

## Testing Indirect Effects for Lower Level Mediation Models in SAS PROC MIXED

Here we provide syntax for fitting the lower-level mediation model using the MIXED procedure in SAS as well as a sas macro, **IndTest.sas** that performs the computations necessary for evaluating the average indirect and total effects. In addition, a simulated data file is provided, named **sim.sas7bdat**, to which the lower level mediation model can be fit. The population model from which the simulated data were generated has the following form:

$$\begin{aligned}M_{ij} &= d_{Mj} + a_j X_{ij} + e_{Mij} \\ Y_{ij} &= d_{Yj} + b_j M_{ij} + c'_j X_{ij} + e_{Yij}\end{aligned}$$

In the population, the fixed-effects are  $d_{Mj} = d_{Yj} = 0$ ,  $a = b = .6$  and  $c' = .2$  and the variances of the random effects are  $VAR(d_{Mj}) = .6$ ,  $VAR(d_{Yj}) = .4$ ,  $VAR(a_j) = VAR(b_j) = .16$  and  $VAR(c'_j) = .04$ . The covariance between  $a_j$  and  $b_j$  is  $\sigma_{a_j, b_j} = .113$ , and all other random effects are uncorrelated. These values imply that the average indirect and total effects in the population are .473 and .673, respectively. Last, the Level 1 residual variances are  $VAR(e_{Mij}) = .65$  and  $VAR(e_{Yij}) = .45$ . In the simulated data, the number of Level 2 units (indicated by  $j$ ) is  $N = 100$ , the number of observations within each Level 2 unit (indicated by  $i$ ) is  $n_j = 8$ . We recommend saving the simulated data file to a directory on the users computer (e.g., `c:\example\`) to be analyzed using the provided syntax. We now show the syntax for fitting the model to the data using the procedures described in Bauer, Preacher, and Gil (2006).

### SAS code for rearranging data:

The data must first be prepared for the analysis through the creation of a single dependent variable ( $Z$ ) from the values of the mediator ( $M$ ) and the distal outcome ( $Y$ ). Two selection variables are also created, labeled  $S_y$  and  $S_m$ , to indicate when  $Z$  represents  $M$  versus  $Y$ . This rearrangement of the data is shown visually in Table 1 of Bauer, Preacher, and Gil (2006). The

following syntax prints the data for the first two Level 2 units in the original data file, creates a new data file called rearrange from the original data file which contains the restructured data, and prints the rearranged data for the first two Level 2 units. The printed data correspond to the values presented in Table 1 of Bauer, Preacher & Gil (2006).

```
*indicating the directory of the sim.sas7bdat data file, change as needed;
libname example 'c:\example\';

*printing data for the first two Level 2 units (corresponding to top of Table 1);
proc print data=example.sim; where id le 2;
  var id Y M X;
run;

*rrearranging the data to permit fitting the lower level mediation model;
data rearrange; set example.sim;
  Z = Y;          *first assigning Z the value of Y;
  Sy = 1;        *setting Sy selection variable to 1 to indicate Z is Y;
  Sm = 0;        *setting Sm selection variable to 0 to indicate Z is not M;
  dv = 'Y';      *creating variable dv also to differentiate Y from M;
  output;        *outputting the first record for the observation;
  Z = M;          *now assigning Z the value of M;
  Sy = 0;        *setting Sy selection variable to 0 to indicate Z is not Y;
  Sm = 1;        *setting Sm selection variable to 1 to indicate Z is M;
  dv = 'M';      *creating variable dv also to differentiate Y from M;
  output;        *outputting the second record for the observation;
run;

*printing the rearranged data for the first two Level 2 units (corresponding to the
bottom of Table 1);
proc print data=rearrange; where id le 2;
  var id Z Sy Sm M X dv;
run;
```

Using this rearranged data set, we may now proceed to fit the lower level mediation model in the MIXED procedure.

