Cross-Lagged Panel Correlation: A Test for Spuriousness

Article *in* Psychological Bulletin · October 1975

Impact Factor: 14.76 · DOI: 10.1037/0033-2909.82.6.887

David A. Kenny
University of Connecticut

Psychological Bulletin
1975, Vol. 82, No. 6, 887–903

Cross-Lagged Panel Correlation: A Test for Spuriousness

David A. Kenny
Harvard University

Cross-lagged panel correlation is a method for testing spuriousness by comparing cross-lagged correlations. True experiments control for spuriousness by random assignment, but random assignment limits true experimental studies to independent variables that can be manipulated. Like any statistical method, cross-lagged analysis is based on a set of assumptions: synchronicity and stationarity. Different forms of stationarity have different consequences for both the changes in the synchronous correlations over time and the difference between cross-lags. Homogeneous stability is a necessary assumption in the identification of both the source and direction of a causal effect. Cross-lagged analysis is a low-power test. It is better adapted than either multiple regression or factor analysis for many questions in panel studies. Multiple regression must assume no errors of measurement in the independent variables and no correlated errors, while factor analysis must specify a particular factor structure. Two extended examples of cross-lagged analysis are discussed with special emphasis placed on the issue of stationarity and the estimation of reliability ratios.

It has now been over 10 years since Campbell (1963; Campbell & Stanley, 1963) suggested the technique of cross-lagged panel correlation. Panel study is the survey term for a longitudinal design. Although cross-lagged analysis has been in use since at least 1901 (Hooker), we owe the first[1] formal presentation of the analysis to Campbell. The rudiments of cross-lagged panel correlation necessitate two constructs, X and Y, measured at two different points in time. As in Figure 1, the two constructs and two times generate four variables (X_1, X_2, Y_1, and Y_2), and the four variables generate six correlations: two autocorrelations ($\rho_{X_1X_2}$ and $\rho_{Y_1Y_2}$), two synchronous correlations ($\rho_{X_1Y_1}$ and $\rho_{X_2Y_2}$), and two cross-lagged correlations ($\rho_{X_1Y_2}$ and $\rho_{X_2Y_1}$).

As the name suggests, cross-lagged panel correlation is the comparison of the cross-lagged correlations which can be expressed as a *cross-lagged differential:* $\rho_{X_1Y_2} - \rho_{X_2Y_1}$. Campbell's original suggestion was that if X caused Y, then the cross-lagged differential would be positive, and if Y caused X, the differential would be negative. (Unless otherwise stated, all correlations here are assumed to be positive.) Campbell and his students (Kenny, 1973; Rickard, 1972; Rozelle & Campbell, 1969) have elaborated the method. Another tradition (Bohrnstedt, 1969; Duncan, 1969; Goldberger, 1971; Heise, 1970; Pelz & Andrews, 1964) has suggested replacing cross-lagged analysis with multiple regression or partial correlation analysis. Still another approach to panel data has been the application of factor analysis (Corballis & Traub, 1970; Jöreskog, 1969; Nesselroade, 1972). This approach usually involves oblique solutions in which factors are time specific. The fourth and oldest formal approach to panel data is the 16-fold table described by Lazarsfeld (1972, Note 1; Yee & Gage, 1968). This approach involves nominal data,

The work on this paper was supported in part by the Spencer and Milton Funds of Harvard University and by National Science Foundation Grant SOC-7103704-03, Donald T. Campbell, principal investigator.

I would like to thank William McGarvey, Louise Kidder, John Berman, Marylee Taylor, Michael Milburn, Robert Rosenthal, Robert Calsyn, Don Feree, and Joanne Martin, who kindly read an earlier version of this paper. Any errors are solely the responsibility of the author. Special thanks are also due to Lenor Kirkeby for assistance in the preparation of the manuscript.

Requests for reprints should be sent to David A. Kenny, Department of Psychology and Social Relations, Harvard University, Cambridge, Massachusetts 02138.

[1] Pelz and Andrews (1964) independently suggested this method.

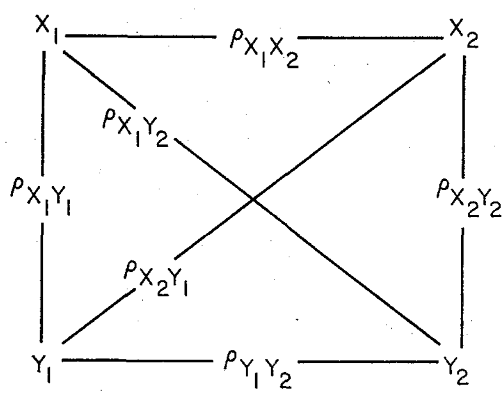

FIGURE 1. Cross-lagged panel correlation paradigm. (X and Y are variables and 1 and 2 are times.)

usually dichotomous, and is currently best represented by the work of Goodman (1973).

The literature on the use of panel data is growing and is contradictory. Indeed, the practitioner would be hard pressed to decide how to analyze panel data. Should one use cross-lagged panel correlation, multiple regression analysis, factor analysis, or the 16-fold table? This paper is a verbal exposition of the approach taken by cross-lagged analysis. It is shown that cross-lagged analysis is a test for *spuriousness,* by which is meant that the relationship between X and Y is not due to the causal effects of either but to the effects of a third variable Z. Alternative terms for spuriousness are third-variable effects, common factoredness, or "cosymptomatic" effects. Argumentation is restricted to a verbal level since the formal mathematical proofs have either already been presented in Kenny (1973) or could easily be derived from Table 1 of that same article.

Description of Cross-Lagged Panel Correlation

Its Logic

Cross-lagged analysis is a quasi-experimental design (Campbell & Stanley, 1963; Kenny, 1975). At the heart of quasi-experimental inference is the attempt to rule out plausible alternative explanations of a causal effect (i.e., biases or artifacts). In correlational analysis the chief alternative explanation of any causal effect is spuriousness. Almost any statistical relationship—be it simple correlation, partial correlation, or regression coefficient—can be attributed not to causality but to spuriousness. Suppes (1970) has even defined a causal relationship negatively as a nonspurious relationship. Ideally these spurious causes should be measured and controlled in the nonexperimental case.

True experiments control for spuriousness by random assignment to treatment conditions. Random assignment guarantees that there is no systematic relationship in the population between the treatment and the dependent variable given the null hypothesis of no-treatment effect. Thus, any relationship between the treatment and the dependent variable that cannot plausibly be explained by chance is attributed to the causal effects of the treatment.

Although random assignment permits us to make causal inferences, it brings with it some potentially burdensome methodological limitations. True experimentation rules out of consideration as independent variables any variable that cannot be manipulated and then randomly assigned. Many important variables, usually individual differences, are not subject to manipulation as simply as the intensity of light. Psychologists spend considerable time theorizing about intelligence, attitude change, extroversion–introversion, and evoked potential, but since these variables are attached to rather than assigned to the organism, they are studied more often as dependent rather than independent variables. To some degree the traditional stimulus–response or input–output orientation in psychology may reflect the limitation of experimental treatments to manipulatable variables. The requirement of manipulating the independent variables also prevents us from examining certain variables because of ethical considerations. For instance, malnutrition has been proposed as an important cause of children's cognitive ability, but it would be highly unethical to randomly assign children to levels of malnutrition. Thus, for practical and ethical reasons it is not always possible to use random assignment to control for spuriousness.

The null hypothesis of cross-lagged panel correlation tested by equality of the cross-lags is that the relationship between X and Y

is due to an unmeasured third variable and not causation. Before causal models are entertained, the third variable explanation should be ruled out. The logic of true experimentation is similar: Before accepting that the treatment has an effect, the null hypothesis of sampling error must be ruled out. The null model for cross-lagged analysis is illustrated in Figure 2. A third variable, Z_1, causes X_1 and Y_1 simultaneously. (Actually Z may cause X and Y with a lag, and the lag would be the same for both X and Y.) Over time, Z changes and at Time 2, Z_2 causes X_2 and Y_2. Given the inapplicability of true experimentation in numerous areas, this analysis can be used to test for spuriousness.

