A SAS/IM9 MATRIX Program to Estimate Indirect Effects and their Standard Errors

David P. MacKinnon and Eric Yu-I Wang
University of Southern California

Abstract

Mediated or indirect effects play a major role in social science theories. The existence of these effects indicates that the relationship between two variables depends on an intervening variable. For example, peer influences may mediate the effect of intelligence on adult socioeconomic status. Joreskog and Sorbom (1981) present matrix formulas for the calculation of indirect effects. Recently, Sobel (1982; 1986) derived matrix equations for the calculation of the standard errors of indirect effects. The purpose of this paper is to present a SAS/IM9 matrix program for the calculation of indirect effects and their standard errors. The program shows the versatility and easy use of SAS/IM9 for matrix calculations. Because SAS software is widely available, the program should be accessible to researchers who need to test the statistical significance of indirect effects. Program input can be obtained from any covariance structure analysis program. Output includes estimates of indirect effects and their standard errors. Several examples are presented.

Introduction

Identification and assessment of mediating variables is a central task in many research fields. For example, psychologists study the mediating effect of family support on depression following bereavement. Epidemiologists seek to determine the extent to which cholesterol mediates the effect of relative weight on coronary heart disease. Health psychologists evaluate positive mental outlook as a possible mediator of cancer treatment success.

A recent application of indirect effect estimation is the analysis of experimental prevention studies (McCaul and Glasgow, 1985; MacKinnon, et al., 1988). The purpose of this analysis is to determine how the prevention program is effective. For example, in school-based drug prevention programs, drug use and mediating variables are measured. The mediating variables measure the psychosocial constructs the program was designed to change. Evaluation of the extent to which program effects are mediated by these psychosocial constructs generates information on how the program was effective. In other words, the program changes a mediating variable which in turn reduces drug use. This information is then used to improve future programs by selecting the most effective program components.

Mediation is summarized in the following figure.

![Simple Mediation Model](image)

\[
\text{INDIRECT EFFECT } = ab = \frac{\text{DIRECT EFFECT}}{c}
\]

The effect of the independent variable on the dependent variable can be decomposed into a direct effect \(c\) and an indirect effect through a mediating variable \(ab\). The decomposition of total, direct, and indirect effects for more complicated models has been described (Alwin and Hauser, 1975; Bollen, 1987). Joreskog and Sorbom (1981) present matrix equations for the calculation of these effects.

While the calculation of the value of the indirect effect is relatively straightforward, determination of the standard error of the indirect effect is more complicated. Sobel (1982) used the multivariate delta method (Bishop, Fienberg, and Holland, 1975; Chapter 14) to derive the asymptotic standard error of the indirect or mediating variable effect. Because asymptotic methods were used, the results hold only in large samples. Based on the multivariate delta method, the standard errors for the indirect effects are obtained by pre and postmultiplying the covariance matrix among the estimated parameters by the partial derivatives of the indirect effects. In Sobel (1982), the partial derivatives for the indirect effects and the ratio of indirect to direct effect are presented for the socioeconomic model described in the paper. Later, in Sobel (1986), matrix equations are given for the standard errors of indirect effects. For the simple model presented above, the equations yield the standard error of the indirect effect \(ab\) equal to the square root of \(a^2\sigma_b^2 + b^2\sigma_a^2\).

Recently, Bollen (1987) described a procedure for obtaining the value and standard error of specific indirect effects. His procedure involves setting parameter estimates and elements in matrices for partial derivatives equal to zero, and recalculating effect matrices and standard errors.

1151
Sobel (1982, 1986) proposed that indirect effects and standard errors could be calculated using any matrix programming language. Programs to calculate indirect effects and their standard errors have been written in FORTRAN (Stone, 1985), the Gauss programming language (as part of LINGO, Schonberg, 1987) and most recently as part of LISREL (Joreskog and Sorbom, 1988). The purpose of this paper is to present the SAS/IMatrix code for the calculation of indirect effects and their standard errors. The program includes an option for user-supplied partial derivatives to obtain standard errors for other functions of the model estimates such as the ratio of indirect to the direct effect or the ratio of the indirect effect to the total effect. The program also allows detailed investigation of specific effects as described in Bollen (1987). The program is applied to the examples described in Sobel (1986) and Bollen (1987), and an example from the substantive area of school-based drug prevention.

