Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers | Cowles Foundation

1-1-1972

A Comparison of Alternative Estimators for a Dynamic Relationship Estimated from a Time Series of Cross-Sections When the Disturbances Are Small

Jon K. Peck

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

Part of the Economics Commons

Recommended Citation

Peck, Jon K., "A Comparison of Alternative Estimators for a Dynamic Relationship Estimated from a Time Series of Cross-Sections When the Disturbances Are Small" (1972). *Cowles Foundation Discussion Papers*. 558.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/558

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 325

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

A COMPARISON OF ALTERNATIVE ESTIMATORS FOR A DYNAMIC RELATIONSHIP ESTIMATED FROM A TIME SERIES OF CROSS-SECTIONS WHEN THE DISTURBANCES ARE SMALL

Jon K. Peck

January 7, 1972

SECTION 1

INTRODUCTION*

Traditionally economists have relied on time series data in the estimation and testing of economic models. Such series are usually short and are frequently highly collinear. This can lead to imprecisely estimated models and render futile the testing process. In response, the econometrician can abandon his attempts to decide subtle questions, or he can resort to sources of empirical evidence other than time series data.

Cross-section data provide alternative sources of information since they typically contain more observations than time series and are less likely to be collinear. Many economic models, however, specify dynamic behavior over time as a part of the model, see, for example, Griliches [5], and Nerlove [13], and, therefore, cross-section data by themselves are inadequate. Perhaps partly in response to this difficulty, data consisting of time-series of cross-sections are becoming more common so that some past history is available for each unit of observation in

---

*I would like to thank Professors David Grether, Joseph Kadane, and Marc Nerlove for their generous assistance and encouragement. Portions of this research were completed while I was a research fellow at the Brookings Institution. Support was also received from the National Science Foundation and from the Ford Foundation. Mrs. Glena Ames did a superb job of typing.

the cross-section. The Current Population Survey sample of the U.S. Bureau of the Census [23] and the New Jersey Negative Income Tax Experiment data [17] are typical examples. Time series of cross-sections can also be appropriate in models which are not dynamic as a means of increasing the number of degrees of freedom.

The additional complexity of this type of data requires care in the specification of the model under investigation, however, especially the assumptions made about the nature of the random disturbances in the model.

This paper analytically investigates several estimators of a single equation model containing the lagged value of the dependent variable as an explanatory variable. Suppose

$$y_{it} = y_{it-1}\alpha + x_{it}\beta + \sigma u_{it} , \quad i = 1, \ldots, N ; \quad t = 1, \ldots, T \quad (1\text{-}1)$$

where y_{it} is the observation on unit i at time t ; y_{it-1} is the observation on unit i at time $t-1$; x_{it} is a $1xk$ vector of exogenous variables, and u_{it} is a random, unobserved, normally distributed disturbance term with zero mean and variance $Eu_{it}^2 = 1$. σ is a constant used to achieve the variance normalization of u_{it} ; α is a fixed scalar, and β is a $kx1$ constant vector of coefficients. There are N observations in each cross-section and there are T cross-sections. Typically N is much larger than T . T is assumed to be at least 2.

Some account must be taken of the structure of the sample in considering the effects of omitted explanatory variables and measurement errors, which constitute the error term in the equation. It is reasonable to assume that some omitted factors are associated with the unit of observation and persist over time while other factors are associated with a time period and affect all of the units of observation to some degree. These omitted effects may be regarded as unknown constants or as random effects whose joint distribution is only incompletely known. The first approach leads to the addition of dummy variables to equation (1), one variable for each group of omitted effects; the second leads to a variance-components model for the errors in which the disturbance term u_{it} is assumed to be the sum of several independent random variables each of which represents a set of omitted effects. Thus

$$u_{it} = \mu_i + \tau_t + \nu_{it} \tag{1-2}$$

where μ_i, τ_t, and ν_{it} are independent random variables each with mean zero and with variances σ_μ^2, σ_τ^2, and σ_ν^2, respectively. If the model-builder's interest is primarily in identifying, say, the individual effects associated with each unit over time, then a dummy variable approach might be appropriate but if his interest is, rather, in the entire population of units from which his sample is drawn, then the variance-components approach might be preferable (see Scheffé [19, 20]). For predicting y_{it} for a time beyond the sample period, dummy variables are appealing, but for predicting y_{it} for a unit not in the original sample, this method clearly cannot be used while the variance-components method is applicable.

The question of which specification to use might be unimportant if both specifications led to similar estimates of the interesting coefficients, but this is not necessarily the case. Balestra and Nerlove [4] in estimating a demand function and Nerlove [14, 15] in Monte Carlo studies found that the choice of the specification made a significant difference in the estimated coefficients.

The estimation problem in the variance-components model when the relationship is dynamic is more difficult than for the classical regression model. Ordinary Least Squares applied to equation (1) no longer has the optimal properties it has in the classical case. If y_{it-1} were not a regressor, Ordinary Least Squares would be consistent but not efficient (see Theil [21]). If y_{it-1} appeared in the regression but the covariance matrix of the disturbances were an identity matrix, Ordinary Least Squares would be biased but consistent and efficient (see Hurwicz [6], and Johnston [7]). However, if both complications occur together, they reinforce each other and Ordinary Least Squares is inconsistent (see Johnston [7]). This is not surprising since an important assumption in the use of Ordinary Least Squares is that the regressions in equation (1) are uncorrelated with the error terms in the same period and y_{it-1} contains u_{it-1} which will be correlated with u_{it} leading to correlation between y_{it-1} and u_{it}. The finite-sample properties of Ordinary Least Squares in this case have not previously been analyzed analytically in the literature, but it is clear from the large-sample asymptotic properties and Monte Carlo studies that this estimator is unsatisfactory, and alternatives must be found.

The estimators which have been considered in the literature as alternatives to Ordinary Least Squares in estimating equation (1) and similar specifications are of four types: (1) Generalized Least Squares estimates using a known or consistently estimated covariance matrix for the errors; (2) Maximum Likelihood procedures or procedures which are intended to be close approximations to Maximum Likelihood; (3) Instrumental Variables estimation; and (4) estimators which knowingly misspecify the mixed-effects model of equation (1) as a fixed-effects model, say, with dummy variables and apply an appropriate estimation technique. Combinations of these procedures such as using methods of the third or fourth types to estimate consistently the covariance matrix and applying a method of the first type using this estimate have also been considered.

Maddala [11] considered Generalized Least Squares (GLS) with a consistently estimated covariance matrix for an equation with a lagged dependent variable and found it asymptotically less efficient than if the covariance matrix were known, regardless of the way the covariance matrix is estimated. He shows that for equation (1), iterating the estimation process between estimates of the covariance matrix and estimates of the parameters, which produces new estimates of the covariance matrix, is not equivalent to the Maximum Likelihood method. Amemiya and Fuller [2] also considered this problem. Amemiya [1] found the large-sample asymptotic distribution for GLS with a consistently estimated covariance matrix, for the Maximum Likelihood estimator (ML) and a misspecification model of the fourth type listed above. All three are found to be asymptotically equivalent as N and T increase to infinity.

Wallace and Hussain [24] consider the estimation problem when the lagged dependent variable is absent and the covariance matrix is estimated with OLS using a variance-components model and using a fixed-effects model. They find that the two methods are asymptotically equivalent. This result is of doubtful use, however, when y_{it-1} occurs in the equation, since the addition of this complication may drastically affect the properties of the estimators.

The results in this literature have two shortcomings. First, they fail to find distinctions among estimators which are quite different in approach. This suggests that the asymptotic properties of these estimators provide only a gross guide to the behavior of the estimators in samples of realistic size; therefore, important differences may remain undetected. A second difficulty is with the very notion of large-sample asymptotic properties for this problem. The sample can get infinitely large by increases in either the cross-section size, N, or the time-series size, T, or some combination of both. The properties of the estimators might be thought to depend on the ratio of N to T in finite samples, but these differences are obscured by the large-sample asymptotic approach.

Since the task of determining the exact finite-sample properties of the estimators used in this problem is intractable, one must proceed either by Monte Carlo methods, or by finding some other way of approximating the distribution of these estimators which will not obscure their differences.

Nerlove has investigated by Monte Carlo methods the properties of various estimators of equation (1) using the disturbance specification (2) with the time effect, τ, omitted. In [14] the estimation problem is considered for the case where equation (1) contains no exogenous variables.

The second paper [15] investigates the case where one exogenous variable is present but also allows for a constant term. The sample size was fixed at $N = 25$ and $T = 10$. For the two-component model, the covariance matrix of the disturbances depends only on one parameter, $\rho = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\nu^2)$ called the intraclass correlation coefficient. Six estimators of the model are considered: (a) Generalized Least Squares (GLS) using the true value of ρ. (b) Ordinary Least Squares (OLS). (c) Ordinary Least Squares applied to equation (1) but with the addition of a separate dummy variable for each observation in the cross section (LSC). This technique assumes a fixed-effects model and is, therefore, a type of specification error. (d) Instrumental Variables (IV) applied to equation (1) using the only available instrument, the lagged values of the exogenous variable. (e) "Two-round" estimates (2RC) found by using an estimate of ρ obtained from LSC in the GLS estimator. (f) "Two-round" estimates (2RI) using as an estimator of ρ, the value obtained from IV estimation. (g) Maximum Likelihood (ML) estimation of the relationship (1) assuming that the initial observations y_{io} are fixed. Each of these estimators was studied for 120 parameter combinations, but the same exogenous variables were used throughout. That study concludes that the 2RC estimator is superior.

> Both in terms of relative bias and mean square error, over a wide range of parameter values, the two-round procedure, using a value of $\hat{\rho}$ estimated from first-round regressions including individual constant terms, compares favorably with all the other estimation techniques investigated....[1]

---

[1] Nerlove [15], p. 381.

There are, however, many difficulties with Monte Carlo studies. Many possible experiments cannot be performed, and generalizations are hazardous without theoretical results as a guide to what may safely be inferred. One cannot even be sure that the crucial parameters of the problem have been varied. Furthermore, the results of Monte Carlo studies are often difficult to summarize in a concise way.

The approach adopted in this study is to determine analytically the approximate bias and mean squared error of several of the estimators which have been discussed where the approximation is in terms of the disturbance variance, σ^2, rather than in terms of the sample size, which is the customary approximation. The focus is placed on the mean squared errors of the estimators since a quadratic loss function in $\begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{pmatrix}$, where $\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}$ is the estimated vector of coefficients, leads to the choice of the mean squared error as the appropriate criterion to use in deciding between two estimators (see Lindgren [10]). The bias calculations may also be of some interest.

The approximation technique used is due to Chernoff and Kadane and has previously been used by Kadane to investigate the properties of k-class estimators of a simultaneous equations system in [8] and [9].

The method is to express the error of the estimator $\begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{pmatrix}$ as an infinite series in powers of σ, the standard deviation of the disturbance. From this expression the squared error of the estimator can be found as a series in σ. Then expectations are taken term-by term in these expansions until sufficient accuracy has been achieved. As the standard error of the disturbance term in the model becomes small, the accuracy of the approximation improves. As σ becomes small, the fit of the model improves and the regression function

$$y_{it} = y_{it-1}\alpha + x_{it}\beta$$

becomes a better description of the data. Of course, this limiting process cannot actually be occurring, unlike the approximations in terms of the sample size where the sample could be envisioned as actually increasing in size, but this does not prevent the use of the approximation. How useful the approximate moments of the estimators found with this technique will be for a particular model will depend to some extent on how accurate the approximations are for that model, but for comparing alternative estimators for a model, it is only necessary that the approximations rank the estimators correctly, which may occur even if the absolute errors of the small-sigma asymptotic moments are large.

Ideally, one would otbain the small-sigma asymptotic moments for all of the estimators discussed above, but even these asymptotic moments are difficult to find; therefore, several of the estimators have not been analyzed. The asymptotic moments are found for the following estimators, GLS, OLS, LSC, IV, and OLS applied to a misspecified model. While these results are derived for the two error components model, they hold for an arbitrary nonsingular covariance matrix for the errors when certain expectation matrices are recalculated for other covariances. In particular, they hold for spherical errors and, therefore, as a special case the approximate bias of OLS due to the presence of a lagged dependent variable in the regression is found. It is shown that the simple conclusion of Johnston that the bias in $\hat{\alpha}$ is negative when α is positive for an equation with no exogenous variables no longer holds when an exogenous regressor is introduced. The sign and magnitude of the bias will depend on the statistical properties of the exogenous variables.

The model and set of assumptions to be used throughout this study are as follows. The equation to be estimated is

$$y_{it} = y_{it-1}\alpha + x_{it}\beta + \sigma u_{it} \qquad (1\text{-}1)$$

with error specification

$$u_{it} = \mu_i + \nu_{it} \qquad (1\text{-}2)'$$

where μ_i and ν_{it} are independent, normally distributed random errors with mean zero and constant variances σ_μ^2 and σ_ν^2 respectively. Specifically,

$$E\mu_i\mu_j = E\nu_{is}\nu_{it} = E\nu_{it}\nu_{jt} = 0 \quad \text{for} \quad i \neq j\ , \quad s \neq t\ .$$

Also $E\mu_i\nu_{jt} = 0$ for all i, j, t . σ is chosen so that $Var(u_{it}) = 1$. Arrange the sample observations first by individual and then by time period and define the vectors

$$y = (y_{11}, y_{12}, \ldots, y_{1T}, y_{21}, \ldots, y_{2T}, \ldots, y_{N1}, \ldots, y_{NT})'$$

$$y_{-1} = (y_{10}, y_{11}, \ldots, y_{1T-1}, \ldots, y_{N0}, \ldots, y_{NT-1})' \quad \text{and}$$

$$u = (u_{11}, u_{12}, \ldots, u_{1T}, \ldots, u_{N1}, \ldots, u_{NT})'$$

$$x_j = (x_{11j}, x_{12j}, \ldots, x_{1Tj}, \ldots, x_{N1j}, \ldots, x_{NTj})' \qquad j = 1, \ldots, k$$

x_j is the vector of observation on the j^{th} exogenous variable. Define the matrix of exogenous variables

$$X = (x_1 \; x_2 \ldots x_k) .$$

Using this notation equation (1) can be written as

$$y = y_{-1}\alpha + X\beta + \sigma u . \qquad (1\text{-}3)$$

The covariance matrix of u is

$$Euu' \equiv \Omega = I_N \otimes \mathcal{A} \qquad (1\text{-}4)$$

where I_N is the NxN identity matrix, $\otimes$, is the Kroneker product operator and $\mathcal{A}$ is the TxT matrix

$$\begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & & & & \\ \rho & \rho & \cdots & \rho & \rho \end{bmatrix}$$

$\rho = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\nu^2)$ is the intraclass correlation coefficient of the disturbance terms. To see (4) observe that

$$Eu_{it}u_{jt} = 0 \; , \quad i \neq j \quad \text{and}$$

$$Eu_{is}u_{it} = E(\mu_i + \nu_{is})(\mu_i + \nu_{it}) = \begin{cases} \sigma_\mu^2 \; , & s \neq t \\ \sigma_\mu^2 + \sigma_\nu^2 \; , & s = t \end{cases} \tag{1-5}$$

The NTxk matrix X is assumed to be nonstochastic and of full rank, k and to contain at least one nonconstant variable. Some assumption must be made about the initial observations on the dependent variable y_{io} , $i = 1, \ldots, N$. Perhaps the most natural assumption would be that they are stochastic and drawn from the same distribution as the other observations y_{it} . But this is a conditional distribution dependent on the values of the regressor variables in periods prior to those included in the sample which are unknown by assumption. A distribution could be assumed for the exogenous variables which could be used to obtain the marginal distribution of the y_{io}'s , but this would not be in the spirit of the model and would violate the assumption that x is nonstochastic.[2] Therefore, the assumption is made that the initial observations on the dependent variable are nonstochastic. The results found below, then, may be taken as conditional on the initial set of observations on y . T is assumed to be at least two since if T equals one the sample contains no information on the dynamic relationship. Denote the set of assumptions just described as θ . They are summarized in Table I.

---

[2] If some prior observations on the regressors were available, the unconditional distribution of the y_{io} might be approximated from them, at least if α is small.

TABLE I

Assumptions (θ)

$y = y_{-1}\alpha + X\beta + \sigma u$.

X is nonstochastic, of full rank, k , and contains at least one non-constant variable.

$E(u|X) = 0$ for all X .

$u \sim n(0,\Omega)$ where $n(0,\Omega)$ is an NT-dimensional normal distribution with mean vector 0 and nonsingular covariance matrix $\Omega = I_N \otimes \mathcal{Q}$.

The initial set of observations y_{io} , $i = 1, \ldots, N$ is nonstochastic.

In Section 2 the GLS estimator is analyzed where the covariance matrix is known. As corollaries to that analysis the bias and mean squared error for OLS are found. Section 3 discussed the problem of specification error and analyzes the LSC estimator, and Section 4 treats the IV estimator. The results are summarized in Section 5.

SECTION 2

In this section the properties of Generalized Least Squares (GLS) are analyzed where the covariance matrix Ω is assumed to be known. Of course, Ω is rarely known and must in practice be estimated, but GLS is still of interest in itself. If it were not for the presence of y_{-1} in the model, GLS would be BLUE (see Theil [21]). With y_{-1} present, however, GLS will be biased and may lack other desirable properties (see Johnston [7] and Hurwicz [6]).

The analysis is carried out using an arbitrary weighting matrix Σ^{-1} in the weighted least-squares estimator. When $\Sigma^{-1} = \Omega^{-1}$ the estimator is GLS and when $\Sigma^{-1} = I$, the estimator reduces to OLS. These derivations are valid for an arbitrary nonsingular covariance matrix Ω with the modifications described below.

In deriving the small-sigma asymptotic properties of estimators of this model it is useful to distinguish between the model expressed in equation (1-3), which includes y_{-1}, and the "reduced form" of the model in which y_{-1} is substituted out of the equation. This is given by

<u>Lemma 1</u>. The reduced form of equation (1-3) is

$$\begin{aligned} y &= (W + \sigma V)\alpha + X\beta + \sigma u \\ &= (Z + \sigma V^*)\begin{pmatrix}\alpha \\ \beta\end{pmatrix} + \sigma u \end{aligned} \qquad (2\text{-}1)$$

where W and V are $NTx1$ column vectors

$$W = (w_{10}, \ldots, w_{1T-1}, \ldots, w_{NO}, \ldots, w_{NT-1})'$$

and

$$V = (v_{10}, \ldots, v_{1T-1}, \ldots, v_{NO}, \ldots, v_{NT-1})'$$

with elements

$$w_{it} = \begin{cases} \sum_{j=1}^{t} \alpha^{t-j} x_{ij}\beta + \alpha^t y_{io} , & i = 1, \ldots, N ; \quad t = 1, \ldots, T-1 \\ y_{io} , & i = 1, \ldots, N ; \quad t = 0 \end{cases}$$

and

$$v_{it} = \begin{cases} \sum_{j=1}^{t} \alpha^{t-j} u_{ij} , & i = 1, \ldots, N ; \quad t = 1, \ldots, T-1 \\ 0 , & i = 1, \ldots, N ; \quad t = 0 . \end{cases}$$

Z and V^* are $NT \times k+1$ matrices

$$Z = (\underset{NTx1}{W} \quad \underset{NTxk}{X}) \text{ and } V^* = (\underset{NTx1}{V} \quad \underset{NTxk}{0}) .$$

Z is the nonstochastic part of the regressors and V^* the stochastic part. It should be noted that W and V correspond to the fixed and random parts, respectively, of the lagged dependent variable; not to y itself. The notation is summarized in Table II.

