ESTIMATION IN PANEL MODELS:

RESULTS ON POOLING CROSS-SECTIONS AND TIME SERIES*

Michael T. Hannan

and

Alice A. Young

Stanford University

Technical Report #51

0 < ? 7 f

*This research was supported by NSF Grant GS-32065. Aage S^rensen, Neil Henry and Arthur S. Goldberg&ftfede helpful comments on an earlier version.

ESTIMATION IN PANEL MODELS:

RESULTS ON POOLING CROSS-SECTIONS AND TIME SERIES

I. INTRODUCTION

Panel designs, in which the same sample of units is observed at more than one point in time, are becoming increasingly more common in sociological research. Yet, many of the central methodological issues which arise in such designs are not yet well understood. We address ourselves to three related issues: (1) alternatives in the formulation of panel designs, (2) estimation problems in conventional panel designs, and (3) estimation in "pooled" panels. The key design issue concerns the treatment of panels with multiple "waves" of observations. The estimation problems of interest arise due to the likelihood that errors in a panel design will be autocorrelated. A "pooled" model in which all waves are analyzed in a single model is advocated as a possible solution to both design and estimation problems for multi-wave panels. Since there is no treatment of "pooled" models in the sociological literature, we discuss the estimation issues in some detail. Finally, we present results of a Monte Carlo simulation of the behavior of alternative estimators for pooled models.

II. PANEL MODELS IN SOCIAL RESEARCH

The increased popularity of panel analysis reflects, in part, wider availability of comparable data over time and, in part, intellectual concerns in social science. We briefly consider two quite different motivations for adopting panel designs.

The first, and perhaps most pervasive, attraction of panel designs is as a methodological strategy for unraveling reciprocal causation. It is widely known that the, presence of "causal loops" greatly complicates cross-sectional analysis. -In particular, ordinary least squares regression is inappropriate for such cases. Researchers faced with suspected reciprocal causal structures are usually advised to use simultaneous equations estimators (e.g., two-stage least squares). These procedures resolve the problem by using "instruments," i.e., variables which are not involved in the "feedback cycle" but which have causal effects on some but not all of the variables in the cycle.* If a set of instruments can be found which exhibit the proper pattern of relationships with variables in the model, simultaneous equations procedures will resolve the analytic difficulty caused by "feedback. But, the value of the instrumental variables strategy depends heavily on the researcher's knowledge of the relationships of instruments to variables in the model.[2] If one does not have such detailed knowledge of the behavior of the system, it is tempting to look for alternative solutions.

It has apparently occurred to many social researchers to use longitudinal (or intertemporal) variation to disentangle reciprocal causal effects. More precisely, it is suggested that lagged values of variables involved in feedback loops be treated as instruments. This is the logic underlying Lazarsfeld's famous sixteenfold table for panel analysis (Lazarsfeld, 1972; Boudon, 1968) and the "cross-lag correlation" design proposed by Pelz and Andrews (1964), and Campbell and Stanley (1963). In both cases one begins with a pair of variables, X and Y, assumed to causally affect each other. To evaluate the effect of X on Y, one regresses Y on lagged X and lagged Y and to evaluate the effect of Y on X, one regresses X on *-laggi_J Y and lagged X.[3] In each case the lagged dependent variable is treated as an instrument.

Duncan (1969, 1972) has correctly deflated the more overblown claims for the power of this method. He showed, for example, that in the presence of both lagged and instantaneous causal effects, the use of only lagged dependent variables as instruments will not suffice to identify the structure (i.e., yield a unique solution in terms of causal or structural parameters). Further, introducing lagged dependent variables gives rise to a number of estimation difficulties. As we will see below, lagged dependent variables are not proper instruments when the disturbances are autocorrelated. Yet, one should not overlook the possible advantages of the strategy of using lagged variables as regressors. The addition of inter-temporal variation cannot decrease one's knowledge of the causal structure. The central methodological problem of panel analysis is to exploit inter-temporal variation in such a way as to simplify causal inference.

A closely related motivation for panel analysis arises from work with models containing unobservable variables. Such models confront measurement and other analytic difficulties by inserting into structural equations models both measured and unmeasured variables. The use of unobservables will ordinarily lead to problems of identification unless strong restrictions are placed on the model. One possibility that occurred to a number of sociologists (Heise, 1969; Blalock, 1970; Duncan, 1972; Hannan et al., 1974) is to measure the same variables at multiple points in time and presume that the causal relations under study are time-invariant. Under a limited number of conditions this strategy leads to identification of multi-variable, multi-wave panel models containing unobservables. The main point for present purposes is that this use of the panel design has essentially the same motivation as Lazarsfeld's: use temporal variation to eliminate identification problems.

t

A second, perhaps deeper, motivation for panel analysis is an interest in dynamics. Sociology has been overwhelmingly preoccupied with static models of social organization and social behavior. In recent years, however, the limitations of a completely static approach have begun to make an impression, and a number of sociologists are experimenting with dynamic formulations. Dynamic models are typically analyzed in other sciences by the study of a single unit over many time periods. Unfortunately, many research areas in social science do not contain observational series sufficiently rich for time series analysis.

Under at least some conditions one can study dynamics with short time series by analyzing panel observations. Coleman (1968) has made this clear in the following way. Consider the following differential equation:

$$| \, | = a + bY + cX \ . \tag{1}$$

This equation relates the rate of change in some variable, Y, to its own level and to some causal variable, X, and is explicitly dynamic. To estimate (1) it is necessary to relate it to observations made at discrete time intervals. The usual procedure is to integrate (1) to yield an expression:

$$y_t = <a/b)(e^{bAt} - 1) + e^{bAt} y_{fc\_k} + <c/b)(e^{bAt} - 1) \ x_{fc\_k} \ . \tag{2}$$

This may be rewritten as

$$y_t = a^* + b^* y_{t\_k} + c^* x_{t\_k} \ . \tag{3}$$

Coleman suggests collecting panel observations for at least two periods, say t and $t-k$, and estimating (3) by ordinary least squares. Then one can use (2) to transform estimates of (3) into estimates of the parameters of the differential equation.**

The estimation form of Coleman's model is identical to that most frequently used in estimating "cross-lag" correlations. The points of similarity are the inclusion of lagged values of the dependent variables and the dependence on panel observations. Given these similarities, the dominant concerns of both strategies can be reasonably well represented by the following two equation models (for N individuals at T time periods):

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 y_{i,t-1} + \beta_2 x_{i,t-1} + u_{it}, \quad (i = 1, \ldots, N; \ldots) \\ x_{it} &= \gamma_0 + \gamma_1 y_{i,t-1} + \gamma_2 x_{i,t-1} + w_{it}. \end{aligned} \tag{4}$$

Here the introduction of lagged dependent variables in both equations gives rise to the dynamic character of the model. Further, they serve as "instruments" (or perhaps only "pseudo-instruments") for estimating and testing. We take the model in (4) as the methodological point of reference. All that follows is addressed to researchers who, for one reason or another, are interested in estimating models similar to (4).

