# The Practice of Imputation Methods with Structural Equation Models

Cherie J. Alf * Michael D. Larsen † Frederick O. Lorenz ‡

**Abstract**
When using survey data, researchers must evaluate how to effectively handle missing data. For social survey data, full information maximum likelihood methods are often implemented when the researcher is interested in structural equation models. This strategy is convenient to implement and provides acceptable results. Yet it does not incorporate any imputation methods for assessing the missing information. We examine the benefits of imputation methods as an alternative for managing missing data, particularly in longitudinal surveys where missingness may be conditioned on previous panel data. Our goal is to outline the practical use of imputation and resulting gains in estimation. We apply these procedures to the Family Transitions Project, a longitudinal survey of more than 550 participants, which focuses on familial relationships and socioeconomic stress induced by economic hardships.

**Key Words:** Missing data, full information maximum likelihood, multiple imputation, longitudinal survey

## 1. Introduction

Longitudinal surveys are often plagued by missing information. Respondents are prone to refuse answering questions or completely miss waves of data collection. This missing data leads to difficulties in providing accurate analysis of the data. A researcher must not only make decisions about what kind of statistical analyses are of interest, but also how to effectively handle the missing information.

The purpose of this paper is to examine three different strategies for handling missing data. This work is motivated by concerns raised by researchers who work with panel data. In particular, our interest is in missing data methods when structural equation models are used in the statistical analysis. The three approaches assessed using the same structural equation model are listwise deletion, full information maximum likelihood, and multiple imputation.

This paper begins by providing some background information on the Family Transitions Project, the motivating data set for this analysis. Next a brief explanation of structural equation models and missing data are provided. This is followed by an overview of the three ways to handle missing data used in the subsequent analysis. Lastly a discussion of the results from the application to the Family Transitions Project and some concluding remarks are given.

## 2. Background

### 2.1 Family Transitions Project

The sample consists of participants in the Family Transitions Project (FTP), a longitudinal study of families in Iowa. The FTP encompasses two earlier projects, the Iowa Youth and Families Project and the Iowa Single Parent Project. The Iowa Youth and Families Project began in 1989 with a total of 451 families. The recruited families were located in rural areas of Iowa and were comprised of two biological parents, a $7^{th}$ grade child (the "target child"), and a sibling within 4 years of the target child. In 1991 the Iowa Single Parent Project began a study of 210 adolescents, their single-parent mothers, and a close-aged sibling. Many of the measures recorded in the two studies were the same. In 1994 these two

*Department of Statistics, Iowa State University
†Biostatistics Center, Department of Statistics, George Washington University
‡Institute of Social and Behavioral Research, Department of Statistics, Iowa State University

projects were combined to begin the FTP. Of the 210 families in the Single Parent Project, 108 of the targets were in $9^{th}$ grade in 1991 and eligible for the new study, resulting in 559 individuals in FTP (Conger and Conger, 2002).

The families included in the panel were contacted in 1991, 1992, 1994, and every other year since 1995. During the data collection process, multiple informants are asked to provide information. These informants include each of the family members participating in the study, trained observers who rate videotapes of family discussions in their home, and close friends and romantic partners of the target individual. Our focus is on the interactions between the targets and their mothers, close friends, and romantic partners. We specifically identify a measure of hostility between the targets and their mothers for the 1991-1994 surveys and between the targets and their partners for the 1995-2005 surveys.

We focus on the information provided from the ratings of the trained observers at the Institute for Social and Behavioral Research at Iowa State University. Using the video-recording of interactions, the trained observer is asked to "Use the following to rate each interaction for the given behaviors." For the measure of hostility used in the analysis, we sum the rating for the following five traits: hostility, angry coercion, escalate hostile, reciprocate hostile, and antisocial. Each of these behaviors is clearly defined by the FTP trained observers. The ratings are from 1 (low) to 9 (high).

### 2.2 Structural Equation Models

Many social science studies are conducted using a panel of individuals. To statistically analyze panel data, one must consider the relationship between data points for an individual and one approach to examining these relationships is by using structural equation models. A structural equation model accounts for the relationship between observed variables through the use of latent variables (Bollen, 1989). For longitudinal data sets researchers often consider growth curve models where the outcome of specific variables have been measured across time. Due to the manner in which the data are collected, these data points are often correlated through time. The growth curve model, used to estimate the rate of growth of the population over a specified time period, can be formulated as a structural equation model and is thus called a latent growth curve model.

