_{it} is a $1\times k$ vector of stochastic exogenous variables, x_{it} are independent and identically distributed across different units i and x_{it}, u_i and w_{is} are mutually independent. Other assumptions are specified in (2.1). For this model, the initial values y_{i0} as well as past values of the exogenous variables x_{i0}, x_{i-1}, etc. have to be taken into account in deriving the exact likelihood function. Whether one can easily handle this initial value problem depends on the nature of the set of exogenous variables. In this section, we consider the case of the x_{it} being time invariant. For example, in household survey data, variables such as race, sex and family background variables will not change over time. The time variant exogenous variables case is considered in the next section.

Consider the case $x_{it} = x_i$ for all t. Equation (3.1) is

$$(3.2) \qquad y_{it} = \rho y_{it-1} + x_i\beta + u_i + w_{it} \qquad \begin{array}{l} i=1,\ldots,N \\ t=1,\ldots,T \end{array} \quad |\rho|<1$$

To derive the likelihood unconditional on y_{i0}, one notes that (3.2) can be rewritten as

$$(3.2)' \qquad y_{it} = \frac{x_i\beta}{1-\rho} + \frac{u_i}{1-\rho} + \frac{w_{it}}{1-\rho L}$$

$$= x_i\beta^* + u_i^* + w_{it}^*$$

where $\beta^* = \beta/(1-\rho)$, $u^*_i = u_i/(1-\rho)$ and $w_{it}^* = w_{it}/(1-\rho L)$. This equation is analogous to (2.1)' with the addition of exogenous variables. The log likelihood function is

(3.3) $$L(\rho, \sigma^2_{u*}, \sigma^2_{w*}, \beta^*)$$

$$= -\frac{NT}{2} \ln(2\Pi) - \frac{N}{2} \ln |\Omega| - \frac{1}{2} \Sigma_{i=1}^{N}(y_i - (I_T \otimes x_i)\beta^*)'\Omega^{-1}(y_i-(I_T \otimes x_i)\beta^*)$$

$$= -\frac{NT}{2} \ln(2\Pi) - \frac{NT}{2} \ln\sigma^2_{w*} - \frac{N}{2} \ln|V| - \frac{N}{2} \ln| 1 + \sigma^2_R \ell_T' V^{-1} \ell_T|$$

$$- \frac{1}{2\sigma^2_{w*}} [\Sigma_{i=1}^{N}(y_i-(I_T \otimes x_i)\beta^*)'V^{-1}(y_i-(I_T \otimes x_i)\beta^*)$$

$$- (\ell_T'V^{-1}\ell_T)^{-1}\Sigma_{i=1}^{N}((y_i-(I_T \otimes x_i)\beta^*)'V^{-1}\ell_T)^2$$

$$+ \quad 1+\sigma^2_R\ell_T'V^{-1}\ell_T)^{-1} \; \Sigma_{i=1}^{N}y_i'V^{-1}\ell_T\ell_T'V^{-1}y_i/(\ell_T'V^{-1}\ell_T)]$$

where the notation Ω, V etc., is the same as defined in the previous section. From the first order conditions, $\partial L/\partial\sigma^2_{w*} = 0$ and $\partial L/\partial\beta^* = 0$, we can solve for σ^2_{w*} and β^* in terms of σ^2_R and ρ.

(3.4) $$\sigma^2_{w*} = \frac{1}{NT}[\Sigma_{i=1}^{N}(y_i - (I_T \otimes x_i)\beta^*)'V^{-1}(y_i - (I_T \otimes x_i)\beta^*)$$

$$- (\ell_T'V^{-1}\ell_T)^{-1}\Sigma_{i=1}^{N}((y_i-(I_T \otimes x_i)\beta^*)'V^{-1}\ell_T)^2$$

$$+ (1+\sigma^2_R\ell_T'V^{-1}\ell_T)^{-1}\Sigma_{i=1}^{N}y_i'V^{-1}\ell_T\ell_T'V^{-1}y_i/(\ell_T'V^{-1}\ell_T)]$$

(3.5) $$\beta^* = [\Sigma_{i=1}^{N}(I_T \otimes x_i')(\sigma^2_R\ell_T\ell_T' + V)^{-1}(I_T \otimes x_i)]^{-1}\Sigma_{i=1}^{N}(I_T \otimes x_i')$$

$$(\sigma^2_R\ell_T\ell_T' + V)^{-1}y_i$$

Hence conditional on ρ and σ_R^2, the maximum likelihood estimates $\hat{\beta}^*(\rho, \sigma_R^2)$ and $\hat{\sigma}_{w*}^2(\rho, \sigma_R^2)$ can be derived from (3.5) and (3.4). Substituting $\hat{\beta}^*(\rho, \sigma_R^2)$ and $\hat{\sigma}_{w*}(\rho, \sigma_R^2)$ into (3.3), we have the concentrated likelihood function $L(\rho, \sigma_R^2)$. The maximum likelihood estimates are derived by searching for the maximum value of $L(\rho, \sigma_R^2)$ on the parameters s ace $(-1, 1) \times (0, \infty)$. For certain search procedures, it is convenient to have a bounded parameter space. For this model, one can reparameterize σ_R^2 into $\lambda = (1 + \sigma_R^2 \ell_T' V^{-1} \ell_T)^{-1}$. Since there is a one-one correspondence between (ρ, σ_R^2) and (ρ, λ), maximum likelihood estimates can be found by searching on the parameter space $(\rho, \lambda) \in (-1,1) \times (0, 1)$. Equations (3.4) and (3.5) can be rewritten as

$$(3.4)' \quad \hat{\sigma}_{w*}^2(\rho,\lambda) = \frac{1}{NT}[\Sigma_{i=1}^N (y_i-(I_T \otimes x_i)\beta^*)' V^{-1}(y_i - (I_T \otimes x_i)\beta^*) - (\ell_T' V^{-1} \ell_T)^{-1} \Sigma_{i=1}^N ((y_i-(I_T \otimes x_i)\beta^*)' V^{-1} \ell_T)^2 + \lambda \Sigma_{i=1}^N (y_i' V^{-1} \ell_T)^2 / \ell_T' V^{-1} \ell_T]$$