Proof: The case $t = 1$ is different since the definition of w_{i0} differs from the definition in the general case.

For $t = 1$

$$y_{i1} = y_{i0}\alpha + X_{i1}\beta + \sigma u_{i1}$$

$$= w_{i0}\alpha + X_{i1}\beta + \sigma u_{i1} .$$

For $t = 2$ the result is obviously true. Assume the result is true for an arbitrary $t = s$. Then

$$y_{is+1} = \alpha y_{is} + X_{is+1}\beta + \sigma u_{is+1}$$

$$= \alpha[\sum_{j=1}^{s} \alpha^{s-j}(X_{ij}\beta + \sigma u_{ij}) + \alpha^{s} y_{i0}] + X_{is+1}\beta + \sigma u_{is+1}$$

$$= \sum_{j=1}^{s+1} \alpha^{s+1-j}(X_{ij}\beta + \sigma u_{ij}) + \alpha^{s+1} y_{i0}$$

which equals $w_{is+1}\alpha + X_{is+1}\beta + \sigma v_{is+1}\alpha + \sigma u_{is+1}$.

q.e.d.

The following notation is used in the remainder of this paper. Let Σ be an arbitrary symmetric positive definite matrix. Σ^{-1} is the weighting matrix for GLS and will usually be taken as Ω^{-1} for GLS or as I for OLS. Let

$$Q = (Z'\Sigma^{-1}Z)^{-1} = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} = (q^{c} \quad q_{2})$$

where q_{11} is a scalar, q_{12} is $1xk$ and q_{21} is its transpose, and

q_{22} is kxk . q^c is a k+lxl column vector and q_2 is k+lxk . Further, let $Q^* = (q^c \ \ 0)$ be a k+l x k+l matrix. Define P_Z as the projection $Z(Z'Z)^{-1}Z'$ into the space spanned by the columns of Z . Thus P_Z is an idempotent and symmetric matrix. Let $\bar{P}_Z = I - P_Z$, the projection into the space orthogonal to Z . The projection notation can be generalized as follows. Let P_Z^Σ be the projection in the metric defined by Σ ,

$$P_Z^\Sigma = \Sigma^{-1/2}Z(Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1/2}$$

P_Z^Σ is, again, symmetric and idempotent. Let $M_1 = \Sigma^{-1}ZQZ'\Sigma^{-1}$ and $\bar{M}_1 = \Sigma^{-1} - M_1$. Then $M_1 = \Sigma^{-1/2}P_Z^\Sigma\Sigma^{-1/2}$ and $\bar{M}_1 = \Sigma^{-1/2}\bar{P}_Z^\Sigma\Sigma^{-1/2}$. If $\Sigma = I$, then M_1 reduces to P_Z . Define also $M_2 = \Sigma^{-1}Zq^cq^rZ'\Sigma^{-1}$ and $\bar{M}_2 = \Sigma^{-1} - M_2$ where q^r is $(q^c)'$. The following relationship holds between M_1 and M_2

$$M_2 = q_{11}(M_1 - \Sigma^{-1/2}P_X^\Sigma\Sigma^{-1/2}) \qquad (2\text{-}2)$$

$$= q_{11}\Sigma^{-1/2}(P_Z^\Sigma - P_X^\Sigma)\Sigma^{-1/2} .$$

This follows from observing that

$$q^cq^r = \begin{bmatrix} q_{11}^2 & q_{11}q_{12} \\ q_{21}q_{11} & q_{21}q_{12} \end{bmatrix} = q_{11}Q - q_{11}\begin{bmatrix} 0 & 0 \\ 0 & q_{22} - \dfrac{q_{21}q_{12}}{q_{11}} \end{bmatrix} \qquad (2\text{-}3)$$

$$= q_{11}\left[Q - \begin{pmatrix} 0 & 0 \\ 0 & (X'\Sigma^{-1}X)^{-1} \end{pmatrix} \right] \qquad (2\text{-}4)$$

TABLE II

Definitions

$y_{it} = y_{it-1}\alpha + X_{it}\beta + \sigma u_{it}$, $i = 1, \ldots, N$, $T = 1, \ldots, T$ or

$$\underset{NTx1}{y} = \underset{NTx1}{y_{-1}}\alpha + \underset{NTxk}{X}\ \underset{kx1}{\beta} + \sigma \underset{NTx1}{u}$$

$$Euu' = \Omega = I_N \otimes \mathcal{Q} \text{ where } \mathcal{Q} = \begin{bmatrix} 1 & \rho & \cdots & & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & & & & \\ \rho & \rho & \cdots & \rho & 1 \end{bmatrix}$$

$y_{-1} = W + \sigma V$ where W is nonstochastic and V is stochastic

$W = (w_{10}, \ldots, w_{1T-1}, \ldots, w_{N0}, \ldots, w_{NT-1})'$

$V = (v_{10}, \ldots, v_{1T-1}, \ldots, v_{N0}, \ldots, v_{NT-1})'$.

$$w_{it} = \sum_{j=1}^{t} \alpha^{t-j} x_{ij}\beta + \alpha^t y_{i0} \ ; \quad v_{it} = \sum_{j=1}^{t} \alpha^{t-j} u_{ij} \ , \quad i = 1, \ldots, N \quad t = 1, \ldots, T-1$$

$$w_{i0} = y_{i0} \ ; \quad v_{i0} = 0 \ , \quad i = 1, \ldots, N$$

$\underset{NTxk+1}{Z} = (W \ \ X)$ is the nonstochastic part of the regressors and

$\underset{NTxk+1}{V^*} = (\underset{NTx1}{V} \ \ \underset{NTxk}{0})$ is the stochastic part

$C = EuV' = I_N \otimes (\rho C_1 + (1-\rho)C_2)$, $C_2^* = I_N \otimes C_2$

$G = EVV' = I_N \otimes (\rho G_1 + (1-\rho)G_2)$

$a = (\alpha^0 \ \ \alpha^1 \ \ \alpha^2 \ \ldots \ \alpha^{T-1})'$

TABLE II (continued)

Σ is an arbitrary symmetric positive definite matrix

$$Q = (Z'\Sigma^{-1}Z)^{-1} = \begin{pmatrix} \underset{1x1}{q_{11}} & \underset{1xk}{q_{12}} \\ \underset{kx1}{q_{21}} & \underset{kxk}{q_{22}} \end{pmatrix} = \begin{pmatrix} \underset{k+1x1}{q^c} & \underset{k+1xk}{q_2} \end{pmatrix}$$

$Q^* = (q^c \quad 0)$

$P_Z = Z(Z'Z)^{-1}Z'$, the projection operator into the space spanned by the columns of Z .

$\bar{P}_Z = I - P_Z$

$P_Z^\Sigma = \Sigma^{-1/2}Z(Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1/2}$

$M_1 = \Sigma^{-1}ZQZ'\Sigma^{-1}$, $\bar{M}_1 = \Sigma^{-1} - M_1$

$M_2 = \Sigma^{-1}Zq^c(q^c)'Z'\Sigma^{-1} = q_{11}\Sigma^{-1/2}(P_Z^\Sigma - P_X^\Sigma)\Sigma^{-1/2}$

$d = (d_{11}, \ldots, d_{NT})'$, an instrumental variable

$D = (d \quad X)$

$H = (D'Z)^{-1}$

where (4) follows from the formula for the inverse of a partitioned matrix. Using (4) M_2 can be written as

$$M_2 = q_{11}\Sigma^{-1}Z\left[Q - \begin{pmatrix} 0 & 0 \\ 0 & (X'\Sigma^{-1}X)^{-1} \end{pmatrix} \right] Z'\Sigma^{-1}$$

which is $q_{11}\Sigma^{-1/2}(P_Z^{\Sigma} - P_X^{\Sigma})\Sigma^{-1/2}$. The following convention will be used on M_1 , M_2 , and Q . These symbols appearing without a superscript refer to an arbitrary Σ . When it is desired to specify Σ , a superscript will be used. For example, $Q = (Z'\Sigma^{-1}Z)^{-1}$ but $Q^{\Omega} = (Z'\Omega^{-1}Z)^{-1}$.

The analysis for a Generalized Least Squares estimator of equation (1-3) is done for an arbitrary weighting matrix Σ^{-1} . As special cases corresponding to $\Sigma = I$ and $\Sigma = \Omega$ the moments of OLS and the usual GLS estimator with the correct covariance matrix are found. The estimator to be analyzed is

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}_{\Sigma} = [(y_{-1} \quad X)'\Sigma^{-1}(y_{-1} \quad X)]^{-1}(y_{-1} \quad X)'\Sigma^{-1}y \ . \qquad (2\text{-}5)$$

To analyze this estimator it is necessary to find the approximate error of the estimator, $\begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{pmatrix}_{\Sigma}$, as

<u>Lemma 2.</u> $\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}_{\Sigma} = Q\{\sigma Z'\Sigma^{-1}u$

$$+ \sigma^2[V^{*\prime}\bar{M}_1 u - Z'\Sigma^{-1}V^* QZ'\Sigma^{-1}u]$$

$$+ \sigma^3[-V^{*\prime}\Sigma^{-1}ZQV^{*\prime}\bar{M}_1 u - V^{*\prime}\bar{M}_1 V^* QZ'\Sigma^{-1}u - Z'\Sigma^{-1}V^* QV^{*\prime}\bar{M}_1 u$$

$$+ Z'\Sigma^{-1}V^* QZ'\Sigma^{-1}V^* QZ'\Sigma^{-1}u]$$

$$+ \sigma^4[-q_{11}V^{*'}\bar{M}_1V^*V^{*'}\bar{M}_1u + 2V^{*'}\bar{M}_1V^*V^{*'}M_2u + q_{11}Z'\Sigma^{-1}V^*V^{*'}\bar{M}_1V^*QZ'\Sigma^{-1}u$$

$$+ V^{*'}M_2V^*V^{*'}\bar{M}_1u + 2q_{11}Z'\Sigma^{-1}V^*V^{*'}\Sigma^{-1}ZQV^{*'}\bar{M}_1u$$

$$- Z'\Sigma^{-1}V^*V^{*'}M_2V^*QZ'\Sigma^{-1}u]\} + o_p(\sigma^5) .$$

Proof: Define A as $(V^{*'}\Sigma^{-1}Z + Z'\Sigma^{-1}V^*)Q$ and B as $V^{*'}\Sigma^{-1}V^*Q$. Using the definition of the estimator (5), the error of the estimator, $\begin{pmatrix}\hat{\alpha} - \alpha \\ \hat{\beta} - \beta\end{pmatrix}_{\Sigma}$, can be written as

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}_{\Sigma} = \sigma[(Z' + \sigma V^{*'})\Sigma^{-1}(Z + \sigma V^*)]^{-1}(Z' + \sigma V^{*'})\Sigma^{-1}u \qquad (2\text{-}6)$$

which is equal to

$$\sigma Q[I + \sigma(A + \sigma B)]^{-1}(Z' + \sigma V^{*'})\Sigma^{-1}u . \qquad (2\text{-}7)$$

If the inverse matrix $[I + \sigma(A + \sigma B)]^{-1}$ exists, (7) can be approximated by a sigma-expansion using the formula

$$(I + \sigma E)^{-1} = I - \sigma E + \sigma^2E^2 - \sigma^3E^3 + \dots . \qquad (2\text{-}8)$$

The existence of the inverse matrix is shown by the following argument. The matrix to be inverted is of the form $(D + \sigma E)$ where D is non-stochastic and E contains random terms. D is nonsingular by assumption. Consider the determinant $|D + \sigma E|$. This determinant is a polynomial expression in σ and is, therefore, a continuous function of σ. Then since $|D| \neq 0$ by assumption, there exists a neighborhood around $\sigma = 0$ where $|D + \sigma E|$ is also nonzero. Therefore, the desired inverse exists for sigma sufficiently small to be in that neighborhood. It is assumed throughout that sigma satisfies this criterion. The neighborhood of sigma around zero in which the inverse exists will be random since E is a random matrix.

Then using formula (8) the error expression (7) can be approximated as

$$\sigma Q[I - \sigma A + \sigma^2(AA - B) + \sigma^3(BA + AB - AAA)](Z' + \sigma V^{*\prime})\Sigma^{-1}u + 0_p(\sigma^5) \quad (2\text{-}9)$$

which is, after grouping terms by powers of sigma,

$$\begin{aligned} Q\{&\sigma Z'\Sigma^{-1}u + \sigma^2[-AZ'\Sigma^{-1}u + V^{*\prime}\Sigma^{-1}u] \\ &+ \sigma^3[(AA - B)Z'\Sigma^{-1}u - AV^{*\prime}\Sigma^{-1}u] \\ &+ \sigma^4[(BA + AB - AAA)Z'\Sigma^{-1}u + (AA - B)V^{*\prime}\Sigma^{-1}u]\} + 0_p(\sigma^5) \ . \end{aligned} \quad (2\text{-}10)$$

To facilitate the proofs to follow, it is convenient to regroup the terms of (10). Consider first the terms of order σ^2. They are

$$-(V^{*\prime}\Sigma^{-1}Z + Z'\Sigma^{-1}V^*)QZ'\Sigma^{-1}u + V^{*\prime}\Sigma^{-1}u = V^{*\prime}\overline{M}_1u - Z'\Sigma^{-1}V^*QZ'\Sigma^{-1}u \ . \quad (2\text{-}11)$$

The terms of order σ^3 are

$$\begin{aligned} &[(V^{*\prime}\Sigma^{-1}Z + Z'\Sigma^{-1}V^*)Q(V^{*\prime}\Sigma^{-1}Z + Z'\Sigma^{-1}V^*)Q - V^{*\prime}\Sigma^{-1}V^*Q]Z'\Sigma^{-1}u \\ &\quad - (V^{*\prime}\Sigma^{-1}Z + Z'\Sigma^{-1}V^*)QV^{*\prime}\Sigma^{-1}u \ . \end{aligned} \quad (2\text{-}12)$$

Multiplying (12) out and collecting terms gives

$$-V^{*\prime}\Sigma^{-1}ZQV^{*\prime}\bar{M}_1u - V^{*\prime}\bar{M}_1V^*QZ'\Sigma^{-1}u - Z'\Sigma^{-1}V^*QV^{*\prime}\bar{M}_1u$$
$$+ Z'\Sigma^{-1}V^*QZ'\Sigma^{-1}V^*QZ'\Sigma^{-1}u \; . \tag{2-13}$$

The terms of order σ^4 are, similarly,

$$V^{*\prime}\Sigma^{-1}V^*QV^{*\prime}M_1u + V^{*\prime}\Sigma^{-1}V^*QZ'\Sigma^{-1}V^*QZ'\Sigma^{-1}u$$
$$+ V^{*\prime}\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}V^*QZ'\Sigma^{-1}u + Z'\Sigma^{-1}V^*QV^{*\prime}\Sigma^{-1}V^*QZ'\Sigma^{-1}u$$
$$- [V^{*\prime}\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}ZQV^{*\prime}M_1u + V^{*\prime}\Sigma^{-1}ZQV^{*\prime}M_1V^*QZ'\Sigma^{-1}u$$
$$+ V^{*\prime}M_1V^*QV^{*\prime}M_1u + V^{*\prime}M_1V^*QZ'\Sigma^{-1}V^*QZ'\Sigma^{-1}u$$
$$+ Z'\Sigma^{-1}V^*QV^{*\prime}\Sigma^{-1}ZQV^{*\prime}M_1u + Z'\Sigma^{-1}V^*QV^{*\prime}M_1V^*QZ'\Sigma^{-1}u$$
$$+ Z'\Sigma^{-1}V^*QZ'\Sigma^{-1}V^*QV^{*\prime}M_1u + Z'\Sigma^{-1}V^*QZ'\Sigma^{-1}V^*QZ'\Sigma^{-1}V^*QZ'\Sigma^{-1}u]$$
$$+ V^{*\prime}\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}u + V^{*\prime}M_1V^*QV^{*\prime}\Sigma^{-1}u + Z'\Sigma^{-1}V^*QV^{*\prime}\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}u$$
$$+ Z'\Sigma^{-1}V^*QV^{*\prime}\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}u + Z'\Sigma^{-1}V^*QZ'\Sigma^{-1}V^*QV^{*\prime}\Sigma^{-1}u$$
$$- V^{*\prime}\Sigma^{-1}V^*QV^{*\prime}\Sigma^{-1}u$$

which reduces to

$$- V^{*\prime}\bar{M}_1V^*QV^{*\prime}\bar{M}_1u + V^{*\prime}\bar{M}_1V^*QZ'\Sigma^{-1}V^*QZ'\Sigma^{-1}u$$
$$+ V^{*\prime}\Sigma^{-1}ZQV^{*\prime}\bar{M}_1V^*QZ'\Sigma^{-1}u + Z'\Sigma^{-1}V^*QV^{*\prime}\bar{M}_1V^*QZ'\Sigma^{-1}u$$
$$+ V^{*\prime}\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}ZQV^{*\prime}\bar{M}_1u + Z'\Sigma^{-1}V^*QV^{*\prime}\Sigma^{-1}ZQV^{*\prime}\bar{M}_1u \tag{2-14}$$
$$+ Z'\Sigma^{-1}V^*QZ'\Sigma^{-1}V^*QV^{*\prime}\bar{M}_1u$$
$$- Z'\Sigma^{-1}V^*QZ'\Sigma^{-1}V^*QZ'\Sigma^{-1}V^*QZ'\Sigma^{-1}u \; .$$

The second term of (14) is

$$V^{*\prime}\bar{M}_1 V^* Q Z'\Sigma^{-1} V^* Q Z'\Sigma^{-1} u$$

$$= V^{*\prime}\bar{M}_1 V q^r Z'\Sigma^{-1} V q^r Z'\Sigma^{-1} u \ . \tag{2-15}$$

Noting that $q^r Z'\Sigma^{-1} V$ is a scalar and, therefore, equals its transpose gives for (15)

$$V^{*\prime}\bar{M}_1 V^* V^{*\prime}\Sigma^{-1} Z q^c q^r Z'\Sigma^{-1} u \tag{2-16}$$

$$= V^{*\prime}\bar{M}_1 V^* V^{*\prime} M_2 u \ . \tag{2-17}$$

The third term of (14) is

$$V^{*\prime}\Sigma^{-1} Z Q V^{*\prime}\bar{M}_1 V^* Q Z'\Sigma^{-1} u$$

$$= V^{*\prime}\Sigma^{-1} Z q^c V'\bar{M}_1 V q^r Z'\Sigma^{-1} u$$

$$= V'\bar{M}_1 V V^{*\prime} M_2 u \ , \tag{2-18}$$

taking advantage of the fact that $V'\bar{M}_1 V$ is a scalar. Then (18) is

$$V^{*\prime}\bar{M}_1 V^* V^{*\prime} M_2 u \ . \tag{2-19}$$

The fifth term of (14) is

$$V^{*\prime}\Sigma^{-1} Z Q V^{*\prime}\Sigma^{-1} Z Q V^{*\prime}\bar{M}_1 u$$

$$= V^{*\prime}\Sigma^{-1} Z q^c V'\Sigma^{-1} Z q^c V'\bar{M}_1 u$$

$$= V^{*\prime}\Sigma^{-1} Z q^c q^r Z'\Sigma^{-1} V V'\bar{M}_1 u \ , \tag{2-20}$$

transposing the scalar $V'\Sigma^{-1} Z q^c$. Then (20) is

$$V^{*\prime} M_2 V^* V^{*\prime}\bar{M}_1 u \ . \tag{2-21}$$

The seventh term of (14) can be similarly rewritten

$$
\begin{aligned}
&Z'\Sigma^{-1}V^*QZ'\Sigma^{-1}V^*QV^{*\prime}\bar{M}_1u \\
&= Z'\Sigma^{-1}Vq^rZ'\Sigma^{-1}Vq_{11}V'\bar{M}_1u \\
&= q_{11}Z'\Sigma^{-1}V^*V^{*\prime}\Sigma^{-1}ZQV^{*\prime}\bar{M}_1u \; . \qquad (2\text{-}22)
\end{aligned}
$$

Rewriting the last term of (14) gives

$$
\begin{aligned}
&-Z\Sigma^{-1}V^*QZ'\Sigma^{-1}V^*QZ'\Sigma^{-1}V^*QZ'\Sigma^{-1}u \\
&= -Z'\Sigma^{-1}V^*V^{*\prime}M_2V^*QZ'\Sigma^{-1}u \; . \qquad (2\text{-}23)
\end{aligned}
$$

Collecting the first term of (10) and expressions (11), (13), and (14) and using (17), (18), (19), (21), and (23) gives the lemma.

