NBER WORKING PAPER SERIES

MULTIPLE TIME SERIES MODELS APPLIED TO PANEL DATA

Thomas E. MaCurdy

Working Paper No. 646

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

March 1981

This research was supported by NSF grant SOC 77-27136. Portions of this paper were presented at the Econometric Society Meetings held in Montreal, Canada, June, 1979. I am grateful for comments from Bronwyn Hall and Tom Mroz. The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #646
March 1981

Multiple Time Series Models Applied to Panel Data

ABSTRACT

This study presents a general methodology for fitting multiple time series models to panel data. The basic statistical framework considered here consists of a dynamic simultaneous equation model where disturbances follow a permanent-transitory scheme with transitory components generated by a multivariate autoregressive-moving average process. This error scheme admits a wide variety of autocovariance patterns and provides a flexible framework for describing the dynamic characteristics of longitudinal data with a minimal number of parameters. It is possible within this framework to consider generally specified rational distributed lag structures involving both exogenous and endogenous variables which includes infinite order lag relationships. This paper outlines the generalizations of standard time series models that are possible when using panel data, and it identifies those instances in which procedures found in the time series literature cannot be directly applied to analyze longitudinal data. Data analysis techniques in the time series literature are adapted for panel data analysis. These techniques aid in the choice of a time series model and prevent one from choosing a specification that is broadly inconsistent with the data. Several estimation procedures are proposed that can be used to estimate all the parameters of a multiple time series model including both regression coefficients and parameters of the covariance matrix. The techniques developed here are robust in the sense that they do not rely on any specific distributional assumptions for their asymptotic properties, and in many cases their implementation requires only standard computer packages.

Professor Thomas E. MaCurdy
Department of Economics
Stanford University
Stanford, California 94305

(415) 497-9694

Introduction

This paper applies the apparatus of stationary time series analysis to the analysis of panel data. Multiple time series models, such as those studied by Zellner and Palm (1974, 1975) and Wallis (1977), are combined with the components of variance models of Balestra and Nerlove (1966) and Hussain and Wallace (1969). Grafting these two distinct models together offers a natural framework for pooling cross section and time series data.

The statistical model considered in this paper is based on what is known in the time series literature as a dynamic simultaneous equation model (DSEM). This model merges multiple time series analysis with the analysis of simultaneous equations. It consists of a system of structural equations where endogenous variables are related not only to one another and to exogenous variables, but also to lagged values of these variables. Generalized variants of this model offer a rich statistical framework for the analysis of panel data. Virtually any empirical specification that is linear in measured variables can be analyzed within this framework, including generally specified distributed lag relationships which may be of infinite order involving either endogenous or exogenous variables.

To model the correlation properties of disturbances over time, this study recognizes a broader class of error processes than has been considered in existing work on panel data. Disturbances are assumed to consist of permanent components and time varying components that follow generally specified auto-regressive-moving averages (ARMA) processes. Previous studies typically impose specific autocorrelation schemes on the data, and they do not perform systematic tests among competing specifications. This study, on the other hand, approaches the problem of choosing an error structure for panel data the same way that

a time series analyst decides on a particular time series specification. Given the class of models considered in this paper, an important part of the statistical problem is to find those specifications that are consistent with the data.

The statistical framework considered below goes beyond the single equation case and proposes the use of vector ARMA processes as a method for pooling cross section and time series data in the multi-equation case. Multiple time series models offer a robust scheme for combining systems of equations like those found in seemingly unrelated or simultaneous equation models. These models allow one to use time series techniques to estimate simultaneously several structural distributed lag relationships involving both endogenous and exogenous variables and still allow for a general autocorrelation pattern for disturbances.

This study provides a general method for fitting multiple time series models to panel data. It is especially tailored for analyzing a panel data set that has a large number of individuals and a relatively short time series. Special problems arise when one uses panel data to estimate time series models, and some procedures found in the time series literature are not directly applicable. On the other hand, panel data permits the consideration of more general parameterization than is possible in standard time series analysis. This study proposes solutions to the special problems, and it identifies the generalizations of empirical specifications possible in the analysis of panel data.

To aid in the choice of model specifications, techniques found in the time series literature for identifying the forms of distributed lag relationships and the orders of ARMA processes are adapted for application in a panel data setting. These techniques prevent one from choosing specifications for

the DSEM and for the error process that are broadly inconsistent with the data. Treating residuals as dependent variables in a seemingly unrelated regression analysis, it is possible to compute estimates of the sample covariogram, correlogram, and partial correlation function associated with disturbances and the standard errors of these estimates. As in the case of standard time series analysis, these estimates provide information to choose among competing ARMA specifications for the error process and they provide the basis for simple tests for several forms of nonstationarity and heteroscedasticity.

This study also develops general methods for estimating the models and the error processes described above. Specifically, these methods provide for the estimation of the parameters of a system of seemingly unrelated regression or simultaneous equations including parameters of the covariance matrix when these parameters are subject to an arbitrary set of nonlinear constraints. These constraints may simultaneously involve regression coefficients and parameters of the covariance matrix. Estimators based on least squares and on quasi-maximum likelihood procedures are proposed that do not rely on any specific distributional assumptions for their asymptotic properties. Both limited information and full information estimation procedures are developed. These procedures are computationally efficient and simple to implement.

Section I presents a general statistical framework and considers a wide range of issues associated with model specification. Section II develops data analysis techniques, and Section III describes estimation procedures.

I. A General Statistical Framework

Panel data offers observations on a sample of "individuals" in more than one time period. An individual here refers to an observational unit such as a household or a firm. The models developed below are designed to estimate the structures relating an individual's variables both within and across time periods using all the available data.

Dynamic simultaneous equations offer a flexible framework for describing the relationships between an individual's measured variables. A set of structural equations from a DSEM, suitably modified for a panel data analysis, may be written as

$$(1) \qquad \sum_{j=0}^{n} \Gamma_j Y_{1i}(t-j) = \sum_{j=0}^{r} \Psi_j Y_{2i}(t-j) + \sum_{j=0}^{s} B_j X_i(t-j) + U_i(t), \quad t = 1,\ldots,T, \quad i = 1,\ldots,N,$$

where

$Y_{1i}(t)$ = g x 1 vector of endogenous variables for individual i in time period t,

$Y_{2i}(t)$ = h x 1 vector of endogenous variables,

$X_i(t)$ = m x 1 vector of exogenous variables,

Γ_j, $j = 0,\ldots,n$, Ψ_j, $j = 0,\ldots,r$, and B_j, $j = 0,\ldots,s$, are coefficient matrices of order g x g, g x h, and g x m, respectively,

and

$U_i(t)$ = g x 1 disturbance vector.

In terms of matrix lag operators, this system of equations may alternatively be written as

$$(2) \qquad \Gamma(L)Y_{1i}(t) = \Psi(L)Y_{2i}(t) + B(L)X_i(t) + U_i(t), \qquad t = 1,\ldots,T, \quad i = 1,\ldots,N,$$

where L is the lag operator (i.e., $L^jX_i(t) = X_i(t-j)$), and $\Gamma(L) \equiv \sum_{j=0}^{n} \Gamma_j L^j$, $\Psi(L) \equiv \sum_{j=0}^{r} \Psi_j L^j$, and $B(L) \equiv \sum_{j=0}^{s} B_j L^j$ are finite order matrix lag operators. Period 1 (i.e., t = 1) in this model refers to the first period of the panel in which one observes all of the current and lagged values of both the endogenous and the exogenous variables. Thus, T = T* - max (n, r, s) where T* = total number of time periods supplied by the panel data source and n, r, and s are the orders of the lag polynomials $\Gamma(L)$, $\Psi(L)$ and B(L). The following analysis assumes that time dummies or polynomials in time are included among the exogenous variables to capture period effects that are common across individuals. Thus, we have N independent sets of T time series observations with which to estimate the parameters of model (2). The above specification assumes that a researcher wishes to analyze g equations at once, and the procedures developed below apply to the analysis of this case. A researcher may desire to consider more than one equation at a time in order to achieve parameter identification, to impose restrictions across equations, or to obtain more efficient estimates.

Sepcification (2) includes virtually all econometric models that are linear in measured variables as special cases. If a researcher chooses to analyze a single structural equation (or one equation at a time), then in (2) one sets g = 1, h = the number of separate endogenous variables on the right-hand side of the equation, and n = the number of separate exogenous variables.[1]

[1]Notice that a variable is said to enter the equation if either its current or its lagged value appears on the right-hand side of equation (2).

If $\Gamma(L) = \Gamma_0$ and $\Psi(L) = \Psi_0$ where Γ_0 and Ψ_0 are coefficient matrices of orders g x g and g x h, respectively, the system of equations given by (2) reduces to a standard simultaneous equation model (i.e., there are no lagged endogenous variables) with g equations per period. If, in addition, $\Gamma_0 = I_g$ and $\Psi_0 = 0$, we obtain a seemingly unrelated or a multivariate regression model with g equations per period.[1] If, still further, g = 1, we have a standard multiple regression model where there is a single equation per period for each individual.

The statistical framework given by (2) permits the consideration of a wide variety of distributed lag relationships between the elements of Y_1, Y_2, and X, including infinite order schemes. The assumption that the lag polynomials $\Gamma(L)$, $\Psi(L)$, and B(L) are of finite order is not as restrictive as it may first appear. It is possible to estimate any infinite order distributed lag relationship which can be written as a ratio of finite order lag polynomials using model (2). Such lag schemes, known as rational distributed lags, admit flexible weight patterns on past variables and contain many well known schemes as special cases.[2]

To see how it is possible to analyze this type of lag structure within the framework of (2), consider a single equation version of the model where Y_2 and X each represent a single variable. Assuming that Y_1 is related to Y_2 and to X through rational distributed lag schemes, we have

---

[1] "Seemingly unrelated regression model" in this paper refers to any system of regression equations whose disturbances are not assumed to be uncorrelated. These equations may or may not contain the same set of explanatory variables, and there may be parameter constraints across equations. Also, there may exist some form of covariance restrictions for disturbances included in a system, and there is no requirement that disturbances are uncorrelated across systems.

[2] See Griliches (1967) for a survey of these distributed lag schemes.

$$(3) \qquad Y_{1i}(t) = \frac{b_1(L)}{b_2(L)} Y_{2i}(t) + \frac{c_1(L)}{c_2(L)} X_i(t) + U_i^*(t)$$

where $b_1(L)$, $b_2(L)$, $c_1(L)$, and $c_2(L)$ are lag polynomials of finite (and typically low) order, and $U_i^*(t)$ is a disturbance.[1] Multiplying both sides of this equation by $b_2(L)c_2(L)$ converts it into the form of equation (2) with $\Gamma(L) = b_2(L)c_2(L)$, $\Psi(L) = b_1(L)c_2(L)$, $B(L) = c_1(L)b_2(L)$, and $U_i(t) = b_2(L)c_2(L)U_i^*(t)$ where $\Gamma(L)$, $\Psi(L)$ and $B(L)$ are all finite order. There is, then, an equivalent relationship between equations (2) and (3). Analyzing rational distributed lag schemes using specification (2) will, in general, imply nonlinear restrictions relating the coefficients of the polynomials $\Gamma(L)$, $\Psi(L)$, and $B(L)$. This, however, introduces no significant complications in the following discussion. Both the data analysis and the estimation procedures developed below permit the imposition of such restrictions. While imposing these constraints will, in general, yield more efficient parameter estimates, it is important to recognize that one can estimate all the parameters needed to construct rational distributed lag structures without imposing any restrictions in the estimation of equation (2). One can always convert the estimates of equation (2) into those of equation (3) by observing that $\frac{\Psi(L)}{\Gamma(L)} = \frac{b_1(L)}{b_2(L)}$ and $\frac{B(L)}{\Gamma(L)} = \frac{c_1(L)}{c_2(L)}$.

The following analysis assumes that the error terms, $U_i(t)$, are distributed independently across individuals (i.e., across the index i), but that these disturbances are autocorrelated over time (i.e., over the index t) for the same individual. It is often useful or necessary to restirct the

[1] The polynomials $b_2(L)$ and $c_2(L)$ are assumed to have roots that lie strictly outside the unit circle.

form of this autocorrelation. Imposing such restrictions reduces the number of parameters in a model; it can create a statistical model that may be used for prediction outside the sample period; or, in the case of a simultaneous equation model, it can aid in securing the identification of structural parameters. A natural specification for the error process in a pandel data setting is one that merges linear multiple time series models with variance component schemes.

