Cross-Lagged Panel Analysis

Michael W. Kearney*

Cross-lagged panel analysis is an analytical strategy used to describe reciprocal relationships, or directional influences, between variables over time. Cross-lagged panel models (CLPM), also referred to as cross-lagged path models and cross-lagged regression models, are estimated using panel data, or longitudinal data where each observation or person is recorded at multiple points in time. The models are considered "crossed" because they estimate relationships from one variable to another and vice-versa. They are considered "lagged" because they estimate relationships between variables across different time points. Taken together, cross-lagged panel models estimate the directional influence variables have on each other over time.

The primary goal of cross-lagged panel models is to examine the causal influences between variables. In essence, cross-lagged panel analysis compares the relationship between variable X at Time 1 and variable Y at Time 2 with the relationship between variable Y at Time 1 and X at Time 2. It is widely used to examine the stability and relationships between variables over time to better understand how variables influence each other over time.

This entry discusses cross-lagged panel analysis, an analytical strategy used in longitudinal communication research. It describes its rationales and origins in research. It also describes modern path-analytic approaches to cross-lagged panel analysis. Finally, this entry discusses some important assumptions and issues with cross-lagged panel analysis.

Directional Influences

Basic methods for testing causality have several limitations. Correlational analysis relies on theoretical inferences to make arguments about causality.

*Ph.D. Candidate, Communication Studies, University of Kansas.

Kearney, M. W. (in press). Cross Lagged Panel Analysis. In M. R. Allen (Ed.), *The SAGE Encyclopedia of Communication Research Methods.* Thousand Oaks, CA: Sage

Since cross-sectional data represents only one moment in time, there is no way to determine if these inferences are correct. The experimental method utilizes randomization and control to provide a more robust method for examining causality. In many cases, however, randomization and control are not practical or even possible. For example, costs associated with recruiting truly random samples for multiple time points are often too expensive. And resources are not the only barriers to randomization. In many cases, randomization creates ethical dilemmas that make studies examining certain variables like aging or illness problematic. In these situations, researchers often turn to longitudinal research and cross-lagged panel analysis.

Cross-Lagged Correlations

Cross-lagged panel analysis is used to compare the relationship between variable X at Time 1 (X_1) and variable Y at Time 2 (Y_2) with the relationship between Y_1 and X_2. In the past, this was accomplished by examining zero-order correlations. Cross-lagged correlations (CLC) were used to make arguments about causal directions between variables. Correlations of the same size indicated a reciprocal relationship. If one of the coefficients was larger, however, it suggested that changes in one variable lead to changes in the other variable and not the other way around. Comparing CLC thus provides some evidence of directional influence, but it has serious flaws.

Several weaknesses have been identified in the cross-lagged correlations method. One weakness is that CLC do not account for contemporaneous relationships between variables. Contemporaneous relationships refer to the correlations between variables within the same time point. Another weakness is that CLC do not account for the stability of each construct across time points. Stability refers to the degree to which values of a variable are unchanging over time. As a result of these shortcomings, the CLC method has largely been discarded in favor of cross-lagged path (or regression) models.

Cross-Lagged Panel Models

Like the CLC method, cross-lagged path models compare cross-lagged relationships. In addition to allowing for the estimation of cross-lagged effects, cross-lagged path models also control for correlations within time-points and autoregressive effects, or stability, across time. Autoregressive effects describe the amount of stability in constructs over time. Smaller autoregressive coefficients (closer to zero) indicate more variance in the construct, meaning less stability or influence from the previous time point. Larger autoregressive coefficients indicate little variance over time, meaning more stability or influence from the previous time point.

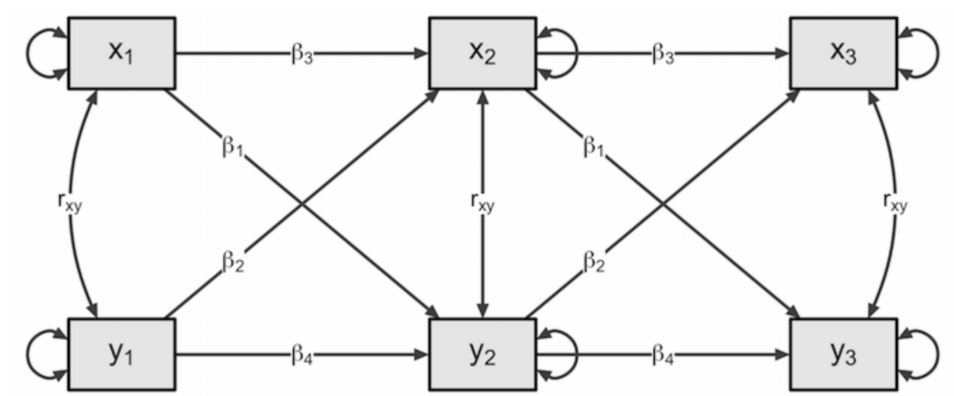
The most basic cross-lagged panel model includes two constructs measured at two time points. Cross-lagged panel models assume that each time a construct is measured is a variable. The simplest model therefore consists of two X variables (x_1, x_2) and two Y variables (y_1, y_2). The model also includes ten parameters, making it just identified. These parameters include exogenous variances ψ_{x_1}, ψ_{y_1}, [synchronous] correlations $r_{x_1 y_1}$, $r_{x_2 y_2}$, cross-lagged paths β_1, β_2, auto-regressive paths β_3, β_4, and endogenous residuals ζ_{x_2}, ζ_{y_2}. Estimates for cross-lagged effects now control for contemporaneous effects and variance across time (stability). Causal predominance can be examined by comparing standardized coefficients of the cross-lagged paths. This basic model can be easily extended for studies with three or more waves of measurement as well.