In latent growth curve analysis all individuals are assumed to have the same functional form of the growth curves. The difference is in allowing the parameters describing their curves to vary for each participant. The goal of analysis is then to obtain a description of the variability between the parameters describing each individual's growth curve. In a latent growth curve model, the factor loadings for the slope(s) are specified to incorporate time into the model. The factor loadings of the intercept are all constrained to be one, indicating that an individual's intercept does not change over time. By way of analysis, the means for the intercept and slope parameters characterize the mean growth of the population. The variability in values for the individual's latent factors are given as the variances and covariances of these factors.

There are a number of standard assumptions made when using latent growth curve models. The error terms are assumed to have zero mean and constant variance for each individual. The error terms are independent of each other and any other variable except the measured variables they directly affect. The error terms are also nonautocorrelated.

In order to find an adequate model for examining the effects of the missing data methods, we examined linear, piecewise, and quadratic latent growth curve models before arriving at our final model. The model we choose is a quadratic latent growth curve model, since there was evidence of a curvilinear pattern. In addition, it is a good model to work with because it is more complex than a linear model but not so complex that it is difficult to compare missing data methods. The quadratic growth curve model for target individual $i$ is given by

$$y_{it} = \beta_{0i} + \beta_{1i}t + \beta_{2i}t^2 + \epsilon_{it}, \tag{1}$$

where $y_{it}$ is the level of hostility measured for individual $i$ at time $t = 0, \ldots, T$, $\beta_{0i}$ is the intercept factor for individual $i$, $\beta_{1i}$ is the linear slope factor, $\beta_{2i}$ is the quadratic slope factor, and $\epsilon_{it}$ is the variability for individual $i$ at time $t$. For the variability in the model, $E(\epsilon_{it}) = 0$ for all $t$, $E(\epsilon_{it}^2) = \sigma_i^2$ for all $t$, $COV(t, \epsilon_{it}) = 0$ for all $t$ and $COV(\epsilon_{it}, \epsilon_{it'}) = 0$ for all $t \neq t'$ except where otherwise noted. In this model the intercept, $\beta_0$, is constant over time for any given individual $i$. The parameters $\beta_{1i}$ and $\beta_{2i}$ are the linear and quadratic slope coefficients, respectively. The error terms have zero mean and constant variance across time for a given individual. Also they are uncorrelated with the time trend variable and nonautocorrelated.

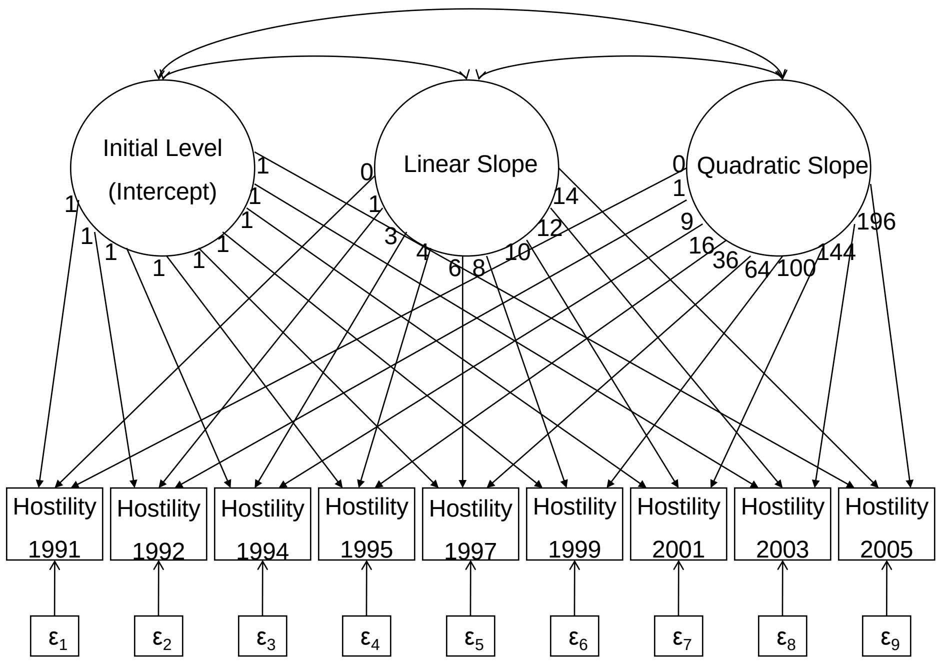


Figure 2.1 Quadratic latent growth curve model.