#### SAS code for fitting lower-level mediation model:

The lower-level mediation model can now be fit to the rearranged data. The model is formulated using the selection variables as

$$Z_{ij} = S_{M_{ij}} (d_{M_j} + a_j X_{ij}) + S_{Y_{ij}} (d_{Y_j} + b_j M_{ij} + c'_j X_{ij}) + e_{Z_{ij}}$$

Distributing the selection variables, this gives the equation

$$Z_{ij} = d_{M_j} S_{M_{ij}} + a_j (S_{M_{ij}} X_{ij}) + d_{Y_j} S_{Y_{ij}} + b_j (S_{Y_{ij}} M_{ij}) + c'_j (S_{Y_{ij}} X_{ij}) + e_{Z_{ij}}$$

Each of the coefficients in the model is assumed to have both a fixed effect (representing the average effect) and a random effect (representing differences across Level 2 units in the effects relative to the average). This is specified in the MIXED syntax below:

```
proc mixed data=rearrange asycov covtest noclprint method=reml maxiter=100;
  class dv;
  model Z = Sm Sm*X
          Sy Sy*M Sy*X
          /noint solution ddfm=kr covb;
  random Sm Sm*X Sy Sy*M Sy*X/g gcorr type=un subject=id;
  repeated / group=dv subject=id;
run;
```

The `proc mixed` statement calls the MIXED procedure with the rearranged data set. The option `asycov` requests the asymptotic covariance matrices of the variance component estimates, which are needed for the computation of standard errors for the average indirect and total effects. The option `noclprint` suppresses printing of the levels of any declared classification variables, `method = reml` calls the restricted maximum likelihood estimator for the model, and `maxiter=100` sets the maximum number of iterations for the maximum likelihood estimator to 100.

Next, the statement `class dv` indicates to SAS that the variable `dv` is a classification variable (or nominal variable). If the data were not already sorted on the Level 2 ID variable (`id`) it would also be necessary to declare this variable as a classification variable.

The `model` statement is used to specify the fixed effects of the model. The outcome variable `Z` is indicated to the left of the equal sign and then, to the right of the equal sign, each of

the variables with fixed effects is listed. Following the backslash, the option `noint` removes the general intercept term from the model, the option `solution` requests that the fixed-effect estimates be displayed, and the option `ddfm=kr` requests the Keward-Rogers method of computing degrees of freedom for the tests of the fixed effects, which uses a Satterthwaite approximation for the degrees of freedom and implements the Kackar & Harville (1984) correction to the standard errors. The additional option `covb` produces the covariance matrix of the fixed effect estimates, which is required for computing the standard errors of the average indirect and total effects.

The `random` statement indicates which of the effects in the model have random components (vary in magnitude across Level 2 units). For the present model, each effect is assumed to be random; hence the same variables previously listed in the `model` statement are also listed here. Following the backslash, the option `g` produces the covariance matrix of the random effects and the option `gcorr` produces the correlation matrix of the random effects. The option `type=un` specifies that the covariance matrix of the random effects is to be unrestricted, so that the covariance for each pair of random effects will be estimated freely. Last, the `subject=id` option indicates that the random effects vary over the Level 2 ID variable `id`. If the Level 2 ID variable is numeric and sorted, then it need not be declared as a classification variable in the `class` statement; otherwise it must be so declared. The advantage to using a numeric and sorted Level 2 ID variable is that this speeds the data processing within the MIXED procedure, saving computing time.

Finally, the `repeated` statement is used to access the Level 1 residual covariance matrix. The default assumption is that the Level 1 residuals are independent and homoscedastic. The option `group=dv` changes this assumption so that a unique variance is estimated for the Level 1

residuals for each level of the dv classification variable. As such, using this option allows the Level 1 residual variance to differ when  $Z = Y$  versus  $Z = M$ . The `subject=id` option again indicates the grouping of observations within Level 2 units.

### Output of MIXED procedure

We now show a selection of the output for the data. First, we obtain output for the covariance matrix of the random effects (referred to as **G** in SAS) and the corresponding correlation matrix:

Estimated G Matrix							
Row	Effect	Subject	Col1	Col2	Col3	Col4	Col5
1	Sm	1	0.6794	0.01816	0.05681	0.009322	-0.00667
2	Sm*x	1	0.01816	0.1205	0.01188	0.09896	-0.02150
3	Sy	1	0.05681	0.01188	0.2703	-0.00428	-0.01828
4	Sy*m	1	0.009322	0.09896	-0.00428	0.1119	0.005422
5	x*Sy	1	-0.00667	-0.02150	-0.01828	0.005422	0.03244