Cross-lagged panel models can include more than two waves of measurement. An example of a three-wave cross-lagged panel model is presented in Figure 1. In the depicted model, paths are constrained to equality across time. Thus, β_1 represents the cross-lagged effects from X_1 on Y_2 and from X_2 on $Y3$. The model also assumes that stability is a function of only the previous time point, meaning a log of one. These assumptions can change depending on the study, but theory should drive these decisions whenever possible. Unlike the two-wave models, models with more than three waves of measurement are typically over-identified, which allows for models with different error structures (autoregressive effects) and assumptions about variance over time to be compared.

Issues and Assumptions in Cross-Lagged Panel Analysis

Cross-lagged panel analysis is a useful tool for describing lagged relationships between two or more variables, though it is sometimes explicitly used as evidence of causality. In its most basic form, cross-lagged panel analysis attempts to identify causal predominance, which occurs when one variable influences another variable without also experiencing a reciprocal influence in return. Causal predominance is indicated when the effect of X at time 1 on Y at time 2 is large, while the effect of Y at time 1 on X at time 2 is zero. In such cases, X is considered the source variable and Y is considered the effect variable. The remainder of this entry covers common issues and

Figure 1



important assumptions that explain why many scholars recommend only using cross-lagged panel analysis for exploratory research.

Synchronicity. Cross-lagged panel analysis makes several important assumptions. The first is the assumption of synchronicity, which assumes that measurements at each time point occurred at the exact same times. Although most studies are designed to measure variables simultaneous, complications during data collection frequently violate this assumption.

Stationarity. Another assumption of cross-lagged panel analysis is that variables and relationships stay the same across time. This assumption, referred to as stationarity, relates to the stability of a construct as well as the nature of the relationships between constructs over time. As was previously discussed in models with three or more time points, there are varying degrees of stationarity, though very few theories offer guidance in this area.

Comparing Cross-Lagged Coefficients. In order to make claims about causal predominance, cross-lagged path analysis typically includes comparing relative sizes of cross-lagged coefficients. This is accomplished by standardizing variables. Although each of the standardized variables can be described in similar terms, standardization does not necessarily address fundamental differences distributions. In some cases it may not be appropriate to assume the variables were measured on the same scale.

Measurement Error. Although cross-lagged panel analysis can also be done using structural equation modeling, many cross-lagged panel models assume that variables are measured without error. Of course, for many variables in communication research, this is clearly not the case, which leads to biased results. Some scholars have also argued that measurement error may also be misidentified as real change when models have only two time points. In these cases, measurement error could still confound results of structural equation models.

Timeframe of Effect. Cross-lagged panel models also assume X_1 occurs before X_1, but it does not explicitly include time. Instead, it assumes the influence of one variable on another is a function of lag, or time between waves of measurement. The amount of lag can be any length of time, though it must be contextually appropriate to have meaningful interpretations. If the lag is too short, measurement will occur before the effects can be observed. If the lag is too long, the effects will dissipate before the next time of measurement.

Omitted Variables. Cross-lagged panel analysis, in theory, assumes all possible variables were measured and included in the model. This definition of causality, which originated from econometrics, is unlikely to hold communication research. Given the uncertainty surrounding many communication variables, scholars phrase interpretations of cross-lagged panel analysis in terms of directions of "influence" rather than "causality."

Stability. Cross-lagged panel models generally lack explicit theories of change. As such, autoregressive parameters are included to account for stability for everyone across time. This assumes there are no inter-individual differences, or differences between people, over time in stability. Inter-individual differences that do exist, such as unobserved trait-like influences or dependencies, may bias results.

References

[1] Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological methods, 20*(1), 102.

[2] Kenny, D. A. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin, 82*(6), 887.

[3] Rogosa, D. (1980). A critique of cross-lagged correlation. *Psychological Bulletin, 88*(2), 245.

[4] Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In Laursen, B., & Little, T. D. (Ed.), *Handbook of developmental research methods* (pp. 265-278). New York: Guilford Press.

[5] Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction.* New York: Taylor & Francis. pp 122-151.