2012-06-08

Estimation of the Effects of Parental Measures on Child Aggression Using Structural Equation Modeling

Jordan Daniel Pyper
Brigham Young University - Provo

Estimation of the Effects of Parental Measures on Child Aggression Using Structural Equation Modeling

Jordan D. Pyper

A selected project submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

G. Bruce Schaalje, Chair
William F. Christensen
Natalie J. Blades

Department of Statistics

Brigham Young University

June 2012

ABSTRACT

Estimation of the Effects of Parental Measures on Child Aggression Using Structural Equation Modeling

Jordan D. Pyper
Department of Statistics, BYU
Master of Science

A child's parents are the primary source of knowledge and learned behaviors for developing children, and the benefits or repercussions of certain parental practices can be long lasting. Although parenting practices affect behavioral outcomes for children, families tend to be diverse in their circumstances and needs. Research attempting to ascertain cause and effect relationships between parental influences and child behavior can be difficult due to the complex nature of family dynamics and the intricacies of real life. Structural equation modeling (SEM) is an appropriate method for this research as it is able to account for the complicated nature of child-parent relationships. Both Frequentist and Bayesian methods are used to estimate the effect of latent parental behavior variables on child aggression and anxiety in order to allow for comparison and contrast between the two statistical paradigms in the context of structural equation modeling.

Estimates produced from both methods prove to be comparable, but subtle differences do exist in those coefficients and in the conclusions to which a researcher would arrive. Although model estimates between the two paradigms generally agree, they diverge in the model selection process. The mother's behaviors are estimated to be the most influential on child aggression, while the influence of the father, socio-economic status, parental involvement, and the relationship quality of the couple also prove to be significant in predicting child aggression.

Keywords: latent variables, manifest variables, structural equation modeling, Bayesian methods, Frequentist methods

CONTENTS

CHAPTER 1

INTRODUCTION

An essential feature of the development of personality and human behavior is the socialization that a child receives from his or her parents. Beginning in infancy, the attachment that a child is able to make with the mother affects the child's disposition (Ainsworth 1989). Children with more despondent mothers struggled to form emotional attachments with others and also suffered from greater stress and behavioral problems. Parents are the primary source of knowledge and learned behaviors for developing children, and the benefits or repercussions of certain parental practices can be long lasting. Ideal outcomes for children occur when parental efforts are harmonious and cooperative (Shaffer 2005). That is to say, parents that share a stable and supportive relationship provide an environment that leads to better outcomes for their children. Conversely, couples dealing with marital discord among other difficulties, such as financial problems, struggle to provide this environment (Kitzmann 2000). Parents affect children directly through their interactions with their children and indirectly through the example they provide. Of these practices, there are some that may be more responsible than others for later problems in children. Identifying particular problematic aspects of parental comportment is useful when an intervention or support is needed for troubled families. One objective of this research is to estimate the effects of various parental practices on child behaviors for 'at risk' families.

Although parenting practices affect behavioral outcomes for children, families tend to be diverse in their circumstances and needs. Research ascertaining causal relationships between parental influences and child behavior can be difficult due to the complex nature of family dynamics and the intricacies of real life. Children raised in unstable family environments, which can involve changes in residence, primary caregiver having many intimate partners, and disadvantaged circumstances exhibit more aggression, disobedient conduct, as well as, depression and anxiety (Ackerman et al. 1999). Consequently, environmental factors such as economic circumstances must be considered to provide an understanding of the contextual effect in which these family dynamics are taking place. The intent of this research is to confirm that practices such as child involvement and relationship quality of the parents will have a positive effect, while more problematic behaviors like substance abuse will adversely affect the child. Additionally, these problematic practices can also affect other parental behaviors like relationship quality.

Structural Equation Modeling (SEM) lends itself to this kind of a research question because it allows the researcher to consider complex relationships between the variables of interest. SEM represents environmental factors and parental variables as latent variables in order to estimate the parental and environmental contribution to a child's behavior. Parental behaviors such as substance abuse or psychological issues are not easily observed as these behaviors are multifaceted concepts and are difficult to accurately measure or even quantify. The use of latent variables in SEM models allows the researcher to get at these difficult concepts by measuring more accessible data that together comprise the desired factor. Structural equation modeling not only allows for the estimation of latent variables, but also accounts for measurement error present in the observed (manifest) variables. Particularly within the social sciences, this can be a problem as most measures for sociological phenomena lack the measurement precision assumed in common statistical analyses like regression. Measurement error in the data leads to attenuation of the effect estimates (Chesher 1991). Due to these properties, SEM provides an effective analysis when attempting to estimate the effect of various parental practices on child behavioral outcomes.

The majority of the work with structural equation models has been done using frequentist methods that estimate model parameters using maximum likelihood, which assumes that the latent variables are normally distributed and relies on asymptotic normal theory to make inferences on those parameters (Lee 2007). Bayesian methods provide a recent approach to SEM that does not rely on normality or asymptotic results. Bayesian methods, however, do require that all parameters be given prior distributions allowing the researcher to incorporate previous knowledge into the analysis. Relevant research exists for most sociologic problems which can be used to elicit prior parameter distributions. Bayesian and Frequentist methods possess useful qualities and both will be implemented for this problem. A second objective of this research is to then compare and contrast these methods to better understand what the two statistical paradigms uniquely contribute to this problem.

CHAPTER 2

LITERATURE REVIEW

2.1 Structural Equation Modeling

Structural equation modeling is an extension of confirmatory factor analysis where specific explanatory relationships are specified among manifest and latent variables (Raykov and Marcoulides 2000). As with confirmatory factor analyses, latent constructs are created by using manifest or observed variables as manifestations of some latent category. Relationships or regression equations are specified between these latent categories and other latent or manifest variables.

Confirmatory factor analysis assumes that

$$\mathbf{y}_i \sim MVN(\boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{f}_i, \boldsymbol{\Sigma}), \tag{2.1}$$

where $\mathbf{y}_i$ is the manifest data for the i^{th} individual, $\boldsymbol{\Lambda}$ is a matrix of factor loadings, $\boldsymbol{f}_i$ is a vector of latent variables, $\boldsymbol{\nu}$ is a vector of intercepts, and $\boldsymbol{\Sigma}$ is the covariance matrix for these manifest variables. The confirmatory factor analysis model, also known as the measurement model, is similar to a standard regression equation,

$$\mathbf{y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{f}_i + \boldsymbol{\psi}_i, \tag{2.2}$$

where, $\boldsymbol{\psi_i}$ is an error term associated with the manifest variables. A structural equation model extends this measurement model by specifying linear relationships among latent and additional manifest variables. This can be expressed as,

$$\boldsymbol{f}_i = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{f}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i. \tag{2.3}$$

The latent variable $\boldsymbol{f}_i$ is regressed on the explanatory latent variable $\boldsymbol{\xi}$ where $\boldsymbol{B}$ describes the relationship among all $\boldsymbol{f}_i$, $\boldsymbol{\Gamma}$ contains the estimated effects of $\boldsymbol{\xi}$ on $\boldsymbol{f}$, and $\boldsymbol{\delta}_i$ is the error from that regression (Lee 2007).

A visual depiction of equations 2.2 and 2.3 is shown below in figure 2.1. Variables x_1 and x_2 are the manifest variables for latent variable $\boldsymbol{\xi}$ and ν_1 and ν_2 are the factor loading for those manifest variables. In this case, $\boldsymbol{\Gamma}$ is not a matrix but one coefficient γ_1 being the estimated linear effect of $\boldsymbol{\xi}$ on $\boldsymbol{f}$. Variables y_1 and y_2 are manifest variables associated with the latent variable $\boldsymbol{f}$ as described in equation 2.2.

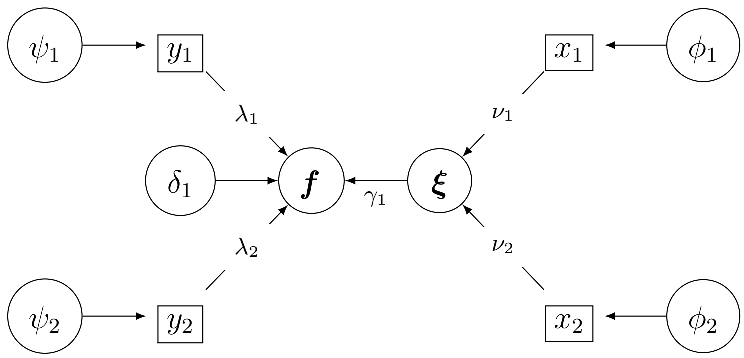

Figure 2.1: A simple SEM

While some SEM's are able to estimate non-linear effects, the most common model assumes linear effects in the variables and that the latent variables are normally distributed. This is known as a LISREL model, which stands for linear structural relations (Raykov and Marcoulides 2000). General assumptions associated with these models are that the elements of the latent variable $\boldsymbol{\xi}_i$ are independent of the elements of $\boldsymbol{\delta}_i$, and error terms associated with the structural equation ($\boldsymbol{\delta}$) are independent of the factor loading error terms ($\boldsymbol{\Psi}$). Also, error terms from the factor loadings of $\boldsymbol{\xi}$ ($\boldsymbol{\phi}$) are independent of factor loading errors of $\boldsymbol{f}$. All errors from both the measurement and structural component of the model are assumed to be distributed as multivariate normal with mean vector $\mathbf{0}$ and respective covariance matrices (Dunson et al. 2005). Identifiability constraints needed for confirmatory factor analysis similarly apply to SEM as latent variables are computed the same way. Bayesian methods share many of these assumptions but do not make any assumptions of asymptotic normality of the latent variables or endogenous error terms.

2.2 Bayesian SEM

Research in the social sciences is primarily based on social theory or previous work that has been done on the topic of interest. Bayesian methods allow for the incorporation of previous knowledge to be specified for all the parameters in the model including the latent variables. Exact distributions can be sampled using MCMC methods, making it unnecessary to derive the theoretical sampling distributions of the model parameters. Additionally, distributions associated with those parameters are not necessarily constrained to be distributed normally and inference based on asymptotic theory as frequentist methods would impose.

