CAN VALID CAUSAL INFERENCES BE MADE FROM PANEL DATA?

Kim Onn Yap

Northwest Regional Educational Laboratory

Portland, Oregon

Peter Dunn-Rankin

University of Hawaii

Honolulu, Hawaii

Introduction

During the past two decades, several methods have been developed for making causal inferences from quasi-experimental data (Campbell, 1963; Campbell & Stanley, 1963; Yee, 1966; Blalock, 1969). Some of these methods are direct "borrowings" from statistical procedures (e.g., regression analysis, part-correlation analysis) originally designed to determine functional relations among variables. Others (e.g., cross-lagged correlation analysis, frequencies-of-shift-across-median method and frequencies-of-change-in-product-moment method) have been developed for the express purpose of disentangling causal relationships. Regardless of their origins, the use of these techniques is generally problematic. Limitations and pitfalls have been pointed out by writers such as Tukey (1954), Richards (1966), Werts & Watley (1968), Linn & Werts (1969), Rozelle & Campbell (1969) and Crano et al. (1972).

A critical problem related to the use of causal methods centers on the validity of causal interpretations produced by these methods. The validity problem is compounded by a number of conditions peculiar to the social sciences. First, the interrelationships among variables in social science settings typically entail a high degree of complexity.

For instance, the causal flow may be reciprocal rather than uni-directional; the causal influence may be congruent, incongruent or both (Yee, 1966). Secondly, researchers in the social sciences typically pay little attention to the specification of assumptions inherent in methods of data reduction and interpretation. This neglect is particularly evident in causal analysis. Thirdly, empirical proof of causal relations can be obtained only through experimental manipulation. Yet, it is in the social sciences that such manipulation is typically not feasible. Finally, it is apparent that much of the controversy over the use of causal methods is due to the lack of a clear conceptualization of causal relationships. The discussion of necessary and sufficient conditions has contributed little to such conceptualization.

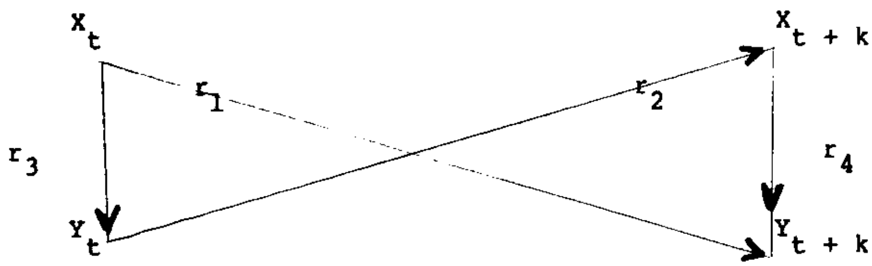
The present study was designed to address the question of whether valid causal inferences can be made from panel data gathered at two time points. A simulative evaluation was conducted to ascertain the validity of five causal methods as they are applied to panel data to yield causal inferences.

The Causal Methods

The five causal methods that were evaluated in the study included: cross-lagged correlation method (CLC), part correlation method (PC), econometric method (EC), frequencies-of-shift-across-median method (FSM) and frequencies-of-change-in-product-moment method (FCP). Brief descriptions of these methods follow:

1. <u>Cross-lagged correlation (CLC)</u>

The method was first proposed by Campbell and Stanley (1963) and is based on the rule of time precedence. When a given event consistently precedes the occurrence of another and the reverse does not hold, two hypotheses become plausible: (a) event 1 is a cause of event 2; or (b) both events are the effects of some more general cause. The method makes causal inferences on the basis of the relative magnitudes of the cross-lagged correlations. This may be illustrated by the following diagram:



If X and Y are the cause and effect variables, respectively, r_1 will be greater than r_2, k being an interval which is close to the causal interval needed for a change in X to effect a change in Y. In addition, cross-lagged correlation r_1 should also be greater than either simultaneous correlation r_3 or r_4 if k is close to the causal interval.

