

NO-A189 078

DETECTING SERIAL CORRELATION IN THE ERROR STRUCTURE OF
A CROSS-LAGGED PANEL MODEL(U) STANFORD UNIV CA DEPT OF
STATISTICS L S HAYER ET AL. 02 FEB 88 TR-401

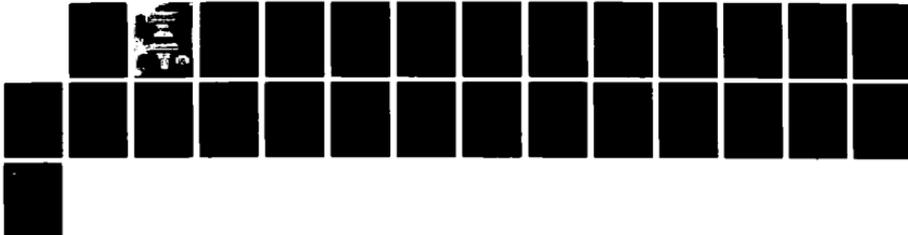
1/1

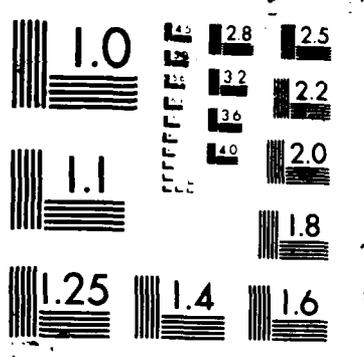
UNCLASSIFIED

NO0014-06-K-0156

F/G 12/3

NL





AD-A189 070

88 2 19 078

DETECTING SERIAL CORRELATION IN THE ERROR STRUCTURE
OF A CROSS-LAGGED PANEL MODEL

BY

LAWRENCE S. MAYER and STEVEN S. CARROLL

TECHNICAL REPORT NO. 401

FEBRUARY 2, 1988

Prepared Under Contract

N00014-86-K-0156 (NR-042-267)

For the Office of Naval Research

Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Code	
Distribution Statement	
Dist	Special
A-1	

1. INTRODUCTION

Cross-lagged panel studies are statistical studies in which two or more variable are measured for a large number of subjects at each of several waves or points in time. The variables divide naturally into two sets and the primary purpose of analysis is to estimate and test the strength of the relationship between the sets. Such studies are found in the mainstream of social, behavioral and business research (e.g. Crano, Kenny and Campbell (1972), Greenberg, Kessler and Logan (1979), Eaton (1978), Polachek and McCutcheon (1983) and Frey (1984)). This paper contributes to these studies by developing and applying procedures for detecting the presence of serial correlation in the error structure of the regression models used in such studies.

Methods of analysis for cross-lagged panel studies have evolved over the last twenty years (Kessler and Greenberg (1981) provide an excellent review of this development). Early methods used correlation statistics and were motivated by the work of Donald Campbell (1963), which, in turn, was motivated, by the seminal work of Paul Lazarsfeld (1948) on the analysis of panel studies involving discrete variables. The most significant advancement in the methodology of cross-lagged panel analysis has been the development of a regression approach. This approach treats the cross-effects as parameters in one or more regression models and then applies standard regression procedures to estimate and test these parameters (for example, see Peltz and Andrews (1964), Duncan (1969), Heise (1970), Hannan and Young (1977), and Rogosa (1980)). Although the cross-lagged panel study includes observations over time, for the most part, this approach has assumed that the errors of the model are independent across waves.

Mayer (1984a; 1984b) extends the regression approach, by incorporating the cross-effects as parameters in a multivariate regression model and develops procedures to estimate and test these parameters. He considered both the model with independent

errors and the model with serially correlated errors. In this paper we extend the applicability of this results by considering the problem of detecting if the multivariate regression model requires the serially correlated error structure as opposed to the independent error structure.

2. PRELIMINARY NOTIONS

To formalize the multivariate regression structure of the cross-lagged panel model let $z_{it} = (x'_{it}; y'_{it})'$ be the measurements made on the k variables for the i th replication at the t th wave ($i = 1, \dots, n$; $t = 0, \dots, T$) where x_{it} and y_{it} are sub-vectors of p and q measurements, respectively. Let $Z_t = (X_t; Y_t)$ be the $n \times k$ response matrix that has z_{it} as the i th row. The regression structure is

$$Z_t = Z_{t-1} B + E_t \quad t = 1, \dots, T \quad (2.1)$$

where B is a $k \times k$ matrix of unknown regression coefficients that does not depend on t ; and E_t is an unobserved random error matrix with structure

$$E_t = E_{t-1} A + \epsilon_t \quad t=1, \dots, T \quad (2.2)$$

where A is a $k \times k$ unknown matrix of autoregressive parameters and ϵ_t is an unobserved random matrix with rows which are independent Gaussian vectors with common mean 0 and common covariance matrix Λ that does not depend on t .