This study considers such an error structure. In particular, $U_i(t)$ is assumed to follow a permanent-transitory scheme of the form

$$U_i(t) = \phi_i + \nu_i(t) \tag{4}$$

where

ϕ_i = g x 1 vector of permanent components with

$$E(\phi_i \phi_j') = \begin{cases} \Phi & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

and the $\nu_i(t)$ is a g x 1 vector of transitory components uncorrelated with ϕ_i and generated by the multiple time series process

$$\nu_i(t) = -\sum_{j=1}^{p} A_j \nu_i(t-j) + \varepsilon_i(t) + \sum_{j=1}^{q} M_j \varepsilon_i(t-j) \tag{5}$$

which may be equivalently written as

$$A(L)\nu_i(t) = M(L)\varepsilon_i(t) \tag{6}$$

where

$A(L) \equiv \sum_{j=0}^{p} A_j L^j$ and $M(L) \equiv \sum_{j=0}^{q} M_j L^j$ are g x g matrix lag operators with $A_0 = M_0 = I_g$ and the roots of $|M(L)| = 0$ are assumed to lie on or outside the unit circle,[1]

and

$\epsilon_i(t)$ = g x 1 vector of white noise with

$$E(\epsilon_i(t)\epsilon_j'(t^*)) = \begin{cases} \Sigma & \text{if } t = t^* \text{ and } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $U_i(t)$ is the sum of a vector of correlated permanent components, ϕ_i, and a vector of individual specific variates, $\nu_i(t)$, which follows a multivariate ARMA process. There are two sources of autocorrelation accounted for in this error specification: one is due to the presence of permanent components which capture the effects of unmeasured characteristics unique to the individual that remain constant over the sample period; and the other source is the time series components which account for the existence of unobserved variables that vary systematically from one period to the next. This error process admits a wide variety of autocorrelation patterns and provides a flexible scheme for describing the time series aspect of panel data with a minimal number of parameters.

Previous studies on longitudinal analysis have considered special cases of the error specification proposed above for a single equation (which implies g = 1 in (4) and (6)). The most popular specification is one that combines a permanent component with a pure autoregressive scheme

[1] The restriction on the coefficients of M(L) is the usual one imposed in the time series literature to guarantee that these coefficients are identified.

(i.e., $p > 0$ and $q = 0$ in (6)).[1] A few studies consider a permanent component and a low order moving average process (i.e., $g = 1$, $p = 0$, and $q > 0$).[2] No study considers a mixture of an autoregressive and a moving average process.

There are several other error specifications found in the literature that can be analyzed using the statistical model for disturbances given by (4) and (6). One such specification attempts to generalize the above model following the suggestions of Nerlove (1967) in his work on "unobserved components." Disturbances are assumed to depend on more than one transitory component. In particular, instead of (4), it is assumed that

$$U_i(t) = \phi_i + \sum_{\ell=1}^{J} \nu_{\ell i}(t) \qquad t = 1,\ldots,T, \quad i = 1,\ldots,N,$$

where the transitory components $\nu_{\ell i}(t)$, $\ell = 1,\ldots,J$, are mutually independent and each follows a restricted ARMA process of the form

$$A_\ell(L)\nu_{\ell i}(t) = M_\ell(L)\varepsilon_{\ell i}(t) \qquad \ell = 1,\ldots,J,$$

where $A_\ell(L)$ and $M_\ell(L)$ are matrix lag operators with the same properties as $A(L)$ and $M(L)$ defined above, and the $\varepsilon_{\ell i}(t)$'s are white noise vectors. According to this specification, the disturbance vector of the DSEM depends on permanent components and a sum of J time varying components that are each generated by a unique multivariate ARMA process. Since this new error process includes the simpler process proposed above as a special case with $J = 1$,

[1] David (1971), Hause (1977), Lillard-Willis (1978) and Lillard-Weiss (1979) are examples of studies that estimate first order autoregressive schemes (i.e., $p = 1$). Ashenfelter (1978) considers higher order processes.

[2] Friedman (1954, p. 353) and Hause (1977) are examples of such studies.

some authors have offered this new process as a way of providing for a wider class of autocorrelation structures for disturbances.[1] This more complex error process, however, does not admit more general autocorrelation structures. This result is a direct consequence of the fact that summing disturbances generated by ARMA processes yields a new disturbance that also follows an ARMA process.[2] Thus, any autocorrelation pattern produced by the complex error process can be duplicated by the simpler process given by (4) and (6).[3]

---

[1] Hause (1977), Lillard-Weiss (1978) and Lillard-Weiss (1979) estimate elementary specifications of this error process for the case J = 2. The most widely estimated specification assumes that $\nu_{1i}(t)$ follows a first order autoregressive process and $\nu_{2i}(t)$ is white noise.

[2] Define the disturbance vector $\eta(t) = \sum_{\ell=1}^{J} \nu_\ell(t)$ where each of the $\nu_\ell(t)$'s follows an ARMA process of the form assumed in the text (i.e., $A_\ell(L)\nu_\ell(t) = M_\ell(L)\varepsilon_\ell(t)$, $\ell = 1,\ldots,J$). It is always possible to represent the ARMA process for $\nu_\ell(t)$ as $|A_\ell(L)|\nu_\ell(t) = A_\ell^*(L)M_\ell(L)\varepsilon_\ell(t) \equiv M_\ell^*(L)\varepsilon_\ell(t)$ where $|A_\ell(L)|$ and $A_\ell^*(L)$ are the determinant and the adjoint matrix associated with $A_\ell(L)$, and $M^*(L)$ is a fine order matrix defined as $A_\ell^*(L)M_\ell(L)$. Thus, premultiplying $\eta(t)$ by $\rho(L) \equiv \prod_{\ell=1}^{J} |A_\ell(L)|$ yields $\rho(L)\eta(t) = \sum_{\ell=1}^{J} \theta_\ell(L)M^*(L)\varepsilon_\ell(t)$ where $\theta_\ell(L) = \rho(L)/|A_\ell(L)|$ is a finite order polynomial. The right-hand side summation expression is known to have a vector moving average representation since its autocorrelation function is zero after finite order (see Hannan (1970, p. 66)). Thus, we see that $\eta(t)$ follows a vector ARMA process.

[3] There is, then, a fundamental identification problem associated with the use of the complex specification of the error process. To estimate this specification, one requires a priori restrictions on each of the ARMA processes generating transitory components.

One can also analyze error specifications in which permanent components are not uncorrelated with "exogenous" variables or in which disturbances depend on individual specific growth rate terms as well as permanent and time series components. For those specifications where ϕ_i is correlated with the $X_i(t)$'s, first-differencing equations (i.e., multiplying both sides of each equation by (1-L)) eliminates permanent components and creates a new model that satisfies the assumptions of the DSEM and error process proposed above. Similarly, if disturbances depend on individual specific growth rate terms and, instead of (4), $U_i(t) = \phi_i + b_i t + v_i(t)$ where b_i is a g x 1 vector of permanent components distributed randomly across individuals,[1] then first-differencing can once again be used to transform this error specification into a model like those proposed above.[2] In the case of an individual specific quadratic trend (i.e., $b_i t^2$), second-differencing puts the model into the appropriate form. Differencing, then, offers a way of collapsing more general DSEM's and error processes into a specification which is nested within the framework considered in this paper. It is important to recognize that differencing changes the specification of the DSEM and the ARMA process for transitory disturbances in a known way and introduces no new parameters. Thus, the effects of differencing can easily be undone in the sense that one can construct the model associated with levels using only the parameter estimates of the differenced equations.

One can estimate rational distributed lag relationships using the strategy outlined above when disturbances are assumed to follow the error

[1] Such error specifications are common in the empirical literature on earnings (see, for example, Hause, 1977 and Lillard and Weiss, 1979).

[2] First differencing in this case reduces random trends to permanent components.

scheme proposed above. If the disturbance $U_i^*(t)$ appearing in (3) follows a permanent-transitory scheme of the sort given by (4) and (6), then the transformed disturbance $U_i(t)$ in specification (2) (which equals $b_2(L)c_2(L)U_i^*(t)$) follows an error scheme of the same form. Thus, using model (2) along with error processes (4) and (6), one can fully estimate rational distributed lag structures while imposing covariance restrictions. Translating from equation (3) to specification (2) will, in general, imply nonlinear restrictions relating the coefficients of the lag polynomials $\Gamma(L)$, $\Psi(L)$ and $B(L)$ associated with measured variables and the polynomials $A(L)$ and $M(L)$ determining the autocorrelation of disturbances. The data analysis techniques and the estimation procedures developed below permit one to consider and to incorporate these types of constraints. Introducing the possibility of covariance restrictions in the analysis of rational lags can lead to an increase in the efficiency of estimation, and it can provide a source of parameter identification which relaxes the need for exclusion restrictions and exogeneity assumptions.

A Familiar Representation

Combining all the structural equations for an individual into one model creates an alternative representation of the above DSEM that is particularly useful for the analysis of panel data. Stacking the equations given by (1) for individual i in descending order starting with the last period yields

$$(7)\qquad \begin{pmatrix} \sum_{j=0}^{n} \Gamma_j Y_{1i}(T-j) \\ \vdots \\ \sum_{j=0}^{n} \Gamma_j Y_{1i}(1-j) \end{pmatrix} = \begin{pmatrix} \sum_{j=0}^{r} \Psi_j Y_{2i}(T-j) \\ \vdots \\ \sum_{j=0}^{r} \Psi_j Y_{2i}(1-j) \end{pmatrix} + \begin{pmatrix} \sum_{j=0}^{s} B_j X_i(T-j) \\ \vdots \\ \sum_{j=0}^{s} B_j X_j(1-j) \end{pmatrix} + \begin{pmatrix} U_i(T) \\ \vdots \\ U_i(1) \end{pmatrix} .$$

In matrix notation, this system of equations may be written as

$$(8)\qquad \underset{\sim}{\Gamma} Y_{1i} = \underset{\sim}{\Psi} Y_{2i} + \underset{\sim}{B} X_i + U_i \qquad i = 1,\ldots,N$$

with

$$E(U_i U_j') = \begin{cases} 0 & i = j \\ 0 & \text{otherwise} \end{cases}$$

where $Y_{1i}' = (Y_{1i}'(T), Y_{1i}'(T-j),\ldots,Y_{1i}'(1-n+1), Y_{1i}(1-n))$,

$Y_{2i}' = (Y_{2i}'(T),\ldots,Y_{2i}'(1-r))$, $X_i' = (X_i'(T),\ldots,X_i'(1-s))$, $U_i' = (U_i'(T),\ldots,$

$U_i'(1))$, and the coefficient matrices $\underset{\sim}{\Gamma}$, $\underset{\sim}{\Psi}$, and $\underset{\sim}{B}$ are block diagonal band matrices with the matrices $[\Gamma_0,\ldots,\Gamma_n]$, $[\Psi_0,\ldots,\Psi_r]$, and $[B_0,\ldots,B_s]$ running

down the diagonal of $\underset{\sim}{\Gamma}$, $\underset{\sim}{\Psi}$, and $\underset{\sim}{B}$, respectively.[1] Written in the form of (8), we have constructed a system of simultaneous equations in which disturbance vectors are independently distributed over observations so that it is possible to estimate the unknown parameters of the coefficient matrices $\underset{\sim}{\Gamma}$, $\underset{\sim}{\Psi}$, and $\underset{\sim}{B}$ using standard simultaneous equation estimation procedures.[2]

The main consequence of assuming a DSEM of the sort presented in equations (1) or (2) when analyzing panel data is the imposition of constraints across equations in different time periods for a given individual. Inspection of equation (7) reveals that the specification assumed in (1) implies equality constraints across the rows of $\underset{\sim}{\Gamma}$, $\underset{\sim}{\Psi}$, and $\underset{\sim}{B}$.

One obvious generalization of the above DSEM that is possible when this model is used to analyze panel data rather than time series data concerns the constancy of matrix lag operators over time. In specifying the DSEM given

---

[1] A matrix Q is a diagonal band or a block diagonal band matrix if it has the form

$$Q = \begin{bmatrix} a & b & c & d & & & & \\ & a & b & c & d & & & \\ & & \ddots & \ddots & \ddots & \ddots & & 0 \\ & & & \ddots & \ddots & \ddots & \ddots & \\ 0 & & & & a & b & c & d \\ & & & & & a & b & c & d \end{bmatrix}.$$

If a, b, c, and d are constants, then Q is a diagonal band matrix. If a, b, c, d are matrices, then Q is a block diagonal band matrix with the matrix [a, b, c, d] running down the diagonal.

[2] These procedures include two stage and three stage least squares methods that assume disturbances are correlated across equations but not across observations. When a nonsimultaneous specification of the DSEM consits of g multiple regression equations per period (i.e., $\Gamma(L) = I_g$ and $\Psi(L) = 0$ in (2)), we have $\underset{\sim}{\Gamma} = I_{Tg}$ and $\underset{\sim}{\Psi} = 0$, and (8) becomes a seemingly unrelated regression model which can be estimated using standard joint generalized least squares procedures. See Section III for further discussion.

by (2), it is not required in a panel data analysis to assume that the matrix lag operators $\Gamma(L)$, $\Psi(L)$ and $B(L)$ are the same across time periods. Instead, one can add a "t" subscript to these matrices indicating that there is a new set of coefficients for each period and, thus, a different distributed lag relationship. The consequence of this generalization in (8) is the relaxation of equality constraints relating the rows of $\underset{\sim}{\Gamma}$, $\underset{\sim}{\Psi}$, and $\underset{\sim}{B}$.

A Specification for the Covariance Matrix

The consequence of assuming that disturbances appearing in (1) follow the error specification given by (4) and (6) is the imposition of restrictions on the covariance matrix $\Theta = E(U_i U_i')$ associated with the stacked representation of the DSEM given by (8). The following analysis derives the exact restrictions on autocovariances implied by the combined variance component-multiple time series process assumed above, and it formulates an explicit parameterization for the covariance matrix Θ.