$$(3.5)' \quad \hat{\beta}(P,\lambda) = [\Sigma_{i=1}^N (I_T \otimes x_i')(V^{-1} - V^{-1}\ell_T \ell_T' V^{-1}/(\ell_T' V^{-1} \ell_T))(I_T \otimes x_i) + \lambda \Sigma_{i=1}^N (I_T \otimes x_i') V^{-1} \ell_T \ell_T' V^{-1} (I_T \otimes x_i)/(\ell_T' V^{-1} \ell_T)]^{-1} \cdot [\Sigma_{i=1}^N (I_T \otimes x_i')(V^{-1} - V^{-1}\ell_T \ell_T' V^{-1}/(\ell_T' V^{-1}\ell_T)) y_i + \lambda \Sigma_{i=1}^N (I_T \otimes x_i') V^{-1} \ell_T \ell_T' V^{-1} y_i/(\ell_T' V^{-1} \ell_T)]$$

and the concentrated likelihood function becomes

$$(3.6) \quad L(\rho,\lambda) = c - \frac{NT}{2} \ell n \hat{\sigma}_{w*}^2(\rho,\lambda) - \frac{N}{2} \ell n |V| + \frac{N}{2} \ell n \lambda$$

where $c = -\frac{NT}{2}(\ln(2\Pi) + 1)$. Let $(\hat{\rho}, \hat{\lambda})$ be the maximum likelihood estimates of (ρ, λ). The maximum likelihood estimate of σ_R^2 can be calculated as

$$(3.7) \qquad \hat{\sigma}_R^2 = (\frac{1}{\hat{\lambda}} - 1)/\ell_T' \hat{V}^{-1} \ell_T)$$

Let $\theta' = (\beta^{*\prime}, \sigma_{w*}^2, \sigma_R^2, \rho)$ and $\hat{\theta}$ be the maximum likelihood estimates of θ. Under the regularity condition that the true parameter vector θ is an interior point of the compact parameter space, the density function of x_i satisfies the regularity conditions in Rao [1973] and $\operatorname{plim}_{N\to\infty} \frac{1}{N} \Sigma_{i=1}^{N} (I_T \otimes x_i) \Omega^{-1} (I_T \otimes x_i') = \Sigma$ exists and is a positive definite matrix, $\hat{\theta}$ is

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{D} N(0, \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 2\Phi_\theta^{-1} \end{bmatrix})$$

where $(\Phi_\theta)_{ij} = \operatorname{tr}(\Omega^{-1} \frac{\partial \Omega}{\partial \theta_i} \,{}^{-1} \frac{\partial \Omega}{\partial \theta_j})$, i, j = 2,..4. Since $\beta^* = \beta/(1-\rho)$, the maximum likelihood estimate of β is

$$(3.7) \qquad \hat{\beta} = (1-\hat{\rho})\hat{\beta}^*.$$

It can be easily shown by the Taylor series expansion that

$$\sqrt{N}(\hat{\beta}-\beta) \xrightarrow{D} N(0, (1-\rho)^2 \Sigma^{-1} + 2(\Phi_\theta^{-1})_{44} \beta^* \beta^{*\prime})$$

Similarly, we can derive the maximum likelihood estimates of σ_u^2 and σ_w^2. Since $\sigma_{w*}^2 = \sigma_w^2/(1-\rho^2)$, $\sigma_{u*}^2 = \sigma_u^2/(1-\rho)^2$ and $\sigma_R^2 = \sigma_{u*}^2/\sigma_{w*}^2$, the maximum likelihood estimates are

$$(3.8) \qquad \hat{\sigma}_w^2 = (1-\hat{\rho}^2)\hat{\sigma}_{w*}^2$$

(3.9) $$\hat{\sigma}_u^2 = (1-\hat{\rho})^2 \hat{\sigma}_{w*}^2 \hat{\sigma}_R^2$$

and

$$\sqrt{N}(\hat{\sigma}_w^2 - \sigma_w^2) \xrightarrow{D} N(0,\ 2[1-\rho^2, 0, -2\sigma_{w*}^2\rho]\Phi_\theta^{-1}[1-\rho^2,\ 0,\ -2\sigma_{w*}^2\rho]')$$

$$\sqrt{N}(\hat{\sigma}_u^2 - \sigma_u^2) \xrightarrow{D} N(0,\ 2[(1-\rho)^2\sigma_R^2,\ (1-\rho)^2\sigma_{w*}^2,\ -2(1-\rho)\sigma_{w*}^2\sigma_R^2]\Phi_\theta^{-1}$$

$$\cdot\ [(1-\rho)^2\sigma_R^2, (1-\rho)^2\sigma_{w*}^2,\ -2(1-\rho)\sigma_{w*}^2\sigma_R^2]')$$

An alternative estimation procedure which is computationally simpler and provides asymptotic efficient estimates is as follows. Let $\varepsilon_{it} = y_{it} - x_i\beta^*$ and $\phi = 1 + \sigma_R^2\ \ell_T' V^{-1}\ell_T$. Similar to the derivation in the last section, conditional on ρ and β^*, the MLE of σ_{w*} and ϕ from maximizing the log likelihood function in (3.3) are

(3.10) $$\hat{\sigma}_{w*}^2(\rho,\ \beta^*) = \frac{1}{N(T-1)}(\Sigma_{i=1}^N \varepsilon_i' V^{-1}\varepsilon_i - \Sigma_{i=1}^N \varepsilon_i' V^{-1}\ell_T\ell_T' V^{-1}\varepsilon_i/\ell_T V^{-1}\ell_T)$$

(3.11) $$\hat{\phi}(\rho, \beta^*) = (T-1)\Sigma_{i=1}^N \varepsilon_i' V^{-1}\ell_T\ell_T' V^{-1}\varepsilon_i/(\Sigma_{i=1}^N \varepsilon_i' V^{-1}\varepsilon_i \ell_T' V^{-1}\ell_T$$

$$- \Sigma_{i=1}^N \varepsilon_i' V^{-1}\ell_T\ell_T' V^{-1}\varepsilon_i)$$

It follows that the concentrated likelihood function is $L(\rho,\ \beta^*)$ which has the same expression in (2.8) except that y_i should be replaced by ε_i . The estimation procedure consists of three steps:

Step 1. Estimate equation (3.2) by ordinary least squares (OLS),

$$\hat{\beta}_L^* = (T\Sigma_{i=1}^N x_i' x_i)^{-1}\ \Sigma_{i=1}^N \Sigma_{t=1}^T x_i' y_{it}$$

Step 2. Estimate ρ by $\hat{\rho}_L$ which maximizes $L(\rho,\ \beta_L^*)$ along (-1, 1). This can be done by dividing the admissible range (-1, 1) by M points

ρ_i $i=1,\ldots,M$ and choose the point $\hat{\rho}_L = \rho_i$ such that $L(\rho_i, \beta_L^*) \geq L(\rho_j, \hat{\beta}_L^*)$ $j=1,\ldots,M$. Estimate σ_{w*}^2 and ϕ by $\hat{\sigma}_{w*}^2(\hat{\rho}_L, \hat{\beta}_L^*)$ and $\hat{\phi}(\hat{\rho}_L, \hat{\beta}_L^*)$.