The rows of E_t are thus independent Gaussian vectors with common mean 0 and common covariance matrix

$$\Sigma_t = A' \Sigma_{t-1} A + \Lambda \quad t=1, \dots, T. \quad (2.3)$$

If we let Σ_t be the covariance matrix of Z_t ,

then

$$\Sigma_t = B' \Sigma_{t-1} B + \Sigma_t = B' \Sigma_{t-1} B + A' \Sigma_{t-1} A + \Lambda \quad (2.4)$$

To complete the specification, we describe the behavior of the response matrix and error matrix at the initial wave. Since the initial observations are generated by the same sampling scheme as the other observations we assume the initial response matrix Z_0 is random with rows that are independent Gaussian vectors with common mean Q and common covariance matrix Σ_0 . Likewise, we let the rows of the initial error matrix E_0 be independent Gaussian vectors with common mean Q and common covariance matrix Σ_0 .

The primary goal in the research approach is to estimate the regression parameter matrix β and to test the hypothesis of no effects ($\beta = \Lambda = Q$) and the hypothesis that the responses in X_{t-1} do not affect the responses in X_t , and conversely, the responses in X_{t-1} do not affect the responses in X_t - the hypothesis that β and Λ are block diagonal. The method of estimation and testing depends heavily on whether serial correlation is present, i.e., whether $\Lambda = Q$. For example, if serial correlation is present, but ignored, the estimator of β derived by applying maximum likelihood (to the wrong model) is not consistent in the number of replications (see Section 4).

Before presenting our tests we comment on the appropriateness of existing procedures for detecting the presence of serial correlation in regression models involving observations over time. The most widely used procedure, the Durbin-Watson procedure, is not appropriate for the cross-lagged panel study because it is not designed for use with models that contain only lagged endogenous predictors and is designed for use with univariate regression models (Malinvaud (1980), Johnson (1972)). The most notable extensions of this procedure are appropriate for models with lagged endogenous predictors but, again, are not designed for multivariate regression models (Durbin (1970), Judge, et. al. (1980)). Finally, we know of one extension of these procedures to a procedure that is appropriate for multivariate models (Guilkey (1975)); but it, like all of these procedures, guarantees that the

advertised alpha level is accurate asymptotically in the number of waves. The relevant asymptotics for cross-lagged panel studies are in the number of replications since the number of waves tends to be very small.

Mayer (1984b) considered two informal procedures for detecting the presence of serial correlation in the errors for the multivariate version of the cross-lagged panel model. His first method is to ignore the possibility of serial correlation, estimate the matrix of regression coefficients by a pooled version of least squares estimation, and then use the residuals to estimate the autoregressive parameter matrix \hat{A} . This method gives a crude estimate of A but does not produce an estimator that is consistent as the number of replications becomes large. His second method is to partition the matrix A conformally with Z_t and set the off-diagonal blocks equal to 0. The diagonal blocks of \hat{A} are estimated by expressing each subresponse vector at wave t as a function of both subresponse vector at wave $t-1$. The other subresponse is considered an exogenous predictor and a multivariate version of a standard method of econometrics is used to estimate the diagonal block of \hat{A} . This method is crude in that ignoring the off-diagonal blocks of A may seriously bias the estimates of the off-diagonal elements of B , and these blocks are a major focus of the regression approach.

3. TESTING FOR SERIAL CORRELATION

We begin with a conditional approach to testing for the presence of serial correlation in the error structure of the multivariate cross-lagged panel model, conditional in the sense that it treats the first two waves as fixed. We then consider an unconditional approach that captures the random nature of all the waves. We complete the section by discussing the impacts of serial correlation on the procedures used to estimate and test the matrix of regression coefficients.

advertised alpha level is accurate asymptotically in the number of waves. The relevant asymptotics for cross-lagged panel studies are in the number of replications since the number of waves tends to be very small.

Mayer (1984b) considered two informal procedures for detecting the presence of serial correlation in the errors for the multivariate version of the cross-lagged panel model. His first method is to ignore the possibility of serial correlation, estimate the matrix of regression coefficients by a pooled version of least squares estimation, and then use the residuals to estimate the autoregressive parameter matrix \hat{A} . This method gives a crude estimate of \hat{A} but does not produce an estimator that is consistent as the number of replications becomes large. His second method is to partition the matrix \hat{A} conformally with ζ and set the off-diagonal blocks equal to 0. The diagonal blocks of \hat{A} are estimated by expressing each subresponse vector at wave t as a function of both subresponse vectors at wave $t-1$. The other subresponse is considered an exogenous predictor and a multivariate version of a standard method of econometrics is used to estimate the diagonal block of \hat{A} . This method is crude in that ignoring the off-diagonal blocks of \hat{A} may seriously bias the estimates of the off-diagonal elements of β , and these blocks are a major focus of the regression approach.

3. TESTING FOR SERIAL CORRELATION

We begin with a conditional approach to testing for the presence of serial correlation in the error structure of the multivariate cross-lagged panel model, conditional in the sense that it treats the first two waves as fixed. We then consider an unconditional approach that captures the random nature of all the waves. We complete the section by discussing the impacts of serial correlation on the procedures used to estimate and test the matrix of regression coefficients.

