Detecting Underlying Emotional Sensitivity in Bereaved Children via a Multivariate Normal Mixture Distribution.

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Nicole DePriest Kelbick, B.S., M.S.

* * * * *

The Ohio State University

2003

Dissertation Committee:

Joseph S. Verducci, Adviser
Elizabeth A. Stasny
Thomas J. Santner

Approved by

_________________________________________________________________
Adviser
Department of Biostatistics
ABSTRACT

A common theme in finite mixture problems involves a random sample taken from a population consisting of an unknown mixture of distributions. The goal is to identify the component distributions using information from the sample. A medical example might entail clinical test results from patients whose true disease status is unknown. Another example pertains to latent class models which attempt to relate observed data to an unseen variable whose possible outcomes correspond to classes of a population.

Although mixture models are conceptually appealing many obstacles arise during their application. Areas of difficulty include complicated likelihoods, lack of clearly defined hypotheses, cumbersome estimating equations and elusive asymptotic properties. Progress in the study of mixture models was hindered by these difficulties until the advent of adequate computational power and numerical methods.

The topic of this thesis is motivated by a longitudinal study conducted at The Ohio State University that focused on the course of grief in children who experienced the loss of a parent. Researchers hypothesize parental loss will have a greater psychological impact on some of the children which will manifest itself over an extended period of time as an increase in the number of symptoms associated with behavioral, anxiety, mood and other psychological disorders. A mixture model approach is used to determine whether or not such a latent group of grieving children exists.
Under the null hypothesis, the bereaved children are a homogenous group and the data is assumed to have a multivariate normal distribution. The alternative hypothesis states the data follow a mixture of two multivariate normal distributions. Data patterns are exploited to develop simple models for the variance and correlation structures. Mean models are formulated to test the statistical hypotheses of interest. This approach has the benefit of reducing the number of model parameters resulting in a simplified fitting process. A consequence of the nonexclusive relationship between the mixing distribution and the model parameters is that both play a fundamental role in the development and outcome of the statistical inference procedure. Results of model fitting are reported and conclusions based on the likelihood ratio test are discussed.
To my husband Matt, whose steadfast support enabled me to succeed,
and to Mom and Dad, whose infinite patience has finally been rewarded.
ACKNOWLEDGMENTS

I would like to thank Dr. Mary Fristad in the Department of Psychiatry at The Ohio State University for providing me the opportunity to work with her as a graduate research assistant. The topic of my dissertation originated from statistical issues encountered during my work with the Grief Research Project. Thanks also goes to Julie Cerel, Ph.D. for dealing with my periodic data requests and for patiently explaining the various psychiatric jargon I encountered. I appreciate the help provided by Dave Fournier of Otter Research Ltd., who is the developer of the AD Model Builder software that enabled me to obtain, with relative ease, estimates of complicated data models. Lastly, I would like to thank my advisor, Dr. Joseph Verducci, without whose guidance I would not have been able to successfully complete my dissertation.
December 24, 1968 .......................... Born - Dayton, Ohio

1991 ........................................ B.S. Mathematics, University of Dayton, Dayton, Ohio
1995 ........................................ M.S. Statistics, The Ohio State University, Columbus, Ohio
1994-1997 ................................. Graduate Research Associate, The Ohio State University, Columbus, Ohio
1998-2000 ................................. Graduate Teaching Associate, The Ohio State University, Columbus, Ohio
2000-2001 ................................. Visiting Faculty, The Ohio State University, Columbus, Ohio

FIELDS OF STUDY

Major Field: Biostatistics
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xiii</td>
</tr>
<tr>
<td><strong>Chapters:</strong></td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 The Grief Research Study</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Summary of Subsequent Chapters</td>
<td>5</td>
</tr>
<tr>
<td>2. Literature Review</td>
<td>6</td>
</tr>
<tr>
<td>2.1 Longitudinal Data Analysis</td>
<td>6</td>
</tr>
<tr>
<td>2.2 Finite Mixture Models</td>
<td>13</td>
</tr>
<tr>
<td>2.2.1 Estimation</td>
<td>16</td>
</tr>
<tr>
<td>2.2.2 Testing for Latent Structure</td>
<td>21</td>
</tr>
<tr>
<td>2.3 The Research Problem</td>
<td>22</td>
</tr>
<tr>
<td>3. Preliminary Analysis</td>
<td>25</td>
</tr>
<tr>
<td>3.1 Missing Data</td>
<td>25</td>
</tr>
<tr>
<td>3.1.1 Theory of Missing Data Mechanisms</td>
<td>25</td>
</tr>
</tbody>
</table>
3.1.2 Missing the Initial Interview .......................... 27
3.1.3 Exploring Missing Data Patterns ......................... 30
3.1.4 Missing Data Mechanisms and the BAMO Data ........... 31
3.2 Demographics .................................................. 39
3.3 Mixture Model Application to the Research Problem .......... 40
  3.3.1 Data Transformation ........................................ 42
  3.3.2 Lindsay’s Method of Moments Procedure .................. 45
  3.3.3 Application to the Grief Data ............................ 49

4. Developing and Fitting Models ................................. 65
  4.1 Model Development ........................................... 65
  4.2 Results of Model Fitting ...................................... 71

5. Discussion and Conclusions ..................................... 73

Appendices:

A. Demographics and Summary Statistics for Subjects With An Initial Interview ........................................... 80
B. Demographics for and Summary Statistics Subjects Without An Initial Interview ...................................... 87
D. Program Code for Estimating a p-Component Univariate Normal Mixture Distribution .............................. 111

Bibliography ......................................................... 119
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>4</td>
</tr>
<tr>
<td>3.1</td>
<td>27</td>
</tr>
<tr>
<td>3.2</td>
<td>29</td>
</tr>
<tr>
<td>3.3</td>
<td>30</td>
</tr>
<tr>
<td>3.4</td>
<td>31</td>
</tr>
<tr>
<td>3.5</td>
<td>32</td>
</tr>
<tr>
<td>3.6</td>
<td>34</td>
</tr>
<tr>
<td>3.7</td>
<td>36</td>
</tr>
<tr>
<td>3.8</td>
<td>39</td>
</tr>
<tr>
<td>3.9</td>
<td>40</td>
</tr>
<tr>
<td>3.10</td>
<td>52</td>
</tr>
<tr>
<td>3.11</td>
<td>53</td>
</tr>
<tr>
<td>3.12</td>
<td>54</td>
</tr>
</tbody>
</table>
3.13 Method of moments and maximum likelihood estimates of component means when the primary axis is CUR25. ................................. 54

3.14 Proportion estimates of the susceptible and non-susceptible groups obtained from using the allocation rule based on Fisher’s discriminate function. ........................................................................... 57

3.15 Comparing the coefficients of the linear combination used in Fisher’s discriminate function for the primary axis choices $\bar{X}_W$ and CUR25. .......................... 59

3.16 Classification comparison of discriminate function allocation according to primary axis. ................................................................. 59

3.17 Classification comparison of discriminate function allocation according to depression status during the CUR1 time period when $\bar{X}_W$ is primary axis. ................................................................. 60

3.18 Classification comparison of discriminate function allocation according to inclusion criteria status when $\bar{X}_W$ is primary axis. ................. 61

3.19 Observed covariance matrix for the SqRt-BAMO data. ................. 62

3.20 Maximum likelihood estimate of the covariance matrix when $X_W$ is the primary axis. ................................................................. 63

3.21 Maximum likelihood estimate of the covariance matrix when CUR25 is the primary axis. ................................................................. 63

4.1 Parameter estimates under the $H_0$ model and corresponding standard errors and 95% confidence intervals. .............................................. 71

4.2 Parameter estimates under the $H_1$ model and corresponding standard errors and 95% confidence intervals. .............................................. 72

A.1 Gender of subjects. ................................................................. 80

A.2 Ethnic background of subjects. ................................................ 80

A.3 Inclusion Criteria Status (ICS) by subject. ................................. 81
B.8 Gender of surviving parent for subjects without an interview.  

B.9 Inclusion Criteria Status (ICS) by family for subjects without an initial interview.  

B.10 Hollingshead score assessing socioeconomic status by family for subjects without an initial interview.  

B.11 Distribution of ages of surviving parents one month after spousal death for subjects without an initial interview.  

B.12 Distribution of ages of deceased parent of subjects without an initial interview.  

B.13 Observed SqRt-BAMO means and variances for each time frame corresponding to the 41 subjects without an initial interview.  

B.14 Observed covariance matrix for the SqRt-BAMO data corresponding to the 41 subjects without an initial interview.  

B.15 Observed correlation matrix for the SqRt-BAMO data corresponding to the 41 subjects without an initial interview.
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 A summary of the time line of the interviewing process for Phase I.</td>
<td>3</td>
</tr>
<tr>
<td>3.1 A boxplot summary of BAMO scores for subjects with initial interview data and for those without.</td>
<td>28</td>
</tr>
<tr>
<td>3.2 Observed mean BAMO response over time according to each missing data pattern.</td>
<td>32</td>
</tr>
<tr>
<td>3.3 Nonresponse rates for the child ((n = 183)) and adolescent ((n = 145)) groups at each of the follow-up interviews.</td>
<td>33</td>
</tr>
<tr>
<td>3.4 Nonresponse rates for the simple ((n = 174)) and complex ((n = 154)) groups at each of the follow-up interviews.</td>
<td>35</td>
</tr>
<tr>
<td>3.5 Non-response rates according to socioeconomic status.</td>
<td>37</td>
</tr>
<tr>
<td>3.6 Mean BAMO scores over time according to socioeconomic status.</td>
<td>38</td>
</tr>
<tr>
<td>3.7 Histograms of the marginal distributions of the BAMO scores.</td>
<td>42</td>
</tr>
<tr>
<td>3.8 Scatterplot matrix of all time comparisons using the BAMO data.</td>
<td>43</td>
</tr>
<tr>
<td>3.9 Histograms of the marginal distributions of the BAMO scores after the square-root transformation.</td>
<td>44</td>
</tr>
<tr>
<td>3.10 Scatterplots of all time comparisons using the SqRt-BAMO data.</td>
<td>45</td>
</tr>
<tr>
<td>3.11 Maximum likelihood estimates of the component means.</td>
<td>55</td>
</tr>
<tr>
<td>3.12 Dotplot of Fisher’s discriminant function results.</td>
<td>58</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>3.13</td>
<td>Histograms of the marginal distributions of the SqRt-BAMO data with density estimate overlay.</td>
</tr>
<tr>
<td>4.1</td>
<td>Graphical displays of the observed variance and correlation structures of the SqRt-BAMO data.</td>
</tr>
<tr>
<td>4.2</td>
<td>Observed means of the SqRt-BAMO data.</td>
</tr>
<tr>
<td>4.3</td>
<td>Plots of the observed means of the SqRt-BAMO data over time according to SES levels and gender.</td>
</tr>
<tr>
<td>A.1</td>
<td>Histogram of the distribution of ages of the 328 subjects with initial interviews.</td>
</tr>
<tr>
<td>A.2</td>
<td>Histogram of the distribution of ages of the 187 surviving parents of subjects with initial interviews.</td>
</tr>
<tr>
<td>A.3</td>
<td>Histogram of the distribution of ages at time of death of the 187 deceased parents of subjects with initial interviews.</td>
</tr>
<tr>
<td>B.1</td>
<td>Histogram of the distribution of ages of the 41 subjects without initial interviews.</td>
</tr>
<tr>
<td>B.2</td>
<td>Histogram of the distribution of ages of the 27 surviving parents of subjects without initial interviews.</td>
</tr>
<tr>
<td>B.3</td>
<td>Histogram of the distribution of ages at time of death of the 27 deceased parents of subjects without initial interviews.</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

It is reasonable to speculate that parental death has a significant psychological impact on surviving children, particularly when the children have yet to reach adulthood. Understanding the psychological effects that the death of a parent has on such children will help determine if psychiatric intervention would be beneficial. Data from a study involving recently bereaved children is analyzed to ascertain whether or not parental death will emotionally impact some of the children more than others. A description of the study is provided followed by a synopsis of subsequent chapters.

1.1 The Grief Research Study

A longitudinal study was undertaken by Mary A. Fristad, Ph.D. of the Department of Psychiatry at The Ohio State University and Elizabeth Weller, M.D. and Ronald Weller, M.D., of the University of Pennsylvania, to explore the course of grief in children from families in which one parent has died. Three groups of children ranging between 5 and 18 years of age were recruited for the Grief Research Study. The bereaved group consists of 369 children from families where one of the parents has died. The remaining two groups are comparison groups. The depressed control group is composed of 111 psychiatric outpatients diagnosed with major depression. The 129...
community control subjects are recruited from local schools, churches and community groups. The bereaved group will be emphasized in the remaining discourse since their data will be the focus of this dissertation.

Each of the children who participated in the study was interviewed as was the surviving parent. A variety of psychological instruments were administered during each interview. Two important instruments were the Diagnostic Interview for Depression in Children and Adolescents (DIDCA) and the Diagnostic Interview for Children and Adolescents - Revised (DICA-R). The DIDCA is a structured interview used to assess depressive symptoms and establish mood disorder in child and adolescents (Weller and Weller 1979). The DICA is a structured interview designed to identify symptomatology meeting diagnostic DSM-III-R criteria of a variety of childhood psychological disorders (Reich and Welner 1988). The diagnostic criteria in DSM-III-R is the basis for clinical diagnosis of an assortment of mental disorders (APA 1987). The DICA-R and DIDCA were administered to the child and parent and used to elicit information about the child’s emotional state.

Four interviews were conducted over a two year period. These interviews were conducted at 1, 6, 13, and 25 months post-death and were carried out by well-trained staff, graduate students and undergraduate students. An interrater reliability of 0.90 was required before independent interviewing was permitted. Each child and parent were assigned different interviewers and the same people were used to conduct follow-up interviews whenever possible. The interviews were conducted either at the homes of the subjects or at the study center located at the Ohio State campus. Interviews lasted between 1 and 4 hours.
Retrospective as well as current responses were obtained at each interview. Figure 1.1 summarizes the timeline and labels the different time periods. Three time frames were reviewed at the initial interview: 1) lifetime up until one month prior to parental death 2) the month prior to the death and 3) the first month after the death. The time between follow-up interviews were divided into two periods: 1) two weeks prior to the current (CUR) interview and 2) the time since the last interview (SLI) up to the prior 2-week mark. For purposes of consistency, at the initial interview CUR1 refers to the first month after the parental death, SLI1 to the one month period before the death and LF to lifetime.

![Timeline Diagram](Figure 1.1: A summary of the time line in weeks of the interviewing process for Phase I. Zero refers to time of death. The vertical lines delineate the time periods associated with each interview. The solid circles indicate the actual time used in analyses for the designated time period.)

A psychometric scale called BAMO was developed by the researchers to measure the most common types of mental disorders encountered in children. Table 1.1 summarizes the specific disorders targeted by the BAMO score. The BAMO score is a weighted sum of psychological indicators of potential problems in the areas of
behavior, anxiety, mood and other disorders. Each component contributes a value between 0 and 1, inclusive, with the exception of those disorders marked by * which contribute a value between 0 and 0.5, inclusive. The BAMO score pools both child and parent information obtained from the DICA and DIDCA. Indicators of behavior, anxiety and other disorders are determined by the DICA whereas symptoms of mood disorder are specified by the DIDCA.

<table>
<thead>
<tr>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol Abuse</td>
</tr>
<tr>
<td>Attention Deficit Hyperactive Disorder (ADHD)</td>
</tr>
<tr>
<td>Cigarette Smoking*</td>
</tr>
<tr>
<td>Conduct Disorder</td>
</tr>
<tr>
<td>Drug Abuse</td>
</tr>
<tr>
<td>Oppositional/Defiant Disorder</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Anxiety</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avoidant Disorder</td>
</tr>
<tr>
<td>Obsessive-Compulsive Disorder</td>
</tr>
<tr>
<td>Overanxious Disorder</td>
</tr>
<tr>
<td>Simple Phobias*</td>
</tr>
<tr>
<td>Post-Traumatic Stress Disorder (PTSD)</td>
</tr>
<tr>
<td>Separation Anxiety</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mania</td>
</tr>
<tr>
<td>Major Depression</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anorexia</td>
</tr>
<tr>
<td>Bulimia</td>
</tr>
<tr>
<td>Encopresis*</td>
</tr>
<tr>
<td>Enuresis*</td>
</tr>
<tr>
<td>Psychosis</td>
</tr>
<tr>
<td>Somatization</td>
</tr>
</tbody>
</table>

Table 1.1: Summary of the disorders assessed by the BAMO score. Each component is scaled between 0 and 1, inclusive, with the exception of those disorders marked by * which are scaled between 0 and 0.5, inclusive.
The BAMO score is calculated for each child at each time period and is used in subsequent analyses to help determine if a subgroup of children exists that is more emotionally vulnerable to the death of a parent. This emotional susceptibility is expected to manifest as a significant increase in the number of behavioral, anxiety and depressive symptoms that are measured by the BAMO scale.

1.2 Summary of Subsequent Chapters

This dissertation describes the challenges of detecting the existence of a group of psychologically susceptible grieving children and explains the statistical techniques developed to address this problem. Chapter 2 provides a literature review of related statistical research. Chapter 3 presents preliminary data analysis. Chapter 4 describes the statistical models developed to test the hypotheses of interest and reports the outcome of model fitting. Results are discussed in Chapter 5 followed by a review of potential topics for future research.
CHAPTER 2

LITERATURE REVIEW

Since the grief data is collected repeatedly for each subject over the course of two years, the data is longitudinal. Secondly, the nature of the problem of detecting a potential group of emotionally susceptible children requires knowledge of mixture models. The relevant literature for each topic will be summarized below. The strengths of existing methods in each area will be discussed as well as their limitations within the context of the current problem.

2.1 Longitudinal Data Analysis

Traditional methods of dealing with serial measurements use multivariate regression techniques. This is a natural extension of the well-developed theory of linear regression for a single response variable. This approach, however, is limited to continuous data having a distribution that is approximately Gaussian. Difficulties arise with unbalanced designs, missing data, as well as discrete responses (Ware 1985). The 1980’s saw several breakthroughs that help overcome these difficulties.

A general multivariate model with unrestricted covariance structure is not conducive to highly unbalanced data. Laird and Ware (1982) address this issue by
introducing a two-stage random-effects model for longitudinal data. Assume each individual \( i, i = 1, \ldots, N \) is observed on \( n_i \) occasions. In the first stage, a distribution is specified for the response vectors \( Y_i \), conditional on the individual random-effects parameters. A distribution is then specified for the random-effects parameters in the second stage. The model can be described in the following fashion:

**Stage 1:**

\[
Y_i | b_i = X_i \beta + Z_i b_i + e_i, \tag{2.1}
\]

where \( X_i \) is the \( n_i \times p \) known design matrix relating \( \beta \) to \( Y_i \) and \( Z_i \) is the \( n_i \times k \) known design matrix relating the \( k \times 1 \) vector of unknown individual effects \( b_i \) to \( Y_i \). The \( e_i \) are independently distributed as \( N(0, R_i) \). It is assumed that the \( n_i \times n_i \) covariance matrix \( R_i \) is positive-definite and the unknown parameters in \( R_i \) are independent of \( i \). That is, \( R_i \) depends upon \( i \) only through its dimension \( n_i \).

**Stage 2:**

\[
b_i \sim N(0, \Lambda_i), \tag{2.2}
\]

where the \( b_i \) are i.i.d. and are independent of the \( e_i \). The \( k \times k \) covariance matrix \( \Lambda \) is assumed to be positive-definite. Marginally, the \( y_i \) have covariance matrix

\[
\Sigma_i = Z_i \Lambda Z_i' + R_i. \tag{2.3}
\]

The remainder of this paper discusses methods of model parameter estimation using maximum likelihood (ML) and restricted maximum likelihood (REML) techniques.

An advantage of this model is that balanced data is not necessary so the number and timing of observations can vary. Specific modeling of between-subject and within-subject variation is also possible. The limitation of this setup in relation to the current problem is that the unobserved variable is assumed to be normally distributed.
The methodology introduced by Liang and Zeger (1986) and Zeger and Liang (1986) transformed longitudinal data analysis. Their work expands the theory of generalized linear models (McCullagh and Nelder 1989) to accommodate longitudinal data. As a result, a general framework is developed that includes both continuous and discrete outcomes.

The authors assume the primary interest is the relationship between the outcome variable and certain independent variables and the time dependence of the outcomes for each subject is considered a nuisance parameter. They focus on obtaining regression parameter estimates using an estimating equations approach. Only minor assumptions are needed to maintain consistency of the regression and variance parameter estimates.

Instead of specifying the joint distribution of the repeated measurements, only the marginal distribution of the outcome at time $t$ is stipulated. Let $Y_i = (Y_{i1}, \ldots, Y_{in_i})'$ and $X_i = (x_{i1}, \ldots, x_{in_i})'$ represent the outcomes and covariate values for the $i$th subject, $i = 1, \ldots, N$, respectively. The form of the marginal density of $Y_{it}$ is assumed to be

$$f(y_{it} | \beta, \phi) = \exp\{[y_{it} \theta_{it} - a(\theta_{it}) + b(y_{it})] \phi\}$$

(2.4)

where $\theta_{it} = h(\eta_{it})$, $\eta_{it} = x_{it}' \beta$. As a result,

$$E(Y_{it}) = a'(\theta_{it}) = \mu_{it}, \quad \text{and} \quad \text{Var}(Y_{it}) = a''(\theta_{it})/\phi = v_{it}. \quad (2.5)$$

A likelihood analysis assuming independence of observations over time as well as between subjects yields score equations having the form

$$S_I(\beta) = \sum_{i=1}^{N} \left( \frac{\partial \mu_i}{\partial \beta} \right)' (Y_i - \mu_i) = 0. \quad (2.6)$$
The solution to Equation 2.6, \( \hat{\beta}_I \), is consistent and \( N^{\frac{1}{2}}(\hat{\beta}_I - \beta) \) is asymptotically multivariate Gaussian with mean zero and covariance matrix \( V_I \), for which a consistent estimator also exists.

Although \( \hat{\beta}_I \) is consistent and can be computed by existing GLIM software, its principal drawback is its inefficiency when the correlation between observations over time is large. Accounting for the correlation between observations over time improves efficiency. Define \( A_i = \text{diag}\{a''(\theta_{it})\} \) and let \( R(\gamma) \) be a \( n_i \times n_i \) symmetric “working” correlation matrix. Note that in the case where \( R(\gamma) \) is the true correlation matrix,

\[
V_i = A_i^{\frac{1}{2}} R(\gamma) A_i^{\frac{1}{2}} / \phi
\]

is equivalent to \( \text{Var}(Y_i) \). Define the general estimating equations as

\[
S_G(\beta, \gamma) = \sum_{i=1}^N \left( \frac{\partial \mu_i}{\partial \beta} \right)' V_i^{-1} (Y_i - \mu_i) = 0. \tag{2.8}
\]

These equations reduce to the score equations of the independence model in Equation 2.6 when \( R(\gamma) \) is the identity matrix.

Equation 2.8 is similar to the quasi-likelihood score function discussed by Wedderburn (1974) and McCullagh and Nelder (1989). The major difference is that the \( V_i \)'s defined here depend on both \( \beta \) and \( \gamma \). In order for the above estimating equations to rely solely on \( \beta \), the parameters \( \phi \) and \( \gamma \) are replaced by \( N^{\frac{1}{2}} \)-consistent estimators \( \hat{\phi}(Y, \beta) \) and \( \hat{\gamma}(Y, \beta, \hat{\phi}) \), respectively. The estimate \( \hat{\beta}_G \) is defined as the solution to the system of generalized estimating equations obtained after substituting \( \hat{\phi} \) and \( \hat{\gamma} \) into Equation 2.8. An iterative computation procedure is used to estimate \( \beta_G \). Liang and Zeger (1986) show that under mild regularity conditions \( N^{\frac{1}{2}}(\hat{\beta}_G - \beta) \) is asymptotically multivariate Gaussian with zero mean and a covariance matrix \( V_G \) that they
derive. A consistent estimate of \( V_G \) can be obtained by substituting \( S_iS_i^T \) for \( V_i \) as well as the estimates \( \hat{\beta}, \hat{\phi} \) and \( \hat{\gamma} \) into the formula for \( V_G \) described in the paper.