Structural equation modeling is understood to be a 'large sample' statistical method, meaning that in order for SEM to provide stable results a sample size of at least a few hundred is typically needed. However, Bayesian methods can provide greater stability with smaller sample sizes, but when less data is available the model estimates will rely more heavily on prior specifications. In general terms, a Bayesian analysis introduces bias into the parameter estimates through the prior distributions with the objective of reducing the mean square error. Intuitively, as the sample size gets large estimates produced from Bayesian and Frequentist methods should closely resemble each other as the prior information will have less influence with larger sample sizes.

Many researchers may prefer the idea of unbiased estimates because of notions that unbiased estimates provide an 'objective' approach to research. Both Frequentist and Bayesian methods arguably introduce bias into the analysis by how the researcher approaches, understands, and hypothesizes the research question (Berger and Berry 1988). Non-informative priors can be in used in order to prevent influencing posterior estimates, which should be implemented in situations where prior information is not available (Lee 2007). Caution should be exercised, however, when using highly disperse prior specifications because overly non-informative priors often produce improper posterior distributions for more complex models rendering essentially meaningless results (Dunson et al. 2005).

Bayesian methods provide posterior distributions for not only the parameters of interest but also for the latent variables. These distributions provide useful insights into non-linearity and other lack of fit issues. Problems with lack of fit can be assessed from the Frequentist paradigm, as well, using a two-stage approach. These estimates, however, are biased and do not capture the uncertainty associated with the model, and may be problematic to estimate (Dunson et al. 2005). Sensitivity of some estimates to prior distributions, particularly variances, can also be of concern. Additionally, when non-informative priors lead to improper posterior densities the added diagnostic benefits of posterior densities for the latent variables are lost.

Although Bayesian SEM's possess the added benefits of providing posterior distributions for the latent variables, there are potential issues that can arise from the Monte Carlo Markov Chain sampling methods. The issue of slow mixing of the MCMC sampler or high autocorrelation in the posterior draws often occurs in hierarchical settings which can also be the case when estimating structural equation models. However, this autocorrelation can be mitigated by choosing alternate parameterizations (Dunson et al. 2005). MCMC within Gibbs sampling tends to be a computationally expensive process in general and, even more so for SEM's due to the high complexity of these models. Consequently, estimating Bayesian structural equation models come at a higher computational cost.

Estimating the effects of parental behaviors on child well-being can be done using both Frequentist or Bayesian methods. While both statistical paradigms differ in their methods of estimation they both can produce a LISREL model which estimates linear effects. Various aspects of these approaches will be compared and contrasted in this project to provide a more exhaustive analysis and to better understand what features of these methods best lend themselves to this sociological problem.

CHAPTER 3

METHODS

Data for this project come from the Fragile Families and Child Wellbeing Study (FFCW), a birth cohort study of nearly 5,000 children born in 20 U.S. cities between 1998 and 2000. The Fragile Family and Child Well-being study comes from the Bendheim-Thoman Center for Research on Child Well-being (CRCW) and is funded through grants and a number of private foundations. Data from this study are made publicly available for the purpose of understanding family dynamics of 'at-risk' households where children are born to unmarried parents who are at a greater risk for breaking up. The data set for this project was provided by Dr.W.Justin Dyer from Brigham Young University and was used for his doctoral dissertation; however, some variables used in this project were not used in Dr.Dyer's research and the statistical analysis used here is distinct.

3.1 Variables

Variables considered from the FFCW data set reflect parental practices, environmental circumstances, and child behaviors. Variables used in the analysis come from the third and fourth waves or time points of collection (third wave when the child was 30 to 36 months old and forth at 5 years of age). The explanatory variables come from the third wave of the study, whereas the response variables come from the fourth. Each wave of the study is separated by one year; that is, the child behaviors were measured one year after the parental variables considered in the model. This temporal gap is necessary because the effect of parental practices occur over a period of time and are not instantaneous. Consequences from these practices in child behavior are subsequent to experiencing those practices over a period of time. Variables used in the analysis are shown in Table 3.1.

Table 3.1: FFCW Variables.

Response Variables	**Description**
Aggress	Mother's responses to the Achenbach Child Behavior Checklist
Explanatory Variables	**Description**
Father's Characteristics	
FDrug	Indicator for father's illicit drug use
F_FInvol	Score indicating father's involvement with his child
FAlcoho	Indicator for father's alcohol use
FDep	Indicator for depressive behavior
FAnx	Indicator for anxiety disorders
Mother's Characteristics	
MDrug	Indicator for mother's illicit drug use
M_FInvol	Score indicating mother's involvement with his child
MAlcoho	Indicator for mother's alcohol use
MDep	Indicator for depressive behavior
MAnx	Indicator for anxiety disorders
Environmental Circumstances	
DEmp	Indicator if father is employed
DadEd	Father's education attainment
Inc2PT3 (*Family income-to-needs*)	family income divided by poverty line
P_BelPov	Percent of families in area below poverty level
Punemp	Percent of families in area that are unemployed
RelQual	Score of relationship quality for each couple

Eight of the fifteen manifest variables involved in the measurement equation for these models are indicator variables which are not normally distributed. Although assuming multivariate normality is not necessarily a good assumption for these eight variables, it does not provide meaningless results. Consequently, data from these Bernoulli distributed variables contain less information than continuous variables. The amount of information provided by binary variables is also affected by the relative prevalence of 1's compared to 0's. If prevalence rates of 1's in the data are low, then little information is known about the observations corresponding to that category and is analogous to making inference from a small sample. Perils stemming from small sample sizes can also apply to these situations where inferences are done with greater trepidation. Prevalence rates of responses of 1 compared to 0 for these eight variables are shown below in table 2.2.

The largest behavioral prevalence rate of 21.1% belongs to the binary variables for mother's depressive status, and the lowest rate of 1.4% corresponds to mother's alcohol use. The average prevalence rate of problematic parental behaviors was 3.1% when excluding the

Table 3.2: Prevalence Rates for Parental Binary Variables

Var.	0	1	rate
FAlcohol	4699	199	0.04
MAlcohol	4831	67	0.01
FDep	4089	800	0.16
MDep	3861	1037	0.21
MAnx	4670	228	0.05
FAnx	4691	207	0.04
FDrug	4768	130	0.03
MDrug	4824	74	0.02

two parental depression variables which have the two highest rates. This demonstrates that despite a large sample size, there is not much information contained in many of these binary variables.

Using variables with sparse information in a complex SEM analysis is far from ideal; however, it is reflective of the reality of the situation when the desired variable or measure, such as parental issues or substance abuse, is both difficult to measure and relatively infrequent in the population. Where Frequentist and Bayesian methods would tend to closely agree in more ideal large sample scenarios, this data problem provides a situation in which the two methods are more likely to diverge. Bayesian estimates based on the data may be more flexible than frequentist estimates derived from the sample covariance matrix, but they may also be more dependent on the priors as is the case with smaller sample sizes. Comparing the two statistical paradigms under more exigent circumstances may provide better insights into the benefits and differences between them.

3.2 Proposed Model

Initial hypotheses regarding the latent structures of the parental variables postulated one latent variable involving the *Father's Characteristics* from Table 1 and another with the *Mother's Characteristics*. These were thought to be measures of the mother's and father's contribution to the child's behavior. However, fit indices from confirmatory factor analysis

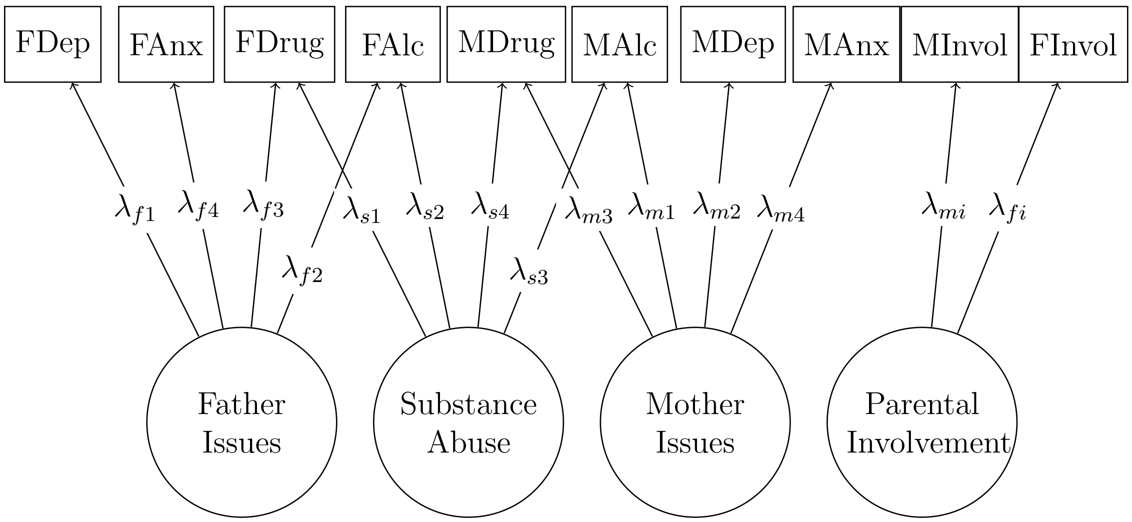

Figure 3.1: Latent Constructs of Parental Variables.

reported a poor fit of the data to these latent constructs. Consequently, exploratory factor analysis was used to better understand relationships among the manifest variables. The maximum likelihood method allowed models to be estimated using different numbers of latent variables. Using more than four latent variables provided no additional insight into the nature of the latent construct so only four were used. Figure 3.1 shows the latent construct to be used for this project.

From the loadings provided by the exploratory factor analysis, four resultant latent variables emerged: parental involvement, substance abuse, father issues, and mother issues. This latent construct should be examined with some skepticism as it was a result of 'data snooping' and may be over fit to the current data set leading to poor out-of-sample prediction.