3.1 Conditional Analysis

Combining equations (2.1) and (2.2) yields

$$z_t = z_{t-1} B + \varepsilon_{t-1} A + \varepsilon_t \quad t=1, \dots, T$$

Using the fact that $\varepsilon_{t-1} = z_{t-1} - z_{t-2} B$

$$z_t = z_{t-1} (B + A) - z_{t-2} (B A) + \varepsilon_t \quad t=2, \dots, T \quad (3.1)$$

which expresses the cross-lagged panel model with first-order autoregressive errors as a second-order vector-valued autoregressive model with independent errors over waves and independent replications at each wave, a model studied by Anderson (1978).

We suggest that the test for the presence of serial correlation be preceded by a test for no effects or independence of the response vector over time, the null hypothesis, $H_1: B = A = Q$. The null hypothesis implies $H_1^*: B + A = Q$ and $BA = Q$, a hypothesis about the models as expressed in (3.1). The model is reexpressed as

$$z_t = z_{t-1} Q + \varepsilon_t$$

$$\text{where } z = [z_T'; \dots; z_2']', \quad z_{-1} = \begin{bmatrix} z_{T-1}' & \dots & z_1' \\ z_{T-2}' & \dots & z_0' \end{bmatrix}' = \begin{bmatrix} z_{-1}^{(1)'} \\ z_{-1}^{(2)'} \end{bmatrix}'$$

$$\varepsilon = [\varepsilon_T'; \dots; \varepsilon_2']' \text{ and } Q = \begin{bmatrix} B + A \\ -B A \end{bmatrix} = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}$$

We define the estimators

$$\hat{Q} = [z_{-1}' \quad z_{-1}]^{-1} z_{-1}' z$$

and

$$\hat{Q} = [n(T-1)]^{-1} (z - z_{-1} \hat{Q})' (z - z_{-1} \hat{Q})$$

and the test statistic

$$t_1 = \text{tr} \hat{Q}' \begin{bmatrix} Z_0' & Z_1' \\ Z_0 & Z_1 \end{bmatrix} \hat{Q} \hat{\Lambda}^{-1}$$

A proof quite similar to those in Anderson (1978) yields

Theorem 1: Under the null hypothesis of no effects (H_1^*) and given Z_0 and Z_1 , the test statistic t_1 is a Lawley-Hotelling trace statistic and has a chi-square distribution with $2k^2$ degrees of freedom as the number of replications becomes large.

If H_1^* is rejected two possibilities are considered. Either β and Λ are both not zero or β alone is not zero. Because of the nature of the model we dismiss the possibility of serially correlated errors ($\Lambda \neq Q$) in a model with no regression effects ($\beta = Q$) and the hypothesis of "countering effects", ($\beta = -\Lambda$).

We differentiate between the two possibilities by testing the hypothesis $H_2: \Lambda = Q$ which is equivalent to testing the hypothesis $H_2^*: \beta \Lambda = Q$ (since we concluded $\beta \neq Q$ as a result of the first test). Let $\hat{Q} = (\hat{Q}_1', \hat{Q}_2')$ where \hat{Q} is partitioned conformally with Q .

Let $Z_{-1}' = [Z_{-1}^{(1)'}; Z_{-1}^{(2)'}]$ and

$$t_2 = \text{tr} \hat{Q}_2' \begin{bmatrix} Z_{-1}^{(2)'} & Z_{-1}^{(2)} \\ Z_{-1}^{(1)'} & Z_{-1}^{(1)} \end{bmatrix} - Z_{-1}^{(2)'} Z_{-1}^{(1)} (Z_{-1}^{(1)'} Z_{-1}^{(1)})^{-1} Z_{-1}^{(1)'} Z_{-1}^{(2)} \hat{Q}_2 \hat{\Lambda}^{-1}.$$

We have

Theorem 2: Under the null hypothesis of no serial correlation (H_2^*) and given Z_0 and Z_1 , the test statistic t_2 is a Lawley-Hotelling trace statistic and has a chi-square distribution with k^2 degrees of freedom, as the number of replications becomes large.

The proof of this result is a direct application of a result of Anderson (1978).

The calculation for this conditional analysis, can be done by a standard statistical package such as SAS. Two limitations

of this conditional analysis are that it sacrifices the $2nk$ degrees of freedom of the first two waves and that it provides no direct estimate of the regression matrix B .

3.2 Unconditional Analysis

Applying the same reasoning as used in the conditional analysis, we begin the unconditional analysis by considering the hypothesis of no effects $H_1: B = A = Q$.

The likelihood function $L(B, A, \theta_0, \xi_0)$ generated by the $nk(T + 1)$ observations is

$$\begin{aligned}
 (2\pi)^{-\frac{nk(T+1)}{2}} & |A|^{-\frac{n(T-1)}{2}} |A' \xi_0 A + A|^{-\frac{n}{2}} |\theta_0|^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \right. \\
 & \left. \{ \text{tr } \xi_0 \theta_0^{-1} \xi_0' + \text{tr} (z_1 - z_0 B) (A' \xi_0 A + A)^{-1} (z_1 - z_0 B)' + \right. \\
 & \left. \sum_{t=2}^T \text{tr} (z_t - z_{t-1} (B + A) - z_{t-2} (-BA)) A^{-1} (z_t - z_{t-1} (B + A) - \right. \\
 & \left. z_{t-2} (-BA))' \} \right) \quad (3.2)
 \end{aligned}$$

which can be maximized numerically.

