MULTIPLE LINEAR REGRESSION VIEWPOINTS
Volume 13, Number 1 Spring 1984

MICROCOMPUTER-RESIDENT PROGRAM FOR THE ANALYSIS OF STRUCTURAL EQUATIONS

Lee M. Wolfle

Virginia Polytechnic Institute and State University

ABSTRACT

A microcomputer-resident program for the analysis of structural equations (PASE) has been designed to provide the causal modeler with all of the usually desired estimates of coefficients in recursive causal models. The program is interactive and self-documenting, and requires only a series of option selections by the user. Output includes all of the usual regression coefficients, plus total causal effects decomposed into direct and indirect causal effects.

Structural equation causal models provide a powerful aid to assist in the substantive interpretation of social and educational processes. Unlike straightforward regression analyses, structural equation analyses permit the measurement, not only of direct causal effects, but also of indirect causal effects through other, causally intervening independent variables (Finney, 1972). For example, it is now well understood that the primary reason father's occupational status is so closely associated with son's occupational status is not that sons directly inherit their father's status, but rather that sons of fathers with high status attain educations of a level that allow entry into occupations of higher status.

Wolfle (1980), among others, showed how application of the basic theorem (Duncan, 1966) or first law (Kenny, 1979) of path analysis could be used to aid in the interpretation of the causal effects of one variable in a model on another. While the application of the first law of path analysis provides a useful aid in interpretation, in many cases its computation is tedious in practice. Alwin and Hauser (1975), followed by Wolfle (1983), showed how a series of relatively simple regression equations could be used to estimate the direct and indirect causal effects in a hierarchical structural equation model.

The present paper describes a microcomputer program designed to provide the causal modeler with all of the usually desired estimates of coefficients in recursive causal models, and also yields all total, direct, and indirect causal effects implied by the hierarchical causal ordering of variables in the equation.

The microcomputer program accomplishes this goal by the simple expedient of requesting the user to supply information about the causal order of variables in the model, and with this information calculates a series of reduced-form equations from which the total causal effects of independent variables are stored in the computer's memory. The differences between the reduced-form coefficients, or total effects, and the fully specified, or direct effect, coefficients are the total indirect causal effects. Algebraic proofs of this relationship have been provided by Griliches and Mason (1972) and Wolfle (1983).

The microcomputer program described in this paper, PASE: Program for Analysis of Structural Equations, was designed to provide the causal modeler with all of the usually desired estimates of coefficients in recursive causal models. The program is user friendly in that it is interactive and self-documenting, and requires only a series of option selections by the user. The program requires the input of a zero-order correlation matrix from either an existing file or the keyboard. Output includes all of the usual regression coefficients, plus total causal effects decomposed into direct and indirect causal effects.

System Requirements

PASE was written for an Apple II or Apple II Plus microcomputer that utilizes Applesoft BASIC. System configuration must be a minimum of 48K RAM, and one disk drive operating on DOS 3.3. The program provides support for, but does not require, a printer for hard-copy output.

The Analysis of Structural Equations

The most important advance in social research methodology in the past 15 years has been the introduction (Duncan, 1966) to the social sciences of causal modeling techniques first worked out over 60 years ago (Wright, 1921, 1925). On the one hand, this development has been important to social theory, for the techniques of causal modeling provide an explicit link between theory and the equations used to test the hypothesized relationships. On the other hand, while the estimation methods for structural equations implied by causal models are not new, the techniques have proven to be invaluable aids in the interpretation of social data. One of the most important of these interpretative aids in causal modeling is the decomposition of zero-order associations among variables into various causal components (see Wolfle, 1980).