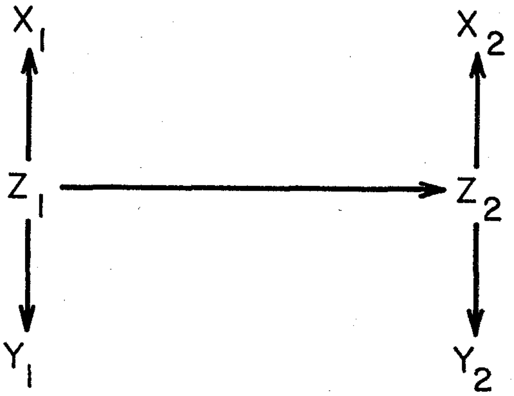

FIGURE 2. Cross-lagged panel correlation null hypothesis. (X, Y, and Z are variables and 1 and 2 are times.)

Its Assumptions

The hypothetical third variable effects can be ruled out by making two assumptions—*synchronicity* and *stationarity*. To test any model, assumptions about that model must first be made. All models rest on a set of assumptions; there are no assumption-free models. For instance, although in experimentation it is assumed that random assignment works, believers in unconscious ESP effects may be skeptical of the results from experimentation because assignment may reflect the experimenter's unconscious manipulation of dice or the random number table. But most of us are willing to *assume* that the procedure of random assignment works. True parametric experimental inference usually involves additional assumptions of normality and homogeneity of errors and additivity. Not surprisingly the assumptions of cross-lagged analysis are much more tentative and less robust than those for true experiments. But like random assignment, their justification lies in the *design* of research.

Synchronicity means that the two constructs X and Y are measured at the same point in time.[2] At first glance synchronicity would seem to be an easy assumption to satisfy. Panel studies are defined as the replication at two different points in time of a cross-sectional survey on the same set of persons. However, because of the problems of *retrospection* and *aggregation*, synchronicity may not be so easily satisfied. Some variables in panel studies ask the subjects to recall behaviors, attitudes, or experiences of the past. These questions either directly or indirectly ask the subjects to retrospect. In some sense the data may not be generated at the time of measurement but at some time prior to measurement.

For example, let variable X be a concurrent measure and let Y be retrospective, as in Figure 3. Construct X is measured at Times 1 and 2 and construct Y at Times .5 and 1.5. This would make the lag between X_1 and $Y_{1.5}$ only .5 units of time and the lag between $Y_{.5}$ and X_2 1.5 units. Given the common empirical finding that variables measured closer together in time correlate more highly than those measured further apart in time, $r_{X_1Y_{1.5}}$ should be greater than $r_{X_2Y_{.5}}$ since the lag for the first correlation is smaller than that for the second.[3] Note in Figure 3 that

[2] The assumption of synchronocity was first explicitly pointed out by Chaffee (Note 2).

[3] A cross-lagged analysis is still possible for the example in Figure 3 by comparing $r_{X_1Y_{1.5}}$ with either $r_{X_1Y_{.5}}$ or $r_{X_2Y_{1.5}}$ and making the additional assumptions about stability. Note that the traditional paradigm of cross-lagged panel correlation is a rectangle as in Figure 1, while in Figure 3 there are two isosceles triangles. Still another possibility is a symmetric trapezoid with the four measurements of X_1, Y_2, X_4, and Y_3. There are two cross-lagged comparisons: $r_{X_1Y_3}$ with $r_{X_4Y_2}$ and $r_{X_1Y_2}$ with $r_{X_4Y_3}$. Perhaps the first use of the isosceles triangle design was that by Calsyn (1973) and of the symmetric trapezoid was that by Seaver (1971). The additional stability assumption is that the causes of both X and Y are equally stable over equal periods of time.

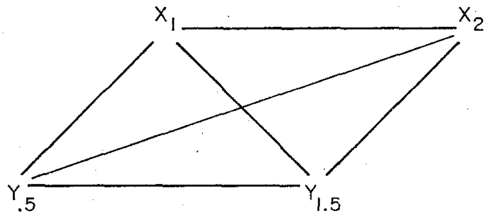

FIGURE 3. Lack of synchronicity. (X is measured at Times 1 and 2 and Y at Times .5 and 1.5.)

we have a parallelogram with diagonals of unequal length, not a rectangle as in Figure 1. Retrospective measures are not at all uncommon in panel studies.

Another problem for synchronicity is aggregation. Many variables are aggregated or averaged over time. A good example of a measure of this type is grade point average. If grade point average is to be taken as a measure of ability, at what point in time does it measure ability? It is actually an aggregation of performances evaluated by a teacher. The aggregation problem is well known in the econometrics literature, in which many of the important economic variables are aggregated across time (e.g., gross national product and unemployment rates).

Although the issue of synchronicity is critical to a cross-lagged analysis, simultaneity of measurement usually insures synchronicity. More central to cross-lagged analysis is the assumption of stationarity. *Stationarity* means that the causal or structural equation for a variable is not different at the two points of measurement. Some degree of stationarity is to be expected in panel studies. Since it is believed that the same variable is repeatedly measured in the panel study, one would suppose that the causes of this variable would not have changed. The issue of stationarity then becomes one of the strengths of those causes. *Stationarity* should be distinguished from *stability*. *Stability*, measured by autocorrelation, refers to a lack of change over time of the empirical values of a variable, while *stationarity* refers to a lack of change over time of the strength and direction of the causes of a variable. Actually only *relative* stability is implied, since all the values could change by a constant and then the autocorrelation would be unity.

Stationarity presumes that the causal processes did not change during the interval measured. This argument would not be reasonable if there were evidence that the subjects moved into a different stage over time, because a change in stage implies that the causal determinants, and therefore the causal *structure*, have changed over time. One would then expect stationarity to be less plausible during periods of rapid growth. The emphasis on stationarity presumes that the causal process is in equilibrium. Cross-lagged analysis is therefore inappropriate for the study of the onset of a causal effect.

There are three different types of stationarity: perfect stationarity, proportional stationarity, and quasi-stationarity.[4] These three different types of stationarity make different assumptions about the changes in causal structure over time and, therefore, have different consequences for the way in which the synchronous correlations change over time and for the difference between cross-lags.

Perfect stationarity means that there is no change in the causal equations of the variables over time. A necessary, though not sufficient, condition of perfect stationarity is that the synchronous correlations should not change over time.

Proportional stationarity means that the causal coefficients of each variable change over time by the same constant. This would be the case if all the variables' reliabilities increased or decreased by the same proportional constant. Given proportional stationarity, the ratio of the Time 2 synchronous correlation over the Time 1 synchronous correlation should be the same across all variable pairs. Given either perfect or proportional stationarity and spuriousness, the cross-lags should be equal.

Quasi-stationarity means that the causal coefficients of each variable change by a proportional constant, but each measured variable has its own unique constant. I call this *quasi-stationarity* since the true scores are perfectly stationary. *Quasi-stationarity* implies that the synchronous correlations would

[4] Proportional stationarity is what has been called a *between-variable proportionality constraint* and quasi-stationarity, a *within variable proportionality constraint* (Kenny, 1973).

be equal if they were corrected for attenuation due to measurement error. The differential reliabilities of the Time 1 and Time 2 measures can greatly bias the comparison of cross-lagged correlations. Consider the example in Table 1. The two variables selected from the Iowa Tests of Basic Skills (ITBS) (Lindquist & Hieronymus, 1964) are Graph and Table Reading and Knowledge and Use of Reference Materials. The data appear to be perfectly stationary since the synchronous correlations only differ because of sampling error. The cross-lag differential indicates that the use of graphs and tables causes the use of reference materials. An alternative explanation is that the Graphs and Tables test measure decreases in reliability and the Reference test measure increases. (This explanation is supported by additional analyses discussed in Crano et al., 1972). The unequal cross-lags would then reflect the fact that the correlation between two relatively unreliable variables (References test in Grade 4 and Graphs and Tables test in Grade 6) is compared to the correlation between two relatively reliable measures (References test in Grade 6 and Graphs and Tables test in Grade 4). The equal synchronous correlations reflect that a reliable and unreliable variable are correlated at each point in time. Campbell (1963) first pointed out that variables which increased their reliability would appear to be effects and variables that decreased in reliability would appear to be causes.

