COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut

COWLES FOUNDATION DISCUSSION PAPER NO. 271

A NOTE ON A VARIANCE-COMPONENTS MODEL USEFUL IN THE STUDY OF CROSS-SECTIONS OVER TIME

Marc Nerlove

April 9, 1969

A NOTE ON A VARIANCE-COMPONENTS MODEL USEFUL IN THE STUDY OF CROSS-SECTIONS OVER TIME

by

Marc Nerlove*

1. Introduction

The purpose of this note is to explore the properties of a variance-covariance matrix and certain estimates which arise in connection with a model which has been found useful in the study of data on a number of individuals observed over several periods of time. See, for example, [2], [3], and [4]. Let there be $i = 1, \ldots, N$ individuals observed over $t = 1, \ldots, T$ time periods, and suppose that there exists a stochastic relationship connecting the observations on some dependent variable, y_{it}, with a number of independent variables, $x_{it}^{(1)}, \ldots, x_{it}^{(K)}$, and certain unobserved random variables u_{it} :

$$(1) \quad y_{it} = x_{it}^{(1)}\beta_1 + \ldots + x_{it}^{(K)}\beta_K + u_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T.$$

If we array our observations first by individual and then by period, we may represent equations (1) by

*The research on which this note is based was supported by a grant from the National Science Foundation to the Cowles Foundation for Research in Economics at Yale University. I am indebted to David Grether and E.J. Hannan for extremely helpful discussions on the subject. They should not be held responsible for errors.

(2) $$y_i = X_i\beta + u_i \; , \quad i = 1, \ldots, N \; ,$$

where y_i is a $T\times 1$ vector, X_i is a $T\times K$ matrix, β is a $K\times 1$ vector, and u_i is a $T\times 1$ vector. We leave aside, temporarily, the question of whether there are any lagged values of y_{it} included among the columns of X_i . Letting

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}, \quad X = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix}, \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}$$

be a $TN\times 1$ vector, a $TN\times K$ matrix, and a $TN\times 1$ vector, respectively, we have, finally

(3) $$y = X\beta + u$$

which is the usual form of regression model. The special nature of the problem, however, is expressed in the assumptions we make about the variance-covariance matrix of the vector u , since we have assumed a homogeneous relation between dependent and independent variables over all time periods and all individuals.

Disturbances are generally supposed to represent the net effects of numerous individually unimportant but collectively significant variables which have been omitted from the analysis. Some of these, we may suppose, are specific to the individuals considered; some are specific to the time periods; and, some are specific to both. We assume that u_{it} may be decomposed into the sum of three independent normal variables, each with zero mean,

(4) $$u_{it} = \mu_i + \lambda_t + \nu_{it} ,$$

such that

(5)
$$E\mu_i\mu_{i'} = \begin{cases} \sigma_\mu^2 & i = i' \\ 0 & i \neq i' \end{cases}$$
$$E\lambda_t\lambda_{t'} = \begin{cases} \sigma_\lambda^2 & t = t' \\ 0 & t \neq t' \end{cases}$$
$$E\nu_{it}\nu_{i't'} = \begin{cases} \sigma_\nu^2 & i = i' \text{ and } t = t' \\ 0 & \text{otherwise .} \end{cases}$$

μ_i represents the individual effects, λ_t the period effects, and ν_{it}, the remainder. (For further justification see [4, p. 2-7].) If we let

(6)
$$\begin{cases} \sigma^2 = \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\nu^2 \\ \rho = \sigma_\mu^2/\sigma^2 \\ \omega = \sigma_\lambda^2/\sigma^2 \end{cases}$$

we may write the variance-covariance matrix associated with the vector u in simple form, which, however, makes use of the Kronecker product notation.

If A is a pxq matrix with typical element a_{ij} and B is an rxs matrix, the Kronecker product of A and B , written $A \otimes B$, is defined as the prxqs matrix

$$\begin{bmatrix} a_{11}B & \cdots & a_{1q}B \\ \vdots & & \vdots \\ q_{p1}B & \cdots & a_{pq}B \end{bmatrix} .$$

The following notation is also useful:

I_{NT} , I_T and I_N are identity matrices of orders NTxNT , TxT , and NxN , respectively.

e_{NT} , e_T , and e_N are vectors consisting entirely of ones of orders NTx1 , Tx1 , and Nx1 , respectively.