The last latent variable considers economic circumstances is shown in Figure 3.2. This variable representing the relative class or socio-economic status (SES) of the families in the study fit well as hypothesized.

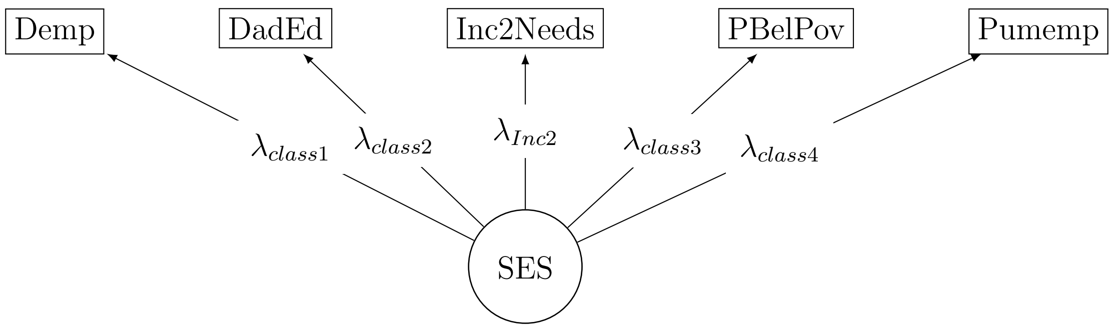

Figure 3.2: Latent class or socioeconomic status variable.

These variables provide an idea into the education, employment, income, economic situation of the communities in which families live. Note that error terms for each manifest variable have been omitted from these figures.

To estimate the structural equation model these latent variables were regressed on the behavioral outcomes for children, aggression and anxiety. Different parental practices may contribute to different behavioral problems requiring outcomes for children to be considered separately instead of treated as a latent variable. Also, interventions or treatments for different behavioral issues are often distinct so it is more informative to analyze these behaviors individually.

The full structural equation model for this data is shown in Figure 3.3 with the manifest loadings on the latent variables omitted for simplicity. Covariances between all of the latent variables were estimated to better capture the complex relationships between these different aspects of parental behaviors and circumstances. These covariance parameters are symbolized using ϕ, and regression effects are notated using β. Latent variables representing mother issues, father issues, and substance abuse are all likely to covary because they share manifest variables indicating mother's and father's drug and alcohol use. Because mother and father issues are correlated and consequently collinear, only one of the parental issues will be used in the mediational relationship.

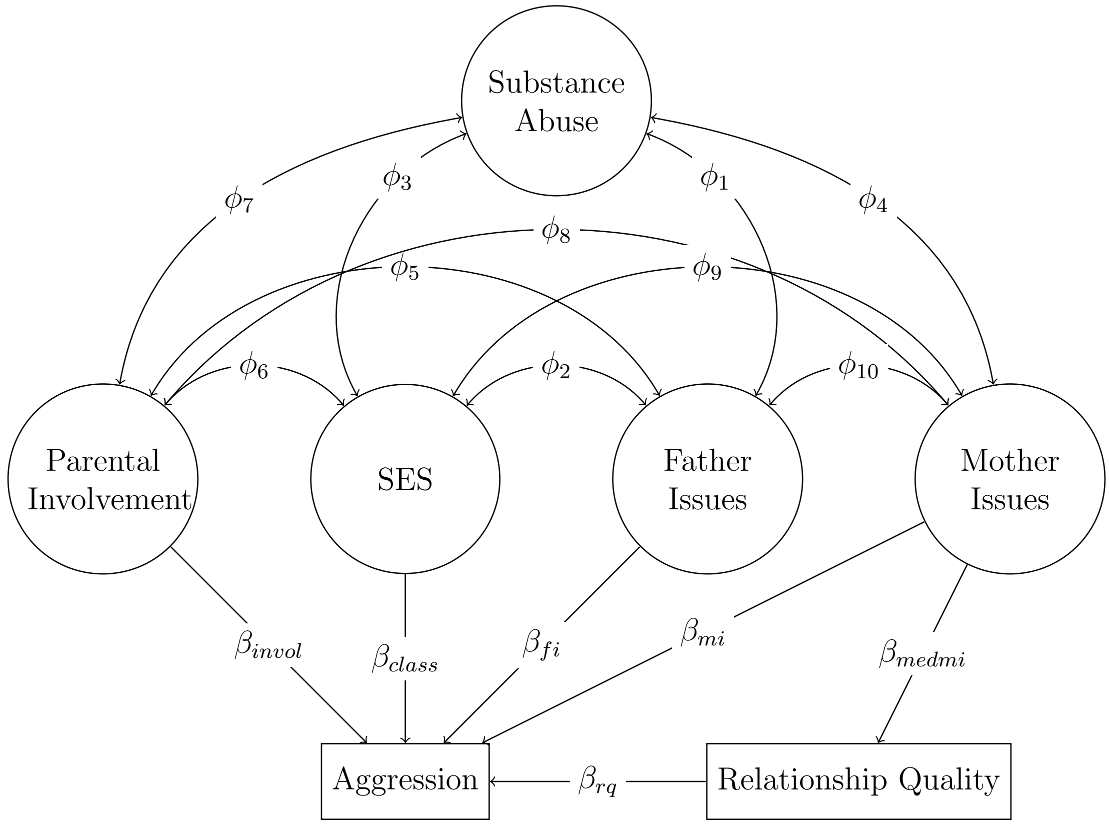

Figure 3.3: SEM of Parental Practices and Class on Aggression.

The reduced model shown in figure 3.4 is used to determine if there is a mediational relationship between mother issues and relationship. Providing a reduced model allows for discussion of model comparison methods of Frequentist and Bayesian SEM's. This will be done using information criteria: AIC, CAIC, and BIC for the Frequentist models and DIC for the Bayesian approach.

The Frequentist SEM model will be estimated using SAS, while the Bayesian SEM will be carried out in WinBUGS.

Prior Elicitation

Although some conceptual knowledge exists regarding the variables of interest, not much is known with respect to the magnitude of effects or the nature of covariances proposed in this

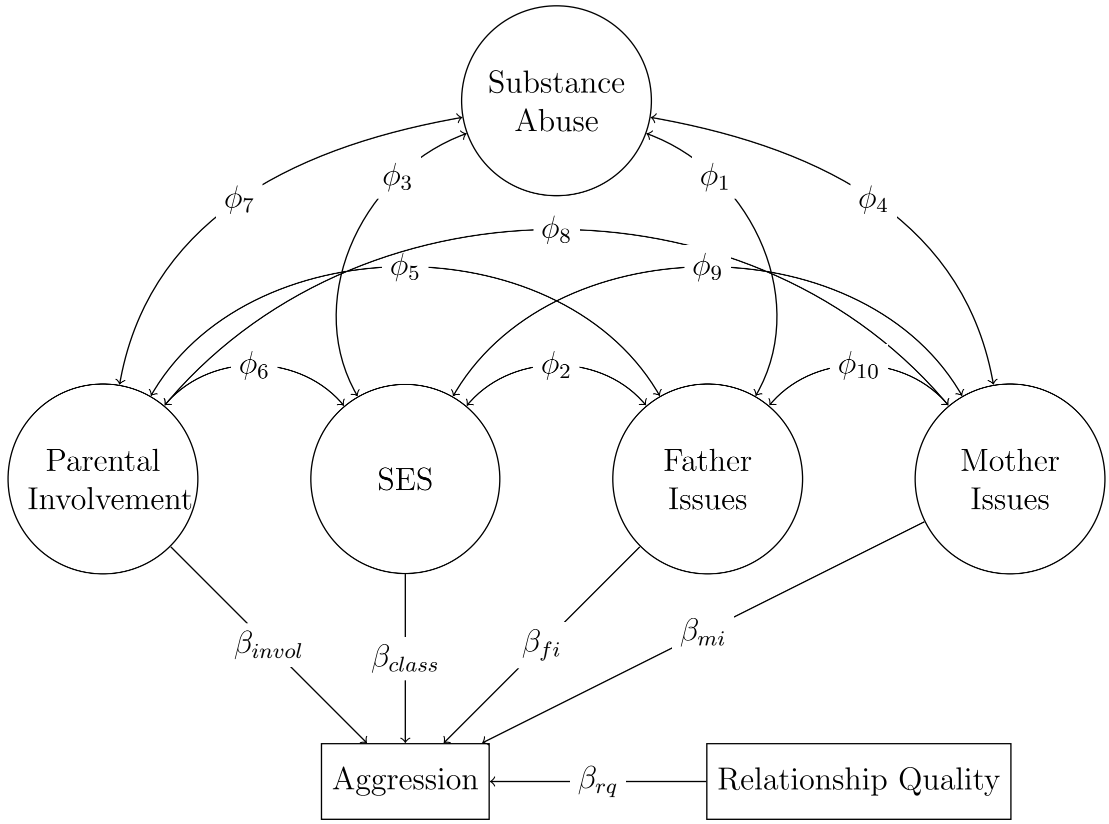

Figure 3.4: SEM of Parental Practices and Class on Aggression.

research. Consequently, prior specifications shown in table 3.3 assume an expected value of zero for the slope effects and covariances, which is a somewhat non-informative approach. Overly disperse variances are not used for the prior distributions in order to avoid improper posteriors. The sample size for this data set (n=4,898) is likely large enough to wash out much of the influence from the prior distributions.