III. CONSTRUCTION OF MULTI-WAVE PANEL MODELS

The sociological literature offers little in the way of didactic treatments of the handling of multi-wave panels.^ Let us consider the major alternatives in the context of a substantive example. Suppose a researcher, interested in the effects of ethnic heterogeneity on levels of political violence, collects observations for a sample of N nations for the years 1950, 1955, 1960, 1965, and 1970. If the appropriate causal lag is five years, the data yield four usable lag periods: 1950-55, 1955-60, 1960-65, and 1965-70. The question is how to utilize all four "waves."

' This situation seems to admit of three main alternatives (other than ignoring some of the data completely):

A. Average observations over time and analyze the averages. For example in the regression of levels of violence on ethnic diversity, the dependent variable for each nation would be the average level of violence for five different periods, etc. Thus the 5N observations are compressed into N observations. Procedures like this always result in a loss of statistical efficiency (i.e., increase the standard error of estimators). Further, grouping over time or over units gives rise in many cases to complicated aggregation problems (cf. Theil, 1954; Hannan, 1971).

B. Conduct separate analyses for each lag period. For example, estimate the model for 1950-55 (with N observations), then again for 1955-60, and so forth. As in the previous case, this procedure sacrifices statistical efficiency. It does, however, avoid aggregation problems. An advantage of this procedure is that it may uncover changes in the causal structure over time. If the nature of the relationship among variables is changing over time, estimated causal effects will differ from period to period. Any such time nonstationarity can be very damaging to inference. To check for such problems, one should conduct analysis in this form before moving to the alternative discussed below. If the causal structure does not appear to be constant over the entire period, there is no point in pooling observations. In such cases, the analyst must first attempt to identify the source of the change in causal structure and, if possible, modify the model to take its effects into account.

There is a major difficulty that cannot be addressed within this and the previous strategy. The source of the difficulty is autocorrelation of disturbances. Disturbances will tend to be autocorrelated if the variables omitted from the model are stable over time (cf. Heise, 1970). For example, we might expect the repressive power of the state to be relatively stable over the period of observation. If this variable were not included in the model, its presence in the disturbance would tend to make the errors autocorrelated.

Whenever a lagged dependent variable is included in the model, the errors will be correlated with at least one of the regressors and ordinary least squares regression will be biased and inconsistent. The problem with the two approaches just discussed is that, although the analyst may acknowledge the existence of autocorrelation in the disturbances, he does not have enough information to test for such effects or to modify the analysis to take the problem into account. This limitation, together with the possibility of gains in efficiency when all the observations are used in a single model, are the main motivations for considering the next alternative.

C. <u>Pool the lag structures into a single model</u> (more generally pooling
g
the time series of cross-sections). For example, conduct a least squares analysis on all 5N observations simultaneously. The data may be arranged such that the "pooled" dependent variable consists of the political violence scores for the first nation in 1955, 1960, 1965, and 1970, then the four scores for the second nation at those same times, and so on for all N nations. Observations for each independent variable are arranged analogously. More generally, there are NT observations for each regressor (N individuals, or other elementary units, each measured at T points in time).

Unlike the previous case, only a. single set of structural or causal parameters is estimated. The single set of parameters are substantively meaningful only if the analyst has reason to believe that the causal structure does not undergo change during the observation period. Whether or not it is desirable to employ such a strong hypothesis depends on the purposes of the analysis.

However, deductively oriented sociologists with abstract theoretical concerns ought to welcome the opportunity to formulate and test such "ahistoric" hypotheses.

If restriction to a single set of causal parameters is appropriate, the pooling method yields a considerable gain in efficiency. It is immediately obvious, however, that one does not have 4N "independent" outcomes. Rather, the amount of independent variation lies somewhere between that contained in N and 4N independent observations. We shall see below the magnitude of autocorrelation of disturbances determines whether it is closer to N or 4N . When the autocorrelation of disturbances approaches unity, one gains very little new information from each additional wave. When the autocorrelation approaches zero, each new wave contains essentially as much new information about the causal structure as the previous wave.

Pooled models allow explicit consideration of autocorrelation problems. This is its main advantage over the other alternatives mentioned. To evaluate the potential contribution of pooled models we must proceed more formally. In the next section we state the conditions under which pooling is a useful procedure and collect analytic results on estimation in pooled models. In so doing, we will continue to contrast pooled models with the other two alternatives considered here.

IV. ANALYTIC RESULTS ON POOLED MODELS

In this section we collect available analytic results on pooled models. The simplest models are considered first. The complications are introduced, first singly and then jointly. The models considered are those which arise in the "cross-lag correlation" and Coleman approaches. We begin with the following one-equation model:

$$y_{it} = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 x_{i,t-1} + u_{it} \ , \quad (i = 1 \ldots N; \ t = 1,\ldots,T) \ , \quad (5)$$

where the $x_{i,t-1}$'s are stochastic,[9] and the u_{it}'s are unobservable disturbances with

$$
\begin{aligned}
& u_{it} = \alpha_i + v_{it} \ , \quad \text{all } i, t \ , \\
& E\alpha_i = Ev_{it} = 0 \ , \quad \text{all } i, \ t \ , \\
& E\alpha_i v_{i't} = 0 \ , \quad \text{all } i, \ i', \text{ and } t \ , \\
& E\alpha_i \alpha_{i'} = \begin{cases} \sigma_\alpha^2 \ , & i = i' \\ 0 \ , & \text{otherwise} \end{cases} \\
& Ev_{it} v_{i't'} = \begin{cases} \sigma_v^2 \ , & i = i' \ , \ t = t' \\ 0 \ , & \text{otherwise} \end{cases} \\
& E\alpha_i x_{i't} = 0 \ , \quad \text{all } i, i', \text{ and } t \ , \\
& Ev_{it} x_{i't'} = 0 \ , \quad \text{all } i, i', t, \text{ and } t' \ , \\
& Ev_{it} y_{i',0} = 0 \ , \quad \text{all } i, \ i', \text{ and } t \ .
\end{aligned}
\qquad (6)
$$

With one exception, these are the usual least squares assumptions. The peculiar feature of this model is the specification of the disturbance. It contains a unit-specific component α_i and the usual random component v_{it}. In other words, the model contains an unobserved variable, α_i, which is constant for each individual but varies across individuals. This term summarizes all of the unobserved causes of Y which are relatively constant over time. In our substantive example, the factors summarized in α_i might include specific historical experiences such as the nature of the elite at the time of state formation, or perhaps features of the legal system. Whatever the substantive context, it is quite likely that some such unobserved variables will be operating. To the extent that they remain constant over time, they will conform to the model outlined above.

It is quite natural to extend the model to incorporate time-specific effects. These are effects which vary from period to period but which are constant across units within any period. For example, level of political violence in every polity might be affected in the same way by widespread economic booms or busts. More generally, the introduction of time-specific error components is likely to be useful whenever all (or most) units are affected by environmental variations in the same sorts of ways.

As long as the time-specific effects behave similarly to the defined above, no additional analytic issues arise when both types of effects are included. Consequently, there is no need to complicate the algebra in the discussion which follows by including time-specific terms.