Using this notation we may write the variance-covariance matrix of u as

$$(7) \quad Euu' = \sigma^2\Omega = \sigma^2\{(1 - \rho - \varpi)I_{NT} + \rho(I_N \otimes e_T e_T') + \varpi(e_N e_N' \otimes I_T)\} .$$

Three problems are of interest: First, we need to obtain Ω^{-1} , or better still $\Omega^{-1/2}$, in order to generate the generalized least-squares estimate of β and/or the likelihood function. The generalized least-squares estimates are

$$(8) \qquad \hat{\beta} = [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}y$$

and may generally be found more simply computationally by first transforming X and y to

$$(9)\qquad \begin{cases} X^* = \Omega^{-1/2}X \\ y^* = \Omega^{-1/2}y \end{cases}$$

and then estimating β from the ordinary least-squares regression of y^* on X^*. The logarithmic likelihood function, assuming $\mu_i \sim n(0, \sigma^2_\mu)$, $\lambda_t \sim n(0, \sigma^2_\lambda)$, and $\nu_{it} \sim n(0, \sigma^2_\nu)$, independently and satisfying conditions (5), is

$$(10)\qquad L(\beta, \rho, \omega, \sigma^2 | y, X) =$$

$$- \frac{NT}{2} \log 2\pi + \frac{1}{2} \log|\Omega^{-1}| - \frac{1}{2}(y - X\beta)'\Omega^{-1}(y - X\beta) .$$

The generalized least-squares estimates, for known Ω, amount, of course, to minimizing the last term in this expression and, thus, to neglecting the second term. Such neglect does not generally matter provided Ω is actually known, but loss of efficiency may result if Ω is first estimated in certain cases.

A second problem involves the derivation of $\Omega^{1/2}$. If one wishes to explore the properties of various sorts of estimation procedures by Monte Carlo experiments, a particularly useful approach when X involves lagged values of y, it is necessary to generate pseudo-random variables distributed normally with mean zero and variance-covariance matrix Ω. Since this must be done many times and for a variety of different values of the parameters ρ, ω, and σ^2, it is important to achieve computational efficiency. The optimal approach from this standpoint is to generate NT pseudo random variables, w_{it}, which are $n(0, 1)$ and "independent," then transform them by

(11) $$u = \sigma\Omega^{1/2}w$$

to obtain the relevant u_{it}'s .

A third problem of considerable interest, but less immediate application, is the determination of the characteristic roots and vectors of Ω .

We shall examine these problems in reverse order, concluding with a discussion of the generalized least-squares estimates and their interpretation.

2. The Characteristic Roots and Vectors of Ω

The matrix Ω has three terms, one in I_{NT} , one in $I_N \otimes e_T e_T'$, and one in $e_N e_N' \otimes I_T$. It is therefore clear that any vector which annihilates both $I_N \otimes e_T e_T'$ and $e_N e_N' \otimes I_T$ must be a characteristic vector of Ω with associated root $1 - \rho - \omega$. Let φ_j , $j = 1, \ldots, N-1$, be $N-1$ vectors, each $N\text{x}1$, which are orthonormal and orthogonal to e_N :

(1) $$e_N'\varphi_j = 0 ,$$
$$\varphi_j'\varphi_{j'} = \begin{cases} 1 & j = j' \\ 0 & j \neq j' , \quad j = 1, \ldots, N-1 . \end{cases}$$

Let ψ_k , $k = 1, \ldots, T-1$, be $T-1$ vectors each $T\text{x}1$, which are orthonormal and each orthogonal to e_T :

(2) $$e_T'\psi_k = 0$$
$$\psi_k'\psi_{k'} = \begin{cases} 1 & k = k' \\ 0 & k \neq k' , \quad k = 1, \ldots, T-1 . \end{cases}$$

Note that we can always find such sets of vectors. The $(N-1)(T-1)$ vectors $\varphi_j \otimes e_T e_T'$ are orthonormal and annihilate both $I_N \otimes e_T e_T'$ and $e_N e_N' \otimes I_T$ as may readily be verified. They are thus characteristic vectors of Ω associated with the characteristic root $1 - \rho - \omega$ which is of multiplicity $(N-1)(T-1)$.

Consider the Kronecker products $\varphi_j \otimes \xi_T$ and $\xi_N \otimes \psi_k$ where ξ_T and ξ_N are, respectively, any T and N dimensional vectors. Clearly, the vector $\varphi_j \otimes \xi_T$ annihilates the term containing $e_N e_N' \otimes I_T$ since the vector φ_j are each orthogonal to the vector e_N . Similarly the $\xi_N \otimes \psi_k$ annihilates the term containing $I_N \otimes e_T e_T'$. It might thus be supposed that $N-1$ additional characteristic vectors of the form $\varphi_j \otimes \xi_T$ and $T-1$ of the form $\xi_N \otimes \psi_k$ may be found. To determine the form of ξ_T and ξ_N more specifically, observe that ξ_T and ψ_k must be orthogonal for $k = 1, \ldots, T-1$, and ξ_N and φ_j must also be for $j = 1, \ldots, N-1$, since the vectors $\varphi_j \otimes \psi_k$, $\varphi_{j'} \otimes \xi_T$ and $\xi_N \otimes \psi_{k'}$ must be orthogonal for any j , k , j' and k' . Thus, for example,