Lemma 2 can be employed to find the bias and mean squared error for GLS and OLS up to terms of order σ^6 . The bias is found first for the estimator with weighting matrix Σ^{-1} as

<u>Theorem 1</u>. Under the set of assumptions (θ) and assuming that Σ is a positive definite and symmetric matrix the small-sigma asymptotic bias of the estimator (5) is

$$
\begin{aligned}
E\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}_{\Sigma} &= \sigma^2 Q\left[\begin{pmatrix} \operatorname{tr} \bar{M}_1C \\ 0 \end{pmatrix} - Z'\Sigma^{-1}C'\Sigma^{-1}Zq^c\right] \\
&+ \sigma^4 Q\left\{\begin{bmatrix} -q_{11}(\operatorname{tr} \bar{M}_1G \operatorname{tr} \bar{M}_1C + 2 \operatorname{tr} \bar{M}_1G\bar{M}_1C) + 2(\operatorname{tr} \bar{M}_1G \operatorname{tr} M_2C + 2 \operatorname{tr} \bar{M}_1GM_2C \\ + \operatorname{tr} M_2G \operatorname{tr} \bar{M}_1C + 2 \operatorname{tr} M_2G\bar{M}_1C \\ 0 \end{bmatrix}\right. \\
&+ Z'\Sigma^{-1}[q_{11}(2G\bar{M}_1(C+C') + 2C'\bar{M}_1G + 2G \operatorname{tr} \bar{M}_1C + C' \operatorname{tr} \bar{M}_1G \\
&\quad \left. - 2GM_2C' - C' \operatorname{tr} M_2G]\Sigma^{-1}Zq^c\right\} + O(\sigma^6)
\end{aligned}
$$

where $C = EuV'$ and $G = EVV'$. The $NT \times NT$ matrices of expectations C and G are derived in the appendix in Lemmas A1 and A3. It is shown there that when the disturbances u have the time series of cross-sections covariance structure described in Table I the expectations are $C = I_N \otimes (\rho C_1 + (1-\rho)C_2)$ where ρ is the intraclass correlation coefficient,

$$C_1 = \frac{1}{1-\alpha}\begin{bmatrix} 0 & 1-\alpha & 1-\alpha^2 & \cdots & 1-\alpha^{T-1} \\ 0 & 1-\alpha & 1-\alpha^2 & \cdots & 1-\alpha^{T-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 1-\alpha & 1-\alpha^2 & \cdots & 1-\alpha^{T-1} \end{bmatrix}$$

$$C_2 = \begin{bmatrix} 0 & 1 & \alpha & \alpha^2 & \cdots & \alpha^{T-2} \\ 0 & 0 & 1 & \alpha & \cdots & \alpha^{T-3} \\ \vdots & & & & & \vdots \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & 1 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \end{bmatrix},$$

and $G = I_N \otimes (\rho G_1 + (1-\rho)G_2)$ where

$$G_1 = \frac{1}{(1-\alpha)^2}\begin{bmatrix} 0 & 0 & 0 & & 0 \\ 0 & (1-\alpha)^2 & (1-\alpha)(1-\alpha^2) & \cdots & (1-\alpha)(1-\alpha^{T-1}) \\ \vdots & \vdots & & & \vdots \\ 0 & (1-\alpha^{T-1})(1-\alpha) & (1-\alpha^{T-1})(1-\alpha^2) & \cdots & (1-\alpha^{T-1})^2 \end{bmatrix},$$

and

$$G_2 = \frac{1}{\alpha^2-1}\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & \alpha^2(1-\alpha^{-2}) & \alpha^3(1-\alpha^{-2}) & \cdots & \alpha^T(1-\alpha^{-2}) \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & \alpha^T(1-\alpha^{-2}) & \alpha^{T+1}(1-\alpha^{-4}) & \cdots & \alpha^{2(T-1)}(1-\alpha^{2(T-1)}) \end{bmatrix} .$$

Defining a as $(1 \ \alpha \ \alpha^2 \ \ldots \ \alpha^{T-1})'$ and y_0 as $(y_{10} \ y_{20} \ \cdots \ y_{N0})$, it can be verified that $W = (I_N \otimes C_2')X\beta + y_0 \otimes a$. C_1 is the component of C due to the nonzero intraclass correlation among the disturbances and the presence of the lagged dependent variable as a regressor while C_2 arises solely from the presence of y_{-1} . The component matrices of G , G_1 and G_2 similarly represent intracorrelation effects and lagged variable effects, respectively. G is the matrix of covariances among the reduced-form errors in y_{-1} and C is the matrix of covariances between these reduced-form errors and the disturbance terms. Clearly, both of these matrices depend on the coefficients α and β of the equation (1-3). G is a symmetric matrix, but C is not.

Proof: The proof is accomplished by finding the expectations of the terms in Lemma 2 term by term. Recall that the stochastic elements in that lemma are V^* and u while $\bar{M}_1$, M_2 , Z , Q , and Σ^{-1} are non-stochastic.

The first term of Lemma 2 has expectation

$$E\sigma QZ'\Sigma^{-1}u = \sigma QZ'\Sigma^{-1}Eu = 0 \qquad (2\text{-}24)$$

since u is assumed to have zero mean. The terms of order σ^2 in the lemma are, except for Q which premultiplies all of the terms in the expression,

$$V^{*'}\overline{M}_1 u - Z'\Sigma^{-1}V^{*}QZ'\Sigma^{-1}u \ . \tag{2-25}$$

The expectation of (25) is

$$E\begin{pmatrix} V'\overline{M}_1 u \\ 0 \end{pmatrix} - Z'\Sigma^{-1}EVq^{r}Z'\Sigma^{-1}u \ . \tag{2-26}$$

Since $V'\overline{M}_1 u$ and $q^{r}Z'\Sigma^{-1}u$ are scalars, (26) can be rewritten as

$$\begin{pmatrix} E \ \ \mathrm{tr}\ \overline{M}_1 uV' \\ 0 \end{pmatrix} - Z'\Sigma^{-1}EVu'\Sigma^{-1}Zq^{c}$$

which is, using Lemma A1 and noting that $\mathrm{tr}\ \overline{M}_1 C' = \mathrm{tr}\ \overline{M}_1 C$,

$$\begin{pmatrix} \mathrm{tr}\ \overline{M}_1 C \\ 0 \end{pmatrix} - Z'\Sigma^{-1}C'\Sigma^{-1}Zq^{c} \ . \tag{2-27}$$

The terms of order σ^3 in Lemma 2 all have expectation zero since each one is an odd product of normally distributed random variables with zero means (see Anderson [3]).

The terms remaining to be evaluated are all of order σ^4 . They are considered one by one.

$$\begin{aligned} &E(-q_{11}V^{*'}\overline{M}_1 V^{*}V^{*'}\overline{M}_1 u) \\ &= -q_{11}\begin{bmatrix} \mathrm{tr}\ \overline{M}_1 G\ \mathrm{tr}\ \overline{M}_1 C + 2\ \mathrm{tr}\ \overline{M}_1 G\overline{M}_1 C \\ 0 \end{bmatrix} \end{aligned} \tag{2-28}$$

using Lemma A9.

The next term of order σ^4 is $2V^{*\prime}\bar{M}_1V^*V^{*\prime}M_2u$ which has expectation

$$2\begin{bmatrix} \text{tr } \bar{M}_1G \text{ tr } M_2C + 2 \text{ tr } \bar{M}_1GM_2C \\ 0 \end{bmatrix} \tag{2-29}$$

from Lemma A9.

The expectation of the third term is found from Lemma A10 as

$$EZ'\Sigma^{-1}V^*QV^{*\prime}\bar{M}_1V^*QZ'\Sigma^{-1}u = q_{11}Z'\Sigma^{-1}[2G\bar{M}_1 + I \text{ tr } \bar{M}_1G]C'\Sigma^{-1}Zq^c \; . \tag{2-30}$$

The expectation of the fourth term is given by Lemma A9 as

$$EV^{*\prime}M_2V^*V^{*\prime}\bar{M}_1u = \begin{bmatrix} \text{tr } M_2G \text{ tr } \bar{M}_1C + 2 \text{ tr } M_2G\bar{M}_1C \\ 0 \end{bmatrix} . \tag{2-31}$$

The expectation of the next term is given by Lemma A11,

$$\begin{aligned} &E2q_{11}Z'\Sigma^{-1}V^*V^{*\prime}\Sigma^{-1}ZQV^{*\prime}\bar{M}_1u \\ &= 2q_{11}Z'\Sigma^{-1}[G \text{ tr } \bar{M}_1C + G\bar{M}_1C + C'\bar{M}_1G]\Sigma^{-1}Zq^c \; . \end{aligned} \tag{2-32}$$

The expectation of the last term in Lemma 2 is found from Lemma A10 as

$$\begin{aligned} &E{-Z'}\Sigma^{-1}V^*V^{*\prime}M_2V^*QZ'\Sigma^{-1}u \\ &= -Z'\Sigma^{-1}[2GM_2C' + C' \text{ tr } M_2G]\Sigma^{-1}Zq^c \; . \end{aligned} \tag{2-33}$$

Having taken all of the necessary expectations, it remains only to collect the terms of order σ^4. Collecting the terms of order σ^4 with only the first element nonzero from (28), (29), and (31) gives

$$\begin{bmatrix} -q_{11}(\text{tr}\ \bar{M}_1G\ \text{tr}\ \bar{M}_1C + 2\ \text{tr}\ \bar{M}_1G\bar{M}_1C) + 2(\text{tr}\ \bar{M}_1G\ \text{tr}\ M_2C + 2\ \text{tr}\ \bar{M}_1GM_2C) \\ + \text{tr}\ \bar{M}_1C\ \text{tr}\ M_2G + 2\ \text{tr}\ M_2G\bar{M}_1C \\ 0 \end{bmatrix} \qquad (2\text{-}34)$$

The remaining terms from (30), (32), and (33) are

$$Z'\Sigma^{-1}[q_{11}(2G\bar{M}_1C' + C'\ \text{tr}\ \bar{M}_1G + 2G\ \text{tr}\ \bar{M}_1C + 2G\bar{M}_1C + 2C'\bar{M}_1G) - 2GM_2C' - C'\ \text{tr}\ M_2G]\Sigma^{-1}Zq^c\ . \qquad (2\text{-}35)$$

The terms of order σ^5 are not shown in Lemma 2, but they would all be odd products of normally distributed random variables and, therefore, would have expectation zero. Then combining (24), (27), (34), and (35) completes the proof of the theorem.

The bias of an estimator may influence the decision of whether or not to use the estimator, but bias alone is not usually an adequate criterion by which to judge an estimator; bias functions are not usually thought to be monotonic functions of bias alone. A more appropriate measure of the adequacy of an estimator may be its mean squared error matrix. Theorem 2 gives the small-sigma asymptotic mean squared error of the general estimator of formula (5).

<u>Theorem 2</u>. Under assumptions (θ) and assuming that Σ is a symmetric positive definite matrix the small-sigma asymptotic mean squared error of the estimator given by (5) is

$$E\begin{pmatrix}\hat{\alpha} - \alpha \\ \hat{\beta} - \beta\end{pmatrix}_{\Sigma}\begin{pmatrix}\hat{\alpha} - \alpha \\ \hat{\beta} - \beta\end{pmatrix}_{\Sigma}' = E(ee')_{\Sigma}$$

$$= \sigma^2 QZ'\Sigma^{-1}\Omega\Sigma^{-1}ZQ + \sigma^4 Q\left\{\begin{bmatrix}(\text{tr }\bar{M}_1C)^2 + \text{tr }\bar{M}_1C\bar{M}_1C + \text{tr }\bar{M}_1G\bar{M}_1\Omega & 0 \\ 0 & 0\end{bmatrix}\right.$$

$$- Z'\Sigma^{-1}[(C+C')\text{tr }\bar{M}_1C + C\bar{M}_1C + C'\bar{M}_1C' + 2C\bar{M}_1C' + G\bar{M}_1\Omega + \Omega\bar{M}_1G + \Omega\text{ tr }\bar{M}_1G]\Sigma^{-1}ZQ^*$$

$$- Q^{*\prime}Z'\Sigma^{-1}[(C+C')\text{tr }\bar{M}_1C + C\bar{M}_1C + C'\bar{M}_1C' + 2C\bar{M}_1C' + G\bar{M}_1\Omega + \Omega\bar{M}_1G + \Omega\text{ tr }\bar{M}_1G]\Sigma^{-1}Z$$

$$+ Z'\Sigma^{-1}[(C+C')\text{tr }M_2C + CM_2C + C'M_2C' + 2C'M_2C + GM_2\Omega + \Omega M_2G + G\text{ tr }M_2\Omega$$

$$- q_{11}((C+C')\text{tr }\bar{M}_1C + C\bar{M}_1C + C'\bar{M}_1C' + G\bar{M}_1\Omega + \Omega\bar{M}_1G)]\Sigma^{-1}Z\}Q + O(\sigma^6) .$$

Four of the terms in the mean squared error are contributed by the bias and the remainder come from the variance. The bias contributes the term $(\text{tr }\bar{M}_1C)^2$ in the first part of the σ^4 expression, the term $-Z'\Sigma^{-1}C'\Sigma^{-1}Zq^c$ tr $\bar{M}_1C$ and its transpose, and the term $Z'\Sigma^{-1}C'M_2C\Sigma^{-1}Z$. The first term of this theorem is similar to the familiar large-sample expression for the covariance matrix of this estimator except that the nonstochastic part of the regressors, Z , occurs in the expression instead of the regressor matrix itself. The next higher order term is, unfortunately, very lengthy.

<u>Proof</u>: The proof proceeds in a manner similar to the proof of Theorem 1. First, the squared error of the estimator is found and then the expectation of this expression is found term by term.

From Lemma 2 the squared error of the estimator is

$$(ee')_{\Sigma} = Q\{\sigma^2 Z'\Sigma^{-1}uu'\Sigma^{-1}Z$$

$$+ \sigma^3[Z'\Sigma^{-1}u(u'\bar{M}_1V^* - u'\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}Z) + (V^{*\prime}\bar{M}_1u - Z'\Sigma^{-1}V^*QZ'\Sigma^{-1}u)u'\Sigma^{-1}Z]$$

$$+ \sigma^4[(V^{*\prime}\bar{M}_1uu'\bar{M}_1V^* + Z'\Sigma^{-1}V^*QZ'\Sigma^{-1}uu'\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}Z$$

$$- Z'\Sigma^{-1}V^*QZ'\Sigma^{-1}uu'\bar{M}_1V^* - V^{*\prime}\bar{M}_1uu'\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}Z)$$

$$+ (-Z'\Sigma^{-1}uu'\bar{M}_1V^*QZ'\Sigma^{-1}V^* - Z'\Sigma^{-1}uu'\Sigma^{-1}ZQV^{*\prime}\bar{M}_1V^* \qquad (2\text{-}36)$$

$$- Z'\Sigma^{-1}uu'\bar{M}_1V^*QV^{*\prime}\Sigma^{-1}Z + Z'\Sigma^{-1}uu'\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}Z)$$

$$+ (-V^{*\prime}\Sigma^{-1}ZQV^{*\prime}\bar{M}_1uu'\Sigma^{-1}Z - V^{*\prime}\bar{M}_1V^*QZ'\Sigma^{-1}uu'\Sigma^{-1}Z$$

$$- Z'\Sigma^{-1}V^*QV^{*\prime}\bar{M}_1uu'\Sigma^{-1}Z + Z'\Sigma^{-1}V^*QZ'\Sigma^{-1}V^*QZ'\Sigma^{-1}uu'\Sigma^{-1}Z)]\}Q + o_p(\sigma^5)\ .$$

Observe that in the terms of order σ^4 the last set of terms in parentheses is the transpose of the preceding parenthesized set of terms and the fourth σ^4 term is the transpose of the third.