According to (4), each component of the disturbance vector U_i is generated by an error model of the form $U_i(t) = \phi_i + \nu_i(t)$, $t = 1,\ldots,T$, where $E(\phi_i\phi_i') = \Phi$, and $\nu_i(t)$ follows the multivariate ARMA process, $A(L)\nu_i(t) = M(L)\varepsilon_i(t)$, which is distributed independently of ϕ_i with $E(\varepsilon_i(t)\varepsilon_i'(t)) = \Sigma$. Defining $\nu_i' = (\nu_i'(T),\ldots,\nu_i'(1))$ and ι as a $T \times 1$ vector of ones, we have $U_i = (\iota \otimes \phi_i) + \nu_i$; so,

$$(9) \qquad \Theta = E(U_i U_i') = (\iota\iota' \otimes \Phi) + E(\nu_i \nu_i').$$

To specify Θ, we need the implied parameterization for $E(\nu_i\nu_i')$.

According to (5), ν_i is determined by the system of equations

$$(10) \qquad \nu_i = \begin{pmatrix} \nu_i(T) \\ \cdot \\ \cdot \\ \cdot \\ \nu_i(1) \end{pmatrix} = - \begin{pmatrix} \sum_{j=1}^{p} A_j \nu_i(T-j) \\ \cdot \\ \cdot \\ \sum_{j=1}^{p} A_j \nu_i(1-j) \end{pmatrix} + \begin{pmatrix} \sum_{j=0}^{q} M_j \varepsilon_i(T-j) \\ \cdot \\ \cdot \\ \sum_{j=0}^{q} M_j \varepsilon_i(1-j) \end{pmatrix}$$

where $M_0 = I$. This system does not represent a one-to-one transformation from the $\varepsilon_i(t)$'s, $t = 1,\ldots,T$, to ν_i. One cannot, then, derive the covariance matrix for ν_i from (10) if given only distributional assumptions for $\varepsilon_i(T),\ldots,\varepsilon_i(1)$. Also appearing in (12) are the variables $\nu_i(0),\ldots,\nu_i(1-p)$ and $\varepsilon_i(0),\ldots,\varepsilon_i(1-q)$ which are known in the time series literature as initial conditions or starting values for the error process. To derive a parameterization for $E(\nu_i \nu_i')$, one requires a treatment for initial conditions.

This paper treats initial conditions for disturbances as random variables. Conventional time series techniques that treat starting values as known constants (usually chosen to be zero) result in inconsistent estimates for the parameters of the error process if applied in a panel data analysis where T is fixed because, in contrast to time series analysis, initial conditions do not become "irrelevant" as the sample size increases. Similarly, time series procedures that "backforecast" or treat initial conditions as parameters introduce an incidental parameters problem in a panel data analysis which under most circumstances also leads to inconsistent estimates for all parameters of the error process.[1] Treating initial conditions as random variables avoids problems with inconsistency by introducing only a finite number of new parameters: those determining the distribution of the starting

[1] This problem of incidental parameters and inconsistent estimation is examined by Anderson-Hsiao (1981).

values, and those relating the distribution of the starting values to the distribution of disturbances realized in periods 1 through T.

There are several complications associated with choosing a distribution for the two sets of initial conditions specified above. If we assume that the stochastic process generating disturbances during the sample period is also operative prior to this period, then one would expect the $\nu_i(\ell)$'s, $\ell = 1\text{-}p,\ldots,0$, to be not only correlated with one another and with the $\varepsilon_i(j)$'s, $j = 1\text{-}q,\ldots,0$, but also with all the $\nu(t)$'s realized after period 0. Furthermore, the correlations relating these variables will, in general, depend directly on parameters of the ARMA process given by (6), and one must account for these restrictions to achieve efficient estimation. It is possible to minimize these complications by specifying the system of equations given by (10) and considering an alternative expression for ν_i.

A moving average representation of an ARMA process provides the basis for this new expression for ν_i. Assuming the multiple time series process given by (7) started sometime in the <u>finite</u> past, say between the periods $\tau\text{-}b$ and τ with $b \geq 0$, it is possible to write each of the $\nu_i(t)$'s realized after period τ as a moving average scheme of the form

$$(11) \qquad \nu_i(t) = \sum_{j=0}^{t-\tau-1} K_j \varepsilon_i(t-j) + \sum_{j=t-\tau}^{t-\tau+b} K_j f_i(t-j) \qquad t > \tau$$

where

$$K_0 = I_g$$

$$K_1 = M_1 - A_1$$

$$K_2 = M_2 - A_2 - A_1 K_1$$

$$\vdots$$

$$K_k = M_k - \sum_{j=1}^{k} A_j K_{k-j}$$

and the $f_i(\ell)$'s, $\ell = \tau-b,\ldots,\tau$, are error vectors distributed independently of $\varepsilon_i(t^*)$ for all $t^* > \tau$. Formally, one can derive a relation like (11) by starting with the ARMA representation $\nu_i(t) = -\sum_{j=1}^{p} A_j\nu_i(t-j) + \sum_{j=0}^{q} M_j\varepsilon(t-j)$ and successively substituting out for past $\nu_i(t-j)$'s using their ARMA representations until $t-j = \tau$. One can readily verify that (11) is indeed a valid expression for $\nu_i(t)$ since premultiplying this equation by $A(L)$ yields $A(L)\nu_i(t) = M(L)\varepsilon_i(t)$ for $t > \tau+q$. The $f_i(\ell)$'s in (11) may be interpreted as the true initial conditions of the ARMA process. Specifying the distribution of these variables determines exactly how and when the ARMA process generating $\nu_i(t)$'s began.

Using (11), it is possible to reformulate the system of equations given by (10). To avoid the need for dealing with several possible cases, it is convenient to introduce the notation $K_j = 0$ for $j < 0$ (for $j \geq 0$, K_j is defined in (11)) and the definition that a summation of the form $\sum_{\ell=0}^{c}$ is equal to zero whenever $c < 0$. Using this notation, equations (10) and (11) imply

$$
\text{(12)}\qquad
\begin{pmatrix}
\sum_{j=0}^{p} A_j\nu_i(T-j) \\
\vdots \\
\sum_{j=0}^{p} A_j\nu_i(p+1-j) \\
\nu_i(p) \\
\nu_i(p-1) \\
\vdots \\
\nu_i(1)
\end{pmatrix}
=
\begin{pmatrix}
\sum_{j=0}^{q} M_j\varepsilon_1(T-j) \\
\vdots \\
\sum_{j=0}^{q} M_j\varepsilon_i(p+1-j) \\
\sum_{j=0}^{q-1} K_j\varepsilon_i(p-j) \\
\sum_{j=0}^{q-1} K_{j-1}\varepsilon_i(p-j) \\
\vdots \\
\sum_{j=0}^{q-1} K_{j-p+1}\varepsilon_i(p-j)
\end{pmatrix}
+
\begin{pmatrix}
0 \\
\vdots \\
0 \\
\eta_i(p) \\
\eta_i(p-1) \\
\vdots \\
\eta_i(1)
\end{pmatrix}
$$

where

$$\eta_i(t) = \sum_{j=t-p+q}^{t-\tau-1} K_j \varepsilon_i(t-j) + \sum_{j=t-\tau}^{t-\tau+b} K_j f_i(t-j), \qquad t = 1,\ldots,p.$$

The first set of T-p sets of equations in (12) are simply the standard representation of the ARMA process generating $\nu_i(p+1),\ldots,\nu_i(T)$, and the second p sets of equations are the moving average representations of the ARMA process for $\nu_i(1),\ldots,\nu_i(p)$ with the $\eta_i(\ell)$'s, $\ell = 1,\ldots,p$, defined to include all disturbances realized prior to period p-q+1 (i.e., the $\varepsilon_i(t)$'s and the $f_i(t)$'s for $t \leq p-q$). The formulation of (12) assumes that $\tau \leq p-q$.[1] In matrix notation, (12) may be written as

$$(13) \qquad F\,\nu_i = G\begin{pmatrix}\varepsilon_i \\ \eta_i\end{pmatrix}$$

with

$$\nu_i = \begin{pmatrix}\nu_i(T) \\ \cdot \\ \cdot \\ \cdot \\ \nu_i(1)\end{pmatrix}, \quad \varepsilon_i = \begin{pmatrix}\varepsilon_i(T) \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_i(p-q+1)\end{pmatrix}, \quad \eta_i = \begin{pmatrix}\eta_i(p) \\ \cdot \\ \cdot \\ \cdot \\ \eta_i(1)\end{pmatrix},$$

and F is a gT x gT matrix and G is gT x g(T+q) matrix defined as

$$F = \begin{bmatrix} F_1 \\ \\ F_{21} \quad F_{22} \end{bmatrix} = \begin{bmatrix} A \\ \\ 0 \quad I_{gp} \end{bmatrix},$$

[1] This assumption concerning the starting time of the ARMA process generating the $\nu_i(t)$'s is a weak restriction and follows immediately from the assumption that $\nu_i(p+1)$ can be represented by the specification given by (5). This restriction on τ ensures that no $f_i(t)$'s appear in the moving average component of (5) for t = p+1.

$$G = \begin{bmatrix} G_{11} & & G_{12} \\ & & \\ G_{21} & G_{22} & G_{23} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{M} & & 0 \\ & & \\ 0 & \underline{K} & I_{gp} \end{bmatrix},$$

where: $F_1 = \underset{\sim}{A}$ is a $g(T-p) \times gT$ block diagonal band matrix with the matrix $[A_0,\ldots,A_p]$ running down the diagonal; $F_{21} = 0$ is a $gp \times g(T-p)$ matrix; $F_{22} = I_{gp}$ is an identity matrix of dimension gp; $G_{11} = \underset{\sim}{M}$ is a $g(T-p) \times g(T-p+q)$ block diagonal band matrix with the matrix $[M_0,\ldots,M_q]$ running down the diagonal; $G_{12} = 0$ is a $g(T-p) \times gp$ matrix; $G_{21} = 0$ is a $gp \times g(T-p)$ matrix, $G_{23} = I_{gp}$; and $G_{22} = \underline{K}$ is a $gp \times gq$ with

$$\underline{K} = \begin{bmatrix} K_0 & K_1 & \cdots & K_{q-2} & K_{q-1} \\ K_{-1} & K_0 & \cdots & K_{q-3} & K_{q-2} \\ \vdots & \vdots & & \vdots & \vdots \\ K_{1-p} & K_{2-p} & \cdots & K_{q-p-1} & K_{q-p} \end{bmatrix}.$$

When forming the partioned matrices associated with F and G, the above analysis assumes that any matrix with an implied dimension equal to zero is deleted from the specification. Thus, when $p = 0$, $F = [F_1]$ and $G = [G_{11}]$; and when $q = 0$, $\underline{K}$ is eliminated and $G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{23} \end{bmatrix}$. The above specification for $\underline{K}$ is written in general terms to handle all possible orders of the ARMA process including $p < q$, $p = q$, and $p > q$. Recognizing $K_j = 0$ for $j < 0$ and $K_j \neq 0$ for $j \geq 0$ reveals that $\underline{K}$ has a fairly simple form with nonzero elements in those K_j's on and above the positions (1, 1), (2, 2),...,(min(q,p), min(q,p)), and zero elements in those K_j's below and left of these positions.[1]

[1]Formally, the matrix $\underline{K}$ has $[K_0, K_1,\ldots,K_{q-1}]$ as its first set of

Given the expression for ν_i implied by (13), the problem of parameterizing $E(\nu_i \nu_i')$ becomes one of specifying a correlation structure for the disturbance vectors ε_i and η_i. Since each of the components of ε_i follow a white noise error process, we have

$$(14) \qquad E(\varepsilon_i \varepsilon_i') = (I_{T-p+q} \otimes \Sigma) \equiv \Sigma^*$$

where $\Sigma = E(\varepsilon_i(t)\varepsilon_i'(t))$ for $t = p-q+1,\ldots,T$ and Σ^* is defined by the Kronecker product. Inspection of the formula for the $\eta_i(\ell)$'s reveals three facts: (i) the $\eta_i(\ell)$'s depend on a common set of disturbances; (ii) all of these disturbances are realized prior to period $p+q-1$; (iii) included among these disturbances are the initial conditions for the ARMA process (i.e., the $f_i(t)$'s). Since each of the components of ε_i are realized during and after period $p+q-1$, fact (ii) implies $E(\eta_i \varepsilon_i') = 0$. Fact (i) implies that the components of η_i are mutually correlated, so $E(\eta_i \eta_i')$ contains no zero elements in general. In addition, without imposing rigorous restrictions on both the number and the correlation structure of initial conditions, fact (iii) indicates that no restrictions will exist on the form of $E(\eta_i \eta_i')$. In the general case, the η_i will possess an arbitrary covariance structure which we may formally express as

$$(15) \qquad E(\eta_i \eta_i') = \Lambda$$

where Λ is any positive definite symmetric matrix.

Combining the above results, we obtain the following specification for Θ

---

g rows, $[0, K_0, \ldots, K_{q-2}]$ as its second set of g rows, and so on until the p^{th} set of g rows is reached, or if $q > p$, until the q^{th} set of g rows is reached after which the rows of $\underline{K}$ contain zeros.

$$(16) \qquad \Theta = E(\nu_i \nu_i') = F^{-1} G \begin{bmatrix} \Sigma^* & 0 \\ 0 & \Lambda \end{bmatrix} G' F^{-1'}.$$

This parameterization imposes all of the restrictions implied by the ARMA process unless one is willing to introduce precise information about how and when this process started.

