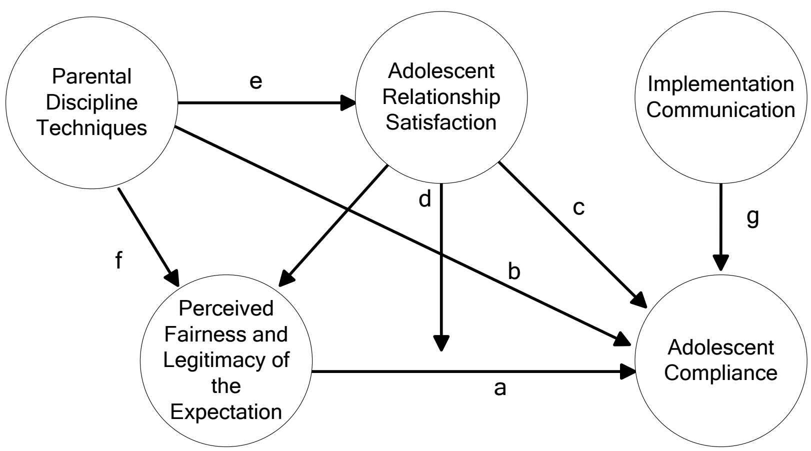
Guidelines for Writing a Proposal that Uses Causal Modeling

Guidelines for Proposal Preparation Using SEM

This handout provides guidelines for writing a dissertation proposal that uses SEM methods. I begin by considering different ways of organizing the introduction section and then I discuss sections that need to be included in the "Analysis" section of the proposal. I then describe how to organize the results section after the data have been collected and analyzed, i.e., how to write up the results section for the finished thesis.

I will use as an example the path model below. I want to develop my proposal around a test of this model. The main outcome variable is adolescent compliance with parental expectations about how they are to behave in dating situations. Compliance is impacted by how fair and legitimate the adolescent perceives the parental expectations to be (path a), the discipline strategies parents use if a transgression were to occur (path b), how satisfied the adolescent is with his or her relationship with the parents (path c), and implementation communication (path g). Implementation communication refers to how much the parent has talked with the child about how to react when pressures not to behave properly are exerted by peers. The other paths in the model are self-explanatory (at least for our purposes here).



Writing the Introduction Section

1. When writing the introduction, one strategy is to present your path diagram to readers early on as an organizing device that guides your literature review. The diagram gives the reader an overview of where you are headed and where you will end up. An alternative strategy is to save the presentation of the diagram for the end of the introduction. The idea in this case is that you review all the literature relevant to the diagram (without presenting the diagram) and this review culminates in a synthesized framework that is captured in the diagram. So, the literature review builds up to the diagram. Either approach is fine and your choice of which direction to pursue depends on what you think will communicate best.

2. Based on the above diagram, I might start my proposal by discussing adolescent compliance with parental expectations (the main outcome variable and topic of interest) and why it is an important area of study. I then write up my literature review that focuses on each link in the theory, with a subheading for each one. For example, I might have a section titled "The Relationship Between Adolescent Compliance and Perceived Fairness/Legitimacy of Expectations," that corresponds to path a. I might have another section titled "Adolescent Compliance and Adolescent Relationship Satisfaction" that corresponds to path c. And so on.

Within a section, I would review relevant literature and develop the logic of why I think there is a link between the two constructs in the heading. I would discuss competing predictions, if they exist. In some cases, it may be natural to consider more than one link within the same section.

3. In doing the above, you will want to keep the measurement model out of the diagram. The introduction is conceptual in nature, so measurement should not be a consideration. The exception, of course, is if your main focus is on measurement issues. Use circles not rectangles in your diagrams. This will make it easier when you introduce latent variables and indicators of those variables in your Method and Results sections.

4. Path diagrams are essentially a set of hypotheses, with one hypothesis per path. [For elaboration of this point, see the main text of theory construction book, Chapter 7]. Consider ending each subheading with a formal statement of a hypothesis describing the link in the path diagram. For example, I might culminate my discussion of the section on “Discipline Strategies and Adolescent Compliance” with the following hypothesis:

Hypothesis 1: The type of discipline strategy a parent uses will be associated with adolescent compliance with parental behavioral expectations. The more the parent relies on reasoning strategies in conjunction with restricting privileges related to peer contact, the more the adolescent will tend to comply with the expectation.