The model stated in (5-6) contains what amounts to a factor-analytic structure for the disturbances (Goldberger, 1973a: 17). In particular, disturbances for the same unit are correlated at the same magnitude no matter how distant they are in time. Unfortunately we do not yet have sufficient experience with panel analysis utilizing varieties of estimation technologies over a wide enough class of substantive situations to know how appropriate this error specification will be to researchers. This assumption may be unrealistic in long time series (where the more usual autoregressive scheme will probably be more appropriate). In relatively short series, as commonly found in panel analysis, the assumption may be a good approximation. One aim of this paper is to stimulate interest in examination of such problems.

It will be convenient to employ matrix expressions for (5-6):

$$Zt = \quad + "t \gg \tag{7}$$

where

y

y_t = ^y11'y12. . . .y1T'y21'*** "*" 'yN1'***'yNT^' *

Zt-1 " (y10'y11'"''y1,T-1'y20. . . .y2,T-1""'yN0""'yN,T-1)1 »

t-l = <xio»xii»'"»'xi}T_i»x20. . . .*2,T-1. . . .*N0*''''xN,T-1)'

"t = 'ui2'*'*'U1T'U21'*"*'U2T'""*'UN1'* *"'UNT^' '

Q_t • Ci»y_{t_1}»x_{t_1}) » {where ' is an NT x 1 vector of ones} ,

and

J ™ ('%'^'

Further,

$$Euu^* = \begin{matrix} A & 0 & & \\ 0 & A & & \\ & & \cdot & \cdot \\ 0 & 0 & \ldots & A \end{matrix}, \text{ an (NT x NT) matrix ,} \quad (8)$$

where

$$a^2 = a^2_{p,} + a^2_v , \quad \underset{R}{p} = \alpha^2 / (a^2 + a^2) \quad , \text{ and } \underline{A} \qquad , \text{ a (T x T) matrix} \quad (9)$$

P P

It is helpful in presenting available analytic results to begin with special cases of (7) before considering the model in its entirety. We begin with the case in which the lagged dependent variable does not appear in the model (i.e., = 0) .

Case 1: ^ = 0

When the lagged dependent variable does not appear in the model, it is .reasonable to expect that the disturbances will be uncorrelated with regressors. As we have formulated the model in (6) this will be the case. Consequently, the ordinary least squares estimator (OLS) of $J3_0LS$ = ^ ^tZt is consistent.

To consider the efficiency of OLS we must refer to the variance-covariance matrix of the disturbances, $E\underline{u}_t\underline{u}_t^*$. OLS is efficient if and only if

2

$^{Eu}t^t$ = a * ' *-s» t*ie variance-covariance matrix of disturbances has constant variance and all zero covariances. Clearly, by (8-9), OLS is efficient if and only if p = 0 .

At this' point it is natural to search for a consistent estimator which avoids the problem in the disturbances. The existence of such an estimator is suggested by the fact that we can transform (8) in such a way as to produce "well-behaved" disturbances. What we need is to find a matrix fi which when applied to (7) yields

$$Q^{-1/2} y_t - a^{1/2}{}_{st} | + o^{1/2} u_t \quad , \tag{10}$$

$$E[ff^{1/2} u\hat{}u^* \; n^{1/2}] = o^2 i \quad . \tag{11}$$

Nothing in the causal structure has been changed and we can apply ordinary 1n*st spires *to* (10r>. Because of (11) OLS applied to the transformed model is now an efficient estimator. The gain in efficiency relative to OLS applied to (7) arises -from the explicit consideration of correlated errur.

The procedure suggested in (10) is an application of the widely useful <u>generalized least squares</u> (GLS) approach to estimation. The application of GLS to pooled models is commonly advocated in the econometric and biometric literatures (Nerlove, 1971; Searle, 1971).

Since we will make continued reference to the GLS estimator, we need a somewhat more formal representation. The GLS estimator is defined as

JGLS - «t£:1 itff1 If (12)

where

A"1 o 0

0 A^{-1} 0

n1 =

0 0 . . . A^{-1}

and (cf. Hannan and Young, 1974)

A = (l/rj)(I_T - _U'/T) + (l/ipCM'/T) , (13)

where T| = (1 - p) , § = (1 - p) + Tp , and J_ is a (T x 1) vector of ones.

The form of the GLS transformation (13) can be intuitively motivated as follows. The peculiar feature of pooled models is the use of both cross-sectional (between-unit) and longitudinal (within-unit) variation to estimate causal parameters. The richness of the data presents an implicit choice: how to weight one type of variation relative to another. Generalized least squares uses p to weight the two types of information. To see this, consider the case where p = 0 . Then $A^{-1/2}$ * —x an(* observations are transformed in (10) by an identity transformation. GLS reduces to OLS where cross-sectional and time series variation are weighted proportionately to N and T (see Maddala, 1971). At the other extreme, when p = 1 , $A^{-1} = \ell\ell'/T^2$. It is easy to show that this transformation averages observations over time for each unit (as in the first of the broad approaches discussed in Section III). The result is a regression on grouped observations where all of the weight is placed on cross-sectional variation. In cases where p takes on a value 0 < p s 1 , GLS weights time series variation inversely to p . Such a weighting seems appropriate since p measures "redundancy" in the time series. The more redundancy, the lower the weight attached to longitudinal variation.

The gain in efficiency of GLS relative to OLS arises because GLS uses a weighting based on population or sample information rather than an arbitrary weighting (cf. Goldberger, 1973b). Such a gain can be realized only if p is known a priori or can be estimated from sample data in a consistent (or unbiased) manner. There are no realistic cases where sociological researchers will have prior knowledge of p . So we turn to a discussion of procedures for estimating p .

The most widely used procedures for estimating p involves introducing dummy variables for each unit into (7):

$$y_t - QfcS + Ay + V \tag{14}$$

where y is an ((N-l)x 1) vector containing the ^ , and = (, »• • • and is an (NT x 1) vector with ones corresponding to observations on unit i and zeroes elsewhere. We point out below that estimates of y can be used Lo construct a useful estimator of p . But before moving to that discussion, we consider (14) more closely. Notice that the y.'s that give rise to the correlation of disturbances have been shifted out of the disturbance. The remaining portion of the disturbance, ^ > well-behaved. In particular, $Ev^j. = o_vI$. So OLS applied to (14) is an asymptotically efficient estimator (cf. Amemiya, 1967). We refer to this estimator as least squares with dummy variables (LSDV).

In cases where N is moderately large, calculation of LSDV by direct methods will be either costly or impossible within conventional regression programs. An alternative approach is more generally useful. Express all observations as deviations from each unit's mean and apply OLS. The estimates of p are identical to those obtained using (13) directly and (1 can be recovered by operations analogous to those used in recovering the constant in regressions with deviation scores (see Hannan and Young, 1974).

The results of LSDV can be used to calculate p as follows. To estimate p we need an estimate of a^2 • Nerlove (1971) suggests

$$s> \sigma - \sum_{i=1}^{N} \hat{} \; X_{i=1}^{N} \; v^{n)2/n} \bullet$$

An obvious estimator of a^2 is the sum of squared residuals from the LSDV regression divided by NT . Then

$$P = \quad \cdot \qquad (15)$$

Nerlove chose p in (15) over a maximum likelihood estimate to avoid negative values of p (which are implausible in most applications). Unfortunately the estimator in (15) is upwardly biased (at least in small samples) with the magnitude of the bias inversely related to p .