Figure 2.1 provides a visual depiction of the quadratic latent growth curve model. The arrows indicate causal relationships between our observed variables (hostility) and the latent variables (intercept and slopes). The numbers indicate the factor loadings, which are chosen to give information about the shape of growth across time. The linear factor loadings are incremented by the number of lapsed time periods since the previous observation. The first regression coefficient is set to zero, which allows the intercept to be identified. The factor loading for the quadratic slope are the square of those for the linear slope.

### 2.3 Missing Data

Dealing with missing data is a challenge most survey statisticians face when analyzing data collected from surveys. Missingness may occur when a participant inadvertently fails to respond to a question, refuses to answer a question, does not complete a timed trial, does not appear for a measurement session, or altogether stops participating in a longitudinal study. There are three forms of missing data to consider; namely, missing completely at random, missing at random, and not missing at random. When there are missing values in a data set, traditional statistical methods should not be implemented without first handling the missingness. The way in which missing data are handled is critical to the statistical analysis. If the missing data are not properly dealt with, the resulting parameter estimates may be bias.

In order to determine the method by which the missing data should be handled, one first needs to speculate the type of missingness present in the data set. It may be the case that there is more than one type of missingness in a data set. Once one considers why data may be missing then the method for handling the missing data needs to be determined.

In the definitions below, the following notation is used. Let $Y = \{Y_1, \ldots, Y_p\} = \{y_{i,j} \mid i = 1, \ldots, n, j = 1, \ldots, p\}$ denote the complete data set, where $p$ is the number of variables observed on $n$ individuals. Let $M = \{m_{i,j} \mid i = 1, \ldots, n, j = 1, \ldots, p\}$ where $m_{i,j} = 1$ if $y_{i,j}$ is missing and zero otherwise. Let $Y_{obs}$ denote the observed values of $Y$ and $Y_{mis}$ denote the missing observations. The type of missingness can be defined by the conditional distribution of $M$ given $Y$, that is $f(M \mid Y, \theta)$ where $\theta$ denotes a set of unknown parameters.

Data are considered to be *missing completely at random (MCAR)* if the missing value does not depend on the variable itself or on the values of other variables in the data set. That is if $f(M \mid Y, \theta) = f(M \mid \theta)$ for all $Y$ and $\theta$. For data to be missing completely at random, the missingness does not have to be a random pattern, but it cannot depend on the data. This is perhaps the simplest case of missing data to deal with in the analysis step. When the missing data for a given variable does not depend on the variable's particular value but is dependent upon other covariates in the data, then the data are considered to be *missing at random (MAR)*. That is, after controlling for other covariates, the missingness on $Y$ is independent of its true value. The conditional distribution is given as $f(M \mid Y, \theta) = f(M \mid Y_{obs}, \theta)$ for all $Y_{mis}$ and $\theta$. The data are *not missing at random (NMAR)* if the missing data are dependent on the values that are missing. This is given as $f(M \mid Y, \theta) = f(M \mid Y = (Y_{obs}, Y_{mis}), \theta)$ for all $Y$ and $\theta$.

In order to better understand these types of missing data, we use a hypothetical example. When respondents are asked to report their income in a survey, this variable often contains missing values. Assume that we have a random sample of individuals from some well-defined population. They are given five categories and asked into which category their income falls. If all individuals with incomes in categories four and five do not respond to this question, then this is NMAR. The data would be considered MAR if all individuals who are self-employed do not report their income. If every fifth individual given the survey does not report their income, then this is considered to be MCAR.