Estimated G Correlation Matrix							
Row	Effect	Subject	Col1	Col2	Col3	Col4	Col5
1	Sm	1	1.0000	0.06348	0.1326	0.03381	-0.04495
2	Sm*x	1	0.06348	1.0000	0.06584	0.8523	-0.3439
3	Sy	1	0.1326	0.06584	1.0000	-0.02463	-0.1952
4	Sy*m	1	0.03381	0.8523	-0.02463	1.0000	0.09001
5	x*Sy	1	-0.04495	-0.3439	-0.1952	0.09001	1.0000

Each random effect is labeled by the variable it refers to; hence the random effect for *Sm* represents the random intercept for the *M* equation, the random effect for *Sy* represents the random intercept for the *Y* equation, the random effect for *Sm\*X* represents the random slope for the regression of *M* on *X*, and so on. Significance tests for the random effect variances and covariances are given in the following output, which labels each element of **G** by its row and column position (e.g., UN(1,1) = row 1, column 1 of the unstructured **G** matrix):

Covariance Parameter Estimates						
Cov Parm	Subject	Group	Estimate	Standard Error	Z Value	Pr >  Z

UN(1,1)	id		0.6794	0.1132	6.00	<.0001
UN(2,1)	id		0.01816	0.04109	0.44	0.6585
UN(2,2)	id		0.1205	0.02932	4.11	<.0001
UN(3,1)	id		0.05681	0.06452	0.88	0.3786
UN(3,2)	id		0.01188	0.02921	0.41	0.6842
UN(3,3)	id		0.2703	0.05409	5.00	<.0001
UN(4,1)	id		0.009322	0.03900	0.24	0.8111
UN(4,2)	id		0.09896	0.02282	4.34	<.0001
UN(4,3)	id		-0.00428	0.02782	-0.15	0.8776
UN(4,4)	id		0.1119	0.02914	3.84	<.0001
UN(5,1)	id		-0.00667	0.03450	-0.19	0.8467
UN(5,2)	id		-0.02150	0.01885	-1.14	0.2541
UN(5,3)	id		-0.01828	0.02333	-0.78	0.4334
UN(5,4)	id		0.005422	0.01788	0.30	0.7618
UN(5,5)	id		0.03244	0.02004	1.62	0.0528
Residual	id	dv M	0.6467	0.03668	17.63	<.0001
Residual	id	dv Y	0.5090	0.03067	16.59	<.0001

The final two lines of this output show the residual variances for the  $M$  and  $Y$  equations.

Last, we show the output for the fixed (average) effects in the model:

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Sm	0.09321	0.08978	99	1.04	0.3017
Sm*x	0.6119	0.04680	101	13.07	<.0001
Sy	-0.09685	0.06288	98.1	-1.54	0.1267
Sy*m	0.6106	0.04628	92.1	13.19	<.0001
x*Sy	0.2208	0.03832	71.1	5.76	<.0001

Again, the effects are labeled by the variable to which they refer. Hence,  $Sm$  represents the average intercept for the  $M$  equation,  $Sm*X$  represents the average slope for the regression of  $M$  on  $X$ ,  $Sy$  represents the average intercept for the  $Y$  equation, and so on. Other relevant output includes the asymptotic covariance matrix of the covariance parameter estimates for the random effects and the asymptotic covariance matrix of the fixed effects estimates. Elements of these matrices are necessary for computing the average indirect effect estimate and average total effect estimate. Although these elements could be used to perform the necessary calculations by hand, we also provide a macro that will perform these computations automatically.

Use of the IndTest SAS Macro

To use the IndTest macro, the **IndTest.sas** file should be downloaded and saved to a directory on the user's computer (e.g., c:\macros\). Then, this macro should be included in the SAS program where the lower level mediation model is fit with the statement

```
%include 'c:\macros\IndTest.sas';
```

where the directory should be changed as needed to where the IndTest.sas file has been saved.