If quasi-stationarity can be assumed, reliability ratios[5] can be estimated and used to correct the cross-lagged correlations for changes in reliability over time (Kenny, 1973). The square root of the reliability ratio of each variable can be shown to be the variable's proportional constant of its causal equation. To estimate these reliability ratios there must be at least three variables, and if there are four or more variables the quasi-stationarity assumption can be tested by comparing the multiple reliability ratio estimates.

[5] Strictly speaking, what are called *reliability ratios* are communality ratios. I use the expression *reliability ratio,* since the concept of reliability is more familiar to most readers and is very close to what is meant by *communality.*

TABLE 1

UNSTATIONARITY EXAMPLE: CORRELATION OF GRAPH AND TABLE READING AND KNOWLEDGE AND USE OF REFERENCE MATERIALS AT GRADES 4 AND 6 FOR THE CORE CHILDREN OF THE MILWAUKEE SCHOOL SYSTEM

Variable	1	2	3	4
1. Graphs (4)	1.00			
2. References (4)	.44	1.00		
3. Graphs (6)	.30	.24	1.00	
4. References (6)	.46	.47	.45	1.00

Note. $N = 1,501$. Number in parenthesis refers to the grade. Data from Crano, Kenny, and Campbell (1972).

Thus, both a lack of synchronicity and a lack of stationarity are potential explanations of a difference between cross-lagged correlations. If the model in Figure 2 is correct, then both stationarity and synchronicity together would imply equal cross-lags. The null hypothesis that the cross-lagged differential is zero is then a test of spuriousness. What if the cross-lagged differential is not zero? Asymmetrical cross-lags may indicate a causal effect; more generally they indicate that there is a factor that causes one of the measured variables and causes the other measured variable at a later point in time.[6] This factor, called the *causal factor,* is one of many factors that make up the causal variable. The phrase "X causes Y" is shorthand for "something in X later causes Y." It need not be the case that the measure of X is valid or that the causal factor is the same as the true score of X. Though this problem of verbally explaining a causal effect is also present in true experiments, it is not as severe. One knows from an experiment that X causes Y, but the experiment does not necessarily tell what in X causes Y. We are all too familiar with how different theoretical perspectives focus on different aspects of an experimental treatment to explain the same causal effect. The problem of interpreting cross-lagged differences centers on the construct validity of measures just as it does in experimental research. The more valid, reliable, and unidimensional the measure, the more straightforward is the interpretation.

[6] Kenny (1973) has called this factor a *cross-lagged common factor.*

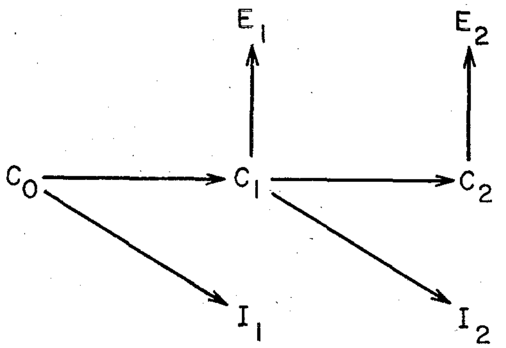

FIGURE 4. Model of the synchronous effect of current economic condition (C) on expectations (E) and its lagged effect on income (I).

To illustrate the difficulties in interpreting cross-lagged differences, consider the example in Table 2, taken from Taylor's (Note 3) analysis of data taken from Campbell, Converse, Miller, and Stokes (1960). In this example there is a significant difference between the cross-lagged correlations. Naively we could interpret the results as indicating expected financial change causing later income. We could begin to concoct a "self-fulfilling prophecy" explanation for this effect. An alternative explanation is that expected financial change is simply a forecast of income based on present economic barometers. The difference between the cross-lags may not reflect a causal effect of expectations on income but may simply be evidence of the validity of a person's forecast. To pictorially illustrate this explanation consider Figure 4. It is assumed that expected financial change is caused by current economic conditions and that income is a lagged effect of these conditions. The cross-lagged difference between expected financial change and income is due to the fact that they have the same unmeasured cause (current economic conditions) and one variable (expected financial change) is affected by this cause before the other (income). The causal factor is not expectations but the economic barometers.

Given causal effects, the synchronous correlations do not change over time if the causal functions are stationary and if the causal factor changes at the same rate over time during intervals of equal length. Both Yee and Gage (1968) and an early article by D. Campbell (1963) are incorrect to assume that causal effects necessarily imply a change in the synchronous correlations over time.

Rozelle and Campbell (1969) and Yee and Gage (1968) have pointed out another difficulty in interpreting cross-lagged differences—competing, confounded pairs of hypotheses. There are two *sources* of a causal effect (X and Y) and two *directions* of that effect (positive and negative), making a total of four possible hypotheses. Finding $\rho_{X_1Y_2} > \rho_{X_2Y_1}$ is consistent with both X causing an increase in Y and Y causing a decrease in X. Finding $\rho_{X_1Y_2} < \rho_{X_2Y_1}$ is consistent with both Y causing an increase in X and X causing a decrease in Y. To illustrate the difficulties of confounded hypotheses, consider the example taken from Kidder, Kidder, and Snyderman (Note 4) in Table 3. The variables are the number of burglaries and the number of police for 724 cities in the United States. At first glance it appears that burglaries cause an increase in the number of police. An alternative "law-and-order" explanation is that the number of police causes a *decrease* in the number of burglaries. Both hypotheses are equally plausible. The data are not consistent with two other hypotheses: Police increase burglaries or burglaries decrease the number of police. These latter hypotheses are not ruled out, but their effects, if they exist, are overwhelmed by the effects of one or both of the two former hypotheses.

Rozelle and Campbell (1969) suggested that a *no-cause baseline* could be computed

TABLE 2

AMERICAN VOTER STUDY: CORRELATION OF INCOME AND EXPECTED FINANCIAL CHANGE MEASURED IN 1956 AND 1960

Variable	1	2	3	4
1. Income (1956)	1.00			
2. Expected financial change (1956)	.11	1.00		
3. Income (1960)	.73	.15	1.00	
4. Expected financial change (1960)	.08	.19	.13	1.00

Note. N = 1,513. Number in parenthesis refers to the year. Example taken from Taylor's (Note 3) analysis of data taken from Campbell, Converse, Miller, and Stokes (1960).

to test both of the confounded hypotheses. Their procedure is as follows: (a) Compute the test–retest correlations of both variables and correct for attenuation. (b) Average these two correlations to obtain a measure of the stability. (c) Multiply the estimate by the average of the two synchronous correlations. (d) The resulting value is a no-cause baseline to which both cross-lags can be compared.

The logic of the no-cause baseline is that the cross-lags should be less than the synchronous correlations by some factor. That factor can be estimated from the autocorrelations. Unfortunately, there are two difficulties with the Rozelle and Campbell baseline. First, it requires that the researcher have estimates of each variable's reliability. Second, and more problematic, is the hidden assumption that all the nonerrorful causes of X and Y change at the same rate over time (i.e., have the same autocorrelation). I call this assumption ***homogeneous stability***. Evidence consistent with this assumption is that the two unattenuated autocorrelations are equal. Given the necessity of reliability estimates and homogeneous stability, I suspect that the Rozelle and Campbell baseline is of limited practical use for longitudinal studies.

Although the sign of the synchronous correlations is neither a necessary or sufficient condition for the direction of the causal effect, it is nonetheless suggestive of the direction. If the synchronous correlations are positive, they are supportive of X causing increases in Y and of Y causing increases in X. Negative synchronous correlations indicate decreases. Also, sometimes the researcher knows the source of causation and the only empirical issue is the direction, or the direction is known and the only empirical issue is the source. In this way the confounded hypotheses can be ruled out a priori.