There are two modifications of the above parameterization for Λ that may be useful in applied work. First, to simplify the construction of the matrix $\underset{\sim}{K}$, one can replace each of the nonzero elements of $\underset{\sim}{K}$ (i.e., all the K_j's, $j \geq 0$) by arbitrary parameters rather than using the coefficients of the ARMA process and the formulas specified above to form these elements. This modification avoids the need for imposing nonlinear restrictions, but it introduces new parameters and reduces the efficiency of estimation.

The second modification concerns the parameterization of Λ defined by (15). This matrix is purely a theoretical construct and represents nuisance parameters. An unattractive feature of this parameterization is that one cannot easily infer an approximate value for Λ using preliminary data analysis techniques or estimation methods that do not require the full estimation of Θ. The availability of approximate values greatly reduces computational burden when used as starting values for parameters in a nonlinear computer program which is invoked to estimate Θ. A way around this problem is to replace Λ by $\Lambda^* = E(\eta_i \eta_i') + \underset{\sim}{K}(I_q \otimes \Sigma)\ \underset{\sim}{K}'$ which is also only restricted to be positive definite and symmetric. Substituting this new parameterization into (16) implies

$$(17) \qquad \Theta = F^{-1} \begin{bmatrix} \underset{\sim}{M} & \Sigma^* & \underset{\sim}{M}' & \underset{\sim}{M} & \Sigma^* \begin{bmatrix} 0' \\ \underset{\sim}{K}' \end{bmatrix} \\ [0 \quad \underset{\sim}{K}] & \Sigma^* & \underset{\sim}{M}' & & \Lambda^* \end{bmatrix} F^{-1'}.$$

According to this new specification for θ, $\Lambda^* = E(v_i^* v_i^{*\prime})$ where the vector $v_i^{*\prime} = (v_i'(p),\ldots,v_i'(1))$ includes the last p components of v_i^*. In contrast to the previous parameterization, Λ^* is directly interpretable and can be easily estimated prior to the full estimation of θ.

There are several ways in which the above specification of θ can be generalized in the analysis of panel data that are not possible in standard time series analysis. Each of these generalizations involves a form of nonstationarity.

First, there is no requirement for the roots of the autoregressive matrix lag operator (i.e., the roots of $(A(L)) = 0$) to lie outside the unit circle. Thus, it is possible to consider such error processes as random walks when using panel data. Whereas in a time series analysis the existence of such nonstationarity has significant consequences on the asymptotic porperties of estimators, it has no such effects in the case of panel data where asymptotic results rely on a large number of individuals rather than a large number of time periods.

Second, it is possible in the analysis of panel data to permit the white noise vectors, $\varepsilon_i(t)$, to be heteroscedastic over time, which introduces yet another form of nonstationarity. To account for this heteroscedasticity in the above analysis, one only needs to define Σ^* appearing in (14) as a block diagonal matrix of the form $\mathrm{Dia}(\Sigma_T,\ldots,\Sigma_{T-p-q})$ where $\Sigma_t = E(\varepsilon_i(t)\varepsilon_i'(t))$. In standard time series analysis this sort of nonstationarity does not necessarily create any conceptional difficulties, but it does require an explicit parameterization of the suspected form of the heteroscedasticity that avoids an incidental parameters problem. In the case of panel data, however, it is possible to allow for arbitrary forms of heteroscedasticity of white noise disturbance vectors over time.

A third form of nonstationarity permitted in panel data analysis concerns the constancy of the autoregressive and the moving average matrix lag operators appearing in the multiple time series error process given by (6). The matrices A(L) and M(L) can be allowed to vary arbitrarily across periods so that there is a new set of autoregressive and moving average parameters for each t. To modify the above analysis to account for A(L) and M(L) being period specific, one must subscript the A_j, the M_j, and the K_j matrices appearing in the specifications of F and G defined by (13) to indicate the time period each matrix is associated with. This subscripting has the effect of relaxing the equality constraints across the rows of the matrices $\underset{\sim}{A}$ and $\underset{\sim}{M}$, and it essentially voids any constraints relating the nonzero elements of $\underline{K}$ to one another or to the coefficients of the ARMA process.

A Reduced Form

In the following analysis on estimation we require a reduced form specification for the simultaneous equation model given by (8). Write this specification as

$$Y_i = \Pi X_i + V_i \qquad i = 1,\ldots,N,$$

$$E(V_i V_j') = \begin{cases} \Omega & i = j \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

where Y_i is a vector that includes all the endogenous variables appearing in (8), the vector X_i contains all exogenous variables for each period including lags, Π is a coefficient matrix, and V_i is a vector of disturbances.

The various specifications of the DSEM considered above imply different restrictions on the Π and the Ω matrices. If considering a non-simultaneous specification (i.e., when $\Gamma(L) = I$ and $\Psi(L) = 0$ in (2)), then (8) is obviously its own reduced form which implies $\Pi = \underset{\sim}{B}$ and $\Omega = \Theta$. If considering the special simultaneous specification where there are no right-hand-side endogenous variables Y_{2i} and no lagged endogenous variables (i.e., when $\Gamma(L) = \Gamma_0$ and $\Psi(L) = 0$ in (2)), then $\underset{\sim}{\Gamma}$ is a nonsingular matrix, and (8) can be solved in the usual way for the reduced form by premultiplying through by $\underset{\sim}{\Gamma}^{-1}$ which implies $\Pi = \underset{\sim}{\Gamma}^{-1}B$ and $\Omega = \underset{\sim}{\Gamma}^{-1}\Theta\underset{\sim}{\Gamma}^{-1'}$.

If analyzing the general specification of the DSEM, however, (8) does not constitute a complete system of equations in the sense that there are more endogenous variables than there are equations; so, it is not possible to solve (8) for reduced form specification and determine the restrictions on Π and Ω without introducing additional equations. The strategy followed here to add the needed equations is the one normally used in limited information analysis of simultaneous equations; namely, a prediction equation is introduced for each endogenous variable that is not determined by the structural model under consideration. There are two sets of endogenous variables that are not directly determined by the DSEM considered above. The elements of the vector Y_{2i} defined by (8) constitute the first set, and the elements of the Y_{1i} realized prior to period 1 (i.e., the initial conditions for the $Y_{1i}(t)$'s) make up the second set which we group into the vector $Y_{1i}^{*\prime} = (Y_{1i}'(0),\ldots,Y_{1i}'(1-n))$. This study assumes that these variables are determined by the equations

$$\begin{pmatrix} Y_{1i}^{*} \\ \\ \\ Y_{2i} \end{pmatrix} = \Pi_2 X_i + V_{2i} \qquad i = 1,\ldots,N \tag{19}$$

where Π_2 is an unrestricted coefficient matrix, and V_{2i} is an error vector that is uncorrelated with all the elements of X_i. These specifications for prediction equations are not restrictive. It is always possible to define the disturbance vector V_{2i} so that it has zero mean, and it is uncorrelated with all the exogenous variables of the model. There is no guarantee that the covariance matrix of this error vector will be independent of X_i, but most of the procedures described below do not require the assumption that V_{2i} is homoscedastic across individuals. We maintain this assumption only to simplify the exposition.

Combining equations (19) with the structural model given by (8) implies a complete system of equations that can be solved for a reduced form like (18) with $Y_i' = (Y_{1i}', Y_{2i}')$ where Y_{1i} and Y_{2i} are defined by (8).[1] The parameter constraints implicit in (8) translate into restrictions on the Π and the Ω matrices. These restrictions can be shown to take the following form. Partitioning $\Pi = \begin{pmatrix} \Pi_1 \\ \Pi_2 \end{pmatrix}$ and $\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$ where Π_2 is the unrestricted set of coefficients and $E(V_{2i}V_{2i}') = \Omega_{22}$ is the unrestricted covariance matrix associated with (19), we have $\Pi_1 = \underset{\sim}{\Gamma}^{-1}(I - \underset{\sim}{\Psi})\Pi_2$, $\Omega_{11} = \underset{\sim}{\Gamma}^{-1}\underset{\sim}{\Theta}\underset{\sim}{\Gamma}'^{-1} - \underset{\sim}{\Gamma}^{-1}\underset{\sim}{\Psi}\Omega_{21} - \Omega_{12}\underset{\sim}{\Psi}\underset{\sim}{\Gamma}'^{-1} - \underset{\sim}{\Gamma}^{-1}\underset{\sim}{\Psi}\Omega_{22}\underset{\sim}{\Psi}'\underset{\sim}{\Gamma}'^{-1}$, and $\Omega_{12} = \Omega_{21}'$ is an unconstrained matrix. These restrictions, of course, collapse to those presented above for the special cases of the DSEM.

---

[1] The complete system of equations can be written as $HY_i = GX_i + \delta_i$ where $H = \begin{pmatrix} \underset{\sim}{\Gamma} & \underset{\sim}{\Psi} \\ \underset{\sim}{0} & \underset{\sim}{I} \end{pmatrix}$, $G = \begin{pmatrix} \underset{\sim}{B} \\ \Pi_2 \end{pmatrix}$ and $\delta_i = \begin{pmatrix} U_i \\ V_{2i} \end{pmatrix}$. The reduced form is obtained by premultiplying by H^{-1} which implies the restrictions $\Pi = H^{-1}G$ and $\Omega = H^{-1}E(\delta_i\delta_i')H'^{-1}$.

II. Techniques for Data Analysis

This section develops simple procedures that provide the basis for choosing the orders of the lag polynomials appearing in the above specifications and for determining whether or not it is reasonable to assume that the coefficients of these polynomials are constant over time. These procedures prevent a researcher from choosing a model specification that is broadly inconsistent with the data. Methods for choosing an appropriate specification for the lag polynomials determining distributed lag relationships (i.e., $\Gamma(L)$, $\Psi(L)$ and $B(L)$ in (2)) are a by-product of results contained in the next section on estimation and will be discussed there. This section focuses on the more complex problem of correctly specifying the form of the autoregressive and the moving average lag polynomials generating the ARMA component of the error process given by (6).

The two principle items used in the time series literature for choosing the specification of an ARMA model are the sample covariogram (or correlogram) and the sample partial correlation function.[1] To provide formal definitions of these concepts, let $U(t)$ denote a random vector which is generated by some time series process. Given a sequence of realizations of $U(t)$, the covariogram is a plot of the covariance or autocovariance between any two elements of this sequence as a function of the number of time periods between realizations. The k^{th} order matrix of autocovariances is $E(U(t)U'(t-k))$. Plotting this matrix as a function of k creates the covariogram. The (j, ℓ) element of this plot, $E(U_j(t)U_\ell(t-k))$, is called an own covariogram when $j = k$ and cross covariogram when $j \neq k$. The partial correlation function,

[1] Granger-Newbold (1977, Chs. 3 and 7) provide an extensive discussion of how to use the covariogram and the partial correlation functions to build time series models. Nelson (1973) provides a more elementary discussion.

on the other hand, is a plot of the partial correlation coefficients against the length of the lag between random variables. The k^{th} order matrix of partial correlation coefficients denoted as A_k is defined by the regression equation $U(t) = -\sum_{j=1}^{k} A_j U(t-j) + \varepsilon(t)$ where $E(\varepsilon(t)U'(t-j)) = 0$ for $j = 1,\ldots,k$.[1] Plotting A_k as a function of k produces the partial correlation function. The covariogram is particularly useful for identifying the presence of a moving average process. The partial correlation function greatly aids in identifying an autoregressive process.

This section formulates simple procedures for estimating the covariogram and the partial correlation function using panel data. The discussion develops these procedures for the analysis of the time series properties of the distrubances $U_i(t)$ appearing in the specification of the DSEM given by (2) and (8).[2] Thus, it provides information that is useful when choosing a specification for the error process given by (4) and (6). To simplify the following exposition, the data analysis procedures are formulated for the situation in which a researcher is investigating the properties of a distrubance from a single equation (i.e., g = 1 in (1)), which implies that $U_i(t)$ is a scalar. These procedures, however, immediately generalize to the multi-equation case.

---

[1] Formally, A_j cannot be interpreted as partial correlation coefficients unless the stochastic process generating U(t) is stationary.

[2] When considering a simultaneous equation specification, it is implicitly assumed that the parameters of the $\underset{\sim}{\Gamma}$, $\underset{\sim}{\Psi}$, and $\underset{\sim}{B}$ matrices in (8) are identified and can be estimated without using any covariance restrictions. Under these circumstances, it is possible to obtain consistent estimates for each $U_i(t)$ using standard two or three stage least squares procedures which neglect covariance constraints. If this assumption is violated, then it is not possible to directly analyze the time series properties of the structural disturbances, and one must apply the following data analysis techniques to the reduced form disturbances given by (17).

Procedures for Estimating the Covariogram and the Partial Correlation Function

Suppose, for the moment, that one knows the true values of the disturbances. The discussion below shows how these disturbances can be used to estimate and test hypotheses concerning the form of the covariogram and the partial correlation function using a standard seemingly unrelated regression framework.