Recall that GLS requires consistent estimates of p . The bias in p does not, however, appear to unduly damage the resulting GLS estimators

(Amemiya, 1967). We study this issue further below. To acknowledge the fact that we are using estimates of p rather than the true values, it is more precise to refer to this estimator j^g_{LS} = Wt^St^"[1] St^f[1] It as modified generalized least squares (MGLS).

We are interested in comparisons between MGLS and LSDV. Amemiya (1967) has proven for the limiting case of N •» 1 that GLS and LSDV are both asymptotically efficient. But, the weighting of cross-sectional and time series variation seems more arbitrary than in MGLS (Maddala, 1971; Nerlove, 1971). Further, since one loses a degree of freedom in the stimation of each in LSDV, LSDV appears to "waste" degrees of freedom relative to GLS. Nerlove's (1961) simulation also suggests that GLS is more efficient than LSDV in small samples. We investigate the small sample behavior of the two estimators below.

Only one further issue remains to be discussed for the case = 0 . We have assumed throughout that disturbances are uncorrelated with lagged X . In many substantive cases this will not be true. When lagged X is correlated with ^ , OLS is inconsistent but both MGLS and LSDV are consistent. Since the correction for p "cleans" the equation of this specification bias, the choice among the estimators involves not only efficiency but also consistency considerations when the factor denoted may be correlated with the regressor.

Case 2: $p_2 = 0$

Since the (J-^'s affect y^$_t$ at every period, it is clear that the lagged dependent variable is correlated with the disturbance. OLS applied to this model is inconsistent and (for p > 0), f^ is biased upwards. OLS gives "credit" to lagged y for the causal effects of ^L . The stronger the effects of [j^ relative to lagged y , the greater the upward bias. In most realistic cases the upward bias is quite high.

Since both MGLS and LSDV take ^ into account (LSDV directly and MGLS via p), they are both consistent estimators of . Again, both estimators are asymptotically efficient and choice between the two depends on knowledge of finite sample properties.

Case 3: ^ ^ o , P_2 ^ 0

The "full" model does not introduce new complications. It is worth pointing out that the presence of ^ tends to lessen upward bias in a estimated by OLS (Malinvaud, 1971:558). And, if , 3_2 and p are all positive, and will be negatively correlated (Johnston, 1971:161). So upward bias in OLS estimates of will tend to bias ^ downward. This means that OLS applied to (7) will ordinarily lead one to overstate the lagged effects of the dependent variable and understate the effects of other (positively correlated) independent variables. As in Case 2, both MGLS and LSDV are consistent and asymptotically efficient estimators for this case.

The full model admits another approach to estimation. Inconsistency in OLS estimates arises only from the correlation of μ, and . Since we have assumed x_{fc} ^ to be uncorrelated with p, , ^ is a proper "instrument" for y_t ^ . Instrumental variables (IV) estimators for and are calculated by first regressing ^ on $x_{t\ 2}$ an(^ then substituting y_{fc} ^ , the predicted or "fitted" values from the first stage, into (7). Then OLS applied to this revised model is IV and gives consistent estimators. Since the IV estimator does not correct for the correlated errors, IV is less efficient than either MGLS or LSDV.^ For these reasons, IV has little to recommend for the model defined here.

In all three cases analytic considerations lead to a clear-cut choice of MGLS or LSDV over OLS (or IV). But analytic results may not be directly useful for social researchers working with small samples. Both MGLS and LSDV are <u>asymptotically</u> efficient (meaning that they attain the minimum possible variability in an infinite sample). But it is important to know how the two compare in moderate and small samples. Further, the consistency results are also large-sample results. It is conceivable that OLS may out-perform MGLS and LSDV in terms on bias and mean squared error in small or moderate samples.

Nerlove (1971) applied a maximum likelihood procedure to a model similar to (8) with very disappointing results. While there is not reason to expect the method of maximum likelihood to be useful in small samples, Nerlove's method may not have been optimal. We are conducting further work in this area but for the moment, we omit any treatment of maximum likelihood.

Case 4: Continue to focus on (7) but assume that (7) is one equation in a two-equation system composed of (7) and = Yq + y_{fc} + x_t + w

The addition of this second equation to (7) completes the "cross-lag" structure so widely used in social research. Our concern is with properties of estimators of and when (7) does not stand alone but is embedded in the full causal structure. The main consequence is to complicate the role of x_t ^ • Because of the causal feedback, x_{fc} ^ will no longer be uncorrelated with u_{fc} (in fact, x can be written as an explicit function of). OLS will no longer necessarily understate the causal importance of Xj. ^ > as in Case 3. Rather, it becomes much more difficult to evaluate complications arising in OLS. Since neither GLS nor LSDV applied to (7) correct for the correlation of ^ with u_{fc} , neither are consistent estimators.

This latest complication is an instance of the familiar simultaneous equations bias (see Goldberger, 1973a: 4). No single-equation method (such as OLS, MGLS, LSDV, or even "true" GLS) copes with the problem. As mentioned earlier, a number of simultaneous equations estimation procedure are available for such situations.

In the case at hand, however, no statistical procedure will resolve the difficulty as the model stands. All of the causal variables are involved in the feedback and thus every regressor is correlated with the two disturbances. No "instruments" exist for consistent "first-stage" estimation as was described for Case 3.

Suppose that a proper instrument for ^ does exist. That is, suppose that the analyst can find a variable, z , that is correlated with ^ , but is uncorrelated with the disturbance . We make reference again to the instrumental variables method. In the first stage, regress x_t ^ on z to produce x_t ^ into (7) and use OLS.

If there were no correlated errors in (7), use of IV would eliminate all inconsistency. But the errors in (7) are correlated and the lagged dependent variable will continue to be correlated with the disturbance. Only if p, = 0 will IV give consistent estimates.

It is, however, true that IV corrects for the "feedback" or simultaneity aspect of the estimation problem. It seems natural, then, to combine IV with one of the methods which corrects for the pooled disturbance correlations, MGLS or LSDV. In fact, both composite estimators, IV-MGLS and IV-LSDV, are consistent. Neither IV-MGLS nor IV-LSDV are maximally efficient. In general, estimators that make use of restrictions sequentially are less efficient than those that use them simultaneously.^ It will be important to develop applications of more efficient estimators for this case. In what follows, we restrict our attention to the two "synthetic" estimators IV-GLS and IV-LSDV.

To summarize this section, we note first that pooling cross-sections in the "full" cross-lag model involves both the "time series" bias discussed earlier

»

and "simultaneous equations bias." The combination of the two types of bias has not (as far as we know) been discussed in the literature. None of the widely used estimators (OLS, LSDV, MGLS, or IV) are consistent in this case. We find that a combination of IV and LSDV or IV and MGLS leads to consistent estimation. Neither synthetic estimator is asymptotically efficient.

Throughout the discussion of, the four cases we have relied on large-sample theory. As we mentioned earlier, it is important for empirical researchers to obtain some information about the behavior of such estimators in small and moderate sized samples. Two issues are important here. We want to compare the efficiency of the various consistent estimators in finite samples. And we also want to compare the performance of the consistent estimators with those of inconsistent estimators (e.g., OLS) which may have smaller mean squared error in small samples (cf. Hurd, 1972). Since these issues have not yet been dealt with analytically, we turn next to a Monte Carlo simulation.