2. Part correlation (PC)

Part correlation has been used by researchers (e.g., Linn & Werts, 1969) as a means of removing spurious association between variables. In a panel situation where two variables, say X and Y, are measured at two time points, four part correlations can be computed from the data: (1) $R_{x1(y2.y1)}$ or the correlation between X_1 and Y_2 with the Y_2 variance due to Y_1 partialled out; (2) $R_{x2(y2.y1)}$ or the correlation between X_2 and Y_2 with the Y_2 variance due to Y_1 partialled out; (3) $R_{y1(x2.x1)}$ or the correlation between Y_1 and X_2 with the X_2 variance due to X_1 partialled out; and (4) $R_{y2(x2.x1)}$ or the correlation between Y_2 and X_2 with the X_2 variance due to X_1 partialled out. Based on the relative magnitudes of the part correlations, causal inferences may be made as follows: (a) If $R_{x1(y2.y1)}$ is greater than $R_{y1(x2.x1)}$, we may infer that X is the source of influence; (b) Conversely, if $R_{y1(x2.x1)}$ is greater than $R_{x1(y2.y1)}$ then we may infer that Y is the source of influence. In either case, the sign of the greater part correlation indicates whether the causal influence is congruent (positive) or incongruent (negative).

3. Econometric (EC).

The econometric methods (Blalock, 1969; Wonnacott & Wonnacott, 1970) were developed from regression analysis and are sometimes simply called the regression methods (Darlington, 1968). While they do not explicitly employ causal terminology, these methods do infer causal relationships between variables, a notable example being the causal influence between supply and demand. Typically the regression coefficients are taken as indices of the strength of causal links between exogenous (independent) and endogenous (dependent) variables. In a two-variables-measured-twice situation, the following equations may be formulated to facilitate causal interpretation of the data:

$$X_2 = B_1X_1 + B_2Y_1 + e_1,$$

$$Y_2 = B_3Y_1 + B_4X_1 + e_2.$$

In the equations, X_1, X_2, Y_1, and Y_2 are measures of variables X and Y taken at two time points (e.g., pretest and posttest). Causal inferences are made on the basis of the relative magnitudes of the beta weights of the exogenous variables--B_2 and B_4 in the given example. If B_2 is statistically significant and B_4 is not, then one could infer that variable Y is the source of causal influence. The sign of B_2 indicates whether the influence is congruent (positive) or incongruent (negative). If, on the other hand, B_4 is statistically significant and B_2 is not, then variable X is inferred to be the source of causal influence. Again, the sign of B_4 suggests the congruity or incongruity of the causal influence. No causal inferences are indicated if both beta weights are significant or non-significant.

4. <u>Frequencies-of-shift-across-median (FSM)</u>.

The method was developed by Yee (1966). It trichotomizes variables measured at two time points by their respective medians. Measures of the variables are said to be (a) above the median or high, (b) on the median, or (c) below the median or low. On the basis of shifts of time-one (pretest) and time-two (posttest) measures across the medians, a variety of response patterns can be identified. The measures can remain without change relative to pretest and posttest medians (high to high, low to low, or median to median). They can shift from pretest medians across posttest medians (median to high or median to low). The measures can also shift to and across posttest medians from positions above or below pretest medians (high to low, low to high, high to median, or low to median). Causal inferences are made on the basis of which variable shifted more relative to the medians of their pretest and posttest measures. For instance, if variable X shifted across its medians more often than variable Y did, then the latter is inferred to be the source of causal influence. In cases where the measures remain unchanged relative to the pretest and posttest medians, causal influence is considered to be uncertain. The direction of causal influence is determined on the basis of the complementarity of the measures (i.e., whether the measures are positively or negatively correlated). If the measures are positively related, the influence is considered to be toward congruity. The influence is considered to be toward incongruity if the reverse is true.

It is noted that in this method, both the source and direction of causal influence are in fact determined for each of the individual cases included in a sample. An overall statement of the source and direction of causal influence is made on the basis of the relative number of cases where variable X or variable Y is inferred to be the source of causal influence. The chi-square statistic is used to determine which variable constitutes the predominant causal factor.

5. Frequencies-of-change-in-product-moment (FCP).

The method, also developed by Yee (1966), converts raw scores of each variable into z scores for each time point. The direction of causal influence (i.e., congruency or incongruency) is determined by the relative magnitudes of the cross-products of pretest z scores and posttest z scores. If the cross-product of the posttest z scores is greater than the cross-product of the pretest z scores, the direction of causal influence is congruent. The direction of influence is incongruent if the reverse is true. The source of influence is determined on the basis of the relative magnitudes of cross-lagged z products. When the direction of influence is congruent, the variable whose pretest measure is contained in the more positive product is the source of influence. When the direction of influence is incongruent, the variable whose pretest measure is contained in the more negative product is the source of the causal influence.