$$(\varphi_j \otimes \psi_k)'(\varphi_{j'} \otimes \xi_T) = \begin{cases} 0 & j \neq j' \\ \psi_k'\xi_T & j = j' , \end{cases}$$

which is necessarily zero only if $\psi_k'\xi_T = 0$. Similarly,

$$(\varphi_j \otimes \psi_k)'(\xi_N \otimes \psi_{k'}) = \begin{cases} 0 & k \neq k' \\ \varphi_j'\xi_N & k = k' , \end{cases}$$

which implies $\varphi_j'\xi_N$ must be zero for $j = 1, \ldots, N-1$. These conditions in turn imply that the vectors $\varphi_j \otimes \xi_T$ and $\xi_N \otimes \psi_k$ are all orthogonal. If they are to have unit length as well, $\xi_N'\xi_N = 1$ and $\xi_T'\xi_T = 1$. Now, by definition the only vector orthogonal to, and independent of ψ_k , $k = 1, \ldots, T-1$, is a scalar multiple of e_T ; hence, the condition $\xi_T'\xi_T = 1$ implies

$$\xi_T = e_T/\sqrt{T} .$$

Similarly,

$$\xi_N = e_N/\sqrt{N} .$$

It follows then that the $N-1$ vectors $\varphi_j \otimes e_T/\sqrt{T}$, $j = 1, \ldots, N-1$, are also characteristic vectors of Ω , and it may be readily verified that the associated characteristic root of multiplicity $N-1$ is $1 - \rho - \varpi + T\rho$. In the same way, it may be seen that the $T-1$ vectors $e_N/\sqrt{N} \otimes \psi_k$, $k = 1, \ldots, T-1$, are characteristic vectors of Ω associated with the root $1 - \rho - \omega + \omega N$ which is of multiplicity $T-1$.

The matrix Ω has all together NT associated characteristic vectors; we have obtained $(N-1)(T-1) + (N-1) + (T-1) = NT-1$, and only one further vector and root remain to be found. Consider the Kronecker product of the two vectors ξ_T and ξ_N determined above: $e_N/\sqrt{N} \otimes e_T/\sqrt{T}$. This vector is clearly orthogonal to all of the characteristic vectors of Ω previously determined, since the latter all contain either φ_j or ψ_k or both, one of which is orthogonal to one of the two terms in the Kronecker product $e_N/\sqrt{N} \otimes e_T/\sqrt{T}$. It follows from the fact that the characteristic vectors span the space of column vectors of Ω that the remaining characteristic vectors must be simply a scalar multiple of $e_N/\sqrt{N} \otimes e_T/\sqrt{T}$. Indeed, it is just that vector, and the associated characteristic root is $1 - \rho - \omega + \rho T + \omega N$ as may readily be verified by expanding $\Omega(e_N/\sqrt{N} \otimes e_T/\sqrt{T})$.

We have proved the following theorem, originally suggested by E.J. Hannan:

<u>Theorem</u>: Let φ_j, $j = 1, \ldots, N-1$, ψ_k, $k = 1, \ldots, T-1$, e_N and e_T be the vectors defined in (1) and (2), and let C be the $NT \times NT$ matrix

$$(3) \qquad C = \begin{bmatrix} e_N'/\sqrt{N} \otimes e_T'/\sqrt{T} \\ e_N'/\sqrt{N} \otimes \psi_1' \\ \vdots \\ e_N'/\sqrt{N} \otimes \psi_{T-1}' \\ \varphi_1' \otimes e_T'/\sqrt{T} \\ \vdots \\ \varphi_{N-1}' \otimes e_T'/\sqrt{T} \\ \varphi_1' \otimes \psi_1' \\ \varphi_1' \otimes \psi_2' \\ \vdots \\ \varphi_{N-1}' \otimes \psi_{T-1}' \end{bmatrix}$$

Then $C'C = I_{NT}$ and

$$C\Omega C' = C\{(1-\rho-\omega)I_{NT} + \rho(I_N \otimes e_T e_T') + \omega(e_N e_N' \otimes I_T)\}$$

$$= \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 I_{T-1} & 0 & 0 \\ 0 & 0 & \lambda_3 I_{N-1} & 0 \\ 0 & 0 & 0 & \lambda_4 I_{(N-1)(T-1)} \end{bmatrix} = D_\lambda$$

where

$$\lambda_1 = 1 - \rho - \omega + \omega N + \rho T$$

$$\lambda_2 = 1 - \rho - \omega + \omega N$$

$$\lambda_3 = 1 - \rho - \omega + \rho T$$

$$\lambda_4 = 1 - \rho - \omega$$

are the four distinct characteristic roots of Ω.