To construct the sample covariogram, we require estimates of autocovariances for each order or length of lag. To estimate the k^{th} order autocovariance, consider the following set of regression equations

$$U_i(t)\, U_i(t-k) = \theta_{kt} + \psi_i(t) \qquad t = k+1,\ldots,T \tag{20}$$

where $U_i(t)\, U_i(t-k)$ is a dependent variable, θ_{kt} is a parameter, and $\psi_i(t)$ is an error term distributed independently across individuals. Stacking these equations for a given individual yields a seemingly unrelated regression model of the form

$$\begin{pmatrix} U_i(T)\, U_i(T-k) \\ \vdots \\ U_i(k+1)\, U_i(1) \end{pmatrix} = \theta_k + \psi_i \qquad i = 1,\ldots,N,$$

$$E(\psi_i \psi_j') = \begin{cases} C & i = j \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

where $\theta_k' = (\theta_{kT},\ldots,\theta_{k(k+1)})$ is a vector of intercepts, $\psi_i' = (\psi_i(T),\ldots,\psi_i(k+1))$ is a disturbance vector, and we have implicitly assumed that the fourth moments of $U_i(t)$ exist and are constant across individuals. The intercepts of these

equations $\theta_{kt} = E(U_i(t)\ U_i(t-k))$ represent the k^{th} order autocovariances associated with periods t, t = k+1,...,T.

Thus, estimating equations like (20) or (21) by least squares or joint generalized least squares using data on individuals provides all the information needed to construct the sample covariogram and to test hypotheses concerning its form. Unconstrained estimation yields multiple estimates of k^{th} order autocovariances (i.e., one for each period t = k+1,...,T) which reflects the fact that in a panel data analysis one can permit the parameters of time series processes to be different in each period.[1] Constrained estimation, on the other hand, of seemingly unrelated regression model given by (21) restricting the intercepts to be constant across equations (i.e., constraining the elements of θ_k to be the same) produces a unique estimate of the k^{th} order autocovariance which uses all available data. Estimating models like (21) for each k, then, with equality constraints on intercepts yields unique autocovariances for each order. Plotting these constrained estimates against k creates the sample covariogram.

Combining the system of equations given by (21) for all values of k yields a model of the form

$$(22) \qquad St(U_i U_i') = \theta + \xi_i \qquad i = 1,\ldots,N$$

$$E(\xi_i \xi_j') = \begin{cases} T & i = j \\ 0 & \text{otherwise} \end{cases}$$

---

[1] As discussed in the previous section, in a panel data analysis where asymptotics rely on a large number of individuals, it is possible to allow the autoregressive and/or the moving average lag polynomials (i.e., A(L) and M(L) in (6)) to differ in each period, or allow for heteroscedastic white noise over time. Permitting variation in parameters of this sort generally implies that the θ_{kt}'s are different for each t.

where $St(\cdot)$ denotes an operator that stacks the rows of a matrix and deletes all the elements that lie below the diagonal, $\theta = St(\Theta) = St(E(U_i U_i'))$ is a vector of intercepts, and ξ_i is an error vector that contains the $\psi_i(t)$ disturbances appearing in (20) for all values of k. This expanded seemingly unrelated regression model provides a framework in which one can simultaneously estimate or test constraints involving autocovariances of different orders. An especially interesting hypothesis in this regard is stationarity of the error process which implies that Θ is a *Topletz matrix.*[1] *This hypothesis translates into simple equality constraints* relating the elements of θ which are easily tested jointly using standard generalized least squares procedures applied to (22).

To construct the sample partial correlation function we require estimates of partial correlation coefficients for each order. To estimate the k^{th} order partial correlation coefficient, consider the following set of regression equations

$$(23) \qquad U_i(t) = \rho_{1t}\, U_i(t-1) \quad + ,\ldots, + \rho_{kt}\, U_i(t-k) \quad + e_i(t) \qquad t = k+1,\ldots,T,$$

---

[1] A matrix Q is a Topletz or a block Topletz matrix if it has the form

$$Q = \begin{pmatrix} a & b & c & d & e \\ b & a & b & c & d \\ c & b & a & b & c \\ d & c & b & a & b \\ e & d & c & b & a \end{pmatrix}.$$

If a, b, c, d, and e are constants, then Q is a Topletz matrix. If a, b, c, d, and e are matrices, then Q is a block Topletz matrix with a, b, c, d, and e as its submatrices. When $U_i(t)$ represents a univariate time series, which is the case considered here (e.g., g = 1), stationarity implies that autocovariances of a given order are constant, or equivalently that Θ is a Topletz matrix. In the multivariate case (i.e., g > 1), stationarity of $U_i(t)$ implies that Θ is a block Topletz matrix.

where the ρ_{jt}'s are parameters and $e_i(t)$ is a disturbance distributed independently across individuals. Stacking these equations for a given individual yields an equation system of the form

$$\begin{pmatrix} U_i(T) \\ \cdot \\ \cdot \\ \cdot \\ U_i(k+1) \end{pmatrix} = \sum_{j=1}^{k} \begin{pmatrix} U_i(T-j) & & 0 \\ & \cdot & \\ & & \cdot \\ 0 & & U_i(k-j+1) \end{pmatrix} \rho_j + e_i \quad i = 1,\ldots,N,$$

$$E(e_i e_j') = \begin{cases} R & i = j \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

where $\rho_j' = (\rho_{jT},\ldots,\rho_{j(k+1)})$, $j = 1,\ldots,k$, are parameter vectors, and $e_i' = (e_i(T),\ldots,e_i(k+1))$ is a disturbance vector. The parameters ρ_{kt} represent the k^{th} order partial correlation coefficients associated with periods t, $t = k+1,\ldots,T$.

Estimating the parameters of the seemingly unrelated regression model given by (24) for the different values of k, then, allows one to form the partial correlation function and to test hypotheses relating to its structure. Unconstrained estimation yields estimates of partial correlation coefficients that are period specific. Estimating the parameter vector ρ_k constraining its elements to be equal creates a unique estimate of the k^{th} order partial correlation coefficient. Graphing these constrained estimates for ρ_k against k produces the sample partial correlation function.

Using Residuals in Place of Disturbances

An apparent difficulty with the preceding discussion is that one does not have the true values of the disturbances available for data analysis. This turns out, however, not to be a problem. All of the estimation procedures and properties of estimators described above remain valid if one uses consistent estimates of the disturbances in place of the true values. Thus, one can use residuals and standard seemingly unrelated regression packages to estimate and to test hypotheses concerning the forms of the covariogram and the partial correlation function.

Verifying this proposition requires two theoretical results. Letting $\hat{U}_i$ denote the vector of residuals associated with the stacked representation of the DSEM given by (8),[1] the needed theorems are

$$\text{plim}\ \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_i U_i' - \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \hat{U}_i \hat{U}_i'\right] = 0 \tag{25}$$

and

$$\text{plim}\ \left[\frac{1}{N}\sum_{i=1}^{N} \text{vec}(U_i U_i')\ \text{vec}(U_i U_i')' - \frac{1}{N}\sum_{i=1}^{N} \text{vec}(\hat{U}_i \hat{U}_i')\ \text{vec}(\hat{U}_i \hat{U}_i')'\right] = 0 \tag{26}$$

where vec(·) denotes an operator that stacks the rows of a matrix into a column vector. Proofs of (25) and (26) are presented in Appendix A.[2]

---

[1] The residuals are defined as $\hat{U}_i = \hat{\Gamma} Y_{1i} - \hat{\Psi} Y_{2i} - \hat{B} X_i$. The estimators $\hat{\Gamma}$, $\hat{\Psi}$ and $\hat{B}$ are assumed to converge in probability at a rate so that $\hat{U}_i - U_i$ is $o_p(N^h)$ for $h < 1/2$ which is satisfied for familiar estimators (e.g., least squares, generalized least squares, two- and three-stage least squares, maximum likelihood, etc.).

[2] Similar results are proved by Hannan (1970, Ch. 7) who considers the use of least squares residuals from a time series regression to estimate autocovariances and the spectrum.

These theoretical results imply that replacing U_i by $\hat{U}_i$ in regression models (20) - (24) yields estimators with the same asymptotic properties as those computed using the true disturbances. Proof of this proposition is presented in Appendix B. The central fact used in this proof is that all unconstrained and constrained joint least squares or generalized least squares estimators of models (20) - (24) are linear functions of the matrix $\sum_{i=1}^{N} U_i U_i'$, and their asymptotic normality depends on the large sample behavior of this matrix multiplied by the normalizing constant $\frac{1}{\sqrt{N}}$. Similarly, when residuals are used in place of true disturbances, the asymptotic normality of the new estimators depends on the behavior of the matrix $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{U}_i \hat{U}_i'$. Condition (25) guarantees that the asymptotic distributions of these new estimators is the same as the estimators computed using true disturbances. Condition (26) guarantees that use of standard techniques to compute the covariance matrix of estimators based on residuals and on true disturbances yield equivalent results.

Therefore, when residuals are used to estimate either model (21), (22) or (24), all parameter estimates, standard errors, and test statistics reported by a standard seemingly unrelated regression package are asymptotically valid.[1] Constrained estimation of models like (21), (22) and (24) using

---

[1]Notice that these results do not require any special distributional assumptions other than the existence of fourth moments and the constancy of moments across individuals. More precisely, the application of the central limit theorems requires the existence of any absolute moment greater than fourth order.

residuals, then, offers a simple way to estimate the covariogram and the partial correlation function and to test hypotheses concerning their structure. It ia also possible to construct estimates of the correlogram which is another data analysis tool found in the time series literature. The correlogram is like the covariogram except that it is a plot of the autocorrelations instead of the autocovariances. It is often used instead of the covariogram in time series analysis because correlations are unit free and normalized to lie between -1 and 1, and, so, they are more easily interpreted. Using the estimated values of θ_k from equations (21) or (22), an estimate of the k^{th} order autocorrelation coefficient equals $\hat{\delta}_k = h(\hat{\sigma})$ where $\hat{\sigma}$ is a vector defined by $\hat{\sigma}' = (\hat{\theta}_k, \hat{\theta}_0)$ and h is a function defined by $h(\hat{\sigma}) = \frac{\hat{\theta}_k}{\hat{\theta}_0}$. A standard application of stochastic limit theorems implies that $\hat{\delta}_k$ is approximately normally distributed with the true value of the k^{th} order autocorrelation coefficient as its mean, and a variance given by $\frac{\partial h}{\partial \hat{\sigma}} S \frac{\partial h'}{\partial \hat{\sigma}}$ where S is the covariance matrix of the estimates contained in $\hat{\sigma}$. Thus no further estimation is required to compute estimates and standard errors for the correlogram.

Using standard seemingly unrelated regression packages and residuals, then, one can test between completing specifications of the time series process generating the structural disturbances. If, for example, the estimated covariogram is not significantly different from zero after a short lag, then a pure moving average process is implied. In this instance, if we further test to see whether autocovariances are constant across years, we can determine if the white noise error process is homoscedastic over time. By testing further to see whether autocovariances are constant across samples composed of individuals of a given characteristic, we can determine if the moving average process is the same across individuals. If, instead, the

covariogram converges to a positive constant after a short lag, then the error terms may be generated by a moving average process and a permanent component.

In a more general context, testing among various specifications of θ_k's in (21) or (22) and ρ_k's in (24) allows one to test for a pure moving average process, a pure autoregressive process, and many kinds of mixed processes. One can also test for several forms of nonstationarity and heteroscedasticity. One can distinguish between a fairly wide class of alternative specifications of the error process if one analyzes the data in first and second differences along with analyzing the data in levels. As in standard time series analysis, identifying the specifications of the θ_k's and the ρ_k's that are consistent with the data only narrows the class of models one needs to consider. Rarely does this type of data analysis identify a unique specification. Several models will often explain the data just as well. In the analysis of panel data, this is likely to be even more of a problem because there is typically available only a short time series.

III. Estimation Methods

This section presents methods for estimating time series models that are especially tailored for panel data where T is fixed and asymptotic results depend on large N = number of individuals. These estimation methods are very general; given all equations are linear in the variables, they can be applied to estimate any simultaneous equations model that involves any set of nonlinear restrictions between parameters including covariance restrictions. Two kinds of estimation procedures are considered: the first is "least squares methods" which includes generalized, and two and three stage least squares procedures; and the second is "quasi-maximum likelihood methods." For each set of procedures, we consider both limited and full information methods.

The following analysis does not present any formal identification conditions. For the standard multivariate ARMA model, Kashyap-Nasburg (1974) develop necessary and sufficient conditions for identification. Hannan (1969) presents sufficient conditions. These conditions are not easily applied in practice. Panel data introduces additional complications. The length of the time series becomes a crucial factor. The treatment of initial conditions reduces the effective length of the panel and at the same time introduces new parameters. Notice, on the other hand, that adding permanent components to a multiple time series model does not complicate the identification conditions. First differencing equations eliminates permanent components, and it does not introduce any new parameters. The standard identification criteria can be applied directly to the first-differenced specification of the model. Introducing

permanent components, then, has the effect of reducing the length of the time series by one period. Identification will be lost only in those cases in which the orders of the autoregressive and the moving average components are sufficiently high to make the length of the time series a crucial factor.