The independent estimating equations yield consistent estimates of \( \beta \) and \( \text{var}(\hat{\beta}) \) as long as the form of the mean is correctly specified. This consistency is maintained in the general case regardless of the choice of \( R(\gamma) \) as long as estimators for \( \gamma \) and \( \phi \) are \( N^{\frac{3}{2}} \)-consistent. To maintain the consistency of \( \hat{\beta}_I \) and \( \hat{\beta}_G \) in the presence of missing data, either dropouts must occur completely at random or the specified \( R(\gamma) \) must be the true correlation matrix.

These assumptions can be relaxed somewhat when the data are considered missing at random. Robins, Rotnitzky and Zhao (1995) extend the GEE method to accommodate random dropouts. The idea is to incorporate into the estimating equations of Equation 2.8 the probability that subject \( i \) has not dropped out by time \( t_j \) given the subject’s previous responses and any pertinent covariate information. Let \( p_{ij} \) denote this probability and let \( P_i \) be the diagonal matrix with non-zero elements \( p_{ij} \). Weighting the responses \( Y_i \) by \( P_i^{-1} \) preserves the consistency of the mean response. This is accomplished by incorporating \( P_i^{-1} \) into Equation 2.8 in the following way:

\[
S_G(\beta, \gamma) = \sum_{i=1}^N \left( \frac{\partial \mu_i}{\partial \beta} \right)' V_i^{-1} P_i^{-1} (Y_i - \mu_i) = 0. \tag{2.9}
\]

A thorough discussion of the exact requirements under which Equation 2.9 yields a consistent estimate of \( \beta \) when estimates \( \hat{p}_{ij} \) are substituted for \( p_{ij} \) is provided by Robins et al. (1995).

The examples of Zeger and Liang (1986) show that when the correlation between time points is small, both \( \hat{\beta}_I \) and \( \hat{\beta}_G \) are quite efficient. As the serial correlation increases, however, \( \hat{\beta}_G \) maintains efficiency whereas \( \hat{\beta}_I \) does not. The same is true when the number of observations, \( n_i \), varies according to individual.
This generalized estimating equation (GEE) approach to estimation of regression parameters has become quite popular. Software has been developed for two common statistical packages, Splus (via `gee` and `yags`) and SAS (via `GENMOD`). This popularity stimulated research that yielded similar extensions of GLMs for both random effects and Markov settings.

Zeger and Liang (1992) review three extensions of GLMs to the longitudinal setting: marginal models, random effects models and transition (Markov) models. They discuss in particular the GEE approach applied to marginal models using examples involving discrete outcomes. A synopsis of each model follows. A more in-depth discussion of these models is provided by Diggle, Hagearty, Liang and Zeger (2002).

**Marginal models**

Let $Y_{it}$ represent the response of the $i$th subject at time $t$ and let $x_{it}$ be the corresponding vector of explanatory variables. A marginal model assumes the following:

1. the relationship between the marginal mean, $\mu_{it} = E(Y_{it})$, and $x_{it}$ at time $t$ is specified by a known link function, $g$,

$$g(\mu_{it}) = x'_{it} \beta; \quad (2.10)$$

2. the marginal variance is a known function depending on the marginal mean, 

$$\text{Var}(Y_{it}) = v(\mu_{it})\phi, \quad (2.11)$$

where $\phi$ is viewed as an over-dispersion parameter that adjusts for additional variability in $Y_{it}$ not accounted for by $v(\mu_{it})$;
3. the correlation between responses for a given subject is a known function of the marginal means and possibly additional parameters, $\gamma$,

$$\text{Corr}(Y_{is}, Y_{it}) = \rho(\mu_{is}, \mu_{it}; \gamma), \quad s < t = 1, \ldots, n_i. \quad (2.12)$$

**Random effects models**

In the random effects model, it is assumed there exist regression coefficients that are heterogenous across subjects. This variability is the source of modelling the correlation among responses for an individual. Assumptions for this model include:

1. the random effects $b_i, i = 1, \ldots, N$, are i.i.d. with common underlying distribution $F(\cdot, \Lambda)$;

2. the conditional distribution of $Y_{it}$ given $b_i$ satisfies the GLM criteria with

$$g(E(Y_{it}|b_i)) = x_{it}'\beta + z_{it}'b_i, \quad (2.13)$$

where $z_{it}$ is a $p \times 1$ vector of subject-varying covariates;

3. the responses $Y_{it}, t = 1, \ldots, n_i$, are conditionally independent given $b_i$.

**Transition (Markov) models**

In transition models, correlation among an individual’s responses is a result of the current observation, $Y_{it}$, depending on past observations $Y_{it-1}, \ldots, Y_{i1}$. Thus, previous responses can be viewed as additional explanatory variables. Let $H_{ij} = \{Y_{it}, t = 1, \ldots, j - 1\}$ represent the response history of subject $i$ at time $j$. Assumptions for the transitional model include the following:

1. the conditional distribution of $Y_{it}|H_{it}$ satisfies GLM criteria and the conditional expectation of $Y_{it}$ given the past is related to $x_{it}$ and $H_{it}$ by a known link
function, $g$,

$$g(\mu_{it}^c) = X_{it}'\beta + \sum_{j=1}^{s} \alpha_j^c f_j(H_{it}), \quad (2.14)$$

where $\mu_{it}^c = E(Y_{it}|H_{it})$ and $f_j, j = 1, \ldots, s$ are known functions;

2. the conditional variance of $Y_{it}$ given the past is a known function of $\mu_{it}^c$, 

$$\text{Var}(Y_{it}|H_{it}) = v(\mu_{it}^c)\phi. \quad (2.15)$$

Model selection depends on the particular research question. Marginal models allow the mean and correlation to be modeled separately whereas transition models and random effects models address regression and within-subject correlation simultaneously. When interest is focused on population averages, marginal models are appropriate. In this setting the regression coefficients have the same interpretation as in a cross-sectional analysis. Random effects models are an obvious choice when individual effects are of interest. If it is believed that prior responses influence current responses, then transition models are suitable.

Estimation techniques vary according to model. The GEE methodology described earlier is commonly used for marginal models and can be used in certain instances with transition models. Maximum likelihood is used in random effects models as well as transition models. For a detailed discussion of estimation in each of the three settings and related difficulties see Diggle et al. (2002).

### 2.2 Finite Mixture Models

Let $f(y|\theta)$ represent a $d$-dimensional probability density function characterized by an $m$-dimensional parameter vector $\theta$. Let $Q(\theta)$ be the distribution function for
A general definition for a mixture density $p(y)$ is

$$p(y) = \int f(y|\theta)dQ(\theta)$$  \hfill (2.16) \tag{2.16}

where $Q$ is called the mixing distribution. The notation $p(y)$ will be used to represent a probability density function even in the discrete case. The remaining discourse will be restricted to the case where $Q$ is a discrete distribution with a finite number of support points, $\theta_j \in \Theta, j = 1, \ldots, p$ each having probability $\pi_j > 0$ of occurring and $\pi_1 + \ldots + \pi_p = 1$. The component densities, $f(y|\theta_j)$, will be members of a common parametric family. The resulting finite mixture density function will then have the form

$$p(y) = \sum_{j=1}^{p} \pi_j f(y|\theta_j).$$  \hfill (2.17)

Mixture models apply naturally to a variety of situations. A medical example might involve results of an assortment of clinical tests from patients whose true disease status is unknown. Another example might pertain to latent class models which attempt to relate observed data to an unseen variable whose possible outcomes correspond to classes of a population. Lindsay (1995) provides a compilation of situations that have a mixture structure.

Mixture models are conceptually appealing but many obstacles can arise during actual implementation. Areas of difficulty include complicated likelihoods, lack of clearly defined hypotheses, cumbersome estimating equations and elusive asymptotic properties. Although the idea of mixture models has been around for over a century, these difficulties have hindered research until recently. An example of the complexity involves the successful but laborious attempt by Pearson (1894) to derive estimates of the five parameters of the two-component univariate normal mixture using method
of moments. The process involved much tedious algebra which ultimately required finding the roots of a ninth degree polynomial. It is not surprising that a resurgence of interest in mixture models these past few decades coincided with the rapid growth of computational power.

A standard analysis of mixture models is typically concerned with estimating the latent distribution $Q$ which involves the likelihood function

$$L(Q) = \prod_{i=1}^{n} L_i(Q) = \prod_{i=1}^{n} \left( \sum_{j=1}^{p} \pi_j f(y_i|\theta_j) \right).$$  \hfill (2.18)

The corresponding log-likelihood function is

$$l(Q) = \ln L(Q) = \sum_{i=1}^{n} \ln L_i(Q).$$ \hfill (2.19)

Everitt and Hand (1981) cite a simple example involving the mixture of two univariate normal densities that reveal some of the difficulties encountered in analyzing mixture models. In this case, $l(Q)$ has the form

$$l(Q) = l(\pi, \mu_1, \mu_2, \sigma_1, \sigma_2)$$

$$= \sum_{i=1}^{n} \ln[\pi N(y_i|\mu_1, \sigma_1) + (1 - \pi)N(y_i|\mu_2, \sigma_2)]$$ \hfill (2.20)

where $N(\cdot)$ denotes the density of the normal distribution. Unlike the unicomponent case, the normal equations, obtained by equating the first derivatives with respect to each parameter to zero, do not have explicit solutions. Thus, iterative techniques are typically necessary to determine parameter estimates. Another problem that occurs is the unboundedness of $l(\cdot)$ resulting in singularities. For instance, setting $\mu_1 = y_1$ and letting $\sigma_1 \to 0$ causes $l(\cdot) \to \infty$. This obstacle can be avoided if the variances are bounded away from zero.
Titterington, Smith and Makov (1985) point out the not-so-straightforward problem of hypothesis testing involving mixture models. Continuing with the previous example, suppose we wish to test whether the distribution of $Y$ consists of a single normal component or two components. This translates to testing the following hypotheses:

\[
H_0 : \quad p(y) = N(y|\mu, \sigma),
\]

\[
H_1 : \quad p(y) = \pi N(y|\mu_1, \sigma_1) + (1 - \pi)N(y|\mu_2, \sigma_2).
\] (2.21)

Traditional hypothesis testing methods employ the likelihood ratio test which is assumed to follow a chi-square distribution with degrees of freedom equal to the number of constraints on the parameters in $H_1$ necessary to produce the nested model in $H_0$. Unfortunately, $H_0$ is not uniquely determined from $H_1$. We could apply the single constraint $\pi = 1$ or we could use the two constraints $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$. This ambiguity makes the appropriate degrees of freedom for the likelihood ratio test indeterminate. Simulation studies of the likelihood ratio test in the mixture model setting suggest the breakdown occurs as a result of the violation of regularity conditions that are the basis of conventional asymptotic theory (Titterington et al. 1985).

2.2.1 Estimation

Due to the intractable nature of the estimating equations, a variety of estimation methods have been investigated which include graphical techniques, method of moments (MOM), maximum likelihood (ML), Bayesian methods and minimum distance methods. The reader is referred to Titterington et al. (1985) for a summary of the latter two methods. Both Everitt and Hand (1981) and Titterington et al. (1985)
review various graphical techniques for detecting mixtures as well as estimating parameters. Examining histograms for multimodality is a natural starting point. The absence of more than one mode does not indicate a mixture structure is inappropriate, however. Behboodian (1970) derives the following unimodal criteria for a mixture of two univariate normals:

$$|\mu_1 - \mu_2| \leq 2 \min(\sigma_1, \sigma_2).$$  \hfill (2.22)

See Fowlkes (1979) and Bhattacharya (1967) for examples of other graphing techniques. Although the interpretation of the results of graphical methods are rather subjective and their helpfulness is limited when separation of components is small, they are a useful exploratory tool.

Method of moments estimation has seen extensive application in the mixture setting despite difficulties which include 1) lack of uniqueness 2) consistency not being guaranteed 3) not achieving asymptotic efficiency and 4) the covariance structure of the estimators having to be approximated (Titterington et al. 1985). The popularity of MOM estimation stems in part from the existence of explicit solutions in most applications as well as the difficulties associated with maximum likelihood estimation prior to the availability of computers.

Although explicit solutions may exist, obtaining them can be difficult as is clearly demonstrated in the examples presented by Pearson (1894) for a mixture of two normal densities. Lindsay (1989a) has done extensive work developing swiftly-computed moment matrices that yield a consistent estimate of $Q$, the mixing distribution (Lindsay 1989b). Lindsay and Basak (1993) extend this method to multivariate normal mixtures and even suggest a way of estimating the covariance matrix, $\Sigma$, assumed to be completely unknown. Even though ML algorithms play a more prominent role
now that powerful computing is accessible, MOM provides a means of producing good starting values for the algorithms which is crucial for successful ML estimation.

Because of the many attractive properties of ML estimation in simpler parametric situations, such as ease of computation, asymptotic normality and the natural relation to likelihood-based inference, its pursuit in the context of mixture modeling is not surprising. Unfortunately, these attributes do not necessarily cross over to the mixture setting.

Numerical methods are typically employed since explicit solutions are rare. Titterington et al. (1985) relate the mixture model to a missing data situation. If the category to which an observation belongs is known, the complete data for the \(i\)th observation can be written as \(z_i = (y_i, w_i), i = 1, \ldots, N\) where \(y_i\) is the observed response and \(w_i\) is a \(k \times 1\) indicator vector with a 1 in the position corresponding to the appropriate classification and zeros elsewhere. In this context, the EM algorithm would be an appropriate numerical method (Dempster, Laird and Rubin 1977). The EM algorithm involves a two-step process:

\begin{align*}
\text{E step:} & \quad \text{Calculate } l_{EM}(Q, Q^{(m)}) = E[\ln L(z \mid Q) \mid y, Q^{(m)}]. \\
\text{M step:} & \quad \text{Find } Q^{(m+1)} \text{ to maximize } l_{EM}(Q, Q^{(m)}). 
\end{align*}

Explicit solutions for both steps are desirable. Titterington et al. (1985) write general formulas for the E step in the univariate setting and note that the particular form of the M step is specific to the given problem. Redner and Walker (1984) derive explicit formulas for both the E step and M step when component densities are members of the multivariate normal family. Once a starting value \(Q^{(0)}\) is provided, the algorithm iterates between the E step and M step until a stopping criteria is met. The chief advantage of the EM algorithm is that the likelihood at each step is
monotonically increasing. That is, for \( m = 0, 1, \ldots \),

\[
\ln L(z \mid Q^{(m+1)}) \geq \ln (z \mid Q^{(m)}).
\] (2.23)

Another advantage is that derivatives of the likelihood are not required. The tradeoff is that the EM algorithm can be very slow.

On the other hand, Newton-Raphson (NR) and quasi-Newton numerical methods are typically much faster. The drawback is that first- and second-order differentiation with respect to the unknown parameters are necessary as well as matrix inversion. NR methods also lack the monotonic property of the EM algorithm. Although convergence is not guaranteed in either the NR or quasi-Newton approach, failure seems to some extent more likely under NR. Titterington et al. (1985) suggest combining both techniques. Iterate five to ten times under the EM algorithm to position high on the surface of the likelihood then switch to the NR method to seek out the maximum.

In some cases, the desirable asymptotic properties of ML estimation in the simpler parametric models are maintained in the mixture model setting. Redner and Walker (1984) present theoretical results and cite the regularity conditions necessary to uphold consistency and asymptotic normality when the number of components is assumed to be known.

Lindsay (1995) provides an in-depth discussion of finding what is known as a nonparametric likelihood estimator (NPMLE) of the latent distribution \( Q \). In this context, \( k \) is unknown so that no parametric assumptions are made regarding \( Q \). The goal is to maximize the likelihood function,

\[
l(Q) = \sum_{i=1}^{n} \ln L_i(Q) = \sum_{s=1}^{D} n_s \ln L_s(Q)
\] (2.24)
over all distributions $Q$ where

$$L_s(Q) = \int L_s(\theta) dQ(\theta).$$

(2.25)

The likelihood has been consolidated to $D$ distinct summands to allow for $n_s$ repeated observations of a single $L_s$. The only assumption is that each $L_i(\theta) = f_i(x_i|\theta)$ is nonnegative and bounded.

Define the path between $Q_0$ and $Q_1$ in the space of distribution functions by $Q_\alpha = (1 - \alpha)Q_0 + \alpha Q_1$, $\alpha \in [0, 1]$. The directional derivative corresponding to the path from $Q_0$ to $Q_1$ is defined as the derivative of $l(Q_\alpha)$ evaluated at $\alpha = 0$:

$$D_{Q_0}(Q_1) = \frac{dl(Q_\alpha)}{d\alpha} \bigg|_{\alpha=0} = \sum_{s=1}^{D} n_s \left[ \frac{L_s(Q_1)}{L_s(Q_0)} - 1 \right].$$

(2.26)

When $Q_1$ is degenerate at the point $\theta$, Equation 2.26 is called the gradient function and is denoted by $D_{Q_0}(\theta)$. The distribution $\hat{Q}$ is designated the MLE by the gradient function if and only if

$$D_{\hat{Q}}(\theta) \leq 0 \quad \forall \theta.$$

(2.27)

The idea is that if the above inequality fails for some $\theta_0$ then the latent distribution candidate is not the MLE and the likelihood must be increasing in the direction of $\theta_0$ so it makes sense to put some mass at this point.

Another attractive feature of the gradient function is that the support points $\hat{\theta}_j$ of the MLE $\hat{Q}$ can be derived from the relation $D_{\hat{Q}}(\hat{\theta}_j) = 0$. Although the MLE of $Q$ is not necessarily unique, the NPMLE theorem states that for any two latent distributions $\hat{Q}_0$ and $\hat{Q}_1$ that maximize the likelihood, the likelihood itself is uniquely determined in the sense that $l(\hat{Q}_0) = l(\hat{Q}_1)$. Lindsay shows that $\hat{Q}$ is in fact unique for
various families of component distributions. Thus, the NPMLE method provides another means of obtaining the MLE with the added advantage of very few assumptions regarding the parameter space of $\theta$.

2.2.2 Testing for Latent Structure

A variety of methods exist for testing latent structure. Titterington et al. (1985) and Lindsay (1995) both review various testing procedures. DerSimonian (1989) examined many tests and concluded that the majority fell into one of two categories. The more general of the two essentially tests

$$H_0 : \text{one component} \quad \text{vs} \quad H_1 : \text{any mixture} \quad (2.28)$$

(Neyman 1959, Neyman and Scott 1966). According to Lindsay (1995) it is a simple yet effective test for overdispersion and is appealing since the test does not depend on the mixing distribution in $H_1$.

The second type of test that is commonly used is the likelihood ratio test (LRT). Its design is geared more for testing hypotheses that stipulate a specific number of components in both hypotheses. For example, a typical test might involve:

$$H_0 : \text{one component} \quad \text{vs} \quad H_1 : \text{two components}. \quad (2.29)$$

Difficulties arise when testing for number of components in a mixture model that are usually avoided when the LRT involves simpler parametric models.

Recall the LRT requires evaluating the likelihood at the MLE under both the null and alternative hypotheses. Let $L_0$ and $L_1$ represent the corresponding likelihoods under $H_0$ and $H_1$, respectively. The likelihood ratio statistic is given by $\Lambda = L_0/L_1$ (Neyman and Pearson 1928). Given certain regularity conditions, Wilks
(1938) showed that $-2 \ln \Lambda$ is asymptotically distributed as chi-square with degrees of freedom equal to the difference in number of parameters between the null and alternative hypotheses.

This is not necessarily the case in the mixture model setting. Wolfe (1971) suggests LRT is not appropriate for testing number of components because of the failure to meet regularity conditions (Serfling 1980). For example, suppose the hypotheses of interest are $H_0 : p(y) = f(y|\theta) \text{ vs } H_1 : p(y) = \pi f(y|\theta_1) + (1 - \pi) f(y|\theta_2)$.

This breakdown in the asymptotic theory has been illustrated in many examples. For instance, Lindsay (1995) describes the situation of having a mixture of two known component densities with $\pi$ being the only unknown. It is shown that the limiting distribution of $-2 \ln \lambda$ is a mixture of two chi-squares, $0.5\chi_0^2 + 0.5\chi_1^2$, where $\chi_0^2$ represents a degenerate distribution at 0. In the case of a two-component normal mixture with $\pi$ specified, however, Titterington et al. (1985) state that under $H_0$, $-2 \ln \lambda$ will be asymptotically distributed $\chi_d^2$ where $d$ is the number of unknown parameters under $H_0$. It appears that the usual asymptotics apply as long as the parameters under $H_0$ are not on the boundary of the parameter space and the dimensions of competing models are clearly defined.

2.3 The Research Problem

The research question under consideration is whether or not there exists a more emotionally impaired subgroup of grieving children. Detection of such a latent class naturally involves mixture models. The longitudinal nature of the data requires using a mixture of multivariate distributions. In this case, the multivariate normal distribution will be used. Although much of the mixture model literature focuses on
the univariate case, Everitt and Hand (1981) provide ML estimates of \( \pi_k, \mu_k \) and \( \Sigma_k \) in the multivariate normal setting that are analogous to the parameter estimates in the unicomponent case except each observation is weighted by the posterior probability of belonging to the \( k \)th component given \( y_i \). These estimates are not explicit, however, and require iterative techniques to obtain solutions.

The current problem deviates from this general setup by specifying models for the mean, variance and correlation structures. In particular, a set of predictor variables will be used to model the mean so that we have \( \mu(\beta) = h(\beta, x) \) where \( h \) will be a nonlinear function of the elements in \( \beta \). The mixture model will have the form

\[
p(y|x, \beta, \delta, \pi) = \pi N[\mu_1(\beta), \Sigma(\gamma)] + (1 - \pi)N[\mu_2(\beta, \delta), \Sigma(\gamma)]. \tag{2.30}
\]

The mean vector, \( \mu_2 \), differs from \( \mu_1 \) by the \( m \times 1 \) parameter vector \( \delta \). This vector contains the added effects of the emotionally susceptible group of bereaved children relative to the non-susceptible children for each time period. Thus, the hypothesis test of the existence of a latent group of emotionally susceptible children becomes

\[
H_0 : \delta = 0 \quad \text{vs} \quad H_1 : \delta \neq 0. \tag{2.31}
\]

The advantage of modeling the mean in this fashion is that there is less ambiguity regarding the formulation of the hypotheses. The model corresponding to the null hypothesis \( H_0 \) is clearly a subset of the model under \( H_1 \).

Parameters \( \beta \) and associated predictors \( x \) are commonly referred to as structural parameters when they appear in addition to the mixing distribution \( Q \). Estimation methods and asymptotic theory for these auxiliary parameters is in the developmental stage (Pfanzagl 1990) and (Bickel, Klaassen, Ritov and Wellner 1993). Lindsay
and Lesperance (1995) summarize the progress in maximum likelihood theory including estimation techniques and inference for the mixing distribution and structural parameters.

The null hypothesis in the current situation is very similar to the marginal model in the longitudinal setting since separate models for the mean and correlation structure will be specified. GEE methodology is not applicable in the mixture distribution setting, however, and cannot be utilized under $H_0$ because a link function does not exist that linearizes the chosen mean model as a function of the elements of $\beta$. Assuming the mixture distribution has two components eliminates NPMLE methods from consideration since this approach assumes the number of components is unknown. Iterative procedures involving the log-likelihood function will be implemented instead to obtain parameter estimates under both the null and alternative hypotheses.
CHAPTER 3

PRELIMINARY ANALYSIS

This chapter presents the preliminary analysis that laid the groundwork for the models to be used in the hypothesis testing. Missing data patterns are explored, demographics of the data are summarized and mixture model diagnostics are investigated.

3.1 Missing Data

The issue of missing data will be addressed since almost half of the 369 subjects missed at least one interview. A brief summary of missing data mechanisms is presented first. Missing data patterns in the BAMO data are then evaluated to ascertain what assumptions can be made when developing models and estimation procedures.

