2006

A Normal-Mixture Model with Random-Effects for RR-Interval Data

Jessica McKinney Ketchum
Virginia Commonwealth University

A NORMAL-MIXTURE MODEL WITH RANDOM-EFFECTS FOR RR-INTERVAL DATA

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

JESSICA MCKINNEY KETCHUM

B.S. Mathematics, New Mexico State University, 2001

Co-Director: VISWANATHAN RAMAKRISHNAN, PH.D.

ASSOCIATE PROFESSOR, DEPARTMENT OF BIOSTATISTICS

Co-Director: AL M. BEST, PH.D

ASSOCIATE PROFESOR, DEPARTMENT OF BIOSTATISTICS

Virginia Commonwealth University

Richmond, Virginia

August 2006

Acknowledgement

Above all I would like to thank my mother and father. From the day I was born you taught me how to aim for the stars. You showed me that nothing is unreachable with simple dedication, hard work, and drive. Mother, I thank you for teaching me independence. You have always expressed how proud of me you are, no matter what. This has given me the strength to take risks in life. And you have taught me to learn from these risks, regardless of their consequences, making all life's experience worthwhile. Father, I thank you for teaching me dedication. You showed me the value of education and the commitment it takes to achieve it. I thank you most for moving our family abroad. I cannot express enough how much that shaped my soul. Despite the outcome, I have no regrets. For my dear brother, Justin, thank you for being my best friend and yet my worst enemy, my partner in crime and the one to turn me in, my complete opposite but the only one in this world just like me.

Second, I would like to thank my future husband, Sid, and his family, for supporting me emotionally through this seemingly impossible accomplishment. Sid, you listened to me go on and on and on about my research, without much clue as to what I was talking about. As painful as this must have been for you, I thank you and love you dearly for this. The daily social interaction I was able to have with you most definitely helped me maintain my sanity. You have been with me through the ups and downs of my research and studies and through this I have seen just how much of my rock you really are. I look forward to growing old with you.

I owe an enormous amount of gratitude to Virginia Commonwealth University and the people who define this great institution. Dr. Best and Dr. Ramesh, thank you for directing me in my research. Without you and your incredible minds, I would not have been able to submit this work. You both devoted countless hours and amounts of energy to this research. We all started knowing very little about this topic, and I am so excited with the result. I really look forward to continuing our research together.

Along with my dissertation advisors, I would like to thank Dr. Wetzel and Dr. Pickler for guiding me in my research of heart rate variability. I look forward to future applications of the model described in this research with the heart rate data we have spent so much time preparing.

I would like to thank the Department of Biostatistics for generously funding my education and teaching me statistics. My experience with our department has been extremely positive. Yvonne and Denise, thank you so much for everything. Beyond the obvious help, you were always there to talk to. To all the professors, thank you for guiding me through my studies. Your doors were always open, and you were all eager to offer assistance and advice.

Finally, Dr. Elswick, thank you for being my friend, my mentor, my advisor, and my teacher. The statistical knowledge and consulting experience I have gained from you over the past five years is priceless.

In his poem, *Ulysses*, Alfred Lord Tennyson writes:

I am a part of all that I have met;
Yet all experience is an arch wherethro'
Gleams that untravell'd world, whose margin fades
For ever and for ever when I move.
How dull it is to pause, to make an end,
To rust unburnish'd, not to shine in use!
As tho' to breathe were life. Life piled on life
Were all too little, and of one to me
Little remains: but every hour is saved
From that eternal silence, something more,
A bringer of new things; and vile it were
For some three suns to store and hoard myself,
And this gray spirit yearning in desire
To follow knowledge like a sinking star,
Beyond the utmost bound of human thought.

These words speak volumes about my own life experiences. Through my travels and from all the faces I have met and learned from throughout my life, I have found myself. I am a part of all of these people. As I age, I will grow with each new face and experience I encounter. The research from this dissertation has opened up many possible doors for me, teaching me how much knowledge is still left undiscovered. The production of this work has inspired me to dedicate my life to education, not only as a teacher, but as an eternal student of science.