3. The Square Root of Ω

The square root of Ω, $\Omega^{1/2}$, may be defined simply in terms of orthogonal matrix C and the matrix D_λ, defined in the Theorem of the previous section:

(1) $$\Omega^{1/2} = C'D_{\sqrt{\lambda}}C ,$$

where

$$D_{\sqrt{\lambda}} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 & 0 \\ 0 & \sqrt{\lambda_2} I_{T-1} & 0 & 0 \\ 0 & 0 & \sqrt{\lambda_3} I_{N-1} & 0 \\ 0 & 0 & 0 & \sqrt{\lambda_4} I_{(N-1)(T-1)} \end{bmatrix} .$$

Let us partition C in (2.3) above as

(2) $$C = \begin{bmatrix} C_1' \\ C_2' \\ C_3' \\ C_4' \end{bmatrix} ,$$

where

(3)
$$\begin{cases} C_1 = e_N/\sqrt{N} \otimes e_T/\sqrt{T} = \frac{1}{\sqrt{NT}} e_{NT} , & \text{an NTx1 vector,} \\ C_2 = \frac{1}{\sqrt{N}} [e_N \otimes \psi_1, \ldots, e_N \otimes \psi_{T-1}] , & \text{an NTxT-1 matrix,} \\ C_3 = \frac{1}{\sqrt{T}} [\varphi_1 \otimes e_T, \ldots, \varphi_{N-1} \otimes e_T] , & \text{an NTxN-1 matrix,} \\ C_4 = [\varphi_1 \otimes \psi_1, \varphi_1 \otimes \psi_2, \ldots, \varphi_{N-1} \otimes \psi_{T-1}] , & \text{an NTx(T-1)(N-1) matrix .} \end{cases}$$

Thus

(4) $$\Omega^{1/2} = C' D_{\sqrt{\lambda}} C$$

$$= [C_1 C_2 C_3 C_4] \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 & 0 \\ 0 & \sqrt{\lambda_2} I_{T-1} & 0 & 0 \\ 0 & 0 & \sqrt{\lambda_3} I_{N-1} & 0 \\ 0 & 0 & 0 & \sqrt{\lambda_4} I_{(N-1)(T-1)} \end{bmatrix} \begin{bmatrix} C_1' \\ C_2' \\ C_3' \\ C_4' \end{bmatrix}$$

$$= \sqrt{\lambda_1} C_1 C_1' + \sqrt{\lambda_2} C_2 C_2' + \sqrt{\lambda_3} C_3 C_3' + \sqrt{\lambda_4} C_4 C_4' .$$

Clearly

(5) $$C_1 C_1' = \frac{1}{NT} e_{NT} e_{NT}' .$$

Consider the product

(6) $$C'C = C_1 C_1' + C_2 C_2' + C_3 C_3' + C_4 C_4' = I_{NT}$$

Premultiply (6) by $e_N e_N' \otimes I_T$. Since

$$(e_N e_N' \otimes I_T) C_1 C_1' = N C_1 C_1' = \frac{e_{NT} e_{NT}'}{T} ,$$

$$(e_N e_N' \otimes I_T) C_2 C_2' = N C_2 C_2' ,$$

$$(e_N e_N' \otimes I_T) C_3 C_3' = [(e_N' \varphi_1) e_T / \sqrt{T}, \ \ldots, \ (e_N' \varphi_{N-1}) e_T / \sqrt{T}] C_3'$$

$$= 0 \quad (\text{because } e_N' \varphi_k = 0 , \quad \text{for } k = 1, \ldots, N-1) ,$$

and

$$(e_N e_N' \otimes I_T) C_4 C_4' = [(e_N' \varphi_1) \psi_1, \ \ldots, \ (e_N' \varphi_{N-1}) \psi_{T-1}]$$

$$= 0 \quad (\text{for the same reason}),$$

it follows that

(7) $$C_2 C_2' = \frac{e_N e_N'}{N} \otimes I_T - \frac{e_{NT} e_{NT}'}{NT} .$$

In the same manner, premultiplication of (6) by $I_N \otimes e_T e_T'$ shows that

$$(8) \qquad C_3C_3' = I_N \otimes \frac{e_Te_T'}{T} - \frac{e_{NT}e_{NT}'}{NT} .$$

Combining (5), (6), (7), and (8), we obtain

$$(9) \qquad C_4C_4' = I_{NT} - \frac{e_Ne_N'}{N} \otimes I_T - I_N \otimes \frac{e_Te_T'}{T} + \frac{e_{NT}e_{NT}'}{NT} .$$