Step 3. Estimate β^* by generalized least squares (GLS)

$$\hat{\beta}_G^* = (T\Sigma_{i=1}^N x_i'\hat{\Sigma}^{-1}x_i)\Sigma_{i=1}^N \Sigma_{t=1}^T x_i'\hat{\Sigma}^{-1}y_{it}$$

where $\hat{\Sigma} = \hat{V}^{-1} - \hat{V}^{-1}\ell_T(\frac{1}{\hat{\sigma}_R^2} + \ell_T'\hat{V}^{-1}\ell_T)^{-1}\ell_T'\hat{V}^{-1}$. $\hat{\sigma}_R^2 = \frac{1}{\ell_T\hat{V}^{-1}\ell_T}(\hat{\phi}(\hat{\rho}_L, \hat{\beta}_L^*)-1)$.

Evidently, the estimates derived in the first two steps are consistent. It follows that $\hat{\beta}_G^*$ is asymptotically efficient (see Zellner [1962]). It remains to show that the estimate $\hat{\theta}_L = (\hat{\rho}_L, \hat{\phi}(\hat{\rho}_L, \hat{\beta}_L^*), \hat{\sigma}_{w*}^2(\hat{\rho}_L, \hat{\beta}_L^*))$ is asymptotic efficient. Let $\hat{\theta}_M$ be MLE of $(\rho, \phi, \sigma_{w*}^2)$. It is enough to show that $\sqrt{N}(\hat{\theta}_L - \theta)$ and $\sqrt{N}(\hat{\theta}_M - \theta)$ have the same limiting distribution. Since conditional on $\hat{\beta}_G^*$, $\hat{\theta}_L$ maximizes $L(\theta, \hat{\beta}_L^*)$, by Taylor expansion

$$0 = \frac{\partial L(\hat{\theta}_t, \hat{\beta}_L^*)}{\partial \theta}$$

$$= \frac{\partial L(\theta, \hat{\beta}_L^*)}{\partial \theta} + \frac{\partial^2 L(\theta_+, \hat{\beta}_L^*)}{\partial\theta\partial\theta'}(\hat{\theta} - \theta)$$

$$= \frac{\partial L(\theta, \beta_L^*)}{\partial \theta} + \frac{\partial^2 L(\theta, \beta_+^*)}{\partial\theta\partial\theta'}(\hat{\beta}_L^* - \beta^*) + \frac{\partial^2 L(\theta_+, \hat{\beta}_L^*)}{\partial\theta\partial\theta'}(\hat{\theta}_L - \theta)$$

where θ_+ lies between $\hat{\theta}_G$ and the true parameter θ; β_+^* lies between $\hat{\beta}_L^*$ and β^*. It follows

$$(3.12)\qquad \sqrt{N}(\hat{\theta}_L-\theta) \overset{D}{=} - [\frac{1}{N}\frac{\partial^2 L(\theta,\beta^*)}{\partial\theta\partial\theta'}]^{-1}(\frac{1}{\sqrt{N}}\frac{\partial L(\theta,\ \beta^*)}{\partial\beta} +$$

$$+ \frac{1}{N}\frac{\partial^2 L(\theta,\ \beta^*)}{\partial\theta\partial\beta^*}\cdot\sqrt{N}(\hat{\beta}^*_L - \beta^*))$$

$$\overset{D}{=} -[\frac{1}{N}\frac{\partial^2 L(\theta,\ \beta^*)}{\partial\theta\partial\theta'}]^{-1}\frac{1}{\sqrt{N}}\frac{\partial L(\theta,\ \beta^*)}{\partial\theta}$$

where $\overset{D}{=}$ means that both sides have the same asymptotic distribution. Similarly we have

$$(3.13)\qquad \sqrt{N}\ (\hat{\theta}_M - \theta) \overset{D}{=} - [\frac{1}{N}\frac{\partial^2 L(\theta,\beta^*)}{\partial\theta\partial\theta'}]^{-1}(\frac{1}{\sqrt{N}}\frac{\partial L(\theta,\ \beta^*)}{\partial\theta} + \frac{1}{N}\frac{\partial^2 L(\theta,\ \beta^*)}{\partial\theta\partial\beta^*}$$

$$\cdot\ \sqrt{N}(\tilde{\beta}^*_M - \beta^*))$$

$$\overset{D}{=} - [\frac{1}{N}\frac{\partial^2 L(\theta,\ \beta^*)}{\partial\theta\partial\theta'}]^{-1}\frac{1}{\sqrt{N}}\frac{\partial L(\theta,\ \beta^*)}{\partial\theta}$$

The second equalities hold in (3.12) and (3.13) since $\operatorname*{plim}_{N\to\infty}\frac{1}{N}\frac{\partial^2 L(\theta,\ \beta^*)}{\partial\theta\partial\beta^*} = 0$ and $\sqrt{N}\ (\hat{\beta}^*_L - \beta^*)$ has a limiting distribution. Hence $\sqrt{N}(\hat{\theta}_L - \theta)$ has the same limiting distribution as $\sqrt{N}(\hat{\theta}_M - \theta)$ and is asymptotic efficient.