Given homogenenous stability, the cross-lags should always be smaller in absolute value than the synchronous correlations given spuriousness, stationarity, and sychronicity. So a cross-lag larger than the synchronous correlations (assumed to be equal given stationarity) is indicative of a causal effect. As an example consider the correlations in Table 4. The cross-lags in both figures are identical but the synchronous correlations are different. For the A portion of Table 4 two hypotheses are plausible: X causes increases in Y, and Y causes decreases in X; but for the B portion only X causes increases in Y is plausible given homogeneous stability. If Y causes decreases in X and homogeneous stability is the case, then both the cross-lags would be smaller than the synchronous correlations. It should be made clear that if X causes increases in Y and homogeneous stability is the case, then the cross-lag from X to Y need not necessarily be larger than the synchronous correlations, since both instability of spurious causes and misspecified causal lag would tend to make the cross-lag smaller than the synchronous correlations.

TABLE 3

CONFOUNDED HYPOTHESES: CORRELATION OF NUMBER OF POLICE AND NUMBER OF BURGLARIES PER CAPITA MEASURED IN 1968 AND 1969

Variable	1	2	3	4
1. Police (1968)	1.00			
2. Burglaries (1968)	.47	1.00		
3. Police (1969)	.86	.43	1.00	
4. Burglaries (1969)	.35	.89	.39	1.00

Note. $N = 724$. Number in parenthesis refers to the year. Example from Kidder, Kidder, and Snyderman (Note 3).

Its Power

What does a nonsignificant difference between the cross-lagged correlations indicate?

TABLE 4

HYPOTHETICAL CORRELATIONS ILLUSTRATING THE INTERPRETATION OF A CROSS-LAG LARGER THAN THE SYNCHRONOUS CORRELATIONS

Variable	1	2	3	4
		A		
1. X_1	1.0			
2. Y_1	.6	1.0		
3. X_2	.7	.3	1.0	
4. Y_2	.5	.7	.6	1.0
		B		
1. X_1	1.0			
2. Y_1	.4	1.0		
3. X_2	.7	.3	1.0	
4. Y_2	.5	.7	.4	1.0

Strictly speaking one should not accept the null hypothesis of spuriousness, that is, the hypothesis that the variables do not cause each other but are cosymptoms of some set of common causes. There are some alternative explanations. First, it may be that both X and Y cause each other in a positive feedback loop. Without a no-cause baseline this model cannot be tested. Second, it may be that X causes Y or vice versa, but the magnitude of the effect is too small to be detected. In my experience it is very difficult to obtain statistically significant differences between cross-lagged correlations even when the sample size is moderate (75 to 300). The cross-lagged differential depends on the stability of the causal factor (cf. Formula 3 in Kenny, 1973). The more stable this factor is the smaller the differential. In the limiting case in which the factor does not change at all, the differential will be zero. Cross-lagged analysis is therefore not appropriate for examining the causal effect of variables that do not change over time. For these variables their effects might best be diagnosed using other quasi-experimental models (Kenny, 1975). (These models actually identify causal effects through unstationarity, that is, as increases in synchronous correlations.) Finally, large cross-lagged differences are difficult to obtain because the measured lag may not correspond to the causal lag. Normally the lag between measurements is chosen because of convenience not theory, since theory rarely specifies the exact length of the causal lag.

Given the low power of cross-lagged panel correlation, the researcher should design the longitudinal study to include many replications. Ideally, a cross-lagged difference should replicate across (a) different time lags, (b) different groups of subjects, and (c) different operationalizations of the same construct. For instance, most of the causal effects in Crano et al.'s (1972) study of intelligence and achievement can be summarized as abstract skills causing concrete skills. In one of the best empirical applications of cross-lagged analysis, Calsyn (1973) demonstrated all three of the above types of replications to show that academic achievement causes academic self-concept.

Its Alternatives

Although multiple regression and cross-lagged panel correlation are usually viewed as competitors for the analysis of panel data, they imply two alternative linear models that approach panel data inference in very different ways. The estimation of structural coefficients by multiple regression presumes a causal relationship between the measured variables and then estimates the parameters of the model. Cross-lagged analysis is a quasi-experimental method designed to test for spuriousness and presumes as a null hypothesis that the relationships between X and Y are spurious (i.e., due to unmeasured third variables).

There are two difficulties with the application of multiple regression to panel data: measurement error and unmeasured third variables. To estimate the effect of X_1 on Y_2, the variable Y_1 must be controlled for because Y_1 is almost certain to be correlated with both X_1 and Y_2. To be able to totally control for Y_1, perfect reliability must be assumed. For most psychological variables this is clearly an unreasonable assumption. Normally the bias in parameter estimates created by error of measurement is an *attenuation bias,* by which is meant that the estimated parameter underestimates the absolute value of the true parameter. Thus, with attenuation there is no bias if the parameter to be estimated is zero. But with partial regression and partial correlation coefficients the bias is not so benign. The parameter estimates may be larger or smaller than the true parameter or even of the opposite sign. For instance, given that (a) X_1 has no effect on Y_2, (b) Y_1 is unreliably measured, and (c) all nonzero causal coefficients are positive, the partial regression of Y_2 on X_1 "controlling" for Y_1 is positive even though its true value is zero. The bias arises because Y_1 is only partially controlled, since it is measured with error. Heise (1970) has shown by simulations that errors of measurement in panel data do not greatly bias regression coefficients. However, his simulations do not include the relatively common case of one variable reliably measured, say X_1, and the other variable unreliably measured, say Y_1. If the true regression coefficients of $b_{Y_2X_1 \cdot Y_1}$ and $b_{X_2Y_1 \cdot X_1}$ are zero, the estimated regression coefficients would show X causing Y and not Y causing X. An example of this differential bias is contained in an unpublished monograph by Underhill (Note 5). The question in Underhill's study is whether persons change their values to conform to their occupational choice or whether they change their occupations to conform to their values. The value measure is a composite of a three-item questionnaire, and the occupational measure is self-report. Most likely the value measure is of low reliability, while the occupation measure has a reliability very close to unity. The autocorrelations in Table 5 bear this out because they show low stability for the value measure and high stability for the occupational measure. Underhill analyzed the data by partial correlational analysis, which revealed that occupational choice caused values and not vice versa. The conclusion may well be an artifact of the differential reliability of the two measures. For example, the cross-lags in Table 5 indicate that values cause career choice, while the partials yield the opposite conclusion.

TABLE 5

Effect of Differential Reliability: Business Occupational Choice and Business Values

Variable	1	2	3	4
1. Business occupational choice (1)	1.00			
2. Business values (1)	.21	1.00		
3. Business occupational choice (2)	.74	.20	1.00	
4. Business values (2)	.15	.23	.16	1.00

Note. $N = 15{,}850$. Number in parenthesis refers to time. All values are actually probabilities and not correlations. Example taken from Underhill (Note 5).

Possible solutions to the unreliability problem are to assume no causal effects of one variable (Duncan, 1976) or to obtain parallel measures of both constructs (Wiley & Hornik, Note 6) or to add additional waves (Coleman, 1968; Pelz & Faith, Note 7). Sometimes the bias can be negligible enough to be ignored, if the true correlations of the controlled variable with other variables are low or if the reliability of the controlled variable is moderate to high.

A more pressing and almost insoluble problem for the use of multiple regression with panel data is the "third-variable problem" or spuriousness. Consider the logic of multiple regression. With multiple regression analysis, there are a set of predictors and a criterion. To interpret a regression analysis causally, the predictors are assumed to be causes of the criterion. The method of least squares imposes a restriction that errors of prediction in the regression equation are uncorrelated with any of the predictors or, as I shall refer to it, that there are *uncorrelated errors*. While having uncorrelated errors makes perfectly good sense to maximize prediction, it can be incorrect for a causal analysis. The assumption of uncorrelated errors implies that all unmeasured causes of the criterion are uncorrelated with any of the predictor variables. Therefore, all common causes of a predictor and criterion must be measured with perfect validity and reliability and be included in the regression equation. Since this is generally impossible, and usually untestable, the causal interpretation of any regression coefficient can almost always be alternatively explained by unmeasured common causes. Spuriousness can only be conclusively ruled out if the multiple correlation is unity or if units have randomly been assigned to groups. Obviously multiple correlations of unity and random assignment are most uncommon for panel studies.