Table of Contents

Page

List of Tables

Page

List of Figures

Page

Abstract

A NORMAL-MIXTURE MODEL WITH RANDOM-EFFECTS

FOR RR-INTERVAL DATA

By Jessica McKinney Ketchum, Ph.D.

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2006

Major Directors

Viswanathan Ramakrishnan

Associate Professor Department of Biostatistics

Al M. Best

Associate Professor Department of Biostatistics

In many applications of random-effects models to longitudinal data, such as heart rate variability (HRV) data, a normal-mixture distribution seems to be more appropriate than

the normal distribution assumption. While the random-effects methodology is well developed for several distributions in the exponential family, the case of the normal-mixture has not been dealt with adequately in the literature. The models and the estimation methods that have been proposed in the past assume the conditional model (fixing the random-effects) to be normal and allow a mixture distribution for the random effects (Xu and Hedeker, 2001, Xu, 1995). The methods proposed in this dissertation assume the conditional model to be a normal-mixture while the random-effects are assumed to be normal. This is primarily to fit the HRV data, which seems to follow a normal-mixture within subjects. Another advantage of this model is that the estimation becomes much simpler through the use of an EM-algorithm. Existing methods and software such as the PROC MIXED in SAS are exploited to facilitate the estimation procedure.

A simulation study is performed to examine the properties of the random-effects model with normal-mixture distribution and the estimation of the parameters using the EM-algorithm. The study shows that the estimates have similar properties to the usual normal random-effects models. The between subject variance parameter seems to require larger numbers of subjects to achieve reasonable accuracy, which is typical in all random-effects models.

The HRV data is used to illustrate the random-effects normal-mixture method. These data consist of 9 subjects who completed a series of five speech tasks (Cacioppo et al., 2002). For each of the tasks, a series of RR-intervals was collected during baseline, preparation, and delivery periods. Information about their age and gender were also available. The random-effects mixture model presented in this dissertation treats the

subjects as random and models age, gender, task, type, and task × type as fixed-effects. The analysis leads to the conclusion that all the fixed effects are statistically significant. The model further indicates a two-component normal-mixture with the same mixture proportion across individuals fit the data adequately.

1 Introduction

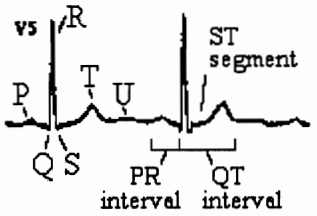
1.1. Introduction to Heart Rate Variability

The human body is a complex structure made up of several systems (respiratory, circulatory, nervous, etc.) which are all designed to function simultaneously. The body must maintain a normal balance among its internal physiological conditions in order to function properly. This balance is referred to as homeostasis. One of the systems that must be controlled and balanced is the heart. The ability of the heart rate to vary is an important indication of the normal homeostatic mechanisms of the cardiovascular system. A person with high variability in their heart rate shows signs of good adaptability and can be considered a healthy individual with well-functioning autonomic control mechanisms. However, a person with low variability in their heart rate often shows signs of poor adaptability of the autonomic nervous system which can be indicative of a variety of conditions (Pumprla et al., 2002).

The measuring of heart rate variability (HRV) is a well studied field dating back to the 1960's. Measuring HRV is noninvasive and simple to acquire, and has the property of being reproducible under standardized conditions (Kleiger et al., 1991). Several different approaches to measuring and analyzing HRV have been used and are reviewed in section 1.3. The term HRV has been associated with describing variations in

the instantaneous heart rate as well as variations in the RR-interval lengths (Task Force, 1996). The RR-interval quantifies the variations in the heart rate by measuring the time, typically in milliseconds (ms), between successive peaks, denoted by R in Figure 1 (Jenkins and Gerred, 1996).

