

VOLUME 9 NUMBER 2

J.D. WILLIAMS

NOVEMBER 1978

PATH ANALYSIS FROM A
REGRESSION PERSPECTIVE

MULTIPLE LINEAR REGRESSION VIEWPOINTS

A publication of the Special Interest Group
on Multiple Linear Regression

MONOGRAPH SERIES #3

MLRV Abstracts appear in CIJE, the ERIC System,
and microform copies are available from
University Microfilms International

MULTIPLE LINEAR REGRESSION VIEWPOINTS

- Chairman** John Williams
University of North Dakota
Grand Forks, ND 58201
- Editor** Isadore Newman, Research and Design Consultant,
The University of Akron, Akron, OH 44325
- Assistant** Diane Vukovich
The University of Akron, Akron, OH 44325
- Executive Secretary** Steve Spaner, Behavioral Studies
University of Missouri, St. Louis, MO 63121
- Chairman-elect** William Connett
State Department of Education, State Capital, MT 59601
- Cover by** David G. Barr

EDITORIAL BOARD

- | | | | |
|---|--|---|---|
| 1 | Dr. William Connett
State Department of Education
State Capital, MT 59601 | 7 | Dr. Michael McShane
Association of Medical Colleges
One Dupont Circle
Washington, D.C. 20036 |
| 9 | Leigh Burstein
Department of Education
University of California
Los Angeles, CA 90024 | | Dr. Isadore Newman
College of Education
The University of Akron
Akron, OH 44325 |
| 3 | Dr. Robert Deitchman
Psychology Department
The University of Akron
Akron, OH 44325 | 2 | Dr. Joe H. Ward, Jr.
Lackland Air Force Base
San Antonio, TX 78228 |
| 4 | Dr. Samuel Houston
University of North Colorado
Greenly, CO 80639 | 6 | Dr. John Williams
University of North Dakota
Grand Forks, ND 58201 |
| 8 | Dennis Leitner
Department of Guidance and
Educational Psychology
Southern Illinois University
Carbondale, IL 62901 | 5 | Dr. Lee Wolfle
Virginia Polytechnic
Institute and State University |

PATH ANALYSIS FROM A REGRESSION PERSPECTIVE

John D. Williams

The University of North Dakota

Copyright 1978

TABLE OF CONTENTS

Chapter		Page
I	Introduction	1
II	Recursive Systems	5
III	Structural Equations	26
IV	Comparing Results from Two Different Data Sets	29
V	Nonrecursive Models	40
VI	Examples of Path Analysis in Educational Research	51
VII	Other Applications of Path Analysis	59
	Tables	66
	References	74

Chapter One

Introduction

Point of View

This little monograph on path analysis is addressed to a somewhat different group of readers than the typical presentation on path analysis. It is assumed that the reader has had at least a modicum of experience with multiple linear regression and the use of full and restricted models. It is presumed that the reader has read or is familiar with at least parts of one of the following texts: Bollen and Wooldredge (1983); Kelly, Bollen and McLeod (1969); Wold and Jöreskog (1973); Williams (1974a); or McLeod, Kelly and McLeod (1975). On the other hand, no presumptions are made regarding prior background in path analysis.

The Path Analysis Literature

Path analysis originated in the work of Sewall Wright (1921, 1934); his work on path analysis emanated from his study of genetics. The use of path analysis was for the most part restricted to genetics for a considerable time; economic applications were given in Wold and Jöreskog (1953), Wold (1954), Goldberger (1964) and Johnston (1972). Blalock (1962, 1964, 1971) was one of the earlier (if not the earliest) advocates of causal analysis in the social sciences. However, Duncan's (1966) paper seems to be the "jump off place" for the barrage of writing

regarding path analysis/structural equations/causal analysis in the social sciences. The American Sociological Association has sponsored a yearbook, Sociological Methodology that has been methodologically dominated by the concerns of path analysis. The editors of the series have been Borgatta (1969), Borgatta and Bohrnstedt (1970), Costner (1971, 1972, 1974), Heise (1974, 1975b, 1977) and presently Schuessler (1977). These yearbooks are an excellent source regarding current sociological concerns (and by inference, their relationships to educational research) of path analysis. Several texts are available that explicate path analysis/causal models, including Blalock (1964), Goldberger and Duncan (1973), Duncan (1975), Namboodiri, Carter and Blalock (1975), Heise (1975a), Li (1975), Hanushek and Jackson (1977) and Asher (1977).

To this point, path analysis has been used as a term synonymous with causal models and structural equations; while the English language is replete with redundancies, different words often are used to convey shades of meaning. So one might describe the shades of meaning among the three terms; to a person intimately familiar with the area, the terms convey a different meaning. Before the term path analysis is defined or contrasted with the two "new" synonyms, a necessary delimitation is made that presently will make sense to readers not in the intended audience, but who are somewhat more familiar with the literature regarding path analysis. The main technique considered in this monograph is termed "recursive models" and has been described in Land (1969) and presented to a regression oriented audience in Williams and Klimpel (1975). The reasons for this delimitation will become obvious only after the recursive models have been considered in detail and contrasted to other techniques in the path analysis oriented literature.

Path Analysis - A Perspective

As attractive as it is to define path analysis in terms of the several regression equations that will likely be involved, such a definition is likely to miss the whole point of path analysis. Users of path analysis will undoubtedly use regression models, but do so only because the underlying substantive theory requires it; some path analysts, such as Goodman (1975), have at times shown that path type models can be studied in contexts other than regression. Rather, users of path analysis emphasize the underlying theory, whether it be social, genetic, economic or psychological. On the other hand, the empirical value of structural equations has sometimes been questioned. Duncan (1975, p. viii) made these curious statements:

It is not my purpose to advocate or defend the use of structural equation models in sociology. Indeed, I hold a rather agnostic view about their ultimate utility and vitality in that discipline, fascinating as the models may be in purely formal terms. One thing that has given me pause is that I did not find it expedient to include here any substantive sociological examples, real or contrived. In thinking about the matter, it seemed to me that most of the persuasive real examples pertain to a narrowly delimited area of sociological inquiry. On the other hand, many contrived examples in the methodological literature of sociology are so utterly implausible as to call into question their very purpose.

Path Analysis - A Link Between Theory and Data Analysis

Perhaps the greatest value of path analysis is that it requires the researcher to make explicit statements of the precise models that are going to be formulated. Clearly, those whose approach to data analysis has been the "shoot it up the flagpole and see if anyone salutes" type of data

snooper will find path analysis a foreign type of approach. To be sure, data snooping has a place in the formulation of theory, but only rarely (if ever) in the testing of a theory. Insofar as theory testing is likely to become more rather than less important in the future of educational research, path analysis seems destined to hold a more important place in the analysis of educational data.

Organization of the Monograph

The major methodological process, recursive structural models, is presented in Chapter Two; the complete data set is given in Appendix A. Structural equations are considered in Chapter Three. In Chapter Four a second set of data is explored. The second set of data was collected on a younger group of students than the first, but the variables reported are identical. The second set of data is included in Appendix B. Various hypotheses are shown and tested as an example of combining the results of two different data sets.

The process of "two-stage least squares" is described at length in Chapter Five. Several educational research applications of path analysis are given in Chapter Six; the studies are by no means a complete listing of such applications, but they should convey some idea of the use of path analysis in educational research. Chapter Seven considers the relationship of path analysis to statistical methodologies other than multiple linear regression, including factor analysis.

Chapter Two

Recursive Systems

In the present chapter, a data set is considered in a path analysis framework. The complete data set is included in Appendix A. The data in Appendix A were collected by Williams and Brekke (in press). They were studying the logical relationships of various Piagetian stages and spatial relations. The variables included in the study were:

X_1 = 1 if a conserver (a Piagetian concept) 0 if a nonconserver;

X_2 = score on the Concrete Operational Reasoning Test (CORT) devised by Burney (1974); X_3 = score on the Logical Reasoning Test (LRT) devised by Ankney and Joyce (1974); X_4 = score on the Revised Minnesota Paper Form Board Test (RMPFBT), a measure of spatial relations (Likert and Quasha, 1970).

Any brief attempt to explain the Piagetian theory is likely to oversimplify to the point of losing most of the richness of the theory. Jean Piaget has published volumes of material on the thinking of children for over half a century; for much of that time, he has run an institute for the study of children's thinking in Geneva, Switzerland. His research went largely unnoticed in the United States until the 1960's. Since that time, there has been an explosion of research regarding Piagetian theory. Expositions of Piaget's theory have been written by Sund (1974) Piaget, Inhelder and Szeminska (1960), Piaget (1930), Elkind (1974), and Furth (1969, 1970), among many others. Over 50 additional references to publications by Piaget can be found in Magary et al. (1977).

Piaget's is a sequential theory; that is, the child's thinking develops in stages. Briefly, the stages are: (1) The sensory-motor stage (0-2 years); (2) The preoperational stage (2-7 years); (3) The concrete-operational stage (7-11 years); and (4) The formal-operational stage (11-14 years).

The measure of conservation (weight), X_1 , is a concrete-operational task that should be achieved near the midpoint of the concrete-operational stage. The Concrete-Operational Reasoning Test (CORT), X_2 , is a paper and pencil task that would be appropriate as a measure at the upper-end of the concrete-operational stage. The Logical Reasoning Test (LRT) is a measure at the formal-operational stage (X_3). The three Piagetian variables are clearly ordered; X_1 should occur first, then X_2 and finally X_3 . The fourth measure, a measure of spatial relations (RMPFBT), X_4 , for present purposes is considered a higher level skill than any of the three Piagetian tasks. Subjects in the present data set (N=116) were high school students. (See Appendix A).

Recursive Equations

Typically a set of recursive equations are used to define the relationship between the four variables:

$$X_2 = a_1 + b_{21}X_1 + e_1, \quad (2.1)$$

$$X_3 = a_2 + b_{31}X_1 + b_{32}X_2 + e_2, \text{ and } (2.2)$$

$$X_4 = a_3 + b_{41}X_1 + b_{42}X_2 + b_{43}X_3 + e_3. \quad (2.3)$$

In the recursive equations, each variable X_i is considered to be a possible cause of the variables occurring after X_i but not before X_i . The values $a_1 - a_3$ represent the intercepts for the corresponding equations. The regression coefficients given in equations 1-3 are indicated as partial regression weights. If a path diagram were made of the recursive

equations, Figure 1 would result.

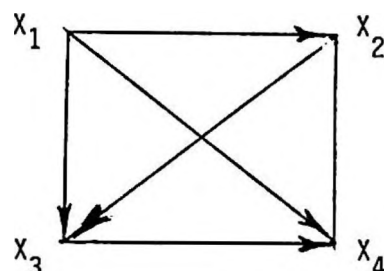


Figure 2.1 Recursive Path Diagram.

Quite often, the interest in path analysis is not in a complete set of paths as given by Figure 1, but rather in restricting the number of paths.

For the present data, it might be conjectured that a path for X_1 to X_4 is unneeded; that is, it does not seem likely that conservation will have any effect on spatial relations except thru the two intermediary Piagetian tasks. Dropping the path from X_1 to X_4 would transform equation 2.3 to $X_4 = a_4 + b_{42.3}X_2 + b_{43.2}X_3 + e_4$. (2.4)

Equation 2.4 presumes $\beta_{41.32}$ vanishes (or is equal to 0). This is equivalent to hypothesizing $\rho_{41.32} = 0$. To test this hypothesis, equation 2.3 can be used as a full model and equation 2.4 can be used as the restricted model:

$$F = \frac{(R_{4.123}^2 - R_{4.32}^2)/1}{(1 - R_{4.123}^2)/(N - k - 1)}, \quad (2.5)$$

where k is the number of paths to X_4 when all paths are represented.

Most users of path analysis focus on the standardized regression weights; these weights (beta coefficients) are called path coefficients. Figure 2.2 shows a path diagram with all paths included for the data in Appendix A.

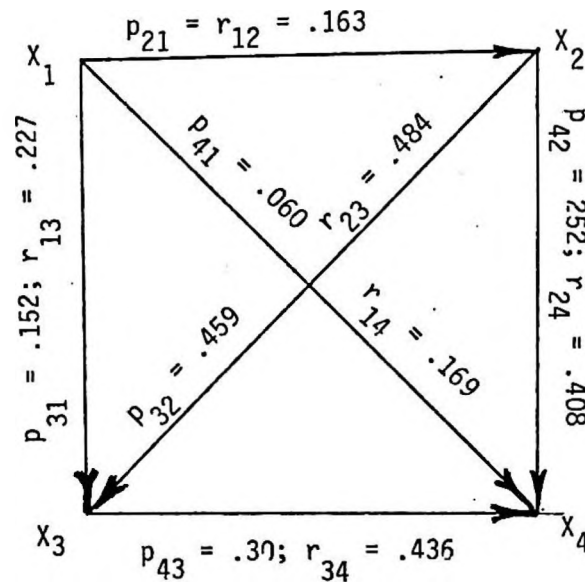


Figure 2.2. Path Diagram for Three Piagetian Tasks and the Minnesota Paper Form Board: X_1 = conservation, X_2 = Concrete-Operational Reasoning Test, X_3 = Logical Reasoning Test and X_4 = Revised Minnesota Paper Form Board Test (spatial relations).

If the path from X_1 to X_4 is eliminated, a new path diagram would result. Such a system would assume that $p_{41.32} = p_{41.32} = p_{41} = 0$. This assumption seems intuitively tenable, given that $p_{41} = .06$. A second path diagram with $p_{41} = 0$ is given in Figure 2.3.

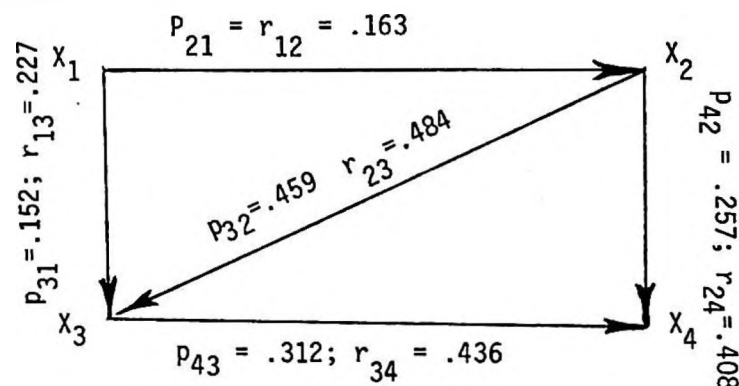


Figure 2.3. Path Diagram with $p_{41} = 0$. X_1 = conservation, X_2 = Concrete-Operational Reasoning Test, X_3 = Logical Reasoning Test and X_4 = Revised Minnesota Paper Form Board Test (spatial relations).

It can be seen that dropping p_{41} has little effect upon the remaining system. Both p_{42} and p_{43} change only slightly from the complete recursive system (Figure 2.2) to the reduced recursive system (Figure 2.3). The test given in equation 2.5 yields

$$F = \frac{(.24421 - .24080)/1}{(1 - .24421)/(116 - 3 - 1)} = .505.$$

This F value is nonsignificant, indicating the hypothesis that $p_{41} = 0$ is tenable. There is yet another way to achieve a test essentially the same as that given by equation 2.5. In that most computer programs also give the computed t values wherein the t value is a test of significance for the partial regression coefficient, the t value can be read from the printout and is $t = .711$. In that $t^2 = .505 = F$, the present result is analogous to the t test being a special case of the analysis of variance.

Tests of More than One Partial Regression Coefficient

The test given by equation 2.5 is appropriate if exactly one path coefficient is being tested. If, however, more than one coefficient is being tested in the model, the probabilities should be adjusted for the specific tests performed. That is, the concerns of multiple comparisons enters into the arena. All too often, multiple F (or t) tests are performed after using a full and restricted model for all restricted models without any adjustments to the probability values. In regard to the usual comparisons, Williams (1971, 1972, 1974b, 1975, 1976) has shown the various multiple comparisons in a regression format while also addressing the over-use of full and restricted models without adjusting the probability values. Bielby and Kluegel (1977) have also addressed this same problem.