Under the null hypothesis H_1 the likelihood function $L(Q, Q, A, \theta_0, Q)$ is

$$\begin{aligned}
 (2\pi)^{-\frac{nk(T+1)}{2}} & |A|^{-\frac{nT}{2}} |\theta_0|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2} \left[\text{tr } \xi_0 \theta_0^{-1} \xi_0' + \right. \right. \\
 & \left. \left. \text{tr} \sum_{t=1}^T (z_t A^{-1} z_t') \right] \right\} \quad (3.3)
 \end{aligned}$$

which is maximized by letting $\hat{\theta}_0 = n^{-1} z_0' z_0$ and $\hat{A} = (nT)^{-1} \sum_{t=1}^T z_t' z_t$.

If we let L_u be the maximum of the likelihood in (3.2) and L_r be the maximum of the likelihood in (3.3) and define

$$t_3 = -2 \log \frac{L_r}{L_u}$$

we have

Theorem 3: Under the hypothesis of no effect $H_1: \beta = \lambda = \rho$ the test statistic t_3 has a chi-square distribution with $(5k^2 + k)/2$ degrees of freedom as the number of replications becomes large.

If we reject $H_1: \beta = \lambda = \rho$ we proceed to test the hypothesis of no serial correlation in the error structure, $H_2: \lambda = \rho$ by applying the likelihood ratio method. Under H_2 the likelihood function $L(\beta, \rho, \lambda, \rho_0, \rho)$ is

$$(2\pi)^{-\frac{nk(T+1)}{2}} |\Lambda_c|^{-\frac{nT}{2}} |\rho_0|^{-\frac{n}{2}} \exp\left(-\frac{1}{2} [\text{tr} \rho_0^{-1} z_0' + \text{tr} \sum_{t=1}^T (z_t - z_{t-1} \beta) \Lambda_c^{-1} (z_t - z_{t-1} \beta)']]\right) \quad (3.4)$$

Applying a result of Anderson (1978) the likelihood function is maximized by

$$\tilde{\beta} = (nT)^{-1} \left[\sum_{i=0}^{T-1} z_i' z_i \right]^{-1} \sum_{i=1}^T z_{i-1}' z_i$$

$$\tilde{\rho}_0 = n^{-1} z_0' z_0$$

and

$$\tilde{\Lambda}_c = (nT)^{-1} \sum_{i=1}^T (z_i - z_{i-1} \tilde{\beta})' (z_i - z_{i-1} \tilde{\beta})$$

Let L_r^* be the maximum of the likelihood function (3.4) and define

$$t_4 = -2 \log \frac{L_r^*}{L_u}$$

where L_u is given in the definition of t_3 ; we have

Theorem 4: Under the hypothesis of no serial correlation ($H_2: A = Q$) the test statistic τ_4 has a chi-square distribution with $(3k^2 + k)/2$ degrees of freedom as the number of replications becomes large.

The tests given in Theorem 3 and Theorem 4 are somewhat difficult to apply in that both require maximization of the likelihood function given in (3.2), a calculation which requires "brute force" numerical optimization. With a little modification, however, tests similar to those in Theorem 3 and Theorem 4, but more easily computed, are provided.

Note that the maximization of the likelihood function given in (3.2) yields maximum likelihood estimators of β , βA , and $\beta + A$ which are mathematically consistent. More explicitly, if we let \hat{H}_1 , \hat{H}_2 , and \hat{H}_3 be these estimators, then they satisfy the constraint

$$\hat{H}_1^{-1} (\hat{H}_2) - \hat{H}_3 = -\hat{H}_1. \quad (3.5)$$

If we relax this constraint then the parameters β , $\beta + A$, and βA can be estimated independently. This relaxation can be achieved by replacing the model of the wave one response matrix in (2.1) by the model

$$\tilde{z}_1 = \tilde{z}_0 \beta_0 + \tilde{E}_1 \quad (3.6)$$

where β_0 is not restricted to be the same as β .

A standard maximum likelihood routine, such as LISREL, can be used to maximize the likelihood function for this relaxed model. Let $L(\beta, \beta_0, A, \Lambda, \tilde{z}_0, \tilde{I}_0)$ be the modified likelihood function and $L_u^* = L(\beta^*, \beta_0^*, A^*, \Lambda^*, \tilde{z}_0^*, \tilde{I}_0^*)$ its maximum.

The hypothesis of independence of the response vector \tilde{z}_{it} over time for the relaxed model is $\hat{H}_1: \beta_0 = \beta = A = Q$. We define the test statistic for \hat{H}_1 to be

$$\tau_3^* = -2 \log \frac{L}{L_u^*}.$$

For the relaxed model the hypothesis of no serial correlation remains $H_2: \Lambda = Q$. We let \tilde{L}_r be the maximum of $L(\beta, \beta_0, \rho, \Lambda, \theta_0, Q)$ and define

$$t_4^* = -2 \log \frac{\tilde{L}_r}{L_u^*}.$$

By the usual theory of likelihood ratio tests we have

Theorem 5: For the relaxed model as the number of replications becomes large under \hat{H}_1 the test statistic t_3^* has a chi-square distribution with $(7k^2 + k)/2$ degrees of freedom, and under H_2 the test statistic t_4^* has a chi-square distribution with $(3k^2 + k)/2$ degrees of freedom.