Similar to FSM, this method determines the source and direction of causal influence for each of the individual cases included in a sample. An overall statement on causal influence is made on the basis of the relative number of cases where variable X or variable Y is inferred to be the source of causal influence. Again, the chi-square statistic is used to determine which variable constitutes the predominant causal factor.

It should be noted that the foregoing descriptions of the five causal methods are in many ways oversimplified. The reader is referred to the various sources cited in the references for a full explication of each of the methods.

The Simulation

The simulation was based on a numerical conceptualization of causal relations in a two-variables-measured-twice situation. The conceptualization was basically an extention of McNemar's (1969) interpretation of the correlation coefficient. In examining the relationship between two variables McNemar suggests that each of the variables may be thought of as a summation of a number of equally potent, equally likely independent elements which can be either present or absent. The magnitude of the correlation is then a function of the number of elements common to both variables.

To this conceptualization of correlation an essential element of causal relations--time precedence--was added, giving rise to the following formulation as a model for generating data sets where causal relations between variables X and Y are implied:

$$X_1 = a + e_1, \quad (1)$$

$$X_2 = a + b + e_2, \quad (2)$$

$$Y_1 = b + e_3, \quad (3)$$

$$Y_2 = b + e_4. \quad (4)$$

In this formulation, X_1 and Y_1 are the time-one measures of variables X and Y; X_2 and Y_2 are the time-two measures of the variables; a, b, and the e's are elements that make up the values of X_1, X_2, Y_1, and Y_2. Some of the relationships among the time-one and time-two measures can be predicted on the basis of McNemar's interpretation of the correlation coefficient. X_1, for instance, correlates zero with Y_1 and Y_2, since these measures have no common elements. The e's, being error terms, are independent of each other. X_2, on the other hand, correlates positively with Y_1 and Y_2, the magnitude of correlation being dependent upon the magnitude of b (the common element) relative to a and the e's.

The key feature of the formulation, however, lies in the inclusion of b in X_2. This element, a part of the time-one and time-two measures of variable Y, now becomes a new element of the time-two measure of variable X. That is, the change in variable X between the two time points is due to the inclusion of an element which is originally a part of variable Y. The formulation suggests, therefore, that the change which occurs in variable X is determined by an element in variable Y. In terminology which will be used in the sections to follow, we would say that Y is the source of causal influence.

Based on this conceptualization of causal relations, it is possible to generate data sets with predetermined causal characteristics. In the simplest case, a set of two-digit random numbers can be selected from a random number table to replace a, b, and the e's in the equations. When this is done, we will have generated four measures (X_1, X_2, Y_1, Y_2) for a hypothetical individual which imply causal relations between variables X and Y (i.e., Y is the source of causal influence). By repeating the same process, we are able to obtain any desired number of such hypothetical cases.

In the present study, the simulation process employed a computer random number generator to provide two-digit numbers with a normal distribution for each data set. A variety of data sets were created by manipulating several key elements in the causal formulation. By multiplying the random numbers by a fudge factor (e.g., multiplying the a's with .50, the b's with .40, and the e's with .10), the relative magnitude of these elements could be varied to produce a particular type of data. For example, by increasing or decreasing the magnitude of the e's relative to the a's and the b's, data sets of varying degrees of reliability were created. Similarly, by varying the magnitude of the b's (the common element) relative to the a's and the e's, it was possible to build into the data greater or lesser amounts of causal influence. Furthermore, by adding another common element, c, to X_1 and Y_1, we obtained data sets in which the time-one measures are correlated. Again, the degree of correlatedness was manipulable by varying the magnitude of c relative to the other elements. Finally, by switching the sign of b in X_2 we were able to build into the data sets causal influence which is congruent (positive), incongruent (negative), or both (i.e., the causal influence is congruent for some hypothetical cases and incongruent for others in the same data set).