Table 3.3: Priors for Model Parameters

Parameter	Prior Dist
$\lambda*$	N(0,prec=0.1)
ψ_{1-19}	Gamma(2,rate=0.1)
ξ	$MVN_5(\mathbf{u}, \mathbf{\Phi}^{-1})$
u_i	N(0,0.1)
$\mathbf{\Phi}^{-1}$	$Wish(\mathbf{I}_5, 5)$
$\beta*$	N(0,prec=0.1)

Prior distributions for factor loadings, notated using $\lambda*$, were all relatively non-informative. Error precisions for the measurement equations are represented with ψ and are given large gamma priors. Error variances are expected to be small so the priors placed on these precisions are large. The five latent variables for this model are assumed to come from the same multivariate normal distribution $\boldsymbol{\xi}$ with precision matrix Φ^{-1}. That precision matrix was given a Wishart prior with the identity matrix and 5 as the initial values for the parameters. Lastly, $\beta*$, notating all linear slope estimates, are given normal non-informative priors.

The logic behind the prior distributions for the reduced model is almost identical. The only change for the reduced model is that an effect for the mediational relationship is not estimated where priors for $\beta*$ remain the same.

CHAPTER 4

RESULTS

4.1 Full Models

Frequentist Full Model

The full model specified in Figure 3.3 was applied to the aggression outcome variable. Estimating this model using SAS provided the effect slopes in Table 4.1. Coefficients for this SEM regression can be interpreted similarly to a standard regression model where positive coefficients correspond to positive effects on aggression exhibited in the child's behavior adjusted for the predictors. However, some main effects appear counterintuitive when accounting for the presence of the mediational relationship of mother issues.

Unstandardized factor loadings from the measurement equations provide reasonable estimates. Father's depression, alcohol, and drug use are all significant and load positively onto the father issues latent variables. Loadings for the mother differ from the father in that only depression loads significantly onto the mother issues variables, likely due to the lack of information in mother's alcohol and drug use. Mother involvement loads positively onto parental involvement and all substance abuse loadings are positive, meaning that these variables share a positive relationship with the manifest variables constrained to be one. For the latent socio-economic variable both father's employment and father's education are positively associated with income-to-needs ratio for the household, and the proportion of families in the neighborhood who live in poverty and are unemployed have negative loadings. Relationships among the manifest variable as described by the respective factor loadings convey what would be expected and make good sociological sense.

Table 4.1: Estimates for Aggression SEM.

Parameter		Estimation Method					
		Frequentist		Bayes			
Type	Label	Est.	p-value	Est.	2.5%	97.5%	dep. fac
Loading	λ_{fi1}	2.0296	0.0001	1.7601	1.5650	1.9730	1.2100
	λ_{fi2}	0.6068	0.0001	0.5567	0.4777	0.6407	1.0800
	λ_{fi3}	0.3537	0.0001	0.3082	0.2510	0.3703	1.0200
	λ_{fi4}	1	—	1	—	—	—
	λ_{mi1}	-0.0404	0.7224	0.0005	-0.0915	0.0932	1.4400
	λ_{mi2}	2.8443	0.0001	1.3923	1.2230	1.5710	1.0100
	λ_{mi3}	-0.1297	0.4002	-0.0620	-0.2028	0.0752	2.4200
	λ_{mi4}	1	—	1	—	—	—
	λ_{minvol}	1.9195	0.0001	1.9115	1.8410	1.9900	8.3800
	λ_{finvol}	1	—	1	—	—	—
	λ_{sa1}	0.1154	0.0504	0.1300	0.0584	0.2055	1.0100
	λ_{sa2}	0.8728	0.0001	0.4447	0.2973	0.5932	3.0000
	λ_{sa3}	1.4257	0.0001	0.6350	0.4933	0.7800	4.1800
	λ_{sa4}	1	—	1	—	—	—
	λ_{class1}	0.0975	0.0001	0.0999	0.0831	0.1181	1.0400
	λ_{class2}	0.6706	0.0001	0.6852	0.5594	0.8199	1.0300
	λ_{class3}	-0.1533	0.0001	-0.1549	-0.1696	-0.1416	5.1800
	λ_{class4}	-0.0737	0.0001	-0.0752	-0.0823	-0.0689	4.2200
	λ_{Inc2}	1	—	1	—	—	—
Coeff.	β_{class}	-0.0357	0.0192	-0.0311	-0.0570	-0.0052	1.0300
	β_{fi}	-0.4643	0.0193	-0.0818	-0.3317	0.1705	1.0700
	β_{invol}	0.1929	0.0001	0.0582	0.0163	0.1007	1.1100
	β_{medmi}	-11.0551	0.0001	-4.9093	-5.2880	-4.5540	2.6500
	β_{mi}	31.2966	0.0006	6.3767	4.0960	8.9491	1.3300
	β_{rq}	1.6499	0.0026	0.6634	0.3624	1.0170	1.2600

Holding all else constant, the direct effect of the latent father issues variable, $\hat{\beta_{fi}}$, indicates a reduction in the level of aggression of the child as the father is increasingly troubled. Although this may seem odd, fathers who are increasingly troubled may be less active in the child's life, which is a good thing when the father is a bad influence. The effect of parental involvement lends credibility to this idea as it has a positive effect on aggression where it is more beneficial to the child when dysfunctional parents are less involved. Mother issues, which essentially represents mother depression and anxiety, according to the factor loadings, has a very large positive effect on child aggression and a substantial negative effect on relationship quality. Considering β_{medmi} and β_{rq} represents the indirect effect of mother issues on aggression. A mother's struggle with depression, anxiety, and substance abuse (including alcohol) has a double-edged effect: as mothers are increasingly affected by these problems a child tends to be more aggressive and relationship quality decreases. When comparing the direct effect of mother issues to the indirect effect on aggression, the two effect pathways are close in magnitude and may to some extent cancel each other out. The linear effect of the latent socio-economic class variable shows a subtle negative effect on child aggression, meaning that as economic conditions improve, aggression decreases. All covariates in this model are shown to be significant using asymptotically derived t-values.

The baseline χ^2-test for this model was very significant (p-value < 0.0001), which indicates some lack of fit as convention dictates the corresponding p-value should be around 0.3 or 0.4 for a good fit. Bentler's CFI for this model is 0.8684 and RMSEA is 0.0602 suggesting that this model provides a superior fit as compared to a baseline model. Although Bentler's CFI is slightly below the typical rule of thumb value of 0.9 and RMSEA slightly above 0.06, this could be due to the data set being particularly noisy as no two people are identical, genetically or environmentally, and the data come from a diversity of regions and social backgrounds.

Bayesian Full Model

The model from figure 3.3 was re-estimated using Bayesian methods with priors shown in table 3.3. Convergence diagnostics for these were done by examining trace plots along with checking dependence factors computed from the Raftery diagnostics. Trace plots of the effects from the structural equation are shown in figure 4.1 and show good exploration of the parameter space. Posterior distributions for these parameters are smooth and unimodal, showing good convergence.

Dependence factors are shown along with point estimates in table 4.1. The majority of which are less than five, except λ_{class3} and λ_{pinvol}. Trace plots for these two parameters were

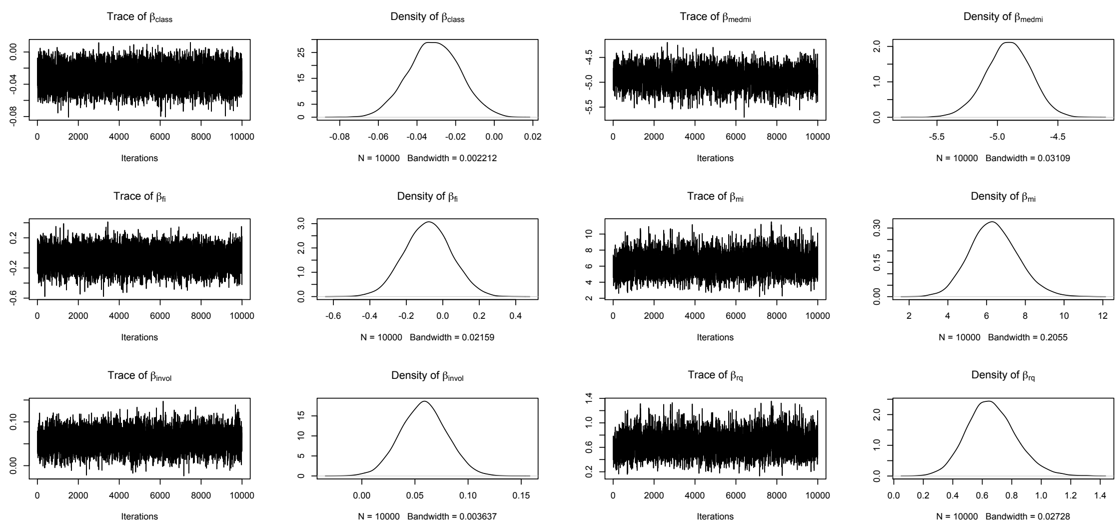

Figure 4.1: Trace Plots for Estimated Effects.

examined and demonstrated good exploration of the parameter space and smooth posteriors like those already shown for the slope coefficients.

Factor loadings for the manifest variables are very similar to those produced in the frequentist model and convey the same intuitive relationships among the manifest variables. One difference, however, is that while the trends are the same the loading for mother depression is nearly half the size in the Bayesian model.

Bayesian estimates for the full aggression model show the same overall trend in the effects as the frequentist model, however, all coefficients are shrunk towards zero. This may be due to influence of the priors as they are all centered at zero. Again, mother issues has a strong positive effect on aggression while exercising a pronounced negative influence on relationship quality. The trend where the direct and indirect effect cancel each other out occurs here, as well, which may be indicative of identifiability issues in the model. The magnitude of the estimates for mother issues differ between the Bayesian and frequentist paradigms, but the loading for mother depression is also larger for the frequentist model, which may account for this difference. Father issues is not a significant predictor, which disagrees with the frequentist model. When adjusting for the mediational relationship other covariates in the model show the same pattern where father issues reduce aggression and parental involvement increases it. Correlations between these estimates may explain why some of the effects behave as they do. Plots in figure 4.2 explore these correlations in the posterior draws for the full aggression model.