3.1.1 Theory of Missing Data Mechanisms

Recall that information was obtained about more than one time period during each interview. Figure 1.1 summarized the various time frames. Analysis will focus on the period beginning one month prior to death (SLI1) and ending 25 months post-death (CUR25) resulting in a total of 8 time periods being studied. Let the $369 \times 8$ matrix $Y$ represent the BAMO scores for each subject at each time. Define
the response indicator matrix $R$ to be the $369 \times 8$ matrix of 0’s and 1’s such that

$$R_{ij} = \begin{cases} 1, & y_{ij} \text{ observed,} \\ 0, & y_{ij} \text{ missing.} \end{cases}$$

(3.1)

The relationship between $R$ and $Y$ will influence which methods of estimation are used so it is important to establish an appropriate missing data mechanism.

Little and Rubin (1987) discuss three primary types of missing data mechanisms: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). MCAR is the simplest yet least realistic setting. Breaking up the hypothetical complete data into its observed and missing components we can write $Y = (Y_{\text{obs}}, Y_{\text{mis}})$. The situation of MCAR assumes the observed values, $Y_{\text{obs}}$, constitute a random sub-sample of the complete data $Y$. Viewing $R$ as a random variable, Little and Schenker (1995) define missing data to be MCAR provided

$$f(R \mid Y, \psi) = f(R \mid \psi) \text{ for all } Y,$$

(3.2)

where $f(R \mid Y, \psi)$ is the density of the conditional distribution of $R$ given $Y$ and $\psi$ parameterizes the rates of response. Equation 3.2 is the equivalent of saying the probability of response is unrelated to the values in $Y$ so that an analysis can proceed using only those observations whose data is complete.

The MAR setting allows missingness to depend on the observed data but not the unobserved data. The probability of response depends solely on the observed data. MAR is characterized as

$$f(R \mid Y_{\text{obs}}, Y_{\text{mis}}, \psi) = f(R \mid Y_{\text{obs}}, \psi) \text{ for all } Y_{\text{mis}}.$$

(3.3)

The MNAR scenario assumes the probability of response depends on unobserved data as well as the observed data. The left-hand side of Equation 3.3 cannot be
simplified. This situation is the most complicated in terms of likelihood-based inferences because it requires modeling the underlying missing data mechanism. In this context, missing data mechanisms falling under the MNAR category are referred to as non-ignorable.

Since MCAR is a special case of MAR, all missing data mechanisms can be classified as either MAR or MNAR. Let $\theta$ be the parameters associated with the distribution of $Y$. Likelihood-based inferences are typically focused on $\theta$, and the $\psi$ are considered nuisance parameters. MAR is preferred because it assumes $Y_{mis}$ has been integrated out of the likelihood and is therefore ignorable. If it can further be assumed that parameters $\theta$ and $\psi$ are distinct, then the joint distribution of $Y$ and $R$ can be written as the product of two densities with distinct sets of parameters:

$$f(Y_{obs}, R \mid \theta, \psi) = f(Y_{obs} \mid \theta) f(R \mid Y_{obs}, \psi).$$  \hspace{1cm} (3.4)

Hence, likelihood-based inferences about $\theta$ using $f(Y_{obs}, R \mid \theta, \psi)$ are equivalent to those relying solely on $f(Y_{obs} \mid \theta)$.

### 3.1.2 Missing the Initial Interview

Table 3.1 summarizes the number of subjects missing at each of the four interviews. An unusual aspect of this study is the inclusion of the 41 subjects who were

<table>
<thead>
<tr>
<th>1st Interview</th>
<th>2nd Interview</th>
<th>3rd Interview</th>
<th>4th Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>41 (11.1%)</td>
<td>58 (15.7%)</td>
<td>103 (27.9%)</td>
<td>125 (33.9%)</td>
</tr>
</tbody>
</table>

Table 3.1: The number of subjects missing at each interview out of 369 subjects.
not interviewed one month post-death but were allowed to participate in the study nonetheless. According to the researchers of the Grief Research Project, the families were contacted after the parental death in a time span consistent with other families in the study. Although the families agreed to participate in the study, the first interview was not obtained until close to 6 months post-death in a majority of the cases. The researcher’s general impression regarding this delay was that these families were too overwhelmed by the death. Additional follow-up interviews were conducted nevertheless, since the families had agreed to do the study.

Figure 3.1 and Table 3.2 provide summaries of the BAMO scores for the 41 subjects who missed the initial interview and the 328 subjects who participated. They appear

![Figure 3.1: Boxplot summaries of the BAMO scores at each time for the two groups of subjects allocated by whether or not an initial interview was conducted.](image-url)
to substantiate the general impressions of the researchers. The BAMO scores for the subjects missing the initial data have consistently higher quartiles across all time periods.

<table>
<thead>
<tr>
<th>Subjects Missing the Initial Interview</th>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Mean</th>
<th>Q3</th>
<th>Max</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>41</td>
</tr>
<tr>
<td>CUR1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>41</td>
</tr>
<tr>
<td>SLI6</td>
<td>0</td>
<td>0.79</td>
<td>1.28</td>
<td>1.76</td>
<td>2.20</td>
<td>7.02</td>
<td>37</td>
</tr>
<tr>
<td>CUR6</td>
<td>0</td>
<td>0.57</td>
<td>0.97</td>
<td>1.42</td>
<td>1.66</td>
<td>5.87</td>
<td>37</td>
</tr>
<tr>
<td>SLI13</td>
<td>0</td>
<td>0.79</td>
<td>1.24</td>
<td>1.63</td>
<td>2.51</td>
<td>4.75</td>
<td>25</td>
</tr>
<tr>
<td>CUR13</td>
<td>0</td>
<td>0.60</td>
<td>1.11</td>
<td>1.36</td>
<td>1.96</td>
<td>4.23</td>
<td>25</td>
</tr>
<tr>
<td>SLI25</td>
<td>0</td>
<td>0.40</td>
<td>0.69</td>
<td>1.15</td>
<td>1.89</td>
<td>4.71</td>
<td>25</td>
</tr>
<tr>
<td>CUR25</td>
<td>0</td>
<td>0.29</td>
<td>0.50</td>
<td>0.85</td>
<td>0.91</td>
<td>4.46</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subjects Participating in the Initial Interview</th>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Mean</th>
<th>Q3</th>
<th>Max</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI1</td>
<td>0</td>
<td>0.36</td>
<td>0.79</td>
<td>1.01</td>
<td>1.43</td>
<td>3.98</td>
<td>328</td>
</tr>
<tr>
<td>CUR1</td>
<td>0</td>
<td>0.67</td>
<td>1.14</td>
<td>1.31</td>
<td>1.79</td>
<td>4.56</td>
<td>328</td>
</tr>
<tr>
<td>SLI6</td>
<td>0</td>
<td>0.30</td>
<td>0.70</td>
<td>0.97</td>
<td>1.30</td>
<td>5.66</td>
<td>274</td>
</tr>
<tr>
<td>CUR6</td>
<td>0</td>
<td>0.22</td>
<td>0.58</td>
<td>0.84</td>
<td>1.14</td>
<td>4.95</td>
<td>274</td>
</tr>
<tr>
<td>SLI13</td>
<td>0</td>
<td>0.24</td>
<td>0.61</td>
<td>0.86</td>
<td>1.15</td>
<td>5.46</td>
<td>241</td>
</tr>
<tr>
<td>CUR13</td>
<td>0</td>
<td>0.18</td>
<td>0.49</td>
<td>0.72</td>
<td>0.97</td>
<td>5.33</td>
<td>241</td>
</tr>
<tr>
<td>SLI25</td>
<td>0</td>
<td>0.23</td>
<td>0.56</td>
<td>0.84</td>
<td>1.21</td>
<td>5.34</td>
<td>219</td>
</tr>
<tr>
<td>CUR25</td>
<td>0</td>
<td>0.12</td>
<td>0.45</td>
<td>0.66</td>
<td>0.90</td>
<td>3.47</td>
<td>219</td>
</tr>
</tbody>
</table>

Table 3.2: Summary statistics of the BAMO scores for each time period for the subjects who missed the initial interview \( (n = 41) \) and those who participated in the initial interview \( (n = 328) \). The labels Q1 and Q3 refer to the first and third quartiles, respectively.

It appears that missing the initial interview is directly related to a heightened emotional state which, in turn, corresponds to higher BAMO scores. This translates to initial missingness being related to unobserved data and thus being non-ignorable.
(Little and Schenker 1995). As a result, the data from the 41 subjects who were not interviewed one month post-death were excluded from the analysis. Subsequent summaries and analysis will be based on the 328 subjects who participated in the initial interview.

3.1.3 Exploring Missing Data Patterns

Each subject had the potential to participate in four interviews. Table 3.3 provides a breakdown of the number of subjects according to number of interviews missed.

<table>
<thead>
<tr>
<th>0 Missed</th>
<th>1 Missed</th>
<th>2 Missed</th>
<th>3 Missed</th>
</tr>
</thead>
<tbody>
<tr>
<td>201</td>
<td>45</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>(61.3%)</td>
<td>(13.7%)</td>
<td>(12.5%)</td>
<td>(12.5%)</td>
</tr>
</tbody>
</table>

Table 3.3: The number of subjects according to number of missed interviews.

Sixty-one percent of the 328 subjects completed all four interviews. The remaining thirty-nine percent who missed interviews were split fairly evenly between missing one, two or three interviews, with slightly more subjects missing a single interview. Table 3.4 displays the missing data patterns in more detail where 0 represents a missed interview and 1 signifies an interview was obtained. Categories 2, 3 and 4 indicate the majority of the subjects who missed an interview were permanent dropouts; once an interview was missed, all subsequent interviews were missed. The nineteen subjects corresponding to the last four categories were intermittent dropouts in the sense that they may have missed one or two interviews but participated again in later interviews.
<table>
<thead>
<tr>
<th>Category</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>201</td>
<td>61.3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>28</td>
<td>8.5</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>39</td>
<td>11.9</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>41</td>
<td>12.5</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>11</td>
<td>3.4</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>1.8</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Table 3.4: Pattern of missing data across four interview periods where 0 represents a missed interview and 1 represents an interview obtained. The ‘n’ column indicates the number of subjects falling into each category.

Figure 3.2 shows the observed mean response as a function of time corresponding to the different missing data patterns listed in Table 3.4. Although the mean response profiles for categories 7 and 8 are clearly distinct from the remaining profiles, particularly at the initial interview, any conclusions made would be suspect since each relies on the contribution of only one subject.

The mean response profiles associated with categories 1 through 6 are very similar. All exhibit elevated emotional responses at the time of death which gradually decrease over time. With the exception of category 4, the mean scores for each category do not increase immediately prior to a missed interview which would have suggested the relationship between the BAMO score and the missing data process is non-ignorable.

### 3.1.4 Missing Data Mechanisms and the BAMO Data

A MCAR assumption would simplify estimation procedures by allowing an analysis of only complete data. Such an assumption implies that missingness in a data
Figure 3.2: Observed mean BAMO response over time according to each missing data pattern. The numbers in the plot correspond to the categories listed in Table 3.4.

set is random and therefore unrelated to external elements. Subsequent analysis will reveal associations between the rate of non-response and several variables, exposing the inadequacy of a MCAR assumption for the BAMO data. For example, the increase in non-response rates over time reported in Table 3.5 indicates missingness is associated with time.

<table>
<thead>
<tr>
<th>1st Interview</th>
<th>2nd Interview</th>
<th>3rd Interview</th>
<th>4th Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (0.0%)</td>
<td>54 (16.5%)</td>
<td>87 (26.5%)</td>
<td>109 (33.2%)</td>
</tr>
</tbody>
</table>

Table 3.5: The number of subjects missing at each interview.
Response rates are also associated with age group. Subjects were classified into two groups based on their age at the time of parental death. A subject was classified as a child if the subject’s age at the time of death was 12 years or less and adolescent if it was 13 years or greater. Figure 3.3 indicates higher non-response rates for adolescents over the course of the four interviews. Factors such as school or work commitments may partially account for the higher rate of non-response exhibited by the adolescent group. If emotional distress is a major factor in higher adolescent non-response rates however, one might expect to see higher BAMO scores for the adolescent group. Table 3.6 summarizes within each age group the mean BAMO scores for subjects with complete data and those with incomplete data. No clear pattern emerges when comparing children with complete data and children with incomplete data. In the adolescent group, however, those with incomplete data have consistently

Figure 3.3: Nonresponse rates for the child \((n = 183)\) and adolescent \((n = 145)\) groups at each of the follow-up interviews.
higher mean scores than those with complete data. This suggests an apparent relationship between missingness and age group. It is interesting to note that for subjects with complete data, the children have consistently higher mean scores compared to the adolescent group. The opposite trend occurs when comparing children and adolescents with incomplete data, with the exception of the results at the 25 month time period which are not as reliable since only nine subjects contribute to the mean scores at that time.

<table>
<thead>
<tr>
<th>Children With Complete Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI1 CUR1 SLI6 CUR6 SLI13 CUR13 SLI25 CUR25</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>SE</td>
</tr>
<tr>
<td>n</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Children With Incomplete Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI1 CUR1 SLI6 CUR6 SLI13 CUR13 SLI25 CUR25</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>SE</td>
</tr>
<tr>
<td>n</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adolescents With Complete Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI1 CUR1 SLI6 CUR6 SLI13 CUR13 SLI25 CUR25</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>SE</td>
</tr>
<tr>
<td>n</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adolescents With Incomplete Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI1 CUR1 SLI6 CUR6 SLI13 CUR13 SLI25 CUR25</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>SE</td>
</tr>
<tr>
<td>n</td>
</tr>
</tbody>
</table>

Table 3.6: Observed mean BAMO scores for those subjects with complete data and those with incomplete data within each age group. Standard errors (SE) and sample sizes corresponding to each time period are included.
Initially, the study protocol only permitted inclusion of children whose primary stressor was the parental death. It soon became apparent, however, that such criteria would exclude many children from the study. The inclusion criteria was relaxed somewhat to admit other children. The children that met the initial criteria were classified as “simple” bereaved while the remaining children were classified as “complex” bereaved. Some examples of being classified as complex are death in the extended family in the two years prior to death or mental health contact in the previous two years by an immediate family member.

Figure 3.4 shows the complex group has higher rates of missingness than the simple group over the course of the study. Table 3.7 summarizes the mean BAMO scores for subjects with complete data and those without for the simple and complex groups. The complex group with incomplete data has consistently higher mean BAMO scores through the first 13 months compared to the complex group with complete data.

Figure 3.4: Nonresponse rates for the simple (n = 174) and complex (n = 154) groups at each of the follow-up interviews.
By 25 months the mean scores are smaller for the complex group having incomplete data but only 8 subjects contribute to the mean score at this time. This pattern suggests a link exists between missingness and the BAMO measure via the inclusion criteria status (ICS). The trend reverses for the simple group when comparing those with missing data and those without. The mean scores for the complex group are generally higher than the simple group whether or not missing data is present.

<table>
<thead>
<tr>
<th>Simple Group With Complete Data</th>
<th>SLI1</th>
<th>CUR1</th>
<th>SLI6</th>
<th>CUR6</th>
<th>SLI13</th>
<th>CUR13</th>
<th>SLI25</th>
<th>CUR25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.03</td>
<td>1.37</td>
<td>1.01</td>
<td>0.89</td>
<td>0.86</td>
<td>0.71</td>
<td>0.92</td>
<td>0.74</td>
</tr>
<tr>
<td>SE</td>
<td>0.07</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
<td>0.07</td>
<td>0.06</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>n</td>
<td>124</td>
<td>124</td>
<td>124</td>
<td>124</td>
<td>124</td>
<td>124</td>
<td>124</td>
<td>124</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Simple Group With Incomplete Data</th>
<th>SLI1</th>
<th>CUR1</th>
<th>SLI6</th>
<th>CUR6</th>
<th>SLI13</th>
<th>CUR13</th>
<th>SLI25</th>
<th>CUR25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.85</td>
<td>1.20</td>
<td>1.07</td>
<td>0.81</td>
<td>1.07</td>
<td>0.93</td>
<td>0.85</td>
<td>0.64</td>
</tr>
<tr>
<td>SE</td>
<td>0.10</td>
<td>0.12</td>
<td>0.20</td>
<td>0.14</td>
<td>0.30</td>
<td>0.29</td>
<td>0.42</td>
<td>0.30</td>
</tr>
<tr>
<td>n</td>
<td>59</td>
<td>59</td>
<td>39</td>
<td>39</td>
<td>22</td>
<td>22</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Complex Group With Complete Data</th>
<th>SLI1</th>
<th>CUR1</th>
<th>SLI6</th>
<th>CUR6</th>
<th>SLI13</th>
<th>CUR13</th>
<th>SLI25</th>
<th>CUR25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.87</td>
<td>1.06</td>
<td>0.83</td>
<td>0.72</td>
<td>0.73</td>
<td>0.61</td>
<td>0.73</td>
<td>0.53</td>
</tr>
<tr>
<td>SE</td>
<td>0.08</td>
<td>0.09</td>
<td>0.09</td>
<td>0.08</td>
<td>0.09</td>
<td>0.08</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>n</td>
<td>77</td>
<td>77</td>
<td>77</td>
<td>77</td>
<td>77</td>
<td>77</td>
<td>77</td>
<td>77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Complex Group With Incomplete Data</th>
<th>SLI1</th>
<th>CUR1</th>
<th>SLI6</th>
<th>CUR6</th>
<th>SLI13</th>
<th>CUR13</th>
<th>SLI25</th>
<th>CUR25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.27</td>
<td>1.56</td>
<td>1.04</td>
<td>0.94</td>
<td>1.16</td>
<td>1.01</td>
<td>0.75</td>
<td>0.54</td>
</tr>
<tr>
<td>SE</td>
<td>0.11</td>
<td>0.11</td>
<td>0.15</td>
<td>0.14</td>
<td>0.28</td>
<td>0.27</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>n</td>
<td>68</td>
<td>68</td>
<td>34</td>
<td>34</td>
<td>18</td>
<td>18</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 3.7: Observed mean BAMO scores for those subjects with complete data and those with incomplete data for each ICS level. Standard errors (SE) and sample sizes corresponding to each time period are included.
The Hollingshead Index (Hollingshead and Redlich 1958) is used to assess a family’s socioeconomic status (SES) where 1 = upper class, 2 = upper middle class, 3 = middle class, 4 = lower middle class and 5 = lower class. Figure 3.5 reveals that a low socioeconomic status is related to a higher rate of non-response. The mean BAMO scores of the families with Hollingshead indices between 1 and 4 are fairly indistinguishable from one another as seen in Figure 3.6. The lowest class has noticeably higher mean BAMO scores at all time periods. These results indicate an association between non-response rates and BAMO scores via economic status.

Figure 3.5: Non-response rates according to socioeconomic status: 1 = upper class, 2 = upper middle class, 3 = middle class, 4 = lower middle class, 5 = lower class. The number of subjects, \( n_i \), with SES level \( i \) is \( n_1=43 \), \( n_2=77 \), \( n_3=89 \), \( n_4=97 \) and \( n_5=22 \).

The association of non-response rates with time, age group, SES and ICS indicate that MCAR is not a valid assumption. The apparent association between response rates and BAMO scores via SES and ICS supports the premise that MNAR is the
Figure 3.6: Mean BAMO scores over time according to socioeconomic status: 1 = upper class, 2 = upper middle class, 3 = middle class, 4 = lower middle class, 5 = lower class. The number of subjects, $n_i$, with SES level $i$ is $n_1=43$, $n_2=77$, $n_3=89$, $n_4=97$ and $n_5=22$.

appropriate missing data mechanism for this data. Sufficient information is not available, unfortunately, to formulate a suitable model for the missing data mechanism.

It is conceivable, however, that the occurrence of missing data depends only on the completely observed variables previously mentioned and not on the actual BAMO data itself. That is, the apparent relationship between the BAMO scores and missingness is just a byproduct of the association between the completely observed variables and the response variable. This suggests that non-response bias may be controlled by adjusting for these completely observed variables. To accomplish this, time, age group, SES classification and ICS will be used as predictor variables in the hope that the underlying missing data mechanism is made closer to MAR (Brand 1999).
3.2 Demographics

At the time of parental death, the 328 subjects ranged in age from 5 years to 18 years with a median age of 12 years. Recall that the subjects are classified into one of two groups based on their age at the time of death: adolescent if they were 13 years or older and child otherwise. The median age of the 145 adolescents is 14 years and the median age of the 183 children is 10 years. There are 165 female and 163 male participants in the study. The ethnic background of the subjects is overwhelmingly Caucasian ($n = 320$). Four of the subjects are African American, two are Asian and another two have other ethnic backgrounds.

<table>
<thead>
<tr>
<th>Number of Siblings</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>81</td>
<td>79</td>
<td>20</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>%</td>
<td>43.32</td>
<td>42.25</td>
<td>10.70</td>
<td>3.21</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Table 3.8: Number of families that participated in the study according to number of siblings whose data was analyzed.

Many of the subjects have siblings partaking in the study. The 328 subjects comprise a total of 187 families. Table 3.8 summarizes various family sizes according to the number of siblings whose data is utilized. The majority of subjects are from single-child and two-child families. Ninety-eight of these families met the initial inclusion criteria while 89 families were admitted using the relaxed criteria. This translated into 174 subjects being classified as simple bereaved and 154 as complex bereaved. The socioeconomic status of the families as measured by the Hollingshead score is displayed in Table 3.9. The majority of the families fall into middle class
<table>
<thead>
<tr>
<th>Hollingshead Score</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>22</td>
<td>46</td>
<td>54</td>
<td>52</td>
<td>13</td>
</tr>
<tr>
<td>%</td>
<td>11.76</td>
<td>24.6</td>
<td>28.88</td>
<td>27.81</td>
<td>6.95</td>
</tr>
</tbody>
</table>

Table 3.9: Hollingshead score assessing socioeconomic status by family: 1=upper class, 2=upper middle class, 3=middle class, 4=lower middle class and 5=lower class.

categories. Less than twenty percent of the families are classified as either upper or lower class. In terms of number of subjects, there are 43 upper class, 77 upper middle class, 89 middle class, 97 lower middle class and 22 lower class subjects. Appendix A contains more demographic details. Appendix B provides a demographic summary of the 41 subjects whose data was omitted from the analysis as a result of missing the initial interview. These two groups are very similar demographically. Some dissimilarities worth noting are the larger percentage of children in the group without initial interviews, 68.3% versus 55.8%. In particular, the percentage of children 5 and 6 years of age in this group is 24.4% compared to only 4.3% of the group having initial interviews.

3.3 Mixture Model Application to the Research Problem

The research question of interest is whether or not a more emotionally traumatized subgroup exists within the population of children who experience death of a parent. Researchers from the grief study speculate that parental death will spark a prolonged period of heightened emotional sensitivity to psychosocial stressors in some of the children resulting in elevated levels of behavioral, anxiety and depressive symptoms. The justification underlying this hypothesis is linked to the theory of “kindling”.

The phenomenon of kindling was introduced by Goddard (1967) while investigating the effects of low intensity electrostimulation of the brain on the learning process in rats. Over time, their brains became so sensitized that they would convulse when stimulated. Goddard, McIntyre and Leech (1969) showed the epileptic-like seizures could also be elicited chemically.

The kindling model within a psychiatric context postulates major psychosocial stressors are the source of the initial episode of recurrent affective disorders such as bipolar disorder. As the disorder progresses, subsequent episodes become more frequent and need not be induced by external triggers. In other words, “episodes beget episodes”. This kindling analogy for recurrent mood disorders was first proposed by Post (1992). His research shows that neurobiological pathways can be altered due to repeated exposure to psychosocial stressors and to the impact of the actual episodes. Post’s claim that changes at the genetic level may be occurring is supported by research in which scientists have recently identified a gene linked to susceptibility to depression (Caspi, Sugden, Moffit et al. 2003).

If parental death triggers a kindling-like effect in some of the children who experience parental death, then within the population of grieving children, an emotionally susceptible group coexists with a less sensitive group of children. A natural model to determine whether two such groups exist is a two-component mixture model.

The longitudinal aspect of the study necessitates using a multivariate distribution in any underlying model. An advantage of utilizing the multivariate normal distribution is the mathematically intractable nature of other multivariate distributions. A second advantage is that parameter estimates and hypothesis testing are typically
more robust to departures from normality of the parent population. Parameter estimation involving a mixture model can be complicated and the multivariate nature of the grief data adds another degree of complexity. Assuming a multivariate normal mixture distribution will simplify the process of mixture identification.