## 3. Methods of managing missing data

### 3.1 Listwise Deletion

Listwise deletion, or complete-case analysis (Little and Rubin, 2002), is perhaps the most traditional and simplest method for handling missing data. It can be implemented in most software packages and is often used in practice. Unfortunately the simplicity of listwise deletion comes at the cost of less precise results. Listwise deletion is when any observation that has missing data on a variable is excluded from the analysis. When listwise deletion is performed on the data set the sample size is often greatly reduced. This reduced data set will generally result in less than optimal results that are biased. Even if the data are MAR parts of the distribution for responses may be lost when listwise deletion is used.

In the FTP there are 143 targets on which hostility was observed in all nine waves of data collection. For the remaining 415 targets in the sample, their level of hostility was not measured in at least one wave of data collection. Handling the missing data by using listwise deletion results in a sample size of 25% of the original sample. This deduction of 75% of the target's will result in less precise estimates as indicated by larger standard errors associated with the estimates.

Pairwise deletion is very similar to listwise deletion. Pairwise deletion attempts to utilize all available information by deleting cases on a variable by variable basis (Allison, 2001). This results in more usable information than in listwise deletion, but there is still often a significant loss in accuracy of the estimates. We do not examine this method specifically because of its similarities with listwise deletion.

### 3.2 Full Information Maximum Likelihood

The idea of full information maximum likelihood (FIML) is to use all of the available data in the full sample to obtain estimates and standard errors assuming the data are missing at random. FIML is designed to identify each pattern of missing data and estimate the means, variances and correlations for each pattern. The estimates for the different patterns are then combined to produce overall estimates of the regression coefficients and their standard errors. That is, FIML maximizes a likelihood function that is the sum of all the likelihood functions for each pattern. FIML is different than traditional methods like pairwise deletion in that it takes into account the number of cases associated with each missing data pattern when computing estimates and standard errors (Enders and Bandalos, 2001). Thus the estimates and standard errors produced by FIML reflect all of the data that are actually observed in the data set. This results in estimates that are unbiased assuming that the data are MCAR or MAR. FIML is now available in many software packages. In practice, with the absence of knowledge of the missing data mechanism, FIML might be a good method for dealing with missing data because of the advantage of convenience and ease of use. A disadvantage of FIML is that it assumes a joint multivariate normal distribution of the variables used in the analysis.

Table 3.1 below displays three of the 116 patterns of missingness identified in the FTP data set. In the table, $O =$ observed, $M =$ missing, and $n$ is the number of targets who have this pattern. Pattern 1 indicates that there are 30 targets whose level of hostility was not observed in 1999. For pattern 2, the level of hostility was only measured in 1991 and 1992. For the nine targets in this pattern, the means and covariances for the latent intercept and slope can be calculated using the information provided by hostility in 1991 and 1992. In pattern 3 estimates of the mean and covariances for the latent intercept and slope can be computed using all years except 1999 and 2001. The likelihood for these different patterns is used to obtain maximum likelihood estimates for the means, variances, and correlations (Anderson, 1957; Enders, 2001).

Table 3.1. Three patterns of missing values.

| | 1991 | 1992 | 1994 | 1995 | 1997 | 1999 | 2001 | 2003 | 2005 | $n$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Pattern 1 | O | O | O | O | O | M | O | O | O | 30 |
| Pattern 2 | O | O | M | M | M | M | M | M | M | 9 |
| Pattern 3 | O | O | O | O | O | M | M | O | O | 16 |

### 3.3 Multiple Imputation

One of the main distinctions between multiple imputation (MI) and FIML is that MI imputes missing values (Rubin, 1987; Schafer, 1997). In MI procedures, $m$ replicates of the data set are created. In each of these copies, if the value is observed for a given variable, then that value is used in the copies of the data set. For the values that are missing, the value for the variable is imputed using information from the other observed values for the variable in the data set. In order to obtain the imputed values, an imputation model has to be formulated. This model takes into account any associations among the variables that may prove important in subsequent analyses and that the researchers believe may be inherent in the data. The imputed missing values are independently drawn from the posterior predictive distribution of the missing data (Little and Rubin, 2002). The important feature of MI is that the random variation that is found among the observed variables is also evident among the imputed values. This ensures that the variability for any given variable remains consistent among the observed and imputed values. Each of these $m$ data sets are then analyzed separately using traditional statistical methods.