Figure 1: Normal Sinus Rhythm



Although other possible measures include using the Q or S peaks, the R peak is most often utilized because it is easier to measure with precision. The choice of peak depends on the types and placements of the leads used to record the ECG signal. Once a consistently detectible peak is identified, an accurate measure of the time between these peaks can be determined. Other terms used within the literature to refer to these intervals include the interbeat interval (IBI) or the heart period (HP). The term RR-interval will be used throughout this dissertation.

The study of HRV is common in both research and clinical studies. In clinical settings, HRV can be used as a tool to monitor the influence of medicine on a variety of subject populations. Visualizing how the heart rate varies is of great interest to clinicians because it lends insight into treatment. The importance of accurately characterizing HRV in different cohorts of patients can have dramatic results in a clinical setting. There has been considerable research into understanding the variations in heart rates within

different cohorts of diseases, including hypertension, congestive heart failure, heart transplantation, sudden death, ventricular arrhythmias, and patients with renal failure. Refer to Task Force (1996) and van Ravenswaaij-Arts et al. (1993) for a thorough summary of the current research involving measures of HRV. Most significantly, research has shown that a reduction of HRV in adult patients can be used as a predictor of risk for acute myocardial infarction (Kleiger et al., 1987 and Odemuyiwa et al., 1991) and as an early indicator of diabetic neuropathy (Malpas and Maling, 1990). Within the fetal and neonatal population, it has been observed that acute hypoxia has been associated with in an increase in HRV (Thaler et al., 1985) and chronic hypoxemia was associated with decreased HRV (Smith et al., 1988 and Ribbert et al., 1991).

1.2. Motivation

Although there has been a great deal of progress made in the field of HRV, it has mainly been in the direction of measuring HRV, while little emphasis has been placed on modeling HRV. The statistical measures that are used to describe HRV are primarily summary statistics. However, the variability in heart rate cannot be completely characterized by these measures alone. The underlying distributional properties of the data must also be explored and understood. The use of modeling in statistics is a valuable tool that can yield a tremendous amount of information to the researcher. However, certain assumptions about the data must be met in order to obtain valid statistical results. For example, the use of the mean as a measure of central tendency is only reasonable if the underlying distribution is symmetric.

Motivation for this dissertation comes from two studies described briefly here. In both studies, heart rates from a sample of several subjects were recorded over varying intervals of time. In the first study, these subjects are a sample of sedated adult hospital inpatients whose heart rates were recorded for periods of up to 24 hours (Grap et al., 2006). This particular data set contains one ECG recording per subject. In the second study, these subjects are a random sample of preterm infants whose heart rates were recorded during several bottle feeding sessions lasting up to 15 minutes (Pickler et al., 2006, Pickler et al., 2006, Pickler et al., 2005). In contrast to the first study, this data set contains multiple recordings for each subject, one for each bottle feeding session.

It is of interest to model the variability in heart rates for the recordings from a sample of several subjects using RR-interval data in terms of the fixed-effects (age, gender, treatment, etc.), while accounting for the random-effects (subjects, visits, etc.). When analyzing either of these data sets, an appropriate model should be used to model the correlations structures of the RR-interval data within subjects by incorporating a random subject effects. In the case of the infant data, the correlation structure of the RR-interval data within each bottle feeding session could also be should also be considered by incorporating a random visit (bottle feeding session) effects. An obvious choice for analyzing such data is the mixed-effects generalized linear models (Laird and Ware, 1982). However, the distribution of the RR-interval has been shown to follow a normal-mixture distribution (Mandrekar et al., 2005; Eckberg, 1983; and Nagaraja et al., 1995). While methodology for mixed-effects models is well developed for a variety of distributions in the exponential family, the case of the normal-mixture has not been

thoroughly considered. Xu and Hedeker (2001) do address a case where the random-effects are assumed to be normal-mixture and the resulting conditional model (conditioned on the random-effects) to be normal. With regards to fitting the HRV data described here, however, the random-effects should be assumed to be normal and the resulting conditional model then assumed to be a normal-mixture, since the series RR-interval data within each subject (and bottle feeding session in the infant data set) is distributed with the normal-mixture distribution and not the subjects themselves. Without the methodology for fitting mixed-effects models to RR-interval data arising from normal-mixture distributions, researchers have been limited and forced to deal with the inherent correlation structure of the data in other ways. These methods are summarized in section 1.6.