It follows that an explicit representation of $\Omega^{1/2}$ is

$$(10) \qquad \Omega^{1/2} = \sqrt{\lambda_1}\,\frac{e_{NT}e_{NT}'}{NT} + \sqrt{\lambda_2}\left\{\frac{e_Ne_N'}{N} \otimes I_T - \frac{e_{NT}e_{NT}'}{NT}\right\} + \sqrt{\lambda_3}\left\{I_N \otimes \frac{e_Te_T'}{T} - \frac{e_{NT}e_{NT}'}{NT}\right\} + \sqrt{\lambda_4}\left\{I_{NT} - \frac{e_Ne_N'}{N} \otimes I_T - I_N \otimes \frac{e_Te_T'}{T} + \frac{e_{NT}e_{NT}'}{NT}\right\} .$$

Let w_{it} , $i = 1, \ldots, N$, $t = 1, \ldots, T$, be NT independent, zero-mean, unit-variance, normal variables. Arrange these variables into an NTx1 vector $w' = (w_{11}, \ldots, w_{1T}, \ldots, w_{N1}, \ldots, w_{NT})$. The vector

$$(11) \quad u = \sigma\Omega^{1/2}w = \sigma\sqrt{\lambda_1}\,\bar{w}_{..}e_{NT} + \sigma\sqrt{\lambda_2}\left\{e_N \otimes \begin{pmatrix}\bar{w}_{.1}\\ \vdots \\ \bar{w}_{.T}\end{pmatrix} - \bar{w}_{..}e_{NT}\right\} + \sigma\sqrt{\lambda_3}\left\{\begin{pmatrix}\bar{w}_{1.}\\ \vdots \\ \bar{w}_{N.}\end{pmatrix} \otimes e_T - \bar{w}_{..}e_{NT}\right\} + \sigma\sqrt{\lambda_4}\left\{w - e_N \times \begin{pmatrix}\bar{w}_{.1}\\ \vdots \\ \bar{w}_{.T}\end{pmatrix} - \begin{pmatrix}\bar{w}_{1.}\\ \vdots \\ \bar{w}_{N.}\end{pmatrix} \otimes e_T + \bar{w}_{..}e_{NT}\right\}$$

is distributed according to a multivariate normal distribution with mean zero and variance-covariance matrix $\sigma^2\Omega$. The barred expressions in (11) are defined as follows:

$$
(12)\quad \begin{cases}
\bar{w}_{..} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} w_{it} , & \text{over-all mean;} \\
\bar{w}_{.t} = \frac{1}{N} \sum_{i=1}^{N} w_{it} , \quad t = 1, \ldots, T , & \text{period means; and} \\
\bar{w}_{i.} = \frac{1}{T} \sum_{t=1}^{T} w_{it} , \quad i = 1, \ldots, N , & \text{individual means.}
\end{cases}
$$

Thus, for a Monte Carlo experiment, we generate the numbers u_{it}, $i = 1, \ldots, N$, $t = 1, \ldots, T$, from NT "independent" pseudo-random numbers $w_{it} \sim n(0, 1)$ according to the formula:

$$
(13)\quad u_{it} = \sigma\left(\frac{\sqrt{\lambda_1} - \sqrt{\lambda_2} - \sqrt{\lambda_3} + \sqrt{\lambda_4}}{NT}\right) \sum_{i=1}^{N} \sum_{t=1}^{T} w_{it} + \sigma\left(\frac{\sqrt{\lambda_2} - \sqrt{\lambda_4}}{N}\right) \sum_{i=1}^{N} w_{it} + \sigma\left(\frac{\sqrt{\lambda_3} - \sqrt{\lambda_4}}{T}\right) \sum_{t=1}^{T} w_{it} + \sigma\sqrt{\lambda_4}\, w_{it}
$$

4. The Inverse of Ω and the GLS Estimates of β

Knowledge of the characteristic roots and vectors of Ω, also permits an easy determination of Ω^{-1}. Since

$$(1) \qquad \Omega = C'D_{\lambda}C ,$$

$$(2) \qquad \Omega^{-1} = C'D_{1/\lambda}C ,$$

where

$$D_{1/\lambda} = \begin{bmatrix} 1/\lambda_1 & 0 & 0 & 0 \\ 0 & 1/\lambda_2 I_{T-1} & 0 & 0 \\ 0 & 0 & 1/\lambda_3 I_{N-1} & 0 \\ 0 & 0 & 0 & 1/\lambda_4 I_{(N-1)(T-1)} \end{bmatrix} ,$$

since $C^{-1} = C'$. By exactly the same reasoning as that used to deduce $\Omega^{1/2}$ in the previous section, we obtain

$$(3) \qquad \Omega^{-1} = \frac{1}{\lambda_1} C_1C_1' + \frac{1}{\lambda_2} C_2C_2' + \frac{1}{\lambda_3} C_3C_3' + \frac{1}{\lambda_4} C_4C_4'$$