Subsequent sections investigate modeling the data as a two-component multivariate normal mixture distribution. The validity of a multivariate normal assumption is explored in the next section. Techniques outlined in Lindsay and Basak (1993) to obtain method of moments (MOM) estimates are described thereafter, followed by a summary of their application to the grief data.

3.3.1 Data Transformation

Recall the BAMO score is a weighted sum of psychological indicators of problems in the areas of behavior, anxiety, mood and other disorders. Scores can range from 0 to 18. The maximum score in this data did not exceed 6 although the majority of scores

Figure 3.7: Histograms of the marginal distributions of the BAMO scores.
are less than 3 so that the data is skewed. Histograms of the BAMO marginal distributions in Figure 3.7 reveal positive skewness, particularly at the follow-up times. This is not surprising since one would expect the emotional upheaval caused by the parental death to ease over time which will result in lower BAMO scores for the majority of the children. The small number of high BAMO scores that remain after two years, however, suggest the possible existence of a subgroup of children who may be unable to adapt to the loss of a parent.

Figure 3.8 displays a scatterplot matrix of all possible combinations of BAMO outcomes from different times. In almost all comparisons the BAMO scores become more
spread-out as they increase in value. The only exceptions occur between responses at times corresponding to the same interview period, which are highly correlated. That the variability grows in relation to the magnitude of the BAMO score is not surprising due to the Poisson-like nature of the data since the BAMO score is derived from symptom counts. In this type of situation, typical approaches to stabilize the variance involve transforming the data via a log, reciprocal or square-root transformation.

The reciprocal transformation of the BAMO data created skewed marginal distributions with long right tails. The log transformation yielded marginal distributions that were more symmetrical but heavy in the tails. The square-root transformation produced the best results. Figure 3.9 displays histograms of the marginal distributions of the square-root transformed BAMO data. A slight amount of skewness remains but is much less pronounced compared to the untransformed data. The scatterplots in Figure 3.10 exhibit the desired elliptical shape indicating variance stabilization was

---

**Figure 3.9**: Histograms of the marginal distributions of the BAMO scores after the square-root transformation.
reasonably successful. An inconsistency is the presence of values near zero in quantities larger than might be expected from a normal distribution. Although normalized margins and elliptical-shaped scatterplots do not guarantee multivariate normality, they do strengthen the validity of that assumption. As a result, the square-root transformed BAMO data will be utilized and hereafter referred to as SqRt-BAMO.

Figure 3.10: Scatterplots of all time comparisons using the SqRt-BAMO data.

3.3.2 Lindsay’s Method of Moments Procedure

Lindsay (1989b) devised a computationally efficient method to obtain unique MOM estimates in the univariate normal mixture setting. The basic strategy is
outlined as follows. Suppose \( X \) is distributed as \( N(Q_p, \sigma) \). Let \( Q_p \) be a discrete distribution with \( p \) points of support, \( \theta_1, \ldots, \theta_p \), having corresponding probabilities, \( \pi_1, \ldots, \pi_p \). Let the \( k \)th moment of \( Q_p \) be denoted by \( m_k \). Use the sample moments of \( X \) to estimate the first \( K \) moments of \( Q_p \). Reconstruct \( \hat{Q}_p \) from the estimated moments \( \hat{m}_k, k = 1, \ldots, K \).

Two years later, Lindsay and Basak (1991) generalized the univariate theory to the bivariate setting. Soon after, they developed a proficient estimation procedure that calculated a method of moments estimate of the multivariate normal mixture distribution in any dimension (Lindsay and Basak 1993). The unbiased parameter estimates are consistent and provide effective starting values for search algorithms seeking to maximize the likelihood function.

Let \( X = (X_1, Y_2, \ldots, Y_d)' \) be a \( d \times 1 \) random vector distributed according to the multivariate normal distribution, \( N(Q_p, \Sigma) \). Define \( \mu = (\theta, \phi')' \) to be the corresponding mean vector. The first coordinate \( X_1 \) will be called the primary axis because of its importance in the estimation procedure. The \( (p + d) \times (p + d) \) primary axis moment matrix is defined as \( M_p = E[V_p(\mu)V_p(\mu)'] \) where \( V_p(\mu) = (1, \theta_1, \ldots, \theta_p, \phi_2, \phi_3, \ldots, \phi_d)' \). In the case of the SqRt-BAMO data, using \( p = 2 \) and \( d = 8 \), \( M_2 \) has the form

\[
M_2 = E \begin{pmatrix}
1 & \theta & \theta^2 & \theta^3 & \theta^4 & \theta^5 & \theta^6 & \theta^7 & \theta^8 \\
\theta & \theta^2 & \theta^3 & \theta^4 & \theta^5 & \theta^6 & \theta^7 & \theta^8 & \theta^9 \\
\theta^2 & \theta^3 & \theta^4 & \theta^5 & \theta^6 & \theta^7 & \theta^8 & \theta^9 & \theta^{10} \\
\theta^3 & \theta^4 & \theta^5 & \theta^6 & \theta^7 & \theta^8 & \theta^9 & \theta^{10} & \theta^{11} \\
\theta^4 & \theta^5 & \theta^6 & \theta^7 & \theta^8 & \theta^9 & \theta^{10} & \theta^{11} & \theta^{12} \\
\theta^5 & \theta^6 & \theta^7 & \theta^8 & \theta^9 & \theta^{10} & \theta^{11} & \theta^{12} & \theta^{13} \\
\theta^6 & \theta^7 & \theta^8 & \theta^9 & \theta^{10} & \theta^{11} & \theta^{12} & \theta^{13} & \theta^{14} \\
\theta^7 & \theta^8 & \theta^9 & \theta^{10} & \theta^{11} & \theta^{12} & \theta^{13} & \theta^{14} & \theta^{15} \\
\theta^8 & \theta^9 & \theta^{10} & \theta^{11} & \theta^{12} & \theta^{13} & \theta^{14} & \theta^{15} & \theta^{16}
\end{pmatrix}
\]
Note the influence of $\theta$, the mean parameter coordinate corresponding to $X_1$. The upper $(p+1) \times (p+1)$ matrix of $M_2$, consisting solely of the moments of $\theta$, is involved in the estimation of the component means and population proportions as well as the elements of the common covariance matrix $\Sigma$. Hence, the primary axis plays an important role in estimation which suggests using discretion during its selection.

The basic steps of the method of moments technique developed by Lindsay and Basak (1993) are summarized below. Each step utilizes mean-centered data and assumes $p = 2$ throughout. Steps 1, 2 and 3 employ the univariate techniques described by Lindsay (1989b) to estimate the marginal distribution of the primary axis resulting in estimates of $\hat{\theta}_1, \hat{\theta}_2, \hat{\pi}_1, \hat{\pi}_2$ and $\hat{\sigma}_{11}$. The quantity $\hat{\sigma}_{11}$ is the variance estimate associated with the primary axis. Denote the estimate of $E(\theta^k)$ by $\hat{m}_k(\hat{\sigma}_{11})$.

**Step 1: Estimate $\sigma_{11}$**

Use the moments of $X_1$ to compute $\hat{\sigma}_{11}$ by solving for the smallest positive root of the cubic polynomial defined by

$$\det \begin{pmatrix} 1 & \hat{m}_1(\sigma_{11}) & \hat{m}_2(\sigma_{11}) \\ \hat{m}_1(\sigma_{11}) & \hat{m}_2(\sigma_{11}) & \hat{m}_3(\sigma_{11}) \\ \hat{m}_2(\sigma_{11}) & \hat{m}_3(\sigma_{11}) & \hat{m}_4(\sigma_{11}) \end{pmatrix} = 0.$$

**Step 1*: Select primary axis**

This step is optional. Repeat Step 1 for each axis and calculate the mixture variance $\hat{m}_2(\hat{\sigma}_{11})$ and axis variance $s_j^2$. The axis producing the largest ratio $\hat{m}_2(\hat{\sigma}_{11})/s_j^2$, $j = 1, \ldots, d$ is selected as the primary axis.

**Step 2: Estimate $\theta_1$ and $\theta_2$**

Letting $\hat{m}_k = \hat{m}_k(\hat{\sigma}_{11})$, $\hat{\theta}_1$ and $\hat{\theta}_2$ are the roots of the polynomial defined by

$$\det \begin{pmatrix} 1 & \hat{m}_1 & 1 \\ \hat{m}_1 & \hat{m}_2 & t \\ \hat{m}_2 & \hat{m}_3 & t^2 \end{pmatrix} = 0.$$
Step 3: Estimate \( \pi_1 \) and \( \pi_2 \)

Solve the following system of equations for \( \hat{\pi}_k, k = 1, \ldots, p \):

\[
\begin{pmatrix}
1 & 1 & \cdots & 1 \\
\hat{\theta}_1 & \hat{\theta}_2 & \cdots & \hat{\theta}_p \\
\hat{\theta}_1^2 & \hat{\theta}_2^2 & \cdots & \hat{\theta}_p^2 \\
\cdots & \cdots & \cdots & \cdots \\
\hat{\theta}_1^{p-1} & \hat{\theta}_2^{p-1} & \cdots & \hat{\theta}_p^{p-1}
\end{pmatrix}
\begin{pmatrix}
\hat{\pi}_1 \\
\hat{\pi}_2 \\
\hat{\pi}_3 \\
\vdots \\
\hat{\pi}_p
\end{pmatrix}
= 
\begin{pmatrix}
1 \\
\hat{m}_1 \\
\hat{m}_2 \\
\vdots \\
\hat{m}_{p-1}
\end{pmatrix}.
\]

Step 4: Estimate \( \sigma_{1k} = \text{Cov}(X_1, Y_k) \) for \( k = 1, \ldots, d \)

\[
\hat{\sigma}_{1k} = \hat{E}(X_1 Y_k) - \frac{6\hat{m}_2 \hat{E}(X_1 Y_k)(\hat{m}_2 + \hat{\sigma}_{11}) + 3\hat{m}_3 \hat{E}(X_1^2 Y_k) - 2\hat{m}_2 \hat{E}(X_1^2 Y_k)}{\hat{m}_4 + 3\hat{m}_2^2}
\]

Step 5: Estimate \( \phi_{jk} \)

Let \( \hat{c}_{jk} = \hat{E}(\theta^j \phi_k) \). Compute the \( k \)th coordinate \( \hat{\phi}_{jk} \) of the support point \( \hat{\mu}_j \), \( j = 1, 2, k = 1, \ldots, d \) by solving

\[
\text{det}
\begin{pmatrix}
1 & \hat{m}_1 & 1 \\
\hat{m}_1 & \hat{m}_2 & \hat{\theta}_j \\
\hat{c}_{0k} & \hat{c}_{1k} & \hat{\phi}_{jk}
\end{pmatrix}
= 0.
\]

Step 6: Estimate the remaining elements of \( \Sigma \)

Assume \( S = \Sigma + \Sigma(Q_p) \) where \( S \) is the sample covariance matrix and \( \Sigma(Q_p) \) is the covariance matrix of \( Q_p \). The remaining elements of \( \hat{\Sigma} \) are estimated by solving

\[
\hat{\Sigma} = S - \hat{\Sigma}(\hat{Q}_p)
\]

where \( \hat{\pi}_1(1 - \hat{\pi}_1)(\hat{\mu}_1 - \hat{\mu}_2)(\hat{\mu}_1 - \hat{\mu}_2)' \) is used in place of \( \hat{\Sigma}(\hat{Q}_p) \).

Step 7: Constrain \( \hat{\Sigma} \) to be positive definite

Let \( \lambda_m \) be the largest eigenvalue of \( S^{-1}\hat{\Sigma}(\hat{Q}_p) \). Define \( \alpha = \min \{1, \sqrt{(1 - \epsilon)/\lambda_m} \} \).

If \( \alpha < 1 \), then “shrink” the current estimates by substituting \( \alpha \hat{\mu}_j \) for \( \hat{\mu}_j \) and by replacing \( \hat{\Sigma} \) with \( S - \alpha \hat{\Sigma}(\hat{Q}_p) \).
Step 8: Refine parameter estimates

Repeat Steps 1 to 7 after rotating the data to make $\hat{\mu}_2 - \hat{\mu}_1$ the primary axis. Choose between the initial set of estimates and those obtained after rotation by selecting the set that yields the largest likelihood value.

3.3.3 Application to the Grief Data

The idea behind Step 1* is to associate the axis having the greatest variation in $Q$ as the best candidate for the primary axis. Applying Step 1* to the SqRt-BAMO data resulted in the selection of the CUR25 time period as the primary axis. The major drawback of this choice is that the percentage of missing data is greatest during the last interview period as indicated in Table 3.5. Since the primary axis plays such a crucial role in the estimation procedure, it is desirable to have the primary axis incorporate the most relevant information. The primary axis was selected after careful consideration of the specific research problem.

The researchers believe that an emotionally susceptible subgroup of children exists prior to parental death, but until the occurrence of a traumatic life event, the subgroup is largely undetectable. This implies the emotional state prior to the death of the parent will be uninformative regarding group affiliation. The psychological condition of the two subgroups of children just after the death will be similarly undifferentiable since all of the children will be greatly distressed.

Viewing parental death as a “psychological trigger”, one might expect that members of the susceptible group will exhibit emotional upheaval that will persist after the death. Over time, most of the bereaving children will adapt to the loss but the emotionally susceptible group will not and elevated BAMO scores will be evident.
for a prolonged period of time. As a consequence, group separation should be more
distinct post death. A logical choice for the primary axis is to use a weighted average
of BAMO scores for each subject. In order to emphasize the post death scores, a
zero weight is applied to the SLI1 and CUR1 responses. The remaining responses are
weighted according to the length of time that has elapsed from the time of death.

Since the primary axis greatly influences the estimation of $\hat{Q}_p$, being able to con-
solidate pertinent information into a single quantity makes the weighted mean a desir-
able choice for the primary axis. Using an average has the added advantage of being
less susceptible to bias since 87.5% of the subjects have follow-up data to contribute
unlike the CUR25 time frame, where one-third of the responses are missing.

Let $X$ represent the $328 \times 8$ SqRt-BAMO data matrix and denote the weighted
average of the follow-up responses by $X_W$. The weights involved in calculating $X_W$ are
\[ w_i = t_i / \sqrt{\sum_{i=1}^{8} t_i^2} \] where $t_i$, $i = 1, \ldots, 8$ represent the time elapsed since the parental
death with $t_1 = t_2 = 0$. Using $X_W$ as the primary axis requires a transformation of
$X$ to $X^* = XW$ via the orthonormal transformation matrix $W$ defined by

\[
W = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
w_3 & 0 & 0 & \sqrt{\frac{w_4^2}{w_3^2+w_4^2}} & 0 & 0 & c_1 & d_1 \\
w_4 & 0 & 0 & -\sqrt{\frac{w_3^2}{w_3^2+w_4^2}} & 0 & 0 & c_2 & d_2 \\
w_5 & 0 & 0 & 0 & \sqrt{\frac{w_6^2}{w_5^2+w_6^2}} & 0 & c_3 & d_3 \\
w_6 & 0 & 0 & 0 & -\sqrt{\frac{w_5^2}{w_5^2+w_6^2}} & 0 & c_4 & d_4 \\
w_7 & 0 & 0 & 0 & 0 & \sqrt{\frac{w_8^2}{w_7^2+w_8^2}} & 0 & d_5 \\
w_8 & 0 & 0 & 0 & 0 & -\sqrt{\frac{w_7^2}{w_7^2+w_8^2}} & 0 & d_6
\end{pmatrix}.
\]
The non-primary axes are designed to minimize the effect of missing data. Columns 2 and 3 in \( W \) will copy SLI1 and CUR1 data directly from \( X \). Neither has missing data. Columns 4 through 6 in \( W \) will not introduce any new missing data into \( X^* \) since each pair of differences corresponds to data obtained at the same interview. For example, SLI6 and CUR6 represent different times, but the responses corresponding to those times were obtained at the same interview. As a result, a subject missing data at SLI6 will also be missing data at CUR6. Conversely, a subject with non-missing values at CUR6 will have non-missing values at SLI6. The same is true for the 13-month and 25-month interviews. The last two columns introduce additional missingness since they involve linear combinations of data from different interview periods. These linear combinations are such that the resulting columns are orthonormal to all the other columns in \( W \).

The weighted mean \( \bar{X}_W \) averages over non-missing values. The advantage of this construction is that using \( \bar{X}_W \) as the primary axis will result in fewer missing values than if CUR25 is the primary axis. Whereas CUR25 has a 33.9% non-response rate, \( \bar{X}_W \) has only 12.5% missingness. The disadvantage is that the transformation matrix \( W \) introduces a small amount of additional missing data to the non-primary axes in \( X^* \). This tradeoff is considered acceptable since the primary axis plays a more prominent role in the estimation procedure than the non-primary axes.

The primary axis candidates \( \bar{X}_W \) and CUR25 are evaluated by comparing model estimates and log-likelihood values. Table 3.10 summarizes the log-likelihood values corresponding to these primary axis choices. Method of moments and maximum likelihood results are reported. The EM algorithm is used to obtain maximum likelihood results (Dempster et al. 1977).
The log-likelihood values corresponding to the columns headed by MOM* and ML* in Table 3.10 are calculated prior to transforming the parameter estimates back to the original axis system. It would have been unnecessary to distinguish between before and after transformation log-likelihood values if the data had been complete since the multivariate normal distribution is invariant to location and scale changes. That is, \( f(x|\mu, \Sigma) = f(x^*|\mu^*, \Sigma^*) \) where \( x^* = W'x, \mu^* = W'\mu, \Sigma^* = W\Sigma W' \) and \( f \) is the multivariate normal density.

<table>
<thead>
<tr>
<th>Primary Axis</th>
<th>Log-likelihood Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{X}_W )</td>
<td>MOM* -32.259</td>
</tr>
<tr>
<td>CUR25</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 3.10: Log-likelihood values resulting from various choices for the primary axis. \( \bar{X}_W \) denotes the weighted average response excluding SLI and CUR1 times. Results for both method of moments and maximum likelihood estimates are reported. The log-likelihood values in the MOM* and ML* columns are calculated prior to transforming parameters back to the original axis system. Values differ within type of estimate as a result of missing data.

Missing values in the original data \( X \) prevented the transformed data \( X^* \) from being derived directly from \( XW \). As a result, the log-likelihood values calculated using \( X^* \) and the corresponding parameter estimates, \( \hat{\pi}^*, \hat{\mu}_1^*, \hat{\mu}_2^* \) and \( \hat{\Sigma}^* \), will be different from the log-likelihood calculated using the original data \( X \) and the new parameter estimates obtained after transforming the initial ML parameter estimates back to the original axis system via \( W^{-1} \).
When CUR25 is the primary axis, the log-likelihood value is 31.220. The log-likelihood corresponding to the weighted average is only 20.880. Given the near singularity of the covariance matrix estimate in both cases, the log-likelihood is not the most reliable measure of comparison for the current data. Comparable log-likelihood values can be obtained with some of the mean parameter values of the susceptible group estimated as less than the non-susceptible values. Such instances of distinctly different parameter values yielding similar log-likelihood values indicate the likelihood function is multimodal resulting in many local maxima.

Table 3.11 compares proportion estimates of the susceptible and non-susceptible groups for the different primary axes. Using \( \bar{X}_W \) as the primary axis results in 9.5\% of the subjects being allocated to the emotionally susceptible group. When CUR25 is the primary axis, the ML estimate is more than double with a value of \( \hat{\pi}_2 = 23.5\% \).

<table>
<thead>
<tr>
<th>Primary Axis</th>
<th>MOM ( \hat{\pi} )</th>
<th>MOM ( 1 - \hat{\pi} )</th>
<th>ML ( \hat{\pi} )</th>
<th>ML ( 1 - \hat{\pi} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{X}_W )</td>
<td>0.814</td>
<td>0.186</td>
<td>0.905</td>
<td>0.095</td>
</tr>
<tr>
<td>CUR25</td>
<td>0.711</td>
<td>0.289</td>
<td>0.765</td>
<td>0.235</td>
</tr>
</tbody>
</table>

Table 3.11: Method of moments and maximum likelihood estimates of component proportions resulting from two choices for the primary axis. See Table 3.10 for detailed descriptions of the different primary axes.

Estimates of the component means \( \hat{\mu}_1 \) and \( \hat{\mu}_2 \) associated with the primary axis choices \( \bar{X}_W \) and CUR25 are displayed in Tables 3.12 and 3.13, respectively. Figure 3.11 displays plots of these results. The MOM estimates provide reasonable starting
points for the EM algorithm. Less than 15 iterations were necessary to obtain the ML estimates in each case.

<table>
<thead>
<tr>
<th>Method of Moments</th>
<th>Maximum Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>$\mu_2$</td>
</tr>
<tr>
<td>SLI1</td>
<td>0.825</td>
</tr>
<tr>
<td>CUR1</td>
<td>0.980</td>
</tr>
<tr>
<td>SLI6</td>
<td>0.734</td>
</tr>
<tr>
<td>CUR6</td>
<td>0.610</td>
</tr>
<tr>
<td>SLI13</td>
<td>0.654</td>
</tr>
<tr>
<td>CUR13</td>
<td>0.548</td>
</tr>
<tr>
<td>SLI25</td>
<td>0.632</td>
</tr>
<tr>
<td>CUR25</td>
<td>0.524</td>
</tr>
</tbody>
</table>

Table 3.12: Method of moments and maximum likelihood estimates of component means when the primary axis is $\bar{X}_W$, the weighted average response excluding SLI1 and CUR1 times.

<table>
<thead>
<tr>
<th>Method of Moments</th>
<th>Maximum Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>$\mu_2$</td>
</tr>
<tr>
<td>SLI1</td>
<td>0.806</td>
</tr>
<tr>
<td>CUR1</td>
<td>0.954</td>
</tr>
<tr>
<td>SLI6</td>
<td>0.734</td>
</tr>
<tr>
<td>CUR6</td>
<td>0.641</td>
</tr>
<tr>
<td>SLI13</td>
<td>0.699</td>
</tr>
<tr>
<td>CUR13</td>
<td>0.607</td>
</tr>
<tr>
<td>SLI25</td>
<td>0.593</td>
</tr>
<tr>
<td>CUR25</td>
<td>0.496</td>
</tr>
</tbody>
</table>

Table 3.13: Method of moments and maximum likelihood estimates of component means when the primary axis is CUR25.
Both tables reveal the ML mean estimate $\hat{\mu}_1$ for the non-susceptible group is decreasing over the follow-up time periods. This is to be expected as the non-susceptible group begins to adapt to parental loss. Except for the CUR6, SLI13 and CUR13 times, the elements of $\hat{\mu}_1$ are slightly smaller when CUR25 acts as primary axis.

Figure 3.11: Maximum likelihood estimates of the component means. Solid circles depict the susceptible group. Open circles depict the non-susceptible group. Solid lines represent results when the primary axis is the weighted mean. Dashed lines represent results when the primary axis is CUR25.
Another interesting behavior is that the follow-up SLI mean scores are consistently higher relative to the CUR mean scores of the same interview period. This trend appears not only in $\hat{\mu}_1$, but $\hat{\mu}_2$ as well, for both MOM and ML estimates alike. The overall pattern of $\hat{\mu}_2$ varies, however, depending on the choice of primary axis. The information elicited from $\bar{X}_W$ as primary axis shows the emotional responses of the susceptible group are noticeably higher at the one-year and two-year anniversaries of parental death compared to the first six months. The opposite is true when CUR25 is primary axis. In this case, the susceptible group appears more agitated during the first six months compared to the first year and second year time frames. In both cases, however, the mean responses obtained at both the SLI25 and CUR25 times are slightly higher than their counterparts at the one year time period.

When CUR25 is the primary axis, group differences, defined by $\hat{\mu}_2 - \hat{\mu}_1$, increase to their maximum size at the six-month interview, decrease by the first year then increase again at the second year. Group separation is visibly greater during the first and six-month periods compared to the same periods when $\bar{X}_W$ is the primary axis. In the latter case, the emotional states between the two groups steadily diverge over the two-year time span.