An additional statement must also be added to the MIXED syntax:

```
proc mixed data=rearrange asycov covtest noclprint method=reml maxiter=100;
  class dv;
  model Z = Sm Sm*X
          Sy Sy*M Sy*X
          /noint solution ddfm=kr covb;
  random Sm Sm*X Sy Sy*M Sy*X/g gcorr type=un subject=id;
  repeated / group=dv subject=id;
  ods output covb=acovfix asycov=acovrand solutionf=estfix
            covparms=estrand convergencestatus=converge;
run;
```

The new statement is the ODS (Output Delivery System) statement that outputs the estimates of the fixed effects and variance components of the model, as well as the covariance matrices of these estimates to output data files. This statement must be input exactly as it is written above with no modification for the IndTest macro to perform properly.

After running the MIXED syntax, the IndTest macro is called with the statement:

```
%Indtest (p1=2,p2=4,p3=5,p4=8,draws=10000,seed=123,alpha=.05);
```

The first three arguments to the macro ( $p_1$ ,  $p_2$ , and  $p_3$ ) indicate the locations of the fixed effect estimates  $a$ ,  $b$ ,  $c'$  in the Solution for Fixed Effects table of the MIXED output. Here, the second row of the Solution for Fixed Effects table is the estimate for  $a$ , the third row is the estimate for  $b$ , and the fourth row is the estimate for  $c'$ . The  $p_4$  argument indicates the position of the covariance estimate for the  $a_j$  and  $b_j$  random effects in the Covariance Parameter Estimates table in the MIXED output. For the current model, this covariance is row 4, column 2 of the covariance matrix of the random effects and is listed in the 8<sup>th</sup> line of the Covariance Parameter

Estimates table. The draws and seed argument determine the number of random draws and random number seed for simulating the sampling distributions of the average indirect and total effects. Last, the alpha argument specifies the nominal alpha level (e.g., setting this to .05 specifies the use of 95% confidence intervals).

The results of running the macro for the current model are shown below. First, the macro returns the estimates it will use to compute the average indirect and total effects and their standard errors:

Estimates of a, b, c prime, and cov(aj,bj)

a	0.6118566
b	0.6105628
c prime	0.2208118
cov(aj,bj)	0.0989553

Asymptotic covariance matrix of a, b, c prime, and cov(aj,bj)

	a	b	c prime	cov(aj,bj)
a	0.0021903	0.0009813	-0.000195	0
b	0.0009813	0.002142	-0.000532	0
c prime	-0.000195	-0.000532	0.0014682	0
cov(aj,bj)	0	0	0	0.0005209

These values should be checked against the MIXED output to ensure that the locations of the relevant estimates have been correctly indicated in the macro arguments.

The next portion of output gives the estimated average indirect and total effects, their analytical standard errors, and  $p$ -values and confidence intervals based on the normal approximation:

Analytical Estimate of Average Indirect Effect and Standard Error

Estimated average indirect effect	0.4725322
Standard error	0.0536481
Critical Ratio	8.807997
P-Value	0
alpha level	0.05
LCL	0.3673839
UCL	0.5776805

Analytical Estimate of Average Total Effect and Standard Error

Estimated average total effect	0.693344
Standard error	0.0588041
Critical Ratio	11.79075
P-Value	0
alpha level	0.05
LCL	0.5780902
UCL	0.8085979

Here, LCL and UCL stand for lower and upper confidence limits. Last, the Monte Carlo method is implemented for simulating the sampling distributions of the average indirect and total effects.

The results from this approach are printed below:

Monte Carlo confidence interval				
effect	alpha	draws	LCL_sim	UCL_sim
average indirect effect	0.05	10000	0.37068	0.58066
average total effect	0.05	10000	0.58028	0.81048

As is the case here, the confidence limits from the two approaches are often very similar.

As one further example, we now provide the SAS syntax used to fit the lower level mediation models with the SCD data. This syntax demonstrates several other useful options as well as the specification of a moderated mediation model.