4. Dynamic Model with Time Variant Exogenous Variables

When the exogenous variables change over time and one does not know the past history of these variables, the initial value problem arises. In this section, we assume that the x_{it} are stationary multivariate normal process with zero mean and provide an approach. The model (3.1) implies

$$(4.1) \qquad y_{i1} = \frac{\alpha}{1-\rho} + x_{i1}\beta + \rho x^{*}_{i0}\beta + u_i/(1-\rho) + w_{i1}/(1-\rho L)$$

$$y_{it} = \alpha + \rho y_{it-1} + x_{it}\beta + u_i + w_{it} \qquad t \geq 2$$

where $x^{*}_{i0} = x_{i0}/(1-\rho L)$. Under the assumption that x_{it} are normal, x^{*}_{i0} is normal; consequently it can be decomposed into two independent components:

$$(4.2) \qquad x^{*}_{i0} = x_{i1}\pi_1 + \ldots + x_{iT}\pi_T + v_i$$

where v_i is normal and independent of $x_{i1},\ldots,x_{iT}$. The π_t are functions of the variances of $x_{i1},\ldots,x_{iT}$ and covariances of x_{io}, $x_{i1},\ldots,x_{iT}$. With the relation in (4.2), equations (4.1) can be rewritten as

$$(4.1)' \qquad y_{i1} = \frac{\alpha}{1-\rho} + x_{i1}\beta + \rho x_{i1}\alpha_1 + \ldots + \rho x_{iT}\alpha_T + u_i/(1-\rho) + w^{*}_{i1}$$

$$y_{it} = \alpha + \rho y_{it-1} + x_{it}\beta + u_i + w_{it} \qquad t \geq 2$$

where $w^{*}_{i1} = v_i\beta\rho + w_{i1}/(1-\rho L)$ and $\alpha_t = \pi_t\beta$. Obviously, for a single time series, there are too many parameters to be identified; however, it is easy to see that the parameters β, ρ, α, α_t, σ^2_u, $\sigma^2_* = E(w^{*2}_{i1})$ are identifiable for panel data with large units of cross sections (at least

$N \geq T$). The parameters π_t, σ_v^2 will not be identified unless a more specific structure is imposed on the process $\{x_{it}\}$; for example, if $\{x_{it}\}$ is specified as an ARMA process with appropriate finite orders. The $T \times T$ variance matrix Ω for each unit i is

$$(4.3) \qquad \Omega = \sigma_u^2 \begin{pmatrix} \frac{1}{1-\rho} \\ 1 \\ 1 \end{pmatrix} [\frac{1}{1-\rho}\ 1 \ldots 1] + \begin{pmatrix} \sigma_*^2 & & & \\ & \sigma_w^2 & & \\ & & \ddots & \\ & & & \sigma_w^2 \end{pmatrix}$$

The inverse matrix and its determinant are

$$\Omega^{-1} = \begin{pmatrix} \frac{1}{\sigma_*^2} & & & \\ & \frac{1}{\sigma_w^2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_w^2} \end{pmatrix} - \left(\frac{1}{\sigma_u^2} + \frac{1}{(1-\rho)^2\sigma_*^2} + \frac{T-1}{\sigma_w^2} \right) \begin{pmatrix} \frac{1}{\sigma_*^2(1-\rho)} \\ \frac{1}{\sigma_w^2} \\ \vdots \\ \frac{1}{\sigma_w^2} \end{pmatrix} \cdot$$

$$\left[\frac{1}{\sigma_{w*}^2(1-\rho)}\ \frac{1}{\sigma_w^2} \cdots \frac{1}{\sigma_w^2}\right]$$

$$|\Omega| = \sigma_*^2 \sigma_w^{2(T-1)} \ \sigma_u^2 \left(\frac{1}{\sigma_u^2} + \frac{1}{\sigma_*^2(1-\rho)^2} + (T-1)/\sigma_w^2\right)$$

Let $\theta' = (\rho,\ \beta',\ \alpha,\ \alpha_1,\ldots,\alpha_T,\ \sigma_u^2,\ \sigma_w^2,\ \sigma_*^2)$ and

$$\varepsilon_i' = (y_{i1}-\alpha/(1-\rho)-x_{i1}\beta-\rho x_{i1}\alpha_1 \cdot\cdot -\rho x_{iT}\alpha_T,\ y_{i2}-\alpha-\rho y_{i1}-x_{i1}\beta,\ldots,$$

$$y_{iT}-\alpha-\rho y_{iT-1}-x_{iT}\beta).$$

The log likelihood function of y conditional on X is

$$(4.4) \qquad L(\theta) = -\frac{NT}{2}\ \ell n 2\Pi - \frac{N}{2}\ \ell n\, \sigma_*^2 - \frac{N(T-1)}{2}\ \ell n\, \sigma_w^2 - \frac{N}{2} \ell n\, \sigma_u^2$$

$$- \frac{N}{2} \ell n \left(\frac{1}{\sigma_u^2} + \frac{1}{\sigma_*^2 (1-\rho)^2} + (T-1)/\sigma_w^2\right) - \frac{1}{2} \Sigma_{i=1}^N \varepsilon_i' \Omega^{-1} \varepsilon_i$$

explicitly,

$$(4.4)' \qquad L(\theta) = -\frac{NT}{2}\ \ell n (2\pi) - \frac{N}{2} \ell n\, \sigma_*^2 - \frac{N(T-1)}{2}\ \ell n\, \sigma_w^2 - \frac{N}{2}\ \ell n\, \sigma_u^2$$

$$- \frac{N}{2} \ell n \left(\frac{1}{\sigma_u^2} + \frac{1}{\sigma_*^2 (1-\rho)^2} + \frac{T-1}{\sigma_w^2}\right)$$

$$- \frac{1}{2}\ \Sigma_{i=1}^N\ \frac{1}{\sigma_*^2} (y_{i1} - \frac{\alpha}{1-\rho} - x_{i1}\beta - \rho x_{i1}\alpha_1 \ldots - \rho x_{iT}\alpha_T)^2$$

$$+ \frac{1}{\sigma_w^2}\ \Sigma_{t=2}^T (y_{it} - \alpha - \rho y_{it-1} - x_{it}\beta)^2 + \frac{1}{2} \left(\frac{1}{\sigma_u^2} + \frac{1}{(1-\rho)^2 \sigma_*^2} + \frac{T-1}{\sigma_w^2}\right)^{-1}$$

$$\cdot\ \Sigma_{i=1}^N ((y_{i1} - \frac{\alpha}{1-\rho} - x_{i1}\beta - \rho x_{i1}\alpha_1\ \ldots - \rho x_{iT}\alpha_T)/\sigma_*^2 (1-\rho)$$

$$+ \frac{1}{\sigma_w^2}\ \Sigma_{t=2}^T (y_{it} - \alpha - \rho y_{it-1} - x_{it}\beta))^2$$

The MLE can then be derived by maximizing (4.4). For this model, initial consistent estimates can be derived for use as starting values for the iteration algorithm.