Equations

The indirect effects are calculated in the following equations (Sobel, 1982).

\[ I_y = (I - B)^{-1} - I - \beta \]
\[ I_y = (I - B)^{-1} - I \]
\[ I_y = I_y (I - B)^{-1} - I_y \]
\[ I_y = I_y (I - B)^{-1} - I_y \]

As described in Sobel (pp. 170-171, 1986) and Bollen (pp. 66-67, 1987) the partial derivatives for each indirect effect matrix are equal to:

\[ \frac{\partial \beta}{\partial \gamma} (I - B)^{-1} \cdot (I - B)^{-1} \cdot \gamma \]
\[ \frac{\partial \beta}{\partial \gamma} (I - B)^{-1} \cdot \gamma \]
\[ \frac{\partial \beta}{\partial \gamma} (I - B)^{-1} \cdot \gamma \]

As the vector of parameter estimates, Vec indicates that the elements in the indirect effect matrices are reshaped into a column vector, and * is the Kronecker or tensor product.

The covariance matrix among the estimates is then pre and post multiplied by the matrix of partial derivatives as shown in the following equation:

\[ N(\partial y/\partial \beta) \cdot \Sigma (\partial y/\partial \beta) \]

Program INDIRECT

A listing of the program code is displayed in Appendix A. The program is divided into INPUT and CALCULATION sections.

INPUT to the program includes the number of different variables, parameter matrices, covariance among the parameter estimates, and a matrix to select partial derivatives. The user inputs the number of \( \eta \) variables (m), the number of \( \xi \) variables (n), the number of y variables (p), and the number of x variables (q). These values are used to construct identity and output matrices in the program.

The \( B, \Gamma, \) and \( \Lambda \) parameter matrices and the covariance among the estimates can be obtained from any covariance structure program output (LINGO, EQS, GODAN, LISREL). The program is based on the Joreskog, Keesling, and Wiley covariance structure model (Joreskog and Sorbom, 1988) so some adjustments need to be made for output from programs such as EQS that are based on the Bentler-Weeks (Bentler and Weeks, 1980) formulation. The \( B \) matrix codes the effects among endogenous variables, \( \Gamma \) codes the effects of exogenous on endogenous variables, and \( \Lambda \) codes the effects of latent endogenous variables on measured variables. The user may supply either (1) the partial derivatives for each indirect effect matrix or (2) the standard errors and the correlation matrix among the estimates and the program will compute the covariance matrix.

The \( V_B, V_T, \) and \( V_T \) matrices select partial derivatives. For most applications these matrices consist of 0s and 1s. For each matrix, the number of columns is equal to the number of parameters estimated. The number of rows is equal to the number of elements in each subscripted matrix. For example, \( V_B \) has \( m^2 \) rows and columns equal to the number of parameters estimated.

Before the CALCULATION section of the program, the user must specify output matrices with the CASE variable. Each case corresponds to a possible use of the program. For example, only the indirect effect among the \( q \) is calculated if \( B \) is the only matrix specified. If \( B, \) and \( \Gamma \) are specified, then only effects from \( \eta \) to \( q, \) and \( \xi \) to \( \eta \) are calculated. If all three matrices are specified, all indirect effects are calculated. If the user supplies his own derivatives, the covariance matrix among the estimates is pre and post multiplied by the user-supplied partial derivatives. The START option is used to modularize program steps based on the value of the CASE variable.
In the **CALCULATION** section of the program, the four different indirect effect matrices and standard errors are calculated. In addition, a sufficiency test for the convergence of \( \beta \) is calculated (Joreskog and Sorbom, 1988; Bentler and Freeman, 1983). The calculations follow the matrix equations presented above.

Program **OUTPUT** consists of the indirect effect matrices and the standard error of indirect effect. For user-supplied partial derivatives, a vector of standard errors is output.

**Examples**

Indirect effects and their standard errors were calculated using program **INDIRECT** for the models in Sobel (1986) and Bollen (1987). The reference for the original data, page numbers for the V matrices, and other studies of the data are presented below. The examples are used to check the program and to illustrate its application. This information should prove useful for those who wish to use the program.