Note that under \hat{H}_1 the relaxed model and the unconditional model are identical. We also remark that although the hypothesis of no serial correlation is the same for the two models, it is not obvious how the difference between the models is reflected in hypothesis tests.

3.3 STATISTICAL INFERENCES WITH SERIAL CORRELATION

Before applying our tests to a panel model, we consider the impact of serial correlation in the error structure on the estimates of the regression parameters and the tests of the hypotheses of no effects and no cross-effects.

Mayer (1984b) has shown that ignoring serial correlation leads to estimates of the regression parameters that do not converge to the true parameters as the number of replications becomes large. Similarly the likelihood ratio tests for the test of no effects and no cross-effects can be highly biased if serial correlation is ignored.

Suppose the conditional analysis of the model is used and serial correlation is detected. Then the parameter matrices $\beta + \Lambda$, $\beta\Lambda$ are estimable but the regression parameter matrix β is not. The hypotheses of no effects and no cross-effects can be tested by testing the corresponding hypothesis about $\beta + \Lambda$ and $\beta\Lambda$.

If the unconditional analysis is used and the constraints given in (3.5) are satisfied by the estimates of β and Δ , then expressions are given for β and Δ . If the relaxed model is used - the constraints are ignored - then the values of β , $\beta + \Delta$ and $\beta\Delta$ can be used to solve for β in any of of three ways. The three solutions are only asymptotically identical.

Whether the constraints are ignored or not the likelihood ratio method can be applied to give tests for the hypotheses of no effects and cross-effects.

Several other methods have been suggested to estimate the regression parameters in the cross-lagged panel model when serial correlation is present in the errors (e.g., Markus (1979), Kessler and Greenberg (1980)). One method is to ignore the off-diagonal elements of Δ and to express each response at time t as a linear function of the responses at time $t-1$ and $t-2$ plus an error term which is independent over time. Then ordinary least squares is applied to that equation and the diagonal elements of Δ are obtained from the estimated coefficients. This method resembles a standard method in econometrics attributed to Durbin (Johnston (1972), p. 263). A second method expresses each response at time t as a function of the responses at time $t-1$ and then applies the instrumental variable method of econometrics.

Neither of these methods associates naturally with a test of the hypothesis of no cross-effects. Furthermore, by ignoring the off-diagonal elements of Δ , the estimates of the off-diagonal elements of β may be highly biased, and these are the coefficients that are used to indicate cross-effects. Consequently, a test based on the asymptotic distribution of these estimators may seriously over estimate the degree of cross-effects.

4. ANALYSIS OF PANEL DATA

We illustrate our methods by applying them to a set of panel data that are taken from a study of the relationship between patients opinion of the concept of Health Maintenance Organizations (HMO)

and their perception of the quality of care they are receiving from the Health Maintenance Organization in which they are enrolled. For the purpose of demonstrating our results, the critical issue is whether a cross-lagged panel model fit to these data appears to have serial correlation in the error structure. Secondary issues include the degree to which such correlation affects the estimates of the regression parameters and the degree to which the simpler conditional analysis produces results that are similar to the results produced by the unconditional analysis.

These data were obtained from interviews of 20 patients of a particular HMO. They were selected at random and interviewed after having seen a primary-care physician at the HMO for the first time. The same 20 patients are interviewed again, after their second and third visits to the HMO. The two variables of interest are each a compilation of several measurements that are made at each interview. The original interviews were not provided by the consulting firm that owns the data.

The first variable measures the degree to which the patient supports the concept of an HMO as a provider of primary medical care. Issues of secondary and emergency care were not addressed in the study. The second variable measures the degree to which the patient is satisfied with the treatment received at the HMO on the particular visit just concluded. The variables are standardized (over a larger sample) to have means of 10 and standard deviations of 3. The panel model adopted is the simplest example of the multivariate model since $p = q = 1$.

The conditional analysis was completed by the multivariate regression routine in SAS and produced the results in Table 1. Displayed are the the Lawley-Hotelling trace statistic for testing the hypothesis of no effects, $t_1 = 342$, which is asymptotically chi-square with 8 degrees of freedom and the Lawley-Hotelling trace statistic for testing the hypothesis of no serial correlation in the error structure, $t_2 = 31.1$, which is asymptotically chi-square with 4 degrees of freedom. Combining these tests gives

TABLE 1
CONDITIONAL ANALYSIS OF HMO PANEL DATA

Maximum Likelihood Estimates

$$\hat{Q}_1 = \begin{bmatrix} .715 & .041 \\ -.149 & .591 \end{bmatrix} \quad \text{estimates } B + A$$

$$\hat{Q}_2 = \begin{bmatrix} .338 & -.137 \\ -.645 & .548 \end{bmatrix} \quad \text{estimates } BA$$