Note in describing this hypothesis, I avoid causal terms. Some committee members get upset if you use causal terminology because they argue you can never demonstrate causality. I personally am less concerned about this.

4a. As a variant on the above, consider grouping several hypotheses together: For example, in my section on “Perceived Legitimacy/Fairness and Adolescent Compliance,” I might end with two hypotheses:

Hypothesis 2: The more legitimate an expectation is perceived as being, the more likely the adolescent will comply with it

Hypothesis 2a: The impact of perceived legitimacy on adolescent compliance is moderated by adolescent relationship satisfaction. The less satisfied an adolescent is, the weaker will be the impact of legitimacy on adolescent compliance. This leads to the prediction that there will be an interaction between perceived legitimacy and relationship satisfaction when predicting compliance. [*Note that it is difficult to avoid causal terminology with moderated relationships*]

4b. As another variant, consider stating mediated relationships

Hypotheses 1a: The effects of reasoning during discipline on adolescent compliance will be partially mediated by relationship satisfaction and perceived legitimacy. However, reasoning will be associated with compliance independent of these two mediators.

5. If you have more than one viable model, either (1) present a “working model” and then discuss alternatives relative to it as you discuss each link, or (2) present the different models at the outset in different diagrams. I usually find the first strategy works best. Competing models can differ in (1) the predicted sign of a path coefficient, (2) the presence/absence of a path, and/or (3) the presence of reciprocal causality. How you work these in will depend greatly on the number of differences, how important those differences are, and the complexity of the models.

5a. Don’t shy away from formally stating competing hypotheses. For example, in my section on “Parental Discipline Techniques and Adolescent Compliance” I might develop logic for and state a hypothesis that parental use of threats will increase compliance. Then I might turn around and develop logic for why parental use of threats might actually reduce compliance (because the adolescent will rebel and try to get away with everything s/he can). After stating the two hypotheses, I might culminate the section by saying something like:

In sum there are two competing hypotheses that are logically reasonable:

Hypothesis 1a: The greater the number of friends an adolescent has, the less likely they will be to experiment with marijuana.

Hypothesis 1b: The greater the number of friends an adolescent has, the more likely they will be to experiment with marijuana.

The present study will test these competing hypotheses.

Writing the Method Section

In the method section, you will want to present a description of the proposed sample, the way the data will be collected and the measures that you will obtain. You can organize the description of measures around each box/circle in the path diagram. At some point, you will need to discuss how you are going to analyze the data. In some proposals, students create a separate “Results” section to do this. In other proposals, it is part of the method section (called “Planned Analyses”). Use whatever approach your major professor advises you to do and/or what you think works best.

For SEM analyses, after stating what software you will use (e.g., R, Mplus) there are 12 topics you should consider discussing in your data analysis section of your proposal. You will not necessarily discuss all of these. It depends if they are relevant. The topics are

1. Discuss how you will handle missing data:
2. Discuss how you will handle outliers.
3. Discuss how you will handle non-normality.
4. Discuss model fit indices to be used
5. Discuss limited Information versus full information estimation, if applicable
6. Discuss statistical power, stability of the sample covariance matrix, and asymptotic theory relative to sample size issues
7. Discuss measurement error and how it will be dealt with
8. Discuss control of familywise error rates
9. Discuss factor structure analyses of multiple item measures
10. Discuss clustering and weights, if applicable
11. Discuss model comparison strategies, if applicable
12. Acknowledge the possibility of redundant/equivalent models and addressing specification error

On the next pages, I have written a sample section on each of these topics. Feel free to adapt these to your own proposal. You will need to make changes in a few places based on your sample size and model, but it should be obvious where this needs to be done.

Missing Data

Most of the current literature favors using a FIML approach for missing data. Assuming you are working within a standard SEM package, this approach should be readily available. Here is a sample of what you might write:

"Missing data are expected to be minimal for most variables. Given missing data, parameter estimates and model tests will be pursued in the context of Full Information Maximum Likelihood (FIML) methods as implemented in Mplus. Missing data bias will be explored by computing a dummy variable reflecting the presence or absence of missing data for each variable in the model and then this dummy variable will be correlated with all other variables in the model as well as selected variables external to the model *[you need to specify what these other variables are]*"

If you use some form of imputation method (either single imputation if the amount of missing data is minimal or multiple imputation using, for example, chained equations) then you would describe this process accordingly. Because most of you will be using FIML, I only consider it.