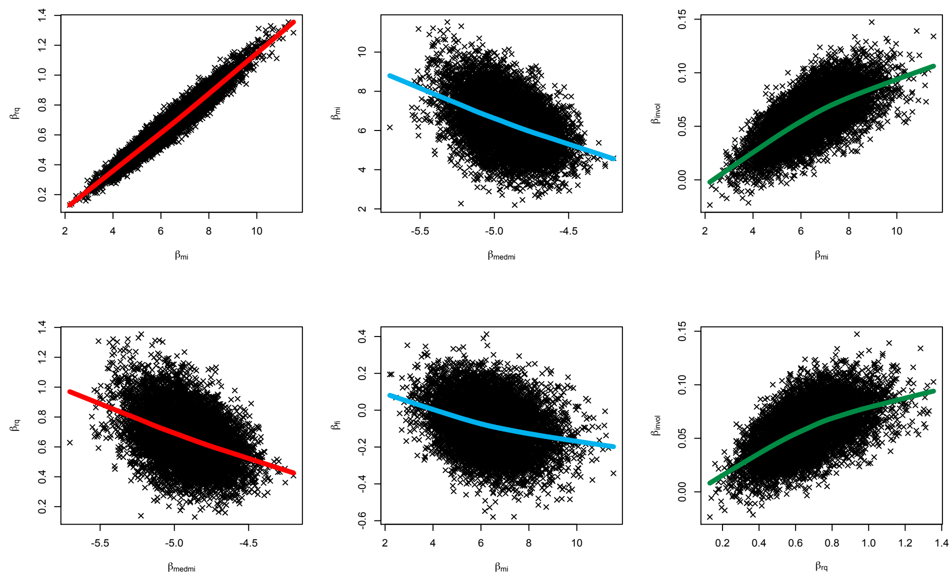

Figure 4.2: Bivariate Plots of Posterior Draws.

The plot in the top left corner of figure 4.2 shows a strong positive correlation between the effect of mother issues and the effect for relationship quality on aggression, which may seem unusual. Although unusual at first glance, it makes more sense when accounting for the other variables. A negative association exists between the direct effect of mother issues and its indirect effect; that indirect effect is negatively related to relationship quality. In light of these other relationships, it becomes more apparent why the effect of mother issues and relationship quality share a positive association. Another odd relationship exists between mother issues and parental involvement, but the effect of parental involvement has a positive association with relationship quality. Many of the effects in this model take on certain relationships when considering the indirect effect of mother issues which results in counterintuitive estimates for some of these explanatory variables.

Overall model performance can be assessed by looking at the residuals from the aggression regression equation. Plots shown in figure 4.3 show the fitted and observed densities for the aggression and relationship quality.

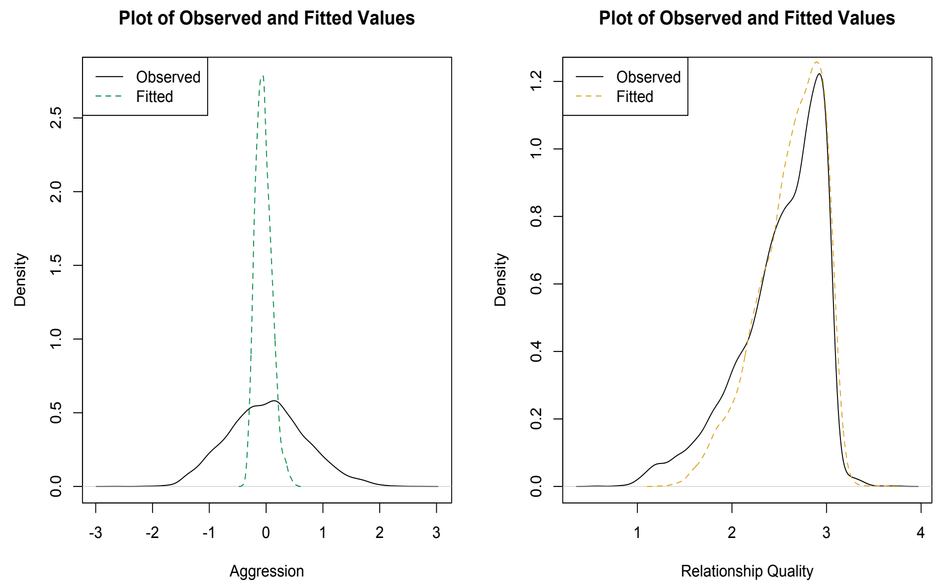

Figure 4.3: Observed versus Fitted Values for Aggression and Relationship Quality.

While the regression on relationship quality appears to provide a reasonable fit, aggression proves to be more problematic to predict. Because of the observational nature of the data set and that the sample comes from multiple regions of the United States, it is not surprising to find that the data turns out to be quite noisy. The structural equation for the full model tends to predict that children on average have a neutral aggression score with little variation around zero. This model lacks the ability to predict those with a greater or lesser propensity towards aggressive behavior, which is likely due to the lack of information in some of the manifest variables, and the noisy and chaotic nature of observational data.

4.2 Reduced Models

Reduced Frequentist Model

When the mediational relationship between mother issues and relationship quality is excluded from the model, estimates become more intuitive. Effects from both statistical paradigms are provided in table 4.2.

Table 4.2: Reduced SEM Model Estimates.

Parameter		Estimation Method					
		Frequentist		Bayes			
Type	Label	Est.	p-value	Est.	2.5%	97.5%	dep. fac
Loading	λ_{fi1}	2.1079	0.0001	1.7991	1.6050	2.0060	1.1600
	λ_{fi2}	0.6587	0.0001	0.5934	0.5126	0.6792	1.0600
	λ_{fi3}	0.3737	0.0001	0.3257	0.2662	0.3876	1.0100
	λ_{fi4}	1	—	1	—	—	—
	λ_{mi1}	0.1288	0.0871	0.0707	-0.0098	0.1514	2.5500
	λ_{mi2}	4.2044	0.0001	2.6462	2.2220	3.1270	2.8000
	λ_{mi3}	0.2812	0.0001	0.1865	0.0666	0.3003	3.0000
	λ_{mi4}	1	—	1	—	—	—
	λ_{minvol}	1.9195	0.0001	1.2803	1.1060	1.4800	4.6600
	λ_{finvol}	1	—	1	—	—	—
	λ_{sa1}	0.1019	0.0973	0.1386	0.0622	0.2192	1.0100
	λ_{sa2}	0.9424	0.0001	0.4962	0.3383	0.6532	4.2700
	λ_{sa3}	1.5703	0.0001	0.6771	0.5306	0.8218	3.7800
	λ_{sa4}	1	—	1	—	—	—
	λ_{class1}	0.0976	0.0001	0.1000	0.0826	0.1183	1.0700
	λ_{class2}	0.6693	0.0001	0.6804	0.5563	0.8113	1.0100
	λ_{class3}	-0.1532	0.0001	-0.1546	-0.1699	-0.1410	4.1700
	λ_{class4}	-0.0737	0.0001	-0.0751	-0.0824	-0.0686	3.1700
	λ_{Inc2}	1	—	1	—	—	—
Coeff.	β_{class}	-0.0232	0.0745	-0.0247	-0.0498	0.0010	1.0400
	β_{fi}	0.1588	0.2542	0.0705	-0.1721	0.3117	0.9820
	β_{invol}	-0.0452	0.0003	-0.0224	-0.0445	-0.0011	1.0700
	β_{mi}	1.1455	0.0001	0.8925	0.6305	1.1630	1.0500
	β_{rq}	-0.0884	0.0042	-0.1359	-0.1900	-0.0829	1.0800

In the reduced model, the latent construct has not changed but some differences do occur in the factor loadings. Where mother's drug use did not load significantly onto mother issues, it now does. Other than this difference, trends in the variables remain the same with mild differences in estimates for the factor loadings.

The effect of socio-economic class provides a smaller and marginally significant reduction in aggression, comparable to the full model. Effects for relationship quality, and parental involvement now behave as would be anticipated where both provide a significant reduction in child aggression. Coefficients for mother and father issues show that parents who struggle with mood disorders or substance abuse tend to have more aggressive children. Father's issues do not provide a significant effect, although positive in general, which makes sense because families inducted into the study are at risk and fathers are more likely not to be active in the child's life. Mothers most often being the primary caretaker in these more broken familial situations wield a much larger and more significant effect on child aggression.

The baseline χ^2-test showed to be very significant like the full model (p-value $<$ 0.0001) implying a similar lack of fit, again, likely due to all the noise in the data. However, goodness of fit indices meet the rule of thumb values with an RMSEA of 0.0517 and a Bentler's CFI of 0.9069.

Reduced Bayesian Model

Trace plots of the slope coefficients given in figure 4.4 show good convergence and exploration with smooth posterior densities for these parameters. All dependence factors for the parameters of interest are less than the rule of thumb value of five.