V. STRUCTURE OF THE SIMULATION

A. The Model

The following model allows us to study both types of bias discussed above:

It + , (16)

*t-i = £t-i Y4 3t-r

where the u_t behave as specified in (3), and

2// 2 ^ 2 .
P • V (C V + a v> *

Cov(u i t W i , t-1) = * > E ^ t = CT w I '

Although we have not simulated the full cross-lag model, allowing the disturbances from the two equations to be correlated (1t ^ 0) introduces exactly the same type of simultaneous equations bias as was identified in the discussion of Case 4.

We chose to restrict our attention to panels with five waves and fifty observations per wave, i.e., $T = 5$, $N = 50$. Two of the parameters of the model, and Y » are fixed (at unity) and the remaining parameters were varied as follows:

$$p - 0, .25, .5, .75, .9, .95 ,$$

$$y - 0, .5, .9 ,$$

$$* \bullet 3, .8 .$$

For each of the thirty-six combinations of values of p , \i , and , we generated one hundred samples with N • 50 , T =» 5 . The properties of the samples is the focus of our investigation. Further details of the simulation and the routines used for calculating estimators are presented in Young, Nielsen and Hannan (1974).

B. The Estimators

The estimators studied are those described above. Here we restate the main asymptotic results (and note methods of calculation where they are not obvious).

1. Ordinary least squares (OLS) applied to (18). Consistent and efficient when p * 0 and ty • 0 ; inconsistent, otherwise.

2. Least squares with individual constants (LSDV) applied to (18) by introducing dummy variables for each unit. Consistent and asymptotically efficient when ty • 0 ; inconsistent otherwise.

3. Modified generalized least squares (MGLS) applied to (18) with p calculated as in (15) from an LSDV first stage. Consistent and asymptotically efficient when \i • 0 ; inconsistent otherwise.

4. "True" generalized least squares (GLS) applied to (18) using known values of p . Minimum variance consistent estimator.

5. Instrumental variables (IV) applied to (18) with Xt-1 replaced by "fitted values" from a regression of Xt-1 on -5t-l • Consistent when p • 0 ; inconsistent otherwise.

6. IV-LSDV by applying LSDV to (18) in which x_{t-1} has been replaced by fitted values. Consistent but not asymptotically efficient regardless of values of p and \|r .

7. IV-MGLS by applying GLS to (18) with xt-1 replaced by fitted values and using p calculated from IV-LSDV estimates of u_t . Consistent but not asymptotically efficient.

VI. FINDINGS

We report our findings in three forms. First, for each combination of p and ty we calculate mean errors (ME) for each estimator. For example, the mean error of ^ is

1Q()

i-1

The ME is a summary of tendency toward directional error. Second, we likewise compute mean squared errors (MSE) : the MSE of is

$$\frac{1}{100} \sum_{i=1}^{100} (Si " M)^2 \cdot$$

The MSE summarizes both absolute amount of error and the variability of the estimator.

Third, to simplify the reporting of tendencies in the behavior of the estimators we present results of regressions of mean squared errors on the parameter values. The regressions take the following form:

MSE - r^j + T^p + + TT_3(p*\jO .

The multiplicative term (p*\p) is introduced as a simple device for locating

12

interaction effects.

1. ty « 0

First we consider the "pure" pooling problem. That is, we examine those conditions in which there is no simultaneous equations bias. Our simulation at this point corresponds closely with Nerlove's (1971). However, he set N • 25 and T = 10 , while we have N = 50 and T = 5 . Few social science panel investigations yield even'five waves of observations.

Attention should be focused on the single-equation methods (OLS, LSDV, and MGLS). First, notice the very poor performances of OLS estimates of both ^ and g_2 . Tables 1 and 2 make clear the very strong effect of p on the behavior of OLS.[13] But the tendency is more marked for estimates of P^ than

Table 1. MEAN ERROR OF ESTIMATE FOR $\hat{ft}_1$
(lp - 0)

6_I - .8

P -	0	.25	.5	.75	.9
OLS	.00*	.09	.13	.16	.17
LSDV	-.13	-.10	-.07	-.04	-.02
MGLS	-.06	-.02	.00	.01	.01

P_1 = .3

P =	0	.25	.5	.75	.9
OLS	.00	.12	.23	.33	.38
LSDV	-.12	-.10	-.08	-.05	-.02
MGLS	-.07	-.04	-.02	-.01	.00

* Figures have been rounded; values listed as .00 actually range between .0003 and .0049.

Table 2. MEAN ERROR OF ESTIMATE FOR $\$_2$
bp = 0)

.8

p =	0	.25	.5	.75	.9
OLS	.01	-.03	-.04	-.05	-.06
LSDV	.00*	.00	.00	.00	.00
MGLS	.01	.00	.00	.00	.00

.3

P «	0	.25	.5	.75	.9
OLS	.01	-.02	-.04	-.05	-.06
LSDV	-.01	-.01	.00	.00	.00
MGLS	.01	.00	.00	.00	.00

*Figures have been rounded; values listed as .00 actually range between .0003 and .0049.

for as expected. OLS estimates of ß^ are biased upwards and estimates of respond with a smaller downward movement. As we noted above, the positive correlation of ^ and x_{fc} ^ yields a negative correlation between and ^2 • Thus any factor which biases upwards will bias f^ downward and the reverse. While results for parameter values • .8 and = .3 are

»

qualitatively the same, the latter case produced, much greater upward bias in estimates of . These results are consistent with those reported by Malinvaud (1971).

Turn next to a consideration of the estimators designed to deal with the pooled model. MGLS always has smaller mean error than LSDV. The difference between the two estimators is small for ^ but quite large in estimating

Examination of mean errors in Table 1 shows that LSDV appears to overcorrect for the presence of autocorrelated error in the sense that estimates of

A

undershoot the mark. The likely explanation is that LSDV attributes to the individual dummy variables jj, part of the causal effect of the lagged % .

There is a slight tendency for MGLS as well to underestimate . The explanation certainly differs from that which applies to LSDV: attention should be focused on the role of p from (15) in MGLS. As may be seen in Table 3, p

Table 3. MEAN ERROR OF ESTIMATE OF
0 => 0)

$\$_1 = .8$

p =	0	.25	.5	.75	-9
MGLS	.32	.32	.22	.20	

B_1 3 .3

P =		.25	.5	.75	.9
MGLS	.23	.22	.15	.07	-02

t

is biased upwards. The upward bias in $\hat{\rho}$ decreases with increases in ρ, and the downward bias in the MGLS estimates of decreases with increasing ρ. This latter observation is consistent with the explanation that upwardly biased $\hat{\rho}$ values produce an overcorrection which will lead to underestimates of P_L.

Table 4 summarizes the quality of the one-equation estimators for the pure pooling problem. The OLS, LSDV and MGLS estimators have the same rank-order of quality as found in Tables 1 and 2. The true GLS estimator is shown as a benchmark, since it is asymptotically efficient and consistent, and it employs the "true" value of ρ. Two items of interest may be found in estimates from the GLS estimation. First, the low values of the mean squared error indicate low amounts of sampling error in the simulation. Second, bias in the estimation of ρ appears to have little effect upon the quality of the MGLS estimator, since the differences between GLS and MGLS are small.