**Socioeconomic Achievement.** In March 1962, a sample of nonblack, nonfarm, males \((N = 3,214)\), aged 35-44, was taken from the experienced civilian labor force in the United States. The data presented in Duncan, Featherman, and Duncan, B. (p. 38; 1972) were also analyzed by Alwin and Hauser (1975). A six variable, three equation model of the achievement process was analyzed in both of Sobel's articles (Sobel, 1982; and Example 1 in Sobel 1986). In both examples, the ordering of the parameters is different from other examples. The \( \Gamma \) elements are presented first, one \( \beta_{11} \) element is seventh, and the last two elements are from the \( \beta \) matrix. Sobel (1982) presents the symbolic form for the partial derivatives, and the paper presented the \( V_{x} \) and \( V_{r} \) matrices on page 174. The analysis revealed a statistically significant direct effect \((ab = 0.1998, se = 0.0564)\) of respondent's education on respondent's income and an indirect effect \((ab = 0.3083, se = 0.214)\) through respondent's occupational status.

**Peer influences.** The data for 329 males and their reported best friend are presented in Hanushek and Jackson (1977, p. 280). The analysis of the correlation matrix for this example appears in the LISREL VI manual (Joreskog, 1985; pp. III-89-III.93). The covariances matrix is analyzed in the second example in Sobel (1986). The model includes seven variables and evaluates the extent to which peer influences mediate the effects of ambition on socioeconomic achievement. In this example, all of the indirect effects and standard errors are computed. The \( V_{x} \) and \( V_{y} \) matrices are presented on page 178 in Sobel (1986). Because all indirect effect matrices are calculated in this example, the program code for this example is listed in the appendix. Sobel's mediation analysis revealed many statistically significant indirect effects. For example, best friend's background has an indirect effect on respondents ambition through best friend's ambition.

**Objective and subjective stratification.** Kluegel, Singleton, and Starnes (1977, p. 603) present the covariance matrix from a sample of 432 white subjects from Gary, Indiana \((N = 432)\). This example was analyzed in Bollen (1987) including detailed analysis of specific effects. When the \( B_{12} \) and \( V_{x1} \) parameter is set to zero, the indirect effect estimates and standard errors on page 65 in Bollen (1987) are obtained. This demonstrates the ease of the **INDIRECT** program to calculate more detailed indirect effect estimates and their standard errors. The \( V_{x} \) and \( V_{y} \) matrices are presented on page 67 in Bollen (1987). The user should read sections in Bollen (1987, pp. 61-62) and Sobel (1986, pp. 180-183) when investigating detailed information on indirect effects.

**Mediation in school-based drug prevention.** A school-level analysis of 42 schools from Kansas City is the final example (MacKinnon et al., 1988). Cigarette smoking and a peer norms mediating variable were measured in the Fall of 1984 and again in the Fall of 1985 after some of the schools received a school-based drug prevention program. Grade and program exposure were the other two variables in the model. One of the hypothesized mediating mechanisms in school-based drug prevention is that changing peer norms regarding drug use will lead to less smoking. The value for the indirect effect was .196 and the standard error of the effect was .117, suggesting reliable mediation of the program effect on cigarette use through changing peer norms. A small simulation study indicated that the standard error based on asymptotic methods was close to the true standard error for a sample size of 42. However, the correctness of asymptotically derived standard errors for indirect effects in small samples is not known.