While many ways of preserving the probability values could be used, the specific approach recommended here follows Dunn's (1961) test. Essentially, Dunn's test requires the researcher to fix the hypotheses to be tested on an a priori basis rather than using a "snooping" procedure such as Scheffe's (1953)

test. Note that in Figures 2.1 and 2.2 there are $\binom{4}{2} = 6$ paths. If these paths are all tested indiscriminately (as would occur with tests that consider all possible comparisons such as Tukey's, 1953, or Scheffe's test) it would seem to be out of character to the conception of path analysis; path analysis is, after all, at least as much an exercise in theory as it is a statistical technique. Few theories would encourage indiscriminate testing of hypotheses; rather, specific hypotheses would be tested for a priori stated reasons.

In regard to the data in Figure 2.2, suppose a researcher had hypothesized that no indirect effects occur; that is, no variable makes any contribution beyond the path to the immediately subsequent variable. See Figure 2.4.

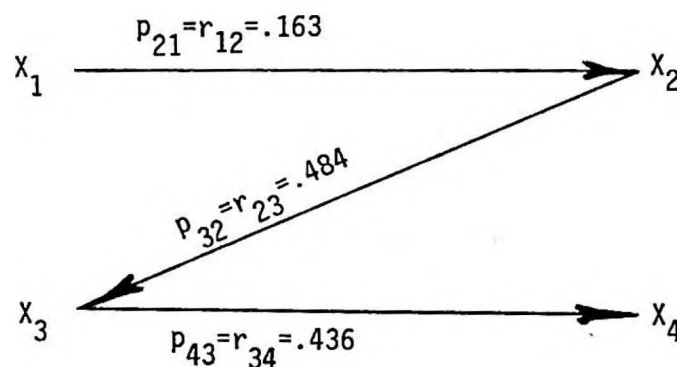


Figure 2.4. Path Diagram with $p_{31} = p_{41} = p_{42} = 0$. X_1 = Conservation, X_2 = Concrete-Operational Reasoning Test, X_3 = Logical Reasoning Test and X_4 = Revised Minnesota Paper Form Board Test (spatial relations).

Table 2.1 contains the path coefficients from the full recursive model (Figure 2.2) and with the corresponding computed t values; the correlation coefficients are also reported. The table is shown in two sections; those paths that are considered necessary by Figure 2.4 (p_{21} , p_{32} , p_{43}) are the top portion and those paths dropped from Figure 2.4 (p_{31} , p_{41} , p_{42}) are the bottom portion.

Table 2.1

Path Coefficients, Correlation Coefficients and Computed t Values for Figure 2.2.

Path Coefficient	r	t
$p_{21} = .163$	.163	1.77
$p_{32} = .459$	.484	5.59
$p_{43} = .300$	.436	3.15
$p_{31} = .152$	.227	1.85
$p_{41} = .060$	.169	.71
$p_{42} = .252$	.408	2.68

To test p_{31} , p_{41} and p_{42} , Dunn's tables (Appendix C and D respectively at the .05 and .01 levels) can be used. The df_w for p_{41} and p_{42} are 112; df_w for p_{31} are 113. There is so little difference in the critical values for such degrees of freedom that they can be ignored for large N; most tables (including those in Appendix C and D) do not provide separate entries for those values so that interpolation is necessary in any case. To two decimal points, the critical value for $m = 3$, $df_w = 113$, is 2.43 (at the .05 level) and 2.99 (at the .01 level). The computed t value for p_{42} , $t = 2.68$, is sufficiently large that we probably would consider retention of this coefficient. Indeed, we might be suspicious of p_{21} , since the corresponding t value is non-significant. For logical reasons, p_{21} would seem necessary in the system even though it is a non-significant path. Hence, a tentative model would include p_{21} , p_{32} , p_{42} and p_{43} , excluding p_{31} and p_{41} . See Figure 2.5.

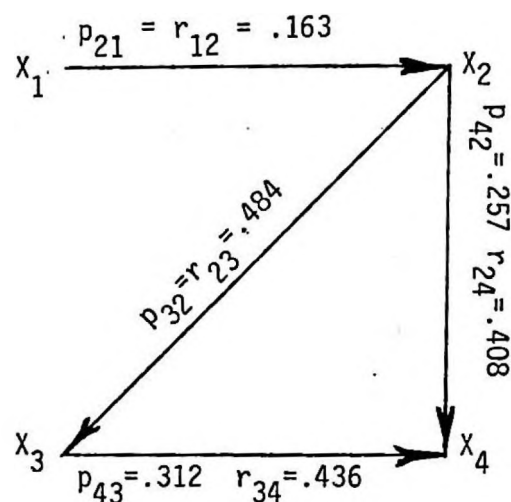


Figure 2.5. Tentative Model for Piagetian data and spatial relations. X_1 = Conservation, X_2 = Concrete-Operational Reasoning Test, X_3 = Logical Reasoning Test and X_4 = Revised Minnesota Paper Form Board Test (spatial relations).

The Problem of Arguing from a Non-Significant Finding

One difficulty with the approach shown in the previous section is that decisions to drop paths is made from a non-significant finding, rather than basing the decision on a significant finding, as is usually the case with statistical inference. While some readers may be satisfied by an inversion of the previous sentence, that is only those paths whose test proves to be significantly different than zero are retained, the use of a statistical test as the major determiner may prove to be inadequate to the task of path analysis. Certainly, paths that vary only slightly from zero but are significant due to a large sample size may still be discarded. This decision can be made on a statistical basis even with a large sample size. A researcher may choose an arbitrary value, g , such that if p is within a confidence band of the value, the path is discarded.

Given that the path coefficient and t value are available, the standard error of the path coefficient is given as $s_p = \frac{p}{t}$. (2.6) As an example, suppose the researcher had decided to drop all path coefficients

whose absolute value are less than .20 or those whose confidence band included .20 (among the coefficients to be tested). To test p_{42} , the confidence band is given by $p \pm t_{\alpha} s_p$, where t_{α} is taken from either Appendix C or D. Because there are three tests involved in the testing, p_{31} , p_{41} , and p_{42} , $m = 3$, $df_w = 113$; the critical value for $\alpha = .05$ is 2.43. Thus, the confidence limits are $.252 \pm .228 = .024, .480$. Under this criterion, p_{42} would be dropped.

Reproducing the Correlation Matrix

One of the more important contributions of path analysis is the reproduction of the correlation matrix. The correlation matrix can be reproduced using the following equation

$$r_{ij} = \sum_k p_{jk} r_{ik}, \quad (2.7)$$

where k includes every variable which has a path leading directly to X_j . For situations such as are depicted in Figure 2.2 wherein every variable is connected to every other variable with a path, equation 2.7 would identically reproduce the correlation matrix. On the other hand, models such as Figures 2.3, 2.4 and 2.5 would tend to have some departure from an exact reproduction of the correlation matrix.

The reproduction of the correlation matrix by successively using equation 2.7, proceeds as follows (using Figure 2.2):

$$r_{12} = p_{21} r_{11} = p_{21} = .163; \quad (2.8)$$

$$r_{13} = p_{31} r_{11} + p_{32} r_{12} = p_{31} + p_{32} r_{12} = .152 + (.459)(.163) = .227; \quad (2.9)$$

$$r_{14} = p_{41} r_{11} + p_{42} r_{12} + p_{43} r_{13} = p_{41} + p_{42} r_{12} + p_{43} r_{13} = (.060) + (.252)(.163) + (.300)(.227) = .169; \quad (2.10)$$

$$r_{23} = p_{31} r_{21} + p_{32} r_{22} = p_{31} r_{21} + p_{32} = (.152)(.163) + (.459) = .484; \quad (2.11)$$

$$r_{24} = p_{41} r_{21} + p_{42} r_{22} + p_{43} r_{23} = p_{41} r_{21} + p_{42} + p_{43} r_{23} = (.060)(.163) + (.252) + (.300)(.484) = .407; \quad (2.12)$$

$$r_{34} = p_{41}r_{31} + p_{42}r_{32} + p_{43}r_{33} = p_{41}r_{31} + p_{42}r_{32} + p_{43} = (.060)(.227) + (.252)(.484) + (.300) = .436. \quad (2.13)$$

All values for the correlation coefficients are within rounding error of the actual correlations.

In that the final model (Figure 2.5) had two paths (p_{31} and p_{41}) deleted it would be useful to compare the predicted correlations with the actual correlations. With p_{31} and p_{41} deleted,

$$\hat{r}_{12} = p_{21}r_{11} = p_{21} = .163; \quad (2.14)$$

$$\hat{r}_{13} = p_{32}r_{12} = (.484)(.163) = .079; \quad (2.15)$$

$$\hat{r}_{14} = p_{42}r_{12} + p_{43}r_{13} = (.257)(.163) + (.312)(.227) = .113; \quad (2.16)$$

$$\hat{r}_{23} = p_{32} = .484; \quad (2.17)$$

$$\hat{r}_{24} = p_{42} + p_{43}r_{23} = (.257)(.312)(.484) = .408; \quad (2.18)$$

$$\hat{r}_{34} = p_{42}r_{32} + p_{43} = (.257)(.484) + (.312) = .436. \quad (2.19)$$

Table 2.2 contains the actual correlations in the portion above the main diagonal in the correlation matrix and the predicted correlations under the deleted paths are found below the main diagonal.

Table 2.2

Actual (Above Main Diagonal) and Predicted (Below Main Diagonal) Correlations for the Piagetian and Spatial Relations Data

	X_1	X_2	X_3	X_4
X_1		.163	.227	.169
X_2	.163		.484	.408
X_3	.079	.484		.436
X_4	.113	.408	.436	

Legend: X_1 = Conservation, X_2 = Concrete-Operational Reasoning Test, X_3 = Logical Reasoning Test and X_4 = Revised Minnesota Paper Form Board Test (spatial relations).

The discrepancy between $r_{13} = .227$ and $\hat{r}_{13} = .079$ might be more than some researchers wish to entertain; they could then add p_{31} back into the system if they so desire.

Further Decomposition of Correlation Equations

The set of equations from 2.8 to 2.13 can be represented as path coefficients only, with the correlations appearing as solutions to the equations. Equation 2.8 is already in such form. If equation 2.9 is inspected,

$$r_{13} = p_{31} + p_{32}r_{12}. \text{ But } r_{12} = p_{21}; \text{ therefore } r_{13} = p_{31} + p_{32}p_{21}. \quad (2.20)$$

In a similar manner, equations 2.10 thru 2.13 can be transformed:

$$r_{14} = p_{41} + p_{42}p_{21} + p_{43}(p_{31} + p_{32}p_{21}); \quad (2.21)$$

$$r_{23} = p_{31}p_{21} + p_{32}; \quad (2.22)$$

$$r_{24} = p_{41}p_{21} + p_{42} + p_{43}(p_{31}p_{21} + p_{32}); \quad (2.23)$$

and

$$r_{34} = p_{41}(p_{31} + p_{32}p_{21}) + p_{42}(p_{31}p_{21} + p_{32}) + p_{43}. \quad (2.24)$$

It can be noticed that each equation (2.8, 2.20 thru 2.24) includes the path coefficient for the path from the two variables involved in the correlation plus some additional material. For example, equation 2.20 for r_{13} includes p_{31} plus $p_{32}p_{21}$. The path coefficient corresponding to the two variables being correlated is called the direct effect. What is meant by a direct effect is precisely defined by equations 2.8 and 2.20 thru 2.24; Darlington (1968) has pointed out the limitations of using beta coefficients and compares several related concepts that, because of similar terminology, sometimes get conceptually confused. Williams and Klimpel (1975) show that the direct effect and independent contribution of the variable are, in fact, different concepts. Wolfle (1977 b), following Finney (1972), extends the Williams and Kimpel paper by separating the components that make up the indirect effect into more definable components. Wolfle (1977 a) extends his earlier paper by giving an explanation in typical sociological

terminology of the different indirect effects. Table 2.3 shows the decomposition of each correlation into four components, the direct effects, indirect effects, spurious correlation, and joint associations. In relation to the decomposition of the total indirect effects into components, see also Anderson and Evans (1974) and Spaeth (1975).

Table 2.3

Assessment of the Effects of the Fully Recursive Model
(All Paths Included)

Variables	Total Association	Direct Effects	Indirect Effects	Spurious	Joint Association
X_1 on X_2	.163	.163	—	—	—
X_1 on X_3	.227	.152	.075	—	—
X_1 on X_4	.169	.060	.109	—	—
X_2 on X_3	.484	.459	—	.025	—
X_2 on X_4	.408	.252	.138	.018	—
X_3 on X_4	.436	.300	—	.136	—

First, in regard to Table 2.3, there are no joint associations, since there is only one variable that serves only as a predictor variable (X_1). No variable can have an indirect effect on the variable it immediately precedes; thus X_1 on X_2 , X_2 on X_3 , and X_3 on X_4 have no indirect effects. In regard to X_1 on X_3 , $r_{13} = p_{31} + p_{32}p_{21}$; $p_{31} = .152$ is the direct effect and $p_{32}p_{21} = (.459)(.163) = .075$ is the indirect effect of X_1 on X_3 thru X_2 . For X_1 on X_4 ,

$$\begin{aligned}
 r_{14} &= p_{41} + p_{42}p_{21} + p_{43}(p_{31} + p_{32}p_{21}); \\
 &= p_{41} + p_{42}p_{21} + p_{43}p_{31} + p_{43}p_{32}p_{21}; \\
 &= (.060) + (.252)(.163) + (.300)(.152) + (.300)(.459)(.163); \\
 &= .060 + .041 + .046 + .022.
 \end{aligned}$$

For r_{14} , .060 represents the direct effect, .041 is the indirect effect of

X_1 on X_4 thru X_2 , .046 is the indirect effect of X_1 on X_4 thru X_3 and .022 is the indirect effect on X_1 on X_4 thru X_2 and X_3 .

$$\begin{aligned}\text{For } r_{23}, r_{23} &= p_{31}p_{21} + p_{32}, \\ &= (.152)(.163) + .459, \\ &= .025 + .459,\end{aligned}$$

where .459 is the direct effect of X_2 on X_3 , and .025 is due to the combined effect of p_{31} and p_{21} . This latter effect cannot be indirect in regard to X_2 on X_3 , because there are no intermediary variables; because X_1 precedes both X_2 and X_3 , this portion of the correlation is called spurious.

For the effect of X_2 on X_4 ,

$$\begin{aligned}r_{24} &= p_{41}p_{21} + p_{42} + p_{43}p_{31}p_{21} + p_{43}p_{32}, \\ r_{24} &= (.060)(.163) + (.252) + (.300)(.152)(.163) + (.300)(.459), \\ r_{24} &= .010 + .252 + .008 + .138.\end{aligned}$$

For r_{24} , .252 represents the direct effect, .138 represents the indirect effect of X_2 thru X_3 and .010 and .008 represent spurious relationships. Finally, .300 represents the direct effect of X_3 on X_4 with the remaining portion due to spurious effects.

One additional value of path analysis should now be seen; each correlation can be decomposed into a direct causal portion, an indirect causal portion and the spurious effects of the variables involved. It would often be of interest to compare the size of these various components. Regarding the present example, X_2 (Concrete-operational Reasoning Test) has a fairly large indirect effect on X_4 (spatial relations) thru X_3 (Logical Reasoning Test). X_1 (conservation) has a larger indirect effect than a direct effect on X_4 ; this indirect effect may be sufficient to include the variable X_1 in the model even though p_{41} is close to zero.

Assessing the Effects Under the Reduced Recursive Model

The model shown in Figure 2.5 has two deleted paths, p_{31} and p_{41} . See Table 2.4.

Table 2.4

Assessing the Effects in the Reduced Recursive Model					
Variables	Total Association	Direct Effects	Indirect Effects	Spurious	Joint Association
X_1 on X_2	.163	.163	—	—	—
X_1 on X_3	.227	—	.079	—	—
X_1 on X_4	.169	—	.067	—	—
X_2 on X_3	.484	.484	—	—	—
X_2 on X_4	.408	.257	.151	—	—
X_3 on X_4	.436	.312	—	.124	—

Using equations 2.8 and 2.20 thru 2.24 with $p_{31} = p_{41} = 0$ and the remaining path coefficients from Figure 2.5,

$\hat{r}_{12} = p_{21} = .163$, the direct effect of X_1 on X_2 ;

$\hat{r}_{13} = p_{31} + p_{32}p_{21}$

$= 0 + (.484)(.163) = .079$, the indirect effect of X_1 on X_3 ; The actual value for r_{13} , .227, is .148 larger than the just arrived at figure. This difference is due to dropping p_{31} .