A zero-order association may develop for one or all of three reasons. The association may be spurious; that is, it can develop because two variables, say X and Y, are related because they are both caused by a prior variable, Z, or a set of Z's. To the extent that the relationship between X and Y is spurious, that portion is called a noncausal component of the zero-order association. The remaining portion of the association between X and Y is causal, and is called the <u>total effect.</u> Total effects may in turn be decomposed into <u>direct</u> causal effects and <u>indirect</u> causal effects. Direct effects in recursive causal models are nothing more than partial regression coefficients of a variable regressed on all causes of it. The indirect causal effect of one variable on another is that portion of the total effect that can be traced through causally intervening variables. Such coefficients, both direct and indirect, can be expressed in either standardized or unstandardized (metric) form; the latter are often preferred, because standardized coefficients are relatively unstable from sample to sample or across populations (Duncan, 1975; Kim and Mueller, 1976).

Users of structural equation techniques need to keep in mind, however, that the interpretations of causal effects are model specific. If the causal model is plausible, the variables within it credibly ordered and accurately measured, then the interpretations of effects within it are plausible. If these conditions are not met, however, then the interpretations based on faulty models are themselves faulty.

Program Input and Output

A new computer program written for the Apple microcomputer, called PASE (Wolfle, 1982), provides a potentially useful tool for estimating hierarchical, recursive causal models. Because such models depend upon least-squares estimation procedures, PASE provides all of the usual regression coefficients. In addition, PASE provides estimates of total causal effects, and decomposes these into direct and indirect components.

PASE permits the input of new correlation matrices along with means and standard deviations. All data matrices can be saved to disk for future analyses. The program thus permits either the input of new matrices or the reading of previously saved data. Data matrices can be reviewed, corrected, truncated, or expanded to the maximum-sized matrix (17 variables) analyzable with the 48K memory limits of the compiled version of PASE.

Once the data have been input, reviewed if desired, changed if necessary, and saved to disk as recommended, the program prompts the user for the number of equations in the causal model. The program next asks the user to specify the dependent variable, followed by a list of the independent variables. The program next requests the user to specify the causal order among the independent variables. With this information, the program proceeds with the calculation of all regression coefficients, both standardized and metric, and decomposes these into direct and indirect causal components. (If one desires, the noncausal component of an association may be calculated by the simple expedient of subtracting the total causal effect from the zero-order association.)

The output of PASE has been organized for easy review. The output menu gives the user the option of reviewing the regression results, the regression ANOVA table, the R-squares among the independent variables, and the decomposition of causal effects. If desired, all of these results may be directed to a printer.

The regression results include all metric slopes, beta weights, standard errors, and t-ratios for the independent variables. The value of the intercept and the R-square for the regression are also included.

The ANOVA table includes the usual regression, residual, and total sums of squares, along with their associated degrees of freedom and mean squares. From these the F-ratio is calculated, and presented along with the standard error of estimate and the regression R-square.

The R-squares among the independent variables may be viewed. A high value among these suggests the presence of multicollinearity, which if present causes regression coefficients to be unstable in the face of slight changes in the zero-order correlation coefficients (see Gordon, 1968). In addition, standard errors are often inflated, and highly correlated independent variables often (and implausibly) have regression coefficients of opposite sign (see, for example, Muffo and Coccari, 1982).

The table of causal decomposition presents the total effect of each independent variable, along with its direct effect and total indirect causal effect. If there are no intervening variables between the causal independent variablo and the caused dependent variable, then the total effect is the direct effect.

<u>An Illustration</u>

To illustrate the use of PASE, refer to the causal model illustrated in Figure 1. The model is based on some analyses presented in Duncan, Featherman, and Duncan (1972), and the data taken from Duncan (1968). Of particular interest in this model is the relationship between ability and earnings; what is tho expected relationship between intelligence and earnings, controlling for social background, educational training, and occupational prestige?

There are three endogenous variables in the model; therefore, there are three equations to be estimated. Focusing attention on the equation for earnings, X(1), one would specify upon request by the program that variable 1 is dependent. The user will then be asked to specify the variable numbers of the causes of X(1); therefore, the user will input variable numbers 2, 3, 4, 5, and 6, since all other variables in the model are hypothesized to cause earnings.