The expectation of the first term of (36) is

$$EZ'\Sigma^{-1}uu'\Sigma^{-1}Z = Z'\Sigma^{-1}\Omega\Sigma^{-1}Z\ . \qquad (2\text{-}37)$$

The terms of order σ^3 are all products of normally distributed random variables each with mean zero; therefore, they all have expectation zero. The expectation of the first group of terms of order σ^4 is found from Lemmas A4, A6, and A5 to be

$$E[V^{*\prime}\bar{M}_1 uu'\bar{M}_1 V^* + Z'\Sigma^{-1}V^*QZ'\Sigma^{-1}uu'\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}Z$$

$$- Z'\Sigma^{-1}V^*QZ'\Sigma^{-1}uu'\bar{M}_1V^* - V^{*\prime}\bar{M}_1uu'\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}Z]$$

$$= \begin{bmatrix} (\mathrm{tr}\ \bar{M}_1C)^2 + \mathrm{tr}\ \bar{M}_1C\bar{M}_1C + \mathrm{tr}\ \bar{M}_1G\bar{M}_1\Omega & 0 \\ 0 & 0 \end{bmatrix}$$

$$+ Z'\Sigma^{-1}(2C'M_2C + G\ \mathrm{tr}\ M_2\Omega)\Sigma^{-1}Z$$

$$- Z'\Sigma^{-1}(C'\ \mathrm{tr}\ \bar{M}_1C + C'\bar{M}_1C' + G\bar{M}_1\Omega)\Sigma^{-1}ZQ^* \qquad (2\text{-}38)$$

$$- Q^{*\prime}Z'\Sigma^{-1}(C\ \mathrm{tr}\ \bar{M}_1C + C\bar{M}_1C + \Omega\bar{M}_1G)\Sigma^{-1}Z\ ,$$

noting that the fourth term is the transpose of the third.

The expectation of the next group of four terms is calculated with the aid of Lemmas A15, A12, A13, and A7 as

$$E[-Z'\Sigma^{-1}uu'\bar{M}_1V^*QZ'\Sigma^{-1}V^* - Z'\Sigma^{-1}uu'\Sigma^{-1}ZQV^{*\prime}\bar{M}_1V^*$$

$$- Z'\Sigma^{-1}uu'\bar{M}_1V^*QV^{*\prime}\Sigma^{-1}Z + Z'\Sigma^{-1}uu'\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}ZQV^{*\prime}\Sigma^{-1}Z]$$

$$= -Z'\Sigma^{-1}[\Omega\bar{M}_1G + C\bar{M}_1C + C\ \mathrm{tr}\ \bar{M}_1C]\Sigma^{-1}ZQ^*$$

$$- Z'\Sigma^{-1}[\Omega\ \mathrm{tr}\ \bar{M}_1G + 2C\bar{M}_1C']\Sigma^{-1}ZQ^*$$

$$- q_{11}Z'\Sigma^{-1}[\Omega\bar{M}_1G + C\bar{M}_1C + C\ \mathrm{tr}\ \bar{M}_1C]\Sigma^{-1}Z \qquad (2\text{-}39)$$

$$+ Z'\Sigma^{-1}[CM_2C + C\ \mathrm{tr}\ M_2C + \Omega M_2G]\Sigma^{-1}Z\ .$$

The expectation of the last set of terms is the transpose of (39).

With all the necessary expectations in hand it remains only to collect terms to prove the theorem. The terms of order σ^2 are given by (37). The terms of order σ^4 are given by (38), (39), and the transpose of (39). The σ^5 terms are not shown in (36), but they would all involve odd products of normal variables and, therefore, have zero expectation as was the case in Theorem 1. This completes the proof of Theorem 2.

In the proofs of Theorems 1 and 2, the only use made of the variance-components structure of the disturbances was in finding the expectations matrices Ω, C, and G. Therefore, the theorems hold for an arbitrary disturbance structure if Ω, C, and G are approximately recalculated. In particular the theorems hold for a three-component error model in which a time effect is allowed for as a component of the error. The theorems generalize so easily because the weighting matrix, Σ^{-1}, is nonstochastic. If the covariance matrix of the disturbances were estimated, then its structure would enter in a more complicated way.

The cases of Theorems 1 and 2 of most interest are for the estimators GLS $(\Sigma = \Omega)$ and OLS $(\Sigma = I)$. These results can be readily seen from the theorem, but they are written out for completeness. The GLS corollaries are

<u>Corollary 1</u>: $$E\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}_{GLS} = \sigma^2 Q^\Omega \left[\begin{pmatrix} \text{tr } \overline{M}_1^\Omega C \\ 0 \end{pmatrix} - Z'\Omega^{-1}C'\Omega^{-1}Z(q^\Omega)^c \right]$$

$$+ \sigma^4 Q^\Omega \left\{ \begin{bmatrix} -q_{11}^\Omega(\text{tr } \overline{M}_1^\Omega G \text{ tr } \overline{M}_1^\Omega C + 2 \text{ tr } \overline{M}_1^\Omega G \overline{M}_1^\Omega C) + 2(\text{tr } \overline{M}_1^\Omega G \text{ tr } M_2^\Omega C + 2 \text{ tr } \overline{M}_1^\Omega G M_2^\Omega C) \\ + \text{tr } M_2^\Omega G \text{ tr } \overline{M}_1^\Omega C + 2 \text{ tr } M_2^\Omega G \overline{M}_1^\Omega C \\ 0 \end{bmatrix} \right.$$

$$+ Z'\Omega^{-1}[q_{11}^\Omega(2G\overline{M}_1^\Omega(C+C') + 2C'\overline{M}_1^\Omega G + 2G \text{ tr } \overline{M}_1^\Omega C + C' \text{ tr } \overline{M}_1^\Omega G$$

$$\left. - 2GM_2^\Omega C' - C' \text{ tr } M_2^\Omega G]\Omega^{-1}Zq^{\Omega c} \right\} + O(\sigma^6)$$

<u>Corollary 2</u>: $$E(ee')_{GLS} = \sigma^2 Q^\Omega$$

$$+ \sigma^4 Q^\Omega \left\{ \begin{bmatrix} (\text{tr } \overline{M}_1^\Omega C)^2 + \text{tr } \overline{M}_1^\Omega C \overline{M}_1^\Omega C + \text{tr } \overline{M}_1^\Omega G & 0 \\ 0 & 0 \end{bmatrix} \right.$$

$$- Z'\Omega^{-1}[(C+C')\text{tr } \overline{M}_1^\Omega C + C\overline{M}_1^\Omega C + C'\overline{M}_1^\Omega C' + 2C\overline{M}_1^\Omega C' + G\overline{M}_1^\Omega \Omega + \Omega \overline{M}_1^\Omega G + \Omega \text{ tr } \overline{M}_1^\Omega G]\Omega^{-1}ZQ^{\Omega *}$$

$$- Q^{\Omega *'} Z'\Omega^{-1}[(C+C')\text{tr } \overline{M}_1^\Omega C + C\overline{M}_1^\Omega C + C'\overline{M}_1^\Omega C' + 2C\overline{M}_1^\Omega C' + G\overline{M}_1^\Omega \Omega + \Omega \overline{M}_1^\Omega G + \Omega \text{ tr } \overline{M}_1^\Omega G]\Omega^{-1}Z$$

$$+ Z'\Omega^{-1}[(C+C')\text{tr } M_2^\Omega C + CM_2^\Omega C + C'M_2^\Omega C' + 2C'M_2^\Omega C + GM_2^\Omega \Omega + \Omega M_2^\Omega G + G \text{ tr } M_2^\Omega \Omega$$

$$\left. - q_{11}((C+C')\text{tr } \overline{M}_1^\Omega C + C\overline{M}_1^\Omega C + C'\overline{M}_1^\Omega C' + G\overline{M}_1^\Omega G)]\Omega^{-1}Z \right\} Q + O(\sigma^6)$$

Corollary 2 follows from setting $\Sigma = \Omega$ and noting that $QZ\Omega^{-1}\Omega\Omega^{-1}ZQ = Q$ and $\text{tr } \overline{M}_1^\Omega G \overline{M}_1^\Omega \Omega = \text{tr } \overline{M}_1^\Omega G$.

The corollaries for OLS can be written using the projection notation.

<u>Corollary 3</u>: $$E\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}_{OLS} = \sigma^2\left[\begin{pmatrix} \mathrm{tr}\ \bar{P}_Z C \\ 0 \end{pmatrix} - Z'C'Z(q^I)^c\right]$$

$$+ \sigma^4 q^I_{11}(Z'Z)^{-1}\begin{bmatrix} -\mathrm{tr}\ \bar{P}_Z G \quad \mathrm{tr}\ \bar{P}_Z C - 2\ \mathrm{tr}\ \bar{P}_Z G \bar{P}_Z C + 2\ \mathrm{tr}\ \bar{P}_Z G\ \mathrm{tr}\ (P_Z - P_X)C + 4\ \mathrm{tr}\ \bar{P}_Z G(P_Z - P_X)C \\ + \mathrm{tr}\ (P_Z - P_X)G\ \mathrm{tr}\ \bar{P}_Z C + 2\ \mathrm{tr}\ (P_Z - P_X)G\bar{P}_Z C \\ 0 \end{bmatrix}$$

$$+ Z'[2G\bar{P}_Z(C+C') + 2C'\bar{P}_Z G + 2G\ \mathrm{tr}\ \bar{P}_Z C + C'\ \mathrm{tr}\ \bar{P}_Z C$$

$$- 2G(P_Z - P_X)C' - C'\ \mathrm{tr}(P_Z - P_X)G]Z(q^I)^c\} + O(\sigma^6)$$

<u>Corollary 4</u>: $$E(ee')_{OLS} = \sigma^2(Z'Z)^{-1}$$

$$+ \sigma^4(Z'Z)^{-1}\begin{bmatrix} (\mathrm{tr}\ \bar{P}_Z C)^2 + \mathrm{tr}\ \bar{P}_Z C \bar{P}_Z C + \mathrm{tr}\ \bar{P}_Z G & 0 \\ 0 & 0 \end{bmatrix}$$

$$- Z'[(C+C')\mathrm{tr}\ \bar{P}_Z C + C\bar{P}_Z C + C'\bar{P}_Z C' + 2C\bar{P}_Z C' + G\bar{P}_Z \Omega + \Omega\bar{P}_Z G + \Omega\ \mathrm{tr}\ \bar{P}_Z G]ZQ^{I*}$$

$$- Q^{I*'}Z'[(C+C')\mathrm{tr}\ \bar{P}_Z C + C\bar{P}_Z C + C'\bar{P}_Z C' + 2C\bar{P}_Z C' + G\bar{P}_Z \Omega + \Omega\bar{P}_Z G + \Omega\ \mathrm{tr}\ \bar{P}_Z G]Z$$

$$+ q_{11}Z'[(C+C')\mathrm{tr}(P_Z - \bar{P}_Z - P_X)C + C(P_Z - \bar{P}_Z - P_X)C + C'(P_Z - \bar{P}_Z - P_X)C'$$

$$+ 2C'(P_Z - P_X)C + G(P_Z - \bar{P}_Z - P_X)\Omega + \Omega(P_Z - \bar{P}_Z - P_X)G$$

$$+ G\ \mathrm{tr}(P_Z - P_X)\Omega]Z\}(Z'Z)^{-1} + O(\sigma^6)\ .$$

This corollary follows from setting $\Sigma = I$ and using the relationship between M_1 and M_2 given at the beginning of this section.

A special case of these results may be of some independent interest. If the cross-section dimension, N, is set to one so that the sample size is T, all of the cross-section structure of the problem vanishes and then setting ρ to zero. Corollary 3 provides an approximate answer to the question of the bias of OLS due solely to the presence of the lagged dependent variable as a regressor when the initial observation y_{10} is considered to be fixed.

Hurwicz [6] and Johnston [7] considered the case where there are no exogenous variables in the equation

$$y_t = \alpha y_{t-1} + \sigma u_t \qquad (2\text{-}40)$$

or

$$y_t = \alpha y_{t-1} + \beta + \sigma u_t \qquad (2\text{-}41)$$

allowing for a constant term. The subscript for the cross-section has been suppressed since it would always be one in this discussion. It was found that the sign of the bias in the least-squares estimate of α is negative if $\alpha > 0$. Hurwicz also calculated the magnitude of the bias: exactly for sample sizes of two and three, and approximately for T greater than three. His approximation is good when α is near zero.

Corollary 3 can be used to find the asymptotic bias when at least one nonconstant exogenous variable is present, but it is not useful for equations (40) or (41). In these equations the notion of a small sigma has no meaning. To see this write the reduced form of (41) with the normalization $Var(u_t) = 1$. This gives

$$y_t = \alpha^t y_0 + \beta \frac{\alpha - \alpha^T}{1-\alpha} + \sigma \sum_{i=1}^{t} \alpha^{t-i} u_i \; . \qquad (2\text{-}42)$$

$$\text{Then}\quad \mathrm{Var}(y_t) = \mathrm{Var}\left(\alpha^t y_0 + \beta\,\frac{\alpha - \alpha^T}{1-\alpha} + \sigma \sum_{i=1}^{t} \alpha^{t-i} u_i\right)$$

$$= \sigma^2\ \mathrm{Var}\left(\sum_{i=1}^{t} \alpha^{t-i} u_i\right)$$

$$= \sigma^2\,\frac{1 - \alpha^{2t}}{1 - \alpha^2} \tag{2-43}$$

since y_0 is a fixed constant. This says that a small value of sigma also means a small amount of variance to be explained and the sigma approximation is not useful. Calculations of the expressions of this section suggest that sigma approximations are good when the disturbance variance is small relative to the variance of the exogenous variables in the model weighted by their coefficients. That is, the approximations are good when the model fits well leaving a relatively small residual unexplained variance. Equations (40) and (41), however, have no source of variation independent of the disturbance terms; therefore, the sigma approximations are uninteresting there.

When a nonconstant exogenous variable is added to the equation, however, the sigma approximation can be used. In this case the simple conclusion of Johnston that the bias in $\hat{\alpha}$ is negative when α is positive no longer holds. The sign and magnitude of the bias will depend on the statistical properties of the exogenous variables. To see this consider the terms of order σ^2 in Corollary 3 under the assumptions that (i) there is just one nonconstant exogenous variable; (ii) $N = 1$; (iii) $\rho = 0$; and (iv) the initial observation y_0 is zero. Then Q is a two by two matrix and

$$Q = (Z'Z)^{-1} = \frac{1}{\Delta}\begin{bmatrix} X'X & -W'X \\ -X'W & W'W \end{bmatrix} \qquad (2\text{-}44)$$

where $\Delta = (X'X)(W'W) - (W'X)^2$. Recalling that under assumption (iv) $W = C_2'X\beta$, Corollary 3 can be rewritten as

$$E\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}_{OLS} = \sigma^2 \left\{ \begin{pmatrix} \operatorname{tr} \bar{P}_Z C_2 \\ 0 \end{pmatrix} - \frac{1}{\Delta}\begin{pmatrix} X'C_2C_2C_2'X\beta^2 & \beta'X'C_2C_2'X \\ X'C_2C_2X\beta & X'C_2X \end{pmatrix}\begin{pmatrix} X'X \\ -X'C_2X\beta \end{pmatrix} \right\}$$

$$+ O(\sigma^4) \ . \qquad (2\text{-}45)$$

The trace of $\bar{P}_Z C_2$ can be further broken down.

$$P_Z C_2 = \frac{1}{\Delta}[X'XWW'C_2 - W'XXW'C_2 - W'XWX'C_2 + W'WXX'C_2] \qquad (2\text{-}46)$$

using (44). The trace of (46) is

$$\frac{1}{\Delta}[X'XW'C_2W - W'XW'C_2X - W'XX'C_2W + W'WX'C_2X] \ . \qquad (2\text{-}47)$$

Then substituting $C_2'X\beta$ for W gives

$$\frac{\beta^2}{\Delta}[X'XX'C_2C_2C_2'X - X'C_2XX'C_2C_2X] \ . \qquad (2\text{-}48)$$

Finally, substituting (48) into (45) and noting that $\operatorname{tr} C_2 = 0$ gives the asymptotic bias of OLS as Corollary 5.

$$E\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}_{OLS} = -\frac{1}{\Delta}\sigma^2\begin{bmatrix} \beta^2[2(X'X)X'C_2^2C_2'X - (X'C_2X)X'C_2(C_2+C_2')X] \\ \beta\ [\ (X'X)X'C_2^2\ \ X - (X'C_2X)^2] \end{bmatrix} + O(\sigma^4)$$

where $\Delta = X'XW'W - (W'X)^2 = \beta^2(X'XX'C_2C_2'X - (X'C_2X)^2) > 0$ since Δ is the determinant of the positive definite matrix $(Z'Z)$.

The signs of the biases in Corollary 5 are ambiguous. The terms $X'X$, $(X'C_2X)^2$, and $X'C_2C_2'X$ are nonnegative, but $X'C_2X$ and $X'C_2^2X$ are indefinite quadratic forms. C_2 is clearly indefinite by inspection, or note that its trace is zero and, therefore, its characteristic roots sum to zero, but it is of rank $T-1$ which means that it must have characteristic roots of both signs. The square of C_2 is similarly seen to be indefinite. Therefore, the sign of the bias for OLS cannot be determined without reference to the particular exogenous variable in the regression, even in the special case considered.

It is, perhaps, interesting to note that while the sign of the bias of $\hat{\beta}$ depends on the sign of β and on α , the sign of the bias in $\hat{\alpha}$ does not depend on the sign of β . In fact, the bias of $\hat{\alpha}$ does not depend on β at all.