$\hat{r}_{14} = p_{41} + p_{42}p_{21} + p_{43}p_{31} + p_{43}p_{32}p_{21}$

$= 0 + (.257)(.163) + (.312)(0) + (.312)(.484)(.163)$

$= 0 + .042 + 0 + .025 = .067$, which represents the indirect effect of

X_1 on X_4 ; $r_{14} - .067 = .102$ is the portion due to deleting p_{31} and p_{41} .

For r_{23} , $\hat{r}_{23} = p_{32} + p_{31}p_{21}$,

$= .484 + 0$, all of which is direct;

$$\begin{aligned} \hat{r}_{24} &= p_{41}p_{21} + p_{42} + p_{43}p_{31}p_{21} + p_{43}p_{32}, \\ &= (0)(.163) + (.257) + (.312)(0)(.163) + (.312)(.484), \\ &= 0 + .257 + 0 + .151; \end{aligned}$$

$p_{42} = .257$ is the direct effect; .151 is the indirect effect of X_2 on X_4 thru X_3 . Finally,

$$\begin{aligned} \hat{r}_{34} &= p_{41}(p_{31} + p_{32}p_{21}) + p_{42}p_{31}p_{21} + p_{42}p_{32} + p_{43} \\ &= 0(0 + (.484)(.163)) + (.257)(0)(.163) + (.257)(.484) + (.312) \\ &= 0 + 0 + .124 + .312. \end{aligned}$$

which represent respectively the spurious effect (.124) and the direct effect (.312).

Relationship of Partitioning the Correlation
Coefficients with Partial and Part Correlation and the Independent
Effect of a Variable

One of the problems of statistics is translating the results of the use of equations to a presentation in English. In common English, what is a "direct effect"? Several authors have attempted to answer in English this question. These renditions in translating statistics has undoubtedly given rise to many misconceptions on the part of unsophisticated readers. Perhaps it is easier to proceed by showing what a direct effect is not to help clear up these misconceptions. It was shown in complete detail in Kerlinger and Pedhazur (1973) and Williams and Klimpel (1975) that a direct effect is not synonymous with "the independent contribution of a predictor." For example, if the effect of variable 3 is considered in relationship to variable 4 from the full recursive model (Figure 2.2; Table 2.3); $R_{4.123}^2 = .24421$; $R_{4.12}^2 = .17711$. Thus, the independent contribution of variable 3 is $R_{4.123}^2 - R_{4.12}^2 = .06710$. It is well established that this value is the square of the part correlation between variable 4 and an adjusted variable 3 such that variables 1 and 2 have been partialled out of variable 3;

$r_{4(3.12)} = .259 = \sqrt{.0671}$. The square of p_{43} is $(.3)^2$ or .09. Thus, neither the direct effect of a variable nor the square of the direct effect is necessarily related to the independent effect of a variable.

This also leaves in some doubt the use of the terminology, "spurious effect". One might suppose that the spurious effect of a variable b on variables a and c would be given by $r_{ac} - r_{ac.b}$ where b might represent one or more prior variables. This relationship does not hold either; $r_{34} = .436$;

$r_{34.12} = .286$. One direct effect is of interest; $p_{32} = r_{23} - p_{31}p_{21}$.

$$p_{32} = r_{23} - (r_{13} - p_{32}r_{12}) r_{12};$$

$$p_{32} = r_{23} - r_{13}r_{12} + p_{32}r_{12}^2;$$

$$p_{32} (1 - r_{12}^2) = r_{23} - r_{13}r_{12};$$

$$p_{32} = \frac{r_{23} - r_{13}r_{12}}{1 - r_{12}^2}, \text{ which is one form of the equation for the part correlation}$$

between variable 2 and variable 3 modified by variable 1 (or $r_{2(3.1)}$).

This finding isn't particularly helpful in decomposing the squared multiple correlation coefficient. To use $r_{2(3.1)}$, an equation could be formed:

$$R_{2.13}^2 = r_{21}^2 + r_{2(3.1)}^2. \quad \text{This formulation presumes that variable 3 is a}$$

predictor of variable 2 which violates our recursive models. What, then, can be said about a direct effect? A direct effect of a variable is that portion of a correlation not due to compound paths in the decomposition of the correlation coefficient. Duncan (1970) addresses many of these same issues from a path analysis perspective (and with different interpretations with those given here) and should be consulted by the interested reader.

Choosing a Solution

Of the models tested, which path diagram should be used as the path diagram to represent the data for reporting purposes? Making this decision is

at least as much an art as it is a science. The decision made by Williams and Brekke was to consider Figure 2.3 with only p_{41} deleted as their model. While this model does include two non-significant paths (p_{21} and p_{31}), their inclusion can be justified by logical reasons and also by considering the effect on the reproduced matrix of correlation coefficients. If the reader prefers another solution, adequate information is presented so that another solution could be investigated.

Identification

There are three possible states of affairs regarding identification in path analysis; an equation (or set of equations) can be just-identified, over-identified, or under-identified. With recursive models, it is not possible for under-identification to take place. While identification takes on a considerably higher interest in nonrecursive models, several points are worth addressing.

For the moment, it is worthwhile to consider a complete nonrecursive model for four variables. See Figure 2.6.

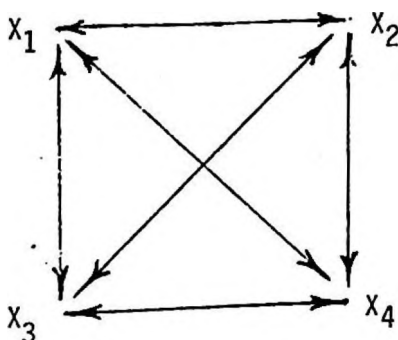


Figure 2.6 A Four Variable Model with all Possible Paths Included

Figure 2.6 would be substantively useless, since it would assume that "everything causes everything else", which is not helpful in any sense. However, it can be seen that there are twelve arrows (equivalently, six double-headed arrows). To include that many paths would cause the system to be under-identified. If judicious choices are made in deleting paths, the system

can become just-identified or over-identified. As it happens, a complete recursive system (see Figure 2.2) is just-identified. That is, all parameters can be estimated, but no tests on the adequacy of the system can be made; by deleting additional paths (see Figure 2.5) thus causing over-identification, tests can be imposed on the system; these tests were given earlier. To see another aspect to over-identification, consider the estimates of r_{14} from two similar equations for Figure 2.5 (with p_{31} and p_{41} deleted). From equation 2.16, $\hat{r}_{14} = .113$; also, $\hat{r}_{14} = .067$ from the use of equation 2.21; $r_{14} = .169$. Thus, not only does over-identification lead to somewhat inaccurate estimates of the correlation coefficients, it can lead to different estimates of the correlation coefficients.

Exogenous and Endogenous Variables

To this point, an effort has been made to present path analysis from a regression framework, using much of the terminology more familiar to the regression formulations. Thus, many of the niceties of path analysis have been omitted. To compensate for this regression emphasis, some of the concepts more familiar to a path analysis view are now considered.

Variables in a path analysis are either exogenous or endogenous. While many authors have described these two terms in considerable detail, one major difference is this: a variable that serves only as an independent (or predictor) variable is considered to be exogenous; exogenous variables are often thought to be pre-existing to a system implied in a path analysis. Accordingly, while an endogenous variable may be an independent variable in part of the system, it must be a dependent variable at least once in the system.

For the example in the present chapter, conservation (X_1) serves as an exogenous variable. Had another variable, such as age been included and had conservation been regressed on age, then conservation would become an endogenous variable.

The usual practice is to consider that all endogenous variables are measured with error, but exogenous variables are measured without error. The usual system in path analysis then includes a residual predictor variable for each endogenous variable, together with an appropriate path coefficient, that accounts for all the remaining variance not attributable to the system. That custom was not followed here; the interested reader is directed to Land (1969) regarding this aspect of path analysis. However, path coefficients for residual variables are easily found as $\sqrt{1 - R^2}$ where R is the multiple correlation of the particular variable with all of its predictor variables. The residual path coefficient is more commonly referred to as the coefficient of alienation in elementary statistics texts. Also, it is usually the case that path diagrams specifically include paths from these residual variables; Figure 2.7 shows the path diagram from Figure 2.5 with p_{13} and p_{14} deleted but with the residual paths included.

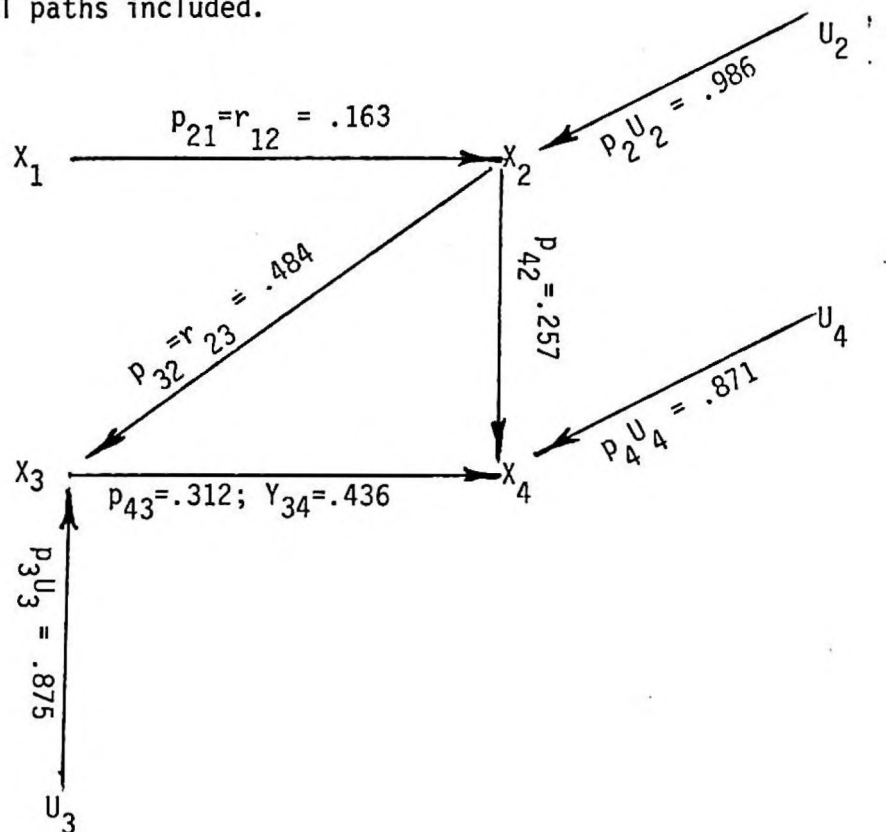


Figure 2.7 Path Diagram with p_{13} and p_{14} Deleted and Residual Paths Included
 X_1 = Conservation, X_2 = Concrete-Operational Reasoning Test, X_3 = Logical Reasoning Test and X_4 = Revised Minnesota Paper Form Board Test (spatial relations)

Specification Error

One major flaw that a researcher may make in path analysis is to incorrectly specify the order of the variables; such a turn of events is called a specification error. As an example, suppose the actual ordering of variables in the data set considered in this chapter is X_1, X_2, X_4, X_3 . That is the formal operations task (X_3) should have had spatial relations (X_4) as a predictor rather than vice-versa. Then many of the estimates of the path coefficients for the system will be irremediably incorrect. Thus, path analysis is not an appropriate tool for those who would prefer to fix the order of the variables on the size of the beta coefficients rather than having decided the order a priori.

Another way that a specification error can occur is by the omission of a variable, either be actual omission or by an assumption of linearity where curvilinearity exists. If the researcher is aware of a variable that should be included that, for whatever reason, is not included, this difficulty can be somewhat alleviated by including indicators of the omitted variable.

In regard to omitting a variable, Woelfle (1978) has pointed out that if a variable X_1 is omitted when it should have been included that the estimate of p_{32} is given as $X_3 = p'_{32} X_2 + e'$ when, in fact, the true model is $X_3 = p_{32} X_2 + p_{31} X_1 + e$. Then, $p'_{32} = p_{32} + p_{31} r_{23}$, so that $p_{31} r_{23}$ represents a specification error. Thus, Woelfle would argue that if a large discrepancy occurs in reproducing the correlation matrix (See Table 2.4), this discrepancy would represent a specification error.

Discussion

Readers who are quite familiar with multiple regression but unfamiliar with path analysis are likely to be perplexed on several points regarding path analysis. Besides the introduction of an almost completely new terminology, words that appear to have commonality between traditional multiple regression analysis and path analysis sometimes refer to different entities. As has been shown here (and previously in Williams and Klimpel, 1975), direct effect has a different meaning from "the independent contribution of a variable". Also "spurious" has a very specific meaning in path analysis that is different from $r_{ac} - r_{ac.b}$.

A justification of sorts could be made for adherents of path analysis in that they could rightly claim that the use of regression addresses only a part of path analysis; path analysis could be seen as a different perspective, much as Bayesian statistics is an entirely fresh approach to understanding data. In this vein, they have coined new words and have attached somewhat different meanings to words already appearing in the statistical literature.

For those who see path analysis as a useful tool to analyze data but prefer to use regression as the means of analysis, there need be no contradiction. As a user (but not necessarily a "true believer") of path analysis, perhaps a suggestion can be made. It appears to me that the terms direct effect, indirect effect and spurious effect are attempts to impose the English language on relationships that are unequivocal in the equation form, but become equivocal by trying to approximate their meaning with English. No analytic loss occurs if the terms are discarded; while this change would be less welcomed by those who prefer verbal presentations, the interests of accuracy would be served.

Chapter Three

Structural Equations

For the researcher not oriented toward the causal model-path analysis-structural equations approach to analyzing data, the addressing of structural equations might appear to be nothing other than a semantic exercise to using simple regression models. Those who have learned the maxim, "You cannot determine causation from correlation", will be skeptical of the meaning addressed to structural models, just as they will likely see path analysis as a violation of their revered maxim.

Before writing down the structural equations, it is worthwhile to examine the reasoning that might lead a researcher to even attempt to find such equations. Briefly, such a researcher (one who wishes to estimate structural equations) has a fairly difficult goal in mind - the finding of a model that has some universality and can be used with a variety of populations. If the structural model were dealing with human behavior, the structural equation should be useful in relating the cause-effect nature of the human behavior rather than simply predict an outcome.

Interestingly, a rather good example of structural equations has already been made in the regression oriented literature (sans the path analysis framework). McNeil (1970) presented a process by which one might, empirically, through the use of regression models, arrive at Newton's law of motion. Those who would be critical of the use of regression might discount McNeil's contribution as "the reinvention of the wheel". Rather, McNeil

was focusing upon the methodology as a way in which so-called "laws" could be empirically found. But these laws are precisely the end goal of those who use structural models. They would hope to find models that have fairly wide application.

Quite clearly, the use of structural equations would most likely be working within a theoretical framework; one such example from an earlier period is the models developed in learning theory by Hull (Hillgard, 1956). Thus, those who might be oriented to approaching data analysis as a data snooper might never appreciate the mind-set of those working from a strong theoretical base.

A Definition of Structural Equations

What is a structural equation (as opposed to a regression equation)? Perhaps it is useful to allow Duncan (1975) to mince a few more words on this issue also:

It does not seem to be possible to give a definition of "structural form" that is other than circular. The structural form of the models is that parameterization-among the various possible ones-in which the coefficients are (relatively) unmixed, invariant, and autonomous. How do you know if you have written a model in its structural form, rather than in some other form? Well, if the coefficients in the model are indeed relatively invariant across populations, somewhat autonomous, and not inseparable mixtures of the coefficients that "really" govern how the world works-then your model is actually in its structural form.

* * *

A strong possibility in any area of research at a given point in time is that there are no structural relations among the variables currently recognized and measured in the area (p. 151-152).

Structural Equations for the Piagetian Data

With some trepidation then, structural models can be expressed for the recursive relationships shown in Chapter Two (with $p_{41}=b_{41.23}=0$).

They are given as

$$X_2 = 21.125 + 1.408X_1 + e_5, \quad (3.1)$$

$$X_3 = 1.574 + 1.236X_1 + .434X_2 + e_6 \quad (3.2)$$

and

$$X_4 = 20.224 + .646X_2 + .831X_3 + e_7. \quad (3.3)$$

The reader will probably recognize these equations as the least squares solutions for the various coefficients. Are the structural coefficients nothing more than simple least squares solutions together with a semantic overload? From the preceding discussion, such criticism would seem unjustified. As will be seen in Chapter Five on nonrecursive relationships, in general, the structural coefficients are not identical with simple (or ordinary) least squares solutions.