The simulation created some 110 data sets, each consisting of 300 hypothetical cases. These data sets were divided into six categories according to the nature of causal influence and correlatedness of the time-one measures. Specifically, 20 data sets were generated in which the causal influence was congruent and the time-one measures were uncorrelated. Another 20 data sets contained data in which the causal influence was incongruent and the time-one measures were uncorrelated. Fifteen data sets were created in which congruent and incongruent causal influences were present simultaneously (i.e., the influence was congruent for one-half of the hypothetical cases and incongruent for the other half) and the time-one measures were uncorrelated. The other 55 data sets were created to parallel the foregoing data types, the only difference being that the time-one measures in the latter data sets were correlated. In all the data sets, variable Y was simulated to be the source of causal influence. The number and types of data sets generated are summarized in Table 1.

Table 1 about here

Upon completion of the simulation process, some of the psychometric characteristics of the data sets were examined. As indicated earlier, the reliability of the data was manipulated by varying the magnitude of the e's relative to the other elements. For example, by multiplying a with .90 and e_1 with .10 in equation (1), the true score is made nine times as great as the error score. As defined by Gulliksen (1950), the reliability coefficient is the ratio of the true variance to the observed variance. Since multiplying an element by a constant will multiply the variance by the square of the constant, the ratio between true variance and error variance (i.e., data reliability) could, therefore, be determined for each of the measures included in the data sets. Based on this procedure, the reliabilities of the various measures were ascertained. They were found to range from .50 to .99.

The correlatedness of the time-one measures was assessed directly by computing correlation coefficients for such measures. As expected, in data sets where the time-one measures had no common elements, these correlations were found to be close to zero. In data sets where common elements were built into the time-one measures such correlations ranged from .30 to .70. These correlation coefficients were consistent with the degree of correlatedness that was built into the data.

Also as indicated earlier, the amount of influence that variable Y had on variable X was manipulated by varying the magnitude of b (the common element) relative to the other elements. Correlation coefficients computed between X_2 and Y_1 varied from .30 to .70. This suggests that the amounts of variance in X_2 that are attributable to Y_1 ranged from approximately one-tenth to one-half. This characteristic, again, is consistent with what was built into the data. Moreover, the amounts of attributable variance are typical of investigations in the social sciences.

Findings

Decision Rules

Each of the five causal methods was applied to all the simulated data sets, resulting in a variety of causal interpretations. The determination of the validity of causal interpretations yielded by each causal method was guided by a set of decision rules. Since the causal interpretations were derived in a variety of ways, it was necessary to devise different decision rules for different causal methods. These rules are described, separately for each causal method, as follows:

1. Cross-lagged correlation method (CLC). If (1) the correlation between Y_1 and X_2 is significant ($P < .05$) and the correlation between X_1 and Y_2 is not and (2) the sign of the correlation between Y_1 and X_2 coincides with the direction of causal influence (i.e., congruent/positive or incongruent/negative) that was built into the data set, then the method is considered to have yielded a valid causal interpretation. With respect to data types E and F, the causal interpreataion is considered to be valid if the source of causal influence is correctly identified, i.e., if rule (1) is satisfied. The decision rules are relaxed because it is impossible for coefficients yielded by this method to bear both positive and negative signs simultaneously.

2. Part correlation method (PC). If (1) $R_{y1(x2.x1)}$ is significant ($P < .05$) and $R_{x1(y2.y1)}$ is not and (2) the sign of the former correlation coefficient coincides with the direction of causal influence that was built into the data set, then the method is considered to have yielded a valid causal interpretation. With respect to data types E and F, the causal interpretation is considered to be valid if the source of causal influence is correctly identified, i.e., if rule (1) is satisfied. The decision rules are relaxed because it is impossible for coefficients yielded by this method to bear both positive and negative signs simultaneously.

3. Econometric method (EC). If (1) the beta weight for Y_1 (when used to predict X_2) is significant ($P < .05$) and the beta weight for X_1 (when used to predict Y_2) is not and (2) the sign of the former beta weight coincides with the direction of causal influence that was built into the data set, then the method is considered to have yielded a valid causal interpretation. With respect to data types E and F, the causal interpretation is considered to be valid if the source of causal influence is correctly identified, i.e., if rule (1) is satisfied. The decision rules are relaxed because it is impossible for coefficients yielded by this method to bear both positive and negative signs simultaneously.