Fisher’s discriminant function (Fisher 1938) is utilized to compare group separation between the two sets of parameter estimates. The idea is to find a linear combination of the multivariate observations resulting in two groups of univariate observations that are segregated as much as possible. Let $\hat{\ell}$ be a vector of coefficients representing some linear combination. Each multivariate observation is projected onto a line in the direction of $\hat{\ell}$ which is then varied until maximum separation is achieved.
The following allocation rule is based on Fisher’s discriminant function. It assigns subjects from the grief study to either the emotionally susceptible group or the non-susceptible group. Let \( \hat{l} = (\hat{\mu}_1 - \hat{\mu}_2)\hat{\Sigma}^{-1} \). Assign the \( i \)th observation, \( i = 1, \ldots, N \), to the non-susceptible group if \( y_i = \hat{l}^\prime x_i \geq \hat{l}^\prime (\hat{\mu}_1 + \hat{\mu}_2)/2 = \hat{m} \); if \( y_i < \hat{m} \), then assign the \( i \)th observation to the susceptible group. For the grief data, cutoff values \( \hat{m}_j, j = 1, \ldots, 8 \) are calculated for each of the missing data patterns in Table 3.4 by omitting the corresponding missing time frames from \( \hat{\mu}_1, \hat{\mu}_2 \) and \( \hat{\Sigma}^{-1} \). As a result, intermingling of the susceptible and non-susceptible observations is possible.

Figure 3.12 displays dot plots of the distribution of \( y \)'s that result from applying the allocation rule to the SqRt-BAMO data using the parameter estimates derived from the two primary axes, \( \bar{X}_W \) and CUR25. The first plot corresponds to results obtained when \( \bar{X}_W \) is primary axis. The second plot pertains to results produced when CUR25 is primary axis. Solid circles represent the susceptible group, open circles the non-susceptible group. The primary axis \( \bar{X}_W \) does a better job of differentiating the two groups as evidenced by the denser cluster of solid circles in the first dot plot.

The proportion estimates of the two groups allocated via Fisher’s discriminate function are displayed in Table 3.14. A greater proportion of observations are allotted

<table>
<thead>
<tr>
<th>Primary Axis</th>
<th>( \hat{\pi}_1 )</th>
<th>( \hat{\pi}_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{X}_W )</td>
<td>0.851</td>
<td>0.149</td>
</tr>
<tr>
<td>CUR25</td>
<td>0.713</td>
<td>0.287</td>
</tr>
</tbody>
</table>

Table 3.14: Proportion estimates of the susceptible and non-susceptible groups obtained from using the allocation rule based on Fisher’s discriminate function.
to the susceptible group for both types of primary axes compared to the ML estimates
in Table 3.11. When CUR25 is primary axis, $\hat{\pi}_2 = 0.287$, which is almost double the
value 0.149 obtained for $\hat{\pi}_2$ when $\bar{X}_W$ is primary axis. This is similar to the trend
seen with the ML estimates.

Figure 3.12: Dotplot of Fisher’s discriminant function results. The top plot corre-
sponds to results obtained when $\bar{X}_W$ is primary axis. The bottom plot corresponds
to results obtained when CUR25 is primary axis. Solid circles represent the values of
$y = \hat{l}'x$ corresponding to the susceptible group; open circles represent the $y$ values
for the non-susceptible group.
According to the allocation rule based on Fisher’s discriminant function, large values of \( y \) are assigned to the non-susceptible group whereas smaller values are assigned to the susceptible group. Table 3.15 reveals notable differences between the two coefficient vectors regarding how the data at each time is weighted. This results in a group allocation discrepancy between the two primary axes as seen in Table 3.16. Most of the difference in classification occurs with the susceptible group. Of the 127 observations that are classified susceptible by the allocation rule for either primary axis, only 16 are labeled as such in both cases.

<table>
<thead>
<tr>
<th>( \hat{t}_{\bar{X}_W} )</th>
<th>SLI1</th>
<th>CUR1</th>
<th>SLI6</th>
<th>CUR6</th>
<th>SLI13</th>
<th>CUR13</th>
<th>SLI25</th>
<th>CUR25</th>
</tr>
</thead>
<tbody>
<tr>
<td>-7.782</td>
<td>-1.348</td>
<td>1.192</td>
<td>1.862</td>
<td>6.149</td>
<td>4.329</td>
<td>5.911</td>
<td>-1.089</td>
<td></td>
</tr>
<tr>
<td>( \hat{t}_{CUR25} )</td>
<td>1.230</td>
<td>3.556</td>
<td>2.827</td>
<td>-4.273</td>
<td>-3.382</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.15: Comparing the coefficients of the linear combination used in Fisher’s discriminate function for the primary axis choices \( \bar{X}_W \) and CUR25.

Table 3.16: Classification comparison of discriminate function allocation according to primary axis.
Initially, the grief researchers thought to predict those who will need psychological intervention based on the presence of major depression right after parental death. Subjects are classified as bereaved-depressed (BD) if they met depression criteria during the CUR1 time period, bereaved-normal (BN) if they did not. Table 3.17 compares this classification to that obtained via Fisher’s allocation rule when using $\bar{X}_W$ as primary axis. The largest discrepancy occurs in the bereaved-depressed category where 110 of these children are labeled non-susceptible. Using CUR25 as primary axis resulted in seventy-eight subjects being similarly labeled but 35 bereaved-normal subjects were assigned to the susceptible group which is 13 more than the results in Table 3.17. The implication is that severe initial emotional distress is not predictive of long-term emotional maladjustment to parental loss.

<table>
<thead>
<tr>
<th></th>
<th>Non-Susceptible</th>
<th>Susceptible</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bereaved-Normal</td>
<td>169</td>
<td>22</td>
<td>191</td>
</tr>
<tr>
<td>Bereaved-Depressed</td>
<td>110</td>
<td>26</td>
<td>137</td>
</tr>
<tr>
<td>Total</td>
<td>279</td>
<td>49</td>
<td>328</td>
</tr>
</tbody>
</table>

Table 3.17: Classification comparison of discriminate function allocation according to depression status during the CUR1 time period when $\bar{X}_W$ is primary axis.

Similar results occur when evaluating the inclusion criteria status of the subjects in relation to emotional vulnerability to parental death. Table 3.18 provides the corresponding breakdown when $\bar{X}_W$ is primary axis. Although there is a slightly greater chance of being designated emotionally susceptible if a subject is classified as complex
rather than simple, 20.8% vs 9.8%, ICS is not predictive of emotional sensitivity to loss of a parent. The corresponding probabilities when CUR25 is primary axis, 29.2% vs 28.2%, yield the same conclusion.

<table>
<thead>
<tr>
<th></th>
<th>Non-Susceptible</th>
<th>Susceptible</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>157</td>
<td>17</td>
<td>174</td>
</tr>
<tr>
<td>Complex</td>
<td>122</td>
<td>32</td>
<td>154</td>
</tr>
<tr>
<td>Total</td>
<td>279</td>
<td>49</td>
<td>328</td>
</tr>
</tbody>
</table>

Table 3.18: Classification comparison of discriminate function allocation according to inclusion criteria status when $\bar{X}_W$ is primary axis.

The marginal densities are calculated using the ML parameter estimates reported in Tables 3.11, 3.12, 3.13, 3.20 and 3.21 for the two primary axes. Figure 3.13 shows histograms of the marginal densities overlayed by the density estimates. The solid line corresponds to the density estimated when $\bar{X}_W$ is the primary axis. The dashed line represents the estimated density associated with the CUR25 primary axis.

The parameter estimates associated with the $\bar{X}_W$ primary axis result in a superior fit compared to those obtained from using CUR25 as the primary axis. The inadequacy can be attributed to the instability of the covariance matrix presented in Table 3.21. Although the determinate of both covariance matrices is near zero, the determinate of the matrix in Table 3.21 is half the determinate of the covariance matrix in Table 3.20. This reduction is enough to yield a very uneven likelihood resulting in parameter estimates that are less than ideal. Note that elements of the covari-
Figure 3.13: Histograms of the marginal distributions of the SqRt-BAMO data with density estimate overlay. The solid line depicts the density estimate that results when $\bar{X}_W$ is the primary axis. The dashed line represents the density estimate when the primary axis is CUR25.

The elements of the estimated covariance matrix in Table 3.21 are consistently lower than the elements of the observed covariance matrix in Table 3.19. This result, coupled with the smaller estimate of $\pi$ produces a higher density peak than the density estimate associated with $\bar{X}_W$.

<table>
<thead>
<tr>
<th></th>
<th>SLI1</th>
<th>CUR1</th>
<th>SLI6</th>
<th>CUR6</th>
<th>SLI13</th>
<th>CUR13</th>
<th>SLI25</th>
<th>CUR25</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI1</td>
<td>0.172</td>
<td>0.151</td>
<td>0.109</td>
<td>0.106</td>
<td>0.091</td>
<td>0.090</td>
<td>0.081</td>
<td>0.083</td>
</tr>
<tr>
<td>CUR1</td>
<td>0.151</td>
<td>0.167</td>
<td>0.120</td>
<td>0.116</td>
<td>0.097</td>
<td>0.094</td>
<td>0.087</td>
<td>0.085</td>
</tr>
<tr>
<td>SLI6</td>
<td>0.109</td>
<td>0.120</td>
<td>0.220</td>
<td>0.212</td>
<td>0.158</td>
<td>0.157</td>
<td>0.129</td>
<td>0.118</td>
</tr>
<tr>
<td>CUR6</td>
<td>0.106</td>
<td>0.116</td>
<td>0.212</td>
<td>0.225</td>
<td>0.150</td>
<td>0.150</td>
<td>0.133</td>
<td>0.122</td>
</tr>
<tr>
<td>SLI13</td>
<td>0.091</td>
<td>0.097</td>
<td>0.158</td>
<td>0.150</td>
<td>0.215</td>
<td>0.199</td>
<td>0.122</td>
<td>0.104</td>
</tr>
<tr>
<td>CUR13</td>
<td>0.090</td>
<td>0.094</td>
<td>0.157</td>
<td>0.150</td>
<td>0.199</td>
<td>0.205</td>
<td>0.118</td>
<td>0.107</td>
</tr>
<tr>
<td>SLI25</td>
<td>0.081</td>
<td>0.087</td>
<td>0.129</td>
<td>0.133</td>
<td>0.122</td>
<td>0.118</td>
<td>0.205</td>
<td>0.179</td>
</tr>
<tr>
<td>CUR25</td>
<td>0.083</td>
<td>0.085</td>
<td>0.118</td>
<td>0.122</td>
<td>0.104</td>
<td>0.107</td>
<td>0.179</td>
<td>0.184</td>
</tr>
</tbody>
</table>

Table 3.19: Observed covariance matrix for the SqRt-BAMO data.
Although the specific trends manifested by $\hat{\mu}_1$ and $\hat{\mu}_2$ for $\bar{X}_W$ and CUR25 are different, both primary axes appear to support the hypothesis of the existence of a subgroup of emotionally susceptible children. The conclusion that CUR25 as primary axis yields more desirable parameter estimates because of a higher likelihood value is suspect. Not only does CUR25 have more than double the percentage of missing

<table>
<thead>
<tr>
<th></th>
<th>SLI1</th>
<th>CUR1</th>
<th>SLI6</th>
<th>CUR6</th>
<th>SLI13</th>
<th>CUR13</th>
<th>SLI25</th>
<th>CUR25</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI1</td>
<td>0.161</td>
<td>0.141</td>
<td>0.089</td>
<td>0.086</td>
<td>0.046</td>
<td>0.040</td>
<td>0.060</td>
<td>0.063</td>
</tr>
<tr>
<td>CUR1</td>
<td>0.141</td>
<td>0.157</td>
<td>0.101</td>
<td>0.096</td>
<td>0.056</td>
<td>0.048</td>
<td>0.067</td>
<td>0.066</td>
</tr>
<tr>
<td>SLI6</td>
<td>0.089</td>
<td>0.101</td>
<td>0.202</td>
<td>0.198</td>
<td>0.105</td>
<td>0.098</td>
<td>0.107</td>
<td>0.099</td>
</tr>
<tr>
<td>CUR6</td>
<td>0.086</td>
<td>0.096</td>
<td>0.198</td>
<td>0.215</td>
<td>0.103</td>
<td>0.098</td>
<td>0.107</td>
<td>0.098</td>
</tr>
<tr>
<td>SLI13</td>
<td>0.046</td>
<td>0.056</td>
<td>0.105</td>
<td>0.103</td>
<td>0.137</td>
<td>0.115</td>
<td>0.069</td>
<td>0.052</td>
</tr>
<tr>
<td>CUR13</td>
<td>0.040</td>
<td>0.048</td>
<td>0.098</td>
<td>0.098</td>
<td>0.115</td>
<td>0.114</td>
<td>0.062</td>
<td>0.049</td>
</tr>
<tr>
<td>SLI25</td>
<td>0.060</td>
<td>0.067</td>
<td>0.107</td>
<td>0.107</td>
<td>0.069</td>
<td>0.062</td>
<td>0.141</td>
<td>0.117</td>
</tr>
<tr>
<td>CUR25</td>
<td>0.063</td>
<td>0.066</td>
<td>0.099</td>
<td>0.098</td>
<td>0.052</td>
<td>0.049</td>
<td>0.117</td>
<td>0.124</td>
</tr>
</tbody>
</table>

Table 3.20: Maximum likelihood estimate of the covariance matrix when $\bar{X}_W$ is the primary axis.

<table>
<thead>
<tr>
<th></th>
<th>SLI1</th>
<th>CUR1</th>
<th>SLI6</th>
<th>CUR6</th>
<th>SLI13</th>
<th>CUR13</th>
<th>SLI25</th>
<th>CUR25</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI1</td>
<td>0.116</td>
<td>0.100</td>
<td>0.038</td>
<td>0.035</td>
<td>0.050</td>
<td>0.048</td>
<td>0.020</td>
<td>0.024</td>
</tr>
<tr>
<td>CUR1</td>
<td>0.100</td>
<td>0.119</td>
<td>0.052</td>
<td>0.048</td>
<td>0.058</td>
<td>0.054</td>
<td>0.028</td>
<td>0.028</td>
</tr>
<tr>
<td>SLI6</td>
<td>0.038</td>
<td>0.052</td>
<td>0.125</td>
<td>0.116</td>
<td>0.104</td>
<td>0.100</td>
<td>0.041</td>
<td>0.034</td>
</tr>
<tr>
<td>CUR6</td>
<td>0.035</td>
<td>0.048</td>
<td>0.116</td>
<td>0.129</td>
<td>0.095</td>
<td>0.093</td>
<td>0.043</td>
<td>0.035</td>
</tr>
<tr>
<td>SLI13</td>
<td>0.050</td>
<td>0.058</td>
<td>0.104</td>
<td>0.095</td>
<td>0.185</td>
<td>0.169</td>
<td>0.083</td>
<td>0.066</td>
</tr>
<tr>
<td>CUR13</td>
<td>0.048</td>
<td>0.054</td>
<td>0.100</td>
<td>0.093</td>
<td>0.169</td>
<td>0.173</td>
<td>0.079</td>
<td>0.069</td>
</tr>
<tr>
<td>SLI25</td>
<td>0.020</td>
<td>0.028</td>
<td>0.041</td>
<td>0.043</td>
<td>0.083</td>
<td>0.079</td>
<td>0.118</td>
<td>0.095</td>
</tr>
<tr>
<td>CUR25</td>
<td>0.024</td>
<td>0.028</td>
<td>0.034</td>
<td>0.035</td>
<td>0.066</td>
<td>0.069</td>
<td>0.095</td>
<td>0.103</td>
</tr>
</tbody>
</table>

Table 3.21: Maximum likelihood estimate of the covariance matrix when CUR25 is the primary axis.
values than $X_W$, but the discriminate analysis results and the density estimates in Figure 3.13 demonstrate that the primary axis $X_W$ yields more realistic parameter estimates. The fact that the log-likelihood value associated with $X_W$ is smaller than the log-likelihood value corresponding to the less suitable CUR25 primary axis directs attention to the instability of the log-likelihood function itself. An attempt is made to address this problem in the next chapter by specifically modeling the association between responses of different time periods to provide structure to the covariance matrix.
CHAPTER 4

DEVELOPING AND FITTING MODELS

Models are developed to assess whether or not there exists a subgroup of grieving children that exhibit prolonged emotional upheaval in response to the death of a parent. Specific characteristics of the SqRt-BAMO data will be incorporated into the models. The results of model fitting are reported.

4.1 Model Development

Figure 4.1(a) plots the variance of the SqRt-BAMO response for each time. The variance is fairly constant over time. The initial responses display slightly less variability than the follow-up responses but the difference is small. A simple variance model for the data is

\[ \sigma_t^2 = \sigma^2 \] at every time \( t \). \hfill (4.1)

Figure 4.1(b) is a plot of the correlation between SqRt-BAMO responses as a function of difference in time. The open square corresponds to the diagonal elements of the correlation matrix. The open circles represent correlations of SLI1 and CUR1 times with all follow-up times. The open triangles depict the correlations between responses of the same interview period such as SLI1 and CUR1 or SLI6 and CUR6.
Figure 4.1: Graphical displays of the observed variance and correlation structures of the SqRt-BAMO data. (a) The sample variances for each time frame of the SqRt-BAMO data. (b) The correlation between responses as a function of difference in time. The open square corresponds to the diagonal elements of the correlation matrix. The open circles represent correlations of SLI1 and CUR1 times with all follow-up time frames. The open triangles depict the correlations between responses of the same interview period. The stars represent the remaining correlations.
The remaining correlations are designated by stars. The last three groups will be referred to as Low, High and Mid, respectively.

Consider the model in Equation 4.2 for the correlation structure. It is a function of the magnitude of the difference in time between responses, denoted by \( r \). The linear pattern of the distinct groupings of observed correlations in Figure 4.1b is modeled as three parallel lines with group dependent intercepts.

\[
\rho(r) = \exp(-\gamma_1 r - \gamma_2 \text{Low} - \gamma_3 \text{Mid} - \gamma_4 \text{High})
\]  

Figure 4.2 plots the mean SqRt-BAMO response as a function of time. The open and solid circles are indicative of the SLI and CUR time periods, respectively. The mean responses gradually decrease over time in an exponential manner.

Figure 4.2: Observed means of the SqRt-BAMO data. Open circles indicate SLI times; solid circles represent CUR times.
Equation 4.3 states the model that will be used for the mean. It is a function of time, denoted by $t$, and includes additional predictor variables $\mathbf{x}$. Capitalized predictor names denote indicator variables. A distinction between CUR and SLI time periods is made to address the different response rates corresponding to the two sets of time periods. The solid circles in Figure 4.2 reveal that the CUR responses decrease more sharply due to the elevated emotional responses at the time of parental death. The extra steepness in the slope for the CUR time frame is measured by $\beta_8$. Including $\beta_1$ in the model allows for a potential disparity in number of symptoms endorsed during the SLI and CUR time periods. With the exception of the initial interview period, the time spanned by the SLI periods are much longer than the two-week CUR periods, allowing an opportunity for the number of symptoms reported between two types of intervals to differ.

Recall from Chapter 3 that a subject’s age, socioeconomic status (SES) and inclusion criteria status (ICS) will be used as predictor variables with the intention that the underlying missing data mechanism will be made closer to missing at random. The parameter $\beta_2$ is interpreted as the added effect of a subject having a “complex” inclusion criteria status. A subject’s socioeconomic status is one of five categories ranging from poor to upper class. Figure 4.3a shows the disparity between those subjects classified as poor and those who were not. The parameter $\beta_3$ represents the additional effect for those subjects with poor economic backgrounds.

Female children reported elevated emotional responses compared to the male children as evidenced by the plot of the mean responses for each group in Figure 4.3b.
Gender is included in the mean model as a predictor and the increased response exhibited by females is measured by $\beta_4$.

Figure 4.3: Plots of the observed means of the SqRt-BAMO data over time according to (a) SES levels and (b) gender.
Over time, many of the grieving children likely adapt to the loss of their parent and their emotional responses eventually stabilize. The parameter $\beta_0$ represents this state of “emotional equilibrium”. The difference between this level and the emotional state at the time of parental death is measured by $\beta_6$.

Under the null hypothesis $H_0$ the children grieving for a deceased parent are considered to be an emotionally homogenous group. The data is assumed to follow a multivariate normal distribution having mean $\mu(t \mid \beta, x)$ defined in Equation 4.3 and covariance matrix $\Sigma = \sigma^2 P$ where $P$ is the correlation matrix composed of elements derived from Equation 4.2. The alternative hypothesis $H_1$ states that an emotionally susceptible subgroup of grieving children exists. Equation 4.4 displays a mixture of two multivariate normal distributions which is used to model this assumption.

$$
p(y \mid x, \beta, \beta^*, \gamma, \sigma, \pi) = \pi N[\mu_1(t \mid \beta, x), \Sigma(\sigma, \gamma)] + (1 - \pi) N[\mu_2(t \mid \beta, \beta^*, x), \Sigma(\sigma, \gamma)] \hspace{1cm} (4.4)
$$

Each multivariate normal component distribution has the same covariance structure $\Sigma = \sigma^2 P$. The first component distribution represents the non-susceptible group and the corresponding mean $\mu_1(t \mid \beta, x)$ is given by Equation 4.3.

The second component distribution is associated with the susceptible group and its mean $\mu_2$ is expected to decrease at a slower rate than $\mu_1$ and allow for greater emotional stress at the time of parental death. In addition, the level at which “emotional equilibrium” is achieved may be higher than the non-susceptible group. Accounting for these potential group differences translates into the following model for $\mu_2$:

$$
\mu_2(t \mid \beta, \beta^*, x) = \beta_0^* + \beta_1 \text{CUR} + \beta_2 \text{ICS} + \beta_3 \text{SES} + \beta_4 \text{GENDER} + \\
\beta_5 \text{Age} + \beta_6^* \exp(-\beta_7^* t - \beta_8^* t \times \text{CUR}).
$$

\hspace{1cm} (4.5)
Note that the coefficients $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are the same in $\mu_1$ and $\mu_2$. The parameters $\beta_0, \beta_6, \beta_7, \beta_8$ and $\beta^*_0, \beta^*_6, \beta^*_7, \beta^*_8$ are allowed to differ.