Bayesian parameter estimates again trend closer to zero as compared to the Frequentist model. Estimates for the effect of socio-economic class and relationship quality are both smaller meaning that the Bayes model estimates that these two covariates have a greater negative influence on aggression. Overall trends between the frequentist and Bayesian models are consistent with disagreement occurring only in terms of the magnitude of the effects and factor loadings. Both methods also agree about what terms are significant in the model with

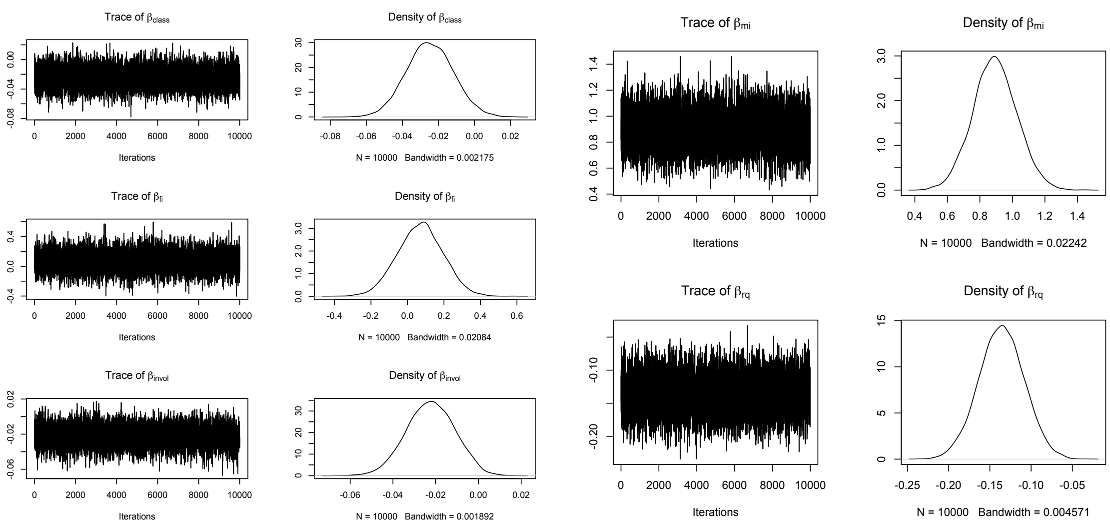

Figure 4.4: Trace Plots for Reduced Model Effects.

the exception of λ_{sa1}. The Bayes model says zero is not contained in the 95% credible interval where maximum likelihood states the loading is only marginally significant (p=0.0973).

Model fit in terms of the fit on aggression appears similar to that of the full model. Figure 4.5 shows the fitted density for the reduced model with the observed distribution of aggression; the right panel shows the fitted values plotted against observed values which generally is a linear relationship.

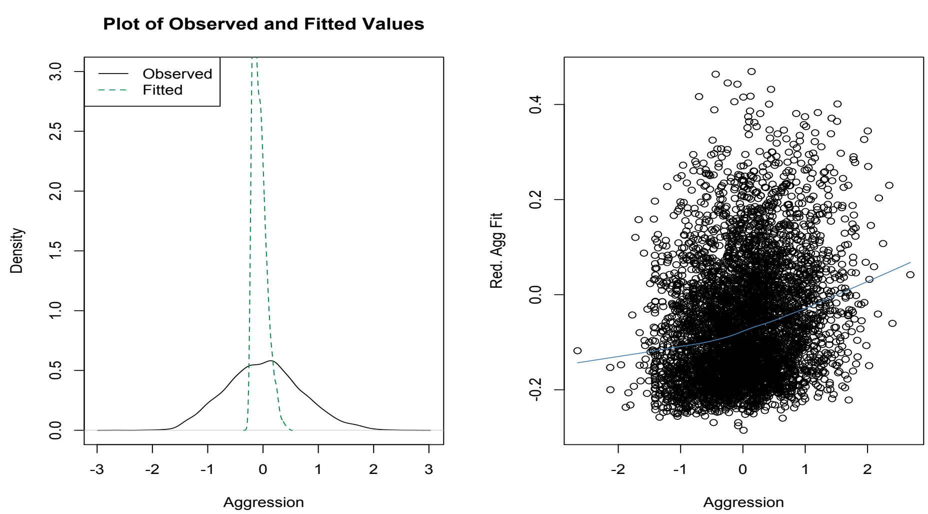

Figure 4.5: Observed Versus Fitted for Reduced Aggression.

Noise in the data is apparent particularly in plotting the fitted and observed data points where the trend would be nearly impossible to decipher without the spline. This trend is difficult to see because the data are so noisy and make observed values difficult to predict. Neither full nor reduced models show apparent differences in prediction performance. Model selection criteria need to be used in order to choose between the full and reduced models.

Model Comparison

Goodness of fit indices such as Bentler's CFI and RMSEA could be used for model selection purposes, but in order to have comparable selection methods between the two statistical paradigms different information criteria specific to those paradigms will be used. SAS provides AIC, CAIC, and BIC, and WinBUGS gives DIC in terms of each response variable where the DIC provided here is with respect to the aggression response variable.

The only real difference between the full and reduced models is the mediational relationship with an indirect effect of mother issues mediated by relationship quality. Even though the difference amounts to two coefficients and an error term, accounting for this extra term affects the rest of the model and adds an additional layer of complexity.

Table 4.3: Table of Information Criteria Comparing Full vs. Reduced.

Information Criteria	Full	Reduced
AIC	2015.5365	1490.3006
CAIC	2397.8622	1902.6126
BIC	2346.8622	1847.6126
DIC	9969.970	10043.100

The three information criteria for the frequentist models unanimously select the reduced model as the one with the better fit, which agrees with the goodness of fit criteria mentioned earlier. Bayesian Deviance Information Criterion, however, selects the full model as the better model for these data. Despite the fact that overall trends between the two methods agree they come to different conclusions with respect to model selection. Although estimates were generally close between the two models, coefficients for mother issues in the full frequentist model appeared unusually large when compared to their Bayesian counterparts. Estimates for β_{medmi} were more than twice as large and β_{mi} nearly five times as large. As has already been mentioned, the factor loading on mother's depression is twice as large in the frequentist model and may account for some of the difference in magnitude between the two methods. Another potential source of the aforementioned disparity could be due to limitations of the data. Manifest loadings involved with the parental variables are binary and most do not contain very much information. Given the complexity of the model and the additional burden placed on these binary variables when estimating a mediational relationship, estimates may have become unstable which would account for the seeming inflated estimates of β_{medmi}, and β_{mi}. Bayesian estimates were not as affected by the identifiability problem because of the added stability from the information contained in the prior distributions. All priors for loadings and slope coefficients were centered at zero with a precision of 0.1 or a standard deviation of 3.1623. It is possible that due to the inflated estimates for some coefficients that the model fit, and consequently, information and fit criteria also suffered in the frequentist model. Problems may also arise from estimated variances for the manifest variables being close to zero which can affect maximum likelihood estimation, particularly because the sample covariance matrix is used to produce said estimates.

The disagreement in model selection between the two paradigms may be a product of the limited information in the dataset, but restrictive priors also influence how extreme Bayesian estimates will go and could also account for the disparity between the two methods.

Sensitivity Analysis

To test if restrictive priors are what is primarily responsible for the more attenuated Bayesian estimates, particularly for the differences found in the direct and indirect effects of mother issues, a quick sensitivity analysis was conducted using another Bayesian full model which was estimated where priors for all loadings and slope coefficients are given ten times the variance. In this case, priors for slope coefficients and factor loadings are $N(0, 100)$, which have ten times the variances as the previous priors.

Table 4.4: Full Model Estimates from Frequentist, Bayesian, and Disperse Bayesian Models.

Parameter		Estimation Method		
		Frequentist	Bayes	
Type	Label	Est.	Disp. Mod.	Full Mod
Loading	λ_{fi1}	2.0296	1.7635	1.7601
	λ_{fi2}	0.6068	0.5563	0.5567
	λ_{fi3}	0.3537	0.3081	0.3082
	λ_{mi1}	-0.0404	-0.0044	0.0005
	λ_{mi2}	2.8443	1.3940	1.3923
	λ_{mi3}	-0.1297	-0.0686	-0.0620
	λ_{minvol}	1.9195	1.9117	1.9115
	λ_{sa1}	0.1154	0.1307	0.1300
	λ_{sa2}	0.8728	0.4472	0.4447
	λ_{sa3}	1.4257	0.6377	0.6350
	λ_{class1}	0.0975	0.0998	0.0999
	λ_{class2}	0.6706	0.6831	0.6852
	λ_{class3}	-0.1533	-0.1548	-0.1549
	λ_{class4}	-0.0737	-0.0752	-0.0752
Coeff.	β_{class}	-0.0357	-0.0316	-0.0311
	β_{fi}	-0.4643	-0.1156	-0.0818
	β_{invol}	0.1929	0.0720	0.0582
	β_{medmi}	-11.0551	-5.0080	-4.9093
	β_{mi}	31.2966	7.7091	6.3767
	β_{rq}	1.6499	0.8398	0.6634

Frequentist and Bayesian estimates provided in Table 4.4 are the same as those in Table 4.1 and are included for convenience in comparing them to those produced in the disperse prior model. Dependence factors show the same good convergence as the initial full model. DIC when using more disperse priors shows a better fit than the first full model with a value of 9920.090.

The largest difference in this new disperse prior model is the increase in β_{mi} where the coefficient increased by 1.3. Other differences can be noted in β_{fi}, β_{pinvol}, and β_{rq}. Although some coefficients saw an increase in magnitude, they do not differ significantly from what the initial Bayesian full model had provided and are contained in those 95%

credible intervals. Restrictive priors are not responsible then for the large disparity between the maximum likelihood and Bayesian estimates. Reasons for the disparity between Bayesian and Frequentist estimates are most likely coming from the already mentioned shortcomings inherent in the data.

CHAPTER 5

DISCUSSION

This data set from the Fragile Families and Child Well-being study presents many challenges to the researcher. Not only does the data not meet the assumption of multivariate normality, but many of the manifest variables contain little information with low prevalence of respondents with a certain trait. It is not clear as to which statistical method appears to be more robust to the violation of multivariate normality. Binary variables contain less information than continuous variables as there is no continuous scale. Creating continuous latent variables from binary loadings does not solve issues stemming from lack of information as they are passed on to those resultant latent variables. Even though the sample size was near 5,000, estimates for mother issues saw some instability and limitations when trying to model a mediational relationship. Likely, because the mother issues variable provides fewer degrees of freedom or less information than is required to estimate the more complex full model and better identify the unique contribution of mother issues to aggression via both direct and indirect pathways. Bayesian models are able to borrow strength from the prior distributions and as unassuming, somewhat non-informative priors were used, estimates saw a shrinkage effect but appeared to be more stable in terms of magnitude for the more complex full model. Despite the fact that prior distributions can provide greater stability when confronted with smaller sample sizes or scarce information, it is still undetermined whether the Bayesian model proved to be more robust to the violation of distributional assumptions.