Maximum Likelihood Estimates of Standard Errors

$$\hat{Q}_1 : \begin{bmatrix} .080 & .116 \\ .061 & .088 \end{bmatrix}$$

$$\hat{Q}_2 : \begin{bmatrix} .089 & .129 \\ .154 & .224 \end{bmatrix}$$

Ratios of the Estimates to Their Estimated Standard Errors

$$\hat{Q}_1 : \begin{bmatrix} 9.0 & .35 \\ -2.4 & 6.7 \end{bmatrix}$$

$$\hat{Q}_2 : \begin{bmatrix} 3.8 & -1.06 \\ -4.2 & 2.45 \end{bmatrix}$$

Lawley-Hotelling Test of H_1^* : $B + A = Q$ and $BA = Q$

$t_1 = 342$ asymptotic chi-square 8df

Lawley-Hotelling Test of H_2^* : $BA = Q$

$t_2 = 31.1$ asymptotic chi-square 4df

evidence that there are temporal effects both in the observed regression and in the error structure.

Also displayed in Table 1 are the maximum likelihood estimates of the elements in $\beta + \lambda$ and $\beta\lambda$, the corresponding estimates of the standard errors of these elements, and the ratios of the estimated coefficients to their estimated standard errors. Examination of these coefficients and the standard errors indicates responses at a given wave are good predictors of the same responses at the next wave. Also indicated is that the patient's perception of the quality of care at wave t may have a stronger affect on his or her opinion of the concept of an HMO at wave $t + 1$ than does his opinion of the concept of an HMO at wave t have on his perception of the quality of his care at wave $t + 1$.

Note again that this conditional analysis provides no direct estimate of the regression parameter matrix or autoregressive parameter matrix.

Before turning to the unconditional analysis we examine Table 2 which displays the results that would have been obtained if the conditional analysis were applied to the model without allowing for serial correlation in the errors. Remembering that the model being analyzed is incorrectly specified, we note that, the Lawley-Hotelling trace statistic for testing the significance of the regression parameters (22.9), which is "advertised" to be asymptotically chi-square with 4 degrees of freedom, is quite a bit smaller than the test statistic t_1 given in Table 1. Examination of the estimated coefficients and the estimated standard errors indicates far less temporal dependence than is indicated by the results in Table 1. Furthermore, these results appear to indicate that the patients opinion of the concept of an HMO at wave t may have a stronger impact on his or her perception of the quality of care received at wave $t + 1$ than does his or her perception of the quality of care received at wave t has on his or her opinion of the concept of HMO at wave $t + 1$; this result is in sharp contrast to the results of Table 1 and indicates that

TABLE 2

CONDITIONAL ANALYSIS OF HMO PANEL DATA:
IGNORING SERIAL CORRELATION

Maximum Likelihood Estimates

$$\hat{\beta} = \begin{bmatrix} .529 & -.313 \\ -.123 & .360 \end{bmatrix}$$

Maximum Likelihood Estimates of Standard Errors

$$\hat{\beta} : \begin{bmatrix} .138 & .153 \\ .140 & .155 \end{bmatrix}$$

Ratios of the Estimates to Their Estimated Standard Errors

$$\hat{\beta} : \begin{bmatrix} 3.8 & -2.0 \\ -.88 & 2.3 \end{bmatrix}$$

"Lawley-Hotelling" test of $H_2: \beta = 0$ [Assuming $\Lambda = 0$]

$t = 22.9$ asymptotic chi-square 4df

ignoring the serial correlation in error structure would have led to misleading conclusions.

The unconditional analysis of the relaxed model was completed with the LISREL program associated with SPSS-X and produced the statistics presented in Table 3. This table gives the approximate likelihood ratio statistic for testing the fit of the model when compared to the model of no effects, $t_3^* = 86.81$, which is asymptotically chi-square with 15 degrees of freedom and the approximate likelihood ratio statistics for testing for the presence of serial correlation in the error structure, $t_4^* = 47.76$, which is asymptotically chi-square with 7 degrees of freedom. The results of the tests are consistent with the results

TABLE 3

UNCONDITIONAL ANALYSIS OF HMO PANEL DATA:
RELAXED MODEL

Maximum Likelihood Estimates (Relaxed Model):

$$\beta_0^* = \begin{bmatrix} -.186 & -.273 \\ .910 & -.019 \end{bmatrix}$$

$$\beta_1^* = \begin{bmatrix} .715 & .041 \\ -.149 & .591 \end{bmatrix} \quad \text{estimates } \beta + \lambda$$

$$\beta_2^* = \begin{bmatrix} .338 & -.137 \\ -.645 & .548 \end{bmatrix} \quad \text{estimates } \beta \lambda$$

$$\beta_0^* = \begin{bmatrix} 7.66 & 2.91 \\ 2.91 & 2.93 \end{bmatrix}$$

$$\lambda^* = \begin{bmatrix} .518 & -.163 \\ -.163 & 1.10 \end{bmatrix}$$

Maximum Likelihood Estimates of Standard Errors

$$\beta_0^*: \begin{bmatrix} .28 & .36 \\ .44 & .58 \end{bmatrix} \quad \beta_2^*: \begin{bmatrix} .08 & .12 \\ .14 & .21 \end{bmatrix}$$

$$\beta_1^*: \begin{bmatrix} .08 & .11 \\ .06 & .08 \end{bmatrix} \quad \lambda^*: \begin{bmatrix} .178 & .187 \\ .187 & .378 \end{bmatrix}$$