The statistical models proposed in Section I involve two sets of parameters: the first set--hereafter called the structural coefficients--consist of all those coefficients appearing in the stacked representation of the DSEM by (9) (i.e., the elements of $\underset{\sim}{\Gamma}$, $\underset{\sim}{\Psi}$ and $\underset{\sim}{B}$);[1] and the second set--hereafter called the covariance parameters--includes those parameters involved in the specification of the covariance matrix of the disturbance vector in (8), $\Theta = E(U_i U_i')$ (i.e., the coefficients of the autoregressive and moving average lag polynomials and the elements of the covariance matrices of the white noise error vectors and initial conditions).

Least Squares Methods

Three estimation procedures based on "least squares methods" are proposed below: one to estimate the set of structural coefficients, another to estimate the set of covariance parameters, and a third to estimate both sets of parameters simultaneously.

Standard procedures can be employed to estimate the set of structural coefficients. If one is analyzing a nonsimultaneous specification

---

[1] Hopefully this terminology will not result in any confusion. When analyzing a nonsimultaneous specification of the DSEM (i.e., $\underset{\sim}{\Gamma} = I$ and $\underset{\sim}{\Psi} = 0$), the set of structural coefficients obviously just includes regression coefficients.

of the DSEM (i.e., $\underset{\sim}{\Gamma} = I$ and $\underset{\sim}{\Psi} = 0$ in (8)), then joint generalized least squares procedures that permit the imposition of equality constraints across equations can be applied to (8) to estimate these coefficients (i.e., the elements of $\underset{\sim}{B}$). If, on the other hand, one is analyzing a simultaneous equation specification of the DSEM, then two or three stage least squares methods that allow for linear restrictions across equations can be applied to estimate the structural coefficients of (8).[1] Recall that the need for imposing equality constraints across equations is a direct implication of the assumption that distributed lag relationships are constant over time. All of the above procedures may be classified as limited information in the sense that they do not simultaneously estimate the parameters of the covariance matrix Θ. They require a consistent estimate of Θ, but this obviously can be constructed without directly estimating any parameters of the time series process generating disturbances. In those cases where one chooses not to introduce any assumptions regarding error processes or covariance restrictions, the above procedures yield estimates of the structural coefficients that use all available information and restrictions.

These estimation procedures provide a natural framework for performing preliminary data analysis to determine the length of the lag polynomials associated with distributed lags (i.e., $\Gamma(L)$, $\Psi(L)$ and $B(L)$ in (2)), and to test whether or not the coefficients of these polynomials are constant

[1] The asymptotic properties of estimators obtained from these procedures do not require the assumption that the disturbances in the prediction equations given by (19) are homoscedastic. Thus, to employ these simultaneous equation estimation procedures, one requires no assumptions in addition to those presented in (8).

across periods. This form of data analysis involves standard tests of linear hypotheses. Identifying the orders of lag polynomials involves tests of whether coefficients on lagged endogenous and exogenous variables are significant or not. Checking for the constancy of distributed lag relationships over time involves tests of equality of coefficients across equations. An attractive feature of the data analysis techniques is that they can be implemented in complete ignorance of the stochastic process generating error terms. An unconstrained estimate of the covariance matrix can be used in the computation of coefficient estimates and test statistics.

The results developed in the previous section offer a general method for estimating the set of covariance parameters. The seemingly unrelated regression model given by (22) is particularly well suited for estimating parameters of the covariance matrix $\Theta = E(U_i U_i')$ associated with the stacked representation of the DSEM given by (8). If one assumes that the components of U_i are generated by the combined variance component-ARMA error process given by (4) and (5), then as shown in Section I, the elements of Θ are functions of the parameters of this error process where the exact functions are implied by the relations (16) or (17). Substituting these functions for the elements of $\Theta = St(\theta)$, one can estimate the entire set of covariance parameters using a standard nonlinear joint generalized least squares procedure applied to (22). As noted in the previous section, conditions (25) and (26) imply that residuals can be used in place of the true disturbances as dependent variables in the estimation of a model like (22) without any need for adjusting the output reported by the computer package; all reported standard errors and test statistics are asymptotically valid.

This, then, provides a simple method for estimating the parameters of any covariance matrix which must satisfy an arbitrary set of nonlinear constraints. This procedure may be classified as limited information in the sense that it does not simultaneously estimate the set of structural coefficients. One requires consistent estimates of these coefficients to compute residuals, but these can be obtained from the methods described above for estimating structural coefficients. The fact that one is not required to estimate all parameters simultaneously is an attractive feature of this procedure since it means that a researcher can concentrate on correctly specifying and estimating the error structure for any given specification of DSEM.

Thus, while the above methods for estimating structural coefficients permit one to neglect specifying the precise form of the covariance structure, this covariance estimation procedure allows one to ignore the specification of the relationships between measured variables once it has been chosen. This, of course, does not mean that the results of one procedure when used as input for the other will not lead to a different set of parameter estimates. Using, for example, estimates from the second procedure to construct a consistent estimate of θ needed in the first will in general produce different estimates of the structural coefficients; and, in turn, using these estimates to form new residuals is likely to imply different estimates for the covariance parameters. In theory, however, these differences should not be statistically significant, and there should be no changes in inferences as a result.

The two limited information procedures outlined above can be combined into one that simultaneously estimates both structural coefficients and covariance parameters. Stacking the system of equations given by (8) on top of those given by (22) yields

$$\begin{pmatrix} \Gamma Y_{1i} \\ St(U_i U_i') \end{pmatrix} = \begin{pmatrix} \Psi Y_{2i} + BX_i \\ St(\Theta) \end{pmatrix} + a_i \qquad i = 1,\ldots,N$$

(27)

$$E(a_i a_j') = \begin{cases} H & i = j \\ 0 & \text{otherwise} \end{cases}$$

where $a_i' = (U_i', \xi_i')$ is a disturbance vector. Substituting $St(\hat{U}_i \hat{U}_i')$ for $St(U_i U_i')$ (i.e., residuals for disturbances), it is possible to estimate this expanded model by three stage least squares techniques that permit the imposition of nonlinear constraints on parameters. In those cases where one is not considering a simultaneous specification of the DSEM (i.e., $\Gamma = I$ and $\Psi = 0$), joint generalized least squares procedures can be applied instead to estimate (27). In Appendix B, it is shown once again that the treating of residuals as if they were the disturbances is appropriate in the sense that all the output reported by the standard estimation procedures applied to (27) is asymptotically valid.

Simultaneously estimating structural coefficients and covariance parameters yields estimates that are more efficient than those obtained from the above limited information methods. There are two sources of increased efficiency. First, in those instances where third moments of U_i are nonzero (which implies that $E(U_i \xi_i') \neq 0$ so $H = E(a_i a_i')$ is not block diagonal), the estimates based on (27) will be more efficient than those obtained from the above procedures for the same reason that generalized least squares estimates are more efficient than ordinary least squares estimates. The second source of efficiency gain arises if there are any constraints involving both structural coefficients and covariance parameters; it is possible to impose these restrictions when estimating (27).

An important assumption maintained in the above discussion of "least squares methods" is that all structural coefficients are identified without the use of any covariance restrictions. If this assumption is not true, it is obviously not possible to estimate structural coefficients using the first procedure, and without these estimates one cannot compute residuals to serve as dependent variables for the second and third procedures. In those cases where covariance restrictions are required for identification, one must work instead with the reduced form specification of the DSEM given by (18) and apply the above procedures to estimate its parameters. The third full information procedure then can be used to obtain a complete set of parameter estimates.

Quasi-Maximum Likelihood Methods

The technique usually applied to estimate models where one is interested in estimating parameters of a covariance matrix is the method of maximum likelihood. Typically, a researcher assumes that disturbances are normally distributed and, then, computes estimates by maximizing the kernel of a multivariate normal density function. Such a procedure is computationally efficient, and it can incorporate nonlinear constraints involving both regression coefficients and covariance parameters. Below we consider the application of these techniques to estimate structural and covariance parameters.

The reduced form specification of the DSEM given by (18) provides the appropriate formulation for application of maximum likelihood methods. As outlined in the discussion following (18), the DSEM implies restrictions on both the elements of the matrix of regression coefficients, Π, and the parameters of the covariance matrix Ω. Denote these restricions by

writing Π and Ω as functions of the form $\Pi(\gamma)$ and $\Omega(\omega)$ where γ and ω are vectors of parameters.[1]

The maximum likelihood estimates of these parameters are defined as those values of γ and ω that maximize the function

$$(28) \quad Q_N(\Pi(\gamma), \Omega(\omega), Y, X) = \frac{1}{N}\sum_{i=1}^{N} q_i$$

$$= \frac{1}{N}\sum_{i=1}^{N} (-\ell n\ |\Omega(\omega)| - (Y_i - \Pi(\gamma)X_i)'\Omega^{-1}(\omega)(Y_i - \Pi(\gamma)X_i)$$

$$= -\ell n\ |\Omega(\omega)| - \frac{1}{N}\sum_{i=1}^{N} (Y_i - \Pi(\gamma)X_i)'\Omega^{-1}(\omega)(Y_i - \Pi(\gamma)X_i)$$

$$= -\ell n\ |\Omega(\omega)| - tr\{\Omega^{-1}(\omega)S(\Pi(\gamma))\}$$

where

$$S(\Pi(\gamma)) = \frac{1}{N}\sum_{i=1}^{N} (Y_i - \Pi(\gamma)X_i)(Y_i - \Pi(\gamma)X_i)'$$

$$= (\frac{1}{N}\sum_{i=1}^{N} Y_iY_i') - (\frac{1}{N}\sum_{i=1}^{N} Y_iX_i')\Pi(\gamma)' - \Pi(\gamma)(\frac{1}{N}\sum_{i=1}^{N} X_iY_i')$$

$$+ \Pi(\gamma)(\frac{1}{N}\sum_{i=1}^{N} X_iX_i')\Pi(\gamma)'$$

and q_i is the function of γ, ω, Y_i and X_i defined by the second expression for Q_N.[2] Under the assumption that reduced form disturbances are normally distributed, it is well known that maximum likelihood estimates of the parameter vector $\alpha' = (\gamma', \omega_1')$ in large samples are approximately distributed according to a normal distribution of the form

---

[1] The following analysis does not rule out the possibility that γ and ω contain common elements.

[2] The reader will immediately recognize that Q_N is proportional to the kernel of a multivariate normal density function. Q_N, then, is the distance function one would use to obtain maximum likelihood estimates assuming that the reduced form disturbance vectors are identically and independently distributed according to a multivariate normal distribution.

(29) $\hat{\alpha} \overset{\cdot}{\sim} N(\alpha_0, \frac{1}{N} G_1^{-1})$

where α_0 denotes the true value of the parameters and $G_1 = \frac{\partial^2 Q_N}{\partial \alpha \partial \alpha'}\Big|_{\hat{\alpha}}$ is an estimate of the matrix of second partials which is also known as the information matrix. All maximum likelihood computer packages report standard errors and test statistics based on (29), and it is these results that many researchers use in their empirical analysis.

An apparent disadvantage of this method of estimation composed with the least squares methods described above is that it relies on specific distributional assumptions. The assumption that reduced form disturbances are normally distributed, however, is not needed in order to use the estimates produced by maximum likelihood procedures to make statistical inferences. Below we briefly describe the properties of these "quasi-maximum likelihood estimates" in absence of the normality assumption and indicate how the output reported by standard maximum likelihood computer packages must be modified to avoid specific distributional assumptions.[1] The proofs of the results summarized below are contained in MaCurdy (1980a), and the reader should consult this reference for further details.

Under fairly weak conditions, it is possible to show that the estimate of α obtained by maximizing Q_N defined by (28) is consistent and asymptotically normally distributed even if reduced form disturbances are not distributed according to a multivariate normal distribution. In particular, it can be shown that the "quasi-maximum likelihood estimate" of α in large samples approximately possesses a normal distribution of the form

[1] In the absence of the normality assumption, estimates obtained by maximizing Q_N defined by (28) are commonly called "quasi-maximum likelihood estimates" (see Malinvaud, 1966).

$$(30) \quad \hat{\alpha} \overset{\cdot}{\sim} N(\alpha_0, \frac{1}{N} G_1^{-1} G_2 G_1^{-1})$$

where $G_2 = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial q_i}{\partial \alpha}\Big|_{\hat{\alpha}} \frac{\partial q_i}{\partial \alpha'}\Big|_{\hat{\alpha}}$ is the matrix of outer partials.[1] This result depends on exactly the same assumptions as those required to prove consistency and asymptotic normality of the estimators based on least squares methods described above.

According to the above results, a researcher may make incorrect inferences if he uses output from a standard maximum likelihood computer program and disturbances are in fact not normally distributed. Many of the reported standard errors and test statistics are invalid without normality.[2] The correct asymptotic distribution is given by (30). If the reduced form disturbance vectors are distributed according to a multivariate normal distribution, then $G_2 = G_1$ in the probability limit and the covariance matrix $G_1^{-1} G_2 G_1^{-1}$ reduces to G_1^{-1} which is the one reported and used by most computer packages. Using instead $G_1^{-1} G_2 G_1^{-1}$ -- which is readily computable -- as the covariance matrix of $\hat{\alpha}$ avoids the requirement of any special distributional assumptions.

There also exists quasi-maximum likelihood techniques that can be applied to estimate subsets of parameters analogous to the limited information least squares methods proposed above. These estimation procedures are

---

[1] In addition to the assumptions implicit in the specification of the reduced form given by(18), two conditions are required to prove this result: (1) the absolute moments of the disturbance vector exists for any order greater than fourth; and (2) plim $\frac{1}{T} \sum_{i=1}^{N} X_i X_i'$ exists and is positive definite. See MaCurdy (1980a) for details.