Chapter Four

Comparing Results from Two Different Data Sets

A second data set (Appendix B) is considered in this chapter; the data are from Brekke and Williams (1978). Several interesting relationships can be drawn between the two data sets in Appendix A and Appendix B.

First, the same variables are available for analysis in set B as were available in set A. One major difference distinguishes the groups; group A is a group of high school students (ages 15 years to 19 years 7 months) whereas group B consists of junior high school students (ages 12 years 3 months to 16 years 3 months); ages given in the Appendices are in months. Three kinds of analyses are performed in the present chapter. First, an analysis procedure somewhat analogous to homogeneity of regression coefficients is tested given that access is available to both data sets. Second, only data set B is considered to be available, and path coefficients from set A are used as a "restricted model". Finally, assuming that the structural coefficients are available (given in Chapter Three) restricted models are employed on the structural equations.

Preliminary Analysis of the Second Data Set

First, because path models are given in Chapter Two and structural equations are given in Chapter Three, corresponding information on data set B are useful for comparison. Because data set A was collected prior to data set B, p_{41} was excluded (were it included, $p_{41} = .043$; thus, its exclusion can be made on both logical and empirical grounds). Figure 4.1 shows the path model for data set B.

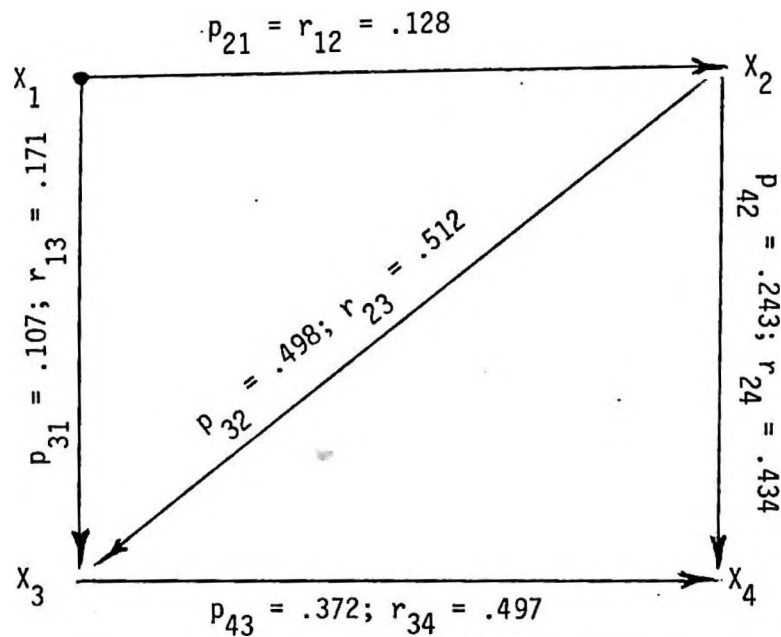


Figure 4.1 Path Diagram with Junior High Subjects, $p_{41} = 0$, ($N = 111$), Data Set B. X_1 = Conservation, X_2 = Concrete-Operational Reasoning Test, X_3 = Logical Reasoning Test and X_4 = Revised Minnesota Paper Form Board Test (spatial relations).

Comparing Figure 4.1 to Figure 2.3, a general remark is that the corresponding path coefficients are in fact remarkably similar; perhaps the relationship of X_1 to X_2 and also of X_1 to X_3 might be seen as being slightly reduced.

The corresponding structural equations for data set B are as follows:

$$X_2 = 21.548 + 1.119X_1 + e_8, \quad (4.1)$$

$$X_3 = .856 + .853X_1 + .455X_2 + e_9 \quad (4.2)$$

and

$$X_4 = 20.961 + .545X_2 + .912X_3 + e_{10}. \quad (4.3)$$

While these coefficients may not seem "unduly" different from those presented in Chapter Three, any conclusions can await a statistical test.

A Preliminary Analysis of the Combined Data Sets

Combining both data sets, analyses like those shown in Figure 4.1 and equations 4.1 - 4.3 are also of interest. Figure 4.2 is a path analysis for the combined data sets.

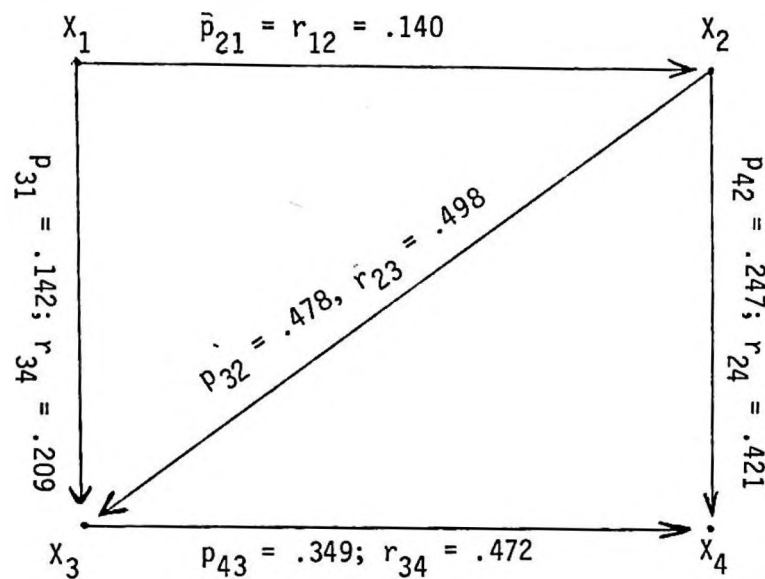


Figure 4.2 Path Diagram for Combined Data Sets ($N = 227$) with $p_{41} = 0$, X_1 = Conservation, X_2 = Concrete-Operational Reasoning Test, X_3 = Logical Reasoning Test and X_4 = Revised Minnesota Paper Form.

As might be expected, the coefficients in Figure 4.2 tend to be intermediate between Figures 2.3 and 4.1. The structural equations for the combined set are as follows:

$$X_2 = 21.394 + 1.196X_1 + e_{11}, \quad (4.4)$$

$$X_3 = 1.174 + 1.125X_1 + .445X_2 + e_{12} \quad (4.5)$$

and-

$$X_4 = 20.678 + .584X_2 + .886X_3 + e_{13} \quad (4.6).$$

While intuitive judgments might be made about the similarity of equations 4.4 - 4.6 to the two earlier sets of structural equations, statistical tests would be more definitive.

Combining Two Data Sets

Path coefficients and structural equations for the combined sets have already been presented; a reasonable question to ask is, "Do the two data sets have homogeneous regression coefficients?" Alternatively, is there any significant loss of information when the groups are combined, other than differences in means? That is, will a single regression line fit both groups, or are separate lines necessary to adequately represent the available information? To test these hypotheses regarding the regression lines, several new variables are useful:

$X_5 = 1$ if a member of set A, 0 if a member of set B;

$X_6 = 1$ if a member of set B, 0 if a member of set A;

$X_7 = X_1 \cdot X_5$;

$X_8 = X_1 \cdot X_6$;

$X_9 = X_2 \cdot X_5$;

$X_{10} = X_2 \cdot X_6$;

$X_{11} = X_3 \cdot X_5$; and

$X_{12} = X_3 \cdot X_6$.

The following models will allow a test of the necessity of separate regression lines for the conservation variable:

$$\text{Model I: } X_2 = b_0 + b_5X_5 + b_7X_7 + b_8X_8 + e_{14}, \quad (4.7)$$

$$\text{Model II: } X_2 = b_0 + b_5X_5 + b_{11}X_{11} + e_{15}, \quad (4.8)$$

$$\text{Model III: } X_2 = b_0 + b_7X_7 + b_8X_8 + e_{16} \quad \text{and} \quad (4.9)$$

$$\text{Model IV: } X_2 = b_0 + b_{11}X_{11} + e_{17}. \quad (4.10)$$

For models I - IV, $R_I^2 = .02069$, $R_{II}^2 = .02042$, $R_{III}^2 = .01988$ and $R_{IV}^2 = .01968$.

Testing Model I against Model II would allow a test of homogeneity of regression; the two groups would be presumed to have the same regression line, but allowed to have different constants. This test is given as

$$F = \frac{(R_I^2 - R_{II}^2)/(3-2)}{(1 - R_I^2)/(227-4)} = .061; \quad (4.11)$$

this is of course non-significant, serving as a partial justification of the non-necessity of more than one regression line.

Testing Model I against Model III would allow a test of differences in means (or constants) without assuming that the two groups fall on the same regression line:

$$F = \frac{(R_I^2 - R_{III}^2)/(3-2)}{(1 - R_I^2)/(227-4)} = .184; \quad (4.12)$$

this non-significant result would be a partial justification that the groups do not differ significantly in their constants. A test of Model I against Model IV would allow a simultaneous test of a need for separate regression lines and difference in constants:

$$F = \frac{(R_I^2 - R_{IV}^2)/(3-1)}{(1 - R_I^2)/(227-4)} = .115. \quad (4.13)$$

This non-significant result would serve as a possible justification that a common regression line and common slope are reasonable in terms of the observed data.

Testing Model II against Model IV would allow a test of a common constant given that the regression lines are homogeneous:

$$F = \frac{(R_{II}^2 - R_{IV}^2)/(2-1)}{(1 - R_{II}^2)/(227 - 3)} = .169. \quad (4.14)$$

Under the assumption that the regression lines are homogenous, the non-significant result would indicate that a common constant is tenable.

Finally, testing Model III against Model IV would allow a test of a common regression line given that the groups are assumed to have a common constant:

$$F = \frac{(R_{III}^2 - R_{IV}^2)/(2-1)}{(1 - R_{III}^2)/(227-3)} = .046. \quad (4.15)$$

Again, the non-significant F value would serve as partial confirmation of a common regression line. It might also be pointed out that the test given by equation 4.11 is equivalent to testing the homogeneity of variance in the analysis of covariance; given that the F in 4.11 is non-significant, the "usual" analysis of covariance is tested by equation 4.14. Models similar to I-IV can be developed for the second stage of path analysis as well:

$$\text{Model V: } X_3 = b_0 + b_5X_5 + b_7X_7 + b_8X_8 + b_9X_9 + b_{10}X_{10} + e_{18}, \quad (4.16)$$

$$\text{Model VI: } X_3 = b_0 + b_5X_5 + b_1X_1 + b_2X_2 + e_{19}, \quad (4.17)$$

$$\text{Model VII: } X_3 = b_0 + b_7X_7 + b_8X_8 + b_9X_9 + b_{10}X_{10} + e_{20} \text{ and} \quad (4.18)$$

$$\text{Model VIII: } X_3 = b_0 + b_1X_1 + b_2X_2 + e_{21}. \quad (4.19)$$

The corresponding R^2 values are:

$$R_V^2 = .27296, R_{VI}^2 = .27238, R_{VII}^2 = .27268 \text{ and } R_{VIII}^2 = .26720.$$

If tests analogous to equations 4.11 to 4.15 are run, no significant findings occur. Other types of tests are conceivable as well. For example, another model that might be of interest is

$$X_3 = b_0 + b_5X_5 + b_7X_7 + b_8X_8 + b_2X_2 + e_{22}. \quad (4.20)$$

Such a model, when compared to Model V, would allow a test of homogeneity on X_2 , given that no other assumptions are made (Perhaps it is possible to test hypotheses for questions that no one would ever ask!).

The third stage of the path analysis would have the following models, similar to Models I-IV and V-VIII:

$$\text{Model IX: } X_4 = b_0 + b_5X_5 + b_9X_9 + b_{10}X_{10} + b_{11}X_{11} + b_{12}X_{12} + e_{23}, \quad (4.21)$$

$$\text{Model X: } X_4 = b_0 + b_5X_5 + b_2X_2 + b_3X_3 + e_{24}, \quad (4.22)$$

$$\text{Model XI: } X_4 = b_0 + b_9X_9 + b_{10}X_{10} + b_{11}X_{11} + b_{12}X_{12} + e_{25} \quad (4.23)$$

and

$$\text{Model XII: } X_4 = b_0 + b_2X_2 + b_3X_3 + e_{26}. \quad (4.24)$$

The corresponding R^2 values are:

$$R_{IX}^2 = .26981, R_X^2 = .26948, R_{XI}^2 = .26976 \text{ and } R_{XII}^2 = .26859.$$

Again, no significant findings occur using tests analogous to those given by equations 4.11 to 4.15; also, tests similar to those suggested by equation 4.20 can be run.

The net effect of the tests run to this point in Chapter Four is that the path diagram with the two sets of subjects is a reasonable way to express the data; using one set of path coefficients and one set of structural coefficients can be done without undue loss of information.

Testing the Path Coefficients from Set A on Set B

Most published research does not allow the luxury of reanalysis with new data sets; if a path analysis has been performed on the data, comparisons of the path coefficients from the two analyses can still be completed. For present purposes, suppose the path diagram from Figure 2.4 is available. Then six variables can be defined:

$$V_1 = z_2 - .128z_1,$$

$$V_2 = z_2 - .163z_1,$$

$$V_3 = z_3 - .107z_1 - .498z_2,$$

$$V_4 = z_3 - .152z_1 - .459z_2,$$

$$V_5 = z_4 - .243z_2 - .372z_3 \text{ and}$$

$$V_6 = z_4 - .252z_2 - .300z_3.$$

V_1 and V_2 express, respectively, the path equation for z_2 from set B and set A. The interest is on the standard deviations (and hence, variances) of $V_1 - V_6$. The squared standard deviations express the error variance for each path equation; V_1 is the least square estimate for the sample data (set B) and V_2 is the least square estimate from the prior data set, and hence serves as a restricted model. Using the methodology of Bottenberg and Ward (1963, p. 47),

$$F = \frac{(q_2 - q_1)/(df_1)}{q_1/(df_2)}, \quad (4.25)$$

where $q_i = (n - 1) s_i^2$.

Also,

$$s_1^2 = .98357, q_1 = 108.193,$$

$$s_2^2 = .98478, q_2 = 108.326,$$

$$s_3^2 = .72696, q_3 = 79.966,$$

$$s_4^2 = .73001, q_4 = 80.301,$$

$$s_5^2 = .70974, q_5 = 78.071, \text{ and}$$

$$s_6^2 = .71269, q_6 = 78.396.$$

For comparing V_1 to V_2 , $F = .134$;

for comparing V_3 to V_4 , $F = .226$;

and, for comparing V_5 to V_6 , $F = .225$.

The conclusion of fitting the prior path coefficients as restrictions on set B is that the fit is not all that bad; the prior path coefficients appear to be well within a reasonable limit of those found in set B.

Testing the Structural Equations from Set A on Set B

If structural equations are available (or if sufficient information is made available to calculate them), these prior structural equations can be imposed upon the new data set; the process is extremely similar to the process given for testing prior path coefficients. Since the error variance for the second set of structural coefficients are readily available, it is necessary only to find the error variance for the imposed set. Three variables can be defined:

$$V_7 = X_2 - 21.125 - 1.408X_1,$$

$$V_8 = X_3 - 1.574 - 1.236X_1 - .434X_2, \text{ and}$$

$$V_9 = X_4 - 20.224 - .464X_2 - .831X_3.$$

V_{10} , V_{11} and V_{12} can be defined to be the similar terms from the structural coefficients given by equations 4.1 - 4.3.

These sum of squared error terms are as follows:

$$q_7 = 1959.906,$$

$$q_8 = 1213.719,$$

$$q_9 = 7108.191,$$

$$q_{10} = 1957.725,$$

$$q_{11} = 1209.464 \text{ and}$$

$$q_{12} = 7092.273.$$

Using equation 4.25, and comparing successively q_7 to q_{10} , q_8 to q_{11} and q_9 to q_{12} , the corresponding F values are

$$F = .060 \text{ (for comparing } q_7 \text{ to } q_{10}),$$

$$F = .126 \text{ (for comparing } q_8 \text{ to } q_{11}) \text{ and}$$

$$F = .080 \text{ (for comparing } q_9 \text{ to } q_{12}).$$

None of these F value begin to approach significance. One major difference between the tests performed on the structural equations and the tests performed on the path coefficients is that the same constant (intercept) is forced upon the second set when using the structural equations.