4. Frequencies-of-change-in-product-moment method (FCP). The decision rules for this method involve the use of the following

notations: YC is the number of cases (out of a sample of 300 cases) where variable Y is inferred to be the source of causal influence and the direction of such influence is congruent; YI is the number of cases where variable Y is inferred to be the source of causal influence and the direction of such influence is incongruent; XC is the number of cases where variable X is inferred to be the source of causal influence and the direction of such influence is congruent; XI is the number of cases where variable X is inferred to be the source of causal influence and the direction of such influence is incongruent. If (1) YC + YI is significantly ($P < .05$) greater than XC + XI (as indicated by a chi-square test) and (2) YC is significantly greater or less than YI according to the direction of causal influence that was built into the data set (i.e., YC is greater than YI if the causal influence is congruent and less than YI if the causal influence incongruent), then the method is considered to have yielded a valid causal interpretation. With respect to data types E and F, YC and YI are expected to be equal. This is because in these data types congruent causal influence was built in for one-half of the hypothetical cases and incongruent causal influence was built in for the other half of the hypothetical cases.

5. <u>Frequencies-of-shift-across-median method (FSM)</u>. The decision rules involve the use of the following notations: YC, YI,

XC, and XI have the same meanings as in FCP; XU is the number of cases where variable X is inferred to be the source of causal influence but the direction of such influence is uncertain; YU is the number of cases where variable Y is inferred to be the source of causal influence but the direction of such influence is uncertain. If (1) YC + YI + YU is significantly ($P < .05$) greater than XC + XI + XU (as indicated by a chi-square test) and (2) YC is significantly greater or less than YI according to the direction of causal influence that was built into the data set (i.e., YC is greater than YI if the causal influence is congruent and less than YI if the causal influence is incongruent), then the method is considered to have yielded a valid causal interpretation. With respect to data types E and F, YC and YI are expected to be equal--for the same reason alluded to in the discussion of decision rules for FCP.

Validity of the Causal Methods

Guided by the decision rules described in the preceding section, causal interpretations provided by the five methods for the 110 data sets were judged as valid or erroneous. Table 2 presents a summary of percentages of valid causal interpretations yielded by each causal method relative to each data type.

Table 2 about here

The cross-lagged correlation method, for instance, yielded 75% valid causal interpretations for type A data sets. When this method was applied to type B data sets, all the causal interpretations were found to be valid. With respect to type C, D, E, or F data sets, the validity of this method was shown to be extremely low, with percentages of valid interpretations ranging from 10% to 0%. The part correlation method yielded high percentages of valid interpretations (85% to 100%) for data sets in types A, B, and C. The chances of obtaining valid causal interpretations with data sets in types D, E, and F are low (45% to 0%). The econometric method provided 90% valid causal interpretations for type A data sets. All causal interpretations yielded by this method for type B data sets were found to be valid. With respect to data sets in types C, D, E, and F, the method was shown to have extremely low validity, with percentages of valid causal interpretations ranging from 25% to 0%. The validity of the frequencies-of-shift-across-median method and the frequencies-of-change-in-product-moment method with respect to data types A, B, C, D, and F was found to be generally low; percentages of valid causal interpretations ranged from 70% to 13%. These two methods, however, provided high percentages (93% and 87%) of valid interpretations for type E data sets.

Based on the results summarized in Table 2, it would appear that the causal methods are most likely to yield valid causal interpretations when they are applied to data sets where the causal influence is congruent and the time-one measures are correlated to some degree. In general, causal interpretations yielded for the other data types are likely to have a low level of validity. With the possible exceptions of the frequencies-of-shift-across-median method and the frequencies-of-change-in-product-moment method, the causal methods cannot be expected to provide valid interpretations with data types where congruent and incongruent influences are present simultaneously. (Although the effect of error variance on causal analysis was not examined in detail in the present study, evidence suggests that it is more difficult to obtain valid causal interpretations with data of low reliability than it is with data of higher reliability.)

To ascertain the overall validity of the causal methods, the total number of valid causal interpretations yielded by each method across all data types was examined. The results (see Table 3) suggest that the most versatile causal method, the part correlation method, can be expected to yield valid causal interpretations approximately two out of three times. The least versatile method, the cross-lagged correlation method, can be expected to provide valid causal interpretations only one out of three times.