Ratio of the Estimates to Their Estimated Standard Errors

$$\beta_0^*: \begin{bmatrix} -.67 & -.75 \\ 2.05 & -.03 \end{bmatrix} \quad \beta_2^*: \begin{bmatrix} 4.07 & -1.13 \\ -4.47 & 2.60 \end{bmatrix}$$

TABLE 3

UNCONDITIONAL ANALYSIS OF HMO PANEL DATA:
RELAXED MODEL
(Continued)

$$B_1^*: \begin{bmatrix} 9.54 & .38 \\ -2.61 & 7.11 \end{bmatrix} \quad A^*: \begin{bmatrix} 2.91 & -.871 \\ -.871 & 2.91 \end{bmatrix}$$

Likelihood Ratio Test of $H_1: B = A = Q$

$$t_3^* = 86.81 \quad \text{asymptotic chi-square 15df}$$

Likelihood Ratio Test of $H_2: A = Q$

$$t_4^* = 47.76 \quad \text{asymptotic chi-square 7df}$$

of the tests from the conditional analysis. We conclude that the panel model has both regression effects and serial correlation in the error structure.

Also displayed in Table 3 are the maximum likelihood estimates of the parameters, their estimated standard errors and the ratio of the estimates to the estimated standard error. These statistics support conclusions almost identical to those obtained from examination of the statistics obtained from the conditional analysis (Table 1). The estimates of $B + A$ and BA are similar although the estimates of the standard errors differ. We suggest that given the ease of calculation and the similarity of results the conditional analysis might be more attractive to the practitioner. Table 4 displays the unconditional analysis of the model with the serial correlation ignored. We note that these results, like the results in Table 2, could be quite misleading. Ignoring the serial correlation in the error of a panel model appears to be a serious error whether the model is treated conditionally or unconditionally.

TABLE 4

UNCONDITIONAL ANALYSIS OF HMO PANEL DATA:
IGNORING SERIAL CORRELATION

(Nominal) Maximum Likelihood Estimates

$$\beta^* = \begin{bmatrix} .536 & -.298 \\ -.106 & .401 \end{bmatrix}$$

$$\lambda^* = \begin{bmatrix} 5.57 & -1.90 \\ -1.90 & 6.66 \end{bmatrix}$$

$$\rho_0^* = \begin{bmatrix} 7.657 & 2.911 \\ 2.911 & 2.983 \end{bmatrix}$$

(Nominal) Maximum Likelihood Estimates of Standard Errors

$$\beta^*: \begin{bmatrix} .147 & .161 \\ .180 & .197 \end{bmatrix}$$

$$\lambda^*: \begin{bmatrix} 1.350 & 1.093 \\ 1.093 & 1.615 \end{bmatrix}$$

Ratio of the Estimates to Their Estimated Standard Errors

$$\beta^*: \begin{bmatrix} 3.6 & -1.8 \\ -5.89 & 2.036 \end{bmatrix}$$

$$\lambda^*: \begin{bmatrix} 4.123 & -1.734 \\ -1.734 & 4.123 \end{bmatrix}$$

Likelihood Test of $H_0: \beta = \rho_0$ [Assuming $\lambda = \rho_0$]

19.72 asymptotic chi-square 4df

5. DISCUSSION

We have developed and demonstrated tests for detecting the presence of serial correlation in the error structure of a cross-lagged panel study. Allowing the error structure to contain serial correlation may extend the usefulness of the multivariate regression model since observations made on a single subject at two waves are rarely independent particularly if the waves are close.

A second method of allowing dependence in the error structure has been formulated by econometricians for the univariate continuous variable panel model (eg, Balestra and Nerlove (1966), and Wallace and Hussain (1969)). In this formulation the error for a given observation can be decomposed into the sum of a pure error and replication effect. The replication effect captures the tendency for sampling units (subjects) that are above the regression line at the first wave to stay above the regression line across waves. Our work with business data has convinced us that this error structure is often more realistic than the independent error structure and is a viable competitor to the serial correlated structure.

Anderson and Tsiao (1981; 1982) have studied the problem of estimating the parameters of the univariate panel model with this decomposable error. We are currently extending their results to the multivariate model and considering the problem of distinguishing between errors that are serially correlated and errors that are decomposable.

*The authors gratefully acknowledge discussions with colleagues and students. Particular thanks go to T. W. Anderson and D. R. Rogosa.