Now there are $m$ different estimates for the means, variances, and correlations of the statistical parameters. These different estimates are combined into a single estimate as follows where the mean from each data set is given by $\mu_j$ and the associated standard error is $s_j$ where $j = 1, \ldots, m$. The mean of the regression coefficients is calculated as the mean of the $m$ estimates, $\bar{\mu} = \frac{1}{m}\sum_{j=1}^{m}\mu_j$. The standard error of the regression coefficients is computed in two pieces. First the within-imputation variance must be computed. This is calculated as the mean of the squared standard deviations from each of the $m$ data sets, $W = \frac{1}{m}\sum_{j=1}^{m}s_j^2$. Second the between-imputation variance is calculated. This is given as the sum of squared differences between the mean for each data set and the overall mean, which is then all divided by $m-1$, $B = \frac{1}{m-1}\sum_{j=1}^{m}(\mu_j - \bar{\mu})^2$. To obtain an estimate of the total variance, the within and between imputation variances are summed where the between-imputation is weighted by $(1 + \frac{1}{m})$. The estimated standard error is then the square root of this quantity, $\bar{s} = \sqrt{W + (1 + \frac{1}{m})B}$. These values can then be used to calculate confidence intervals and perform hypothesis tests with a t-distribution with degrees of freedom equal to $(m-1)(1+\frac{m\bar{\mu}}{(m+1)B})^2$. There are a number of standard statistical software packages that can be used to combine these estimates.

## 4. Results and Discussion

In this section we give the results from our analysis of the FTP data using the three previously discussed methods for handling missing data. The interest of the analysis was to look at the latent factors predicting the hostility of targets toward their mothers and friends or partners. We begin by discussing the missingness within the FTP in Tables 4.1 and 4.2. Then we look at the quadratic latent growth curve model for the FTP. We discuss the results from each of the missing data methods individually and then compare the results across methods. The estimated means and variances, along with their corresponding standard errors, are given for all methods in Tables 4.3 and 4.4.

There are a number of different ways in which data may be missing in the FTP. A family could not have participated in a wave of data collection, the target individual did not participate with their mother, friend, or partner in the observational portion of data collection, or a specific item may be missing in the observer's report. For the purposes of this paper, we are assuming that the missing values are missing at random and are ignoring item missingness on any one of the traits. That is, hostility will only be considered as missing if all five traits are missing. Table 4.1 displays the proportion of targets who participated in the observational data collection for each wave.

Table 4.1. Proportion of targets observed in each wave of data collection.

| Year | 1991 | 1992 | 1994 | 1995 | 1997 | 1999 | 2001 | 2003 | 2005 |
|---|---|---|---|---|---|---|---|---|---|
| Proportion Observed | 0.91 | 0.88 | 0.79 | 0.83 | 0.84 | 0.51 | 0.59 | 0.57 | 0.59 |

Beginning in 1999 the number of targets participating in the observational data collection significantly drops. One explanation for this may be that the majority of the targets graduated high school in 1994 and college in 1998. So in 1999, they were five years out of high school and settling into different parts of the country. This has resulted in lower numbers of targets participating in the observational data collection in the more recent waves.

There are a total of 116 different patterns of missingness across the waves of data collection. Rather than report these 116 different patterns for missingness, Table 4.2 shows the proportion of targets participating in a given number of waves. This indicates that slightly more than 25% of the targets have participated in all nine waves of observational data collection. Of the 559 families in the FTP, 528 (or almost 95%) have participated in at least 3 waves of data collection.

Table 4.2. Proportion of targets participating in waves of data collection.

| # Waves | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|
| Proportion Observed | 0.26 | 0.18 | 0.14 | 0.11 | 0.13 | 0.08 | 0.04 | 0.03 | 0.02 | 0.01 |