To derive a consistent estimate of (α, ρ, β), we follow the suggestion of Balestra and Nerlove [1967] and Nerlove [1971]. Using x_{it-1} as an instrument for y_{it-1}, we can estimate $\hat{\alpha}$, $\hat{\rho}$ and $\hat{\beta}$ as

$$(4.5) \qquad \begin{pmatrix} \hat{\alpha} \\ \hat{\rho} \\ \hat{\beta} \end{pmatrix} = \left[\sum_{i=1}^N \sum_{t=2}^T \begin{pmatrix} 1 \\ x_{it-1} \\ x_{it} \end{pmatrix} (1 \quad x_{it-1} \quad x_{it}) \right]^{-1} \Sigma_{i=1}^N \Sigma_{t=2}^T \begin{pmatrix} 1 \\ x_{it-1} \\ x_{it} \end{pmatrix} y_{it}$$

Since $\operatorname*{plim}_{N\to\infty} \frac{1}{N} \Sigma_{i=1}^N \Sigma_{t=2}^T \begin{pmatrix} 1 \\ x_{it-1} \\ x_{it} \end{pmatrix} (1 \quad x_{it-1} \quad x_{it})$ exists and is positive definite, the estimate $(\hat{\alpha}, \hat{\rho}, \hat{\beta})$ is consistent. Substituting the estimates $\hat{\alpha}, \hat{\rho}, \hat{\beta}$ into the following equation

$$(y_{i1} - \frac{\alpha}{1-\rho} - x_{i1}\beta)/\rho = x_{i1}\alpha_1 + \ldots + x_{iT}\alpha_T + \varepsilon_{i1} \qquad i=1,\ldots,N$$

the parameters $(\alpha_1,\ldots,\alpha_T)$ can be estimated by OLS,

$$(4.6) \qquad \begin{pmatrix}\hat{\alpha}_1\\ \vdots \\ \alpha_T\end{pmatrix} = [\Sigma_{i=1}^{N} \begin{pmatrix}x_{i1}\\ \vdots \\ x_{iT}\end{pmatrix}(x_{i1}\ldots x_{iT})]^{-1}\Sigma_{i=1}^{N}\begin{pmatrix}x_{i1}\\ \vdots \\ x_{iT}\end{pmatrix}(y_{i1} - \frac{\hat{\alpha}}{1-\hat{\rho}} - x_{i1}\hat{\beta})/\hat{\rho}$$

The variance $\sigma^2_{\varepsilon_1} = \sigma^2_u/(1-\rho) + \sigma^2_*$ can be estimated as

$$(4.7) \qquad \hat{\sigma}^2_{\varepsilon_1} = \frac{1}{N-T}\Sigma_{i=1}^{N}(y_{i1} - \frac{\hat{\alpha}}{1-\hat{\rho}} - x_{i1}\hat{\beta} - \hat{\rho}(x_{i1}\hat{\alpha}_1 + \ldots + x_{iT}\hat{\alpha}_T))^2 .$$

The variances σ^2_u and σ^2_w can be estimated in a manner similar to the analysis of variance in Graybill [1961, 1976] and Wallace and Hussain [1971]:

$$(4.8) \qquad \hat{\sigma}^2_w = \frac{1}{N(T-2)}\Sigma_{i=1}^{N}\Sigma_{t=2}^{T}(\tilde{\varepsilon}_{it} - \frac{1}{T-1}\Sigma_{t=2}^{T}\tilde{\varepsilon}_{it})^2$$

$$(4.9) \qquad \hat{\sigma}^2_u = \frac{1}{N-1}\Sigma_{i=1}^{N}(\frac{1}{T-1}\Sigma_{t=2}^{T}\tilde{\varepsilon}_{it} - \frac{1}{N(T-1)}\Sigma_{i=1}^{N}\Sigma_{t=2}^{T}\tilde{\varepsilon}_{it})^2 - \frac{\hat{\sigma}^2_w}{T}$$

where $\tilde{\varepsilon}_{it} = y_{it} - \hat{\rho}y_{it-1} - \hat{\alpha} - x_{it}\hat{\beta}$.

It follows that σ^2_* can be estimated as

$$(4.10) \qquad \hat{\sigma}^2_* = \hat{\sigma}^2_{\varepsilon_1} - \hat{\sigma}^2_u/(1-\hat{\rho})$$

The consistency of all these estimates can be easily shown and is omitted.

With these consistent estimates, simpler two steps maximum likelihood estimates (2SMLE) can be derived. Let $\hat{\theta}$ be the consistent estimates of θ as derived above. The 2SMLE is

$$\hat{\theta}_M = \hat{\theta} - [\frac{\partial^2 L(\hat{\theta})}{\partial\theta\partial\theta'}]^{-1} \frac{\partial L(\hat{\theta})}{\partial\theta}$$

As shown in Zacks [1971], $\hat{\theta}_M$ is consistent, asymptotically normal and asymptotically efficient:

$$\sqrt{N}(\hat{\theta}_M - \theta) \overset{D}{\rightarrow} N(0, \Sigma)$$

where $\Sigma = \text{plim}\, [-\frac{1}{N} \frac{\partial^2 L(\theta)}{\partial\theta\partial\theta'}]^{-1}$ which can be consistently estimated by $\hat{\Sigma} = -[\frac{1}{N} \frac{\partial^2 L(\hat{\theta})}{\partial\theta\partial\theta'}]^{-1}$.

Finally, if the vector of exogenous vectors x_{it}, for each i, can be represented as a finite order ARMA (p, q) process; such that, the number of distinct parameters, $p + q + 1$, is less than T, the π_t in (4.2) will be a function of those $p + q + 1$ basic parameters ω and are identifiable. In this case, one can improve the estimates by maximizing the joint likelihood function.

$$(4.11) \qquad L_*(\theta,\omega) = L(\theta) + \Sigma_{i=1}^{N} L(\omega | x_{i1},\ldots,x_{iT})$$

where $L(\theta)$ is in (4.4) and $L(\omega | x_{i1},\ldots,x_{iT})$ is the log likelihood of $x_i' = (x_{i1},\ldots,x_{iT})$.

5. The Likelihood Function of Multivariate Time Series Error Components Model

In the previous sections, we have considered only single equation models. In this section, we consider dynamic multiple equation error components models.