This inclusion of the constant could cause additional stress on the models; for the present data, the prior constants, as well as the prior regression coefficients, fit reasonably well.

Use of Comparing Path or Structural Models in the Research Literature

Having become familiar with the methodology in the present chapter, one is inclined to wonder, "How often are such models actually tested and reported?" The answer would appear to be, until fairly recently, "Not very often." Specht and Warren (1975) review the comparison of causal models to previous data. For the most part, the literature is methodological rather than being substantive applications. Specht and Warren prefer testing structural models rather than path models (that is

unstandardized coefficients rather than standardized coefficients). On the other hand, because testing structural models can be seen to be a standard regression technique, many substantive applications might escape notice because the author(s) might have used standard regression terminology.

Chapter Five

Nonrecursive Models

To this point, only recursive models have been considered. Non-recursive models, that is, those that contain reciprocal causation, are of special interest. Nonrecursive models might be a meaningful means of analysis for variables that could be seen as changing together over time. The two data sets (sets A and B) include two variables that are excellent candidates for reciprocal causation; the latter two tests could conceivably be viewed as showing reciprocal causation. That is, as people become more proficient at formal operation tasks they may also become better on the spatial relations tasks (this is the approach that was considered in Chapters 1-4); conversely, as a person becomes more proficient in spatial relations tasks, they will also become more proficient at formal operation tasks. Some might view this as just another way to express a correlation coefficient; actually, a reciprocal causation is implied here. That is, a person could study formal operational type materials through such mechanisms as taking coursework in finite mathematics, courses in logic or related courses. One possible expected outcome might be improvement in spatial relations, even though spatial relations have not formally been learned. Again, one might posit that taking coursework in the various geometries should not only be helpful in improving the spatial relations score, but also help improve the formal operations score. Before this model is analyzed, a "complete" model is of some interest.

A Model Containing All Possible Causations

Figure 5.1 shows a diagram wherein all possible paths are included. Comparison of Figure 5.1 with the earlier recursive diagrams would help point out that the recursive models actually involved considerably more assumptions regarding the non-existence of paths than was acknowledged then. Implicit in the recursive models was that all paths from a higher numbered variable to a lower numbered variable were zero. That is, implicit in Chapters 2-4 was that p_{12} , p_{13} , p_{14} , p_{23} , p_{24} and p_{34} were all zero.

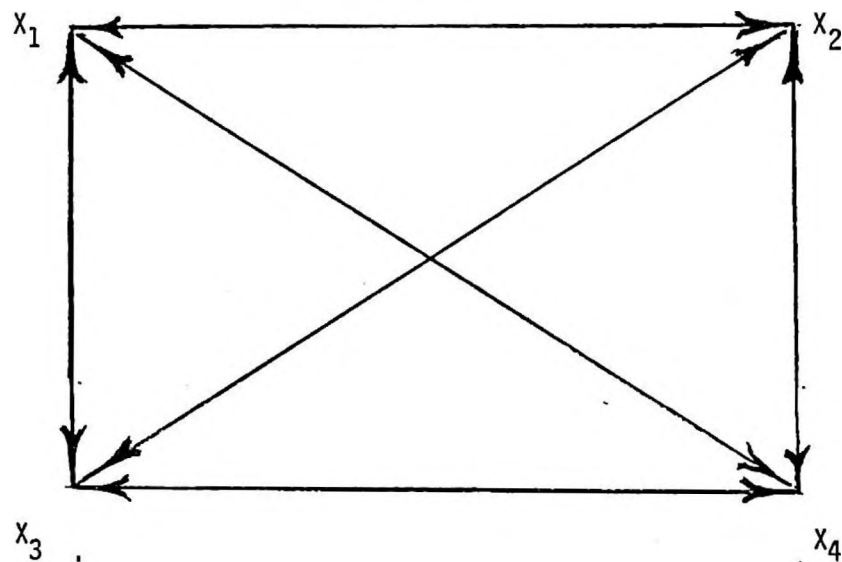


Figure 5.1 Path Diagram With All Possible Paths Included.

The assumptions regarding the non-existence of certain paths brings up again concerns aired earlier in Chapter 2 regarding identification. As will be seen, Figure 5.1 is hopelessly under-identified.

Identification in Nonrecursive Systems

To determine whether over-identification, under-identification or just-identification is present, consider the following diagram:

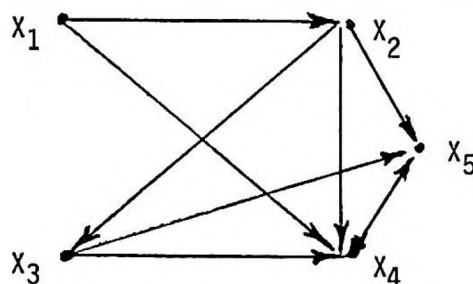


Figure 5.2 A Diagram Demonstrating Three Types of Identification.

To begin, it is simpler to decide upon the identification for that portion of the diagram not involved in reciprocal causation. The equation leading to X_2 is just-identified, as it has a path from its only predecessor variable. On the other hand, the equation leading to X_3 is over-identified, as p_{31} is presumed equal to zero. The variables involved in the reciprocal causation are not quite so easily determined regarding identification. Many references on path analysis include the so-called "order" and "rank" conditions to decide upon the degree of identification in a nonrecursive system. A variant of the "order" condition is Duncan's (1975) "counting" procedure. For a given variable X_i , let G = the number of paths to X_i . Then find H = the number of exogenous variables + all predetermined prior endogenous variables. If $H < G$ the equation is under-identified; if $H = G$ the equation is just-identified and if $H > G$ the equation is over-identified. Strictly speaking, the latter two conditions ($H=G$ or $H>G$) are necessary but not sufficient conditions for identification. For X_4 , $G = 4$, $H = 3$; thus, the equation associated with X_4 is under-identified:

$$X_4 = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_5X_5 + e_{27}. \quad (5.1)$$

However, for X_5 , $G = 3$ and $H = 3$, indicating that the equation $X_5 = b_0 + b_2X_2 + b_3X_3 + b_4X_4 + e_{28}$ (5.2) is "probably" just-identified. The rank procedure tests for linear dependencies involving each equation. While the "rank" condition is not given here in detail (but see Duncan, 1975; Asher, 1976; or Namboodiri, Carter and Blalock, 1975), the equation for X_5 is in fact just-identified.

A Nonrecursive Representation of the Piagetian Tasks and Spatial Relations

Figure 5.3 shows an alternative path diagram for the Piagetian tasks and the spatial relations tests from data set A ($N = 116$). A reciprocal causation is considered regarding formal operations and spatial relations.

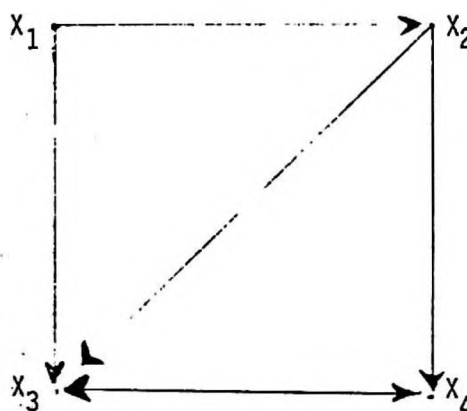


Figure 5.3 Nonrecursive Path Diagram for Piagetian Tasks and Spatial Relations, Set A ($N=116$): X_1 = conservation, X_2 = Concrete-Operational-Reasoning Test, X_3 = Logical Reasoning Test and X_4 = Revised Minnesota Paper Form Board Test (spatial relations).

Now, the degree of identification of the equations for X_2 , X_3 and X_4 as dependent variables is necessary. X_2 is in a simple recursive relationship to X_1 and hence is just-identified (as was true previously). For X_3 , $G = 3$, $H = 2$; thus, the equation associated with X_3 is under-

identified. For $X_4 = G = 2$, $H = 2$, and the equation for X_2 is linearly independent of all other variables, hence, the equation associated with X_4 as a dependent variable is just-identified.

When considering recursive equations, just-identification and over-identification do not cause any special estimation problems in non-recursive models. Under-identification is a much more serious problem; estimation may not take place when under-identification is present. If the problem is to be solved, the under-identification has to be removed. One rather useful way to approach the problem is through the use of two-stage least squares.

Two-Stage Least Squares

Before using two-stage least squares (2SLS), the under-identification must be removed in one of two ways; either existing paths to those variables whose equation is under-identified are to be dropped, or additional exogenous variables can be added; combinations of these two approaches can also be workable. In general, adding exogenous variables is to be preferred over dropping paths, unless the removed paths can be reasonably presumed to be zero.

For data set A, one possible additional exogenous variable is available, age in months (X_a). Unfortunately, X_a fails to be a useful variable for the 2SLS process. To be a useful exogenous variable, high intercorrelations with the various endogenous variables are necessary; X_a does not have this property. If we would proceed with X_a in a 2SLS process, the failure of X_a would lead to collinearity (or near collinearity) among the transformed variables. Thus, we are confronted with the last alternative, the dropping of paths. Here also we encounter an additional problem; we will have to alter our theory to the needs of the 2SLS process

so that estimation may take place. First, we need at least two exogenous variables for 2SLS (otherwise all transformed variables would be simple linear functions of one another). Clearly, p_{21} will have to be dropped so that X_2 may be an exogenous variable. Now either p_{31} or p_{32} will have to be dropped, otherwise X_3 will remain under-identified (p_{43} could be dropped, but of course that eliminates the nonrecursiveness and completely eliminates the problem!). If p_{31} is dropped, X_1 ceases to be an exogenous variable. Thus, for estimation purposes, p_{32} must be dropped; it is a pity, as p_{32} was the largest path coefficient in the models shown in Chapter Two. Figure 5.4 shows the resulting nonrecursive model.

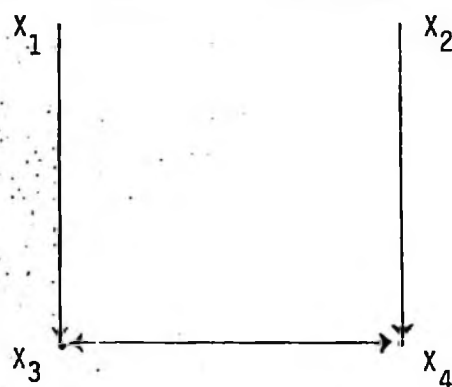


Figure 5.4 Nonrecursive Path Diagram for Set A (N=116):
 X_1 = conservation, X_2 = Concrete-Operational Reasoning Test,
 X_3 = Logical Reasoning Test and X_4 = Revised Minnesota Paper
 Form Board Test (spatial relations).

To accomplish the 2SLS process, the endogenous variables, X_3 and X_4 , are separately regressed on the exogenous variables. The resulting equations are called reduced form equations:

$$X_3 = b_0 + b_1X_1 + b_2X_2 + e_{30} \quad \text{and (5.4)}$$

$$X_4 = b_0 + b_1X_1 + b_2X_2 + e_{31}. \quad (5.5)$$

The reduced form equations constitute the first stage of the 2SLS process. The resulting equations for set A, shown as prediction equations, are

$$\hat{X}_3 = 1.574 + 1.236X_1 + .434X_2$$

and

$$\hat{X}_4 = 21.062 + 2.294X_1 + .983X_2.$$

Now, the second stage of 2SLS can take place. When X_3 is used as the criterion, $\hat{X}_4$ (instead of X_4) is used as a predictor; when X_4 is used as the criterion, $\hat{X}_3$ (instead of X_3) is used as a predictor.

The resulting structural equations are

$$X_3 = -7.732 + .223X_1 + .442\hat{X}_4 + e_{32} \quad (5.6)$$

and

$$X_4 = 18.142 + .178X_2 + 1.855\hat{X}_3 + e_{33} \quad (5.7)$$

Figure 5.5 shows the model with path coefficients.

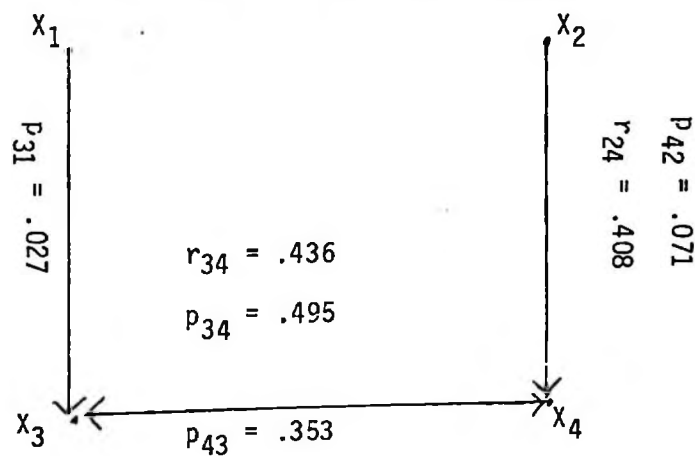


Figure 5.5 Nonrecursive Path Diagram with Path Coefficients Shown (N=116): X_1 = conservation, X_2 = Concrete-Operational Reasoning Test, X_3 = Logical Reasoning Test and X_4 = Revised Minnesota Paper Form Board Test (spatial relations).

It can be seen that the "purifying" or "washing" of the data has improved p_{43} somewhat but allows $p_{34} = .495$ to exceed the correlation between X_3 and X_4 . Were it not for the loss of p_{32} , Figure 5.5 could be a fairly attractive model for interpretive purposes.

Correcting for Bias in the Path Coefficients in 2SLS

Hout (1977) has pointed out that the path coefficients are underestimated for the endogenous variables involved in the reciprocal causation. This bias does not effect the structural equation models given in equations 5.6 and 5.7. The bias is due to the way in which computer programs calculate path (beta) coefficients. In equation 5.6, the path coefficient for $\hat{X}_4$ is found as

$$p_{34} = b_{3.14}(s_{\hat{4}}/s_3) = .442(3.714/3.315) = .495.$$

Hout maintains that the correct value would be found as

$$p_{34} = b_{3.14}(s_{\hat{4}}/s_{\hat{3}}) = .442(3.714/1.679) = .978.$$

Alternatively, p_{34} can be found as the path coefficient found without correction (.495) divided by $R_{3.12}$ (.50675), which is within rounding error of .978. It is necessary to adjust p_{43} for the same reason;

$$p_{43} = b_{4.13}(s_{\hat{3}}/s_{\hat{4}}) = 1.856(1.679/3.714) = .839.$$

Clearly, the dramatic change in the path coefficients of the variables involved in the reciprocal causation would tend to give different interpretations to the data. The corrected values for p_{43} and p_{34} both greatly exceed the correlation between X_3 and X_4 ; $r_{34} = .436$. One might conjecture that the probable inflation due to the correction process would sometimes lead an unwary researcher to allow for a reciprocal

causation because the costs involved are low and the payoff is high. Perhaps Hout is on sound ground in insisting on correcting the resulting path coefficients; the outcomes may leave several readers of the substantive interpretations of corrected path coefficients in a 2SLS process unconvinced of the merits of the correcting process in specific and the 2SLS process in general.

Why is the "Washing" of Data Necessary?

For the nonrecursive relationships, it was necessary to use "washed" variables rather than the variables themselves. Were the variables used themselves, then it is quite likely that, for example, X_3 will be correlated with the error term (sometimes called disturbance) in X_4 . By using $\hat{X}_3$ instead of X_3 when generating the equation for X_4 , this correlation is removed.

Whenever both X depends on Y and Y depends on X , we cannot assume that the error terms from the two equations (for X and Y) are uncorrelated. The use of 2SLS as outlined herein allows the analysis to proceed under the usual assumptions.

So that the assumptions (and their possible violations) can be clearly shown, it is worthwhile to consider finding the residuals for the various equations dealing with data set A. For the fully recursive model wherein no paths are deleted, the error terms for predicting X_2 , X_3 and X_4 were respectively e_1 , e_2 and e_3 . All intercorrelations of these error terms were 0 ($r = 0$). When p_{41} was deleted (e_4), the resulting correlations are $r_{e_1e_4} = -.011$ and $r_{e_2e_4} = -.012$.