$$= \frac{1}{\lambda_1} \frac{e_{NT}e_{NT}'}{NT} + \frac{1}{\lambda_2} \left\{ \frac{e_N e_N'}{N} \otimes I_T - \frac{e_{NT}e_{NT}'}{NT} \right\} + \frac{1}{\lambda_3} \left\{ I_N \otimes \frac{e_T e_T'}{T} - \frac{e_{NT}e_{NT}'}{NT} \right\}$$

$$+ \frac{1}{\lambda_4} \left\{ I_{NT} - \frac{e_N e_N'}{N} \otimes I_T - I_N \otimes \frac{e_T e_T'}{T} + \frac{e_{NT}e_{NT}'}{NT} \right\} .$$

Using the vector y and the matrix X defined in section 1, the generalized least-squares estimates of β are found by solving the normal equations

$$(4) \qquad X'\Omega^{-1}X\hat{\beta} = X'\Omega^{-1}y$$

or

$$(5)\quad \left\{ \frac{1}{\lambda_1} \frac{X' e_{NT} e'_{NT} X}{NT} + \frac{1}{\lambda_2} \left[X' \left(\frac{e_N e'_N}{N} \times I_T \right) X - \frac{X' e_{NT} e'_{NT} X}{NT} \right] \right.$$

$$+ \frac{1}{\lambda_3} \left[X' \left(I_N \otimes \frac{e_T e'_T}{T} \right) X - \frac{X' e_{NT} e'_{NT} X}{NT} \right]$$

$$\left. + \frac{1}{\lambda_4} \left[X'X - X' \left(\frac{e_N e'_N}{N} \otimes I_T \right) X - X' \left(I_N \otimes \frac{e_T e'_T}{T} \right) X + \frac{X' e_{NT} e'_{NT} X}{NT} \right] \right\} \hat{\beta}$$

$$= \left\{ \frac{1}{\lambda_1} \frac{X' e_{NT} e'_{NT} y}{NT} + \frac{1}{\lambda_2} \left[X' \left(\frac{e_N e'_N}{N} \otimes I_T \right) y - \frac{X' e_{NT} e'_{NT} y}{NT} \right] \right.$$

$$+ \frac{1}{\lambda_3} \left[X' \left(I_N \otimes \frac{e_T e'_T}{T} \right) y - \frac{X e_{NT} e'_{NT} y}{NT} \right]$$

$$\left. + \frac{1}{\lambda_4} \left[X'y - X' \left(\frac{e_N e'_N}{N} \otimes I_T \right) y - X' \left(I_N \otimes \frac{e_T e'_T}{T} \right) y + \frac{X' e_{NT} e'_{NT} y}{NT} \right] \right\}$$

Suppose that the relation (1.1) included only two independent variables $x_{it}^{(1)} = 1$, for all i and t, and $x_{it}^{(2)} = x_{it}$. β_1 may then be interpreted as the constant term; β_2 is the slope of the relation between y_{it} and x_{it}. If $\bar{x}_{..}$ and $\bar{y}_{..}$ denote the over-all means of x_{it} and y_{it}, respectively, $\bar{x}_{i.}$ and $\bar{y}_{i.}$ the individual means, and $\bar{x}_{.t}$ and $\bar{y}_{.t}$ the period means, the normal equations reduce to particularly suggestive form:

$$(6)\quad \left\{ \frac{NT}{\lambda_1} \bar{x}_{..}^2 + \frac{N}{\lambda_2} \sum_{t=1}^{T} (\bar{x}_{.t} - \bar{x}_{..})^2 + \frac{T}{\lambda_3} \sum_{i=1}^{N} (\bar{x}_{i.} - \bar{x}_{..})^2 \right.$$

$$\left. + \frac{1}{\lambda_4} \sum_{t=1}^{T} \sum_{i=1}^{N} (x_{it} - \bar{x}_{.t} - \bar{x}_{i.} + \bar{x}_{..})^2 \right\} \hat{\beta}_2$$

$$= \left\{ \frac{NT}{\lambda_1} \bar{x}_{..}\bar{y}_{..} + \frac{N}{\lambda_2} \sum_{t=1}^{T} (\bar{x}_{.t} - \bar{x}_{..})(\bar{y}_{.t} - \bar{y}_{..}) \right.$$

$$+ \frac{T}{\lambda_3} \sum_{i=1}^{N} (\bar{x}_{i.} - \bar{x}_{..})(\bar{y}_{i.} - \bar{y}_{..}) +$$

$$\left. \frac{1}{\lambda_4} \sum_{t=1}^{T} \sum_{i=1}^{N} (x_{it} - \bar{x}_{.t} - \bar{x}_{i.} + \bar{x}_{..})(y_{it} - \bar{y}_{.t} - \bar{y}_{i.} + \bar{y}_{..}) \right\}$$

and

(7) $$\hat{\beta}_1 + \bar{x}_{..}\hat{\beta}_2 = \bar{y}_{..} \ .$$

5. An Interpretation of $\hat{\beta}$ and Its Asymptotic Distribution*

The form of the normal equations in the simple case of one independent variable and a constant term, exhibited in the previous section, suggest an interesting interpretation of the generalized least-squares estimates. Let us write the relation to be estimated as