[2] The standard errors and test statistics associated with regression coefficients remain valid without the normality assumption. The standard errors and test statistics associated with covariance parameters, however, are all invalid. See MaCurdy (1980a) for details.

particularly useful if the aim of an analysis is to estimate only covariance parameters (i.e., only the elements of Ω or ω) or only regression coefficients (i.e., only the elements of Π or γ).

To estimate only covariance parameters, it can be shown that evaluating the distance function Q_N defined by (28) at a consistent estimate of Π and maximizing the resulting function with respect to ω yields estimates of the covariance parameters that are consistent and asymptotically normally distributed. Evaluating the function Q_N at $\hat{\Pi}$, which is any consistent estimate for the true value of the regression coefficients,[1] creates a new function Q_N^* that looks like Q_N except that the matrix $S = \frac{1}{N}\sum_{i=1}^{N} V_i V_i'$ is replaced by the matrix $\frac{1}{N}\sum_{i=1}^{N} \hat{V}_i \hat{V}_i'$ where $\hat{V}_i = Y_i - \hat{\Pi} X_i$ is a vector of reduced form residuals. The function Q_N^* may be interpreted as a "likehood function" that treats residuals as if they were the true values of the disturbances. Given the same conditions assumed for each method of estimation proposed above, one can prove that the estimates of ω obtained by maximizing Q_N^* in large samples are approximately distributed according to a normal probability law of the form

$$(31) \qquad \hat{\omega} \overset{\cdot}{\sim} N\left[\omega_0, \frac{1}{N}\left(-\frac{\partial^2 Q_N^*}{\partial\omega\partial\omega'}\Big|_{\hat{\omega}}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial q_i^*}{\partial\omega}\Big|_{\hat{\omega}}\frac{\partial q_i^*}{\partial\omega'}\Big|_{\hat{\omega}}\right)\left(-\frac{\partial^2 Q_N^*}{\partial\omega\partial\omega'}\Big|_{\hat{\omega}}\right)^{-1}\right]$$

where ω_0 is the true value of ω, and q_i^* is the function q_i defined by (28) evaluated at $\hat{\Pi}$. Notice that the estimate of ω derived from this computationally simpler procedure has the same asymptotic distribution as the full information quasi-maximum likelihood estimate for ω proposed above. Thus, there is no efficiency gain in simultaneously estimating either Π and ω or γ and

[1] By "consistent" I mean that convergence in probability is $o_p(N^h)$ where $h < \frac{1}{2}$.

ω when using quasi-maximum likelihood methods. These results are important because they imply that a researcher can use maximum likelihood computer routines to compute estimates of covariance parameters using only the matrix of the outer product of residuals as input. Modifying the standard error and test statistic output reported by this routine along the lines described above avoids the requirement that disturbances are normally distributed.

A quasi-maximum likelihood procedure analogous to the one described above for estimating covariance parameters also exists for only estimating regression coefficients. Evaluating the function Q_N at a consistent estimate of Ω, rather than Π, and maximizing the resulting function with respect to γ yields estimates of the regression coefficients that are consistent and asymptotically normally distributed. This proposition is obvious once one recognizes that this is completely equivalent to a joint generalized least squares procedure. It is well known that the estimates produced by such a procedure have the same asymptotic properties as the full information estimates proposed above where one simultaneously estimates all parameters.

Quasi-maximum likelihood techniques, then, offer an attractive method of estimating the parameters of the DSEM and the error processes considered in Section I. They are not only competitive with the above least squares methods in terms of their computational efficiency, they are also as robust in the sense that they rely on the same assumptions as the least squares methods to produce consistent parameter estimates and to test hypotheses. For the nonsimultaneous specification of the DSEM, the computationally simpler procedures that condition on consistent estimates

of subsets of parameters can be used without loss of estimation efficiency. For the simultaneous specification of the model, one can apply the full information method in which one simultaneously estimates all parameters and imposes all constraints. This full information method not only allows one to estimate all structural coefficients directly, it also produces more efficient estimates if there are constraints involving both regression coefficients and elements of the reduced form covariance matrix.

IV. Summary

This paper presents specifications of a dynamic simultaneous equations model that can be applied to analyze panel data. This model allows for generally specified error structures and rational distributed lag relationships involving both endogenous and exogenous variables. One has a wider choice of specifications in the analysis of panel data than in standard time series analysis: one can permit parameters to vary freely over time in a panel data setting; permanent components can be combined with multiple time series error processes; and it is possible to relax many stationarity and homoscedasticity assumptions maintained in time series analysis. To derive explicit parameterizations for the covariance matrix associated with disturbances, this study presents a general treatment for initial conditions in a panel data framework.

For purposes of data analysis, simple procedures for estimating the covariogram and the partial correlation function are developed. These procedures use residuals as dependent variables in a seemingly unrelated regression framework. It is shown that the estimates based on residuals have the same asymptotic properties as estimates based on the true disturbances. Thus, using standard computer packages, it is possible to narrow the choice of time series models and test among competing specifications.

General nonlinear estimation procedures are formulated to estimate the full set of parameters of the dynamic simultaneous equations model and the error process. Both "least squares" and "quasi-maximum likelihood" methods of estimation are discussed. These procedures permit any form of nonlinear relationship between parameters in a simultaneous equation model, including restrictions involving both regression coefficients and parameters of the covariance matrix. Simple limited information estimators are proposed to estimate only regression coefficients or only parameters of the covariance matrix. All of these estimation methods generate estimators that are consistent and asymptotically normally distributed without any specific assumptions regarding the distribution of the disturbances.

APPENDIX A

The purpose of this appendix is to verify propositions (25) and (26).

The disturbance associated with the stacked specification of the DSEM given by (8) can be written as

$$U_i = (\underset{\sim}{\Gamma}, -\underset{\sim}{\Psi}, -\underset{\sim}{B}) \begin{pmatrix} Y_{1i} \\ Y_{2i} \\ X_i \end{pmatrix} \equiv R \; Z_i$$

where R is defined as a matrix of coefficients and Z_i is defined as the vector of observed variables for individual i. Let $\hat{U}_i = \hat{R} \; Z_i$ denote the vector of residuals for individual i where $\hat{R}$ is a consistent estimator for R so that $\hat{R} - R$ is $o_p(N^k)$ for $k < \frac{1}{2}$ (i.e., $\hat{R} - R$ is $o_p(N^k)$ if $\text{plim}(N^k(\hat{R} - R)) = 0$). Defining vec(·) as an operator that stacks the columns of a matrix into a column vector, it can be shown that vec(P C Q) = (Q' ⊗ P) vec(C). Thus,

$$\text{vec}(U_i U_i') = \text{vec}(R \, Z_i Z_i' \, R')$$

$$= (R \otimes R) \, \text{vec}(Z_i Z_i')$$

$$\equiv H \, \text{vec}(Z_i Z_i')$$

where the matrix H is defined as the Kronecker product between R and R. The analogous expression for the residuals is $\text{vec}(\hat{U}_i \hat{U}_i') = \hat{H} \, \text{vec}(Z_i Z_i')$ where $\hat{H} = (\hat{R} \otimes \hat{R})$. Since $\hat{R} - R$ is $o_p(N^k)$ for $k < \frac{1}{2}$, it is easy to verify that $\hat{H} - H$ is $o_p(N^{k+\frac{1}{2}})$.

Proposition (25) follows from the observation that

$$\text{(A.1)}\quad \operatorname{plim}\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\operatorname{vec}(\hat{U}_i\hat{U}_i') - \operatorname{vec}(U_iU_i')\right)\right\}$$

$$= \operatorname{plim}\left\{\frac{1}{\sqrt{N}}(\hat{H} - H)\left\{\sum_{i=1}^{N}\operatorname{vec}(Z_iZ_i')\right\}\right\}$$

$$= \operatorname{plim}\left\{\sqrt{N}(\hat{H} - H)\right\}\operatorname{plim}\left\{\frac{1}{N}\sum_{i=1}^{N}\operatorname{vec}(Z_iZ_i')\right\}$$

$$= 0 \cdot \operatorname{plim}\left\{\frac{1}{N}\sum_{i=1}^{N}\operatorname{vec}(Z_iZ_i')\right\} = 0$$

where the last line uses the fact that $\hat{H} - H$ is $o_p(N^{k+\frac{1}{2}})$ for $k < \frac{1}{2}$ which implies $\sqrt{N}(\hat{H} - H)$ is $o_p(N^k)$, and the assumption that fourth moments of observed variables exist which implies $\operatorname{plim}\{\frac{1}{N}\sum_{i=1}^{N} Z_iZ_i\} < \infty$.

Noting that $\operatorname{vec}(\operatorname{vec}(U_iU_i')\operatorname{vec}(U_iU_i')') = \operatorname{vec}(H\operatorname{vec}(Z_iZ_i')\operatorname{vec}(Z_iZ_i')'H')$ $= (H \otimes H)\operatorname{vec}(\operatorname{vec}(Z_iZ_i')\operatorname{vec}(Z_iZ_i')')$, proposition (26) follows from the observation that

$$\text{(A.2)}\quad \operatorname{plim}\{\frac{1}{N}\sum_{i=1}^{N}[\operatorname{vec}(\operatorname{vec}(\hat{U}_i\hat{U}_i')\ \operatorname{vec}(\hat{U}_i\hat{U}_i')') - \operatorname{vec}(\operatorname{vec}(U_iU_i')\operatorname{vec}(U_iU_i')')]\}$$

$$= \operatorname{plim}\{\frac{1}{N}[(\hat{H} \otimes \hat{H}) - (H \otimes H)]\sum_{i=1}^{N}\operatorname{vec}(\operatorname{vec}(Z_iZ_i')\operatorname{vec}(Z_iZ_i')')\}$$

$$= 0$$

where the last line uses the fact that $((\hat{H} \otimes \hat{H}) - (H \otimes H))$ is $o_p(N^{(k+\frac{3}{2})})$ and fourth moments of observed variables exist.

In Appendix B we require a generalization of (A.1). Let B_i, $i = 1,\ldots N$, denote a set of matrices satisfying the condition

$$\text{(A.3)} \quad \operatorname{plim} \left\{ \frac{1}{N} \sum_{i=1}^{N} B_i \operatorname{vec}(Z_i Z_i') \right\} < \infty.$$

We have

$$\sum_{i=1}^{N} B_i(\operatorname{vec}(\hat{U}_i \hat{U}_i') - \operatorname{vec}(U_i U_i')) = \sum_{i=1}^{N} B_i(\hat{H} - H)\operatorname{vec}(Z_i Z_i')$$

$$= \sum_{i=1}^{N} \operatorname{vec}(B_i(\hat{H} - H)\operatorname{vec}(Z_i Z_i))$$

$$= \left[\sum_{i=1}^{N} (\operatorname{vec}(Z_i Z_i')' \otimes B_i) \right] \operatorname{vec}(\hat{H} - H)$$

where the last line uses the matrix algebra theorem stated above. Thus,

$$\text{(A.4)} \quad \operatorname{plim} \left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} B_i(\operatorname{vec}(\hat{U}_i \hat{U}_i') - \operatorname{vec}(U_i U_i')) \right\}$$

$$= \operatorname{plim} \left\{ \frac{1}{N} \sum_{i=1}^{N} (\operatorname{vec}(Z_i Z_i')' \otimes B_i) \right\} \cdot \operatorname{plim} \left\{ \sqrt{N} \operatorname{vec}(\hat{H} - H) \right\}$$

$$= 0$$

where the last line follows from (A.3) and the fact that $(\hat{H} - H)$ is $O_p(N^{k+\frac{1}{2}})$.

APPENDIX B

The purpose of this appendix is to prove that when residuals are used in place of true disturbances to estimate models (22), (24), or (27), the output reported by a standard application of estimation procedures based on "least squares methods" is asymptotically valid. To simplify the exposition, proofs are only given for the case in which there are no constraints on parameters. Proofs of the following propositions when constraints are present (including nonlinear constraints) involve more algebra, but they are conceptually equivalent to those presented below.

Consider first the use of residuals in estimating model (22) and the parameters of the covariance matrix. Since the seemingly unrelated regression models proposed for estimating the covariogram given by (21) are nested in (22), the following results apply to these models as well. Replacing disturbances by residuals in (22) yields

$$\text{(B.1)} \quad St(\hat{U}_i\hat{U}_i') = \theta + \xi_i + (St(\hat{U}_i\hat{U}_i') - St(U_iU_i'))$$

$$E(\xi_i\xi_j') = \begin{cases} T & i = j \\ 0 & i \neq j \end{cases}$$

where the ξ_i's are independently distributed error vectors.

Estimating (B.1) by a joint generalized least squares computer program yields an estimate of θ equal to

$$\text{(B.2)} \quad \hat{\theta}_{GLS} = \hat{\theta}_{LS} = \frac{1}{N}\sum_{i=1}^{N} St(\hat{U}_i\hat{U}_i')$$

where we have used the fact that the generalized and the ordinary least squares estimators are equivalent since all equations contain the same exogenous variables.