It has been argued by Heise (1970) that by partialing Y_1 out of the relationship between X_1 and Y_2, spuriousness is controlled. However, controlling for Y_1 only partially partials out the effect of spurious causes since Y_1 is only an imperfect indicator of all the common causes of X_1 and Y_2. Even if Y_1 is measured without error, it still controls the spurious effects of Y_1, and only of Y_1.

The effects of errors of measurement and unmeasured third variables make inference from panel data by multiple regression problematic. Although multiple regression is a powerful method for nonexperimental inference, cross-lagged panel correlation is better adapted for panel data analysis, for which it was explicitly developed; it assumes measurement error and unmeasured third variables. The strength of cross-lagged analysis lies in its narrowness; it has been developed for and adapted to panel data analysis. The more common forms of statistical analysis like multiple regression, analysis of variance, and factor analysis though very general may not be easily adaptable to panel studies.

Although the model of cross-lagged panel correlation is a factor model in that it assumes that unobserved variables (factors) bring about the observed relationships, cross-lagged panel correlation does not use factor analysis in the estimation of factor loadings and factor correlations. The focus of cross-lagged analysis is model testing, not parameter estimation. Cross-lagged analysis tests a model of spurious effects that implies equal cross-lagged correlations given the assumptions of synchronicity and stationarity. The goal of the application of factor analysis to panel studies (cf. Corballis & Traub, 1970; Duncan, 1972; Jöreskog, 1969; Nesselroade, 1972) is to *estimate* factor loadings and factor correlations. The models I have discussed are indeterminate from a factor analysis point of view in that they allow for any number of factors. The orientation of cross-lagged panel correlation is to put constraints not on the number of factors, as in factor analysis, but to put constraints on the pattern of loadings over time. Cross-lagged analysis assumes invariant factor structure over time.

USE OF CROSS-LAGGED PANEL CORRELATION

Significance Tests

The hypotheses tested in a cross-lagged analysis are the equality of synchronous correlations to test for stationarity and the equality of cross-lags to test for spuriousness. One cannot use Fisher's z transformation (McNemar, 1969, pp. 157–158) to test for the significance of the differences between these correlations, since the correlations are themselves correlated. One can, however, use a rather bulky but easily programmable test cited by Peters and Van Voorhis (1940) and attributed to Pearson and Filon.[7]

Background Variables

Very often panel studies contain measures like sex, ethnicity, social class, and other background variables that are potential sources of spuriousness. There are two different strategies for handling these background or social grouping variables. The first is to perform separate analyses of each sex, race, or social group. The second is to subtract out the effects of the background variable. The first strategy is preferred if different causal patterns are expected for different social groups. For instance, Crano et al. (1972) found contrasting causal relationships for lower- and middle-class children in the relationship between intelligence and achievement. However, sample size often prohibits this strategy.

The second strategy—subtracting out the effects of background variables—can be done by computing partial correlations between the relevant variables controlling for the background variables. If the background variables are nominally coded, then dummy variables can be created for them. This procedure assumes that the causal processes are the same within social groups, though the groups may differ in mean level. I have found that controlling for background variables often increases the stationarity of the data. Background variables can be controlled for by examining the cross-lagged *partial* correlations.[8]

---

[7] Since the formula is not easily accessible, it is reproduced here. Let 1, 2, 3, and 4 be variables, N be sample size, and k equal

$$(r_{12} - r_{24}r_{14})(r_{34} - r_{24}r_{23}) + (r_{13} - r_{12}r_{23})(r_{24} - r_{12}r_{14}) + (r_{12} - r_{13}r_{23})(r_{34} - r_{13}r_{14}) + (r_{13} - r_{14}r_{34})(r_{24} - r_{34}r_{23}),$$

the following then has approximately a standard normal distribution:

$$z = \frac{\sqrt{N}(r_{14} - r_{23})}{\sqrt{(1 - r_{14}^2)^2 + (1 - r_{23}^2)^2 - k}}.$$

[8] There are two helpful rules in choosing variables to partial out. First, the variable to partial out should independently explain at least a moderate amount of variance. Otherwise nothing is changed by the partialing. Second, D. Campbell has suggested that any control variable should in principle be able to explain as much variance of the Time 1 variables as of the Time 2 variables. For instance, imagine a study of cognitive skills that had only a Time 2 measure of intelligence. Given instability, the intelligence measure will more highly correlate with the Time 2 measures than with the Time 1 measures, and therefore, it would be inappropriate as a variable to partial on.

Missing Data

A problem that plagues panel studies with long time lags is the attrition of subjects (Crider, Willits, & Bealer, 1973; Schaie, 1973). I suspect that sources of bias for means (like attrition and practice effects) do not similarly bias correlations, since means do not affect the size of a correlation. I would suggest that all correlations be based on the sample for which there is complete data, since comparing correlations based on different samples is problematic.

Examples

Let us consider two examples using cross-lagged analysis. The first is again taken from Crano et al. (1972). This example is chosen primarily to illustrate the importance of stationarity in the analysis. The second example is taken from Eron, Huesmann, Lefkowitz, and Walder (1972). This is perhaps the best-known use of cross-lagged analysis, and it has already been reproduced in at least two methods texts (Neale & Liebert, 1973; Rosenthal & Rosnow, 1975). The article by Eron et al. discusses the effect of viewing television violence on male adolescent aggression.

The three variables in the example from Crano et al. (1972) are all taken from the ITBS. They are Vocabulary, Knowledge and Use of Reference Materials, and Arithmetic Problem Solving. The sample consists of 1,501 inner-city Milwaukee children tested at fourth and sixth grades. The correlations among the variables are presented in Table 6.

Perhaps the first thing to do in examining cross-lagged panel correlation diagrams is to note the test–retest or autocorrelations. These correlations indicate both the stability and the reliability of the variables. Note that the autocorrelations in Table 6 are reasonable. Vocabulary has the higher autocorrelation, which is consistent with the fact that it had the most items. (Occasionally test–retest correlations can be too large because, as Carver [1974] has argued, the emphasis on stability as a measure of reliability makes it difficult to study change.)

TABLE 6

MILWAUKEE EXAMPLE: CORRELATION OF VOCABULARY, KNOWLEDGE AND USE OF REFERENCES, AND ARITHMETIC PROBLEM SOLVING FOR GRADES 4 AND 6

Variable	1	2	3	4	5	6
1. Vocabulary (4)	1.00					
2. References (4)	.46	1.00				
3. Arithmetic (4)	.51	.47	1.00			
4. Vocabulary (6)	.60	.40	.43	1.00		
5. References (6)	.52	.47	.49	.53	1.00	
6. Arithmetic (6)	.42	.35	.44	.44	.58	1.00

Note. N = 1,501. Number in parenthesis refers to the grade. Data taken from Crano, Kenny, and Campbell (1972).

All three synchronous correlations show a rather large statistically significant change[9] over time (the minus sign indicates a decrease): $z = 2.91$ for Vocabulary and References, $z = -2.71$ for Vocabulary and Arithmetic, and $z = 4.47$ for References and Arithmetic. As in Table 1 it might be that the cross-lagged differentials in the example are due to the changes in reliability of the measures. For instance, it appears that both Vocabulary and Arithmetic Problem Solving cause Knowledge and Use of Reference Materials, but these cross-lagged differentials may be due to the increasing reliability of the References test measure. The synchronous correlations are consistent with this increasing reliability interpretation, since the synchronous correlations of both the Vocabulary and Arithmetic tests with the References test increase over time.