Outliers

Here is what you might write about outliers:

"Traditional multivariate outlier analysis often uses the Mahalanobis D statistic, but such statistics are vulnerable to the very outliers they are intended to detect. The present study will use a robust outlier method based on a projection type method described in Wilcox (2017). This method calculates a robust measure of location for each of the k variables in the data, such as a median. For a given data point, a line is projected though the centroid, with the line extending through the full multidimensional space of the data. All of the remaining data points are then perpendicularly projected onto this line and values are assigned to each data point based on where along the line it falls relative to the centroid. This process reduces the multivariate data to a univariate variable consisting of values representing distance from the centroid. A MAD-median rule is then applied to identify outliers. Analyses will be conducted both with and without the identified outliers to determine if conclusions are potentially impacted by them."

Non-normality

I assume you will be using robust maximum likelihood from Mplus or bootstrapping. Write the following:

"Issues of non-normality will be addressed by using the Huber-White robust method of estimation as implemented in Mplus (option MLR). If the data patterns are incompatible with this method, percentile bootstrapping will be used."

Indices of Fit

Consider writing something like the following:

"Following the recommendations of Bollen and Long (1993), a variety of global fit indices will be used, including indices of absolute fit, indices of relative fit and indices of fit with a penalty function for lack of parsimony. These include the traditional overall chi square test of model fit (which should be statistically non-significant), the Root Mean Square Error of Approximation (RMSEA; which should be less than 0.08 to declare satisfactory fit), the p value for the test of close fit (which should be statistically non-significant), the Comparative Fit Index (CFI; which should be greater than 0.95); and the standardized root mean square residual (which should be less than 0.08). In addition to the global fit indices, more focused tests of fit will be pursued. These include examination of the standardized residual covariances (which should be between -2.00 and 2.00) and modification indices (which should be less than 4.00). The parameter estimates also will be examined for Heywood cases. "

Limited Information Estimation versus Full Information Estimation

If you use a limited information estimation approach, consider the following:

The theoretical questions posed in this research are framed in the path diagram in Figure 1. It is natural to think of applying traditional structural equation modeling (SEM) strategies to such models. Traditional SEM uses full information estimation approaches where all of the path coefficients (and their standard errors) are estimated simultaneously in the context of the full system of linear equations implied by the model. The same statistical algorithm (e.g., maximum likelihood estimation) is applied throughout. An alternative approach is to use a limited information estimation strategy. This approach uses the path diagram to identify the structural relationships of interest and to define the relevant linear equations. However, the overall model is broken up into pieces and estimates of the coefficients are derived within each piece separately using statistical methods that are appropriate for that piece. Full information estimation approaches can yield more efficient parameter estimates and also yield more perspectives about goodness of model fit. However, the full information estimation approach also has disadvantages. For example, model misspecification in one part of the model can yield biased estimates in another part of the model. By contrast, in limited information estimation, specification error is compartmentalized. Limited information estimation also allows one to tailor the analytic method to the nature of the variables involved in a given piece of the overall model (e.g., logistic regression, ordinal regression, OLS regression, Poisson regression). Full information estimation strategies will be pursued but, where necessary, limited information estimation approaches will be used.

Statistical Power and Sample Size Considerations

For statistical power and sample size, the best way to determine power for a given path coefficient is to pursue a Monte Carlo simulation (see Muthén & Muthén, 2002). If you would rather use a more informal approach, consider the following:

"To determine an appropriate sample size, structural equation modeling requires that in addition to statistical power, issues of the stability of the covariance matrix and the use of asymptotic theory be taken into account. In terms of power, it is difficult to evaluate the power associated with specific path coefficients in complex SEM models because of the large number of assumptions about population parameters that must be made. A rough approximation of power can be obtained by using a limited information approach with single indicators of the path models implied by Figure 1. This permits the use of traditional power analysis software to gain a sense of sample size demands (Jaccard & Wan, 1996). In all examples below, we assume an alpha level of 0.05 and a two tailed test.