The fitted quadratic growth curve model indicates that there is a concave relationship between the target's level of hostility and time. That is, hostility levels increase over the first few years of data collection and then begin to level off and perhaps even decrease. In addition to the concave relationship, there is also an increasing level of hostility across time, as indicated by the positive linear slope. Additionally, we also correlate the errors for the hostility in 1995 with the errors for the hostility in 1991. The year 1995 is the first year in which hostility was measured with the targets' friend or partner and the target was one year out of high school. The hypothesis is that the hostility depicted in the target's family of origin continues in future relationships. That is, the target's hostility toward their friend or partner is related to the measured hostility toward their mother. By correlating the error terms for the hostility level in 1995 and the level in 1991, the model is able to depict this hypothesis. The second factor loading is freed in our growth curve model. This is because there is an increase of almost 10 points in the mean level of hostility between 1991 and 1992. The remainder of the mean hostility levels are between 12.1 and 15.6. By allowing the second factor loading to be estimated, we can capture this large change in our model. Freeing the second factor loading does not change the parameter estimates for the latent means, standard errors, and correlation, it simply allows our model to fit better. This model supports the hypothesis that hostility learned and exhibited in the family-of-origin is carried forward into the target's adult romantic relationships. We are going to use this growth curve structural equation model to evaluate different ways for handling the missing data in the FTP.

Table 4.3. Comparison of Estimates.

| | Mean Estimates (S.E.) | | |
|---|---|---|---|
| | Listwise | FIML | MI |
| Intercept | 11.959 (1.221) | 12.951 (0.789) | 12.796 (0.861) |
| Linear Slope | 1.003 (0.315) | 0.499 (0.208) | 0.577 (0.228) |
| Quadratic Slope | -0.075 (0.018) | -0.041 (0.012) | -0.046 (0.013) |

Table 4.4. Comparison of Variance Estimates.

| | Variance Estimates (S.E.) | | |
|---|---|---|---|
| | Listwise | FIML | MI |
| Intercept | 27.134 (7.986) | 24.116 (6.220) | 25.569 (6.001) |
| Linear Slope | 1.536 (0.411) | 1.584 (0.389) | 1.696 (0.368) |
| Quadratic Slope | 0.005 (0.002) | 0.005 (0.001) | 0.006 (0.001) |

When listwise deletion is used to handle the missing data, any individual that is not observed on all variables is deleted from the data set. There are only 143 targets who have participated in all nine waves of data collection for the FTP. This means that the number of individuals used to fit the model using listwise deletion decreases from 559 to 143. This is a 75% decrease in the number of observations used to estimate the latent factors.

The estimates of the means and variances of the latent growth curve factors using listwise deletion to handle the missing data are significant at the 1% level for all three of the factors. The linear slope is positive and the quadratic slope is negative. This indicates that there is a concave relationship of the targets' hostility across time. The estimated correlation between the errors for the level of hostility in 1995 and the hostility observed in 1991 is -0.572 with a standard error of 1.103 and a test statistic of -0.519. This regression is not significant ($p = 0.604$) to the model when listwise deletion is used for the missing data. The estimated correlations are significant for the correlation between the intercept and linear slope and between the two slopes. The correlation between the intercept and the quadratic slope is not significantly different than zero at the 5% level ($p = 0.137$). The correlation is estimated to be -0.413 (S.E. = 0.159 and t-value = -2.602) between the intercept and linear slope. The correlation between the linear and quadratic slopes is estimated as -0.970 with a standard error of 0.015 and test statistic of -64.584.

Using FIML to deal with the missing data allows for all of the observed data to be used in the analysis. Studies have found that this results in estimates that tend to be more accurate than those produced using listwise deletion. Another advantage to this method is that it is readily available in many standard software packages.

Applying FIML to handle the missing data and compute estimates for the latent factors in the FTP produces estimates that are found to be significant at the 2% level. There is a negative estimated mean for the quadratic slope, which indicates that there is a concave relationship for the targets' levels of hostility over time. This follows the conclusion in the listwise deletion method. All of the correlations are found to be significantly different than zero at the 2% level of significance. The correlation between the intercept and the linear slope is -0.492 (S.E. = 0.132, t-value = -3.722) and quadratic slope is 0.413 (S.E. = 0.143, t-value = 2.885). The correlation between the two slopes is -0.969 with a standard error of 0.011 and test statistic of -89.039. The estimated correlation between the errors for the level of hostility in 1995 and hostility in 1991 is -0.249 with a standard error of 0.103 and test statistic equaling -2.433, which indicates that this correlation is significant. This is a contradictory conclusion from the listwise deletion method, which is partically due to the fact that with fewer cases in the listwise deletion method it is more difficult to find a significant correlation between these two years.