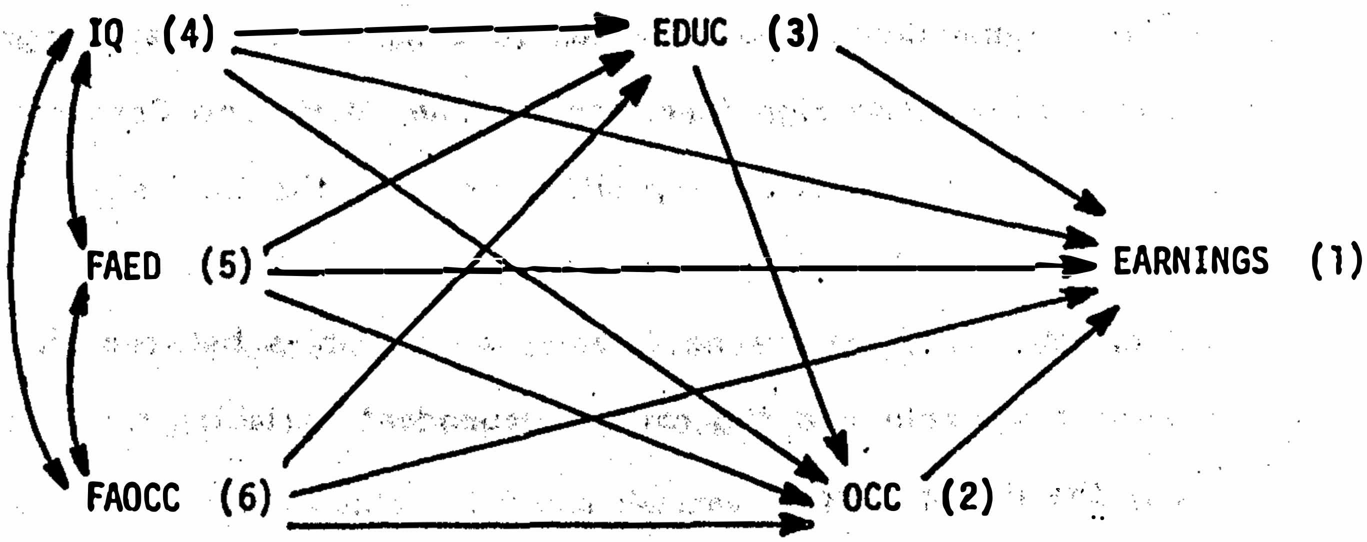

Figure 1. Path Diagram Representing Dependence of Earnings on Status Attainment, Intelligence, and Family Background.

The program will next ask the user to identify the causal order of the independent variables. In this case, X(4), X(5), and X(6) occur simultaneously in a single causal block of exogenous variables. The user would therefore input variable numbers 4, 5, and 6 as constituting the variables in Block 1 (followed by the value 99 to terminate the Block listing). Educational attainment, X(3), is the single variable in Block 2, followed by occupational prestige, X(2), in Block 3. With this blocking information, the program proceeds with the calculation of the coefficients for X(1).

The regression results for this equation are shown in Table 1. These indicate by examination of the beta weights that the most important effect of earnings is the prestige level of the respondent's occupation, X(2). The relative effects of education and intelligence are less than half that of prestige, while the influence of father's education and occupation are statistically indistinguishable from zero.

The decompositions of causal effects for this equation are shown in Table 2. Examination of the total causal effects indicate that intelligence, educational attainment, and occupational prestige all have about equal total effects on earnings. The indirect effects indicate that about half of the total effect of education on earnings occurs indirectly through occupation; that is, those people with higher levels of educational attainment not only receive higher earnings ceteris paribus, but also tend to enter occupations of higher prestige which in turn lead to higher earnings.