In the nonrecursive case, some of the predictor variables will be correlated with the error term; the use of 2SLS removes this part of the problem. Interestingly, even with the use of 2SLS the error terms themselves will tend to be correlated; $re_{32}e_{33} = .286$. Perhaps a few more intercorrelations would be of interest:

$$re_{30}e_{32} = 1 \text{ and } re_{31}e_{33} = 1.$$

Thus, the significance of the coefficient for the transformed variable as a predictor of the other variable in the nonrecursive relationship was already tested in the first stage of the 2SLS process. This rather interesting finding points out the importance of the exogenous variables. To illustrate, the significance of p_{43} was known before X_3 was regressed on X_1 and X_2 . The test for p_{43} was already given by the test for b_1 in equation 5.5. Similarly, the significance of p_{34} is given by the test of b_2 in equation 5.4. One might argue that 2SLS appears to have "thrown out the baby and saved the wash." This is particularly so when it is considered that the model was changed to accomodate the statistical needs of the model; that is, p_{32} and p_{21} were dropped, losing some of the richness of the path analysis process. It should be pointed out that Hout (1977) maintains that inappropriate error variances are found by the just described 2SLS process; he gives rather complex equations to correct the error variances. While these data should not be construed to mean that 2SLS should be avoided as an analytic technique, it should be clear that a considerable number of pitfalls are involved with the technique. Further, the use of nonre-

cursive models should not be undertaken if the researcher is using the nonrecursiveness to avoid a specification error. Actually, such usage is a specification error. Thus, nonrecursiveness should be entertained as a model only if the underlying theory specifically predicts mutual causation.

Chapter Six

Examples of Path Analysis in Educational Research

Introduction

The choice of reporting examples of any phenomenon is likely to be idiosyncratic. If the entire population is not reported, or if random sampling is not employed to choose the examples, then the selection process is at the discretion of the selector; so it is in the present case. The educational research literature is, despite the intentions of the various abstracting systems (notably ERIC), too diverse to be fully documented. Part of the difficulty (if difficulty is indeed a correct interpretation) is that educational research is perhaps the least bounded of the social and/or behavioral sciences; no core set of a few journals is likely to include research that will be seen as being "most important" to the discipline; some articles by those identifiable as educational researchers will appear in what might seem extremely unlikely sources and yet be seen as important additions to the research literature. Beyond this, important implications to educational research can readily be produced by researchers who manifestly are not identifiable as educational researchers.

Having given an appropriate caveat, the present selections can be described. First, the elaborate analysis of Jencks et al. (1972) is considered. Jencks reanalyzed Coleman et al.'s (1966) data on the

equality of educational opportunity and threw in a reassessment of Jensen's (1969) controversial article for good measure. Perhaps it was fore-ordained that Jencks would also come to very controversial conclusions, but not before his analysis became abstruse; one might guess only a small portion of those who read Jencks' book actually understood the complex manipulations. Additionally, several examples of path analysis are taken from retrievable sources, either the archival journals or the ERIC system.

Jencks' Inequality

Briefly, Jencks used several available data sources, including the Coleman et al. (1966) report, Project Talent and the 1960 census data and arrived at a somewhat controversial conclusion; the most important variable in economic success is luck. There is no need to again review Jencks' conclusions here; in addition to the several reviews that have been made, including Levine and Bane (1975), several papers appeared in the Harvard Educational Review and were republished as Perspectives on Inequality (1973).

The present interest is in the various path models used by Jencks. They are perhaps as complicated as any that have appeared in the non-technical literature. Several path models with up to 13 variables are shown to measure intergenerational mobility. Jencks also uses path models to investigate the various values for h^2 (the so-called heritability index). Using Jencks' data, but using a much simpler approach, Li (1975) shows an estimate of $h^2 = .61$; Li maintains that Jencks' analysis, despite its complications, fails to yield any meaningful estimates. Li

also considered the importance of an unobserved variable; consider the following diagram.

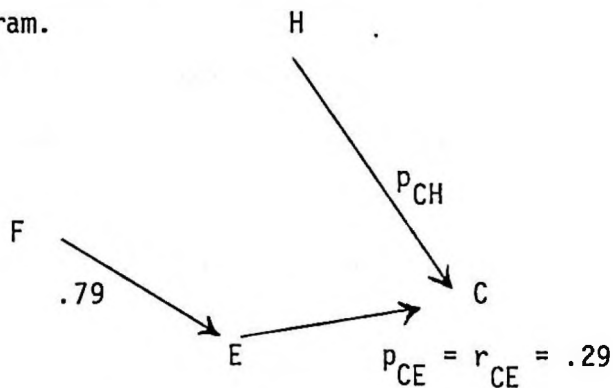


Figure 6.1 Path Diagram for Li's Example: H = unobserved heredity variable; C = child's IQ score; E = a measure of environment; F = midparent (for foster parents) IQ score (Data from Burks, 1928).

Not shown in Figure 6.1 is r_{FC} ; $r_{FC} = .23$. Li derived r_{FE} as $r_{FC} = r_{FE}p_{CE} = .23$; $r_{FE} = .23/.29 = .793$. The effect of heredity was determined by solving $p_{CE}^2 + p_{CH}^2 = 1.00$; thus, $p_{CH} = \sqrt{1 - .29^2} = .957$; this is made under the assumption that heredity is a residual factor that accounts for everything except for the portion caused by the environment. Without detracting from Li's presentation, many would argue that his assumption regarding heredity as the "cause" of the remaining variation is highly untenable. Loehlin, Lindzey and Spuhler (1975) estimate that over a third of the variance in intelligence is due to the joint influence of heredity and environment. Additionally, error variance due to lack of reliability artificially increases the estimate for the effect of heredity using Li's approach. For additional criticism of Jencks' path models, see Taylor (1973) and Asbury (1973). Jencks' (1973, 1974) rebuttals may also be of interest.

An interesting review of Inequality is made by Hauser and Dickinson

(1974). They reanalyze a 13 variable model that contains over-identification. By making different assumptions than Jencks, they arrive at a model with less disturbance between the actual correlations and those implied by the path model. They end by calling the analysis in Jencks' Appendix B as "technically sloppy". Further, the appendix "... gives no evidence for the textual claims that the heritability of IQ is unimportant in socioeconomic achievement". (p. 168).

Other Examples of Path Analysis in Educational Research

Picou et al. (1977) showed, in a six variable analysis, that several hypothesized paths differed significantly for black and white students when academic self concept was used as the criterion. Their methodology was similar to that employed in Chapter Two. Unfortunately, when as many as six variables are used, even in a recursive model, the number of path coefficients (and perhaps also correlation coefficients) may be beyond the ability of many readers to digest. Unfortunately for the consumers of research, six variables in a path analysis is fewer than usual, rather than more than usual. This may not be problematical for the fellow (sister) researcher, but decision makers rarely tend to be attuned to current research methodologies. Thus, decisions may be implemented on extremely shaky findings, but high in the currency of public popularity.

What might be called a "purely educational" application (vis-a-vis either sociological or sociology of education application) was conducted by Madaus et al. (1973). Using the six major taxonomic levels of Bloom's (1956) taxonomy, a hierarchical structure is empirically tested.

Gimmel (1974) used path analysis for determining course achievement. A recursive model was used deleting non-significant paths. Not surprisingly, Gimmel found that past performance had the highest direct effect. Sex was found to be an important variable, but also showed a strong indirect effect through past achievement. While an obvious interaction effect was taking place between sex and past achievement, Gimmel did not place any interpretation on this interaction.

In an attempt to compare the efficacy of both path analysis and a causal analysis approach, Braungart (1975) investigated university students regarding family status, socialization and student politics. His apparent orientation was methodological rather than substantive; he concluded that both methods could serve as reliable checks on one another. One might comment that these results were to be expected; replication would serve as a more serious test of the adequacies of the two approaches.

In a methodologically oriented paper, Feldman (1971) shows how path analysis could be used in researching various college environments. Schmitt, Stone and Rabinowitz (1977) show, in a cross-lagged experiment with three college classes, that quite different interpretations occur when the data is pooled from the three classes and when the classes are individually analyzed, using student ratings and course achievement as the variables. The issue of student ratings would appear of considerable complexity.

Two symposia have focused on path analysis-causal analysis. The first, published in Goldberger and Duncan (1973), is mainly methodological. One article of educational interest was completed by Joreskog (1973). Scores were reported on the Scholastic Aptitude Test for both the mathematics and science sections for students studied on a longitudinal basis in the fifth, seventh, ninth and eleventh grades. One interesting outcome is that science scores become increasingly less dependent on mathematics scores at the higher grades. Joreskog interprets this to be the result of science courses in the lower grades being based on logical reasoning, whereas in the higher grades memorization of facts becomes more important.

The second symposium is published in Sewell, Hauser and Featherman (1976) and is focused on measuring school achievement from an interdisciplinary (but sociologically based) viewpoint. Sewell and Hauser (1976) show that

the effect of measured ability on educational attainment is not merely a reflection of socioeconomic status, obviously conflicting with the point of view of many social scientists. T. Williams (1976) explored the possibility of a triple advantage model, i.e., children from high status backgrounds benefit genetically, have more favorable environments and interact more effectively with their environment. A surprising finding is that the least important part of the triple advantage is the family environment. Featherman and Carter (1976) found that the longer a person takes to get a college degree (in continuous attendance) the higher their social status is likely to be in their first position upon graduation. Wiley (1976) found that the quantity of schooling (in days attendance per year) was positively related to achievement. An extensive application of two-stage least squares was made by Levin (1976) in comparing the importance of family background by ordinary least squares and two-stage least squares; the effect of family background is considerably reduced in the two-stage least squares process. Finally, one of the methodological conclusions of Hauser, Sewell and Alwin (1976) is that differences among high schools (in Wisconsin) is quite small, and that research should focus on issues other than school-to-school variations in aspirations and achievement.

Summary

Path analysis is seen as a potentially powerful technique to be used in educational research. Most actual applications in educational research have been sociologically oriented, reflecting the prior usage of path analysis in sociology. Educational researchers who might otherwise become familiar with path analysis from a sociological viewpoint are encouraged to read a source such as Li (1975) which addresses many of the same concerns but from a geneticist viewpoint; in that path analysis had its origin in genetics, Li is well worth the reading.

Chapter Seven

Other Applications of Path Analysis

The applications of path analysis given in the previous six chapters is by no means exhaustive; indeed, an exhaustive survey is impossible. The bounds of path analysis are essentially only defined in so far as users of the concept apply paths to some phenomena. In that sense, path analysis is not bounded by multiple linear regression, but easily moves into multivariate considerations. Four additional general areas of application of path analysis will be briefly considered in this chapter: (1) panel studies; (2) the concept of unmeasured variables, including factor analysis; (3) canonical analysis, and (4) time series analysis.

Panel Studies

Description of the panel studies, particularly the two wave, two variable (2W2V) models are quite common in the sociological literature (see Heise 1970, Duncan 1972, Kenny 1973, Hannan and Young 1977 and Wheaton et al. 1977). Basically, a panel study involves measuring a set of y variables at t points in time on the same subjects. In practice, few panel studies are reported that are not either a 2W2V, three wave, two variable (3W2V) or two wave, multiple variable (2WMV) application. The use of panel studies was thought to be a way in which reciprocal causation problems could be resolved, but Duncan (1972) has pointed out the limitations of such claims. Figure 7.1 shows a 2W2V diagram.

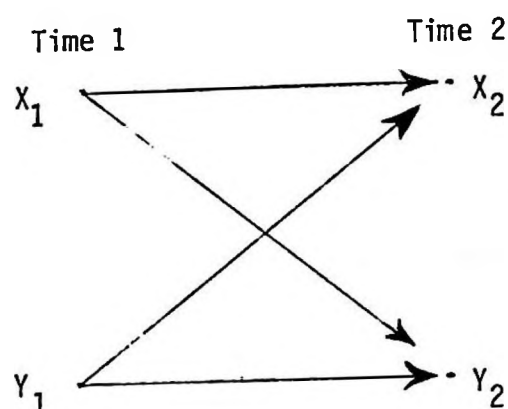


Figure 7.1 A 2W2V Model.

The correlations among the variables have specific names: $r_{X_1X_2}$ and $r_{Y_1Y_2}$ are termed autocorrelations since they involve measuring the same variables at different points in time; $r_{X_1Y_1}$ and $r_{X_2Y_2}$ are termed synchronous correlations since they involve relationships at a given point in time; $r_{X_1Y_2}$ and $r_{X_2Y_1}$ are termed cross-lagged correlations since they involve lagging the variables over time and crossing them.

When panel studies were first envisioned, it was thought that the larger cross-lagged correlations would help determine causation between the two variables. Later it was thought that the largest path coefficient between the two cross-lagged variables would still allow this inference. However, the effect of an unobserved variable could well affect the system in such a way to alter this outcome also. Another possibility is that the time intervals between measurements may not be sufficient to give the variables a chance to have their full effect. As an example of this latter problem, suppose a select group of individuals are followed throughout their young adulthood. All are interested in electronics and pursue careers in some phase of this field. Suppose that the X variables are years of education and the Y variables are income since age 18 (corrected to dollars

in the year age 18 occurs); measurements in terms of the two variables are made at two year intervals to age 30. Those individuals who enter an apprenticeship program immediately upon graduation from high school are likely to have the least amount of education and perhaps the most aggregated income at age 30. Individuals who pursued a graduate degree program, particularly the doctorate, will, by age 30, tend to have a smaller aggregate income. One might infer from this that education has a negative impact on income. If the model is further complicated by using reality, that is the vagaries of the economic system together with demographic change are included in the model, the inference becomes even more difficult to make. That is, the payoff matrix will differ for people in different age cohorts. Undoubtedly, the outcomes of differential education will be substantially different for those who attained the age of 18 in the years 1958, 1968 and 1978 respectively.

Unmeasured Variables

Concern for unmeasured variables has been frequently made in the path analytic literature. The example from Li (1975) shown in Chapter Six of the present monograph regarding the use of heredity as an unmeasured variable to predict the child's intelligence score is one such type of use; a second type of use of unmeasured variables concerns factor analysis. Figure 7.2 gives one possible approach to considering the use of factor analysis.

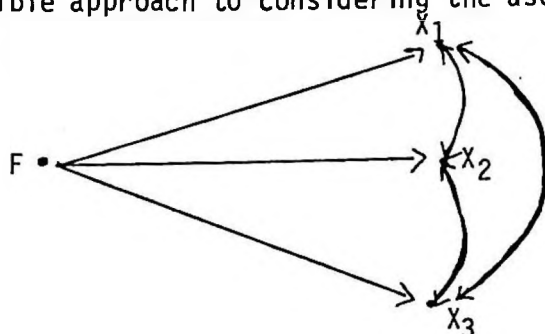


Figure 7.2 Prototype of Using an Unobserved Variable with Three Measured Variables.

Now, an interesting approach can be taken; we might either assume that a single (or perhaps more than one) factor accounts for the intercorrelations among a set of variables and that the factor F is causal to the various X_i variables, or we might alternately assume the X_i variables are causal to the unobserved factor F . In either case, we could attempt to find a composite variable F through factor analysis. While the final path coefficients in the two different approaches would differ, the factor analytic process would remain the same.

Allen (1974) presents compelling reasons for the use of Rao's (1955) canonical factor analytic technique, particularly if there is interest in estimating factor scores. Canonical factors are unique (not always the case for several factor analytic techniques) because the canonical correlation interpretation of the factors guarantees that the canonical factors are maximally related to the observed variables.

When using the canonical factor regression method (see also Harmon 1967, p. 363) to generate composite factor scores, several useful outcomes are available (see Allen 1974, p. 61). The squared linear weight for each X_i indicates the direct contribution of that variable to the variance in the unobserved first common factor score. Also, the reliability coefficient is the squared multiple correlation coefficient between the first common factor and the observed (X_i) variables.

Canonical Analysis

Canonical analysis has not often been employed in path analysis; Hauser's (1973) attempt to disaggregate educational attainment is a pointed exception. Mayer and Younger (1974) also consider the canonical

relationship. Hence what follows is speculative and subject to revision. If a variable in a path analysis is replaced by a set of variables, we would still be interested in investigating the relationships among the variables. To put the question in perspective, see Figure 7.3.