If something is known about the behavior of the exogenous variable, however, more can be said. In particular, suppose there is one exogenous variable, X_t , which grows geometrically and is one in period zero so that $X_t = \gamma^t$, $t = 1, 2, \ldots, T$. Assume for convenience that $0 < |\alpha| < 1$, $\alpha \neq \gamma$, $\gamma \neq 1$ and $\alpha\gamma \neq 1$. Since the bias expression for β is slightly simpler, attention is focused on the bias of $\hat{\beta}$ in order to compute some numerical results. Several quantities must be computed. $X'X$ is

$$(\gamma \quad \gamma^2 \quad \gamma^3 \quad \ldots \quad \gamma^T)\begin{pmatrix} \gamma \\ \gamma^2 \\ \vdots \\ \gamma^T \end{pmatrix} = \frac{\gamma^2 - \gamma^{2T+2}}{1 - \gamma^2} \;, \qquad (2\text{-}49)$$

summing the geometric series.

$$X'C_2X = (\gamma \quad \gamma^2 \quad \ldots \quad \gamma^T)\begin{bmatrix} 0 & 1 & \alpha & \alpha^2 & \ldots & \alpha^{T-2} \\ 0 & 0 & 1 & \alpha & \ldots & \alpha^{T-3} \\ 0 & 0 & 0 & 1 & \alpha \ldots & \alpha^{T-4} \\ \vdots & & & & & \\ 0 & 0 & . & . & . & 0 \quad 1 \\ 0 & 0 & . & . & . & 0 \quad 0 \end{bmatrix}\begin{bmatrix} \gamma \\ \gamma^2 \\ . \\ . \\ . \\ \gamma^T \end{bmatrix}$$

$$= \frac{1}{1 - \alpha\gamma}(\gamma \quad \gamma^2 \quad \ldots \quad \gamma^T)\begin{pmatrix} \gamma^2(1 - (\alpha\gamma)^{T-1}) \\ \gamma^3(1 - (\alpha\gamma)^{T-2}) \\ \vdots \\ \gamma^{T-1}(1 - (\alpha\gamma)^2) \\ \gamma^T(1 - \alpha\gamma) \\ 0 \end{pmatrix}$$

$$= \frac{1}{1 - \alpha\gamma}\left\{\frac{\gamma^3 - \gamma^{2T+1}}{1 - \gamma^2} - \alpha^{T-1}\gamma^{T+2}\left(1 + \frac{\gamma}{\alpha} + \left(\frac{\gamma}{\alpha}\right)^2 + \ldots + \left(\frac{\gamma}{\alpha}\right)^{T-2}\right)\right\}$$

$$= \frac{1}{1 - \alpha\gamma}\left\{\frac{\gamma^3 - \gamma^{2T+1}}{1 - \gamma^2} - \alpha^{T-1}\gamma^{T+2}\,\frac{1 - \left(\frac{\gamma}{\alpha}\right)^{T-1}}{1 - \frac{\gamma}{\alpha}}\right\} \qquad (2\text{-}50)$$

$X'C_2^2X$ must also be found. It is

$$(\gamma \;\; \gamma^2 \;\; \ldots \;\; \gamma^T) \begin{bmatrix} 0 & 0 & 1 & 2\alpha & 3\alpha^2 & 4\alpha^3 & \ldots & (T-2)\alpha^{T-3} \\ 0 & 0 & 0 & 1 & 2\alpha & 3\alpha^2 & \ldots & (T-3)\alpha^{T-4} \\ \vdots & & & & & & & \vdots \\ 0 & 0 & & & \ldots & & 1 & 2\alpha \\ 0 & 0 & & & \ldots & & 0 & 1 \\ 0 & 0 & & & \ldots & & 0 & 0 \end{bmatrix} \begin{pmatrix} \gamma \\ \gamma^2 \\ \vdots \\ \\ \gamma^T \end{pmatrix}$$

$$= (\gamma \;\; \gamma^2 \;\; \ldots \;\; \gamma^T) \begin{bmatrix} \gamma^3(1 + 2\alpha\gamma + 3(\alpha\gamma)^2 + \ldots + (T-2)(\alpha\gamma)^{T-3} \\ \gamma^4(1 + 2\alpha\gamma + 3(\alpha\gamma)^2 + \ldots + (T-3)(\alpha\gamma)^{T-4} \\ \vdots \\ \gamma^{T-1}(1 + 2\alpha\gamma) \\ \gamma^T \\ 0 \\ 0 \end{bmatrix}$$

$$= \gamma^4[1 + 2\alpha\gamma + 3(\alpha\gamma)^2 + \ldots + (T-2)(\alpha\gamma)^{T-3}] + \gamma^6[1 + 2\alpha\gamma + 3(\alpha\gamma)^2 + \ldots + (T-3)(\alpha\gamma)^{T-4}] + \ldots + \gamma^{2T-2} \; . \tag{2-51}$$

To sum the terms in (51), recall that the sum

$$\sum_{i=1}^{n} i r^{i-1} = \frac{d}{dr} \sum_{i=1}^{n} r^i = \frac{1 - (n+1)r^n + nr^{n+1}}{(1-r)^2} \; . \tag{2-52}$$

Using this formula (51) is

$$\frac{1}{(1-\alpha\gamma)^2}\{\gamma^4[1-(T-1)(\alpha\gamma)^{T-2}+(T-2)(\alpha\gamma)^{T-1}]+\gamma^6[1-(T-2)(\alpha\gamma)^{T-3}$$

$$+(T-3)(\alpha\gamma)^{T-2}]+\ldots+\gamma^{2T-2}[1-2\alpha\gamma+(\alpha\gamma)^2]\} \qquad (2\text{-}53)$$

$$=\frac{1}{(1-\alpha\gamma)^2}\left\{\frac{\gamma^4-\gamma^{2T}}{1-\gamma^2}-\gamma^{2T}\left[2\left(\frac{\alpha}{\gamma}\right)+3\left(\frac{\alpha}{\gamma}\right)^2+\ldots+(T-2)\left(\frac{\alpha}{\gamma}\right)^{T-3}+(T-1)\left(\frac{\alpha}{\gamma}\right)^{T-2}\right]\right.$$

$$\left.+\gamma^{2T}\alpha^2\left[1+2\left(\frac{\alpha}{\gamma}\right)+\ldots+(T-3)\left(\frac{\alpha}{\gamma}\right)^{T-4}-(T-2)\left(\frac{\alpha}{\gamma}\right)^{T-3}\right]\right\} \qquad (2\text{-}54)$$

This expression can be summed in a similar manner to give

$$X'C_2^2X=\frac{1}{(1-\alpha\gamma)^2}\left\{\frac{\gamma^4-\gamma^{2T}}{1-\gamma^2}-\frac{\gamma^{2T}}{\left(1-\frac{\alpha}{\gamma}\right)^2}\left[2\frac{\alpha}{\gamma}-\left(\frac{\alpha}{\gamma}\right)^2-T\left(\frac{\alpha}{\gamma}\right)^{T-1}+\left(\frac{\alpha}{\gamma}\right)^T\right]\right.$$

$$\left.+\frac{\gamma^{2T}\alpha^2}{\left(1-\frac{\alpha}{\gamma}\right)^2}\left[1-(T-1)\left(\frac{\alpha}{\gamma}\right)^{T-2}+(T-2)\left(\frac{\alpha}{\gamma}\right)^{T-1}\right]\right\} \qquad (2\text{-}55)$$

The last quantity which must be computed is $X'C_2C_2'X$. This is

$$\frac{1}{1-\alpha^2}(\gamma \;\; \gamma^2 \;\; \ldots \;\; \gamma^T)\begin{bmatrix} 1-\alpha^{2T-2} & \alpha-\alpha^{2T-3} & \alpha^2-\alpha^{2T-4} & \ldots & \alpha^{T-2}-\alpha^T & 0 \\ \alpha-\alpha^{2T-3} & 1-\alpha^{2T-4} & \alpha-\alpha^{2T-5} & \ldots & \alpha^{T-3}-\alpha^{T-1} & 0 \\ \vdots & & & & & \\ \alpha^{T-2}-\alpha^T & \alpha^{T-3}-\alpha^{T-1} & \alpha^{T-4}-\alpha^{T-2} & \ldots & 1-\alpha^2 & 0 \\ 0 & 0 & 0 & & 0 & 0 \end{bmatrix}\begin{pmatrix}\gamma \\ \gamma^2 \\ \\ \\ \gamma^T\end{pmatrix}$$

$$= \frac{1}{1-\alpha^2}(\gamma\ \gamma^2\ \ldots\ \gamma^T)\begin{bmatrix} \gamma + \gamma^2\alpha + \ldots + \gamma^{T-1}\alpha^{T-2} - (\gamma\alpha^{2T-2} + \gamma^2\alpha^{2T-3} + \ldots + \gamma^{T-1}\alpha^T) \\ \gamma\alpha + \gamma^2 + \ldots + \gamma^{T-1}\alpha^{T-3} - (\gamma\alpha^{2T-3} + \gamma^2\alpha^{2T-4} + \ldots + \gamma^{T-1}\alpha^{T-1}) \\ \vdots \qquad\qquad\qquad\qquad \vdots \\ \gamma\alpha^{T-2} + \gamma^2\alpha^{T-3} + \ldots + \gamma^{T-1} - (\gamma\alpha^T + \gamma^2\alpha^{T-1} + \ldots + \gamma^{T-1}\alpha^2) \\ 0 \end{bmatrix}$$

(2-56)

The product in (56) is the sum of two sets of terms, those involving the left-hand set of terms in the column vector and those involving the parenthesized terms in the vector. The second part of the product is easily computed. It is

$$\frac{-1}{1-\alpha^2}\left[\gamma^2\alpha^{2T-2}\left(1 + \frac{\gamma}{\alpha} + \left(\frac{\gamma}{\alpha}\right)^2 + \ldots + \left(\frac{\gamma}{\alpha}\right)^{T-2}\right) + \gamma^3\alpha^{2T-3}\left(1 + \frac{\gamma}{\alpha} + \ldots + \left(\frac{\gamma}{\alpha}\right)^{T-2}\right) + \ldots + \gamma^T\alpha^T\left(1 + \frac{\gamma}{\alpha} + \ldots + \left(\frac{\gamma}{\alpha}\right)^{T-2}\right)\right]$$

$$= \frac{-1}{1-\alpha^2}\,\frac{1-\left(\frac{\gamma}{\alpha}\right)^{T-1}}{1-\frac{\gamma}{\alpha}}\,\gamma^2\alpha^{2T-2}\left(1 + \frac{\gamma}{\alpha} + \ldots + \left(\frac{\gamma}{\alpha}\right)^{T-2}\right)$$

$$= \frac{-\gamma^2\alpha^{2T-2}}{(1-\alpha^2)}\,\frac{\left(1-\left(\frac{\gamma}{\alpha}\right)^{T-1}\right)^2}{\left(1-\frac{\gamma}{\alpha}\right)^2} \qquad (2\text{-}57)$$

The first part of the product in (56) is

$$\frac{1}{1-\alpha^2}\left\{\begin{array}{cccccc} \gamma^2 & + \gamma^3\alpha & + \gamma^4\alpha^2 & + \ldots & + \gamma^T\alpha^{T-2} & + \\ \gamma^3\alpha & + \gamma^4 & + \gamma^5\alpha & + \ldots & + \gamma^{T+1}\alpha^{T-3} & + \\ \gamma^4\alpha^2 & + \gamma^5\alpha & + \gamma^6 & + \ldots & + \gamma^{T+2}\alpha^{T-4} & + \\ \vdots & & & & \vdots & \\ \gamma^T\alpha^{T-2} & + \gamma^{T+1}\alpha^{T-3} & + \gamma^{T+2}\alpha^{T-4} & + \ldots & + \gamma^{2(T-1)} & + \end{array}\right\} . \quad (2\text{-}58)$$

Taking advantage of the symmetry of (58), it can be written as the sum of two geometric series. This gives

$$\frac{1}{1-\alpha^2}\left\{\frac{\gamma^2-\gamma^{2T}}{1-\gamma^2}+\frac{2\alpha}{1-\gamma\alpha}\left[\gamma^3(1-(\gamma\alpha)^{T-2})+\gamma^5(1-(\gamma\alpha)^{T-3})+\ldots+\gamma^{2T-3}(1-\gamma\alpha)\right]\right\}$$

which sums to

$$\frac{1}{1-\alpha^2}\left\{\frac{\gamma^2-\gamma^{2T}}{1-\gamma^2}+\frac{2\alpha}{1-\gamma\alpha}\left[\frac{\gamma^3-\gamma^{2T-1}}{1-\gamma^2}-\frac{\gamma^{T+1}\alpha^{T-2}\left(1-\left(\frac{\gamma}{\alpha}\right)^{T-3}\right)}{1-\frac{\gamma}{\alpha}}\right]\right\} \quad (2\text{-}59)$$

Collecting (57) and (59) gives for $X'C_2C_2'X$

$$\frac{-\gamma^2\alpha^{2T-2}}{1-\alpha^2}\frac{\left(1-\left(\frac{\gamma}{\alpha}\right)^{T-1}\right)^2}{\left(1-\frac{\gamma}{\alpha}\right)^2}+\frac{1}{1-\alpha^2}\left\{\frac{\gamma^2-\gamma^{2T}}{1-\gamma^2}+\frac{2\alpha}{1-\gamma\alpha}\right.$$

$$\left.\left[\frac{\gamma^3-\gamma^{2T-1}}{1-\gamma^2}-\frac{\gamma^{T+1}\alpha^{T-2}\left(1-\left(\frac{\gamma}{\alpha}\right)^{T-3}\right)}{1-\frac{\gamma}{\alpha}}\right]\right\} . \quad (2\text{-}60)$$

With these formulas in hand, the asymptotic bias of OLS can be numerically evaluated up to terms of order σ^4 . Table III presents the results of

TABLE III. THE APPROXIMATE BIAS OF OLS WITH A GEOMETRICALLY GROWING EXOGENOUS VARIABLE

$\beta = 1.0, \quad \sigma^2 = .5$

	T = 10			T = 20			T = 30			T = 40		
α \ γ	.9	1.05	1.15	.9	1.05	1.15	.9	1.05	1.15	.9	1.05	1.15
-.5	.00635	.0135	.00859	.000539	.00347	.000368	.0000653	.00133	.0000225	.0000079	.000509	.00000137
-.3	.00813	.0287	0	.000898	.0129	0	.000108	.00590	0	.0000131	.00181	0
.1	.0211	.105	.180	.00221	.0674	.0412	.000265	.0455	.00211	.000059	*	*
.3	.0346	.133	.129	.00362	.0635	.0109	.000433	.0322	.000686	.000053	.00917	.0000263
.5	.0555	.160	.108	.00672	.0712	.00870	.000802	.0306	.000537	.000097	.0124	.0000329
.6	.0164	.0833	.0310	.00990	.0855	.00949	.00122	.0360	.000609	.000148	.0143	.0000373
.7	*	*	*	.00577	.0783	.00221	.00200	.0482	.000722	.000262	.0195	.0000506

*Loss of precision in calculation.

these calculations for several values of alpha and gamma values of .9, 1.05, and 1.15. The results are calculated for sample sizes of 10, 20, 30, and 40. Since the bias of $\hat{\beta}$ is, from Corollary 5, proportional to $1/\beta$ and, therefore, the effects of varying beta are obvious, the calculations are done for only one value of beta. Beta is fixed at one and σ^2 is taken to be .5.

The first fact which can be observed from the table is that the bias decreases rapidly with increases in the sample size. By the time the sample size reaches thirty, the largest bias in the table is 4.8% and at a sample size of forty the largest bias is 2%. For a sample of ten observations the bias reaches 16% at $\alpha = .3$ and $\gamma = 1.05$ and is 18% for $\alpha = .1$ and $\gamma = 1.15$.

The effect on the bias of varying gamma can be substantial. For a sample of size ten and $\alpha = .1$, the bias varies from 2% to 18% as gamma varies from .9 to 1.15. For the same alpha and a sample size of twenty, the size of the bias at $\gamma = 1.05$ is thirty times the size of the bias at $\gamma = .9$. In most cases the largest biases occur at $\gamma = 1.05$ for all sample sizes and values of alpha.

As alpha varies from -.5 to .7, the bias generally rises until about .6 and then in some cases declines again, but the bias always remains positive for the cases calculated, even when alpha is negative.

SECTION 3

The results of the previous section can be used to find the properties of OLS under a type of misspecification of the equation estimated. The true equation is, as before,

$$y = y_{-1}\alpha + X\beta + \sigma u \ , \quad Euu' = \Omega \qquad (3\text{-}1)$$

but instead of estimating (1), OLS is used on the equation

$$y = y_{-1}\alpha + X\beta + \psi\gamma + \sigma u \qquad (3\text{-}2)$$

where ψ is an $NT \times \vartheta$ matrix of nonstochastic variables and is assumed to be of full rank $\vartheta \geq 1$. It is further assumed that the nonstochastic component of the right-hand side of (2), $(W \ X \ \psi)$ is of rank $1+k+\vartheta$. It then follows by the argument used earlier to establish the existence of the necessary inverse matrix in the estimation of (1) that for sufficiently small sigma the inverse matrix required in the estimation of (2) exists.

Estimation of (2) corresponds to including variables in the relationship which do not belong there. One case in which this type of error arises is in estimating a variance-components model by regressing, instead, the corresponding fixed-effects model with dummy variable regressors. This procedure will be analyzed after the general case of (2) is discussed.

If this were a classical regression model with nonstochastic regressors and spherical disturbances, the properties of the misspecified estimator would be clear. The estimator of the true parameters would remain unbiased, but there would be some loss in efficiency. This case is formally covered by Theil's analysis of specification error [21, 22] which, however, goes on to consider the case of variables incorrectly omitted from the list of regressors. However, where y_{-1} appears as a regressor and the disturbances do not satisfy the classical assumptions this analysis no longer applies. In fact, the misspecified estimator discussed below (LSC) may do better than the classical procedure.

Denote OLS applied to equation (2) when (1) is the true model as Least Squares under Misspecification (LSM). The asymptotic bias and mean squared error of LSM will be found from Corollaries 2-3 and 2-4. The LSM estimator of $(\alpha \quad \beta' \quad \gamma')$ is

$$\widehat{\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}}_{LSM} = \begin{bmatrix} (Z' + \sigma V^{*'})(Z + \sigma V^*) & (Z' + \sigma V^{*'})\psi \\ \psi'(Z + \sigma V^*) & \psi'\psi \end{bmatrix}^{-1} \begin{pmatrix} Z' + \sigma V^{*'} \\ \psi' \end{pmatrix} y \qquad (3\text{-}3)$$

where the reduced form of Lemma 2-1 has been substituted into the expression for the estimator. The systematic part of y_{-1} , W , is unchanged from the original model since equation (1) is still assumed to be the correct specification. V^* also remains the same.

The concern here is with the properties of $\hat{\alpha}$ and $\hat{\beta}$; therefore, an expression is derived for the error of these estimates alone. Using the formula for the inverse of a partitioned matrix, the first $k+1$ rows of the inverse matrix in (3) can be written as

$$[(Z' + \sigma V^{*\prime})\bar{P}_\psi(Z + \sigma V^*)]^{-1}[I - (Z' + \sigma V^{*\prime})\psi(\psi'\psi)^{-1}] \; . \qquad (3\text{-}4)$$

From (3) and (4) the error of the estimator for $(\alpha \;\; \beta')'$ is

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}_{LSM} = \sigma[(Z' + \sigma V^{*\prime})\bar{P}_\psi(Z + \sigma V^*)]^{-1}(Z + \sigma V^*)\bar{P}_\psi u \; . \qquad (3\text{-}5)$$

But (5) is just the error of the estimator given by (2-6) with the weighting matrix Σ^{-1} taken as $\bar{P}_\psi$. Therefore LSM is a GLS estimator of the form (2-5) using an incorrect weighting matrix. Then Theorems 1 and 2 apply directly to give the asymptotic bias and mean squared error of LSM as

<u>Corollary 1</u>:
$$E\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}_{LSM} = \sigma^2 Q\left[\begin{pmatrix} \mathrm{tr}\,\bar{M}_1 C \\ 0 \end{pmatrix} - Z'\bar{P}_\psi C'\bar{P}_\psi Z q^c\right]$$

$$+ \sigma^4 Q\left\{\begin{bmatrix} -q_{11}(\mathrm{tr}\,\bar{M}_1 G\;\mathrm{tr}\,\bar{M}_1 C + 2\,\mathrm{tr}\,\bar{M}_1 G\bar{M}_1 C + 2\,\mathrm{tr}\,\bar{M}_1 G\;\mathrm{tr}\,M_2 C \\ + 4\,\mathrm{tr}\,\bar{M}_1 G M_2 C + \mathrm{tr}\,M_2 G\;\mathrm{tr}\,\bar{M}_1 C + 2\,\mathrm{tr}\,M_2 G\bar{M}_1 C \\ 0 \end{bmatrix}\right.$$

$$+ Z'\bar{P}_\psi[q_2(2G\bar{M}_1(C+C') + 2C'\bar{M}_1 G + 2G\,\mathrm{tr}\,\bar{M}_1 C + C'\,\mathrm{tr}\,\bar{M}_1 C)$$

$$\left. - 2GM_2C' - C'\,\mathrm{tr}\,M_2 G]\bar{P}_\psi Z q^c\right\} + O(\sigma^6) \; .$$

The superscripts on Q and its components, $\bar{M}_1$ and M_2 which indicate that the weighting matrix is $\bar{P}_\psi$ are omitted in this corollary and the following one to avoid excessive notational complexity.