Table 4. MEM SQUARED ERROR AVERAGED OVER ALL CONDITIONS (where p t 0, if = 0, # of cases = 1000)

	$\hat{\beta}_1$	$\hat{\beta}_2$
OLS	.059	.005
LSDV	.005	.001
MGLS	.003	.001
GLS	.002	.001

It is simple to summarize our findings for the pure pooling case. Modified generalized least squares consistently outperforms LSDV. Both MGLS and

>

LSDV are far superior to OLS. The use of the convenient p from (15) does not appear to incur great costs, particularly with high true values of p . Presumably most sociological investigators do face high p values.

2. in + 0

We next introduce simultaneous equations bias into the structure just examined. Since our attention is focused on the estimation of panel models we do not investigate "pure" simultaneous equations bias but only the combination of the two sources of error.

In broad terms we find that allowing \t t 0 , that is, introducing simultaneous equations bias involving x_{fc} ^ , produces very large magnitudes of error, particularly in estimates of • ^ comparison of the single-equation estimators in Tables 4 and 5 shows that the presence of the simultaneity problem

Table 5. MEAH SQUARED ERROR AVERAGED OVER ALL CONDITIONS
(where p i 0, ip 1 0, # of cases = 2000)

	6_1	3_2
OLS	.038	.144
LSDV	.003	.124
MGLS	.001	.131
IV	.065	.156
IV-LSDV	.074	.266
IV-MGLS	.019	.089

produces relatively small changes in the average mean squared error of ,
but quite large increases in the MSE of P_2 • There exists no benchmark estimator comparable to GLS in Table 4 with which to compare the results of this part of the simulation. Here GLS is no longer consistent, as already noted. One could construct an IV-GLS estimator using fitted values of and known "true" values of p , but this estimator is not asymptotically efficient. This latter estimator would thus not allow one to distinguish between the potential inefficiency of an IV estimator and sampling error within the simulation.

We begin with another contrast of the single-equation methods. First, OLS is inferior over the range of parameter values to MGLS and to LSDV (in Table 5). Thus, the choice among single-equation methods is between MGLS and LSDV. Both estimators, in these cases, tend to underestimate 3]_ and overestimate p_2 » as may be seen in Tables 6 and 7. But, LSDV underestimates 6j_ more than does MGLS while MGLS overestimates g_2 more than does LSDV.

The behavior of the estimators over combinations of p and ir is at least as interesting as the gross comparisons between estimators. It is most conveniait to focus on the regression results presented in Tables 8 and 9. Since results of regressions for the case • .3 were similar to those for 3j_ • -8 , only the latter are presented here.

Consider first MSE in estimates of §j_ . As in the case ij » 0 , OLS reacts differently to increases in p than does LSDV and MGLS. As p increases, MSE (in) for OLS increases but the corresponding MSE for LSDV and MGLS decreases. One would expect that an estimator such as OLS which does not correct for an. analytical difficulty such as p would do increasingly worse in the face of greater values of p . Further, increases in ty do not have a pari.i«,iiiariy strong direct effect on any of the three single-equation estimators. But, the combined effects of p and ty cannot be completely accounted

Table 6. MEAN ERROR OF ESTIMATE FOR $\hat{A}$
($\$_1$ =• .8)

\jj^m	0	0	0	0	0
P ^m	0	.25	.5	.75	.9
OLS	.00	.09	.13	.16	.17
LSDV	-.13	-.10	-.07	-.04	-.02
MGLS	-.06	-.02	.00	.01	.01
IV	.00	.10	.14	.17	.18
IV-LSDV	-.30	-.28	-.27	-.25	-.24
IV-MGLS	-.17	-.14	-.12	-.10	-.09
* =	.5	.5	.5	.5	.5
P =	0	.25	.5	.75	.9
OLS	-.02	.07	.11	.14	.15
LSDV	-.12	-.10	-.08	-.05	-.03
MGLS	-.07	-.04	-.02	-.01	-.01
IV	.00	.10	.14	.17	.18
IV-LSDV	-.33	-.32	-.31	-.28	-.26
IV-MGLS	-.20	-.17	-.15	-.13	-.11
=	.9	.9	.9	.9	.9
P -	0	.25	.5	.75	.9
OLS	-.03	.05	.09	.12	.14
LSDV	-.09	-.08	-.07	-.05	-.03
MGLS	-.06	-.04	-.04	-.03	-.03
IV	.00	.10 <	.14	.16	.18
IV-LSDV	-.36	-.34	-.33	-.31	-.28
IV-MGLS	-.22	-.19	-.17	-.15	-.12

t

Table 7. **MEAN ERROR OF ESTIMATE FOR $\hat{8}_2$**
(6_i - .8)

=	0	0	0	0	0
P -	0	.25	.5	.75	.9
OLS	.01	-.03	-.04	-.05	-.06
LSDV	.00	.00	.00	.00	.00
MGLS	.01	.00	.00	.00	.00
IV	-.02	-.24	-.32	-.37	-.38
IV-LSDV	.62	.59	.56	.52	.50
IV-MGLS	.37	.31	.27	.23	.21
4>-	.5	• .5	.5	.5	.5
P =	0	.25	.5	.75	.9
OLS	.43	.37	.29	.19	.11
LSDV	.41	.35	.29	.21	.13
MGLS	.43	.38	.31	.22	.14
IV	-.03	-.24	-.33	-.38	-.40
IV-LSDV	.69	.67	.64	.59	.55
IV-MGLS	.42	.36	.32	.28	.24
rp«	.9	.9	.9	.9	.9
P =	0	.25	.5	.75	.9
OLS	.77	.69	.57	.41	.25
LSDV	.74	.65	.53	.38	.24
MGLS	.77	.67	.54	.38	.24
IV	-.04	-.25	-.33	-.38	-.41
IV-LSDV	.71	.69	.66	.62	.57
IV-MGLS	.43	.38	.34	.30	.26

Table 8. REGRESSION OF MEAN SQUARED ERROR IN $ Oil THE PARAMETERS
(// of cases = 1800, 3 = .8)[1]

Mean Squared Error in;		R^2
OLS	= .0002 + .0309p + .0027^- .0131(p*|0	.92
	(.0003) (.0004) (.0005)	
LSDV	= .0171 - .0189p - .0088ij> + .0114(p*iJO	.68
	(.0004) (.0004) (.0006)	
MGLS	= .0032 - .0038pf+ .0003^ + .0005(p*iJ,)[T]	.45
	(.0002) (.0002) (.0003)	
IV	= .0024 + .0325p - .0008^[f] + .0013(p*i|>)[f]	.83
	(.0005) (.0005) (.0008)	
IV-LSDV	= .0919 - .0383p + .0481^ - .0235(p*ij>)	.33
	(.0033) (.0038) (.0057)	
IV-MGLS	- .0293 - .0211p + .0213^ - .0156(p*iJ/)	.36
	(.0016) (.0018) (.0028)	