(1) $$y_{it} = \beta_1 + \beta_2 x_{it} + \mu_i + \lambda_t + \nu_{it} \ .$$

Taking deviations from the over-all mean annihilates β_1 :

(2) $$y_{it} - \bar{y}_{..} = \beta_2(x_{it} - \bar{x}_{..}) + \mu_i + \lambda_t + \nu_{it} \ , \quad i=1, \ldots, N \ , \quad t=1, \ldots, T \ ,$$

where we set $\frac{1}{N} \sum_{i=1}^{N} \mu_i = 0$, $\frac{1}{T} \sum_{t=1}^{T} \lambda_t = 0$, and $\frac{1}{NT} \sum_{i,t} \nu_{it} = 0$, their expected values. The ordinary least-squares estimates are given by solving

*This section was inspired by a similar treatment of the two-component model contained in an unpublished paper by G.S. Maddala.

the equation

$$(3) \qquad \sum_{i,t} (y_{it} - \bar{y}_{..})(x_{it} - \bar{x}_{..}) = b_2^{OLS} \Sigma (x_{it} - \bar{x}_{..})^2 ,$$

but, even in the case in which x_{it} is non-stochastic, this estimate is inefficient, although it is, in this case, unbiased.

To see how the inefficiency results, take means in equation (2) over t and set $\frac{1}{T} \sum_t \lambda_t = 0$ and $\frac{1}{T} \sum_t \nu_{it} = 0$, their expected values:

$$(4) \qquad \bar{y}_{i.} - \bar{y}_{..} = b_2(\bar{x}_{i.} - \bar{x}_{..}) + \mu_i , \quad i = 1, \ldots, N .$$

Equations (4) suggest still another estimate of β_2 may be obtained by solving the normal equation

$$(5) \qquad \sum_{i=1}^{N} (\bar{y}_{i.} - \bar{y}_{..})(\bar{x}_{i.} - \bar{x}_{..}) = b_2^{(i)} \sum_{i=1}^{N} (\bar{x}_{it} - \bar{x}_{..})^2 .$$

This "regression" makes use of the variation in the sample across individuals but not over time. Similarly, take means in equations (2) over individuals and set $\frac{1}{N} \sum_i \mu_i = 0$ and $\frac{1}{N} \sum_i \nu_{it} = 0$, their expected values:

$$(6) \qquad \bar{y}_{.t} - \bar{y}_{..} = b_2(\bar{x}_{.t} - \bar{x}_{..}) + \lambda_t , \quad t = 1, \ldots, T .$$

Equations (6) suggest an additional estimate of β_2 may be obtained by solving the normal equations

$$(7) \qquad \sum_{t=1}^{T} (\bar{y}_{.t} - \bar{y}_{..})(\bar{x}_{.t} - \bar{x}_{..}) = b_2^{(t)} \sum_{t=1}^{T} (\bar{x}_{.t} - \bar{x}_{..})^2 .$$

This "regression" makes use of the variation in the sample over time periods but not across individuals. Finally, consider estimating both μ_i and λ_t as parameters in the relation (2). Doing so amounts to running a multiple regression involving both individual and period dummies. Alternatively, think of estimating μ_i from (4) by replacing b_2 by an estimate of β_2, and λ_t from (6) in the same way. Insert the resulting estimates of μ_i and λ_t in (2) and obtain

$$(8) \qquad (y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{y}_{..}) = b_2(x_{it} - \bar{x}_{i.} - \bar{x}_{.t} + \bar{x}_{..}) + \nu_{it} ,$$

$$i = 1, \ldots, N , \quad t = 1, \ldots, T .$$

Equations (8) suggest estimating β_2 from the normal equation

$$(9) \qquad \sum_{i,t} (y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{y}_{..})(x_{it} - \bar{x}_{i.} - \bar{x}_{.t} + \bar{x}_{..})$$

$$= b_2^{LSC} \sum_{i,t} (x_{it} - \bar{x}_{i.} - \bar{x}_{.t} + \bar{x}_{..})^2 .$$

This estimate is denoted by superscribing LSC since it corresponds to the estimate obtained by treating μ_i and λ_t as constant parameters and estimating them as well as the slope and over all intercept of (1) by ordinary least squares. This "regression" makes use of the variation of the sample observations <u>about</u> their individual and period means, but not of the variation <u>among</u> individual means or period means. In addition, note that an estimate of β_2 , of sorts, may be obtained from the ratio of the overall means of y_{it} and x_{it} .