The model we will consider is specified as follows:

$$(5.1)\qquad \underset{m\times 1}{y_{it}} = \underset{m\times m}{\Phi}\ \underset{m\times 1}{y_{it-1}} + \underset{m\times k}{\beta}\ \underset{k\times 1}{x_i} + \underset{m\times 1}{u_i} + \underset{m\times 1}{w_{it}} \qquad \begin{array}{l} i=1,\dots,N \\ t=1,\dots,T \end{array} \quad (T \geq 4)$$

$$\underset{m\times 1}{w_{it}} = \underset{m\times 1}{a_{it}} - \underset{m\times m}{\theta}\ \underset{m\times 1}{a_{it-1}}$$

where u_i ~ i.i.d. $N(0, \Sigma_u)$, a_{it} ~ i.i.d. $N(0, \Sigma)$, u_i and a_{it} are indepedent and the x_i are time invariant exogenous variables. We assume further that the y_{it} are stationary for each unit i and the identification conditions for multiple ARMA processes given in Hannan [1969] are satisfied so as to eliminate redundant parameters in this model.

The multiple time series model in panel data has been considered in Anderson [1978]. Anderson's models do not have the error components structure; consequently heterogeneity across units are excluded. Anderson considered conditional ML estimation procedures. As we have pointed out in previous sections, for panel data with short time periods, conditional maximum likelihood estimates are inconsistent. To have efficient estimates, the exact likelihood function is needed. For single time series ARMA(p, q) processes the exact likelihood function has been derived in Newbold [1974] and Dent [1977]. Newbold's procedure generalizes the backward forecasting procedure of Box and Jenkins [1970] and has been extended to derive the exact likelihood function for multiple ARMA processes by Nicholls and Hall [1978].

To derive efficient estimates for model (5.1), we have to derive the exact likelihood function which takes into account the error components structure and the presence of exogenous variables. The derivation proceeds in two steps. First, we will derive the likelihood function of $y_i' = (y_{i1},\ldots,y_{iT})$ conditional on the u_i's as fixed parameters. The derivation parallels the procedure in Newbold, Nicholls and Hall. For each unit i, let $y_i = (y_{i1}',\ldots,y_{iT}')'$, $a_i = (a_{i1}',\ldots, a_{iT}')'$, $a_i^* = (a_{i0}', y_{i0}')'$,

$$\underset{Tm\times Tm}{H} = \begin{bmatrix} I_m & & & 0 \\ -\Phi & I_m & & \\ & \ddots & \ddots & \\ 0 & & -\Phi & I_m \end{bmatrix}, \quad \underset{Tm\times Tm}{F} = \begin{bmatrix} I_m & & & \\ -\theta & \ddots & & \\ & \ddots & \ddots & \\ & & -\theta & I_m \end{bmatrix}$$

and $\underset{Tm\times 2m}{G} = \begin{bmatrix} -\theta & \Phi \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}$. The equation (5.1) implies that

$$\text{(5.2)} \qquad Hy_i = I_t \otimes (Bx_i + u_i) + Fa_i + Ga_i^*$$

Since $y_{i0} = (I_m-\Phi)^{-1}(Bx_i+u_i)+(I_m-\Phi L)^{-1}w_{it}$, $\mu_i^* = E(a_i^*|u_i) =$

$$= \begin{bmatrix} 0 \\ (I_m-\Phi)^{-1}(Bx_i+u_i) \end{bmatrix}$$

Equation (5.2) implies

(5.2)' $\quad Hy_i = P(Bx_i + u_i) + Fa_i + G(a_i^* - \mu_i^*)$

$$\text{where} \quad \underset{Tm\times m}{P} = \begin{pmatrix} I_m+\Phi(I_m-\Phi)^{-1} \\ I_m \\ \vdots \\ I_m \end{pmatrix} = \begin{pmatrix} (I_m-\Phi)^{-1} \\ I_m \\ \vdots \\ I_m \end{pmatrix}$$

Let T and R be $2m\times 2m$, $m\times m$ nonsingular matrices such that

$$TVT' = I_{2m}, \qquad R\Sigma R' = I_m$$

where $V = E((a_i^* - \mu_i^*)(a_i^* - \mu_i^*)'|u_i)$ is the conditional variance of a_i^* .

It follows from (5.2)' that

$$\text{(5.3)} \qquad e_i = \begin{pmatrix} 0 \\ (I_T\otimes R)F^{-1}H \end{pmatrix} y_i - \begin{pmatrix} 0 \\ (I_T\otimes R)F^{-1}P \end{pmatrix}(Bx_i+u_i) + \begin{pmatrix} I_{2m} \\ -(I_T\otimes R)F^{-1}GT^{-1} \end{pmatrix} e_i^*$$

where $e_i = \begin{pmatrix} T(a_i^*-\mu_i^*) \\ (I_T\otimes R)a_i \end{pmatrix}$ and $e_i^* = T(a_i^*-\mu_i^*)$. Let $\Psi = \begin{pmatrix} 0 \\ (I_T\otimes R)F^{-1}H \end{pmatrix}$,

$$C = \begin{pmatrix} 0 \\ (I_T\otimes R)F^{-1}P \end{pmatrix} \quad \text{and} \quad \Lambda = \begin{pmatrix} I_{2m} \\ -(I_T\otimes R)F^{-1}GT^{-1} \end{pmatrix} .$$

Since $E(e_i e_i'|u_i) = I$ and the jacobian is $|(I_T\otimes R)F^{-1}H| = |R|^T = |\Sigma|^{-T/2}$ the joint likelihood function of $(e_i^{*\prime}, y_i')'$ is

$$L(y_i, e_i^*|u_i) = (2\pi)^{-\frac{1}{2}(T+2)m}|\Sigma|^{-\frac{T}{2}} \exp\{-\tfrac{1}{2}(\Psi y_i - C(Bx_i+u_i) + \Lambda e_i^*)' \cdot$$

$$(\Psi y_i - C(Bx_i + u_i) + \Lambda e_i^*)\}$$

Let $\hat{e}_i^* = -(\Lambda'\Lambda)^{-1}\Lambda'(\Psi y_i - CBx_i - Cu_i)$. It follows that

$$(\Psi y_i - CBx_i - Cu_i + \Lambda e_i^*)'(\Psi y_i - CBx_i - Cu_i + \Lambda e_i^*)$$