#### SAS code for fitting initial model with SCD data:

The initial model fit to the SCD data, shown in Equation 19 of Bauer, Preacher, and Gil (2006) was of the form:

$$EMOT_{ij} = d_{Ej} + a_j PHYS_{ij} + e_{Eij}$$

$$STRESS_{ij} = d_{Sj} + b_j EMOT_{ij} + c'_j PHYS_{ij} + e_{Sij}$$

This model was fit with the MIXED syntax shown below:

```
proc mixed data=rearrange asycov covtest noclprint method=reml maxiter=100;
  class dv; *indicates whether Z=Pain_Plea or Z=Stress;
  model Z = Sp Sp*Pain_Rate
         Ss Ss*Pain_Plea Ss*Pain_Rate
         /noint solution covb ddfm=bw;
  random Sp Sp*Pain_Rate Ss Ss*Pain_Plea Ss*Pain_Rate /g gcorr type=un subject=id;
  repeated / type=sp(pow)(day) group=dv subject=id;
run;
```

The variable `Pain_Rate` is the rating of physical pain, `Pain_Plea` is the rating of the emotional unpleasantness of the pain, and `sp` and `ss` are the selection variables. Note that the addition of the option `type=sp(pow) (day)` in the `repeated` line. This option produces a continuous time autoregressive structure for the Level 1 residuals (by `day`), and the `group=dv` specification requests different Level 1 variances and autoregressive parameters for the two dependent variables. In concert, these options in the `repeated` statement produce the Level 1 variance/covariance structure specified in Equations 20 and 21 of Bauer, Preacher, and Gil (2006).

SAS code for fitting moderated mediation model with SCD data:

The model was then extended to include the number of SCD symptoms at baseline as a moderator of the mediation effect. The Level 1 equations of the model remained unchanged from those above, but for this model each random effect was predicted by the variable *ACUTE*:

$$\begin{aligned}
 d_{E_j} &= \gamma_{d_e0} + \gamma_{d_e1} ACUTE_j + u_{d_ej} \\
 a_j &= \gamma_{a0} + \gamma_{a1} ACUTE_j + u_{aj} \\
 d_{S_j} &= \gamma_{d_s0} + \gamma_{d_s1} ACUTE_j + u_{d_sj} \\
 b_j &= \gamma_{b0} + \gamma_{b1} ACUTE_j + u_{bj} \\
 c'_j &= \gamma_{c'0} + \gamma_{c'1} ACUTE_j + u_{c'j}
 \end{aligned}$$

Substituting these terms into the Level 1 equations shows that the random effects of the model remain the same but that there are several additional fixed effects. Specifically, there is now a main effect of *ACUTE* on both outcomes, an interaction between *ACUTE* and *PHYS* in the prediction of *EMOT*, and interactions of *ACUTE* with *PHYS* and *ACUTE* with *EMOT* in the prediction of *STRESS*. Using the selection variable approach, each of these effects is multiplied by the selection variable for the equation in which it appears when predicting the combined outcome variable *Z*. The specification of this model in `MIXED` is shown below:

```

proc mixed data=rearrange asycov covtest noclprint method=reml maxiter=100;
  class dv;
  model Z = Sp Sp*Pain_Rate Sp*Acute Sp*Acute*Pain_Rate
           Ss Ss*Pain_Plea Ss*Pain_Rate Ss*Acute Ss*Acute*Pain_Rate Ss*Acute*Pain_Plea
           /noint solution covb ddfm=bw;
  random Sp Sp*Pain_Rate Ss Ss*Pain_Plea Ss*Pain_Rate /g gcorr type=un subject=id;
  repeated / type=sp(pow)(day) group=dv subject=id;
  contrast "Acute -> abj" Sp*Acute*Pain_Rate 1, Ss*Acute*Pain_Plea 1 / e chisq;
run;

```

Notice that the random statement is unchanged and that the primary modification from the prior model is to include the additional fixed effects in the `model` statement. One other addition is the contrast statement, which is used to conduct a test of the multiparameter null hypothesis that the effect of Acute on both the  $a_j$  and  $b_j$  random slopes is zero.

### Use of the IndTest SAS Macro with Moderated Mediation Models

Finally, we note that the IndTest macro can also be used with moderated mediation models like the one above. Specifically, for such a model, the IndTest macro will return a test of the simple indirect effect and simple total effect where the moderator is zero. To assess simple effects at a different value of the moderator, the moderator should be recentered at that value in a data step. The MIXED syntax must then be rerun with the newly centered version of the moderator. Last, the simple effects can be calculated by rerunning the IndTest macro from the new results. For a more detailed explanation of this approach to assessing simple effects, see Aiken and West (1991).