Significant differences between $\beta^*_0, \beta^*_6, \beta^*_7, \beta^*_8$ in $\mu_2$ and their counterparts in $\mu_1$ indicate the existence of a group of susceptible children. To formalize an appropriate statistical hypothesis, define $\Delta = (\beta^*_0 - \beta_0, \beta^*_6 - \beta_6, \beta^*_7 - \beta_7, \beta^*_8 - \beta_8)'$. Whether or not a susceptible group of children exists corresponds to testing the pair of hypotheses,

$$H_0 : \Delta = 0 \quad \text{vs} \quad H_1 : \Delta \neq 0.$$  \hspace{1cm} (4.6)

### 4.2 Results of Model Fitting

Since explicit solutions are not available, iterative techniques involving a quasi-Newton optimization routine are used to obtain parameter estimates by minimizing

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% C. I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUR</td>
<td>$\beta_0$</td>
<td>0.3772</td>
<td>0.0808</td>
<td>(0.219, 0.536)</td>
</tr>
<tr>
<td>ICS</td>
<td>$\beta_1$</td>
<td>0.1998</td>
<td>0.0133</td>
<td>(0.174, 0.226)</td>
</tr>
<tr>
<td>SES</td>
<td>$\beta_2$</td>
<td>0.1227</td>
<td>0.0397</td>
<td>(0.045, 0.201)</td>
</tr>
<tr>
<td>GENDER</td>
<td>$\beta_3$</td>
<td>0.3337</td>
<td>0.0816</td>
<td>(0.174, 0.494)</td>
</tr>
<tr>
<td>Age</td>
<td>$\beta_4$</td>
<td>0.0417</td>
<td>0.0395</td>
<td>(-0.036, 0.119)</td>
</tr>
<tr>
<td>Time</td>
<td>$\beta_5$</td>
<td>0.0051</td>
<td>0.0063</td>
<td>(-0.007, 0.017)</td>
</tr>
<tr>
<td>Time×CUR</td>
<td>$\beta_6$</td>
<td>0.3544</td>
<td>0.0223</td>
<td>(0.309, 0.400)</td>
</tr>
<tr>
<td>Time Difference</td>
<td>$\sigma$</td>
<td>0.4289</td>
<td>0.0119</td>
<td>(0.406, 0.452)</td>
</tr>
<tr>
<td>Low</td>
<td>$\gamma_1$</td>
<td>4.8959e-8</td>
<td>1.0938e-5</td>
<td>(-2.1e-5, 2.2e-5)</td>
</tr>
<tr>
<td>Mid</td>
<td>$\gamma_2$</td>
<td>0.6990</td>
<td>0.0728</td>
<td>(0.556, 0.841)</td>
</tr>
<tr>
<td>High</td>
<td>$\gamma_3$</td>
<td>0.4969</td>
<td>0.0466</td>
<td>(0.406, 0.588)</td>
</tr>
<tr>
<td></td>
<td>$\gamma_4$</td>
<td>0.0798</td>
<td>0.0058</td>
<td>(0.068, 0.091)</td>
</tr>
</tbody>
</table>

Table 4.1: Parameter estimates under the $H_0$ model and corresponding standard errors and 95% confidence intervals.
the negative of the log-likelihood. The optimization routines were developed using
AD Model Builder and AUTODIF software, both of which are based on the C++
programming language (Otter Research Ltd. 2000). Tables 4.1 and 4.2 summarize
the parameter estimates, standard errors and corresponding 95% confidence intervals
under $H_0$ and $H_1$, respectively. Substituting the estimates from Table 4.1 into the log-
likelihood function yields the value $\ln L_0 = -10.1162$ under $H_0$. A similar substitution
using the estimates from Table 4.2 produces the log-likelihood value $\ln L_1 = -0.6361$
under $H_1$. The resulting likelihood ratio statistic is $-2 \ln \Lambda = -2 (\ln L_0 - \ln L_1) = 18.9602$.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUR</td>
<td>$\beta_0$</td>
<td>0.3643</td>
<td>0.0749</td>
<td>(0.216, 0.511)</td>
</tr>
<tr>
<td>ICS</td>
<td>$\beta_1$</td>
<td>0.2879</td>
<td>0.0525</td>
<td>(0.185, 0.391)</td>
</tr>
<tr>
<td>SES</td>
<td>$\beta_2$</td>
<td>0.1135</td>
<td>0.0394</td>
<td>(0.036, 0.191)</td>
</tr>
<tr>
<td>GENDER</td>
<td>$\beta_4$</td>
<td>0.0345</td>
<td>0.0393</td>
<td>(-0.043, 0.112)</td>
</tr>
<tr>
<td>Age</td>
<td>$\beta_5$</td>
<td>0.0003</td>
<td>0.0060</td>
<td>(-0.012, 0.012)</td>
</tr>
<tr>
<td>Time</td>
<td>$\beta_7$</td>
<td>0.0021</td>
<td>0.0010</td>
<td>(1.0e-4, 0.004)</td>
</tr>
<tr>
<td>Time×CUR</td>
<td>$\beta_8$</td>
<td>0.2749</td>
<td>0.1068</td>
<td>(0.066, 0.484)</td>
</tr>
<tr>
<td>Same as above</td>
<td>$\beta^*_0$</td>
<td>0.3700</td>
<td>0.0004</td>
<td>(0.369, 0.371)</td>
</tr>
<tr>
<td>but represents</td>
<td>$\beta^*_6$</td>
<td>0.5169</td>
<td>0.0400</td>
<td>(0.439, 0.595)</td>
</tr>
<tr>
<td>the susceptible</td>
<td>$\beta^*_7$</td>
<td>4.7535e-8</td>
<td>1.0649e-5</td>
<td>(-2.1e-5, 2.1e-5)</td>
</tr>
<tr>
<td>group</td>
<td>$\beta^*_8$</td>
<td>0.0407</td>
<td>0.0124</td>
<td>(0.016, 0.065)</td>
</tr>
<tr>
<td>Time Difference</td>
<td>$\gamma_1$</td>
<td>4.3450e-9</td>
<td>9.9112e-7</td>
<td>(-1.9e-6, 1.9e-6)</td>
</tr>
<tr>
<td>Low</td>
<td>$\gamma_2$</td>
<td>0.7266</td>
<td>0.0810</td>
<td>(0.568, 0.885)</td>
</tr>
<tr>
<td>Med</td>
<td>$\gamma_3$</td>
<td>0.5035</td>
<td>0.0507</td>
<td>(0.406, 0.605)</td>
</tr>
<tr>
<td>High</td>
<td>$\gamma_4$</td>
<td>0.0726</td>
<td>0.0064</td>
<td>(0.060, 0.085)</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0.7028</td>
<td>0.0461</td>
<td>0.612, 0.793</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: Parameter estimates under the $H_1$ model and corresponding standard er-
rors and 95% confidence intervals. Parameter estimates are assumed to be asymptot-
ically normally distributed.
Although mixture models arise naturally in many settings, actual implementation has been limited by the difficulty of parameter estimation as well as the lack of a general method for testing hypotheses. The methodology discussed in Chapter 4 attempts to address these issues.

Instead of estimating group means as in the method of moments approach used in Chapter 3, advantage is taken of patterns in the data to develop models. These data patterns are utilized in conjunction with characteristics of the particular research problem to formulate models tailored specifically for the statistical hypotheses of interest. The manner in which the mean models are designed allows the null hypothesis $H_0$ to be clearly and concisely written as a subset of the alternative hypothesis $H_1$.

Another benefit of specifically modeling the mean and covariance structure is a reduction in the number of parameters. Using an unconstrained two-component multivariate normal mixture distribution to model the SqRt-BAMO data would have involved 16 mean and 36 covariance matrix parameters as well as $\pi$ for a total of 53 parameters. The models in Equations 4.1, 4.2, 4.3, 4.4 and 4.5 reduce the total number of parameters under $H_1$ to 19. A similar reduction, from 44 to 14 parameters,
occurs under the $H_0$ model. Curtailing the number of parameters in such a fashion streamlines the fitting process considerably.

A typical problem encountered in parameter estimation involving a likelihood function $L$ composed of the product of mixture distributions is that the function contains many local maxima resulting in a multimodal likelihood. In fact $L$ is often unbounded. For example, suppose $x_1, x_2, \ldots, x_n$ is a sample from a two-component univariate normal mixture. Each sample point introduces a singularity into the likelihood function since $L(\mu_1, \mu_2, \sigma_1, \sigma_2, \pi) = \infty$ whenever $(\mu_1, \sigma_1) = (x_i, 0)$ or $(\mu_2, \sigma_2) = (x_i, 0)$ for $i = 1, \ldots, n$. A similar event occurs in the multivariate normal case when the determinate of the covariance matrix $|\Sigma|$ approaches zero. Such solutions do not produce meaningful mixtures and it is desirable to avoid them. This issue can be addressed by assigning structure to the covariance matrix as well as assuming the covariance matrices are the same for each component distribution. Both methods are utilized in the current problem in an attempt to smooth the likelihood function.

Parameter estimates and standard errors associated with the models under $H_0$ and $H_1$ are reported in Tables 4.1 and 4.2, respectively. These solutions yield a likelihood ratio test (LRT) statistic of $-2 \ln \Lambda = 18.9602$. Such a large value strongly favors $H_1$. Ninety-five percent confidence intervals indicate age and gender under $H_1$ are not important predictors since these confidence intervals contain zero. The predictors CUR, ICS and SES appear to be useful predictors at the 95% confidence level.

Figure 4.1 illustrates the BAMO responses are positively correlated over time. This means that subjects with high scores will be more likely to have high scores at nearby points of time. A similar result will occur for subjects with low scores. What appears to be two discernible groups of responses is actually just a by-product of the
correlation structure. Over time this artificially induced separation will gradually dis-
appear as correlations between responses from distant points in time become smaller.
An indication that two latent groups exist arises when these correlations do not tend
to zero but remain steady as the difference in time between responses increases.

The value of $\gamma_1$ in the correlation model defined in Equation 4.2 provides a way of
assessing if group separation is a temporary phenomenon or not. A large $\gamma_1$ results in
a rapidly decreasing correlation between responses of distant time periods indicating
a perceived disparity in responses is just a consequence of the correlation structure.
A small $\gamma_1$ suggests that any group differences detected are genuine. The latter
scenario seems to be supported by the data since the value of $\hat{\gamma}_1$ under $H_0$ and $H_1$
is essentially zero providing further evidence that latent groups consisting of susceptible
and non-susceptible grieving children exist.

A small number of subjects partaking in the grief study exhibit very few if any
symptoms of the mental disorders measured by BAMO, resulting in extremely low
scores for this group. Under $H_1$, the optimization routine attempts to closely fit this
emotionally healthy group of children. Results showed $\hat{\beta}_0^*$ near zero with $\hat{\beta}_0^* < \hat{\beta}_0$ and
$\hat{\beta}_6^* < \hat{\beta}_6$. The intention, however, is for $\beta_0^*$, $\beta_6^*$, $\beta_7^*$ and $\beta_8^*$ to reflect an emotionally
vulnerable group of grieving children.

Recall that $\beta_0^*$ is the “equilibrium” state that the susceptible group will eventually
attain, $\beta_6^*$ is the jump in emotional response from this state of equilibrium at the
time of parental death, $\beta_7^*$ is the rate of change in emotional responses over time
and $\beta_8^*$ distinguishes the rate of change between SLI and CUR time frames. In the
scenario where susceptible and non-susceptible groups exist, one expects the state
of equilibrium for the susceptible group to be no less than that of the level of the
non-susceptible group. The rate of change in emotional responses for the susceptible
group will likely be slower too. In terms of elements from the mean models, this
translates into \( \beta_0^* \geq \beta_0, \beta_6^* \geq \beta_6, \beta_7^* \leq \beta_7 \) and \( \beta_8^* \leq \beta_8 \).

To avoid fitting the well-adapted group, a lower bound for \( \beta_0^* \) is chosen. The
scores of the community control comparison group mentioned in Chapter 1 are used
to obtain a meaningful value for the lower bound. The community control group
represents children who have not experienced the death of a parent and their responses
provide a baseline comparison. Average scores over time for this group did not fall
below 0.39. The value 0.37 is used as a conservative estimate of the equilibrium state
the susceptible group could eventually attain. Coincidentally, \( \hat{\beta}_0 \) under \( H_0 \) is 0.3772.
The fitted model changed little when the lower bound was varied between 0.37 and
0.39.

The solution presented in Table 4.2 is intriguing. Approximately thirty percent of
the children are estimated to belong to the susceptible group \( (\hat{\pi} = 0.7028) \). The emo-
tional equilibrium level of the non-susceptible group is \( \hat{\beta}_0^* = 0.3643 \) which is slightly
less than the susceptible group level \( \hat{\beta}_0 = 0.3700 \). The difference is small and does
not appear to be statistically significant since 0.37 falls within the 95% confidence
interval for \( \hat{\beta}_0 \). Although the estimate \( \hat{\beta}_6^* = 0.5169 \) exceeds \( \hat{\beta}_6 = 0.3972 \) their corre-
sponding 95% confidence intervals overlap somewhat. The parameter estimate \( \hat{\beta}_7^* \) is
approximately zero whereas \( \hat{\beta}_7 = 0.0021 \). The difference between \( \hat{\beta}_8^* = 0.0407 \) and
\( \hat{\beta}_8 = 0.2749 \) is large. Although close, the 95% confidence intervals for the latter two
pairs of parameters do not overlap implying statistically significant differences in each
case.
The outcomes signify the susceptible group may be more traumatized initially by the death of a parent ($\hat{\beta}_6^* \geq \hat{\beta}_0$). In the long term though, the susceptible and non-susceptible groups become indistinguishable ($\hat{\beta}_6^* \approx \hat{\beta}_0$). It will take the susceptible group longer to adapt to the parental death ($\hat{\beta}_7^* < \hat{\beta}_7$). An interesting feature that arises is the difference between SLI and CUR responses appears more dramatic for the non-susceptible group ($\hat{\beta}_8^* < \hat{\beta}_8$). One way of interpreting this phenomenon is that the outlook of susceptible children is steadily pessimistic whereas the current outlook of the non-susceptible children tends to be more optimistic.

An underlying assumption up to this point is the parameter estimates of the mean parameters under $H_1$ have asymptotic normal distributions. The distribution of the LRT is assumed to be chi-squared with degrees of freedom equal to the number of constraints imposed on $H_1$ to produce $H_0$, which is 5 based on the hypotheses appearing in Equation 4.6. Unfortunately, methodology for estimating mixing distributions is more evolved than the techniques of formal inference and the related asymptotic theory.

A general format for quantitative inference in problems involving mixture distributions is lacking. One of the difficulties of testing hypotheses via the likelihood ratio test is the ambiguity of defining $H_0$. Suppose the alternative hypothesis of interest involves a two component univariate mixture distribution with density $\pi f(x; \theta_1) + (1 - \pi) f(x; \theta_2)$. The null hypothesis that the true distribution has only one component can be stated as $\pi = 1$, $\pi = 0$ or $\theta_1 = \theta_2$. As a result, a specific distribution under $H_0$ can be approximated by several points in the parameter space associated with $H_1$. 

77
Results obtained in special cases suggest the limiting distribution of likelihood ratio tests entails mixtures of chi-squared distributions with differing degrees of freedom, called “chi-bar-squared” distributions. For example, in the case of a two component mixture with \( \pi \) the only unknown, Lindsay (1995) shows that the limiting distribution of the LRT is \( 0.5 \chi^2_0 + 0.5 \chi^2_1 \) where \( \chi^2_k \) is a chi-squared distribution with \( k \) degrees of freedom. The symbol \( \chi^2_0 \) refers to a degenerate distribution with all of its mass concentrated at 0. Testing involving mixtures of normals when the mixing proportions are known also resulted in chi-bar-squared limiting distributions (Goffinet, Loisel and Laurent 1992).

In situations where the mixing distribution is finite and has a known number of support points, Redner and Walker (1984) were able to show that standard maximum likelihood theory applies to parameter estimation. In other words, under moderate assumptions, the solution to the likelihood equations, defined to be the derivative of \( \ln L \) with respect to the unknown mixing parameters set to zero, is unique and strongly consistent. In addition, the solution maximizes \( L \) at least locally and is asymptotically normally distributed.

Including auxiliary parameters, such as \( \beta \) and \( \beta^* \) in Equations 4.3 and 4.5, into the mixture framework introduces another level of complexity. In these situations the primary focus is typically estimation and inference of the structural parameters and the mixing distribution \( Q \) is viewed as incidental (Lindsay and Lesperance 1995). In the current problem, however, the structural parameters and the mixing distribution are intrinsically related and both play a pivotal role in the hypothesis testing.

The limiting distribution of the likelihood ratio test statistic corresponding to the hypotheses given in Equation 4.6 is unknown. The alternative hypothesis to reduces
to the null hypothesis by assuming the following: \( \Delta = (\beta_0^* - \beta_0, \beta_6^* - \beta_6, \beta_7^* - \beta_7, \beta_8^* - \beta_8)’ = 0 \) and \( \pi = 0 \). Extrapolating from the results by Lindsay (1995) and Goffinet et al. (1992) suggests a mixture of \( \chi^2_4 \) and \( \chi^2_5 \) may be an appropriate limiting distribution.

The results of model fitting yield \(-2 \ln \Lambda = 18.96\) as the value of the LRT statistic. Although such a large value favors the alternative hypothesis, results are inconclusive until the issue of the limiting distribution is resolved. The sampling distributions of the LRT statistic and the mean parameters need to be investigated within the context of the current problem before claiming that the data supports the existence of a susceptible group of bereaved children.

The Grief Research Project recently finished collecting five-year data. Incorporating these responses into the current data set should aid in the mixture analysis. The susceptible group may be more distinguishable from the non-susceptible group by this time. If this is the case, it may not be necessary to bound \( \beta_0^* \) from below, a shortcoming of the current procedure. The possible existence of an emotionally well-adapted group suggests that a three-component mixture distribution may be a more suitable model of the population of grieving children. Such a representation may also eliminate the need to bound \( \beta_0^* \). Once model estimates are obtained with the inclusion of the five-year data, group allocation can be explored and characteristics of the susceptible group examined.
Number of subjects: 328

<table>
<thead>
<tr>
<th>Gender</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>165</td>
<td>163</td>
</tr>
<tr>
<td>%</td>
<td>50.30</td>
<td>49.70</td>
</tr>
</tbody>
</table>

Table A.1: Gender of subjects.

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Caucasian</th>
<th>African American</th>
<th>Asian</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>320</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>%</td>
<td>97.56</td>
<td>1.22</td>
<td>0.61</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Table A.2: Ethnic background of subjects.
<table>
<thead>
<tr>
<th>ICS</th>
<th>Simple</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>174</td>
<td>154</td>
</tr>
<tr>
<td>%</td>
<td>53.05</td>
<td>46.95</td>
</tr>
</tbody>
</table>

Table A.3: Inclusion Criteria Status (ICS) by subject.

<table>
<thead>
<tr>
<th>Hollingshead Score</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>43</td>
<td>77</td>
<td>89</td>
<td>97</td>
<td>22</td>
</tr>
<tr>
<td>%</td>
<td>13.11</td>
<td>23.48</td>
<td>27.13</td>
<td>29.57</td>
<td>6.71</td>
</tr>
</tbody>
</table>

Table A.4: Hollingshead score assessing socioeconomic status by subject: 1=upper class, 2=upper middle class, 3=middle class, 4=lower middle class and 5=lower class.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Child</th>
<th>Adolescent</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>183</td>
<td>145</td>
</tr>
<tr>
<td>%</td>
<td>55.79</td>
<td>44.21</td>
</tr>
</tbody>
</table>

Table A.5: Subjects classified by age group. A subject was considered an adolescent if he or she was at least 13 years old at time of parental death, a child otherwise.

<table>
<thead>
<tr>
<th>Age</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>2</td>
<td>12</td>
<td>18</td>
<td>30</td>
<td>28</td>
<td>27</td>
<td>38</td>
</tr>
<tr>
<td>%</td>
<td>0.61</td>
<td>3.66</td>
<td>5.49</td>
<td>9.15</td>
<td>8.54</td>
<td>8.23</td>
<td>11.59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>28</td>
<td>35</td>
<td>40</td>
<td>21</td>
<td>24</td>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>%</td>
<td>8.54</td>
<td>10.67</td>
<td>12.2</td>
<td>6.4</td>
<td>7.32</td>
<td>5.79</td>
<td>1.83</td>
</tr>
</tbody>
</table>

Table A.6: Distribution of ages of subjects at first interview.
Number of families represented: 187

<table>
<thead>
<tr>
<th>Number of Siblings</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>81</td>
<td>79</td>
<td>20</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>( % )</td>
<td>43.32</td>
<td>42.25</td>
<td>10.70</td>
<td>3.21</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Table A.7: Number of families that participated in the study according to number of siblings whose data was analyzed.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>137</td>
<td>50</td>
</tr>
<tr>
<td>( % )</td>
<td>73.26</td>
<td>26.74</td>
</tr>
</tbody>
</table>

Table A.8: Gender of surviving parent.

<table>
<thead>
<tr>
<th>ICS</th>
<th>Simple</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>98</td>
<td>89</td>
</tr>
<tr>
<td>( % )</td>
<td>52.41</td>
<td>47.59</td>
</tr>
</tbody>
</table>

Table A.9: Inclusion Criteria Status (ICS) by family.

<table>
<thead>
<tr>
<th>Hollingshead Score</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>22</td>
<td>46</td>
<td>54</td>
<td>52</td>
<td>13</td>
</tr>
<tr>
<td>( % )</td>
<td>11.76</td>
<td>24.6</td>
<td>28.88</td>
<td>27.81</td>
<td>6.95</td>
</tr>
</tbody>
</table>

Table A.10: Hollingshead score assessing socioeconomic status by family: 1=upper class, 2=upper middle class, 3=middle class, 4=lower middle class and 5=lower class.
<table>
<thead>
<tr>
<th>Age of Surviving Parent</th>
<th>23</th>
<th>27</th>
<th>28</th>
<th>30</th>
<th>31</th>
<th>32</th>
<th>33</th>
<th>34</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>(%)</td>
<td>0.53</td>
<td>1.07</td>
<td>0.53</td>
<td>2.14</td>
<td>3.74</td>
<td>1.07</td>
<td>2.14</td>
<td>3.21</td>
</tr>
<tr>
<td>Age of Surviving Parent</td>
<td>35</td>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td>40</td>
<td>41</td>
<td>42</td>
</tr>
<tr>
<td>(n)</td>
<td>10</td>
<td>8</td>
<td>14</td>
<td>13</td>
<td>8</td>
<td>13</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>(%)</td>
<td>5.35</td>
<td>4.28</td>
<td>7.49</td>
<td>6.95</td>
<td>4.28</td>
<td>6.95</td>
<td>4.81</td>
<td>6.95</td>
</tr>
<tr>
<td>Age of Surviving Parent</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>47</td>
<td>48</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td>(n)</td>
<td>8</td>
<td>15</td>
<td>16</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>(%)</td>
<td>4.28</td>
<td>8.02</td>
<td>8.56</td>
<td>2.67</td>
<td>3.74</td>
<td>2.14</td>
<td>1.60</td>
<td>3.21</td>
</tr>
<tr>
<td>Age of Surviving Parent</td>
<td>51</td>
<td>52</td>
<td>53</td>
<td>54</td>
<td>55</td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(%)</td>
<td>3.21</td>
<td>0.53</td>
<td>0.53</td>
<td>1.07</td>
<td>0.53</td>
<td>0.53</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A.11: Distribution of ages of surviving parent at first interview.

<table>
<thead>
<tr>
<th>Age of Deceased Parent</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>30</th>
<th>31</th>
<th>32</th>
<th>33</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>(%)</td>
<td>1.07</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>1.07</td>
<td>1.07</td>
<td>2.14</td>
</tr>
<tr>
<td>Age of Deceased Parent</td>
<td>34</td>
<td>35</td>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td>40</td>
</tr>
<tr>
<td>(n)</td>
<td>7</td>
<td>3</td>
<td>9</td>
<td>7</td>
<td>10</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>(%)</td>
<td>3.74</td>
<td>1.6</td>
<td>4.81</td>
<td>3.74</td>
<td>5.35</td>
<td>6.95</td>
<td>6.42</td>
</tr>
<tr>
<td>Age of Deceased Parent</td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>47</td>
</tr>
<tr>
<td>(n)</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>9</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>(%)</td>
<td>5.35</td>
<td>5.35</td>
<td>5.88</td>
<td>6.42</td>
<td>4.81</td>
<td>6.95</td>
<td>3.74</td>
</tr>
<tr>
<td>Age of Deceased Parent</td>
<td>48</td>
<td>49</td>
<td>50</td>
<td>51</td>
<td>52</td>
<td>53</td>
<td>54</td>
</tr>
<tr>
<td>(n)</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>(%)</td>
<td>2.67</td>
<td>3.21</td>
<td>3.21</td>
<td>3.74</td>
<td>1.6</td>
<td>2.14</td>
<td>0.53</td>
</tr>
<tr>
<td>Age of Deceased Parent</td>
<td>55</td>
<td>56</td>
<td>57</td>
<td>58</td>
<td>59</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>(n)</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(%)</td>
<td>1.07</td>
<td>0.53</td>
<td>1.07</td>
<td>0.53</td>
<td>1.07</td>
<td>0.53</td>
<td></td>
</tr>
</tbody>
</table>

Table A.12: Distribution of ages of deceased parent.
Figure A.1: Histogram of the distribution of ages of the 328 subjects with initial interviews.

Figure A.2: Histogram of the distribution of ages of the 187 surviving parents of subjects with initial interviews.
Figure A.3: Histogram of the distribution of ages at time of death of the 187 deceased parents of subjects with initial interviews.