Now consider the normal equation (4.6) of the preceeding section. If we set $\rho = \omega = 0$, we find $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 1$ and equation (5.6) yields the ordinary least-squares estimate of β_2 , b_2^{OLS} . In effect, the ordinary least-squares estimate weights the various bits of information contained in the sample about β_2 strictly in proportion to the number of observations used in constructing them; thus, the overall mean is weighted by NT , the sum of squared deviations of individual means about the overall mean by N and so on. Common sense, however, would dictate weighting each bit in inverse proportion to the amount of misinformation or error about β_2 likely to be contained in that bit. This is, of course, just what the generalized least-squares estimates accomplish by weighting inversely to the characteristic roots λ_1 , λ_2 , λ_3 and λ_4 . In a sense, this is the underlying rationale behind the greater efficiency of the generalized least-squares estimate of β_2 in comparison with the ordinary least-squares estimate.

By the same token, least-squares regressions introducing dummies for both period and individual are also inefficient, for such estimates give all the weight to the last term in the expressions occurring within both sets of curly brackets in (4.6). It is interesting, however, to note that when N and T are very large, as long as ρ and ω are not zero, this term does indeed dominate and the estimates $\hat{\beta}_2$ and b_2^{LSC} will be very close. The reason is easy to see from the expressions relating the characteristic roots of Ω to the parameters ρ and ω :

$$\lambda_1 = 1 - \rho - \omega + \omega N + \rho T ,$$

$$\lambda_2 = 1 - \rho - \omega + \omega N ,$$

$$\lambda_3 = 1 - \rho - \omega + \rho T ,$$

$$\lambda_4 = 1 - \rho - \omega .$$

Only λ_4 does not depend on N or T , as long as $\omega \neq 0 \neq \rho$, and thus $1/\lambda_1$, $1/\lambda_2$, and $1/\lambda_3$ all tend to zero as both N and T increase, <u>irrespective of the manner in which they increase and irrespective of the value of ρ or of ω as long as neither is zero</u>. Indeed, this is the essence of the theorem proved by Amemiya [1] for a slightly simpler case. His statement, however, contains a minor error, since the values of ρ and ω are irrelevant only if neither are zero.

A rigorous statement of this result requires also that something be said concerning the values of the moments appearing in (4.6), or more generally in (4.5). It is sufficient if $\frac{1}{NT}$ times each of the three terms on the left tends to a finite positive-definite matrix.

It remains something of a puzzle as to why treating μ_i and λ_t as parameters, rather than as random variables, should become asymptotically unimportant. After all as N, $T \to \infty$ there are an infinite number of such parameters; their numbers increase just as fast as the number of pieces of new information available as the sample size increases. The solution to the puzzle is simply that we are not, in fact, estimating them but only β when we solve the normal equations

$$\text{(10)} \quad \left\{ X'X - X'\left(\frac{e_N e_N'}{N} \otimes I_T\right)X - X'\left(I_N \otimes \frac{e_T e_T'}{T}\right)X + \frac{X' e_{NT} e_{NT}' X}{NT} \right\} b^{LSC} = \left\{ X'y - X\left(\frac{e_N e_N'}{N} \otimes I_T\right)y - X'\left(I_N \otimes \frac{e_T e_T'}{T}\right)y + \frac{X' e_{NT} e_{NT}' y}{NT} \right\} ,$$

which corresponds to the <u>relevant portion</u> of the normal equations obtained by treating μ_i and λ_t as constants to be estimated. Indeed, it is clear by examining the likelihood function (1.10), that ρ and ω are not identified asymptotically, although they are in finite samples. While the maximum-likelihood estimate of β does exist asymptotically, and does correspond asymptotically to b^{LSC}, maximum-likelihood estimates of ρ and ω do not exist asymptotically.

REFERENCES

[1] Amemiya, T., "A Note on the Estimation of Balestra-Nerlove Models," Tech. Rept. 4 (NSF-GS-1440), Inst. for Math. Stud. in Soc. Sci., Stanford University (August 14, 1967).

[2] Balestra, P., and M. Nerlove, "Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," Tech. Rept. 8 (NSF-GS-142), Inst. for Math. Stud. in Soc. Sci., Stanford University (December 21, 1964); published in revised form in Econometrica, 34: 585-612 (July 1966).

[3] Nerlove, M., "Experimental Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections," Economic Studies Quarterly, 18: 42-74 (December 1967).

[4] Nerlove, M., "Further Evidence on the Estimation of Dyanmic Economic Relations from a Time Series of Cross-Sections," Cowles Foundation Discussion Paper No. 257 (November 6, 1968).