Full models estimated from both methods had identifiability concerns regarding the direct and indirect effects of mother issues. The magnitude of the coefficients provided by both statistical methods were larger than the range of the response variable 'aggression', which ranged approximately from negative three to three. In addition to the aforementioned lack of information and distributional violations, multicollinearity may also be an issue when attempting to identify the unique effect pathways of mother issues. Although DIC would suggest a better fit with the full model and the p-value for β_{medmi} would corroborate that, estimates when considering the more complex relationship become unstable and lose interpretability. Although Information criteria and p-values may suggest a certain relationship, caution needs to be exercised as the data, in this case, are not able to confidently bear out a more complex mediational model. Slope estimates in the reduced model appear reasonable given the range of aggression and consequently, retain their interpretability. Model parameters in the reduced model are more practical and stable. DIC did not convey this and although frequentist information criteria did prefer the reduced model, mixed messages were given when examining the p-values for the indirect effects. Identifiability issues in the full model does not mean that the effect of mother issues is not potentially mediated by relationship quality, but that statistical issues and data limitations prevent one from confidently making conclusions despite seemingly decisive estimates from the full model.

Regardless of which statistical paradigm was used for this analysis, overall trends in either the Bayesian or frequentist models were in concordance. Disagreements between the two methods came into play during the model selection process. According to AIC, CAIC, and BIC, the reduced model was the more appropriate model, whereas DIC selected the full model. Differences in model selection may be a consequence of inflated estimates produced by the frequentist model, but would require further study to confirm this. It is worth mentioning that the frequentist full model suggested that all coefficients were significant, providing evidence of identifiability concerns. Depending on which statistical method is used, a researcher would come to different conclusions about how a mother's problems with mood disorders or substance abuse influence child aggression. The frequentist approach to this model suggests that relationship quality is not necessarily mediated by the mother's problems and that parental involvement, relationship quality of the couple, and mother issues are significant predictors of child aggression. Socio-economic class may be influential but not as significant as the aforementioned covariates. Bayesian methods suggest that mother issues is the foremost contributor to child aggression and indirectly affects aggression by negatively influencing the relationship quality of the couple.

Many practical reasons exist for choosing to use either a Bayesian or Frequentist approach to structural equation modeling. Computational costs of estimating a Bayesian SEM are far greater than that associated with their counterparts. Specific limitations with the WinBUGS software packages do not allow the researcher to monitor large numbers of parameters. This became problematic when attempting to get posterior distribution for draws of the latent variables and the residuals associated with each latent regression. The time investment for the Bayesian SEM model is considerable, but does provide more information by giving the posterior distributions of all parameters including latent variables. Issues with the model not apparent in parameter estimates can often be diagnosed using this additional information (Dunson et al. 2005). Frequentist SEM's are well established and offer a slew of goodness-of-fit indices and packages that can estimate them. In this research, neither method appears more robust to the problems presented by these data when estimating a LISREL model and more research is needed. Bayesian methods do offer a greater flexibility when selecting the data likelihood, but distributional assumptions were kept the same between the two paradigms to allow for a more straightforward comparison.

5.1 Child Outcomes

From the estimated models, the most detrimental effect on child aggression comes from mothers when they are struggling with mood disorders and substance abuse problems. Parents, particularly mothers, who struggle with mood disorders or substance abuse of one form or another significantly contributed to a child's problematic behavior. These problematic behaviors also undermine the effect of relationship quality implying that they also are harmful towards a harmonious and supportive relationship. Individuals under the influence of illicit substances or alcohol can be more unpredictable and even abusive making it more difficult for growing children to trust or form emotional attachments. Aggressive behaviors can also be learned while parents are under the influence. The effect of problematic behaviors like substance abuse are two-fold because they also destabilize parental relationships providing a more hostile and unstable family environment which is consistent with Ackermann's findings (2002). While the effect of substance abuse, anxiety, and depression are mediated by relationship quality, it more likely that those struggling with substance abuse and mood disorders have more unstable and contentious relationships. Children raised in abusive family environments are likely to repeat those same behaviors when they have families of their own (Egeland 1993). While it is likely that fathers dealing with these same issues affect a child's development as suggest by Egeland, the FFCW data comes from at risk families where fathers are more likely to be uninvolved or negligent. The effect of fathers does not come through as significant for this analysis, but this may be a feature of this particular data set and does not infer that fathers do not play a significant role in the comportment of their children.

The effect of class or economic circumstances also played a role in child behaviors. Monetary shortcomings and problems can be a major source of anxiety and stressful for any household facing rough economic times, potentially exacerbating anxiety, depression, or even substance abuse problems a parent may be facing. The effect of socio-economic status on aggression may be due to exposure to bad influences in rougher neighborhoods or that dysfunctional behaviors are prevalent in more impoverished communities. Additionally, families in more disadvantaged circumstances may also lack resources and knowledge, or access thereto, should a child begin to demonstrate a propensity towards aggressive or antisocial conduct.

5.2 Future Research

Data from observational studies tend to be noisy, particularly within social studies as no two people are identical in genetic traits, environments, or personality. Data for this study is no exception as fitted values for the estimated models failed to account for much of the variance in aggression (shown in figures 4.3 and 4.5). Although much of noise may be due to the inability to carry out a true experimental design, the human experience and real life are for more complex than the models presented in this project. It is likely that there are more covariates that can be added and other, perhaps more complex, latent constructs that may better capture sociological trends and human behavior.

Estimating the measurement equation in structural equation modeling assumes the multivariate normality of the manifest covariates. Nine of which were indicator variables. One of the advantages of Bayesian analysis is greater flexibility in the selection of the data likelihood. For this project a Bernoulli likelihood would have been more appropriate but was not used because of how that would affect the latent variables. Structural equation models using dichotomous variables and exponential families would be more appropriate for this data (Lee 2007).

One of the great shortcoming of these data was the lack of information contained in some of the manifest variables, specifically those addressing substance abuse problems. Many respondents may not admit to abusing alcohol or illicit substances due to the ignominious nature of the behavior. Because of this, respondents who in reality abuse alcohol and/or illicit drugs may deny having done so. Some of the lack of information (low prevalence of those admitting to substance abuse) in the manifest variables involving substance abuse could be attributed to this issue. Structural equation models have exogenous error terms for each manifest variable but this does not account for misrepresentation of data. Profitable research may be done in getting accurate responses from loaded questions.

Analyses for this project focused on the effect of parental variables on children, but research has shown that boys often are affected by adversity and traumatic experiences differently than girls (Shaffer 2005). Future directions could estimate the effects of parental behaviors on boys and girls separately to understand how the different genders are affected and react to difficult familial difficulties.

Trends shown in this project could change depending on the response variables. Models can be for other child well-being outcomes such as anxiety, health, or attention deficit disorder, also included in the FFCW dataset.

BIBLIOGRAPHY

Ackerman, B., Brown, E., Schoff D'Eramo, K., and Izard, C. (2002), "Maternal relationship instability and school behavior of children from disadvantaged families," *Developmental Psychology*, 38, 694–704.

Ackerman, B., Kogos, J., Youngstrom, E., Schoff, K., and Izard, C. (1999), "Family instability and the problem behaviors of children from economically disadvantaged families," *Developmental Psychology*, 35, 258–268.

Ainsworth, M. (1989), "Attachments beyond infancy," *American Psychologist*, 38, 694–704.

Berger, J. O., and Berry, D. A. (1988), "Statistical Analysis and the Illusion of Objectivity," *American Scientist*, 376, 159–165.

Chesher, A. (1991), "The Effect of Measurement Error," *Biometrika*, 78, 451–462.

Dunson, D. B., Palomo, J., and Bollen, K. (2005), "Bayesian Structural Equation Modeling," Tech. Rep. 2005-5, Statistical and Applied Mathematical Sciences Institute (SAMSI), Computer Science Department, Fanstord, California.

Egeland, B. (1993), *A history of abuse is a major risk factor for abusing the next generation* (1st ed., Current Controversies on Family Violence), Thoasand Oaks, CA: Sage.

Kitzmann, K. (2000), "Effects of marital conflict on subsequent triadic family interactions and parenting," *Developmental Psychology*, 36, 3–14.

Lee, S.-Y. (2007), *Structural Equation Modeling: A Bayesian Approach*, West Sussex, England: Wiley.

Raykov, T., and Marcoulides, G. (2000), *A First Course In Structural Equation Modeling*, Mahwah, NJ: Lawrence Erlbaum Associates.

Shaffer, D. R. (2005), *Social and Personality Development* (5th ed.), Belmont, CA: Thomson Wadsworth.