Table 3 about here

Discussion and Conclusions

Two limitations of the study should be noted. First, a small number of data sets were used to evaluate the causal methods. As a result, the study lacks the vigor of a true Monte Carlo study which might enable one to assess the probability of success of each method in yielding valid causal interpretations relative to a particular data type. Secondly, all decision rules were based on the statistical significance (or non-significance) of the various causal coefficients. To the extent that sample size affected the statistical significance of these estimates, the efficacy of the decision rules was dependent upon the particular sample size used in the study.

The results of the study suggest that the five causal methods could not be regarded as comparable to each other. They can, in fact, be expected to yield divergent interpretations when applied to the same data set. Thus, given a set of data, only certain methods may yield valid causal inferences. The use of the other methods would probably result in erroneous causal interpretations. Perhaps more importantly, similarity of results yielded by different methods does not necessarily imply convergent validity. For example, all five methods could conceivably produce the same erroneous causal interpretation when applied to data types where congruent and incongruent causal influences co-exist.

Overall, causal interpretations yielded by the five causal methods are as likely to be erroneous as they are likely to be valid. The findings relating to the validity of the cross-lagged correlation method are particularly interesting. This method is probably the most widely used in panel analysis, yet it was shown to be the least versatile of all existing methods.

References

Blalock, H. M. Theory Construction. Englewood Cliffs: Prentice-Hall, 1969.

Campbell, D. T. "From description to experimentation: interpreting trends as quasi-experiments." In C. W. Harris (Ed.), Problem in Measuring Change. Madison: University of Wisconsin Press, 1963, pp. 212-242.

Campbell, D. T., & J. C. Stanley. "Experimental and quasi-experimental designs for research on teaching." In N. L. Gage (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963, pp. 171-246.

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

Darlington, R. B. "Multiple regression in psychological research and practice." Psychological Bulletin, 69 (1968), pp. 161-182.

Gulliksen, H. Theory of Mental Tests. New York: Wiley and Sons, 1950.

Linn, R. L., & C. E. Werts. "Assumptions in making causal inferences from part correlations, partial correlations, and partial regression coefficients." Psychological Bulletin, 72 (1969), pp. 307-310.

McNemar, Q. Psychological Statistics (4th Ed.). New York: John Wiley & Sons, 1969, pp. 145-146.

Richards, J. M. "A simple analytic model for college effects." The School Review, 74 (1966), pp. 380-392.

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

Tukey, J. W. "Causation, regression, and path analysis." In O. Kempthorn, et al. (Eds.), Statistics and Mathematics in Biology. Ames: Iowa State College Press, 1954, pp. 35-66.

Werts, C. E., & Watley, D. J. "Analyzing college effects: Correlation vs. regression." American Educational Research Journal, 5 (1968), pp. 585-598.

Wonnacott, R. J., & Wonnacott, T. H. Econometrics. New York: Wiley & Sons, 1970.

Yee, A. H. Factors Involved in Determining the Relationship Between Teachers' and Pupils' Attitudes. U. S. Department of Health, Education, and Welfare, Office of Education Research Project No. 5-8346. Austin: The University of Texas, 1966.

Table 1

Number and Types of Data Sets
Generated in the Simulation Study

Congruent Influence Only		Incongruent Influence Only		Congruent and Incongruent Influence	
A	B	C	D	E	F
20	20	20	20	15	15

Note: Each of the data sets consists of 300 hypothetical cases. Time-one measures in data types A, C, and E are uncorrelated; time-one measures in data types B, D, and F are correlated.

Table 2

Validity of Five Causal Methods
Relative to Each Data Type

Causal Method	Data Type					
	A	B	C	D	E	F
CLC	75	100	10	0	0	0
PC	95	100	85	45	0	40
EC	90	100	25	10	0	0
FSM	50	60	35	25	93	13
FCP	70	55	40	35	87	40

Note: Figures in the table represent percentages of valid causal interpretations yielded by each method. Data types are designated as in Table 1.

Table 3

Overall Validity of Five Causal Methods

Method	No. and % of valid interpretations		Rank
PC	71	(64%)	1
FCP	59	(54%)	2
FSM	50	(45%)	3
EC	45	(41%)	4
CLC	37	(34%)	5

Note: Percentages are based on a total of 110 simulated data sets.