For a multiple regression analysis with 4 predictors where the squared multiple correlation is 0.30 and where one wants to detect a predictor that accounts for at least 5% unique variance in the outcome, the required sample size to achieve power of 0.80 is approximately 115. For a logistic regression analysis where the target predictor is a continuous predictor with four other predictors in the equation, where the event rate at the mean of all predictors is 0.20 and where the multiple correlation of the predictor with the other predictors is 0.30, the sample size needed to detect an odds ratio of 1.75 expressed in standardized metrics is about 170 and for an odds ratio in standardized metrics of 2.00 is about 110. For a simple zero order correlation of 0.30 in the population, the sample size needed to achieve power of 0.80 is approximately 80. For a contrast of means between two independent groups and an effect size corresponding to Cohen's definition of a medium effect (a d value of 0.50), the sample size needed to achieve power of 0.80 is approximately 65 per group. For a contrast of dependent means, the corresponding required sample size is about 35. For a percentage difference between two independent groups where the population percentage in the first group is 30 percent and in the second group it is 15 percent, the required sample size for power of 0.80 is about 120 per group. The proposed sample size for this study seems adequate in terms of power.

In terms of asymptotic theory and covariance stability, simulation studies tend to suggest that sample sizes of 100 to 125 or larger often yield adequate results given that reasonably reliable measures are used (reliabilities greater than 0.65) and with a reasonable number of indicators per latent variable (Jackson, 2003; Jaccard & Wan, 1996). The sample size in the proposed study exceeds this standard."

Again, the best way to conduct formal analyses is through simulations and I recommend you do so using that approach. In addition to statistical power, you likely should discuss sample size issues related to asymptotic theory, stability of covariance matrices, and margins of error. See the document on "Methodological Rules" on the website for Chapter 15 for details on this topic.

Measurement Error

Consider writing the following:

"Measurement error will be taken into account through the use of multiple indicators of constructs. In cases where only a single indicator is available, we will adopt the strategy suggested by Joreskog and Sorbom (1996). This involves constraining the error/unique variances for each measure to values corresponding to a priori determined levels of reliability. The reliability levels for the measures will be based on composite reliability indices for multi-item scales or previous research."

Familywise Error Rates and Multiple Contrasts

Consider writing the following:

"At times, multiple significance tests will be conducted within a family of contrasts and there will be concern for inflated familywise error rates. The robustness of conclusions will be compared both with and without statistical corrections for multiple tests (using the strategy discussed in Jaccard & Guilamo-Ramos, 2002). In general, a Holm adjusted modified Bonferroni method (Jaccard, 1998) will be used for controlling familywise error rates, which is more powerful than traditional Bonferroni methods."

Factor Structure of Multiple Items Measures

Consider writing the following:

"For all multi-item measures, the composite reliabilities and factor structures of the measures will be evaluated to ensure that they are behaving in a way that one would expect based on their psychometric histories. Some of the variables in the path diagrams reflect variable categories with multiple variables or dimensions. The intercorrelations of variables will routinely be examined, and coupled with substantive criteria and the results of exploratory or confirmatory factor analyses, decisions will be made about combining indices or introducing latent constructs into the analysis."

Sample Weights

If you are using sample weights, consider writing the following:

"The use of sampling weights in complex model evaluation is controversial (e.g., Lohr & Liu, 1994; Winship & Radbill, 1994). Winship and Radbill (1994) note that if a model is specified correctly and sampling is not outcome based, then use of unweighted estimation strategies are preferred over weighted estimation strategies because they yield smaller standard errors. In practice, models are almost always misspecified to some degree, so the more realistic question is whether the degree of misspecification is consequential (Kott, 1991; DuMouchel & Duncan, 1983). Feinberg (1989) argues that outcome based sampling is the only situation in which weights should be used for multivariate analyses. A range of perspectives on the use of sample weights can be found in Saphire (1984), Rubin (1985), Little (1991), Lohr and Liu (1994), Winship and Randall (1994) and Scott and Wild (1989). DuMouchel and Duncan (1983) propose a test of whether the results of a weighted solution differs significantly from those of unweighted solutions. Asparouhov and Muthen (2009) present a comparable test for SEM modeling. In the absence of such differences, one can report either weighted or unweighted results. The results of the weighted analyses are reported if one wants to be conservative with respect to bias and specification error whereas the results of the unweighted analyses are reported if one wants to maximize efficiency of the estimators.

Both unweighted and weighted analyses will be pursued. The weighted analyses will be performed in M Plus using the procedures discussed in Asparouhov (2005)."

See the document on "Methodological Rules" on the website for Chapter 15 for details on this topic.

Clustering

If you have clustering, consider adapting the following:

"The data will be collected in different organizations/schools with a substantial number of persons within each organization/school, so there is the possibility of clustering effects. The degree of clustering will be evaluated by examining intraclass correlations and adjusting for clustering if the ICCs suggest it necessary to do so, either by the introduction of covariates reflecting organization/school units, or the use of robust estimators available in the M Plus computer programs."