<u>Corollary 2</u>: $E(ee')_{LSM} = \sigma^2 QZ'\bar{P}_\psi \Omega \bar{P}_\psi ZQ$

$$+ \sigma^4 Q \left\{ \begin{bmatrix} (\text{tr } \bar{M}_1 C)^2 + \text{tr } \bar{M}_1 C \bar{M}_1 C + \text{tr } \bar{M}_1 G \bar{M}_1 \Omega & 0 \\ 0 & 0 \end{bmatrix} \right.$$

$$- Z'\bar{P}_\psi [(C+C')\text{tr } \bar{M}_1 C + C\bar{M}_1 C \qquad + C'\bar{M}_1 C' + 2C\bar{M}_1 C' + G\bar{M}_1 \Omega + \Omega \bar{M}_1 G$$

$$+ \Omega \text{ tr } \bar{M}_1 G] \bar{P}_\psi ZQ^*$$

$$- Q^{*\prime} Z'\bar{P}_\psi [(C+C')\text{tr } \bar{M}_1 C + C\bar{M}_1 C + C'\bar{M}_1 C' + 2C\bar{M}_1 C' + G\bar{M}_1 \Omega + \Omega \bar{M}_1 G$$

$$+ \Omega \text{ tr } \bar{M}_1 G] \bar{P}_\psi Z$$

$$+ Z'\bar{P}_\psi [(C+C')\text{tr } M_2 C + CM_2 C + C'M_2 C' + 2CM_2 C' + GM_2 \Omega + \Omega M_2 G + G \text{ tr } M_2 \Omega$$

$$\left. - q_{11}((C+C')\text{tr } \bar{M}_1 C + C\bar{M}_1 C + C'\bar{M}_1 C' + G\bar{M}_1 \Omega + \Omega \bar{M}_1 G)] \bar{P}_\psi Z \right\} Q + O(\sigma^6) \ .$$

It is difficult to say much about these corollaries, but if some structure is imposed on ψ, further results can be obtained. One way to estimate the variance components model of equation (1) with error covariance matrix $\Omega = I_N \otimes (\rho i_T i_T' + (1-\rho) I_T)$, where i_T is a $T \times 1$ vector of ones, is to estimate the corresponding fixed effects model. That is, to let $\psi = I_N \otimes i_T$ be a matrix of dummy variables, one for each individual in the cross-section, i.e., $\psi_{it,j} = 1$, $i = j$ and is zero otherwise where $\psi_{it,j}$ is the element of ψ in column j and row it corresponding to the observation on individual i at time t. Clearly the columns of ψ span the N-dimensional space of the columns of $I_N \otimes i_T i_T'$. Denote this estimator as LSC.

The projection matrix $\bar{P}_\psi$ is easily computed as

$$I_{NT} - (I_N \otimes i_T)[(I_N \otimes i_T)'(I_N \otimes i_T)]^{-1}(I_N \otimes i_T)'$$

$$= I_{NT} - I_N \otimes \frac{i_T i_T'}{T} \quad . \qquad (3\text{-}6)$$

Let $\tilde{Z} = \bar{P}_\psi Z$. Then $\tilde{Z}$ is

$$\begin{pmatrix} z_{11} - \bar{z}_{1\cdot} \\ \vdots \\ z_{1T} - \bar{z}_{1\cdot} \\ \vdots \\ z_{N1} - \bar{z}_{N\cdot} \\ \vdots \\ z_{NT} - \bar{z}_{N\cdot} \end{pmatrix} \text{ where } \bar{z}_{i\cdot} = \frac{1}{T}\sum_{j=1}^{T} z_{ij} \ .$$

Therefore, $\tilde{Z}'\tilde{Z}$ is the within group sum of squares matrix (WSS) where the group is the set of T observations on individual i . Then M_1 becomes $\bar{P}_\psi Z(Z'\bar{P}_\psi Z)^{-1}Z'\bar{P}_\psi$ which is $B_{\tilde{Z}}$, the projection on the nonstochastic portion of the regressors of (3-1) where the interindividual variation has been removed.

Some of the other terms occurring in the bias and mean-squared error expressions can also be further analyzed. Let $a' = (a^0 \ \ a^1 \ \ a^2 \ \ \ldots \ \ a^{T-1})'$. Recall that $C = I_N \otimes (\rho C_1 + (1-\rho)C_2)$. $C_1 = i_T(i_T - a)'$; therefore $\bar{B}_{\tilde{Z}}(I_N \otimes C_1) = I_N \otimes C_1$ as $I_N \times i_T$ is orthogonal to $\tilde{Z}$ and $\text{tr}\ \bar{B}_{\tilde{Z}}(I_N \otimes C_1) = N\left[T - \frac{1-\alpha^T}{1-\alpha}\right]$. Similarly, since

$$G = I_N \otimes (\rho G_1 + (1-\rho)G_2) \quad \text{and}$$

$$G_1 = \frac{1}{1-\alpha}(i_T - a)(i_T - a)' ,$$

$$\bar{P}_{\mathcal{Z}} G_1 = \frac{1}{1-\alpha}(i_T - \bar{P}_{\mathcal{Z}} a)(i_T - a)' .$$

Further note that $M_2 = \mathcal{Z} q^c q^r \mathcal{Z}' = q_{11}(P_{\mathcal{Z}} - P_{\mathcal{X}})$. To restate Corollary 1 for the case where $\psi = I_N \otimes i_T$ consider the terms of order σ^2 in that corollary. They are, except for the factor Q ,

$$\sigma^2 \left[\begin{pmatrix} \text{tr } \bar{M}_1 C \\ 0 \end{pmatrix} - Z'\bar{P}_\psi C'\bar{P}_\psi Z q^c \right] \tag{3-7}$$

$$= \sigma^2 \left[\begin{pmatrix} \text{tr } \bar{P}_{\mathcal{Z}}[I_N \otimes (\rho C_1 + (1-\rho)C_2)] \\ 0 \end{pmatrix} - \mathcal{Z}'C'\mathcal{Z} q^c \right]$$

which becomes

$$\sigma^2 \left[\begin{pmatrix} \rho N\left(T - \frac{1-\alpha^T}{1-\alpha}\right) - (1-\rho)\text{tr } P_{\mathcal{Z}}(I_N \otimes C_2) \\ 0 \end{pmatrix} - (1-\rho)\mathcal{Z}'(I_N \otimes C_2')\mathcal{Z} q^c \right] \tag{3-8}$$

since $\mathcal{Z}$ is orthogonal to C_1 and $\text{tr } C_2 = 0$.

$$\text{tr } P_{\mathcal{Z}}(I_N \otimes C_2) = \text{tr } \mathcal{Z}'(I_N \otimes C_2)\mathcal{Z} Q$$

$$= \mathcal{W}'(I_N \otimes C_2)\mathcal{Z} q^c + \text{tr } \mathcal{X}'(I_N \otimes C_2)\mathcal{Z} q_2 . \tag{3-9}$$

Substituting (9) into (8) gives the bias of the LSC estimator as

Corollary 3:

$$E\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}_{LSC} = \sigma^2 Q \begin{bmatrix} \rho N\left(T - \frac{1-\alpha^T}{1-\alpha}\right) - (1-\rho)[2\mathbb{W}'(C_2^* + C_2^{*\prime})\mathbb{Z}q^c + \text{tr } \mathbb{X}'C_2^*\mathbb{Z}q_2] \\ - (1-\rho)\mathbb{X}'C_2^*\mathbb{Z}q^c \end{bmatrix} + O(\sigma^4)$$

(3-10)

where $C_2^* = I_N \otimes C_2$.

The signs of the biases given by (10) are ambiguous, but the factors affecting the bias can be seen. The effect of varying ρ can be seen from (10). Here Q does not depend on ρ because the weighting matrix is fixed. As ρ becomes larger the terms in ρ become more important than those in $(1-\rho)$ and when ρ is near one, the bias will be approximately

$$E\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \sigma^2 Q \begin{bmatrix} N\left(T - \frac{1-\alpha^T}{1-\alpha}\right) \\ 0 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} q_{11}N\left(T - \frac{1-\alpha^T}{1-\alpha}\right) \\ q_{21}N\left(T - \frac{1-\alpha^T}{1-\alpha}\right) \end{bmatrix} . \qquad (3\text{-}11)$$

Since q_{11} is positive (Q is positive definite) the sign of the bias of $\hat{\alpha}$ will be independent of N and except for an interval for α near one whose size depends on T , the bias will be positive.

The bias in $\hat{\beta}$ will depend also on the sign of q_{21} which is a function of the behavior of the exogenous variables and the signs of α and β. It can be seen most clearly in the case where ρ is near one how the different dimensions of the sample, N and T, affect the properties of the estimator differently even apart from the different statistical properties which the exogenous variables will have in the different dimensions. Since adding more observations in the cross-section does not compound the interaction between the lagged variable over time and the disturbance structure, it is intuitively reasonable to expect its effects to be simpler than the time-series effects, which is clearly the case in expression (13).

When ρ is different from one, the analysis is more difficult. The expression (10) contains terms of opposing signs. Consider the second row of (10). It is, again ignoring y_0

$$-(1-\rho)\tilde{X}'C_2^*\tilde{Z}q^c$$

$$= -(1-\rho)[(\tilde{X}'(I_N \otimes C_2C_2')\tilde{X}\beta \quad \tilde{X}'(I_N \otimes C_2\tilde{X}]q^c \tag{3-12}$$

recalling that $\tilde{W} = (I_N \otimes C_2')\tilde{X}\beta + \tilde{y}_0 \otimes a$, C_2C_2' is positive semidefinite which means that $\tilde{X}'(I_N \otimes C_2C_2')\tilde{X}$ is positive semidefinite. The second term of (12) is, however, indefinite since C_2 was shown above to be indefinite. The vector q^c has a positive first element, but the other elements could have either sign.

The first row of (10) also contains terms whose effect is ambiguous. Also, in comparing $\tilde{W}'\ C_2^*\ \tilde{Z}$ with $\tilde{X}'\ C_2^*\ \tilde{Z}$, it should be noted that the first expression may be more sensitive to higher order time correlations among the exogenous variables than the second term. For where terms like $\mathfrak{X}'\ C_2^*\ \mathfrak{X}$ appear in the second expression, the first will contain instead $\mathfrak{X}'\ C_2^{*2}\ \mathfrak{X}$. In computing the second expression more weight is put on cross-products of terms farther apart in time.

It is clear from Corollary 3 that the bias of LSC, as is also the case for the other estimators, will be significantly affected both by the values of the parameters of the model and by the statistical properties of the exogenous variables; the interactions of these factors will be complex.

The asymptotic mean squared error of LSC follows from Corollary 2 using the relationships

$$\overline{P}_{\psi}\Omega = (1-\rho)\overline{P}_{\psi} \qquad \text{and} \qquad \overline{P}_{\psi}C = (1-\rho)\overline{P}_{\psi}C_2^* .$$

<u>Corollary 4</u>: $E(ee')_{LSC} = \sigma^2(1-\rho)(\mathfrak{Z}'\mathfrak{Z})^{-1}$

$$+ \sigma^4(\mathfrak{Z}'\mathfrak{Z})^{-1}\left\{\begin{bmatrix} (\text{tr } \overline{P}_{\mathfrak{Z}}C)^2 + \text{tr } \overline{P}_{\mathfrak{Z}}C\overline{P}_{\mathfrak{Z}}C + \text{tr } \overline{P}_{\mathfrak{Z}}G\overline{P}_{\mathfrak{Z}}C & 0 \\ 0 & 0 \end{bmatrix}\right.$$

$$- (1-\rho)\mathfrak{Z}'[(C_2^* + C_2^{*\prime})\text{tr } \overline{P}_{\mathfrak{Z}}C + (1-\rho)(C_2^*\overline{P}_{\mathfrak{Z}}C_2^* + C_2^{*\prime}\overline{P}_{\mathfrak{Z}}C_2^{*\prime} + 2C_2^*\overline{P}_{\mathfrak{Z}}C_2^{*\prime}) + \overline{P}_{\mathfrak{Z}}G$$

$$+ G\overline{P}_{\mathfrak{Z}} + I\ \text{tr } \overline{P}_{\mathfrak{Z}}G]\mathfrak{Z}Q^*$$

$$- (1-\rho)Q^{*\prime}\mathfrak{Z}'[(C_2^* + C_2^{*\prime})\text{tr } \overline{P}_{\mathfrak{Z}}C + (1-\rho)(C_2^*\overline{P}_{\mathfrak{Z}}C_2^* + C_2^{*\prime}\overline{P}_{\mathfrak{Z}}C_2^{*\prime} + 2C_2^{*\prime}\overline{P}_{\mathfrak{Z}}C_2^*)$$

$$+ \overline{P}_{\mathfrak{Z}}G + G\overline{P}_{\mathfrak{Z}}]\mathfrak{Z} +$$

$$q_{11}\mathfrak{Z}'[(1-\rho)(C_2^* + C_2^{*\prime})\text{tr } (P_{\mathfrak{Z}} - P_{\mathfrak{X}})C + (1-\rho)^2(C_2^*(P_{\mathfrak{Z}} - P_{\mathfrak{X}})C_2^*$$

$$+ C_2^{*\prime}(P_{\mathfrak{Z}} - P_{\mathfrak{X}})C_2^{*\prime} + 2C_2^*(P_{\mathfrak{Z}} - P_{\mathfrak{X}})C_2^{*\prime}) + (1-\rho)(G(P_{\mathfrak{Z}} - P_{\mathfrak{X}})$$

$+ (P_{\tilde{Z}} - P_{\tilde{X}})G) + G\ \mathrm{tr}(P_{\tilde{Z}} - P_{\tilde{X}})\Omega$

$- q_{11}(1-\rho)((C_2^* + C_2^{*\prime})\mathrm{tr}\ P_{\tilde{Z}}G + (1-\rho)(C_2^*\bar{P}_{\tilde{Z}}C_2^* + C_2^{*\prime}\bar{P}_{\tilde{Z}}C_2^{*\prime}) + G\bar{P}_{\tilde{Z}} + \bar{P}_{\tilde{Z}}G)]\tilde{Z}\}(\tilde{Z}'\tilde{Z})^{-1}$

$+ O(\sigma^6)$.

The mean squared errors of GLS and LSC can be compared using Corollaries 2-2 and 3-4. The terms of order σ^2 in these expressions are $\sigma^2(Z'\Omega^{-1}Z)^{-1}$ and $\sigma^2(1-\rho)(Z'\bar{P}_\psi Z)^{-1}$, respectively. Ω^{-1} can be written as $I_N \otimes \mathcal{A}^{-1}$ when $\rho \neq 1$ where $\mathcal{A}^{-1}$ is $\left(I_{NT} - \frac{\rho}{1 + (T-1)\rho} I_N \otimes i_T i_T'\right)$ (see Nerlove [16]). $\bar{P}_\psi$ is $\left(I_{NT} - \frac{1}{T} I_N \otimes i_T i_T'\right)$. The difference between GLS and LSC is reflected in the difference between $\frac{\rho}{1 + (T-1)\rho}$ and $\frac{1}{T}$.

The two factors are equal when ρ equals one (but Ω is singular at this point). The estimators differ in the way they use the between group variation in estimating α and β . The LSC estimator uses only the within group variation while GLS uses both with the relative weights depending on ρ and T ; the mean squared errors reflect this difference. This analysis is carried further in Nerlove [16].

SECTION 4

Since the heart of the difficulty with OLS in the model where nonspherical disturbances are present and the lagged dependent variable occurs as a regressor is the correlation between y_{-1} and the error vector, an alternative to OLS which attempts to deal directly with this problem is worth considering. One could deal with this problem by using an instrumental variable for the lagged dependent variable. The practical difficulty, of course, is where to find a satisfactory instrument for y_{-1}.

Some lagged exogenous variable could be used, or perhaps a linear combination of the lagged variables. The adequacy of such an estimator will depend on the correlation between the lagged variables and the lagged endogenous variable and on the autocorrelation of the regressors. Therefore, statements about the relative desirability of this estimator without reference to the properties of the regressors should not be expected.

Even in the best of estimation problems the instrumental variables approach is justified by large-sample-asymptotic properties such as consistency although the estimator may be badly biased in finite samples (See Sargan [18]).

As was true for the other estimators considered, the theorems found for IV on the assumption that $\Omega = I_N \otimes \mathcal{Q}$ are valid for models with an arbitrary nonsingular covariance matrix Ω if the matrices C and $\mathcal{C}$ are appropriately recalculated.

The relationship to be estimated is, as before,

$$y = y_{-1}\alpha + X\beta + \sigma u \ . \qquad (4\text{-}1)$$

Suppose that a nonstochastic instrument

$$d = (d_{11}, d_{12}, \ldots, d_{1T}, d_{21}, \ldots, d_{2T}, \ldots, d_{N1}, \ldots, d_{NT})'$$

is available for the lagged dependent variable and let $D = (d \;\; X)$. The instrument is assumed to have the properties that

$$E(u|d) = 0 \text{ for all values of } d \text{ , and} \qquad (4\text{-}2a)$$

$$\text{the matrix } D'Z \text{ is of full rank.} \qquad (4\text{-}2b)$$

Thus d is related to the nonstochastic part of y_{-1} , which is W , and is not related to V , the stochastic part of y_{-1} . A lagged vector of an exogenous regressor will satisfy these assumptions in general (except for the constant term, of course). The assumptions employed in the theorems on GLS are also maintained for this estimator. Define H to be the $k+1 \times k+1$ matrix $(D'Z)^{-1}$ and denote the first column of H as h^c and its first row as h^r . The instrumental variables estimator (IV) is

$$\widehat{\begin{pmatrix} \alpha \\ \beta \end{pmatrix}}_{IV} = [D'(y_{-1} \;\; X)]^{-1}D'y \ . \qquad (4\text{-}3)$$

Separating the stochastic and nonstochastic parts of y_{-1} into W and V as before gives for the error of the IV estimator

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}_{IV} = \sigma \begin{bmatrix} d'(W + \sigma V) & d'X \\ X'(W + \sigma V) & X'X \end{bmatrix}^{-1} \begin{bmatrix} d' \\ X' \end{bmatrix} u \qquad (4\text{-}4)$$

$$= \sigma(D'Z + \sigma D'V^*)^{-1} D'u \; . \qquad (4\text{-}5)$$

For a sufficiently small value of σ the inverse matrix in (5) exists by the same type of argument which guaranteed the existence of corresponding inverses for the other estimators. Then writing the inverse matrix in (5) as $H(I + \sigma D'V^*H)^{-1}$ and approximating it by expanding the inverse term as a power series in σ gives an approximate expression for the error of IV as

<u>Lemma 1</u>: $$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}_{IV} = H(\sigma D'u - \sigma^2 D'V^*HD'u$$

$$+ \sigma^3 D'V^*HD'V^*HD'u - \sigma^4 D'V^*HD'V^*HD'V^*HD'u) + O_p(\sigma^5) \; .$$

Lemma 1 will be used to find the approximate bias and mean squared error of IV. The bias is found as

<u>Theorem 1</u>: Under assumptions (θ) and 2

$$E\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}_{IV} = -\{\sigma^2 D' + \sigma^4 h_{11} HD'[2GS + I \text{ tr } GS]\} C'Dh^{r'} + O(\sigma^6)$$

where $S = DHD' - P_X$.