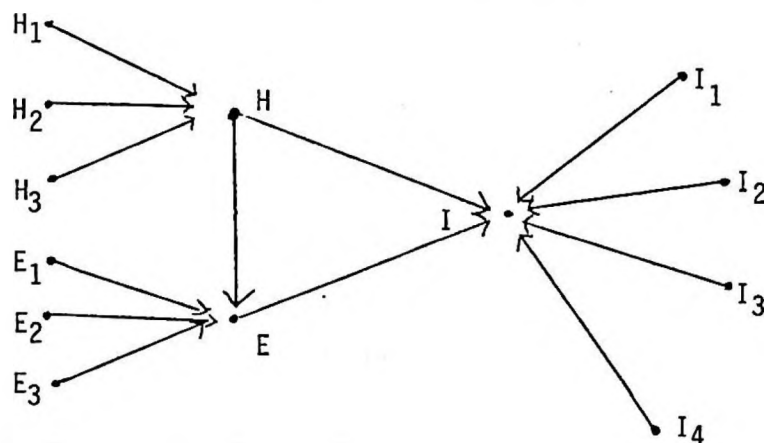


Figure 7.3 An Example of Canonical Path Analysis with Three Sets of Variables.

In Figure 7.3, H_1 , H_2 and H_3 are indicator variables of an unmeasured variable H ; similarly, E_1 , E_2 and E_3 are indicators of E and I_1 , I_2 , I_3 and I_4 are indicators of I . The problem is that the exact procedure to follow is not well defined by the paths shown in Figure 7.3. One strategy would be to factor analyze each set (probably by Rao's procedure) and use at least the first factor (actually, the factor scores on the first factor) for each of the three sets of data. Alternatively, set H and set E might first be used in a canonical analysis. The resulting canonical variable(s) could then be used as predictors of $I_1 - I_4$ in a second canonical analysis. A third strategy would be to complete two canonical analyses; first set H and set E are related, and then sets H and E , using the original variables, could be combined into a single set and related with I . Other strategies are possible also. Canonical analyses could be performed respectively for: (1) set H with set E ; (2) set H with set I ; (3) set E with set I ; and (4) sets H and E with set I . In any case, interpretations are likely to be faulted by other researchers on various grounds, including a predilection

for an analysis procedure not used by the original researcher.

Time Series

The analysis of time series would seem to be a natural area of application of the causal model approach; because of the nature of serial correlations, such has not been the case. Perhaps the major limitation has been the non-familiarity of social researchers with the Box and Jenkins (1970) methodology and the availability of appropriate software programs for multiple linear regression. Hibbs (1974, 1977) has helped bridge the gap between the Box-Jenkins approach and the use of structural equations; he indicates the major difference between the two approaches when confronted with time series data is that the structural equations approach additionally involves specification and estimation of intervention effects in the context of the specified structural models.

Recent explications of the Box-Jenkins approach are given by Nelson (1973) and Glass, Willson and Gottman (1975). Other sociological uses have been shown by Doreian and Hummon (1977) and Burt and Lin (1977).

Path Analysis --- An Unbounded Arena

Hopefully, one outcome of this monograph would be that potential users of path analysis are discouraged from trying to put blinders on path analysis and try to implement a three semester hour course in "all there is to know about path analysis" --- surely they would be destined to failure. In a sense, path analysis is more a systematic way of looking at data rather than being a recipe book for getting answers to questions. In that sense,

path analysis - structural equations is as large as the body of study of the entire area of statistics. Thus, while it might be helpful to have courses devoted to path analysis, it is less than clear where precisely the right point might be to study such courses. Perhaps one or more courses in path analysis might supplant (or partially supplant) such courses as "Research Methods in Educational Research" or similarly titled courses.

APPENDIX A

ID	AGE	CONSERVER	CONCRETE OPERATIONS	FORMAL OPERATIONS	SPATIAL RELATIONS
1	186	1	23	13	49
2	189	1	23	11	55
3	189	1	26	12	52
4	189	1	27	10	25
5	190	1	26	16	48
6	190	1	14	8	39
7	191	1	22	10	47
8	191	1	21	15	48
9	191	1	20	9	43
10	180	1	19	13	41
11	187	1	24	17	43
12	189	1	25	10	50
13	190	1	19	11	52
14	190	1	23	12	56
15	191	1	25	20	54
16	191	1	23	15	53
17	187	0	19	8	47
18	192	1	25	11	51
19	194	1	26	20	48
20	196	1	23	12	43
21	196	1	23	10	46
22	197	1	22	13	43
23	198	1	23	10	25
24	198	1	27	11	45
25	199	1	23	19	48
26	199	1	21	7	25
27	200	1	27	17	58
28	200	1	27	14	59
29	200	1	28	9	44
30	201	1	18	10	37
31	201	1	27	15	44
32	203	1	29	15	50
33	192	0	23	16	53
34	192	0	11	7	28
35	195	0	23	13	36
36	197	0	28	12	50
37	198	0	24	9	36
38	200	0	26	13	44
39	200	0	14	8	34
40	192	1	19	9	50
41	193	1	25	13	42
42	193	1	18	13	44
43	193	1	22	12	37
44	195	1	27	14	29

APPENDIX A

ID	AGE	CONSERVER	CONCRETE OPERATIONS	FORMAL OPERATIONS	SPATIAL RELATIONS
45	195	1	26	15	58
46	196	1	22	9	51
47	196	1	24	14	43
48	197	1	22	14	39
49	197	1	20	10	28
50	198	1	16	7	43
51	200	1	23	6	48
52	201	1	25	16	51
53	202	1	16	9	33
54	202	1	15	13	38
55	202	1	20	9	51
56	192	0	16	11	52
57	197	0	24	16	57
58	203	0	23	16	43
59	204	1	22	12	51
60	204	1	23	11	39
61	207	1	21	15	54
62	209	1	18	10	33
63	209	1	26	17	42
64	209	1	25	14	51
65	209	1	20	11	43
66	209	1	20	11	49
67	210	1	23	13	50
68	215	1	22	10	45
69	210	0	26	13	50
70	204	1	22	15	48
71	206	1	24	16	44
72	206	1	23	17	43
73	206	1	26	14	54
74	210	1	22	11	49
75	211	1	19	15	60
76	212	1	22	14	43
77	212	1	25	18	43
78	213	1	21	15	37
79	215	1	24	11	42
80	204	0	23	11	46
81	206	0	17	5	31
82	206	0	27	13	51
83	215	0	22	9	42
84	216	1	18	11	46
85	216	1	17	4	41
86	217	1	24	10	49
87	217	1	23	16	53
88	217	1	27	13	52

APPENDIX A

ID	AGE	CONSERVER	CONCRETE OPERATIONS	FORMAL OPERATIONS	SPATIAL RELATIONS
89	218	1	17	12	50
90	220	1	24	11	53
91	220	1	25	10	36
92	221	1	19	12	33
93	221	1	21	17	55
94	221	1	20	14	27
95	222	1	25	13	60
96	224	1	25	11	49
97	227	1	25	15	46
98	218	0	18	12	34
99	221	0	23	11	45
100	222	0	21	13	31
101	216	1	24	14	32
102	217	1	22	19	58
103	218	1	25	14	46
104	218	1	23	12	36
105	220	1	21	14	48
106	220	1	23	13	56
107	221	1	19	15	48
108	222	1	18	7	26
109	223	1	29	17	57
110	218	0	21	7	32
111	219	0	22	12	56
112	221	0	14	5	20
113	223	0	24	10	55
114	226	0	18	8	31
115	229	1	21	4	48
116	235	1	21	13	45

APPENDIX B

ID	AGE	CONSERVER	CONCRETE OPERATIONS	FORMAL OPERATIONS	SPATIAL RELATIONS
1	188	1	28	16	52
2	188	0	24	14	48
3	180	1	18	13	56
4	178	1	22	12	42
5	185	0	17	13	27
6	177	0	17	9	42
7	191	1	21	10	41
8	185	1	27	21	49
9	168	1	20	17	42
10	181	0	25	12	54
11	177	1	21	11	34
12	182	0	24	13	42
13	195	1	15	11	50
14	180	1	28	16	62
15	184	1	24	18	47
16	184	0	18	14	53
17	180	1	25	17	47
18	172	1	28	20	52
19	187	1	24	14	46
20	186	0	23	11	35
21	190	1	26	9	34
22	180	0	27	16	38
23	188	1	27	12	58
24	180	0	28	14	60
25	157	0	22	6	37
26	163	1	24	8	42
27	158	1	25	17	53
28	159	0	4	5	15
29	173	0	13	7	43
30	161	1	25	14	49
31	163	1	25	12	37
32	155	1	20	10	51
33	156	0	20	9	44
34	183	1	22	9	38
35	163	1	23	10	42
36	155	0	29	11	28
37	159	0	24	6	40
38	160	1	19	12	51
39	168	0	22	10	40
40	160	1	20	5	40
41	167	0	23	10	48
42	163	1	24	12	32
43	149	1	24	9	32
44	161	1	21	12	47

APPENDIX B

ID	AGE	CONSERVER	CONCRETE OPERATIONS	FORMAL OPERATIONS	SPATIAL RELATIONS
45	161	1	20	7	45
46	150	1	25	14	50
47	157	0	20	4	40
48	160	1	20	13	48
49	155	1	18	9	40
50	195	0	4	3	13
51	169	1	25	10	26
52	165	0	26	9	50
53	156	0	21	11	46
54	147	1	17	6	41
55	155	0	21	10	44
56	147	1	20	7	41
57	163	1	26	17	45
58	157	0	26	13	55
59	157	0	22	10	38
60	156	0	17	12	52
61	158	0	25	15	44
62	155	1	26	5	35
63	150	0	25	14	56
64	176	1	24	16	41
65	173	0	15	7	31
66	185	0	21	11	49
67	175	0	26	15	59
68	167	1	26	13	51
69	176	0	21	2	27
70	169	1	27	11	52
71	176	1	26	16	46
72	178	1	26	13	36
73	167	0	24	10	62
74	174	1	21	11	49
75	173	1	22	10	33
76	173	1	23	14	51
77	171	1	23	17	54
78	173	1	21	8	52
79	173	0	27	18	40
80	161	1	24	15	46
81	171	1	22	9	34
82	168	1	27	6	44
83	167	1	24	15	56
84	167	1	23	9	40
85	170	1	20	7	35
86	180	1	21	5	28
87	175	1	27	14	59
88	176	1	18	7	40

APPENDIX B

ID	AGE	CONSERVER	CONCRETE OPERATIONS	FORMAL OPERATIONS	SPATIAL RELATIONS
89	171	1	22	13	56
90	168	1	16	12	37
91	175	1	23	16	30
92	173	1	23	13	57
93	174	0	26	16	40
94	181	0	15	6	37
95	169	0	24	8	45
96	173	0	23	14	50
97	172	1	14	6	29
98	166	0	25	16	43
99	169	1	22	14	53
100	167	0	22	17	41
101	165	1	27	11	43
102	167	1	22	14	59
103	168	1	23	14	55
104	166	1	23	14	38
105	161	1	16	8	36
106	167	1	24	16	42
107	182	0	21	9	24
108	174	1	16	11	47
109	173	0	25	6	44
110	174	0	23	11	37
111	173	1	25	17	51

APPENDIX C
VALUES FOR DUNN'S TEST
(.05 LEVEL)

m \ v	2	3	4	5	6	7	8	9	10	20
5	3.17	3.54	3.81	4.04	4.22	4.38	4.53	4.66	4.78	5.60
7	2.84	3.13	3.34	3.50	3.64	3.76	3.86	3.95	4.03	4.59
10	2.64	2.87	3.04	3.17	3.28	3.37	3.45	3.52	3.58	4.01
12	2.56	2.78	2.94	3.06	3.15	3.24	3.31	3.37	3.43	3.80
15	2.49	2.69	2.84	2.95	3.04	3.11	3.18	3.24	3.29	3.62
20	2.42	2.61	2.75	2.85	2.93	3.00	3.06	3.11	3.16	3.46
24	2.39	2.58	2.70	2.80	2.88	2.94	3.00	3.05	3.09	3.38
30	2.36	2.54	2.66	2.75	2.83	2.89	2.94	2.99	3.03	3.30
40	2.33	2.50	2.62	2.71	2.78	2.84	2.89	2.93	2.97	3.23
60	2.30	2.47	2.58	2.66	2.73	2.79	2.84	2.88	2.92	3.16
120	2.27	2.43	2.54	2.62	2.68	2.74	2.79	2.83	2.86	3.09
∞	2.24	2.39	2.50	2.58	2.64	2.69	2.74	2.77	2.81	3.02

APPENDIX D
VALUES FOR DUNN'S TEST
(.01 LEVEL)

v \ m										
	2	3	4	5	6	7	8	9	10	20
5	4.78	5.25	5.60	5.89	6.15	6.36	6.56	6.70	6.86	8.00
7	4.03	4.36	4.59	4.78	4.95	5.09	5.21	5.31	5.40	6.08
10	3.58	3.83	4.01	4.15	4.27	4.37	4.45	4.53	4.59	5.06
12	3.43	3.65	3.80	3.93	4.04	4.13	4.20	4.26	4.32	4.73
15	3.29	3.48	3.62	3.74	3.82	3.90	3.97	4.02	4.07	4.42
20	3.16	3.33	3.46	3.55	3.63	3.80	3.76	3.80	3.85	4.15
24	3.09	3.26	3.38	3.47	3.54	3.61	3.66	3.70	3.74	4.04
30	3.03	3.19	3.30	3.39	3.46	3.52	3.57	3.61	3.65	3.90
40	2.97	3.12	3.23	3.31	3.38	3.43	3.48	3.51	3.55	3.79
60	2.92	3.06	3.16	3.24	3.30	3.34	3.39	3.42	3.46	3.69
120	2.86	2.99	3.09	3.16	3.22	3.27	3.31	3.34	3.37	3.58
∞	2.81	2.94	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.48

REFERENCES

- Allen, M. P. Construction of composite measures by the canonical-factor-regression method. In Costner, H. L. Sociological methodology, 1973-74. San Francisco: Jossey-Bass, 1974, 51-78.
- Anderson, J. G. and Evans, F. B. Causal models in educational research: Recursive models. American Educational Research Journal, 1974, 11, No. 1, 29-39.
- Ankney, P. and Joyce, L. The development of a Piagetian paper-and-pencil test for assessing concrete-operational reasoning. Doctoral dissertation, University of Northern Colorado, 1974.
- Asbury, C. A. The methodology used in the Jencks report: A critique. Journal of Negro Education, 1973, 42, No. 4, 530-537.
- Asher, H. B. Causal modeling. Beverly Hills: Sage, 1977.
- Bielby, W. T. and Klugel, J. R. Statistical inference and statistical power in applications of the general linear model. In Heise, D. R. (Ed.) Sociological methodology 1977. San Francisco: Jossey-Bass, 1977, 283-312.
- Blalock, H. M. Four-variable causal models and partial correlations. American Journal of Sociology, 1962, 68: 182-194.
- Blalock, H. M. Causal inferences in nonexperimental research. Chapel Hill: University of North Carolina Press, 1964.
- Blalock, H. M. (Ed.), Causal models in the social sciences. Chicago: Aldine, 1971.
- Bloom, B. S. (Ed.) Taxonomy of educational objectives, Handbook I. Cognitive domain. New York: David McKay, 1956.
- Borgatta, E. F. (Ed.) Sociological methodology 1969. San Francisco: Jossey-Bass, 1969.
- Borgatta, E. F. and Bohrnstedt, G. W., (Eds.), Sociological methodology 1970. San Francisco: Jossey-Bass, 1970.
- Bottenberg, R. A. and Ward, J. H. Applied multiple linear regression. Lackland Air Force Base, Texas: Personnel Research Laboratory PRL-TDR-63-6, 1963.
- Box, G. E. P. and Jenkins, G. M. Time series analysis; forecasting and control. San Francisco: Holden-Day, 1970.