'Null hypothesis it = 0 not rejected at .001 level.

t

Table 9. REGRESSION OF MEAN SQUARED ERROR IN 3 ON THE PARAMETERS
of cases = 1800, £ = .8)
1

<u>Mean Squared</u> 2
Error in: R

OLS - -.0419 + ,0414p + .6612\J> - .6529(p*i|>) .92
(.0055) (.0060) (.0092)

LSDV - -.0377 - .0388p + .6043^ - .6013(p*i|j) .92
(.0048) (.0052) (.0080)

MGLS = -.0381 + .0383p + .6380^ - .6539(p*iJ>) .93
(.0047) (.0052) (.0079)

IV = .0378 +:.1408p + .0202^ + .0142(p*^)f .19
(.0120) (.0132) (.0202)

IV-LSDV = .4471 - .1366p + .191(ty - .0997(p^)t .10
(.0311) (.0343) (.0523)

IV-MGLS = .1667 - .1124p + .0098^ - .0536(p*^)'i" .15
(.0139) (.0154) (.0234)

'Null hypothesis it = 0 not rejected at .001 level.

for by the two additive effects. That is, p and i|f "interact" to produce consequences which cannot be anticipated from consideration of the distinct effects of each. And, these interaction effects are opposite in sign for OLS as opposed to LSDV and MGLS. Combinations of high p and \|j values decrease MSE (of) for OLS but increase MSE for the other two estimators. This means that the quality of the LSDV and MGLS estimators are damaged by the joint presence of the two analytic difficulties (autocorrelation and simultaneity) while that of OLS is actually improved. But, recall that LSDV and MGLS are still superior to OLS even at high levels of p and i|f . In other words, while the differences between OLS and either MGLS or LSDV decrease as p and \|t increase, the ranking of the estimators does not change.

Estimation of f^ involves a very different pattern. Increases in both p and \|i increase mean squared error for all three estimators. But the i|i effects are many magnitudes greater. Surprisingly, we find strong negative interaction effects for all three estimators. In fact, the slopes of the interaction effects are in each case similar in magnitude to those of ty . This is quite interesting because it suggests that the biasing effects of simultaneity can be considerably offset by the presence of autocorrelation. In other words, although each problem is damaging in isolation from the other, the combination of the two tends to be less damaging to inference than at least the pure simultaneity. Examination of Table 7 makes this plain.

Next, consider the instrumental variables (IV) estimator. It is consistent only when p = 0 . Tables 8 and 9 show that increases in p increase the MSE for both and ^ f°r IV estimator. It seems reasonable that IV is sensitive to increases in the problem for which it does not correct (i.e., increases in p). Conversely, IV is designed to deal with nonzero values of i|f , and the MSE of the IV estimator does not appear to depend systematically on either \|i or (p*ijt)*.

The overall performance of the IV estimator is in sharp contrast with the single-equation methods, as may be seen by a return to Table 5. Although OLS, LSDV, MGLS, and IV are all inconsistent when p ^ 0 and f ^ 0 , they differ considerably in quality. The IV estimator is the least efficient of the four methods.

Finally, we turn to the estimators which are consistent when both p 0 and ty ^ 0 , IV-LSDV and IV-MGLS. The most important result is the astounding contrast between IV-LSDV and IV-MGLS. While IV-MGLS has by far the lowest average MSE for A and A ^ across conditions, IV-LSDV has the highest MSE averages for both A (3^ and A ^ • Thus, over the combinations of parameter values studies, IV-LSDV has nothing to recommend it. At the same time IV-MGLS is clearly superior to all other methods studied in the condition p / 0 , M O .

Despite the radical differences in quality, IV-LSDV and IV-MGLS appear to have qualitatively similar performances in the face of changes in parameter values. Both IV-LSDV and IV-MGLS systematically underestimate ^ and over-estimate > as n^y be seen in Tables 6 and 7. In each case, according to Tables 8 and 9, p has a negative effect, ty a positive effect, and (p*i|f) an insignificant effect.

VII. CONCLUSIONS

The cross-lag panel model is of central importance to research which focuses upon dynamic causal processes. This model involves lagged X , reciprocal causation and lagged dependent variables. Most often, the estimation of panel roo^e1/; involves both time series and simultaneous equations problems. Social researchers attracted to the cross-lag model must learn to deal with the combination of the two problems.

We began with the premise that multi-wave data such as that used in the cross-lag panel model can often be best exploited when pooled into a single estimation. We limited our attention to pooling of cross-sections where there are individual-specific but not time period-specific components to error. This model of error components deals with stability in disturbances due only to time-invariant properties of sample units. It is likely to be a good approximation when the omitted unit-specific causes of Y change considerably more slowly than X and Y .

Whether pooling results in improved inference depends on at least three things. It depends on the stability of the causal processes over all waves of observations. It depends on one's ability to model the peculiar error characteristics of pooled data. And, finally, it depends on one's ability to design and modify statistical models to cope with the complicated error structures. Our research focused on the third issue.

The disturbance structure of a pooled model makes OLS inappropriate even when there is no reciprocal causation (i.e., no possibility of simultaneous equations bias). In this restricted case, both LSDV and MGLS are consistent and asymptotically efficient estimators. Choice between the two (and between either and the inconsistent OLS) depends on their small-sample properties. Our simulation yields results which agree with Nerlove's (1971) simulation even though here $N = 50$ and $T = 5$, while he used $N =» 25$ and $T = 10$. We find, in particular, that the performance of OLS is poor enough to cast doubt on the many published panel studies which do not correct for autocorrelation. Further, while both LSDV and MGLS surpass OLS, the modified generalized least squares procedure is superior. In fact, MGLS estimates, using biased estimates of p , are quite close to those of "true" GLS.

We also analyzed the intersection *of* the autocorrelation problem with simultaneity. Both IV-LSDV and IV-MGLS procedures are consistent but not asymptotically efficient estimators for the "full" cross-lag model. Since the behavior of the two "synthetic" estimators has apparently not been previously studied, we have no a priori information concerning their small sample behavior. We find the IV-MGLS procedure vastly superior to IV-LSDV over the whole range of parameter combinations. Further, the IV-MGLS procedure very plainly outperforms the inconsistent estimators studied.

Our results offer strong support for the view that application of the modified generalized least squares approach to panel estimation is preferable to the available alternatives. This appears to be true both when MGLS is used to correct for the "pure" pooling problems and when MGLS is used in combination with an instrumental variables estimator to deal with the combination of time series and simultaneous equations complications.

FOOTNOTES

good nontechnical discussion of the identification problem in both static and dynamic models is presented by Blalock (1969: Chapters 4-5).

[2] It appears that one pays a very heavy price for using "imprc; ^r" instruments. More precisely, instrumental variables methods seem to sutler more seriously from specification error than do the more straightforward ordinary least squares procedures. See, for example, the simulations reported by Blalock, Wells, and Carter (1970) and Hurd (1972).

[3] Whether or not a lagged dependent variable will serve as a useful instrument for unraveling patterns of reciprocal causation depends entirely on the autocorrelation of disturbances. When the disturbances are uncorrelated, the lagged values are perfectly appropriate instruments.