<table>
<thead>
<tr>
<th></th>
<th>SLI1</th>
<th>CUR1</th>
<th>SLI6</th>
<th>CUR6</th>
<th>SLI13</th>
<th>CUR13</th>
<th>SLI25</th>
<th>CUR25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.915</td>
<td>1.067</td>
<td>0.866</td>
<td>0.785</td>
<td>0.804</td>
<td>0.719</td>
<td>0.798</td>
<td>0.688</td>
</tr>
<tr>
<td>Variance</td>
<td>0.172</td>
<td>0.167</td>
<td>0.220</td>
<td>0.225</td>
<td>0.215</td>
<td>0.205</td>
<td>0.205</td>
<td>0.184</td>
</tr>
</tbody>
</table>

Table A.13: Observed SqRt-BAMO means and variances for each time period.
<table>
<thead>
<tr>
<th></th>
<th>SLI1</th>
<th>CUR1</th>
<th>SLI6</th>
<th>CUR6</th>
<th>SLI13</th>
<th>CUR13</th>
<th>SLI25</th>
<th>CUR25</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI1</td>
<td>0.172</td>
<td>0.151</td>
<td>0.109</td>
<td>0.106</td>
<td>0.091</td>
<td>0.090</td>
<td>0.081</td>
<td>0.083</td>
</tr>
<tr>
<td>CUR1</td>
<td>0.151</td>
<td>0.167</td>
<td>0.120</td>
<td>0.116</td>
<td>0.097</td>
<td>0.094</td>
<td>0.087</td>
<td>0.085</td>
</tr>
<tr>
<td>SLI6</td>
<td>0.109</td>
<td>0.120</td>
<td>0.220</td>
<td>0.212</td>
<td>0.158</td>
<td>0.157</td>
<td>0.129</td>
<td>0.118</td>
</tr>
<tr>
<td>CUR6</td>
<td>0.106</td>
<td>0.116</td>
<td>0.212</td>
<td>0.225</td>
<td>0.150</td>
<td>0.150</td>
<td>0.133</td>
<td>0.122</td>
</tr>
<tr>
<td>SLI13</td>
<td>0.091</td>
<td>0.097</td>
<td>0.158</td>
<td>0.150</td>
<td>0.215</td>
<td>0.199</td>
<td>0.122</td>
<td>0.104</td>
</tr>
<tr>
<td>CUR13</td>
<td>0.090</td>
<td>0.094</td>
<td>0.157</td>
<td>0.150</td>
<td>0.199</td>
<td>0.205</td>
<td>0.118</td>
<td>0.107</td>
</tr>
<tr>
<td>SLI25</td>
<td>0.081</td>
<td>0.087</td>
<td>0.129</td>
<td>0.133</td>
<td>0.122</td>
<td>0.118</td>
<td>0.205</td>
<td>0.179</td>
</tr>
<tr>
<td>CUR25</td>
<td>0.083</td>
<td>0.085</td>
<td>0.118</td>
<td>0.122</td>
<td>0.104</td>
<td>0.107</td>
<td>0.179</td>
<td>0.184</td>
</tr>
</tbody>
</table>

Table A.14: Observed covariance matrix for the SqRt-BAMO data.

<table>
<thead>
<tr>
<th></th>
<th>SLI1</th>
<th>CUR1</th>
<th>SLI6</th>
<th>CUR6</th>
<th>SLI13</th>
<th>CUR13</th>
<th>SLI25</th>
<th>CUR25</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI1</td>
<td>1.000</td>
<td>0.893</td>
<td>0.558</td>
<td>0.540</td>
<td>0.461</td>
<td>0.468</td>
<td>0.445</td>
<td>0.483</td>
</tr>
<tr>
<td>CUR1</td>
<td>0.893</td>
<td>1.000</td>
<td>0.614</td>
<td>0.588</td>
<td>0.503</td>
<td>0.499</td>
<td>0.477</td>
<td>0.492</td>
</tr>
<tr>
<td>SLI6</td>
<td>0.558</td>
<td>0.614</td>
<td>1.000</td>
<td>0.952</td>
<td>0.697</td>
<td>0.711</td>
<td>0.626</td>
<td>0.611</td>
</tr>
<tr>
<td>CUR6</td>
<td>0.540</td>
<td>0.588</td>
<td>0.952</td>
<td>1.000</td>
<td>0.654</td>
<td>0.671</td>
<td>0.612</td>
<td>0.595</td>
</tr>
<tr>
<td>SLI13</td>
<td>0.461</td>
<td>0.503</td>
<td>0.697</td>
<td>0.654</td>
<td>1.000</td>
<td>0.949</td>
<td>0.636</td>
<td>0.568</td>
</tr>
<tr>
<td>CUR13</td>
<td>0.468</td>
<td>0.499</td>
<td>0.711</td>
<td>0.671</td>
<td>0.949</td>
<td>1.000</td>
<td>0.642</td>
<td>0.608</td>
</tr>
<tr>
<td>SLI25</td>
<td>0.445</td>
<td>0.477</td>
<td>0.626</td>
<td>0.612</td>
<td>0.636</td>
<td>0.642</td>
<td>1.000</td>
<td>0.921</td>
</tr>
<tr>
<td>CUR25</td>
<td>0.483</td>
<td>0.492</td>
<td>0.611</td>
<td>0.595</td>
<td>0.568</td>
<td>0.608</td>
<td>0.921</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table A.15: Observed correlation matrix for the SqRt-BAMO data.
APPENDIX B

DEMOGRAPHICS AND SUMMARY STATISTICS FOR SUBJECTS WITHOUT AN INITIAL INTERVIEW

Number of subjects: 41

<table>
<thead>
<tr>
<th>Gender</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>23</td>
<td>18</td>
</tr>
<tr>
<td>%</td>
<td>56.10</td>
<td>43.90</td>
</tr>
</tbody>
</table>

Table B.1: Gender of subjects without an initial interview.

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Caucasian</th>
<th>African American</th>
<th>Asian</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>38</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>%</td>
<td>92.68</td>
<td>7.32</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table B.2: Ethnic background of subjects without an initial interview.
Table B.3: Inclusion Criteria Status (ICS) of subjects without an initial interview.

<table>
<thead>
<tr>
<th>ICS</th>
<th>Simple</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>22</td>
<td>19</td>
</tr>
<tr>
<td>%</td>
<td>53.66</td>
<td>46.34</td>
</tr>
</tbody>
</table>

Table B.4: Hollingshead score assessing socioeconomic status for subjects without an initial interview: 1=upper class, 2=upper middle class, 3=middle class, 4=lower middle class and 5=lower class.

<table>
<thead>
<tr>
<th>Hollingshead Score</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>5</td>
<td>11</td>
<td>8</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>%</td>
<td>12.20</td>
<td>26.83</td>
<td>19.51</td>
<td>29.27</td>
<td>12.20</td>
</tr>
</tbody>
</table>

Table B.5: Subjects without an initial interview classified by age group. A subject was considered an adolescent if he or she was at least 13 years old at time of parental death, a child otherwise.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Child</th>
<th>Adolescent</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>28</td>
<td>13</td>
</tr>
<tr>
<td>%</td>
<td>68.29</td>
<td>31.71</td>
</tr>
</tbody>
</table>

Table B.6: Distribution of ages of subjects without initial interviews at the time the first interview would have taken place.

<table>
<thead>
<tr>
<th>Age</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>%</td>
<td>12.20</td>
<td>12.20</td>
<td>4.88</td>
<td>4.88</td>
<td>14.63</td>
<td>4.88</td>
<td>9.76</td>
</tr>
<tr>
<td>Age</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>n</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>%</td>
<td>4.88</td>
<td>7.32</td>
<td>14.63</td>
<td>0.00</td>
<td>7.32</td>
<td>2.44</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table B.6: Distribution of ages of subjects without initial interviews at the time the first interview would have taken place.
Number of families represented by subjects without initial interviews: 27

<table>
<thead>
<tr>
<th>Number of Siblings</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>15</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>(%)</td>
<td>55.56</td>
<td>37.04</td>
<td>7.41</td>
</tr>
</tbody>
</table>

Table B.7: Number of families that participated in the study according to number of siblings for subjects without an initial interview. Note: The numbers above are somewhat misleading. A few families overlap between subjects having an initial interview and those subjects that did not. Specifically, a subject from each of three two-sibling families and one three-sibling family had an initial interview whose data was used in the analysis.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td>(%)</td>
<td>74.07</td>
<td>25.93</td>
</tr>
</tbody>
</table>

Table B.8: Gender of surviving parent for subjects without an interview.

<table>
<thead>
<tr>
<th>ICS</th>
<th>Simple</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>(%)</td>
<td>55.56</td>
<td>44.44</td>
</tr>
</tbody>
</table>

Table B.9: Inclusion Criteria Status (ICS) by family for subjects without an initial interview.
<table>
<thead>
<tr>
<th>Hollingshead Score</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>%</td>
<td>14.81</td>
<td>25.93</td>
<td>22.22</td>
<td>25.93</td>
<td>11.11</td>
</tr>
</tbody>
</table>

Table B.10: Hollingshead score assessing socioeconomic status by family for subjects without an initial interview: 1=upper class, 2=upper middle class, 3=middle class, 4=lower middle class and 5=lower class.

<table>
<thead>
<tr>
<th>Age of Surviving Parent</th>
<th>27</th>
<th>29</th>
<th>30</th>
<th>32</th>
<th>34</th>
<th>35</th>
<th>36</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>%</td>
<td>3.70</td>
<td>3.70</td>
<td>11.11</td>
<td>14.81</td>
<td>11.11</td>
<td>7.41</td>
<td>7.41</td>
</tr>
</tbody>
</table>

Table B.11: Distribution of ages of surviving parents one month after spousal death for subjects without an initial interview.

<table>
<thead>
<tr>
<th>Age of Deceased Parent</th>
<th>27</th>
<th>28</th>
<th>30</th>
<th>31</th>
<th>32</th>
<th>33</th>
<th>34</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>%</td>
<td>3.70</td>
<td>3.70</td>
<td>3.70</td>
<td>0.00</td>
<td>7.41</td>
<td>3.70</td>
<td>7.41</td>
</tr>
</tbody>
</table>

Table B.12: Distribution of ages of deceased parent of subjects without an initial interview.
Figure B.1: Histogram of the distribution of ages of the 41 subjects without initial interviews.

Figure B.2: Histogram of the distribution of ages of the 27 surviving parents of subjects without initial interviews.
Figure B.3: Histogram of the distribution of ages at time of death of the 27 deceased parents of subjects without initial interviews.

<table>
<thead>
<tr>
<th></th>
<th>SLI6</th>
<th>CUR6</th>
<th>SLI13</th>
<th>CUR13</th>
<th>SLI25</th>
<th>CUR25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.761</td>
<td>1.415</td>
<td>1.630</td>
<td>1.357</td>
<td>1.154</td>
<td>0.855</td>
</tr>
<tr>
<td>Variance</td>
<td>2.295</td>
<td>1.928</td>
<td>1.586</td>
<td>1.181</td>
<td>1.287</td>
<td>0.955</td>
</tr>
</tbody>
</table>

Table B.13: Observed SqRt-BAMO means and variances for each time frame corresponding to the 41 subjects without an initial interview.
<table>
<thead>
<tr>
<th></th>
<th>SLI6</th>
<th>CUR6</th>
<th>SLI13</th>
<th>CUR13</th>
<th>SLI25</th>
<th>CUR25</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI6</td>
<td>2.295</td>
<td>2.032</td>
<td>0.857</td>
<td>0.897</td>
<td>0.826</td>
<td>0.674</td>
</tr>
<tr>
<td>CUR6</td>
<td>2.032</td>
<td>1.928</td>
<td>0.811</td>
<td>0.847</td>
<td>0.677</td>
<td>0.597</td>
</tr>
<tr>
<td>SLI13</td>
<td>0.857</td>
<td>0.811</td>
<td>1.586</td>
<td>1.335</td>
<td>1.316</td>
<td>1.028</td>
</tr>
<tr>
<td>CUR13</td>
<td>0.897</td>
<td>0.847</td>
<td>1.335</td>
<td>1.181</td>
<td>1.110</td>
<td>0.861</td>
</tr>
<tr>
<td>SLI25</td>
<td>0.826</td>
<td>0.677</td>
<td>1.316</td>
<td>1.110</td>
<td>1.287</td>
<td>1.054</td>
</tr>
<tr>
<td>CUR25</td>
<td>0.674</td>
<td>0.597</td>
<td>1.028</td>
<td>0.861</td>
<td>1.054</td>
<td>0.955</td>
</tr>
</tbody>
</table>

Table B.14: Observed covariance matrix for the SqRt-BAMO data corresponding to the 41 subjects without an initial interview.

<table>
<thead>
<tr>
<th></th>
<th>SLI6</th>
<th>CUR6</th>
<th>SLI13</th>
<th>CUR13</th>
<th>SLI25</th>
<th>CUR25</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI6</td>
<td>1.000</td>
<td>0.966</td>
<td>0.465</td>
<td>0.547</td>
<td>0.507</td>
<td>0.487</td>
</tr>
<tr>
<td>CUR6</td>
<td>0.966</td>
<td>1.000</td>
<td>0.446</td>
<td>0.524</td>
<td>0.419</td>
<td>0.435</td>
</tr>
<tr>
<td>SLI13</td>
<td>0.465</td>
<td>0.446</td>
<td>1.000</td>
<td>0.976</td>
<td>0.806</td>
<td>0.719</td>
</tr>
<tr>
<td>CUR13</td>
<td>0.547</td>
<td>0.524</td>
<td>0.976</td>
<td>1.000</td>
<td>0.780</td>
<td>0.690</td>
</tr>
<tr>
<td>SLI25</td>
<td>0.507</td>
<td>0.419</td>
<td>0.806</td>
<td>0.780</td>
<td>1.000</td>
<td>0.951</td>
</tr>
<tr>
<td>CUR25</td>
<td>0.487</td>
<td>0.435</td>
<td>0.719</td>
<td>0.690</td>
<td>0.951</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table B.15: Observed correlation matrix for the SqRt-BAMO data corresponding to the 41 subjects without an initial interview.
APPENDIX C

PROGRAM CODE FOR ESTIMATING A TWO-COMPONENT MULTIVARIATE NORMAL MIXTURE DISTRIBUTION

The following code is based on the paper by Lindsay (1995) which outlines a method of moments technique for estimating a two-component multivariate normal mixture distribution. The main program and subroutines were written using the R statistical software package (Ihaka and Gentleman 1996).

"MVNMixtureDensity" = function(data.df, p = 2, FindPrimaryAxis = F, PA = 1, n.int = 10, lower = -2, upper = (max(data.df,na.rm=T)+3)){

  # Estimate a 2-component multivariate normal mixture distribution,
  # p(x; Q_2, sigma^2), using the method of moments technique
  # described in
  #
  # Bruce G. Lindsay (1993) "Multivariate Normal Mixtures: A Fast,
  # Consistent Method of Moments". JASA Vol. 88, No. 422,   
  # pp. 468-476.
  #
  # and
  #
  # Bruce G. Lindsay (1989b) "Moment Matrices: Applications
  # in Mixtures". The Annals of Statistics vol. 17, no. 2,  
  # 722-740.
  #
# External Variables:
# data.df - Nxd data matrix thought to be a p-component multivariate normal mixture
# p - number of components
# FindPrimaryAxis - if TRUE, program determines primary axis
# PA - integer indicating which column to use as primary axis if FindPrimaryAxis is FALSE; default is one
# n.int - number of subintervals to search for roots in bracket()
# lower - lower bound of possible values of primary axis coordinates
# upper - upper bound of possible values of primary axis coordinates
#
# Local Variables:
# N - number of observations (rows) in data.df
# d - dimension (no. of columns) of data
# n - 1xd vector containing number of non-missing observations per data column in data.df
# eps - accuracy of final answer in bisection()
# mn.df - 1xd vector of column means of "data.df"
# x - data such that each column is centered about the origin
# m.hat - (2p+1)xd matrix of the estimates of E[theta^k), k=0,1...2p for data column data.df[,j], j = 1...d
# PrimaryAxis - indicates which column of "data.df" is to be used as the primary axis
# sd.x - sample standard deviations for each column of "x"
# sp - pxd matrix, storing for each time frame, the univariate sd estimates s1, s2, ..., sp, where sp is the estimated sd assuming p components.
# sigma11.hat - MOM variance estimates corresponding to each column of "x" if it was assumed to be the primary axis
# sigma1k.hat - covariance estimates between the primary axis and remaining data data columns
# sm - matrix storing sample moments of the for mean(X^j*Y_k), j=1...3, k=2...d
# theta.hat - the 1xd vector containing the mean parameters estimated via the primary axis; i.e the first coordinate of each of the p component d*1 mean vectors
# Pi - 1xp vector of population proportion estimates
# phi.hat - (p-1)xd matrix containing mean parameter estimates of coordinates 2 through d for the p component mean vectors
# c.hat - 1x(d-1) vector of estimates of E[theta*phi_k], k=2...d
# S - sample covariance matrix
# S.inv - inverse of S
# Sigma.Q.hat - estimated covariance matrix of mixture
distribution Q
# Sigma.hat - component covariance matrix;
# NOTE: S = Sigma.hat + Sigma.Q.hat
# alpha - if less than 1, Sigma.hat is recalculated ("shrunk") to
# be positive definite
# eps - small value used in calculating alpha
# eigen.max - eigen value used to calculated alpha
# support - estimated support points of the mixing distribution
# ll - value of log-likelihood evaluated using estimated parameters

# Initialization
N = dim(data.df)[1]
d = dim(data.df)[2]
n = apply(data.df, 2, FUN=function(x) sum(!is.na(x)))
eps = sqrt(.Machine$double.eps)
PrimaryAxis = PA
x = data.df[,c(PrimaryAxis, (1:d)[PrimaryAxis])]
mn.df = apply(x, 2, mean, na.rm = T)
for(i in 1:d) x[,i] = x[,i] - mn.df[i]
sd.x = sqrt(apply(x, 2, var, na.rm=T))

# Step 1: Estimate sd of primary axis (sigma_11)

# Estimate moments and sigma for each column of data via Lindsay's
# (1989) univariate analysis. The basic idea s1 >= s2 >= ... sp=0
# where s_k is the estimated sd for a k-component mixture
distribution. A bisection algorithm will be repeatedly applied
# until sp is obtained.
sigma11.hat = numeric(d)
sp = matrix(0,p,d)
colnames(sp) = names(x)
rownames(sp) = character(p)
for(i in 1:p) rownames(sp)[i] = paste("s",i, sep="")
m.hat = matrix(0,(2*p+1),d)
colnames(m.hat) = colnames(x)
rownames(m.hat) = character((2*p+1))
for(i in 1:(2*p+1))
rownames(m.hat)[i] = paste("theta^",i-1,sep="")
for(i in 1:d){
    for(j in 2:p){
        sp[1,i] = sd.x[i]
        intervals = bracket(detMp.2, x1=0, x2=sp[j-1,i], n.int, 1, x = x[,i], p=j)
        sp[j,i] = bisection(detMp.2, x1=intervals[1], x2=intervals[2], eps, p=j, x=x[,i])
    }
    sigma11.hat[i] = sp[p,i]
    m.hat[,i] = PrimaryAxisMoments(sigma11.hat[i], x[,i], p)
}

#Step 1*: Find primary axis and make it the first column of x.
if(FindPrimaryAxis){
    ratio = m.hat[3,]/sd.x^2
    PrimaryAxis = (1:d)[rank(ratio) == d]
    new.order = c(PrimaryAxis, (1:d)[-PrimaryAxis])
    x = x[, new.order]
    sd.x = sd.x[new.order]
    sp = sp[, new.order]
    m.hat = m.hat[, new.order]
    sigma11.hat = sigma11.hat[new.order]
    mn.df = mn.df[new.order]
}

#Step 2: Solve for theta[i], i=1...d (coordinates corresponding to primary axis)

#Find support points for p=2. If p > 2, then use bisection routine to find roots based on the nesting property of roots for (p-1) vs p in lemma 2D in Lindsay’s 1989b paper.
    r = theta.hat = numeric(p)
    M = matrix(0, 3, 3)
    for(i in 1:3) M[i,] = m.hat[i:(i+2),1]
    a = M[3,1] - M[2,1]^2
\begin{align*}
r[1] &= \frac{-b + \sqrt{b^2 - 4 \times a \times e}}{2 \times a} \\
r[2] &= \frac{-b - \sqrt{b^2 - 4 \times a \times e}}{2 \times a}
\end{align*}

if (sum(is.na(r)) == 0)
\quad \text{theta.hat}[1:2] = \text{sort}(r[1:2])
else {
\quad \text{cat}("Support points have imaginary values.\n")
\quad \text{return}(\text{NaN})
}

if (p > 2) {
\quad \text{for}(i \text{ in } 3:p) \{ \\
\qquad m = \text{PrimaryAxisMoments}(\text{sp}[i,1], x[,1], p=i) \\
\qquad M = \text{matrix}(0, (i+1), (i+1)) \\
\qquad \text{for } (k \text{ in } 1:(i+1)) \ M[k,] = m.hat[k:(k+i),1] \\
\qquad \text{for }(j \text{ in } 1:i) \\
\qquad \qquad \text{if } (j == 1) \\
\qquad \qquad \quad r[1] = \text{bisection}(\text{Sp.2}, \text{lower}, \text{theta.hat}[1], \text{eps}, \\
\qquad \qquad \quad \quad p = i, \text{mp} = M[,-1]) \\
\qquad \quad \text{else if } (j == i) \\
\qquad \qquad \quad r[1] = \text{bisection}(\text{Sp.2}, \text{theta.hat}[i-1], \text{upper}, \text{eps}, \\
\qquad \qquad \quad \quad p = i, \text{mp} = M[,-1]) \\
\qquad \quad \text{else} \\
\qquad \qquad \quad r[j] = \text{bisection}(\text{Sp.2}, \text{theta.hat}[j-1], \text{theta.hat}[j], \\
\qquad \qquad \quad \quad \text{eps}, p = i, \text{mp} = M[,-1]) \\
\qquad \quad \text{theta.hat}[1:i] = r[1:i] \\
\quad \}\n\}

\text{theta.hat} = \text{t(as.matrix(theta.hat))}
\text{col.labels} = \text{character}(p)
\text{for}(i \text{ in } 1:p) \ \text{col.labels}[i] = \text{paste}("\text{mu}",i, \text{sep=""})
\text{dimnames(theta.hat)} = \text{list(names(x)[1], col.labels)}

#Step 3: Solve for the population proportions
\text{V} = \text{matrix}(1, p, p)
\text{for } (i \text{ in } 1:(p-1)) \ V[(i+1),] = \text{theta.hat}^i
\text{Pi} = \text{solve(V, m.hat[1:p,1])}

98
#Step 4: Estimate covariance terms $E[\theta*\phi_k]$, $k=2...d$, where $X$ is the primary axis. NOTE: Formulas are for $p = 2$ only.

```r
sigma1k.hat = numeric(d-1)
names(sigma1k.hat) = names(x)[-1]
sm = matrix(0,3,d-1)
colnames(sm) = character(d-1)
for(i in 2:d) colnames(sm)[i-1] = paste("Y",i,sep="")
rownames(sm) = character(3)
for(i in 1:3) rownames(sm)[i] = paste("X^",i,sep="")
for(i in 1:3)
  sm[i,] = apply((x[,1]^i*x[,-1]), 2, mean, na.rm=T)
for(k in 1:(d-1))
  sigma1k.hat[k] = cov1k(m.hat[3,1], m.hat[4,1], m.hat[5,1],
                        sm[1,k], sm[2,k], sm[3,k], sigma11.hat[1])
```

#Step 5: Solve for coordinates 2...p of remaining support points. NOTE: Solutions below are for $p=2$ only.

```r
phi.hat = matrix(0,(d-1),p)
dimnames(phi.hat) = list(names(x)[-1], c("mu1","mu2"))
c.hat = ThetaPhiMoments(c(sigma11.hat[1], sigma1k.hat), x, j=1,
                        p=2, dm=d)
for(j in 1:p)
  phi.hat[,j] = theta.hat[j] * c.hat / m.hat[3,1]
```