- Braungart, R. G. Path analysis and causal analysis: Variations of a multivariate technique of measuring student politics. Research in Higher Education, 1975, 3, No. 1, 87-97.
- Brekke, B. W. and Williams, J. D. Relationships among different stages of Piagetian tasks and spatial relationships with young adolescents, paper presented at the 8th Annual UAP Conference on Piagetian Theory and the Helping Professions, Los Angeles, January, 1978.
- Burks, B. S. The relative influence of nature and nurture upon mental development: A comparative study of foster parent-foster child resemblance and true parent-true child resemblance. Yearbook-National Society for the Study of Education, 1928, 27, Part I, 219-316.
- Burney, G. M. The construction and validation of an objective formal reasoning instrument. Doctoral dissertation, University of Northern Colorado, 1974.
- Burt, R. S. and Lin, N. Network time series from archival records. In Heise, D. R. (Ed.) Sociological methodology 1977, 1977, 224-254.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfield, E. D., and York, R. L. Equality of educational opportunity. Washington: U. S. Dept. HEW, Government Printing Office, 1966.
- Costner, H. L. (Ed.) Sociological methodology 1971. San Francisco: Jossey-Bass, 1971.
- Costner, H. L. (Ed.) Sociological methodology 1972. San Francisco: Jossey-Bass, 1972.
- Costner, H. L. (Ed.) Sociological methodology 1973-74. San Francisco: Jossey-Bass, 1974.
- Darlington, R. B. Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 69, 161-182.
- Doreian, P. and Hummon, N. P. Estimates for differential equation models of social phenomena. In Heise, D. R. (Ed.) Sociological methodology 1977. San Francisco: Jossey-Bass, 1977, 180-208.
- Duncan, O. D. Path analysis: Sociological examples. American Journal of Sociology, 1966, 72, 1-16.
- Duncan, O. D. Partials, partitions, and paths. In Borgatta, E. F., and Bohrnstedt, G. W. Sociological methodology 1970. San Francisco: Jossey-Bass, 1970, 38-47.
- Duncan, O. D. Unmeasured variables in linear models for path analysis. In Costner, H. L. Sociological methodology 1972. San Francisco: Jossey-Bass, 1972, 36-82.

- Duncan, O. D. Introduction to structural equation models. New York: Academic Press, 1975.
- Dunn, O. J. Multiple comparisons among means. Journal of the American Statistical Association, 1961, 56, 52-64.
- Elkind, D. Children and adolescents. 2nd Ed. New York: Oxford Univ. Press, 1974.
- Featherman, D. L. and Carter, T. M. Discontinuities in schooling and the socioeconomic life cycle. In Sewell, W. H., Hauser, R. M. and Featherman, D. L. (Eds.) Schooling and achievement in American society. New York: Academic Press, 1976, 133-160.
- Feldman, K. A. Measuring college environments: Some uses of path analysis. American Educational Research Journal, 1971, 8, No. 1, 51-70.
- Finney, M. Indirect effects in path analysis. Sociological Methods and Research, 1972, 1, 175-86.
- Furth, H. Piaget and knowledge. Englewood Cliffs, N. J.: Prentice-Hall, 1969.
- Furth, H. Piaget for teachers. Englewood Cliffs, N. J.: Prentice-Hall, 1970.
- Glass, G. V., Willson, V. L. and Gottman, J. M. Design and analysis of time-series experiments. Boulder, Colo.: Colorado Associated University Press, 1975.
- Gimmel, D. R. A path model for course achievement. Journal of Experimental Education, 1974, No. 2, 1-7.
- Goldberger, A. S. Econometric theory. New York: Wiley, 1964.
- Goldberger, A. S. and Duncan, O. D. (Eds.). Structural equation models in the social sciences. New York: Academic Press, 1973.
- Goodman, L. A. The relationship between modified and usual multiple-regression approaches to the analysis of dichotomous variables. In Heise, D. R. (Ed.) Sociological methodology 1976. San Francisco: Jossey-Bass, 1975, 83-110.
- Hannan, M. T. and Young, A. A. Estimation in panel models: Results on pooling cross-sections and time series. In Heise, D. R. (Ed.) Sociological methodology 1977. San Francisco: Jossey-Bass, 1977, 52-83.
- Hanushek, E. A. and Jackson, J. E. Statistical methods for social scientists. New York: Academic Press, 1977.

- Harmon, H. H. Modern factor analysis. (Rev. Ed.) Chicago: Univ. of Chicago Press, 1967.
- Hauser, R. M. Disaggregating a social-psychological model of educational attainment. In Goldberg, A. S. and Duncan, O. D. (Eds.) Structural equation models in the social sciences. New York: Seminar Press, 1973, 255-284.
- Hauser, R. M. and Dickinson, P. J. Inequality on occupational status and income. American Educational Research Association Journal, 1974, 11, No. 2, 161-168.
- Hauser, R. M., Sewell, W. H. and Alwin, D. F. High school effects on achievement. In Sewell, W. H., Hauser, R. M. and Featherman, D. L. (Eds.) Schooling and achievement in American society. New York: Academic Press, 1976, 309-341.
- Heise, D. R. Causal inference from panel data. In Borgatta, E. F. and Bohrnstedt, G. W. (Eds.) Sociological methodology 1970. San Francisco: Jossey-Bass, 1970, 3-27.
- Heise, D. R. (Ed.), Sociological methodology 1975. San Francisco: Jossey-Bass, 1974.
- Heise, D. R. Causal analysis. New York: Wiley, 1975a.
- Heise, D. R. (Ed.) Sociological methodology 1976. San Francisco: Jossey-Bass, 1975b.
- Heise, D. R. (Ed.) Sociological methodology 1977. San Francisco: Jossey-Bass, 1977.
- Hibbs, D. A. Problems of statistical estimation and casual inference in time-series regression models. In Costner, H. L. (Ed.) Sociological methodology 1973-74. San Francisco: Jossey-Bass, 1974, 252-308.
- Hibbs, D. A. On analyzing the effects of policy interventions: Box-Jenkins and Box-Tiao versus structural equation models. In Heise, D. R. (Ed.) Sociological methodology 1977. San Francisco: Jossey-Bass, 1977, 137-179.
- Hillgard, E. R. Theories of learning. 2nd Ed. New York: Appleton-Century Crofts, 1956.
- Hout, M. A cautionary note on the use of two-stage least squares. Sociological Methods and Research, 1977, 5, 335-346.
- Jencks, C. The methodology of inequality. Sociology of Education, 1973, 46, No. 4, 451-470.
- Jencks, C. Comments. American Educational Research Journal, 1974, 11, No. 2, 169-176.

- Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B. and Michelson, S. Inequality. New York: Basic Books, 1972.
- Jensen, A. R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123.
- Johnston, J. Econometric methods, 2nd Ed. New York: McGraw-Hill, 1972.
- Joreskog, K. G. A general method for estimating a linear structural equation system. In Goldberger, A. S. and Duncan, O. D. (Eds.) Structural equation models in the social sciences. New York: Seminar Press, 1973, 85-112.
- Kelly, F. J., Beggs, D. L. and McNeil, K. A. Multiple regression approach. Carbondale, Ill.: Southern Illinois University Press, 1969.
- Kenny, D. A. Cross-lagged and synchronous common factors in panel data. In Goldberger, A. S. and Duncan, O. D. (Eds.) Structural equation models in the social sciences, New York: Seminar Press, 1973, 153-165.
- Kerlinger, F. N., and Pedhazur, E. J. Multiple regression in behavioral research. New York: Holt, Rinehart and Winston, 1973.
- Land, K. C. Principles of path analysis. In Borgatta, E. F. (Ed.) Sociological methodology 1969. San Francisco: Jossey-Bass, 1969, 3-37.
- Levin, H. M. A new model of school effectiveness. In Sewell, W. H., Hauser, R. M. and Featherman, D. L. (Eds.) Schooling and achievement in American society. New York: Academic Press, 1976, 267-289.
- Levine, D. M. and Bane, M. J. (Eds.) The Inequality Controversy. New York: Basic Books, 1975.
- Li, C. C. Path analysis. Pacific Grove, Calif.: Boxwood Press, 1975.
- Likert, R. and Quasha, W. H. Revised Minnesota paper form board test manual. New York: The Psychological Corporation, 1970.
- Loehlin, J. C., Lindsey, G. and Spuhler, J. N. Race difference in intelligence. San Francisco: Freeman, 1975.
- McNeil, K. A. Meeting the goals of research with multiple linear regression. Multivariate Behavioral Research, 1970, 5, 375-386.
- McNeil, K. A., Kelly, F. J. and McNeil, J. T. Testing research hypotheses using multiple linear regression. Carbondale, Ill.: Southern Illinois University Press, 1975.
- Madaus, G. F., Woods, E. M. and Nuttall, R. L. A causal model analysis of Bloom's taxonomy. American Educational Research Journal, 1973, 10, No. 4, 253-262.
- Magary, J. F., Poulsen, M. K., Levinson, P. J., Taylor, P. A. (Eds.) Proceedings from the Sixth Conference on Piagetian Theory and the Helping Professions. Los Angeles: Univ. Southern California, 1977.

- Mayer, L. S. and Younger, M. S. Multiple indicators and the relationship between abstract variables. In Heise, D. R. (Ed.) Sociological methodology 1975. San Francisco: Jossey-Bass, 1974, 191-211.
- Namboodiri, N. K., Carter, L. E. and Blalock, H. M. Applied multivariate analysis and experimental designs. New York: McGraw-Hill, 1975.
- Nelson, C. R. Applied time-series analysis. San Francisco: Holden-Day, 1973.
- Perspectives on Inequality. Harvard Educational Review. Reprint Series, No. 8, 1973.
- Piaget, J. The child's conception of physical causality. London: Routledge and Kegan Paul, 1930.
- Piaget, J., Inhelder, B. and Szeminska, A. The child's conception of geometry. New York: Basic Books, 1960.
- Picou, J. S., Cosby, A. G., Curry, E. W. and Wells, R. H. Race and the formation of academic self concept: A causal analysis. The Southern Journal of Educational Research, 1977, 11, No. 2, 57-70.
- Rao, C. R. Estimation and tests of significance in factor analysis. Psychometrika, 1955, 20, 93-112.
- Schëffe, H. A. A method for judging all contrasts in the analysis of variance. Biometrika, 1953, 40, 87-104.
- Schmitt, N., Stone, E. F. and Rabinowitz, S. Single class versus pooled causal correlational analyses of the student rating-student outcome relationship. Paper presented at the American Psychological Association Annual Meeting, San Francisco, September, 1977.
- Schuessler, K. (Ed.) Sociological methodology 1978. San Francisco: Jossey-Bass, 1977.
- Sewell, W. H. and Hauser, R. M. Causes and consequences of higher education: Models of the status attainment process. In Sewell, W. H., Hauser, R. M. and Featherman, D. L. (Eds.) Schooling and achievement in American society. New York: Academic Press, 1976, 9-27.
- Sewell, W. H., Hauser, R. M. and Featherman, D. L. (Eds.) Schooling and achievement in American society. New York: Academic Press, 1976.
- Spaeth, J. L. Path analysis. In Amick, D. J. and Walberg, H. J. (Eds.) Introductory Multivariate Analysis. Berkeley, California: McCutchan Publishing Corp., 1975, 53-89.
- Specht, D. A., and Warren, R. D. Comparing causal models. In Heise, D. R. (Ed.) Sociological methodology 1976. San Francisco: Jossey-Bass, 1975, 46-82.

- Sund, R. B. Piaget for educators. Columbus, Ohio: Charles E. Merrill Publishing Company, 1976.
- Taylor, H. F. Playing the dozens with path analysis: Methodological pitfalls in Jencks et al.--Inequality. Sociology of Education, 1973, 46, No. 4, 433-450.
- Tukey, J. W. The problem of multiple comparisons. Dittoed, Princeton Univ., 1953.
- Ward, J. H. and Jennings, E. Introduction to linear models. Englewood Cliffs, N. J.: Prentice-Hall, 1973.
- Wheaton, B., Muthen, B., Alwin, D. F. and Summers, G. F. Assessing reliability and stability in panel models. In Heise, D. R. (Ed.) Sociological methodology 1977. San Francisco: Jossey-Bass, 1977, 84-136.
- Wiley, D. E. Another hour, another day. In Sewell, W. H., Hauser, R. M. and Featherman, D. L. (Eds.) Schooling and achievement in American society. New York: Academic Press, 1976, 225-265.
- Williams, J. D. A multiple regression approach to multiple comparisons for comparing several treatments with a control. Journal of Experimental Education, 1971, 39, 93-96.
- Williams, J. D. Multiple comparisons in a regression approach. Psychological Reports, 1972, 30, 639-647.
- Williams, J. D. Regression analysis in educational research. New York: MSS Information Corp., 1974a.
- Williams, J. D. A simplified regression formulation of Tukey's test. Journal of Experimental Education, 1974b, 42, No. 4, 80-82.
- Williams, J. D. A regression formulation of Dunn's and Scheffe's test. Multiple Linear Regression Viewpoints, 1975, 6, No. 1, 74-82.
- Williams, J. D. Multiple comparison by multiple linear regression. Multiple Linear Regression Viewpoints, Monograph Series #2, 1976, 7, No. 1.
- Williams, J. D. and Brekke, B. W. Relationships among different stages of Piagetian tasks and spatial relations in adolescents. Journal of Genetic Psychology (in press).
- Williams, J. D. and Klimpel, R. J. Path analysis and causal models as regression techniques. Multiple Linear Viewpoints, 1975, 5, No. 3, 1-20.

- Williams, T. Abilities and environments. In Sewell, W. H., Hauser, R. M. and Featherman, D. L. (Eds.) Schooling and achievement in American society. New York: Academic Press, 1976, 61-101.
- Wold, H. Causality and econometrics. Econometrics, 1954, 22, 162-177.
- Wold, H. and Juneen, L. Demand analysis. New York: Wiley, 1953.
- Wolfle, L. M. An introduction to path analysis. Multiple Linear Regression Viewpoints, 1977a, 8, No. 1, 36-61.
- Wolfle, L. M. Path analysis and causal models as regression techniques: A comment. Multiple Linear Regression Viewpoints, 1977b, 7, No. 2, 33-40.
- Wolfle, L. M. Personal communication, 1978.
- Wright, S. Correlation and causation. Journal of Agricultural Research, 1921, 20, 557-585.
- Wright, S. The method of path coefficients. Annals of Mathematical Statistics, 1934, 5, 161-215.

If you are submitting a research article other than notes or comments, I would like to suggest that you use the following format, as much as possible:

Title

Author and affiliation

Indented abstract (entire manuscript should be single spaced)

Introduction (purpose—short review of literature, etc.)

Method

Results

Discussion (conclusion)

References

All manuscripts should be sent to the editor at the above address. (All manuscripts should be camera-ready copy.)

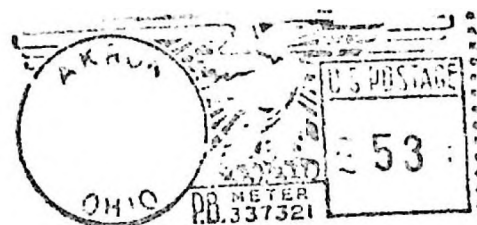
It is the policy of the *multiple linear regression* and of *Viewpoints* to consider for publication articles dealing with the theory and the application of multiple linear regression. Manuscripts should be submitted to the editor as an original, single-spaced typed copy. A cost of \$1 per page should be sent with the submitted paper. Reprints are available to the authors from the editor. Reprints should be ordered at the time the paper is submitted and 20 reprints will cost \$.50 per page of manuscript. Prices may be adjusted as necessary in the future.

"A publication of the *Multiple Linear Regression Special Interest Group* of the American Educational Research Association, published primarily to facilitate communication, authorship, creativity, and exchange of ideas among the members of the group and others in the field. As such it is not sponsored by the American Educational Research Association nor necessarily bound by the Association's regulations.

"Membership in the *Multiple Linear Regression Special Interest Group* is renewed yearly at the time of the American Educational Research Association Convention. Membership dues pay for a subscription to the *Viewpoints* and are divided into two categories: individual=\$3.00; and institutional (libraries and other agencies)=\$12.50. Membership dues and subscription requests should be sent to the Executive Secretary of the MLRSIG."

THE UNIVERSITY OF AKRON
AKRON, OHIO 44325

THIRD-CLASS



157SPAIN
SPARKER STEVEN D.
BEHAVIORAL STUDIES
UNIV OF MO-ST LOUIS
ST LOUIS, MISSOURI 63121

TABLE OF CONTENTS

Chapter		Page
I	Introduction	1
II	Recursive Systems	5
III	Structural Equations	26
IV	Comparing Results from Two Different Data Sets	29
V	Nonrecursive Models	40
VI	Examples of Path Analysis in Educational Research	51
VII	Other Applications of Path Analysis	59
	Tables	66
	References	74