The proof proceeds, as in the GLS proof, by taking the expectation of each term in Lemma 1.

The expectation of the first term in Lemma 1 is

$$E\sigma HD'u = 0 \tag{4-6}$$

since H and D are nonstochastic. The expectation of the term of order σ^2 is

$$\begin{aligned} -E\sigma^2 D'V^* HD'u &= -\sigma^2 D'EVh^r D'u \\ &= -\sigma^2 D'EVu'Dh^{r'} \\ &= -\sigma^2 D'C'Dh^{r'} \end{aligned} \tag{4-7}$$

where the second line is valid because $h^r D'u$ is a scalar and thus equals its transpose. The third term of Lemma 1 has zero expectation since it is an odd product of normal variables each with mean zero. The last term in Lemma 1 has expectation

$$\begin{aligned} -\sigma^4 EHD'V^*HD'V^*HD'V^*HD'u &= -\sigma^4 HD'EVh^r D'Vh^r D'Vh^r D'u \\ &= -\sigma^4 HD'EV^*V^{*'}Dh^{r'} h^r D'V^*HD'u \ . \end{aligned} \tag{4-8}$$

Then by Lemma A10, (8) is

$$-\sigma^4 HD'[2GDh^{r'} h^r D' + I \text{ tr } GDh^{r'} h^r D']C'Dh^{r'} \ . \tag{4-9}$$

The expression $Dh^{r'} h^r D'$ can be more informatively expressed as

$$h_{11} D\left[\begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & h_{22} - \dfrac{h_{21}h_{12}}{h_{11}} \end{pmatrix}\right] D' \tag{4-10}$$

where $H = \begin{pmatrix} \underset{1\times 1}{h_{11}} & \underset{1\times k}{h_{12}} \\ \underset{k\times 1}{h_{21}} & \underset{k\times k}{h_{22}} \end{pmatrix}$.

Expression (10) equals

$$h_{11}(DHD' - P_X) \tag{4-11}$$

since $h_{22} - \frac{h_{21}h_{12}}{h_{11}}$ is $(X'X)^{-1}$.

The term of order σ^5 are not shown in Lemma 1, but it would clearly be an odd product of normal variables again and would, therefore, have expectation zero. Finally, combining (6), (7), and (9) produces the theorem.

The effect of the intraclass correlation on the bias can be seen by rewriting Theorem 1. Considering only the terms of order σ^2 gives

$$E\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}_{IV} = -\sigma^2[D'(I_N \otimes C_2')Dh^{r'} + \rho D'(I_N \otimes (C_1' - C_2'))Dh^{r'}] + O(\sigma^4) \tag{4-12}$$

$$= -\sigma^2 D'[I_N \otimes (C_2' + \rho(C_1' - C_2'))]Dh^{r'} . \tag{4-13}$$

Clearly, in this approximation, the bias changes linearly with ρ , but the terms of order σ^4 contribute a factor which depends on ρ^2 . The matrix in square brackets in (13) is indefinite for any values of ρ since each of the components of it is indefinite. There are no values of ρ and α for which that matrix is the zero matrix. Therefore, the bias of the IV estimator will depend on the statistical properties of both the instrument and the exogenous variables; even the direction of the bias is indeterminate.

Expression (13) can be written more explicitly in terms of X , d , W , and C . It is trivial to show that

$$h^r = \left(\frac{1}{d'W} - \left[X'\left(I - \frac{Wd'}{d'W} \right) X \right]^{-1} \frac{d'XX'W}{(d'W)^2} \left[X'\left(I - \frac{Wd'}{d'W} \right) X \right]^{-1} \frac{d'X}{d'W} \right) . \quad (4\text{-}14)$$

Using this, the bias of the IV estimate of $\hat{\alpha}$ is

$$-\sigma^2 \left\{ d'Cd \left[\frac{1}{d'W} - \left[X'\left(I - \frac{Wd'}{d'W} \right) X \right]^{-1} \frac{d'XX'W}{(d'W)^2} \right] + d'CX \frac{X'd}{d'W} \left[X'\left(I - \frac{dW'}{d'W} \right) X \right]^{-1} \right\}$$

$$+ O(\sigma^4) . \quad (4\text{-}15)$$

Expression (15) makes clear the way in which the bias depends on the properties of the variables. The correlations between d and W , and d and X enter through $d'W$ and $d'X$ while serial correlation properties of d and X enter through $d'Cd$ and $d'CX$ with weights depending on α .

The bias is not the appropriate criterion to employ in choosing an estimator, so the mean squared error of the Instrumental Variables estimator will now be derived.

Theorem 2: Under assumptions (θ) and 2, the mean squared error of IV is

$$E(ee')_{IV} = \sigma^2 HD'\Omega DH'$$

$$+ \sigma^4 HD'\{CSC + C'SC' + (C+C')\text{tr } SC + \Omega SG + GS\Omega$$

$$+ 2C'SC + G \text{ tr } S\Omega\}DH' + O(\sigma^6) .$$

<u>Proof</u>: The squared error of the estimator is, from Lemma 1

$$(ee')_{IV} = \sigma^2 HD'uu'DH' - \sigma^3(HD'uu'DH'V^{*\prime}D - HD'V^*HD'uu'DH)$$

$$+ \sigma^4(HD'uu'DH'V^{*\prime}DH'V^{*\prime}DH' + HD'V^*HD'V^*HD'V^*HD'uu'DH' \tag{4-16}$$

$$+ HD'V^*HD'uu'DH'V^{*\prime}DH') + o_p(\sigma^5) .$$

The second term of order σ^4 is the transpose of the first term.

The proof proceeds by finding the expectation of (16) term by term. The expectation of the first term of (16) is

$$E\sigma^2 HD'uu'DH' = \sigma^2 HD'\Omega DH' . \tag{4-17}$$

The terms of order σ^3 have expectation zero since they are odd products of normal variables. The terms of order σ^4 are evaluated as follows

$$E\sigma^4 HD'uu'DH'V^{*\prime}DH'V^{*\prime}DH'$$

$$= \sigma^4 HD'[CDh^{r\prime}h^r D'C + C \text{ tr } Dh^{r\prime}h^r D' + \Omega Dh^{r\prime}h^r D'G]DH' \tag{4-18}$$

from an application of Lemma A7. The second σ^4 term is the transpose of the first. The expectation of the remaining σ^4 term is found from Lemma A6 as

$$EHD'V^*HD'uu'DH'V^{*\prime}DH'$$

$$= HD'[2C'Dh^{r\prime}h^r D'C + G \text{ tr } Dh^{r\prime}h^r D'\Omega]DH' . \tag{4-19}$$

The terms of order σ^5, which are not shown in Lemma 1, are odd products of normal variables and, therefore, have zero expectation. Finally, collecting (17), (18), and (19) produces the theorem.

This theorem is considerably simpler than the corresponding theorems for the other estimators. The first term, $\sigma^2(D'Z)^{-1}D'\Omega D(Z'D)^{-1}$, is the same as the large sample approximation of the variance-covariance matrix of IV for the case where all the regressors are nonstochastic. The expectations matrices C and G, which arise from the stochastic part of the reduced form, V, enter in the σ^4 term.

Theorem 2 suggests a way of choosing which variable or linear combination of variables to use as an instrument for y_{-1}. Suppose, first, that $\Omega = I$. Then the best instrument up to terms of order σ^4, were it available, would be W under the criterion of minimum mean squared error of the estimator. That is, the difference of the mean squared errors of the IV estimator using an arbitrary instrument and the IV estimator using W is positive semidefinite. For the difference is

$$\sigma^2(D'Z)^{-1}D'D(Z'D)^{-1} - \sigma^2(Z'Z)^{-1} . \qquad (4\text{-}20)$$

If (20) is to be positive semidefinite, then

$$(Z'D)(D'D)^{-1}(D'Z) - (Z'Z) \qquad (4\text{-}21)$$

must be negative semidefinite. But (21) is $-Z'\overline{P}_D Z$ which clearly satisfies that condition. Of course, W cannot be used since it is unobservable, but that observable variable which is most like W should be used. The IV estimator might be iterated using an estimate of W as the instrument although there is no guarantee that the iterated estimator would be optimal in finite samples. This would be similar to the two-round procedures suggested by Nerlove in [15].

When the covariance matrix Ω is not an identity matrix the optimality of W, even if it were available, is no longer clear. In this case inefficiency still remains, as for OLS, because IV ignores Ω. Then the optimal choice of the instrument will also involve Ω. Minimizing $\sigma^2(D'Z)^{-1}D'\Omega D(Z'D)^{-1}$ corresponds to maximizing its inverse

$$Z'D(D'\Omega D)^{-1}D'Z \ . \tag{4-22}$$

Let R be the nonsingular matrix for which $R'\Omega R = I$ and let $\Delta = RD$. Then (22) becomes $Z'RP_{\Delta}R'Z$. This expression is maximized for $\Delta = R'Z$ or $D = R^{-1}R'$. R is the matrix of characteristic vectors of Ω transformed by dividing each vector by the square root of its characteristic value. The characteristic values of Ω are $1-\rho$ and $1 + (T-1)\rho$ and the corresponding vectors are given in Nerlove [16]. The optimal instrument vector D is still unobservable because W is unknown and because ρ is not generally known. This suggests that IV will do best when the instrument chosen is a hybrid of the best "true" IV choice and an appropriate choice involving generalized least squares considerations.

One might be able to do still better in choosing an instrument vector by considering the terms of order σ^4 in Theorem 2, but this would be difficult to do in practice.

CONCLUSION

Four estimators have been analyzed in this study by means of small-sigma asymptotic approximations. The asymptotic bias and mean squared error have been found for Generalized Least Squares, Ordinary Least Squares, Least Squares with Constants, and Instrumental Variables. The expressions obtained are in most cases extremely complex and depend importantly on the unknown parameters α, β, and ρ as well as on the statistical properties of the independent variables. It seems unlikely that strong conclusions can be drawn about the relative suitability of the various estimators without explicit consideration of the unknown parameter values. For a particular problem, if no prior knowledge on the parameters is available, a consistent estimator could be used to estimate the parameters and then the mean squared error expressions could be evaluated at these estimates to choose the best procedure. Whether this would be worth the computational effort involved is uncertain, particularly since several estimators which seem intuitively appealing have not been evaluated here because of the complexity of the expectations which would have to be analyzed.

Other types of approximations were explored for some of these estimators, such as large and small ρ expansions and expansions in the characteristic roots of Ω, but they either gave rise to intractable expressions or discarded too much of the expressions and produced uninteresting results. The ultimate usefulness of these kinds of expansions in analyzing the estimation of dynamic relationships requires further study, but it is clear that the problem is a difficult one. Large-sample asymptotics, however, do not distinguish among some estimators which may be quite different in finite samples, and Monte Carlo studies may fail to adequately explore the appropriate parameter variations. This study points up the necessity of considering a variety of types of independent variables before drawing firm conclusions about the choice of an estimator for this problem.

BIBLIOGRAPHY

[1] Amemiya, Takeshi. "A Note on the Estimation of Balestra-Nerlove Models," Technical Report 4, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1967.

[2] _________ and Wayne Fuller. "A Comparative Study of Alternative Estimators in a Distributed Lag Model," *Econometrica*, Vol. 35, No. 3-4, 1967.

[3] Anderson, T.W. *Introduction to Multivariate Statistical Analysis*. New York: John Wiley and Sons, Inc., 1958.

[4] Balestra, Pietro and Marc Nerlove. "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," *Econometrica*, Vol. 34, No. 3, 1966.

[5] Griliches, Zvi. "Distributed Lags: A Survey," *Econometrica*, Vol. 35, No. 1, 1967.

[6] Hurwicz, Leonid. "Least Squares Bias in Time Series," in *Statistical Inference in Dynamic Economic Models*, edited by Tjalling C. Koopmans. New York: John Wiley and Sons, Inc., 1950.

[7] Johnston, J. *Econometric Methods*. New York: McGraw Hill Book Company, Inc., 1963.

[8] Kadane, Joseph B. "Testing Overidentifying Restrictions when the Disturbances are Small," *Journal of the American Statistical Association*, Vol. 65, No. 329, 1970.

[9] _________. "Comparison of k-Class Estimators when the Disturbances are Small," *Econometrica* (forthcoming).

[10] Lingren, B.W. *Statistical Theory*, Second Edition. New York: The Macmillan Company, 1968.

[11] Maddala, G.S. "Generalized Least Squares with an Estimated Variance Covariance Matrix," *Econometrica*, Vol. 39, No. 1, 1971.

[12] _________. "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," *Econometrica*, Vol. 39, No. 2, 1971.

[13] Nerlove, Marc. Distributed Lags and Demand Analysis for Agricultural and Other Commodities. Washington, D.C.: U.S. Department of Agriculture, Agricultural Handbook, No. 141, 1958.

[14] ________. "Experimental Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections," Economic Studies Quarterly, Vol. 18, 1967.

[15] ________. "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections," Econometrica, Vol. 39, No. 2, 1971.

[16] ________. "A Note on Error Components Models," Econometrica, Vol. 39, No. 2, 1971.

[17] Office of Economic Opportunity. Further Preliminary Results of the New Jersey Graduated Work Incentive Experiment, Office of Economic Opportunity, Washington, D.C., 1971.

[18] Sargan, J.D. "The Estimation of Economic Relationships Using Instrumental Variables," Econometrica, Vol. 26, 1958.

[19] Scheffé, Henry. "Alternative Models for the Analysis of Variance," Annals of Mathematical Statistics, Vol. 27, 1956.

[20] ________. The Analysis of Variance. New York: John Wiley and Sons, Inc., 1959.

[21] Theil, Henri. Principles of Econometrics. New York: John Wiley and Sons, Inc., 1971.

[22] ________. "Specification Errors and the Estimation of Economic Relationships," Review of the International Statistical Institute, Vol. 25, 1957.

[23] U.S. Bureau of the Census. Current Population Reports, Washington, D.C.

[24] Wallace, T.D. and Ashiq Hussain. "The Use of Error Components Models in Combining Cross Section with Time Series Data," Econometrica, Vol. 37, No. 1, 1969.

APPENDIX

This appendix contains the derivations of the lemmas on expectations used in the body of the thesis. Lemmas A1, A2, and A3 depend on the specific form of the covariance matrix, Ω, assumed for the distribution of the disturbances in the model, and would, therefore, have to be recalculated under alternative assumptions on Ω. The other lemmas depend only on the normality of the errors and the assumption that they have a zero mean. That is, they hold for any alternative assumptions about Ω if the expectations in Lemmas A1 and A3 below are appropriately recalculated.

<u>Lemma A1</u>: $$EuV' \equiv C = I_N \otimes (\rho C_1 + (1-\rho)C_2)$$

where

$$C_1 = \frac{1}{1-\alpha}\begin{bmatrix} 0 & 1-\alpha & 1-\alpha^2 & & 1-\alpha^{T-1} \\ 0 & 1-\alpha & 1-\alpha^2 & \dots & 1-\alpha^{T-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 1-\alpha & 1-\alpha^2 & & 1-\alpha^{T-1} \end{bmatrix} = i_T[1-\alpha^0 \quad 1-\alpha^1 \quad 1-\alpha^2 \quad \dots \quad 1-\alpha^{T-1}]$$

and

$$C_2 = \begin{bmatrix} 0 & 1 & \alpha & \alpha^2 & \cdots & \alpha^{T-2} \\ 0 & 0 & 1 & \alpha & \cdots & \alpha^{T-3} \\ \vdots & & & \vdots & & \vdots \\ 0 & 0 & . & . & . & 0 \end{bmatrix}$$

C_1 and C_2 are TxT . C_1 is the component of C due to the nonzero intraclass correlation among the errors and the presence of the lagged dependent variable as a regressor while C_2 arises solely from the second fact.

<u>Proof</u>:

$$EuV' = E \begin{bmatrix} u_{11} \\ \vdots \\ u_{1T} \\ \vdots \\ u_{N1} \\ \vdots \\ u_{NT} \end{bmatrix} [V_{10} \ \cdots \ V_{1T-1} \ \cdots \ V_{N0} \ \cdots \ v_{NT-1}]$$

$$= E \begin{bmatrix} 0 & u_{11}v_{11} & \cdots & u_{11}v_{1T-1} & \cdots & 0 & \cdots & u_{11}v_{NT-1} \\ 0 & . & & & & & & \\ \vdots & . & & & & & & \\ & . & & & & & & \\ 0 & u_{NT}v_{11} & & . & & . & . & u_{NT}v_{NT-1} \end{bmatrix}$$

$$Eu_{it}v_{js} = Eu_{it}\sum_{k=1}^{s}\alpha^{s-k}u_{jk} \quad , \quad i = 1, \ldots, N \; ; \quad j = 1, \ldots, N \; ;$$

$$= E(\mu_i + \nu_{it})\sum_{k=1}^{s}\alpha^{s-k}(\mu_j + \nu_{jk}) \; , \quad t = 1, \ldots, T \; , \quad s = 1, \ldots, T-1 \qquad \text{(A1)}$$

using the definition of u_{it}. Since μ_i, μ_j, ν_{it}, and ν_{jk} are assumed to be mutually independent for $i \neq j$ and $k \neq t$, the expectation (A1) is zero for $i \neq j$. If $i = j$, (A1) equals

$$E \sum_{k=1}^{s} \alpha^{s-k}(\mu_i + \nu_{it})(\mu_i + \nu_{ik}) = \sum_{k=1}^{s} \alpha^{s-k}\sigma_\mu^2 + \{s \geq t\}\alpha^{s-t}\sigma_\nu^2 , \qquad (A2)$$

where $\{s \geq t\}$ is the indicator function which is equal to 1 if $s \geq t$ and is zero otherwise. Then, expression (A2) equals

$$\sigma_\mu^2 \frac{1 - \alpha^{s-1}}{1 - \alpha} + \{s \geq t\}\alpha^{s-t}\sigma_\nu^2 . \qquad (A3)$$

The first term of (A3) is the appropriate element of ρC_1 and the second term belongs to $(1-\rho)C_2$.

q.e.d.