#Step 6: Solve for remaining elements of Sigma.hat using the relationship $S = \text{Sigma.hat} - \text{Sigma.Q.hat}$ where $S$ is the sample covariance and Sigma.Q.hat is the covariance matrix of the mixing distribution Q.hat.

```r
support = as.data.frame(rbind(theta.hat, phi.hat))
mu1 = support$mu1
mu2 = support$mu2
Sigma.Q.hat = Pi[1]*Pi[2]*(mu1 - mu2) %o% (mu1 - mu2)
S = cov(x, use="pairwise")
Sigma.hat = S - Sigma.Q.hat
Sigma.hat[1,1] = sigma11.hat[1]
Sigma.hat[1,2:d] = Sigma.hat[2:d,1] = sigma1k.hat
```

99
#Step 7: Force Sigma.hat to be positive definite.

alpha = 1
if(det(Sigma.hat) <= 0){
    S.inv = solve(S)
    eigen.max = max(eigen((S.inv %*% Sigma.Q.hat), symmetric=T, 
                      only.values=T)$values)
    eps = sqrt(.Machine$double.eps)
    alpha = min(1, sqrt((1-eps)/eigen.max))
    if (alpha < 1){
        support = alpha*support
        Sigma.hat = S - alpha^2 * Sigma.Q.hat
    }
}

if(det(Sigma.hat) <= 0)
cat("\n**Estimated covariance matrix is negative definite**\n")

col.labels = character(d)
col.labels[1] = "theta"
for(i in 2:d) col.labels[i] = paste("phi",i, sep="")
dimnames(Sigma.Q.hat) = list(col.labels, col.labels)
dimnames(Sigma.hat) = list(colnames(x), colnames(x))
support = support + mn.df
rownames(support) = colnames(x)

if(PrimaryAxis > 1) {
    old.order = 1:d
    old.order[PrimaryAxis]=1
    old.order[1:(PrimaryAxis-1)] = 2:PrimaryAxis
    support = support[old.order,]
    Sigma.hat = Sigma.hat[old.order, old.order]
    Sigma.Q.hat = Sigma.Q.hat[old.order, old.order]
}

ll = ll.mvn.mix(data.df, mu1=support$mu1, mu2=support$mu2,
                 Pi=Pi[1], Sigma=Sigma.hat)
# Output

cat("Primary Axis is column ", PrimaryAxis,
  " (", names(data.df)[PrimaryAxis], ").\n\n", sep="")

return(list(support = support, Pi = Pi, ll = ll,
  PrimaryAxis = PrimaryAxis, alpha=alpha,
  Sigma.hat = Sigma.hat, Sigma.Q.hat = Sigma.Q.hat))
}

*******************************************************************************

"PrimaryAxisMoments" = function(sigma, x, p=2){

  #Calculate estimated moments, E[theta^k], k=1...2p

  #External Variables:
  # sigma - standard deviation of primary axis
  # x - data vector
  # p - number of mixture components

  out = c(1,rep(0,(2*p)))
  labels = vector("character", 2*p+1)
  for(i in 0:(2*p)) labels[i+1] = paste("theta^",i,sep="")
  names(out) = labels
  for(k in 1:(2*p)){
    ans=0
    for(m in 0:k){
      d = k - m
      if(d %% 2 == 0) {
        coef = ifelse((m==0 | m==k), 1,
          prod(seq(k-m+1, k))/prod(seq(1, m)))
        EX = mean(x^m, na.rm=T)
        EZ = ifelse(d == 0, 1, prod(seq(1,d,2)) * sigma^(d/2))
        i = ifelse((d/2) %% 2 == 0, 1, -1)
        ans = ans + coef * i * EX * EZ
      }
    }
    out[k+1] = ans
  }
  return(out)
}
"ThetaPhiMoments" = function(cov.1k, x, j=1, p=2, dm=dim(x)[2]){

  #Calculate moments of the form E[theta^j * phi_k], k=2...dm
  #External Variables:
  # cov.1k - covariance estimates between x[,1] and columns of x
  # x - data matrix, first column being the primary axis
  # j - what power to raise theta by
  # p - number of mixture components
  # dm - number of columns in x

  out = numeric(dm)

  for(k in 2:dm){
    ans = 0
    for(m in 0:j){
      d = j - m
      coef = ifelse((m==0 | m==j), 1,
                     prod(seq(j-m+1, j))/prod(seq(1, m)))
      i = ifelse(d %% 2 == 0, (-1)^((d/2)), (-1)^((d+1)/2))
      EX = ifelse((d %% 2) == 0, mean(x[,k]*x[,1]^m, na.rm=T),
                  mean(x[,1]^m, na.rm=T))
      if(d == 0)
        EZ = 1
      else {
        EZ = ifelse(((d %% 2) == 0),
                     prod(seq(1,d,2)) * cov.1k[1]^(d/2),
                     prod(seq(1,(d+1),2)) * cov.1k[1]^-((d+1)/2) *
                     cov.1k[k]/cov.1k[1])
      }
      ans = ans + coef * i * EX * EZ
    }
    out[k] = ans
  }
  out = t(as.matrix(out[-1]))
  row.labels = paste("theta","^",j,sep="")
  col.labels = character(dm-1)
  for (i in 2:dm) col.labels[i-1] = paste("phi_",i-1,sep="")
  dimnames(out)=list(row.labels, col.labels)
  return(out)
}
"detMp.2" = function(sigma, x, p=2){
    #Calculate determinate of the moment matrix

    #External Variables:
    # sigma - estimated standard deviation of data x
    # x - data vector
    # p - number of mixture components

    m.hat = PrimaryAxisMoments(sigma, x, p)
    Mp = matrix(0, (p+1), (p+1))
    for (i in 1:(p+1)) Mp[i,] = m.hat[i:(i+p)]
    detMp = det(Mp)
    return(detMp)
}

"Sp.2" = function(t, p=2, mp){
    #Construction of polynomial used to determine the points of
    #support along the primary axis.

    #External Variables:
    # t - polynomial variable
    # p - number of mixture components
    # mp - the first p columns of the primary axis moment matrix

    M = cbind(mp, t^(0:p))
    detM = det(M)
    return(detM)
}

"cov1k" = function(m2, m3, m4, EXY, EX2Y, EX3Y, sigma11) {
    #Calculate covariance terms between primary axis and
    #remaining variables

    #External Variables:
    # m2, m3, m4 - the 2nd, 3rd, and 4th primary axis moments
    # EXY - estimate of mean of product of the primary axis and
# other variable
# EX2Y - estimate of mean of product of the primary axis^2 and
# other variable
# EX3Y - estimate of mean of product of the primary axis^3 and
# other variables
# sigma11 - standard deviation associated with the primary axis

ans = EXY - (6*m2*EXY*(m2+sigma11) + 3*m3*EX2Y - 2*m2*EX3Y)/(m4 + 3*m2^2)

return(ans)

"bracket" = function(fx, x1, x2, m, max.in, ...) {

# Determine subintervals of the interval [x1, x2] where the function
# fx = 0. The algorithm used was obtained from
#
# New York, NY. p. 352
#
# External Variables:
# fx - function to be evaluated
# x1, x2 - endpoints of the interval [x1, x2] over which fx is
evaluated
# m - number of equally-sized subintervals to evaluate
# max.in - maximum number of roots sought
# ... - optional arguments to be supplied to "fx"; must be in the
# form "arg1 = value1, arg2 = value2, etc..." (quotes omitted)
#
# Local Variables:
# max.out - actual number of bracketing pairs found
# x1b, x2b - vectors of lower/upper endpoints of intervals
# containing roots to function fx.
# dx - length of subintervals to be searched

# Initialize
max.out = 0
x1b = x2b = numeric(max.in)
if (x1 > x2) return("Must have x1 < x2."

dx = (x2 - x1)/m
x = x1
f.lo = fx(x, ...)

#If sign change occurs record bracketing values, otherwise proceed
#to the next subinterval.

for (i in 1:m) {
  x = x + dx
  f.hi = fx(x, ...)
  if (f.lo * f.hi <= 0){
    max.out = max.out + 1
    x1b[max.out] = x - dx
    x2b[max.out] = x
    if (max.in == max.out) return(cbind(x1b = x1b, x2b = x2b))
  }
  f.lo = f.hi
}

return(cbind(x1b = x1b[1:max.out], x2b = x2b[1:max.out]))

"bisection" = function(fx, x1, x2, epsilon, maxiter=40, ...) {

  #Bisection method for finding the root of a function known to lie
  #in the interval [x1, x2]. The algorithm used was obtained from
  #New York, NY. p. 354
  #
  #External Variables:
  #  fx - function to be evaluated
  #  x1, x2 - endpoints of interval known to contain a root of "fx"
  #  epsilon - the final answer will have accuracy to within
  #            +/- epsilon of the true root.
  #  ... - optional arguments to be supplied to "fx"; must be in the
  #        form "arg1 = value1, arg2 = value2, etc..." (quotes omitted)
  #  maxiter - maximum number of bisections performed.
  #
  #Local Variables:
# root - a point in [x1, x2] where fx = 0
# xmid - the midpoint of the current interval being evaluated
# f1, fmid - value of "fx" evaluated at x1 and xmid, respectively
# dx - length of current interval being evaluated
#

# Initialize
root = xmid = 0
f1 = fx(x1, ...)
fmid = fx(x2, ...)
if (f1 * fmid > 0) return("[x1, x2] must contain root.")

# Orient search to begin at endpoint where fx < 0.
if (f1 < 0) {
dx = x2 - x1
root = x1
}
else {
dx = x1 - x2
root = x2
}

# Bisection loop
for (i in 1:maxiter) {
dx = dx/2
xmid = root + dx
fmid = fx(xmid, ...)
if (fmid <= 0) root = xmid
if(abs(dx) < epsilon || fmid == 0) return(root)
}

cat("Performed",maxiter,
    "bisections without achieving desired accuracy.\n")
return(root)
}
"ll.mvn.mix" = function(x, mu1, mu2, Pi, Sigma){
    #Calculate log-likelihood for a 2-component multivariate normal
    #mixture distribution.
    #External Variables:
    # x - data thought to be a mixture
    # mu1, mu2 - component means
    # Pi - proportion associated with first component
    # Sigma - common component covariance matrix
    ll = sum(log(apply(x, 1, mvn.mix, mu1=mu1, mu2=mu2, Pi = Pi,
                      Sigma = Sigma)))
    return(ll)
}

"mvn.mix" = function(x, mu1, mu2, Pi, Sigma){
    #This calculates the density of a 2-component multivariate normal
    #mixture assuming equal covariance structures.
    #External Variables:
    # x - data vector
    # mu1, mu2 - component mean vectors;
    # Pi - weight
    # Sigma - covariance structure
    ans = Pi * mvn(x, mu1, Sigma) + (1-Pi) * mvn(x, mu2, Sigma)
    return(ans)
}

"mvn" <- function(x, mu=c(0,0,0), sigma=diag(c(1,1,1))) {
    #This function computes the multivariate normal distribution.
    #External Variables:
    # x - data point vector
    # mu - mean vector
    # sigma - covariance matrix
}
keep = (1:length(x))[!is.na(x)]
p = length(keep)
X <- as.matrix(x[keep] - mu[keep])
if(dim(X)[1] != 1) X = t(X)
sigma.inv <- solve(sigma[keep,keep])
ans <- (2*pi)^(-p/2) *
det(sigma.inv)^(1/2) * exp(-0.5 * X %*% sigma.inv %*% t(X))
return(ans)

"EM.MVNmix" = function(x, mu1, mu2, Pi, Sigma, max.iter=100,
eps=0.0001){

#EM algorithm used to obtain maximum likelihood estimates for a
two-component multivariate mixture distribution.
#
#External Variables:
# x - data thought to be a mixture
# mu1, mu2 - component mean estimates
# Pi - proportion associated with first component distribution
# Sigma - estimated component covariance matrix
# eps - maximum allowable error between old and new parameters

N = 0
error = eps + 1
ll = ll.max = 0
mu1.new = mu1.max = mu1
mu2.new = mu2.max = mu2
Pi.new = Pi.max = Pi
Sigma.new = Sigma.max = Sigma

while(error > eps & N <= max.iter){

N = N + 1
#E-Step: Assign each observation to one of two groups.

n = dim(x)[1]
group = numeric(n)
ratio = numeric(n)
for(i in 1:n){
    ratio[i] = Pi.new * mvn(x[i,], mu1.new, Sigma.new)/
    ((1-Pi.new) * mvn(x[i,], mu2.new, Sigma.new))
    group[i] = ifelse(ratio[i] > 1, 1, 2)
}

#M-Step: Calculate MLE's based on grouping

mu1.old = mu1.new
mu2.old = mu2.new
Pi.old = Pi.new
Sigma.old = Sigma.new

n1 = sum(group==1)
n2 = sum(group==2)
S1 = cov(x[group==1,], use = "pairwise")
S2 = cov(x[group==2,], use = "pairwise")

mu1.new = apply(x[group==1,], 2, mean, na.rm = T)
mu2.new = apply(x[group==2,], 2, mean, na.rm = T)
Pi.new = n1/(n1+n2)
Sigma.new = ((n1-1)*S1 + (n2-1)*S2)/(n1 + n2 - 2)

#Since missing group data is "discrete", ll is not guaranteed
#to consecutively increase. Stopping rule may not necessarily
#coincide with MLE.

if(ll > ll.max){
    ll.max = ll
    mu1.max = mu1.new
    mu2.max = mu2.new
    Pi.max = Pi.new
    Sigma.max = Sigma.new
}
	error = max(abs(c(mu1.new-mu1.old, mu2.new-mu2.old,
    as.vector(Sigma.new-Sigma.old))))

ll = ll.mvn.mix(x, mu1=mu1.new, mu2=mu2.new, Pi=Pi.new,
    Sigma=Sigma.new)

cat("Error =", error, " ll =", ll, "\n")
}
if(N > max.iter)
    cat("\nConvergence unsuccessful after", max.iter,
        " iterations.\n\n")

ll = ll.mvn.mix(x, mu1=mu1.new, mu2=mu2.new, Pi=Pi.new,
                Sigma=Sigma.new)
weights = c(Pi.new, 1-Pi.new)
support.max = as.data.frame(cbind(mu1=mu1.max, mu2=mu2.max))
weights.max = c(Pi.max, 1-Pi.max)

return(list(support=support, Pi=weights, Sigma.hat=Sigma.new,
           ll = ll, group=group, support.max = support.max,
           Pi.max = weights.max, Sigma.hat.max = Sigma.max,
           ll.max = ll.max, error=error, eps=eps, iter=N))
}
APPENDIX D

PROGRAM CODE FOR ESTIMATING A P-COMPONENT UNIVARIATE NORMAL MIXTURE DISTRIBUTION

The following code is based on two papers, Lindsay (1989a) and Lindsay (1989a), which outline a method of moments technique for estimating a p-component univariate normal mixture distribution. The main program and subroutines were written using the R statistical software package (Ihaka and Gentleman 1996).

"MixtureDensity" = function(x, p = 1, lower.limit = -2, upper.limit = 7, n.int = 10) {
  #
  #Estimate the p-component mixture distribution, p(x; Q_p, sigma^2), using the method of moments technique
  # described in
  #
  #
  #External Variables:
  #  x - data thought to be a p-component mixture
  #  p - number of components
  #  n.int - number of subintervals to search for roots in bracket()
  #  lower.limit - lower bound for possible values of support points
  #  upper.limit - upper bound for possible values of support points
  #
# Local Variables:
# n - number of observations in "x"
# s.p.hat - component variance estimates for a k-component mixing
distribution for k = 1, 2, ..., p
# Mp.hat - (p+1) x (p+1) moment matrix
# intervals - matrix with two columns such that each row contains
# endpoints of an interval where the function
# det(Mp.hat) = 0 occurs
# m - 1x(2p+1) vector of estimated moments, E[(X + i*sigma*Z)^j]
# j = 0, 1,..., 2p, where i=sqrt(-1), X~Normal(Q,sigma) and
# Z~Normal(0,1)
# r - storage of current values of support points
# support - final estimate of support points
# weights - probabilities corresponding to the support points
# V - Vandermonde matrix of the support points used to estimate
# the weights via a linear equation
#

#(1) Define m_p(G) to be the pth moment of some distribution G.
# Then m_p(Q) = E[(X + i*sigma*Z)^p | X = x] when
# X ~ Normal(Q,sigma). These unbiased estimates of the moments
# of Q are the Hermite polynomials which have the recursive
# relationship g_p(x) = x*g_{p-1}(x) - (p-1)*s^2*g_{p-2}(x)
# with g_1(x) = x and g_2(x) = x^2 - sigma^2. The pth moment
# matrix of Q, denoted by Mp(Q), has (i,j)th element m_{i+j}(Q),
i, j = 0, 1, 2, ..., p.

if (p < 2) return("p must be an integer > 1.")
n = length(x)
m = numeric(2*p+1)
eps = sqrt(.Machine$double.eps)

#(2) Estimate sigma by assigning it the smallest non-negative
# root of the equation d(sigma) = det(M_p) = 0. A Bisection
# algorithm will be used since s = s.1 >= s.2 >= ... s.n = 0.
# Use this estimate and the empirical distribution, Fn, to
# estimate the moment matrix Mp(Q).

s.p.hat = numeric(p)
\[ s.p.hat[1] = \sqrt{\text{var}(x) \times (n-1)/n} \]

for (i in 2:p) {
    intervals = bracket(detMp, 0, s.p.hat[i-1], n.int, 1,
                        p=i, data=x)
    s.p.hat[i] = bisection(detMp, intervals[1,1],
                           intervals[1,2], eps, p=i, data=x)
}

Mp.hat = Mp(s.p.hat[p], p, x)
m[1:p] = Mp.hat[1:p,1]
m[(p+1):(2*p+1)] = Mp.hat[(p+1),]

#(3) The support points of the mixing distribution Q are found
# by solving \( S_p(t) = 0 \) where \( S_p(t) \) is the determinate of the
# estimated \((p+1) \times (p+1)\) moment matrix \( M_p(Q) \) with its last
# column replaced by the vector of polynomial elements
# \((1, t, t^2, \ldots, t^p)\).
#
# Find roots for \( p = 2 \). If \( p > 2 \), then use bisection routine
# to find roots based on the nesting property of roots for
# \((p-1)\) vs \( p \) in lemma 2D in Lindsay’s 1989 paper.

r = support = numeric(p)
M = Mp(s.p.hat[2], 2, x)
a = M[3,1] - M[2,1]^2
r[1] = (-b + sqrt(b^2 - 4 * a * d)) / (2*a)
r[2] = (-b - sqrt(b^2 - 4 * a * d)) / (2*a)

if (sum(is.na(r)) == 0)
    support[1:2] = sort(r[1:2])
else {
    cat("MixtureDensity: roots have imaginary values.\n")
    return(NaN)
}

if (p > 2) {
    for(i in 3:p) {
        M = Mp(s.p.hat[i], i, x)[,1:i]
for(j in 1:i){
    if (j == 1)
        r[1] = bisection(Sp, lower.limit, support[1], eps, 
        p = i, mp = M)
    else if (j == i)
        r[i] = bisection(Sp, support[i-1], upper.limit, eps, 
        p = i, mp = M)
    else
        r[j] = bisection(Sp, support[j-1], support[j], eps, 
        p = i, mp = M)
}
support[1:i] = r[1:i]
}

V = matrix(1, p, p)
for (i in 1:(p-1)) V[(i+1),] = support^i
weights = solve(V, m[1:p])
return(list(support=support, weights=weights, 
           sigma.hat = s.p.hat[p]))

**********************************************************************
.bracket" = function(fx, x1, x2, m, max.in, ...) {

    #Determine subintervals of the interval [x1, x2] where the function
    #fx = 0. The algorithm used was obtained from
# External Variables:
#   fx - function to be evaluated
#   x1, x2 - endpoints of the interval \([x1, x2]\) where \(fx\) is evaluated
#   m - number of equally-sized subintervals to evaluate
#   max.in - maximum number of roots sought
#   ... - optional arguments to be supplied to "fx"; must be in the
#           form "arg1 = value1, arg2 = value2, etc..." (quotes omitted)
#
# Local Variables:
#   max.out - actual number of bracketing pairs found
#   x1b, x2b - vectors of lower/upper endpoints of intervals
#              containing roots to function \(fx\).
#   dx - length of subintervals to be searched

# Initialize
max.out = 0
x1b = x2b = numeric(max.in)
if (x1 > x2) return("Must have x1 < x2.")
dx = (x2 - x1)/m
x = x1
f.lo = fx(x, ...)

# If sign change occurs record bracketing values, otherwise proceed
# to the next subinterval.
for (i in 1:m) {
  x = x + dx
  f.hi = fx(x, ...)
  if (f.lo * f.hi <= 0){
    max.out = max.out + 1
    x1b[max.out] = x - dx
    x2b[max.out] = x
    if (max.in == max.out) return(cbind(x1b = x1b, x2b = x2b))
  }
  f.lo = f.hi
}

return(cbind(x1b = x1b[1:max.out], x2b = x2b[1:max.out])))
"bisection" = function(fx, x1, x2, epsilon, maxiter=40, ... ) {
    
    #Bisection method for finding the root of a function known to lie
    #in the interval [x1, x2]. The algorithm used was obtained from
    #
    #New York, NY. p. 354
    #
    #External Variables:
    # fx - function to be evaluated
    # x1, x2 - endpoints of interval known to contain a root of "fx"
    # epsilon - the final answer will have accuracy to within
    # +/- epsilon of the true root.
    # ... - optional arguments to be supplied to "fx"; must be in the
    # form "arg1 = value1, arg2 = value2, etc..." (quotes omitted)
    # maxiter - maximum number of bisections performed.
    #
    #Local Variables:
    # root - a point in [x1, x2] where fx = 0
    # xmid - the midpoint of the current interval being evaluated
    # f1, fmid - value of "fx" evaluated at x1 and xmid, respectively
    # dx - length of current interval being evaluated
    #
    #Initialize
    root = xmid = 0
    f1 = fx(x1, ...)
    fmid = fx(x2, ...)
    if (f1 * fmid > 0) return("[x1, x2] must contain root."")
    
    #Orient search to begin at endpoint where fx < 0.
    if (f1 < 0) {
        dx = x2 - x1
        root = x1
    }
    else {
        dx = x1 - x2
        root = x2
    }
# Bisection loop
for (i in 1:maxiter) {
  dx = dx/2
  xmid = root + dx
  fmid = fx(xmid, ...)
  if (fmid <= 0) root = xmid
  if(abs(dx) < epsilon || fmid == 0) return(root)
}

cat("Performed",maxiter,
    
    "bisections without achieving desired accuracy.\n")
return(root)

"Mp" = function(sigma, p, data) {
# # A moment matrix is constructed via equation 4.1 in Lindsay (1989)
# # Bruce J. Lindsay (1989). "Moment Matrices: Applications in
# #
# # External Variables:
# # sigma - variance of component distributions
# # p - number of components in mixture distribution
# # data - data vector
# #
# # Local Variables:
# # x.unique - unique values in "data"
# # fx.unique - empirical probabilities for each value in "x.unique"
# # hp - m x 2p matrix where row k consists of the Hermite polynomial
# # values E[(X + i*sigma*Z)^j|X=x.unique[k]] for j = 1...2p
# # where "m" is the length of x.unique
# # Ehp - vector of expected values of "hp" with respect to the
# # empirical distribution
# # which is used to estimate the moments for a normal mixture
# # Mp - (p+1) x (p+1) moment matrix containing elements of "Ehp"
#
# Empirical Distribution
x.unique = unique(data)
fx.unique = table(data)/length(data)

# Calculate moment matrix using Hermite polynomials
hp = sapply(x.unique,"Hermite", p=(2*p), sigma=sigma)
Ehp = c(1, fx.unique %*% hp)
Mp = matrix(0, (p+1), (p+1))
for(i in 1:(p+1)) Mp[i,] = Ehp[i:(i+p)]
return(Mp)

"detMp" = function(sigma, p, data) {
  # Calculate the determinate of the moment matrix for normal mixtures
  # described in Lindsay (1989).
  #
  # Bruce J. Lindsay (1989). "Moment Matrices: Applications in
  #
  # External Variables:
  #  sigma - variance of component distributions
  #  p - number of components in mixture distribution
  #  data - data vector
  #
  # Local Variables:
  #  mp - (p+1) x (p+1) moment matrix
  #  dp - determinate of the moment matrix
  
  mp = Mp(sigma, p, data)
dp = det(mp)
return(dp)
}