See the document on "Methodological Rules" on the website for Chapter 15 for details on this topic.

Model Comparisons

Consider adapting the following:

"Comparisons of nested models will use either the traditional nested chi square test for robust algorithms or information fit indices (AIC and BIC; see Raftery, 1995)."

Equivalent Models and Specification Error

Consider adapting the following:

"It is recognized that there may be equivalent models that can account for the data relative to the models being tested. Equivalent models will be described and used to qualify conclusions in the discussion section. In addition, attention will be given to the analysis of diagnostics relevant to specification error in the model focusing on non-linear relationships and interaction effects."

References

Asparouhov, T. (2005). Sampling weights in latent variable modeling. Structural Equation Modeling, 12, 411-434.

Bollen, K. & Long, S. (1993). Testing structural equation models. Newbury Park: Sage.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Matwah, NJ: Erlbaum.

DuMouchel, W.H. & Duncan, G. J. (1983). Using sampling survey weights in multiple regression analyses of stratified samples. Journal of the American Statistical Association, 78, 535-543.

Feinberg, S. E. (1989). Modeling considerations: Discussion from a modeling perspective. In Kaspryzk, D., Duncan, G., Kalton, G., & Singh, M.P. (Eds.) Panel surveys. New York: Wiley.

Honaker, J., Joseph, A., King, G., Scheve, K. & Singh, N. (2003). Amelia: A program for missing data. Department of Government, Harvard University.

Jaccard, J. (1998). Interaction effects in factorial analysis of variance. Newbury Park: Sage.

Jaccard, J. and Wan, C. (1996). LISREL analyses of interaction effects in multiple regression. Newbury Park: Sage.

Jaccard, J. (2006). Multiple regression analysis in clinical child and adolescent psychology. Journal of Clinical Child and Adolescent Psychology, In press.

Jaccard, J. and Guilamo-Ramos (2002). Analysis of variance frameworks in clinical child and adolescent psychology: Basic issues and recommendations. Journal of Clinical Child and Adolescent Psychology, 31, 130-146.

Jackson, D. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. Structural Equation Modeling, 10, 128–141.

Joreskog, K. & Sorbom, D, (1996). Users manual for LISREL. Chicago: Scientific Software.

King, G., Honaker, J., Joseph, A. & Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. American Political Science Review, 95, 49-69.

Kott, P.S. (1991). A model based look at linear regression with survey data. American Statistician, 45, 107-112.

Lee S. Y. & Xu, L. (2003). Case-deletion diagnostics for factor analysis models with continuous and ordinal categorical data. Sociological Methods and Research, 31, 389–419.

Liang, K. & Zeger, S. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 73, 13-22.

Little, R.J.A. (1991). Inference with survey weights. Journal of Official Statistics, 7, 405-424.

Lohr, S. L & Liu, J. (1994). A comparison of weighted and unweighted analyses in the National Crime Victimization Survey. Journal of Quantitative Criminology, 10, 343-360.

Maxwell, S. (2000). Sample size and multiple regression analysis. Psychological Methods, 5, 434-458.

Muthén, B.O. (2004). Mplus technical appendices. Los Angeles, CA: Muthén & Muthén

Muthen, L. & Muthen, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. Structural-Equation-Modeling. 9, 599-620.

Raftery, A. (1995). Bayesian model selection in social science research. Sociological Methodology, 25, 111-163.

Rubin, D.B. (1985). The use of propensity scores in applied Bayesian inference. In Bernardo, J., DeGroot, M.H., Lindley, D.V. & Smith, A. (Eds.). Bayesian statistics. Amsterdam: North Holland.

Saphire, D.G. (1984). Estimation of victimization prevalence using data from the National Crime Survey. New York: Springer-Verlag.

Scott, A.J. & Wild, C.J. (1989). Selection based on the response variable in logistic regression. In Skinner, C.J., Holt, D. & Smith, T.M. (Eds.). Analysis of complex surveys. New York: Wiley.

Wilcox, R. (1999). Fundamentals of modern statistical methods. New York: Springer.

Wilcox, R. (2003). Applying contemporary statistical techniques. San Diego: Academic Press.

Wilcox, R. (2012). Introduction to robust estimation and hypothesis testing. San Diego: Academic Press. Third edition.

Wilcox, R. (2017). Introduction to robust estimation and hypothesis testing. San Diego: Academic Press. Fourth edition.