The advantage of multiple imputation is that complete data sets are generated for use by traditional statistical techniques. To use multiple imputation as a means to handle the missing data, it is necessary to first obtain these multiply imputed data sets. For the FTP, these data sets were generated using `PROC MI` in *SAS*. According to some literature it was believed that 5-10 data sets are sufficient to obtain accurate estimates for the model parameters. There have been some recent studies indicating that more than 10 data sets may be necessary. Thus, we chose to generate 25 imputed data sets for our analysis. Although the hostility scores recorded from the trained observer are integers, we decided not to round the imputed values so that the random variability is preserved. Also, there were no restrictions placed on the range of values taken by the hostility levels. These decisions were made so that no additional bias was introduced into the imputed data sets. In order to present final parameter estimates, *Mplus* was used to analyze each data set individually and then combine the results as previously outlined in the missing data section.

The results for the model using multiple imputation as the method for dealing with missing data show that the estimated means and variances are significantly different than zero for all factors at the 1% level of significance. Again, the estimated mean of the quadratic slope is negative indicative of a concave relationship between hostility and time. When the errors for level of hostility in 1995 is correlated with the errors for the level of hostility in 1991, the estimated correlation is -0.285 with a standard error of 0.126 and a test statistic equaling -2.261. Thus this regression is significant to this model, as seen with FIML. The correlations between the latent factors are all significantly different than zero. The estimated correlation between the intercept and the linear slope is -0.510 (S.E. = 0.112 and t-value = -4.568) and for the quadratic slope it is estimated as 0.428 (S.E. = 0.123 and t-value = 3.494). The correlation between the two slopes is estimated as -0.969 with a standard error of 0.010 and test statistic of -98.865.

Next we present a brief comparison of the three methods used to handle the missing data in the FTP. When the estimated means and variances are compared across the different methods, the FIML and multiple imputation results are similar but the listwise deletion estimates differ. This is as expected since listwise deletion is using a smaller data set to compute the estimates and results in larger standard errors. The standard errors under multiple imputation are slightly larger than those for FIML. This is because of the important property of multiple imputation, which adds the random variability in the observed values to those imputed. The estimated variance for multiple imputation is larger than for either listwise deletion or FIML. This is once again believed to be a result of the random variability added to the imputed values based on that found in the observed data. But the standard errors for the estimated variances are largest for the listwise deletion, again due to the decreased sample size. The estimated correlation between the errors of the level of hostility in 1995 and the hostility in 1991 is not significant when listwise deletion is used to create a data set for analysis. For this method, the correlation between the intercept and the quadratic slope is also not significantly different than zero and the values for the correlations are different than for the other two methods. The estimated correlations are slightly larger and have a smaller standard error for MI in comparison to FIML.

There are many existing methods that may be used to assess the fit of a model. These methods evaluate the degree to which the data fit the hypothesized model and whether the fit can be improved through alternative models. We will report three of the commonly used indices of fit, the chi-square test statistic, the comparative fit index (CFI), which is based on the chi-square test statistic and the null model of uncorrelated or independent variables, and the root mean square error of approximation (RMSEA). The RMSEA is given by

$$\sqrt{\frac{\chi_k^2 - df_k/(n-1)}{df_k}}, \tag{2}$$

where $\chi_k^2$ is the chi-square test statistic for model $k$ with degrees of freedom equal to $df_k$.

The CFI (Bentler and Wu, 1995) is computed as

$$1 - \frac{\tau_k}{\tau_i}, \tag{3}$$

where $\tau_k = \max\left[(n\chi^2_k) - df_k, 0\right]$ is based on the model of interest and $\tau_i = \max[(n\chi^2_i - df_i),$ $(n\chi^2_k - df_k), 0]$. This index takes into consideration the chi-square variates and the degrees of freedom for the model and is restricted to values between zero and one. Values of the CFI close to one indicate an adequate fit of the model. The CFI avoids the underestimation of fit often noted in small samples.