[4] The usual assertion is that at least thirty time periods must be observed for meaningful time series analysis.

arc assuming here that X is fixed over time. For other cases, see Coleman (1968).

Coleman's method involves substituting ordinary least squares estimates of (3) into (2). This procedure involves non-linear operations on the estimators. Unfortunately, least squares estimators do not retain optimal statistical properties under such transformations. Maximum likelihood estimation is clearly preferable.

^Recently a number of researchers (Duncan, 1972; Kenny, 1973; Hannan, Rubinson and Warren, 1974) have applied path analysis to the estimation of multi-wave panels containing unobservable variables. But even three-wave panels produce unwieldy algebraic structures. We fear that this work has done nothing to encourage researchers to employ multiple waves.

g

The pooling of cross-sections and time series has been discussed for some time in the econometrics literature. See, for example, Kuh (1959), Balestra and Nerlove (1966) , Wallace and Hussein (1969), Maddala (1971) and Nerlove (1971). We rely heavily on this literature in the analytic discussion that follows.

9

It is more usual to begin with a framework in which the are fixed. But, our point of reference is with the cross-lag model in which x^t is necessarily a stochastic variable. All of the same results hold at least for special cases when we allow the to be stochastic but we must evaluate not expectations but the limits of probability distributions (denoted plim).

^Note that also IV wastes a whole wave of observations in the first stage. This exacerbates the inefficiency of IV.

^Estimators which use constraints sequentially are termed "limited-information" estimators while those that use them jointly are called "full-information methods." The latter are always more efficient than the former when the constraints hold in the population.

[12] Other regression equations containing terms in p^2 , ty^2 and (p*i|0^2 were also examined. These equations did not change the analysis below and are therefore not reported here.

[13] Results for P= .95 were so similar to those for p^ ~ .9 that they are not shown in Tables 1, 2, 3, 6, 7.

I

REFERENCES

Amemiya, Takeshi

1967 "A Note on the Estimation of Balestra-Nerlove Models." Technical Report No. 4. Stanford University: Institute for Mathematical Studies in the Social Sciences.

Balestra, P. and Marc Nerlove

1966 "Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas." Econometrica 34 (July): 585-612.

Blalock, H. M., Jr.

1969 Theory Construction: From Verbal to Mathematical Formulations. New Jersey: Prentice-Hall, Inc.

Blalock, H. M., Jr., C. S. Wells, and L. F. Carter

1970 "Statistical Estimation with Random Measurement Error" in E. F. Borgatta and G. W. Bohrenstedt,(eds), Sociological Methodology 1970, San Francisco: Jossey-Bass.

Boudon, Raymond

1968 "A New Look at Correlation Analysis," pp. 199-235. In H. M. Blalock Jr. and A. Blalock (eds.) Methodology in Social Research, New York: McGraw-Hill.

Campbell, Donald T. and Julian C. Stanley

1963 Experimental and Quasi-Experimental Design for Research. Chicago: Rand McNally & Company.

Coleman, James S.

1968 "The Mathematical Study of Change" in H. M. Blalock Jr. (ed.) Methodology in Social Research.. New York: McGraw-Hill, pp. 428-478.

Duncan, Otis Dudley

1969 "Some Linear Models for Two-Variable Panel Analysis," Psychological Bulletin 72 (September): 177-182.

Duncan, Otis Dudley

1972 "Unmeasured Variables in Linear Models for Panel Analysis" in H. Costner (ed.), Sociological Methodology 1972. San Francisco: Jossey-Bass, pp. 36-82.

Goldberger, Arthur S.

1973a "Structural Equation Models: An Overview" in A. S. Goldberger and O. D. Duncan (eds.), Structural Equation Models in the Social Sciences. New York: Seminar Press.

Goldberger, Arthur S.

1973b "Rfticient Estimation in Overidentified Models: An Interpretive Analysis" in A. S. Goldberger and O. D. Duncan (eds.), Structural Equation Models in the Social Sciences. New York: Seminar Press.

Hannan, Michael T.
1971 Aggregation and Disaggregation in Sociology. Lexington, Mass.: Heath Lexington.

Hannan, Michael T., Richard Rubinson, and Jean T. Warren
1974 "The Causal Approach to Measurement Error in Panel Analysis: Some Further Contingencies" in H. M. Blalock (ed.), Measurement in the Social Sciences. Chicago: Aldine, in press.

Hannan, Michael T. and Alice A. Young
1974 "Estimation of Pooled Cross-Section and Time Series Models: Preliminary Monte Carlo Results." Paper presented at Conference on Policy Research: Methods and Implications, University of Wisconsin, Madison.

Heise, David R.
1969 "Separating Reliability and Stability in Test-Retest Correlations," American Sociological Review 34 (February):93-101.

Heise, David R.
1970 "Causal Inference from Panel Data" in E. Borgatta and G. Bohrenstedt (eds.), Sociological Methodology 1970. San Francisco: Jossey-Bass, pp. 3-27.

Henderson, C. R.
1971 "Comment on 'The Use of Error Components Models in Combining Cross-Section with Time Series Data.'" Econometrica 39 (March): 397-401.

Hurd, M. D.
1972 "Small-Sample Estimation of a Structural Equation with Autocorrelated Errors," Journal of the American Statistical Association 67 (September): 567-573.

Johnston, J.
1972 Econometric Methods. New York: McGraw-Hill.

Kenny, David A.
1973 "Cross-Lagged and Synchronous Common Factors in Panel Data" in A. S. Goldberger and 0. D. Duncan (eds.), Structural Equation Models in the Social Sciences. New York: Seminar Press.

Knuth, D. E.
1969 The Art of Computer Programming. Vol. 2, Cambridge, Mass.: Addison-Wesley.

Kuh, E.
1959 "The Validity of Cross-Sectionally Estimated Behavior Equations in Time Series Applications," Econometrica 27 (April): 197-214.

Lazarsfeld, P. F., Ann K. Pasanella, and Morris Rosenberg
1972 Continuities in the Language of Social Research, Section IV. New York: The Free Press.

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

Malinvaud, E.
1970 Statistical Methods of Econometrics. New York: American Elsevier Publishing Company, Inc.

Nerlove, Marc
1971 "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections," Econometrica 39 (March): 359-382.

Pelz, Donald C. and Frank M. Andrews
1964 "Causal Priorities in Panel Study Data," American Sociological Review 29 (December):836-848.

Searle, S. R.
1968 "Another Look at Henderson's Methods of Estimating Variance Components," Biometrics 24 (December): 749-778.

Searle, S. R.
1971 "Topics in Variance Component Estimation," Biometrics 27 (March): 1-76.

Theil, H.
1954 Linear Aggregation of Economic Relations. Amsterdam: North-Holland Publishing Company.

Wallace, T. D. and A. Hussain
1969 "The Use of Error Components Models in Combining Cross-Section with Time Series Data," Econometrica 37 (January):55-72.

Wiley, David E. and James A. Wiley
1970 "The Estimation of Measurement Error in Panel Data," American Sociological Review 35 (February):112-117.

Young, Alice A., Francois Nielsen and Michael T. Hannan
1974 "Experimental Study of the Pooling of Cross-Sections and Time Series," Laboratory for Social Research Technical Report (forthcoming).