<u>Lemma A2</u>: $$EV'u = \frac{N\rho}{(1-\alpha)^2} (T-1 - \alpha T + \alpha^T)$$

<u>Proof</u>: $$EV'u = E \text{ tr } V'u = E \text{ tr } uV'$$

$$= \frac{N\rho}{(1-\alpha)^2} (T-1 - \alpha T + \alpha^T)$$

q.e.d.

<u>Lemma A3</u>: $EVV' \equiv G = I_N \otimes (\rho G_1 + (1-\rho)G_2)$ where

$$G_1 = \frac{1}{(1-\alpha)^2}\begin{bmatrix} 0 & & 0 & \cdots & 0 \\ 0 & (1-\alpha)^2 & (1-\alpha)(1-\alpha^2) & \cdots & (1-\alpha)(1-\alpha^{T-1}) \\ \vdots & \vdots & & & \vdots \\ 0 & (1-\alpha^{T-1})(1-\alpha) & & \cdots & (1-\alpha^{T-1})^2 \end{bmatrix},$$

$$G_2 = \frac{1}{\alpha^2-1}\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & \alpha^2(1-\alpha^{-2}) & \alpha^3(1-\alpha^{-2}) & \cdots & \alpha^T(1-\alpha^{-2}) \\ \vdots & \vdots & & & \vdots \\ 0 & \alpha^T(1-\alpha^{-2}) & & & \alpha^{2(T-1)}(1-\alpha^{-2(T-1)}) \end{bmatrix}$$

and $$\operatorname{tr} G = \frac{N\rho}{(1-\alpha)^2(1-\alpha^{-2})}(T-1 - T\alpha^2 - 2\alpha + 2\alpha^T + 2\alpha^{T+1} - \alpha^{2T})$$

$$+ \frac{N(1-\rho)}{(1-\alpha^2)^2}(T-1 - T\alpha^2 + \alpha^{2T}) .$$

The matrices G_1 and G_2 are TxT and represent the intracorrelation effect and the other effects in G as C_1 and C_2 do in C.

<u>Proof</u>:

$$VV' = \begin{bmatrix} 0 & v_{11}v_{11} & v_{11}v_{12} & \cdots & v_{11}v_{1T-1} & 0 & v_{11}v_{21} & \cdots & v_{11}v_{NT-1} \\ 0 & & & & & & & & \cdot \\ \vdots & & & & & & & & \cdot \\ & & & & & & & & \cdot \\ 0 & v_{NT-1}v_{11} & & & v_{NT-1}v_{1T-1} & 0 & v_{NT-1}v_{21} & \cdots & v_{NT-1}v_{NT-1} \end{bmatrix}$$

For $i \neq j$, $Ev_{is}v_{jt} = 0$ since v_{is} and v_{jt} have no components in common and the component errors are assumed to be independent. For $i = j$

$$Ev_{is}v_{it} = E[\sum_{j=1}^{s} \alpha^{s-j}(\mu_i + \nu_{ij})][\sum_{k=1}^{t} \alpha^{t-k}(\mu_i + \nu_{ik})]$$

$$= \rho(\sum_{j=1}^{s} \alpha^{s-j})(\sum_{k=1}^{t} \alpha^{t-k}) + (1-\rho) \sum_{\ell=1}^{m} \alpha^{s+t-2\ell} \qquad \text{(A4)}$$

where $m = \min(s,t)$ since μ and ν are independent each with expectation zero. Then expression (A4) equals

$$\rho \frac{1 - \alpha^s}{1 - \alpha} \frac{1 - \alpha^t}{1 - \alpha} + (1-\rho)\alpha^{s+t} \frac{1 - \alpha^{-2m}}{\alpha^2 - 1} \quad . \qquad \text{(A5)}$$

The first term of (A5) is the s, t-element of G_1 and the second term is the corresponding element of G_2.

q.e.d.

The group of lemmas below are derived from the following result (see Anderson [3], p. 39). Let X_i , $i = 1, 2, 3, 4$ be random variables with a joint normal distribution $N(0,\Sigma)$. Then

$$EX_iX_jX_kX_\ell = \sigma_{ij}\sigma_{k\ell} + \sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk} \quad \text{where}$$

$$\Sigma = (\sigma_{ij}) \ .$$

Recall that $\underset{NTx2}{V^*} = (\underset{NTx1}{V} \quad \underset{NTx1}{0})$ and let D , F , and L be arbitrary (conformable) constant matrices with elements d_{ij} , f_{ij} , and ℓ_{ij} .

<u>Lemma A4</u>: $E(V^{*\prime}Duu'FV^*) = \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}$ where

$$a = \text{tr } CD \text{ tr } C'F + \text{tr } CDCF' + \text{tr } GD\Omega F \ .$$

<u>Proof</u>: $V^{*\prime}Duu'FV^* = \begin{pmatrix} V'Duu'FV & 0 \\ 0 & 0 \end{pmatrix}$. Then

$$EV'Duu'FV = E \sum_{ijk\ell} v_i d_{ij} u_j u_k f_{k\ell} v_\ell$$

$$= \sum_{ijk\ell} E(v_i u_i) E(u_k v_\ell) d_{ij} f_{k\ell} + \sum_{ijk\ell} E(v_i u_k) E(u_j v_\ell) d_{ij} f_{k\ell}$$

$$+ \sum_{ijk\ell} E(v_i v_\ell) E(u_j u_k) d_{ij} f_{k\ell} \qquad \text{(A6)}$$

$$= \sum_{ijk\ell} c_{ji} d_{ij} c_{k\ell} f_{k\ell} + \sum_{ijk\ell} c_{ki} d_{ij} c_{ji} f_{k\ell} + \sum_{ijk\ell} g_{i\ell} d_{ij} \omega_{jk} f_{k\ell} \qquad \text{(A7)}$$

using the expectations $C = (c_{ij})$ and $G = (g_{ij})$ which are given by Lemmas A1 and A3 and the assumed covariance matrix $Euu' = \Omega = (\omega_{ij})$. Then (A7) can be written as

$$\text{tr } CD \text{ tr } C'F + \text{tr } CDCF' + \text{tr } CD\Omega F .$$

q.e.d.

The proofs of Lemmas A5-A14 are similar to the proof of Lemma A4 and are, therefore, presented in less detail. For an arbitrary matrix M , define m^r and m^c as its first row and first column, respectively.

<u>Lemma A5</u>: $EV^*Duu'FV^* = [C'd^{r\prime} \text{ tr } CF' + C'FC'd^{r\prime} + GF'\Omega d^{r\prime} \quad 0]$

<u>Proof</u>: $V^*Duu'FV^* = (Vd^r uu'FV \quad 0)$.

The i^{th} element of the vector $Vd^r uu'FV$ has expectation

$$E \sum_{jk\ell} v_i d_j^r u_j u_k f_{k\ell} v_\ell$$

$$= \sum_{jk\ell} E(v_i u_j)E(u_k v_\ell)d_j^r f_{k\ell} + \sum_{jk\ell} E(v_i u_k)E(u_j v_\ell)d_j^r f_{k\ell}$$

$$+ \sum_{jk\ell} E(v_i v_\ell)E(u_j u_k)d_j^r f_{k\ell}$$

$$= \sum_{jk\ell} c_{k\ell} f_{k\ell} d_j^r c_{ji} + \sum_{jk\ell} d_j^r c_{j\ell} f_{k\ell} c_{ki} + \sum_{jk\ell} d_j^r \omega_{jk} f_{k\ell} g_{i\ell}$$

which gives the vector $C'd^{r'}$ tr $CF' + C'FC'd^{r'} + GF'\Omega d^{r'}$.

Lemma A6: $EV^*Duu'FV^{*'} = C'f^c d^r C + C'd^{r'} f^{c'} C + G$ tr $f^c d^r \Omega$.

Proof: $V^*Duu'FV^{*'} = Vd^r uu' f^c V'$.

The $i\ell^{th}$ element has expectation

$$E \sum_{jk} v_i d_j^r u_j u_k f_k^c v_\ell$$

$$= \sum_{jk} E(v_i u_j)E(u_k v_\ell)d_j^r f_k^c + \sum_{jk} E(v_i u_k)E(u_j v_\ell)d_j^r f_k^c + \sum_{jk} E(v_i v_\ell)E(u_j u_k)d_j^r f_k^c$$

$$= \sum_{jk} d_j^r c_{ji} f_k^c c_{k\ell} + \sum_{jk} d_j^r c_{j\ell} f_k^c c_{ki} + \sum_{jk} d_j^r \omega_{jk} f_k^c g_{i\ell} .$$

Combining these expectations for all the elements gives the lemma.

Lemma A7: $Euu'DV^{*'}FV^{*'} = Cf^c d^{c'} C + d^{c'} Cf^c C + \Omega d^c f^{c'} G$

Proof: $uu'DV^{*'}FV^{*'} = uu'd^c f^c V'$

The $i\ell^{th}$ element of $uu'd^c f^c V'$ has expectation

$$E\sum_{jk} u_i u_j d_j^c v_k f_k^c v_\ell$$

$$= E\sum_{jk} E(u_i u_j)E(v_k v_\ell)d_j^c f_k^c + \sum_{jk} E(u_i v_k)E(u_j v_\ell)d_j^c f_k^c + \sum_{jk} E(u_i v_\ell)E(u_j v_k)d_j^c f_k^c$$

$$= \sum_{jk} \varpi_{ij} g_{k\ell} d_j^c f_k^c + \sum_{jk} c_{ik} c_{j\ell} d_j^c f_k^c + \sum_{jk} c_{i\ell} c_{jk} d_j^c f_k^c \ .$$

Combining these expectation proves the lemma.

Lemma A8: $Eu'DV^{*'}FV^{*}Lu = d^{c'}C(F + F')C'\ell^{r'} + d^{c'}\Omega\ell^{r'}$ tr GF

Proof: $u'DV^{*'}FV^{*}Lu = u'd^c V'FV\ell^r u$ which has expectation

$$E\sum_{ijkm} u_i d_i^c v_j f_{jk} v_k \ell_m^r u_m$$

$$= \sum_{ijkm} E(u_i v_j)E(v_k u_m)d_i^c f_{jk}\ell_m^r + \sum_{ijkm} E(u_i v_k)E(v_j u_m)d_i^c f_{jk}\ell_m^r$$

$$+ \sum_{ijkm} E(u_i u_m)E(v_j v_k)d_i^c f_{jk}\ell_m^r$$

$$= \sum_{ijkm} c_{ij} c_{mk} d_i^c f_{jk}\ell_m^r + \sum_{ijkm} c_{ik} c_{mj} d_i^c f_{jk}\ell_m^r + \sum_{ijkm} \varpi_{im} g_{jk} d_i^c f_{jk}\ell_m^r$$

$$= d^{c'}C(F + F')C'\ell^{r'} + d^{c'}\Omega\ell^{r'} \text{ tr } GF$$

which proves the lemma.

Lemma A9: $$EV^{*'}DV^{*}V^{*'}Fu = \begin{bmatrix} \text{tr } GD \text{ tr } CF + \text{tr } GFCD' + \text{tr } GFCD \\ 0 \end{bmatrix}$$

Proof: $$V^{*}DV^{*}V^{*'}Fu = \begin{bmatrix} V'DVV'Fu \\ 0 \end{bmatrix}$$

The expectation of $V'DVV'Fu$ is

$$E \sum_{ijk\ell} v_i d_{ij} v_j v_k f_{k\ell} u_\ell$$

$$= \sum_{ijk\ell} E(v_i v_j) E(v_k u_\ell) d_{ij} f_{k\ell} + \sum_{ijk\ell} E(v_i v_k) E(v_j u_\ell) d_{ij} f_{k\ell}$$

$$+ \sum_{ijk\ell} E(v_i u_\ell) E(v_j v_k) d_{ij} f_{k\ell}$$

$$= \sum_{ijk\ell} g_{ij} d_{ij} c_{\ell k} f_{k\ell} + \sum_{ijk\ell} g_{ik} f_{k\ell} c_{\ell j} d_{ij} + \sum_{ijk\ell} c_{\ell i} d_{ij} g_{jk} f_{k\ell}$$

which shows the lemma.

<u>Lemma A10</u>: $EV^*V^{*\prime}DV^*Fu = (GD + GD' + I \text{ tr } GD)C'f^{r\prime}$

<u>Proof</u>: $V^*V^{*\prime}DV^*Fu = VV'DVf^r u$.

The i^{th} element of the vector $VV'd^cV'Fu$ has expectation

$$E \sum_{jk\ell} v_i v_j d_{jk} v_k f^r_\ell u_\ell$$

$$= \sum_{jk\ell} E(v_i v_j) E(v_k u_\ell) d_{jk} f^r_\ell + \sum_{jk\ell} E(v_i v_k) E(v_j u_\ell) d_{jk} f^r_\ell$$

$$+ \sum_{jk\ell} E(v_i u_\ell) E(v_j v_k) d_{jk} f^r_\ell$$

$$= \sum_{jk\ell} g_{ij} d_{jk} c_{\ell k} f^r_\ell + \sum_{jk\ell} g_{ik} d_{jk} c_{\ell j} f^r_\ell + \sum_{jk\ell} g_{jk} d_{jk} c_{\ell i} f^r_\ell$$

Assembling these expectations gives the lemma.

<u>Lemma A11</u>. $EV^*V^{*\prime}DV^{*\prime}Fu = Gd^c \text{ tr } CF + CFCd^c + C'F'Gd^c$

Proof: $V^*V^{*\prime}DV^{*\prime}Fu = VV'd^cV'Fu$, the i^{th} element of which has expectation

$$E\sum_{jk\ell} v_i v_j d_j^c v_k f_{k\ell} u_\ell$$

$$= \sum_{jk\ell} E(v_i v_j)E(v_k u_\ell)d_j^c f_{k\ell} + \sum_{jk\ell} E(v_i v_k)E(v_j u_\ell)d_j^c f_{k\ell}$$

$$+ \sum_{jk\ell} E(v_i u_\ell)E(v_j v_k)d_j^c f_{k\ell}$$

$$= \sum_{jk\ell} g_{ij} d_j^c c_{\ell k} f_{k\ell} + \sum_{jk\ell} g_{ik} f_{k\ell} c_{\ell j} d_j^c + \sum_{jk\ell} c_{\ell i} f_{k\ell} g_{jk} d_j^c \,.$$

Collecting these expectations produces the lemma.

Lemma A12: $Euu'DV^{*\prime}FV^* = [(\Omega \text{ tr } GF' + CFC' + CF'C')d^c \quad 0]$

Proof: $uu'DV^{*\prime}FV^* = (uV'FVu'd^c \quad 0)$. The i^{th} element of $uV'FVu'd^c$ has expectation

$$E\sum_{jk\ell} u_i v_j f_{jk} v_k u_\ell d_\ell^c$$

$$= \sum_{jk\ell} E(u_i v_j)E(v_k u_\ell)f_{jk}d_\ell^c + \sum_{jk\ell} E(u_i v_k)E(v_j u_\ell)f_{jk}d_\ell^c + \sum_{jk\ell} E(u_i u_\ell)E(v_j v_k)f_{jk}d_\ell^c$$

$$= \sum_{jk\ell} c_{ij} c_{\ell k} f_{jk} d_\ell^c + \sum_{jk\ell} c_{ik} c_{\ell j} f_{jk} d_\ell^c + \sum_{jk\ell} \omega_{i\ell} g_{jk} f_{jk} d_\ell^c \,.$$

Collecting these expectations yields the lemma.

Lemma A13: $Euu'DV^*V^{*\prime} = \Omega DG + CD'C + C \text{ tr } CD'$

Proof: $uu'DV^*V^{*\prime} = uu'DVV'$. The expectation of the $i\ell^{th}$ element is

$$E\sum_{jk} u_i u_j d_{jk} v_k v_\ell$$

$$= \sum_{jk} E(u_i u_j)E(v_k v_\ell)d_{jk} + \sum_{jk} E(u_i v_k)E(u_j v_\ell)d_{jk} + \sum_{jk} E(u_i v_\ell)E(u_j v_k)d_{jk}$$

$$= \sum_{jk} \omega_{ij} d_{jk} g_{k\ell} + \sum_{jk} c_{ik} d_{jk} c_{j\ell} + \sum_{jk} c_{i\ell} c_{jk} d_{jk} .$$

Arranging this matrix of expectations gives the result.

<u>Lemma A14</u>: $EV^{*'}DV^{*'}FV^{*'}Lu = \begin{bmatrix} d^{c'}(C'L'G + GLC + G \text{ tr } CL)f^c \\ 0 \end{bmatrix}$

<u>Proof</u>: $V^{*'}DV^{*'}FV^{*'}Lu$

$$= \begin{pmatrix} V'd^cV'f^cV'Lu \\ 0 \end{pmatrix} . \qquad V'd^cV'f^cV'Lu \text{ has expectation}$$

$$E\sum_{ijkm} v_i d_i^c v_j f_j^c v_k \ell_{km} u_m$$

$$= \sum_{ijkm} E(v_i v_j)(v_k u_m) d_i^c f_j^c \ell_{km} + \sum_{ijkm} E(v_i v_k)E(v_j u_m) d_i^c f_j^c \ell_{km}$$

$$+ \sum_{ijkm} E(v_i u_m)E(v_j v_k) d_i^c f_j^c \ell_{km}$$

$$= \sum_{ijkm} d_i^c g_{ij} f_j^c c_{mk} \ell_{km} + \sum_{ijkm} d_i^c g_{ik} \ell_{km} c_{mj} f_j^c + \sum_{ijkm} d_i^c c_{mi} \ell_{km} g_{jk} f_j^c$$

$$= d^{c'}Gf^c \text{ tr } CL + d^{c'}GLCf^c + d^{c'}C'L'Gf^c .$$

<u>Lemma A15</u>: $Euu'DV^*FV^* = [(\Omega DG + CD'C + C \text{ tr } CD')f^{r'} \quad 0]$

<u>Proof</u>: $uu'DV^*FV^*$

$$= (uu'DVf^rV \quad 0)$$

$$= (uu'DVV'f^{r'} \quad 0)$$

since $f^r V$ is a scalar and, therefore, equals its transpose, $V'f^{r'}$. Then $Euu'DVV'f^{r'} = Euu'DV^*V^{*'}f^{r'}$ and an application of Lemma A13 shows that this expectation is $(\Omega DG + CD'C + C \text{ tr } CD')f^{r'}$ which proves the lemma.