A method was mentioned earlier that corrects the cross-lagged correlations for changes in reliability. The requisite numerical computations are as follows: First create a set of q_{ij} values where q_{ij} is defined as the Time 2 synchronous correlation of variables i with j divided by the Time 1 synchronous correlation of variables i with j. For example, the values of q are as follows:

$$q_{\mathrm{VR}} = .53/.46 = 1.16,$$
$$q_{\mathrm{VA}} = .44/.51 = .87,$$
$$q_{\mathrm{RA}} = .58/.47 = 1.23,$$

where V = Vocabulary, R = Knowledge and Use of Reference Materials, and A = Arithmetic Problem Solving. Now *directly* com-

[9] All computations were calculated to more significant digits than shown. They were also checked by a Fortran program, PANAL (Kenny, Note 8).

pute the reliability ratios as follows:

$$k_V^2 = \frac{q_{VR}q_{VA}}{q_{RA}} = .81,$$

$$k_R^2 = \frac{q_{VR}q_{RA}}{q_{VA}} = 1.65,$$

$$k_A^2 = \frac{q_{VA}q_{RA}}{q_{VR}} = .92.$$

A reliability ratio is defined as a variable's Time 2 reliability divided by its Time 1 reliability. Thus, a reliability ratio larger than 1 indicates an increase in reliability over time, a value less than 1 a decrease, and a value of 1 no change in reliability.

The hunch about an increase in the reliability of the R measure is confirmed by its reliability ratio of 1.65. Do the changes in reliability over time explain the cross-lagged differentials? With the reliability ratios the cross-lagged correlations can be corrected for changes in reliability in a way similar to correcting a correlation for attenuation. The correction for cross-lag $r_{i_1j_2}$ is simply $\sqrt{k_i/k_j}$, where i and j are variables and 1 and 2 are points in time; and for cross-lag $r_{i_2j_1}$, the correction is $\sqrt{k_j/k_i}$. Using this correction for the example, corrected cross-lags are:

$$\sqrt{\frac{k_V}{k_R}}\, r_{V_4R_6} = .44 \qquad .47 = r_{V_6R_4}\sqrt{\frac{k_R}{k_V}}$$

$$\sqrt{\frac{k_V}{k_A}}\, r_{V_4A_6} = .41 \qquad .44 = r_{V_6A_4}\sqrt{\frac{k_A}{k_V}}$$

$$\sqrt{\frac{k_R}{k_A}}\, r_{R_4A_6} = .40 \qquad .42 = r_{R_6A_4}\sqrt{\frac{k_A}{k_R}}.$$

For this example the cross-lag differentials substantially diminish after correcting for changes in reliability. As can be seen in Table 7, the z test of the difference between cross-lags is statistically significant before correction in two cases, while none are significant after correction. This example illustrates the necessity of considering stationarity as a plausible rival hypothesis of a cross-lag difference. Correcting for changes in reliability need not always decrease the cross-lag differential; it can sometimes reveal a hidden differential. Before moving on to the next example, topics related to the stationarity question are discussed.

Correcting for shifts in reliability by using estimated reliability ratios almost surely increases sampling error, and the Pearson–Filon test then becomes an approximate significance test. The correction should then only be employed on correlations taken from large samples (N must at least be equal to 100).

As stated earlier, the assumption that allows for estimation reliability ratios is called *quasi-stationarity,* which means that the synchronous correlations would be equal if corrected for attenuation due to measurement error. The estimated reliability ratios can be used to correct the synchronous correlations for differential measurement error and to test the viability of the quasi-stationarity assumption. The appropriate correction term for the Time 1 synchronous correlation between variables i and j is $\sqrt{k_ik_j}$, and for the Time 2 synchronous correlation it is $1/\sqrt{k_ik_j}$. Using this correction the following corrected synchronous correlations are obtained for the example:

$$r_{V_4R_4}\sqrt{k_Vk_R} = .49 = r_{V_6R_6}/\sqrt{k_Vk_R},$$

$$r_{V_4A_4}\sqrt{k_Vk_A} = .47 = r_{V_6A_6}/\sqrt{k_Vk_A},$$

$$r_{R_4A_4}\sqrt{k_Rk_A} = .52 = r_{R_6A_6}/\sqrt{k_Rk_A}.$$

Since there are three pairs of synchronous correlations and three reliability ratios, there

TABLE 7

TEST OF CROSS-LAG DIFFERENTIAL USING VOCABULARY, KNOWLEDGE AND USE OF REFERENCE MATERIALS, AND ARITHMETIC PROBLEM SOLVING

Variable pair	Uncorrected		Corrected	
	Differential	z value	Differential	z value
Vocabulary and references	.12	4.74	−.04	−1.62
Vocabulary and arithmetic	−.01	−.26	−.03	−1.28
References and arithmetic	−.14	−5.32	−.02	−.77

Note. N = 1,501.

are no "degrees of freedom" left, and the corrected synchronous correlations must be equal. However, as the number of variables increases there are multiple estimates of the reliability ratios. There are $(n-2)(n-1)/2$ estimates, where n is the number of variables. Given these multiple estimates there should be some way to pool the reliability ratios. The following approach is useful. Each estimate of the reliability ratio can be viewed as the product of three correlations divided by the product of three other correlations. Sum the absolute value of the numerators of each reliability ratio estimate and also sum the absolute value of the denominators of each estimate. The pooled estimate would be the ratio of the summed numerator and denominator. This pooled estimate has two desirable properties. First, its value does not depend on whether the reliability ratio is defined as the Time 2 reliability over the Time 1 or vice versa. A simple arithmetic pooling of the reliability ratios themselves would not have this property. Second, this solution weights more heavily reliability ratio estimates that are made up of larger, more reliable correlations. This pooled reliability ratio can be used to correct the synchronous correlation for changes in reliability. When the number of variables is greater than three, the equality of the corrected synchronous correlations can be used to gauge the viability of the quasi-stationarity assumption.

The method described above of estimating reliability ratios can be called a *direct method.* Crano et al. (1972) also suggested an indirect approach—first estimating each time's reliability or communality and then taking the ratio of those estimates. It might be argued that the ratio of squared multiple correlations is a better statistic than the direct solution advocated here. The squared multiple correlations are obtained by regressing a variable on all variables measured at the same point in time. The direct solution has the advantage of being unbiased in the population, while the ratio of squared multiple correlations is ordinarily not unbiased; however, it may be that squared multiple correlations are more efficient estimates than the direct solution. Table 8 contains the reliability ratio estimates for the Milwaukee study (total sample), using both squared multiple correlations and the direct solution. It is clear that solutions hardly differ in relative magnitude, but the squared multiple correlation estimates are always larger than the direct solution. The squared multiples are larger because they are biased by the higher average correlation at Grade 6 over Grade 4.

TABLE 8

Direct Versus Indirect Solution for Reliability Ratios

Variable	Reliability ratio: Direct	Reliability ratio: Squared multiple correlation ratio
Vocabulary	1.02	1.06
Reading	1.05	1.12
Spelling	1.01	1.05
Punctuation	1.00	1.06
Capitalization	1.25	1.25
Sentence comprehension	1.02	1.05
Map reading	1.03	1.06
Use of graphs	.76	.81
Use of references	1.33	1.35
Arithmetic problem solving	1.01	1.02
Arithmetic comprehension	1.03	1.05
Verbal IQ	.97	.99
Nonverbal IQ	1.11	1.17

Note. $N = 5{,}495$. Data from Crano, Kenny, and Campbell (1972).

In Table 9 there is the second example taken from Eron et al. (1972). The variables are television violence viewing level and aggression. The time points are Grade 3 and Grade 13 (a year after high school graduation). The sample consists of 211 boys from upper New York State. The television violence viewing measure was mother's report at Grade 3 and self-report at Grade 13, while the aggression measure was based on peer nomination procedures.