Winship, C. & Radbill, L. (1994). Sampling weights and regression analysis. Sociological Methods and Research, 23, 230-257.

Guidelines for Writing a Results Section in a Thesis

This section provides guidelines for writing the results section of a thesis once data analysis is complete. You will want to cover the same basic sections as your proposal, but now the description of what you did will be integrated with the presentation of results.

A common approach is to have three sections, each labeled with a separate heading, (1) preliminary analyses, (2) main analyses, and (3) supplementary analyses.

Preliminary Analyses

In this section, you describe how much missing data occurred, what biases were isolated in missing data patterns (if any), how much non-normality there was and what happened in your analysis of outliers. In addition you describe any psychometric analyses that were done on your measures.

Here is a section titled "Preliminary Analyses" from an article I published that addresses most of these issues and illustrates this portion of a results sections.

Preliminary Analyses

Descriptive Statistics. Table 1 presents means and standard deviations for all of the continuous variables used in the models. The median values for each of the variables (not reported) were close to the mean values. The mean value of T-STAI corresponds to T scores of 55 (female) and 56 (male) for normal adults age 40-49.

Outliers. Outlier analyses used methods suggested by Wilcox (2017). Specifically, a projection type robust algorithm was used to identify multivariate outliers. None outliers were identified. Analyses were conducted with and without the outliers and none of the significance patterns changed. Results are presented for the full sample rather than the outlier-reduced sample.

Missing Data. There were small amounts of missing data amounting to no more than a few cases on any given variable. There was no coherent pattern to the missing data. Given the small number of instances of missing data, concerns surrounding estimation with missing information are moot. Missing data were accommodated using full information maximum likelihood (FIML) methods.

Non-Normality. Traditional maximum likelihood methods of SEM assume that the continuous variables in the model are multivariately normally distributed. Skewness and kurtosis indices for each variable are presented in Table 1. Troublesome skewness and kurtosis values are evident for the measure of psychopathology. Model estimation was pursued using robust maximum likelihood algorithms with Huber-White robust estimators as implemented in Mplus.

Main Analyses

In this section you want to convey what happened in your primary analyses, following the basic write-ups we used in class. If your initial model did not fit and you made modifications to it, you will want to note this, as discussed in class. You do not want to take the reader through all the gory details, but you do want to let them know how the analysis progressed from beginning to end.

Supplemental Analyses

In this section, you want to address supplemental issues, such as assuring the reader that the study was sufficiently powered, the biasing effects of measurement error, and exploring specification error. Here is an example from the same article I mentioned above.

Supplemental Analyses

In addition to the above model tests, we conducted supplementary analyses to explore potential problems of model misspecification and parameter bias induced by the presence of measurement error. For the former, we used traditional regression methods in conjunction with product terms to test for possible interaction effects between predictors of each endogenous variable in the model (Jaccard & Wan, 2003). The regression equations were dictated by the limited information estimation approach to SEM described by Bollen (1996) and did not suggest the presence of any meaningful interaction effects. In terms of measurement error, we re-estimated the model but imposed an a priori determined amount of measurement error onto the observed measures using the strategy described by Joreskog and Sorbom (1996). The amount of unreliability imposed was based on the composite reliability for each scale. None of the major conclusions drawn from the original significance tests changed.

It also is useful to provide perspectives on statistical power for the tests of the path coefficients so that one can better appreciate the possibility of a Type II error for statistically non-significant path coefficients. Power analyses for SEM models are complicated and often rest on assumptions that are impractical or not viable. We followed the practice recommended by Jaccard and Turrisi (2003) that provides a rough sense of statistical power by applying power analytic methods for OLS regression as applied to selected linear equations from the set of linear equations implied by the model in question. Given a sample size of 168 and a two tailed alpha level of 0.05, we evaluated the statistical power associated with a path coefficient that represents 5% explained variance over and above a set of five additional covariates. Based on the residuals in Figure 1, we evaluated three scenarios where the initial set of covariates accounted for 10% of the variance, 20% of the variance, or 40% of the variance. The approximate statistical power in these three scenarios was 0.87, 0.89, and 0.97. For a path coefficient that represents 3% additional explained variance in the same scenarios, the approximate statistical power was 0. 66, 0.72, and 0.84. Overall, the approximate power seems adequate for detecting paths that account for at least 5% of the variance of an outcome variable and in some cases, it also is adequate for coefficients that reflect only 3% unique explained variance.