$$= (\Psi y_i - CBx_i - Cu_i + \Lambda \hat{e}_i^*)'(\Psi y_i - CBx_i - Cu_i + \Lambda \hat{e}_i^*) + (e_i^* - \hat{e}_i^*)'\Lambda'\Lambda(e_i^* - \hat{e}_i^*)$$

and

$$L(y_i, e_i^* | u_i) = (2\pi)^{-\frac{1}{2}Tm} |\Sigma|^{-\frac{T}{2}} |\Lambda'\Lambda|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\Psi y_i - CBx_i - Cu_i + \Lambda \hat{e}_i^*)' \cdot$$

$$(\Psi y_i - CBx_i - Cu_i + \Lambda \hat{e}_i^*)\} \cdot L(e_i^* | y_i)$$

where $L(e_i^*\ y_i) = (2\rho)^{-\frac{1}{2}2m} |\Lambda'\Lambda|^{\frac{1}{2}} \exp\{-\frac{1}{2}(e_i^* - \hat{e}_i^*)'\Lambda'\Lambda(e_i^* - \hat{e}_i^*)\}$. Hence the density function of y_i conditional on u_i is

$$L(y_i | u_i) = (2\rho)^{-\frac{1}{2}Tm} |\Sigma|^{-\frac{T}{2}} |\Lambda'\Lambda|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\Psi y_i - CBx_i - Cu_i)' \cdot$$
$$(I - \Lambda(\Lambda'\Lambda)^{-1}\Lambda')(\Psi y_i - CBx_i - Cu_i)\} \tag{5.4}$$

The likelihood function can be simplified slightly. Let

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = I - \Lambda(\Lambda'\Lambda)^{-1}\Lambda .$$

It follows

$$\{\psi y_i - CBx_i - Cu_i)'M(\Psi y_i - CBx_i - Cu_i) \tag{5.5}$$

$$= (Hy - PBx_i - Pu_i)'F'^{-1}(I_T \otimes R')M_{22}(I_T \otimes R)F^{-1}(Hy_i - PBx_i - Pu_i)$$

$$= (y_i - H^{-1}PBx_i - H^{-1}Pu_i)'H'F'^{-1}(I_T \otimes R')M_{22}(I_T \otimes R)F^{-1}H(y_i - H^{-1}PBx_i - H^{-1}Pu_i)$$

As $\Lambda'\Lambda = I_{2m} + T'^{-1}G'F'^{-1}(I_T \otimes \Sigma^{-1})F^{-1}GT^{-1}$

$$M_{22} = I_{Tm} - (I_T \otimes R)F^{-1}GT^{-1}(I + T'^{-1}G'F'^{-1}(I_T \otimes \Sigma^{-1})F^{-1}GT^{-1})^{-1}T'^{-1}G'F'^{-1}(I_T \otimes R')$$

$$= (I_T \otimes R)Q(I \otimes R')$$

where $Q = I_T \otimes \Sigma - F^{-1}G(V^{-1}+G'F'^{-1}(I_T \otimes \Sigma^{-1})F^{-1}G)^{-1}G'F'^{-1}$

With equation (4.5), we have

$$y_i = H^{-1}PBx_i + H^{-1}Pu_i + \xi_i \tag{5.6}$$

where ξ_i is independent with u_i, $\xi_i \sim N(0, \Sigma_\xi)$ and

$$\Sigma_\xi^{-1} = H'F'^{-1}(I_T \otimes \Sigma^{-1})Q(I_T \otimes \Sigma^{-1})F^{-1}H \tag{5.7}$$

Hence, it follows

$$E(y_i) = H^{-1}PBx_i \tag{5.8}$$

$$\begin{aligned} \Sigma_y &= \mathrm{var}(y_i) \\ &= H^{-1}P\Sigma_u P'H'^{-1} + \Sigma_\xi \end{aligned} \tag{5.9}$$

The inverse matrix and the determinant of Σ_y are

$$\begin{aligned} \Sigma_y^{-1} &= \Sigma_\xi^{-1} - \Sigma_\xi^{-1}H^{-1}P\{\Sigma_u^{-1}+P'H'^{-1}\Sigma_\xi^{-1}H^{-1}P\}^{-1}P'H^{-1'}\Sigma_\xi^{-1} \\ &= \Sigma_\xi^{-1} - H'F'^{-1}(I_T \otimes \Sigma^{-1})Q(I_T \otimes \Sigma^{-1})F^{-1}P \cdot \\ &\quad \{\Sigma_u^{-1} + P'F'^{-1}(I_T \otimes \Sigma^{-1})\, Q(I_T \otimes \Sigma^{-1})F^{-1}P\}^{-1} \cdot \\ &\quad P'F'^{-1}(I_T \otimes \Sigma^{-1})Q(I_T \otimes \Sigma^{-1})F^{-1}H \end{aligned}$$

$$\begin{aligned} |\Sigma_y| &= |\Sigma_\xi||\Sigma_u||\Sigma_u^{-1} + P'H^{-1'}\Sigma_\xi^{-1}H^{-1}P| \\ &= |\Sigma_\xi||\Sigma_u||\Sigma_u^{-1} + P'F'^{-1}(I_T \otimes \Sigma^{-1})Q(I_T \otimes \Sigma^{-1})F^{-1}P| \end{aligned}$$

The determinant of Σ_ξ can be recovered from (5.4) which is

$$|\Sigma_\xi| = |\Sigma|^T |\Lambda'\Lambda|$$

$$= |\Sigma|^T |I_{2m} + T'^{-1}G'F'^{-1}(I_T \otimes \Sigma^{-1})F^{-1}GT^{-1}|$$

$$= |\Sigma|^T |V^{-1} + G'F'^{-1}(I_T \otimes \Sigma^{-1})F^{-1}G| |V|$$

The exact likelihood function for the whole sample $y_1, \ldots, y_N$ is

$$(5.10) \qquad L(y) = (2\pi)^{-\frac{1}{2}TNm} |\Sigma_\xi|^{-\frac{N}{2}} |\Sigma_u|^{-\frac{N}{2}} |\Sigma_u^{-1} + P'F'^{-1}(I_T \otimes \Sigma^{-1})Q(I_T \otimes \Sigma^{-1})F^{-1}P|^{-\frac{N}{2}} \cdot$$

$$\cdot \exp\{-\frac{1}{2} \Sigma_{i=1}^N (y_i - H^{-1}PBx_i)'\Sigma_y^{-1}(y_i - H^{-1}PBx_i)\}$$