**Summary**

Program **INDIRECT** successfully returned the correct estimates of indirect effects and their standard errors for several examples. The **SAS**/**IML** language simplifies programming complex matrix equations.
Program Listing

```plaintext
options p=60 l=130;
PROC INL;

*******PROGRAM INDIRECT**********;
*PEER INFLUENCES MODEL;
*ANALYZED IN SIMON 1986;
*DATA IN BRANDIMET AND JACKSON 1977, P. 283;
*SEVENTEEN YEAR OLD SCHOOLBOYS IN LENAWEE COUNTY MICHIGAN;
*NUMBER OF OBSERVATIONS=529;

*** INPUT ****;
*INPUT THE NUMBER OF VARIABLES;
P=4; *P is the number of y variables;
Q=6; *q is the number of x variables;
M=2; *M is the number of eta or endogenous variables;
N=6; *n is the number of eta or exogenous variables;

*PARAMETER MATRICES;
B=0.0000 0.157177 0.187179 0.157199 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

*COVARIANCE MATRIX AMONG THE PARAMETER ESTIMATES;
*IF THE COVARIANCE MATRIX IS INPUT THEN DELETE THE NEXT LINE;
*CASE1* RE * SF ;
**SELECT OUTPUT****;
*INDICATE MATRICES TO BE OUTPUT BY SELECTING A CASE;
* CASE1: IF NN MATRIX;
* CASE2: IF NA YH MATRICES;
* CASE3: IF NA ME MATRICES;
* CASE4: IF NA ME YH MATRICES;
* CASE5: IF USER WILL SUPPLY PARTIAL DERIVATIVES;
* CASE6: IF DETAILED INDIRECT EFFECTS WILL BE EXAMINED;
* IF USER WILL SUPPLY MATRICES THEY SHOULD BE WRITTEN
* IN THE DUS
* SECTION OF THE PROGRAM;
* IF DETAILED INDIRECT EFFECTS WILL BE COMPUTED GO TO
* THE OR SECTION
* OF THE PROGRAM AND ZERO REQUIRED MATRIX ELEMENTS;
CASE6;

*** CALCULATIONS ****;
START DUS;
*SUFFICIENCY TEST FOR STABILITY;
DUS=0;
EVAL=SCALAR(DUS);
INPUT=INP4(IN=0); PRINT "STABILITY INDEX: SUFFICIENT CONDITION FOR CONVERGENCE";
PRINT "THE LARGEST EIGENVALUE SHOULD BE LESS THAN ONE";
PRINT EVAL;
FINISH;

START DUS;
*INPUT USER SUPPLIED PARTIAL DERIVATIVES IN THE DUS MATRIX;
DUS=0;
VUS=DUS*CPUS;
SECOND=SCALAR(vecdeg(YDNAM));
PRINT "STANDARD ERRORS FROM USER SUPPLIED PARTIAL DERIVATIVES";
PRINT VUS;
FINISH;

START GET;
*SELECT PARAMETERS AND ELEMENTS IN THE PARTIAL
*DERIVATIVE MATRIX HERE;
*CASE1: RE; DUS=0, DUS=0, M=0, N=0, ETA=0;
*CASE2: RE; DUS=0, DUS=0, M=0, N=0, ETA=0;
*CASE3: RE; DUS=0, DUS=0, M=0, N=0, ETA=0;
*CASE4: RE; DUS=0, DUS=0, M=0, N=0, ETA=0;

VUS=0 0 0 0 0
0 1 0 0 0
0 0 1 0 0
0 0 0 1 0
0 0 0 0 1
0 0 1 0 0
0 1 0 0 0
0 0 0 0 1
0 0 0 0 1
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0

0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0

0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0

0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0

END;
OUTPUT;
PRINT 'INV MATRIX OF INDIRECT EFFECTS';
PRINT INV;
PRINT 'INV MATRIX STANDARD ERRORS';
PRINT SEINV;
FINISH;

START ME;
*ETA ON KSI;
INV=(INV(INVI*INV)+)*GAMA-GAMA;
ONE=ONE-(INV*GAMA * KSI);
ONE=ONE*ONE;
SEINV=INV*(INV*INV),M,N);
*OUTPUT;
PRINT 'INV MATRIX OF INDIRECT EFFECTS';
PRINT INV;
PRINT 'INV MATRIX STANDARD ERRORS';
PRINT SEINV;
FINISH;

START YE;
*ETA ON Y;
INV=(INV(INVI*INV)+)*LY;
INV=(INV(INVI*INV)+)*IP;
INV=(INV(INVI*INV)+)*LY;
LY=(INV(INVI*INV)+);
SEYE=INV*(INV*INV),M,P);
*OUTPUT;
PRINT 'INV MATRIX OF INDIRECT EFFECTS';
PRINT INV;
PRINT 'INV MATRIX STANDARD ERRORS';
PRINT SEINV;
FINISH;

IF CASE=6 THEN DO;
RUN PET; RUN SUN; RUN YH; RUN YE; RUN WE; RUN WI;
END;
IF CASE=5 THEN DO;
RUN RUS;
END;
IF CASE=4 THEN DO;
RUN SUN; RUN YH; RUN YE; RUN WE; RUN WI;
END;
IF CASE=3 THEN DO;
RUN SUN; RUN RUS; RUN YH; RUN WE; RUN WI;
END;
IF CASE=2 THEN DO;
RUN SUN; RUN YH; RUN WE; RUN WI;
END;
IF CASE=1 THEN DO;
RUN SUN; RUN YH; RUN WE; RUN WI;
END;

References


Authors' Address
David P. MacKinnon Ph.D.
Eric Yu-I Wang M.S.
Institute for Prevention Research
University of Southern California
Department of Preventive Medicine
35 North Lake Avenue, 200
Pasadena, California 91101.