In Table 4.5 is a summary of the model fits for these three methods. Using MI to handle the missing data allows the model to fit better than when FIML is used. Although the $\chi^2$ value is smaller for the listwise deletion method than MI, which indicates better fit, the CFI is smaller and RMSEA is larger, which is indicative of poorer fits. All of the statistics show that the MI method fits the data slightly better than the FIML method. We are able to gain some of information about the missingness in the data set by using imputation for the missing values. In conclusion, using either FIML or MI to obtain estimates for the latent factors yields similar results. Of these two approaches, using MI to deal with the missing data gives a slightly better overall model fit.

Table 4.5. Comparison of model fit.

| | $\chi^2$ | df | CFI | RMSEA |
|---|---|---|---|---|
| Listwise deletion | 75.088 | 32 | 0.875 | 0.097 |
| FIML | 128.275 | 32 | 0.864 | 0.074 |
| MI | 100.745 | 32 | 0.884 | 0.062 |

## 5. Conclusion

We used data from the FTP to illustrate the differences between methods used to handle missing data with structural equation models. For our analysis we considered the level of hostility displayed by the target individuals toward their mother for survey years 1991-1994 and toward their friend or partner in 1995-2005. Using this variable, we built a quadratic structural equation model and evaluated the use of listwise deletion, full information maximum likelihood, and multiple imputation. These methods were compared based on their parameter estimates and a selection of fit indices.

An advantage to using listwise deletion is that it is simple to implement and available in most software packages. Since this method deletes any case that is not completely observed, this often result in biased estimates and large standard errors. Both full information maximum likelihood and multiple imputation use all of the information available in the data set. FIML computes parameter estimates for each pattern of missingness and then combines all of these results using the likelihood for each pattern. This is the method that is commonly used for statistical analysis in the social sciences. In contrast to FIML, multiple imputation finds estimates for those values that are missing. A number of complete data sets are created and for each of these data sets estimates are obtained. These results are then combined to produce final estimates. An advantage of multiple imputation over FIML is that complete data sets are obtained where the variability in the observed data is similar to that of the imputed data set. These complete data sets are then available for traditional statistical analysis.

When the listwise deletion method was applied to the FTP data set, the results were more dissimilar than those for FIML and multiple imputation and the standard errors were always larger. The standard errors for multiple imputation were slightly larger than those for FIML, which is due to the preserved variability of the imputed data values. The parameter estimates were similar for FIML and multiple imputation. According to the model fit statistics used, the multiple imputation model fits slightly better than the FIML model.

All of the parameter estimates indicated that there is a concave relationship of the target's level of hostility across time. For the FIML and multiple imputation methods, there was a significant correlation between the errors for the target's level of hostility toward their mother in 1991 and toward their friend or partner in 1995. This indicates that, for these missing data methods, there is a significant relationship between an individual's behavior in their family of origin and their behavior toward others later in life.

Future work on this topic will include an examination of the effects of rounding and restricted range on multiple imputation estimates. Also we want to look into the number of replicate data sets needed using multiple imputation to produce acceptable results. We would like to see if results can be improved by imputing item missingness on the five traits comprising the total hostility variable. In addition to the methods we have already examined, we want to study fractional imputation methods and nonignorable missing data models.

## References


[1] Allison, P.D. (2001). *Missing Data.* Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-136. Thousand Oaks, CA: Sage.

[2] Anderson, T.W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, **52**, 200203.

[3] Bentler, P.M. and E.J.C. Wu. (1995). *EQS/Windows Users Guide.* Los Angeles, CA: BMDP Statistical Software.

[4] Bollen, K. A. (1989). *Structural equations with latent variables.* J. Wiley & Sons, New York.

[5] Conger, R.D. and K.J. Conger. (2002). Resilience in Midwestern Families: Selected Findings from the First Decade of a Prospective, Longitudinal Study. *Journal of Marriage and the Family*, **64**, 361-373.

[6] Enders, C.K. and D.L. Bandalos. (2001). The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models. *Structural Equation Modeling*, **8**, 430-457.

[7] Enders, C.K. (2001). The Performance of the Full Information Maximum Likelihood Estimator in Multiple Regression Models. *Educational and Psychological Measurement*, **61**, 713-740.

[8] Little, R.J.A. and D.B. Rubin. (2002). *Statistical Analysis with Missing Data.* J. Wiley & Sons, New York.

[9] Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys.* J. Wiley & Sons, New York.

[10] Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* New York: Chapman & Hall.