The cross-lags show a large, statistically significant difference ($z = 3.25$) which seems to indicate either that television violence viewing causes an increase in aggression or that aggression causes a decrease in television violence viewing. Since the second hypothesis is not very plausible, only the first hypothesis remains. But by now we should know that the first thing to do in interpreting cross-lagged panel correlation diagrams is not to

TABLE 9

TELEVISION VIOLENCE VIEWING AND AGGRESSION MEASURED AT GRADES 3 AND 13

Variable	1	2	3	4
1. Television violence viewing (3)	1.00			
2. Aggression (3)	.21	1.00		
3. Television violence viewing (13)	.05	.01	1.00	
4. Aggression (13)	.31	.38	−.05	1.00

Note. $N = 211$. Number in parenthesis refers to grade. Data from Eron, Huesmann, Lefkowitz, and Walder (1972).

look at the cross-lags but to examine the shell around them.

The autocorrelation for the aggression measure is respectable and consistent with the literature. However, the autocorrelation for television violence viewing is not significantly different from zero. The low correlation could be due to the fact that either its stability is low (i.e., television violence viewing is a state and not a trait) or the stationarity of the measure is low. Although the first explanation has considerable merit, there is strong empirical support for the second interpretation. Note that television violence viewing in Grade 13 does not correlate with any other measures in the panel, suggesting a high degree of unstationarity of the television violence measure. Clearly the data do not appear to be stationary. There is a large decrease in synchronous correlations over time ($z = 2.77$), and therefore the simple cross-lag comparison is not valid.

The data are potentially salvageable given the fact that the cross-lag for Grade 3 television violence viewing to Grade 13 aggression is larger than the synchronous correlation between Grade 3 television violence viewing to Grade 3 aggression. Although this difference is not statistically significant[10] ($t = 1.37$), it is supportive of a television-violence-causing-aggression hypothesis given two very critical assumptions.

First, assume that the aggression measure is stationary over time. If, for instance, the aggression measure increased in reliability over time, the cross-lag correlation could be larger than the synchronous correlation. This explanation is not plausible for reasons given in Eron et al. (1972) and Kenny (1972).

The second assumption is one mentioned earlier: homogeneous stability. Assume that all the causes of aggression change at the same rate over time. This assumption was overlooked by Kenny (1972) and Eron and his coauthors (1972). Becker (1972), with an ingenious though somewhat implausible argument, indirectly pointed out the necessity of this assumption. He argued that the data are consistent with the catharsis view of television violence and aggression. Assume that there is a set of unknown background factors that bring about a positive correlation between aggression and television violence viewing at Grade 3. These background factors do not change over time. Also, at Grade 3 there is an instantaneous effect of television violence causing a decrease in aggression, but this effect reduces only the synchronous correlation, since television violence viewing is unstable. Becker's explanation implies different rates of change for the different sources of correlation between television violence viewing and aggression: The background factors are stable while the cathartic effects of violence are very unstable.

Although the data are consistent with both a social learning view and a catharsis view, they do force each perspective to make additional specifications. The social learning view must now hypothesize a lagged effect on aggression, while the catharsis interpretation implies a very short lag.

Even if we accept the view that television violence viewing causes aggression, we must remember that we can only accept that something in Grade 3 television violence viewing later causes Grade 13 aggression. Chaffee (1972) suggested a possible name for this factor: a personality trait called *attraction to adult forms of aggression*. For a young child this variable causes the viewing of violent television shows, while in adolescence it causes overt violence.

If nothing else, it is now clear that the correlations in Table 9 do not call for a glib interpretation. They are consistent with tele-

[10] The significance test we have employed here is the test in which one array is shared (McNemar, 1969, p. 158).

vision violence viewing causing increases in aggression, but they are also consistent with other interpretations. There are some additional problematic aspects with the Eron et al. (1972) data: differences in the measures over time, the 10-year lag which intuitively is just too long, and the skewedness of the aggression measures that presents possible problems of outliers influencing the correlations. Nonetheless, Eron and his coauthors have given us a landmark study in its attempt to test opposing causal models.

Conclusion

Cross-lagged panel correlation is a valuable technique for ruling out the plausible rival hypothesis of spuriousness. It should not be viewed as only an intuitive approach but as a formal method with assumptions. This paper has inordinately emphasized alternative explanations of cross-lag differences in order to present the reader with a list of problems much in the same way that Campbell and Stanley (1963) did for quasi-experimental designs.

Cross-lagged panel correlation, however, is largely an exploratory strategy of data analysis. My suspicion is that its main use will be in uncovering simple causal relationships between uncontrolled variables. What would then follow is either the refinement of both the measures and the causal process in controlled settings or the estimation of causal parameters of the system by structural equation models (Duncan, 1975). The career of a hypothesized causal relationship might be as follows: first, the consistent replication of a cross-sectional relationship; second, the finding of time-lagged relationships between cause and effect; third, the finding of cross-lagged differences; and fourth, an experiment in which the causal variable is manipulated. Obviously, these steps will often overlap, some may be omitted, and the order may be different. I hope to emphasize that cross-lagged panel correlation plays only an intermediary role in social science, between the correlation and a well-elaborated structural model.

McGuire (1973) in social psychology and Wohlwill (1970) in developmental have pointed out that cross-lagged analysis could be a valuable addition to psychology's list of methods. It has been applied to many problems besides the ones previously mentioned. In industrial psychology Wanous (1974) has shown that performance causes intrinsic satisfaction, while extrinsic satisfaction causes performance. Lawler (1968) investigated the relationship between performance and expectancy. In developmental psychology Clarke–Stewart (1973) examined mother–child relationships. In educational psychology Dyer and Miller (1974), Schmidt and Crano (1974), and Crano (1973) have followed up the work of Crano et al. (1972). In social psychology Curry and Kenny (1974) have investigated the effects of actual and perceived similarity in values and personality on interpersonal attraction. Duvall and Welfling (1973) have used cross-lagged analysis in international relations. Cross-lagged panel correlation is proving itself to be a useful quasi-experimental method.

Many important topics have been neglected. Three or more wave models have not been discussed, nor have categorical variables in panel studies nor the application of cross-lagged analysis to time series data. Also, problems of correlated errors of measurement and nonlinearity have not been discussed. Finally, it has often been suggested to me that a cross-lagged analysis should focus on covariances and not on correlations. This suggestion needs further exploration.

Cross-lagged panel correlation need not be limited to longitudinal and panel studies in which the lag between measurements is usually at least months, if not years. A very common design in experimental work is the pretest–posttest design. Researchers usually collect many dependent variables, and the causal relations between the variables could be uncovered by cross-lagged analysis.

Finally, it is not often realized that cross-lagged panel correlation is a special case of the multitrait–multimethod matrix (Campbell & Fiske, 1959). There are two traits, X and Y, and two methods, Times 1 and 2. Much of the logic of Campbell's early articles on longitudinal analysis can be understood in this context. The concept of stationarity might be usefully applied to the analysis of the multitrait–multimethod matrix.

REFERENCE NOTES

1. Lazarsfeld, P. F. *The mutual effects of statistical variables.* New York: Bureau of Applied Social Research, 1947.
2. Chaffee, S. H. *Longitudinal designs for communications research: Cross-lagged correlation.* Paper presented at the meeting of the Association for Education in Journalism, Carbondale, Illinois, August 1972.
3. Taylor, M. *Potential applications of the technique of cross-lagged panel correlation analysis to the effects of low income.* Unpublished manuscript, Harvard University, 1974.
4. Kidder, L. H., Kidder, R. L., & Snyderman, P. *Do police cause crime? A cross-lagged panel analysis.* Duplicated Research Report, Temple University, 1974.
5. Underhill, R. *Occupational values and post-college career choice.* National Opinion Research Center, University of Chicago, 1967.
6. Wiley, D. E., & Hornik, R. *Measurement error and the analysis of panel data. Studies of Educative Processes* (No. 5), University of Chicago, 1973.
7. Pelz, D. C., & Faith, R. W. *Causal analysis of panel data: Progress report January 1971 to present.* Duplicated Research Report, University of Michigan, 1973.
8. Kenny, D. A. *PANAL: Panel data analysis of the first year of "Sesame Street."* Research Report, Harvard University, 1975.