APPENDICES

APPENDIX A

WINBUGS CODE

```
model Fullmod{
for(i in 1:4898){

#1=MAlcho 2=MAnx 3=MDep 4=MDrug 5=FAnx
#6=FDep 7=FAlcho 8=FDrug 9=MInvol 10=FInvol
#11=Femp 12=DadEd 13=Inc2Needs 14=PBelPov 15=PUnemp

#measurement equation model
for(j in 1:15){
y[i,j]~dnorm(mu[i,j],psi[j])
#ephat[i,j]<-y[i,j]-mu[i,j]
}
mu[i,9]<-lamp*xi[i,1]+alp[1]
mu[i,10]<-xi[i,1]+alp[2]
mu[i,1]<-lamm[1]*xi[i,2]+lamsa[3]*xi[i,5]+alp[3]
mu[i,2]<-xi[i,2]+alp[4]
mu[i,3]<-lamm[2]*xi[i,2]+alp[5]
mu[i,4]<-lamm[3]*xi[i,2]+xi[i,5]+alp[6]
mu[i,5]<-xi[i,3]+alp[7]
mu[i,6]<-lamf[1]*xi[i,3]+alp[8]
mu[i,7]<-lamf[2]*xi[i,3]+lamsa[2]*xi[i,5]+alp[9]
mu[i,8]<-lamf[3]*xi[i,3]+lamsa[1]*xi[i,5]+alp[10]
mu[i,11]<-lamc[1]*xi[i,4]+alp[11]
mu[i,12]<-lamc[2]*xi[i,4]+alp[12]
mu[i,13]<-xi[i,4]+alp[13]
mu[i,14]<-lamc[3]*xi[i,4]+alp[14]
mu[i,15]<-lamc[4]*xi[i,4]+alp[15]

#structural equation model
xi[i,1:5]~dmnorm(u[1:5],phi[1:5,1:5])
Aggress[i]~dnorm(Amu[i],ysd)
Amu[i]<-int[1]+bclass*xi[i,4]+binvol*xi[i,1]+brq*RelQual[i]+bmi*xi[i,2]+bfi*xi[i,3]
RelQual[i]~dnorm(nu[i],psd)
nu[i]<-int[2]+bmedmi*xi[i,2]
drq[i]<-RelQual[i]-nu[i]
dag[i]<-Aggress[i]-Amu[i]
} #end of i

for(i in 1:5){u[i]~dnorm(0,0.1)}

#priors on intercepts
for(j in 1:15){alp[j]~dnorm(0,0.1)}
for(k in 1:2){int[k]~dnorm(0,0.1)}

#priors on loadings and coefficients
lamp~dnorm(0,0.1)
lamm[1]~dnorm(0,0.1)
lamm[2]~dnorm(0,0.1)
lamm[3]~dnorm(0,0.1)
lamf[1]~dnorm(0,0.1)
lamf[2]~dnorm(0,0.1)
lamf[3]~dnorm(0,0.1)
lamc[1]~dnorm(0,0.1)
lamc[2]~dnorm(0,0.1)
lamc[3]~dnorm(0,0.1)
lamc[4]~dnorm(0,0.1)
```

```
lamsa[1]~dnorm(0,0.1)
lamsa[2]~dnorm(0,0.1)
lamsa[3]~dnorm(0,0.1)

bclass~dnorm(0,0.1)
binvol~dnorm(0,0.1)
brq~dnorm(0,0.1)
bmi~dnorm(0,0.1)
bfi~dnorm(0,0.1)
bmedmi~dnorm(0,0.1)

#priors on precisions
for(j in 1:15){psi[j]~dgamma(2,0.05)}

psd~dgamma(2,0.1)
#psd<-1/relvar
ysd~dgamma(2,0.1)
#ysd<-1/aggvar
phi[1:5,1:5]~dwish(R[1:5,1:5],5);
phcov[1:5,1:5]<-inverse(phi[1:5,1:5])

} #end of model

model Redmod{
for(i in 1:4898){

#1=MAlcho 2=MAnx 3=MDep 4=MDrug 5=FAnx
#6=FDep 7=FAlcho 8=FDrug 9=MInvol 10=FInvol
#11=Femp 12=DadEd 13=Inc2Needs 14=PBelPov 15=PUnemp

#measurement equation model
for(j in 1:15){
y[i,j]~dnorm(mu[i,j],psi[j])
#ephat[i,j]<-y[i,j]-mu[i,j]
}
mu[i,9]<-lamp*xi[i,1]+alp[1]
mu[i,10]<-xi[i,1]+alp[2]
mu[i,1]<-lamm[1]*xi[i,2]+lamsa[3]*xi[i,5]+alp[3]
mu[i,2]<-xi[i,2]+alp[4]
mu[i,3]<-lamm[2]*xi[i,2]+alp[5]
mu[i,4]<-lamm[3]*xi[i,2]+xi[i,5]+alp[6]
mu[i,5]<-xi[i,3]+alp[7]
mu[i,6]<-lamf[1]*xi[i,3]+alp[8]
mu[i,7]<-lamf[2]*xi[i,3]+lamsa[2]*xi[i,5]+alp[9]
mu[i,8]<-lamf[3]*xi[i,3]+lamsa[1]*xi[i,5]+alp[10]
mu[i,11]<-lamc[1]*xi[i,4]+alp[11]
mu[i,12]<-lamc[2]*xi[i,4]+alp[12]
mu[i,13]<-xi[i,4]+alp[13]
mu[i,14]<-lamc[3]*xi[i,4]+alp[14]
mu[i,15]<-lamc[4]*xi[i,4]+alp[15]

#structural equation model
xi[i,1:5]~dmnorm(u[1:5],phi[1:5,1:5])
Aggress[i]~dnorm(Amu[i],ysd)
Amu[i]<-int+bclass*xi[i,4]+binvol*xi[i,1]+brq*RelQual[i]+bmi*xi[i,2]+bfi*xi[i,3]
#RelQual[i]~dnorm(nu[i],psd)
#nu[i]<-int[2]+bmedmi*xi[i,2]
#drq[i]<-RelQual[i]-nu[i]
dag[i]<-Aggress[i]-Amu[i]
} #end of i

for(i in 1:5){u[i]~dnorm(0,0.1)}

#priors on intercepts
for(j in 1:15){alp[j]~dnorm(0,0.1)}
int~dnorm(0,0.1)

#priors on loadings and coefficients
```

```
lamp~dnorm(0,0.1)
lamm[1]~dnorm(0,0.1)
lamm[2]~dnorm(0,0.1)
lamm[3]~dnorm(0,0.1)
lamf[1]~dnorm(0,0.1)
lamf[2]~dnorm(0,0.1)
lamf[3]~dnorm(0,0.1)
lamc[1]~dnorm(0,0.1)
lamc[2]~dnorm(0,0.1)
lamc[3]~dnorm(0,0.1)
lamc[4]~dnorm(0,0.1)
lamsa[1]~dnorm(0,0.1)
lamsa[2]~dnorm(0,0.1)
lamsa[3]~dnorm(0,0.1)

bclass~dnorm(0,0.1)
binvol~dnorm(0,0.1)
brq~dnorm(0,0.1)
bmi~dnorm(0,0.1)
bfi~dnorm(0,0.1)

#priors on precisions
for(j in 1:15){psi[j]~dgamma(2,0.05)}

psd~dgamma(2,0.1)
#psd<-1/relvar
ysd~dgamma(2,0.1)
#ysd<-1/aggvar
phi[1:5,1:5]~dwish(R[1:5,1:5],5);
phcov[1:5,1:5]<-inverse(phi[1:5,1:5])

} #end of model
```

APPENDIX B

SAS CODE

```
data proj;
infile 'C:\udrive\research\pyper\MProj_Dat.csv' delimiter=',' firstobs=2;
input idnum male Vagg Vanx health Vattprb MAlcho MInvolv MAnx MDep MDrug relQual Dincar1
Dincar2 DAnx DDep DInvolv DAlcho DDrug everincar Demp DedT1 DadEd Inc2pvty Inc2need
Pbpov Punemp;
run;

ods graphics on;
title 'Full Aggression SEM Model';
proc calis method=ml data=proj maxiter=5000 outstat=fulaggmod plots=(all residual);
   lineqs
DInvolv = Fpinvol + e1,
MInvolv = lamp1 Fpinvol + e2,
DDep = lamf1 Fissue + e3,
DAnx = Fissue + e4,
Ddrug = lamf2 Fissue + lamd1 Fsabuse + e5,
DAlcho = lamf3 Fissue + lamd2 Fsabuse+ e6,
MDep = lamm1 Fmissue + e7,
MAnx = Fmissue + e8,
MDrug = lamm2 Fmissue + Fsabuse + e9,
MAlcho = lamm3 Fmissue + lamd3 Fsabuse + e10,
Demp = lam1 Fclass + e11,
DadEd = lam2 Fclass + e12,
Inc2need =  Fclass + e13,
Pbpov = lam4 Fclass + e14,
Punemp = lam5 Fclass + e15,
relQual = b6 Fmissue + e16,
Vagg = b1 Fissue + b2 Fmissue + b4 Fclass + b5 Fpinvol + b3 relQual + e17;
  std
e1-e17 = psi1-psi17,
Fpinvol Fclass Fissue Fmissue Fsabuse= phi1 phi2 phi3 phi4 phi5;
cov
/*Fpinvol Fissue = phi13, SAS seems to do all covariances parameters automatically
Fpinvol Fmissue = phi14,*/
Fsabuse Fpinvol = phi15,
Fissue Fsabuse = phi35,
Fmissue Fsabuse = phi45;
bounds
0<=phi1 phi2 phi3 phi4 phi5,
0<=psi1-psi17;
run;
ods graphics off;

title 'Alt reduced Aggression SEM Model';
proc calis method=ml data=proj maxiter=5000 outstat=redaggmod plots=(all residual);
   lineqs
DInvolv = Fpinvol + e1,
MInvolv = lamp1 Fpinvol + e2,
DDep = lamf1 Fissue + e3,
DAnx = Fissue + e4,
Ddrug = lamf2 Fissue + lamd1 Fsabuse + e5,
DAlcho = lamf3 Fissue + lamd2 Fsabuse+ e6,
MDep = lamm1 Fmissue + e7,
MAnx = Fmissue + e8,
MDrug = lamm2 Fmissue + Fsabuse + e9,
MAlcho = lamm3 Fmissue + lamd3 Fsabuse + e10,
Demp = lam1 Fclass + e11,
DadEd = lam2 Fclass + e12,
Inc2need =  Fclass + e13,
```

```
Pbpov = lam4 Fclass + e14,
Punemp = lam5 Fclass + e15,
Vagg = b1 Fissue + b2 Fmissue + b4 Fclass + b5 Fpinvol + b3 relQual + e16;
  std
e1-e16 = psi1-psi16,
Fpinvol Fclass Fissue Fmissue Fsabuse= phi1 phi2 phi3 phi4 phi5;
cov
/*Fpinvol Fissue = phi13, SAS seems to do all covariances parameters automatically
Fpinvol Fmissue = phi14,*/
Fsabuse Fpinvol = phi15,
Fissue Fsabuse = phi35,
Fmissue Fsabuse = phi45;
bounds
0<=phi1 phi2 phi3 phi4 phi5,
0<=psi1-psi16;
run;
```