Table 1. Regression Results

Dependent Variable: 1

Var.	B	Beta	St.Err	T
2	.2625	.2625	.0381	6.8821
3	.1069	.1069	.0423	2.5277
4	.1013	.1013	.0344	2.9436
5	.0306	.0306	.0342	.8958
6	.0183	.0183	.0348	.5263

Variables:
1 = 1964 earnings,
2 = 1964 occupation,
3 = education,
4 = "early" intelligence,
5 = father's education,
6 = father's occupation.

Table 2. Decomposition of Causal Effects

(Standardized)

FROM	TOTAL	DIRECT	INDIRECT
VAR. 4	.2273	.1013	.1261
VAR. 5	.0881	.0306	.0574
VAR. 6	.1032	.0183	.0849
VAR. 3	.2454	.1069	.1385
VAR. 2	.2625	.2625	0

Variables: See Table 1.

The same may be said for the effects of intelligence on earnings. There is a direct causal effect of intelligence on earnings (the higher one's intelligence <u>ceteris paribus</u>, the higher one's earnings), but there is also a set of indirect effects wherein people of higher intelligence acquire higher levels of education and possess occupations of higher prestige, which also have positive effects on earnings. The combined direct and indirect effects of intelligence make it equally important to the explanation of earnings as is either education or occupational prestige.

In sum, PASE not only permits the causal modeler to examine the straightforward regression results, but, further, PASE also allows one to examine the decomposition of causal effects into their direct and indirect components. These latter examinations often prove to be very useful in revealing how causal effects are manifested in the model.

<u>Availability of PASE</u>

PASE is available from the author, College of Education, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061. Please enclose one blank 5.25-inch, soft-sectored floppy disk compatible with the Apple disk operating system. A users' guide is also available; to cover duplication costs, please enclose a check in the amount of $1.00 made out to VPI&SU College of Education.

REFERENCES

Alwin, Duano F., and Hauser, Robert M., "The Decomposition of Effects In Path Analysis." American Sociological Review, 1975, 40, 37-47.

Duncan, Otis Dudley, "Path Analysis: Sociological Examples." American Journal of Sociology, 1966, 72, 1-16.

Duncan, Otis Dudley, "Ability and Achievement." Eugenics Quarterly, 1968, 15, 1-11.

Duncan, Otis Dudley, Introduction to Structural Equation Models. New York: Academic Press, 1975.

Duncan, Otis Dudley, Featherman, David L., and Duncan, Beverly, Socioeconomic Background and Achievement. New York: Seminar Press, 1972.

Finney, John M., "Indirect Effects in Path Analysis." Sociological Methods & Research, 1972, 1, 175-186.

Gordon, Robert A., "Issues In Multiple Regression." American Journal of Sociology, 1968, 73, 592-616.

Griliches, Zvi, and Mason, William M., "Education, Income, and Ability." Journal of Political Economy, 1972, 80, S74-S103.

Kenny, David A., Correlation and Causality. New York: Wiley, 1979.

Kim, Jae-On, and Mueller, Charles W., "Standardized and Unstandardized Coefficients In Causal Analysis: An Expository Note." Sociological Methods & Research, 1976, 4, 423-438.

Muffo, John A., and Coccari, Ronald L., "Predictors of Outside Funding for Research Among AASCU Institutions." Research in Higher Education, 1982, 16, 71-80.

Wolfle, Lee M., "Strategies of Path Analysis." American Educational Research Journal, 1980, 17, 183-209.

Wolfle, Lee M., "PASE: Program for Analysis of Structural Equations." Behavior Research Methods & Instrumentation, 1982, 14, 548-550.

Wolfle, Lee M., "A Method for Estimating Indirect Effects In Path Analysis." Multiple Linear Regression Viewpoints, 1983 (in press).

Wright, Sewell, "Correlation and Causation." Journal of Agricultural Research, 1921, 20, 557-585.

Wright, Sewell, _Corn and Hog Correlations_. Washington, D.C.: U.S. Department of Agriculture Bulletin 1300, 1925.