REFERENCES

Becker, G. Causal analysis in R–R studies: Television violence and aggression. *American Psychologist,* 1972, *27,* 967–968.

Bohrnstedt, G. W. Observations on the measurement of change. In E. F. Borgatta (Ed.), *Sociological Methodology 1969.* San Francisco: Jossey-Bass, 1969.

Calsyn, R. *The causal relation between self-esteem, locus of control, and achievement: Cross-lagged panel analysis.* Unpublished doctoral dissertation, Northwestern University, 1973.

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. W. *The American voter.* New York: Wiley, 1960.

Campbell, D. T. From description to experimentation: Interpreting trends as quasi-experiments. In C. W. Harris (Ed.), *Problems in measuring change.* Madison: University of Wisconsin Press, 1963.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin,* 1959, *56,* 81–105.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research and teaching. In N. L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963.

Carver, R. P. Two dimensions of tests. *American Psychologist,* 1974, *29,* 512–518.

Chaffee, S. H. Television and adolescent aggressiveness. In G. A. Comstock & E. A. Rubinstein (Eds.), *Television and social behavior* (Vol. 3: *Television and adolescent aggressiveness*). Washington, D.C.: U.S. Government Printing Office, 1972.

Clarke–Stewart, K. A. Interactions between mothers and their children: Characteristics and consequences. *Monographs of the Society for Research in Child Development,* 1973, *38*(6–7, Serial No. 153).

Coleman, J. S. The mathematical study of change. In H. M. Blalock, Jr. & A. B. Blalock (Eds.), *Methodology in social research.* New York: McGraw-Hill, 1968.

Corballis, M. C., & Traub, R. E. Longitudinal factor analysis. *Psychometrika,* 1970, *35,* 79–98.

Crano, W. D. Causal analyses of the effects of socioeconomic status and initial intellectual endowment on patterns of cognitive development and academic ahievement. In D. R. Green (Ed.), *The aptitude–achievement distinction.* New York: McGraw-Hill, 1973.

Crano, W. D., Kenny, D. A., & Campbell, D. T. Does intelligence cause achievement?: A cross-lagged panel analysis. *Journal of Educational Psychology,* 1972, *63,* 258–275.

Crider, D. M., Willits, F. K., & Bealer, R. C. Panel studies, some practical problems. *Sociological Methods & Research,* 1973, *2,* 3–19.

Curry, T. J., & Kenny, D. A. The effects of perceived and actual similarity in values and personality in process of interpersonal attraction. *Quality and Quantity,* 1974, *8,* 29–44.

Duncan, O. D. Some linear models for two-wave, two-variable panel analysis. *Psychological Bulletin,* 1969, *70,* 177–182.

Duncan, O. D. Unmeasured variables in linear models for panel analysis: Some didactic examples. In H. L. Costner (Ed.), *Sociological Methodology 1972.* San Francisco: Jossey-Bass, 1972.

Duncan, O. D. *Structural equation models.* New York: Academic Press, 1975.

Duncan, O. D. Some linear models for two-wave, two-variable panel analysis with one-way causation and measurement error. In H. M. Blalock et al. (Eds.), *Quantitative sociology.* New York: Academic Press, 1976.

Duvall, R., & Welfing, M. Determinants of political institutionalization in black Africa. In J. A. Caporaso & L. L. Roos, Jr. (Eds.), *Quasi-experimental approaches.* Evanston, Ill.: Northwestern University Press, 1973.

Dyer, J. L., & Miller, L. B. Note on Crano, Kenny, and Campbell's "Does intelligence cause achievement?" *Journal of Educational Psychology,* 1974, *66,* 49–51.

Eron, L. D., Huesmann, L. R., Lefkowitz, M. M., & Walder, L. O. Does television violence cause aggression? *American Psychologist,* 1972, *27,* 253–263.

Goldberger, A. S. Econometrics and psychometrics: A survey of communalities. *Psychometrika,* 1971, *36,* 83–107.

Goodman, L. A. Causal analysis of data from panel studies and other kinds of surveys. *American Journal of Sociology,* 1973, *78,* 1135–1191.

Heise, D. Causal inference from panel data. In D. F. Borgatta & G. Bohrnstedt (Eds.), *Sociological Methodology 1970.* San Francisco: Jossey-Bass, 1970.

Hooker, R. W. Correlation of marriage-rate with trade. *Journal of the Royal Statistical Society,* 1901, *64,* 485–492.

Jöreskog, K. G. Factoring the multitest-multi-occasion correlation matrix. *Educational Testing Service Research Bulletin,* RB-69-62, 1969.

Kenny, D. A. Threats to the internal validity of a cross-lagged panel inference as related to "Television violence and child aggression: A follow-up study." In G. A. Comstock & E. A. Rubinstein (Eds.), *Television and social behavior* (Vol. 3: *Television and adolescent aggressiveness*). Washington, D.C.: U.S. Government Printing Office, 1972.

Kenny, D. A. Cross-lagged and synchronous common factors in panel data. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences.* New York: Seminar Press, 1973.

Kenny, D. A. A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design. *Psychological Bulletin,* 1975, *82,* 345–362.

Lawler, E. E. A correlational causal analysis of the relationship between expectancy attitudes and job performance. *Journal of Applied Psychology,* 1968, *52,* 462–468.

Lazarsfeld, P. F. Mutual relations over time of two attributes: A review and integration of various approaches. In M. Hammer, K. Salzinger, & S. Sutton (Eds.), *Psychopathology.* New York: Wiley, 1972.

Lindquist, E. F., & Hieronymus, A. N. *Iowa tests of basic skills: Manual for administrators, supervisors, and counselors.* Boston: Houghton Mifflin, 1964.

McGuire, W. J. The yin and yang of progress in social psychology: Seven koan. *Journal of Personality and Social Psychology,* 1973, *26,* 446–456.

McNemar, Q. *Psychological statistics* (4th ed.). New York: Wiley, 1969.

Neale, J. M., & Liebert, R. M. *Science & behavior: An introduction to methods of research.* Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Nesselroade, J. R. A note on the "longitudinal factor analysis" model. *Psychometrika,* 1972, *37,* 187–191.

Pelz, D. C., & Andrews, F. Detecting causal priorities in panel study data. *American Sociological Review,* 1964, *29,* 836–848.

Peters, C. C., & Van Voorhis, W. R. *Statistical procedures and the mathematical bases.* New York: McGraw-Hill, 1940.

Rickard, S. The assumption of causal analyses for incomplete causal sets of two multilevel variables. *Multivariate Behavioral Research,* 1972, *7,* 317–359.

Rosenthal, R., & Rosnow, R. L. *Primer of methods for the behavioral sciences.* New York: Wiley, 1975.

Rozelle, R. M., & Campbell, D. T. More plausible rival hypotheses in the cross-lagged panel correlation technique. *Psychological Bulletin,* 1969, *71,* 74–80.

Schaie, K. W. Methodological problems in descriptive developmental research on adulthood and aging. In J. R. Nesselroade & H. W. Reese (Eds.), *Life-span developmental psychology.* New York: Academic Press, 1973.

Schmidt, F. L., & Crano, W. D. A test of the theory of fluid and crystallized intelligence in middle- and low-socioeconomic-status children. *Journal of Educational Psychology,* 1974, *66,* 255–261.

Seaver, W. B. *Effects of naturally induced teacher expectancies on the academic performance of pupils in the primary grades.* Unpublished doctoral dissertation, Northwestern University, 1971.

Suppes, P. *A probabilistic theory of causality.* Amsterdam: North-Holland Publishing, 1970.

Wanous, J. P. A causal–correlational analysis of the job satisfaction and performance relationship. *Journal of Applied Psychology,* 1974, *59,* 139–144.

Wohlwill, J. R. The age variable in psychological research. *Psychological Review,* 1970, *77,* 49–64.

Yee, A. H., & Gage, N. L. Techniques for estimating the source and direction of causal inference in panel data. *Psychological Bulletin,* 1968, *70,* 114–126.

(Received September 19, 1974)