BIBLIOGRAPHY

- Anderson, T.W. (1978) "Repeated Measures on Autoregressive Processes", *Journal of the American Statistical Association*, 73, 371-378.
- Anderson, T.W. and Tsiao, C. (1981) "Estimation of Dynamic Models with Error Components", *Journal of the American Statistical Association*, 76, 598-606.
- Anderson, T.W. and Tsiao, C. (1982) "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47-82.
- Balestra, P. and Nerlove, M. (1966) "Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand of Natural Gas", *Econometrica*, 34, 585-612.
- Campbell, D.T. (1963) "From Description to Experimentation: Interpreting Trends as Quasi-experiments", in *Problems in the Measurement of Change* (C.W. Harris, ed), Madison: University of Wisconsin Press, 212-254.
- Crano, W.D., Kenny, D.A. and Campbell, D.T. (1972) "Does Intelligence Cause Achievement: A Cross-Lagged Panel Analysis", *Journal of Educational Psychology*, 63, 258-275.
- Duncan, O.D. (1969) "Some Linear Models for Two-Wave, Two-Variable Panel Analysis", *Psychological Bulletin*, 72, 177-182.
- Durbin, J. (1970) "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors are Lagged Dependent Variables", *Econometrica*, 38, 410-421.
- Eaton, W. (1978) "Life Events, Social Supports, and Psychiatric Symptoms", *Journal of Health and Social Behavior*, 19, 230-234.
- Frey, R.S. (1984) "Does n-Achievement Cause Economic Development?" A Cross-Lagged Panel Analysis of the McClelland Thesis, *Journal of Social Psychology*, 122, 67-70.
- Greenberg, D.G., Kessler, R.C. and Logan, C.H. (1979) "A Panel Model of Crime Rates and Arrest Rates", *American Sociological Review*, 44, 843-850.
- Guilkey, D.K. (1975) "A Test for the Presence of First Order Vector Autoregressive Error When Lagged Endogenous Variables are Present," *Econometrica*, 43, 711-717.

- Hannan, M.T. and Young, A.A. (1977) "Estimation in Panel Models: Results on Pooling Cross-Sections and Time Series," Sociological Methodology, 1977 (D.E. Heise, eds.) San Francisco: Jossey-Bass, 52-83.
- Heise, D. (1970) "Causal Inference from Panel Data, Sociological Methodology", 1970 (E. Borgatta and G. Bohrnstedt, eds.) San Francisco: Jossey-Bass, 3-27.
- Johnston, J. (1972) Econometric Methods, 2nd Edition, New York: McGraw-Hill.
- Judge, G., Griffiths, W., Hill, A., Lee, T. (1980) The Theory and Practice of Econometrics, New York: Wiley.
- Kessler, R.C. and Greenberg, D.F. (1981) Linear Panel Analysis: Models of Quantitative Change, New York: Academic Press.
- Lazarsfeld, P.F. (1948) "The Use of Panels in Social Research", in Continuities in the Language of Social Research (P.F. Lazarsfeld, et.al. eds.) New York: Free Press, 330-337.
- Malinvaud, E. (1980) Statistical Methods of Econometrics, Amsterdam: North Holland.
- Markus, G. (1979) Analyzing Panel Data, Beverly Hills: Sage.
- Mayer, L.S. (1984a) Statistical Inferences in Cross-Lagged Panel Studies, Technical Report, Department of Statistics, Stanford University.
- Mayer, L.S. (1984b) On Cross-Lagged Panel Model with Serially Correlated Errors, Technical Report, Department of Statistics, Stanford University.
- Pelz, D.C. and Andrews, F.M. (1964) "Detecting Causal Priorities in Panel Study Data", American Sociological Review, 29, 836-848.
- Polachek, S.W. and McCutcheon, E.P. (1983) "Union Effects on Employment Stability: A Comparison of Panel Versus Cross-Sectional Data," Journal of Labor Research, 4, 273-285.
- Rogosa, D.R. (1980) "A Critique for Cross-lagged Correlation", Psychological Bulletin, 88, 245-258.
- Wallace, T. and Hussain, A. (1969) "The Use of Error Components Models in Combining Cross Section with Time Series Data", Econometrica, 37, 55-72.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 401	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Detecting Serial Correlation In The Error Structure Of A Cross-Lagged Panel Model		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Lawrence S. Mayer and Steven S. Carroll		8. CONTRACT OR GRANT NUMBER(s) N00014-86-K-0156
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 1111		12. REPORT DATE February 2, 1988
		13. NUMBER OF PAGES 24
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Regression; Panel analysis; Serial correlation; Autoregressive errors		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) PLEASE SEE FOLLOWING PAGE.		

20. ABSTRACT

Cross-lagged panel studies are statistical studies in which two or more variables are measured for a large number of subjects at each of several waves or points in time. The variables divide naturally into two sets and the primary purpose of analysis is to estimate and test the cross-effects between the sets. Such studies are found in the mainstreams of social, behavioral and business research. One approach formulates a multivariate regression model in which the cross-effects are parameters. We contribute to this approach by considering the problem of testing whether the regression model should allow for serial correlation in the error structure. We demonstrate the tests developed by considering a panel study of the attitudes of patients toward the health maintenance organization in which they are enrolled.

END

DATE

FILMED

MARCH

1988

DTIC