This procedure prints standard errors assuming the covariance matrix of $\hat{\theta}_{GLS}$ is

$$(B.3) \quad V(\hat{\theta}_{GLS}) = \frac{1}{N}\hat{T} = \frac{1}{N^2}\sum_{i=1}^{N}(St(\hat{U}_i\hat{U}_i') - \hat{\theta}_{LS})(St(\hat{U}_i\hat{U}_i') - \hat{\theta}_{LS})';$$

and, it reports test statistics assuming that $\hat{\theta}_{GLS}$ is approximately normally distributed with mean θ and covariance matrix $V(\hat{\theta}_{GLS})$ or, equivalently,

$$(B.4) \quad \hat{\theta}_{GLS} \overset{\cdot}{\sim} N(\theta, V(\hat{\theta}_{GLS})).$$

Using (A.1), (A.2) and (B.1), standard applications of asymptotic theory yields

$$\text{plim}\{\hat{T}\} = T$$

and

$$\begin{aligned}\text{dlim}\left\{\sqrt{N}\,(\hat{\theta}_{GLS} - \theta)\right\} &= \text{dlim}\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(St(\hat{U}_i\hat{U}_i') - St(\hat{U}_i\hat{U}_i'))\right\} \\ &= \text{dlim}\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i\right\} + \text{plim}\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(St(\hat{U}_i\hat{U}_i') - St(U_iU_i)')\right\} \\ &= N(0, T)\end{aligned}$$

where dlim denotes convergence in distribution as opposed to plim which is convergence in probability. So, in large samples, we have

$$\hat{\theta}_{GLS} \overset{\cdot}{\sim} N(\theta, \frac{1}{N}\hat{T}).$$

This result verifies that the estimator given by (B,2) is consistent and the reported output of the generalized least squares procedure given by (B,3) and (B.4) is valid asymptotically.

Consider next the use of residuals to estimate the seemingly unrelated regression model given by (24) proposed for estimating the partial correlation function. A convenient representation of (24) is

$$(B.5) \quad (I \otimes U_i')J = (I \otimes U_i')Q\rho + e_i \qquad i = 1,\ldots,N$$

$$E(e_i e_j') = \begin{cases} R & i = j \\ 0 & i \neq j \end{cases}$$

where I is the identity matrix, J is a vector and Q is a matrix of known constants, ρ is a vector of parameters, and e_i is an independently distributed error vector. With the appropriate choice of J and Q and dimensioning of I, ρ contains the partial correlation coefficients for a prespecified order. Suppose, for example, we are interested in estimating the k^{th} order partial correlation coefficient. To do this, we set the dimension of I equal to (T-k); define J so that $(I_{(T-k)} \otimes U_i')J = (U_i(T),\ldots,U_i(k+1))'$; and, treat Q as a block matrix of the form $Q = dia(Q_1,\ldots,Q_{T-k})$ where the T x k matrices Q_j, j = 1,...,T-k, are defined so that $U_i'Q_j = (U_i(T-j),\ldots,U_i(T-j-k+1))$. The implied parameter vector for this specification is $\rho' = (\rho_T',\ldots,\rho_{k+1}')$ where $\rho_t' = (\rho_{1t},\ldots,\rho_{kt})$, t = k+1,...,T, with ρ_{kt} being interpreted as the k^{th} order partial correlation coefficients associated with the period t. As discussed in the text, constraining ρ_{kt} to be constant for all t produces a unique estimate for the k^{th} order partial correlation coefficient, but such constraints are not explicitly considered here.

Using residuals to estimate model (B.5), the estimator for R and the generalized least squares estimator for ρ are

$$\hat{R} = \frac{1}{N} \sum_{i=1}^{N} (I \otimes \hat{U}_i)(J + Q\hat{\rho}_{LS})(J + Q\hat{\rho}_{LS})'(I \otimes \hat{U}_i')$$

$$\hat{\rho}_{GLS} = \left[Q' \left(\sum_{i=1}^{N} (I \otimes \hat{U}_i)\hat{R}^{-1}(I \otimes \hat{U}_i') \right) Q \right]^{-1} Q' \left(\sum_{i=1}^{N} (I \otimes \hat{U}_i)\hat{R}^{-1}(I \otimes \hat{U}_i') \right) J$$

where

$$\hat{\rho}_{LS} = \left[Q' \left(I \otimes \sum_{i=1}^{N} \hat{U}_i \hat{U}_i' \right) Q \right]^{-1} Q' \left(I \otimes \sum_{i=1}^{N} \hat{U}_i \hat{U}_i' \right) J.$$

Using (A.1), we see that $\hat{\rho}_{LS}$ is consistent for ρ. The properties of $\hat{R}$ and $\hat{\rho}_{GLS}$ depend on matrices of the form $\sum_{i=1}^{N} (I \otimes \hat{U}_i)C(I \otimes \hat{U}_i')$ where $\hat{C}$ is a consistent estimate of some positive definite matrix C. The (g, h) block of this quadratic form is $\hat{C}_{gh} \sum_{i=1}^{N} \hat{U}_i \hat{U}_i'$ where $\hat{C}_{gh}$ is the (g, h) element of $\hat{C}$. Since $\text{plim}(\hat{C}_{gh}) = C_{gh}$, we know from (A.1) that this (g, h) block divided by N or $\sqrt{N}$ has the same asymptotic properties as the (g, h) block of $\sum_{i=1}^{N} (I \otimes U_i)C(I \otimes U_i')$ divided by the same normalizing factor. It directly follows, then, that the quantities of $\hat{R}$, $\hat{\rho}_{GLS}$, $\hat{\rho}_{LS}$, and $Q' \left[\sum_{i=1}^{N} (I \otimes \hat{U}_i)\hat{R}^{-1} I \otimes \hat{U}_i' \right] Q$ have the same asymptotic properties as the analogous quantities computed using the true values of the disturbances. Thus, standard applications of asymptotic theory yield the conclusion

$$\text{(B.6)} \quad \hat{\rho}_{GLS} \overset{\cdot}{\sim} N\left(\rho, \left[Q' \left(\sum_{i=1}^{N} (I \otimes \hat{U}_i)\hat{R}^{-1}(I \otimes \hat{U}_i') \right) Q \right]^{-1} \right).$$

The result given by (B.6) is exactly the one assumed by a joint generalized least squares computer package when it reports output.

Finally, consider the use of residuals in estimating model (27) which includes both structural coefficients and covariance parameters. Replacing disturbances by residuals in (27), one may write this system of equations as

$$(B.7)\qquad \begin{pmatrix} \underline{Y}_{1i} \\ St(\hat{U}_i\hat{U}_i') \end{pmatrix} = \begin{pmatrix} Z_i & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} \gamma \\ \theta \end{pmatrix} + a_i + \begin{pmatrix} 0 \\ (St(\hat{U}_i\hat{U}_i') - St(U_iU_i')) \end{pmatrix} \quad i = 1,\ldots,N,$$

$$E(a_i a_j') = \begin{cases} H & i = j \\ 0 & i \neq j \end{cases}$$

where $\underline{Y}_{1i}' = (Y_{1i}'(T),\ldots,Y_{1i}'(1))$, γ is a parameter vector, Z_i is a matrix containing both endogenous and exogenous variables (i.e., the elements of Y_{1i}, Y_{2i}, and X_i), and the disturbance vectors a_i are independently distributed.

Defining $\alpha_i' = (\underline{Y}_i', St(\hat{U}_i\hat{U}_i')')$, $W_i = \begin{pmatrix} Z_i & 0 \\ 0 & I \end{pmatrix}$, and $\delta' = (\gamma', \theta')$,

a three stage least squares procedure applied to (B.7) yields an estimate of δ equal to

$$\hat{\delta}_{3LS} = \left(\sum_{i=1}^{N} \hat{W}_i' \hat{H}^{-1} \hat{W}_i\right)^{-1} \sum_{i=1}^{N} \hat{W}_i' \hat{H}^{-1} \alpha_i$$

where $\hat{W}_i$ is the matrix W_i with all endogenous variables replaced by their predicted values,

$$\hat{H} = \frac{1}{N}\sum_{i=1}^{N} (\alpha_i - W_i\hat{\delta}_{2LS})(\alpha_i - W_i\hat{\delta}_{2LS})',$$

and

$$\hat{\delta}_{2LS} = \left(\sum_{i=1}^{N} \hat{W}_i'\hat{W}_i\right)^{-1} \sum_{i=1}^{N} \hat{W}_i'\alpha_i.$$

This procedure prints standard errors and test statistics assuming that

$$(B.8)\qquad \hat{\delta}_{3LS} \dot{\sim} N\left(\delta, \left(\sum_{i=1}^{N} \hat{W}_i' \hat{H}^{-1} \hat{W}_i\right)^{-1}\right).$$

To determine the asymptotic properties of $\hat{\delta}_{2LS}$ and $\hat{\delta}_{3LS}$, observe that using (B.7)

$$(\hat{\delta}_{2LS} - \delta) = M_1^{-1} \sum_{i=1}^{N} \hat{W}_i' a_i + M_1^{-1} \sum_{i=1}^{N} B_{1i}(St(\hat{U}_i\hat{U}_i) - St(U_iU_i'))$$

and

$$(\hat{\delta}_{3LS} - \delta) = M_2^{-1} \sum_{i=1}^{N} \hat{W}_i'\hat{H}^{-1} a_i + M_2^{-1} \sum_{i=1}^{N} B_{2i}(St(\hat{U}_i\hat{U}_i') - St(U_iU_i'))$$

where $M_1 = \sum_{i=1}^{N} \hat{W}_i'\hat{W}_i$, $M_2 = \sum_{i=1}^{N} \hat{W}_i'\hat{H}^{-1}\hat{W}_i$, $B_{1i} = \hat{W}_i' \begin{pmatrix} 0 \\ I \end{pmatrix}$, and $B_{2i} = \hat{W}_i'\hat{H}^{-1} \begin{pmatrix} 0 \\ I \end{pmatrix}$. Assuming the existence of fourth moments, it can be shown that the matrices B_{1i} satisfy condition (A.3) of Appendix A and $0 < \text{plim}\{\frac{1}{N} M_1\} < \infty$. Thus, $\text{plim}\{\hat{\delta}_{2LS}\} = \delta$ and $\text{plim}\{\hat{H}\} = H$. Using standard stochastic limit theorems (e.g., Theorem 2 of Mann-Wald), it can be further shown that the matrices B_{2i} also satisfy condition (A.3) and that $P = \text{plim}\{\frac{1}{N} M_2\}$ exists and is positive definite. Therefore,

$$\text{dlim}\{\sqrt{N}(\hat{\delta}_{3LS} - \delta)\}$$

$$(B.9) \qquad = P^{-1} \text{ dlim} \left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{W}_i'\hat{H}^{-1} a_i \right\} + P^{-1} \text{ plim} \left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} B_{2i}(St(\hat{U}_i\hat{U}_i') - St(U_iU_i')) \right\}$$

$$= P^{-1} \text{ dlim} \left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{W}_i'H^{-1} a_i \right\}$$

$$= N(0, P^{-1})$$

where the last lines use asymptotic results typically applied in deriving the properties of three stage least squares estimators. Since the matrix

$\frac{1}{N}\sum_{i=1}^{N} \hat{W}'_i \hat{H}^{-1} \hat{W}_i$ is a consistent estimate for P, (B.9) implies that (B.8)is valid as an approximation in large samples; and, so, the output reported by standard simultaneous equation procedures is valid asymptotically.

REFERENCES

Anderson, T. W. The Statistical Analysis of Time Series. New York: John Wiley & Sons, 1971.

Anderson, T. W. and Hsiao, C. "Formulation and Estimation of Dynamic Models Using Panel Data." Journal of Econometrics, forthcoming, 1981.

Ashenfelter, O. "Estimating the Effects of Training Programs on Earnings." Review of Economics and Statistics 40, no. 1 (February 1978): 47-57.

Balestra, P. and Nerlove, M. "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model." Econometrica 34 (July 1966): 585-612.

David, M. "Lifetime Income Variability and Income Profiles." Proceedings of the Annual Meeting of the American Statistical Association, August, 1971, pp. 285-92.

Friedman, M. and Kuznets, S. Income from Independent Professional Practice. New York: National Bureau of Economic Research, 1945.

Granger, C.W.J. and Newbold, P. Forecasting Economic Time Series. New York: Academic Press, 1977.

Griliches, Z. "Distributed Lags: A Survey." Econometrica (1967): 16-49.

Hannan, E. Multiple Time Series. New York: John Wiley & Sons, 1970.

_____. "The Identification of Vector Auto-regressive-Moving Average Systems." Biometrika (1969): 223-25.

Hause, J. "The Covariance Structure of Earnings and the On-the-Job Training Hypothesis." Annals of Economic and Social Measurement (Fall 1977): 335-66.

Hussian, A. and Wallace, T. "The Use of Error Components in Combining Cross Section with Time Series Data." Econometrica 37 (1969): 52-72.

Kashyap, . and Nasburg, . "Parameter Estimation in Multivariate Stochastic Difference Equations." IEEE Transactions on Automatic Control AC-19 (1974): 784-97.

Lillard, L. and Weiss, Y. "Components of Variation in Panel Earnings Data: American Scientists 1960-1970." Econometrica (1979): 437-54.

Lillard, L. and Willis, R. "Dynamic Aspects of Earnings Mobility." Econometrica (1978): 985-1012.