It should be noted that

$$H^{-1} = \begin{pmatrix} I_m & & & & \\ \Phi & I_m & & & \\ \vdots & \ddots & \ddots & & \\ \Phi^{T-1} & \cdots & & \Phi & I_m \end{pmatrix}, \qquad F^{-1} = \begin{pmatrix} I_m & & & \\ \theta & I_m & & \\ \vdots & \ddots & \ddots & \\ \theta^{T-1} & \cdots & \theta & I_m \end{pmatrix}$$

which implies

$$H^{-1}P = I_T \otimes (I_m - \Phi)^{-1}$$

Finally let us consider the variance matrix V of a_i^*. Let Ω_0 be the variance of y_{i0} conditional on u_i.

$$V = \begin{pmatrix} \Sigma & \Sigma \\ \Sigma & \Omega_0 \end{pmatrix} .$$

Ω_0 can be derived as follows. Let

$$v_{it} = (I_m - \Phi L)^{-1}(a_{it} - \Theta a_{it-1})$$

$$= \Phi v_{it-1} + a_{it} - \Theta a_{it-1}$$

It follows

$$\Omega_0 = \Phi\Omega_0\Phi' + \Sigma + \theta\Sigma\theta' - \theta\Sigma\Phi' - \Phi\Sigma\theta'$$

and hence

$$\begin{aligned} vec\Omega_0 &= (\Phi\otimes\Phi)vec\Omega_0 + vec(\Sigma + \theta\Phi\theta' - \theta\Sigma\Phi' - \Phi\Sigma\theta') \\ &= (I_m\otimes I_m - \Phi\otimes\Phi)^{-1}vec(\Sigma + \theta\Sigma\theta' - \theta\Sigma\Phi' - \Phi\Sigma\theta') \end{aligned}$$

which provides an expression for Ω_0 in terms of the basic parameters.

Efficient Estimation can then be derived by maximizing the log likelihood function in (5.10).

6. Summary

In this article, we have considered unconditional maximum likelihood estimates of dynamic error components models for panel data with large units of cross sections but short time periods. We have considered models with or without exogenous variables. For the univariate first order autoregressive error components time series model, unconditional maximum likelihood estimates can be derived from procedures which involve maximization of the concentrated likelihood function with a single parameter in a bounded interval. For models with time invariant exogenous variables, one needs to maximize the concentrated likelihood function with regard to only two parameters in a bounded interval. Alternative simpler procedures are also introduced. Asymptotic properties are derived for those models. For dynamic models with time variant exogenous variables, one has an initial values problem for the processes generating the exogenous variables. For the case that the exogenous variables' processes are stationary and normal, we were able to solve this problem and derived the exact likelihood function. Simple initial consistent estimates can be derived for this model. With these consistent estimates, two step maximum likelihood estimates which are computationally simple and asymptotically efficient were derived. Finally, we introduced a general multiple ARMA dynamic model with error components structure. The exact likelihood function was derived.

References

Anderson, T. W. (1978), "Repeated Measurements on Autoregressive Processes", Journal of the American Statistical Association, 73, p. 371- 78.

Balestra, P. and M. Nerlove (1967), "Pooling Cross Section and Time Series Data in the Estimation of Dynamic Model: The Demand for Natural Gas", Econometrica 34(3), p. 585-612.

Box, G.E.P. and G. M. Jenkins (1970), Time Series Analysis: Forecasting and Control, San Francisco: Holden Day.

Chamberlain, G. (1978), "Analysis of Covariance with Qualitative Data", manuscript, paper presented in the Econometric Society Summer meeting, Boulder, Colorado, June 1978.

Dent, W. T. (1977), "Computation of the Exact Likelihood Function for an ARIMA Process , J. Statist. Comp. & Simul., 5, p. 193-206.

Graybill, F. A. (1961), An Introduction to Linear Statistical Models, McGraw-Hill Book Company, Inc., New York.

Graybill, F. A. (1976), Theory and Application of the Linear Model, Duxberry Press, U.S.A.

Hannan, E. J. (1969), "The Identification of Vector Mixed Auto-Regressive Moving Average System", Biometrika 56, pp. 223-225.

Heckman, J. (1974), "Problems in Estimating a Discrete Time, Discrete Data Stochastic Process", forthcoming in Structural Analysis of Discrete Data, ed. C. Manski and D. McFadden, MIT Press, January 1980.

Jennrich, H. O. (1969), "Asymptotic Properties of Non-Linear Least Squares Estimators", Annals of Mathematical Statistics, 40 p. 633-643.

Maddala, G. S. (1971), "The Use of Variance Components Models in Pooling Cross Section and Time Series Data", Econometrica, 39(2), p. 341-358.

Magnus, J. R. (1978), "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix", Journal of Econometrics, 7, p. 281-312.

Nerlove, M. (1971), "Further Evidence on the Estimation of Dynamic Relations from a Time Series of Cross Sections", Econometrica, 39(2), p. 359-382.

Newbold, P. (1974), "The Exact Likelihood Function for a Mixed Autoregressive Moving Average Process", Biometrika, 61, p. 423-426.

Nicholls, D. F., and A. D. Hall (1978), "The Exact Maximum Likelihood Function of Multivariate Autoregressive Moving Average Models", Working Paper No. 70, Australian National University, 1978, Faculty of Economics and Research School of Social Sciences, forthcoming in Biometrika.

Pierce, D. A. (1971), "Least Squares Estimation in the Regression Model with Autoregressive-Moving Average Errors", Biometrica 58, p. 299-312.

Rao, C. R. (1973), Linear Statistical Inference and Its Applications second edition, John Wiley and Sons, New York.

Trognon, A. (1978), "Miscellaneous Asymptotic Properties of Ordinary Least Squares and Maximum Likelihood Estimators in Dynamic Error Components Models", Annales De L'insee - no. 30-31, p. 632-657.

Wallace, T. D. and A. Hussain (1969), "The Use of Error Component Models in Combining Cross-Section and Time Series Data", Econometrica, p. 55-72.

Zacks, S. (1971), The Theory of Statistical Inference, John Wiley & Sons, Inc., New York.

Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", Journal of the American Statistical Association, 57, p. 348-368.