The research presented in this dissertation provides the foundations for the methodology used to estimate parameters in the mixed-effects models when the conditional data arise from a mixture of normal densities, in particular when modeling HRV data. Before further discussion regarding modeling HRV data, the various measures of HRV and methods used to acquire them are detailed.

1.3. Measuring Heart Rate Variability

Heart rate variability has been used describe variations in both the instantaneous heart rate and the RR-intervals. Measurements of variations in heart rate can be made in a number of ways. These methods have primarily been classified as time domain measurements or frequency domain measurements. In this section, these methods will

briefly be described in accordance with the standards recommended by the Task Force of the European Society of Cardiology and The North American Society of Pacing and Electrophysiology (1996).

1.3.1 Time Domain Methods

Time domain methods are used to summarize the variations in heart rate. They are often utilized because of their simplicity. These methods involve determining either the series of instantaneous heart rates or the series of intervals lengths between successive R peaks. Instantaneous heart rate is typically measured in beats per minute (BPM) and RR-interval length in milliseconds (ms). There is an inverse relationship between RR-interval length and instantaneous heart rate:

$$\text{RR-interval length (ms)} = \frac{60000}{\text{heart rate (BPM)}}.$$

For example, a heart rate of 100 beats per minute (BPM) corresponds to an RR-interval length of 600 ms or 0.6 seconds. There has been some research into which of these two measures should be used when studying variations in heart rate (Khachaturian et al., 1972 and Graham and Jackson, 1970). In 1974, Jennings, Stringfellow, and Graham assessed the normality of (instantaneous) heart rate and heart period (RR-interval) distributions in adult subjects. They found that neither heart period nor heart rate measures were consistently normally distributed. In the 10 samples they examined, 60% of the heart period samples and 100% of the heart rate samples failed to pass normality tests. Thus, they advocated the use of heart period data rather than heart rate because it is "more" normally distributed. This argument is not very convincing as to choosing either of these

measures when restricted to the assumption of normality. A better choice perhaps is to consider a model which allows for the true nature of the distribution to be assumed.

Typical variables calculated with time domain methods are the mean RR-interval, the mean heart rate, the difference between the longest and shortest RR-interval (range), and the difference between night and day heart rate. A common measurement made over long recording periods is the SDNN. This variable represents the standard deviation of all RR-intervals over the entire recording period and is a measure of the variance of the RR-intervals, representing all the cyclic components responsible for variability in the recording period. SDNN is not a well defined statistical quantity because it depends on the length of the recording period (Task Force. 1996). As the length of recording increases, the total variance will also increase. Thus, it is inappropriate to compare the SDNN values from recording periods of different lengths without standardization.

When looking at segments of a recording period, several other variables have been considered, including the SDAAN and SDNN index. The SDAAN is the standard deviation of the average RR-intervals calculated over short periods, usually 5 minutes, from a longer recording, usually 24 hours. This measure estimates changes in heart rate due to cycles longer than 5 minutes. The SDNN index is the average of all standard deviations of the RR-intervals calculated from the 5 minute segments of the entire 24 hour recording. This measure estimates changes in heart rate due to cycles shorter than 5 minutes.

Often times, differences in successive RR-intervals are examined. In these cases the most commonly used measures include the RMSSD, NN50, and p-NN50. The

RMSSD represents the square root of the mean squared differences of successive RR-intervals. The NN50 is defined as the number of interval differences of successive RR-intervals greater than 50 ms, while the p-NN50 is the proportion derived from dividing NN50 by the total number of RR-intervals.

Many of the statistical time domain measurements are highly correlated, thus it is strongly recommended (Task Force, 1996) to use the following three:

- SDNN (estimate of overall HRV);
- SDANN (estimate of long-term components of HRV); and
- RMSSD (estimate of short-term components of HRV).

1.3.2 Frequency Domain Methods

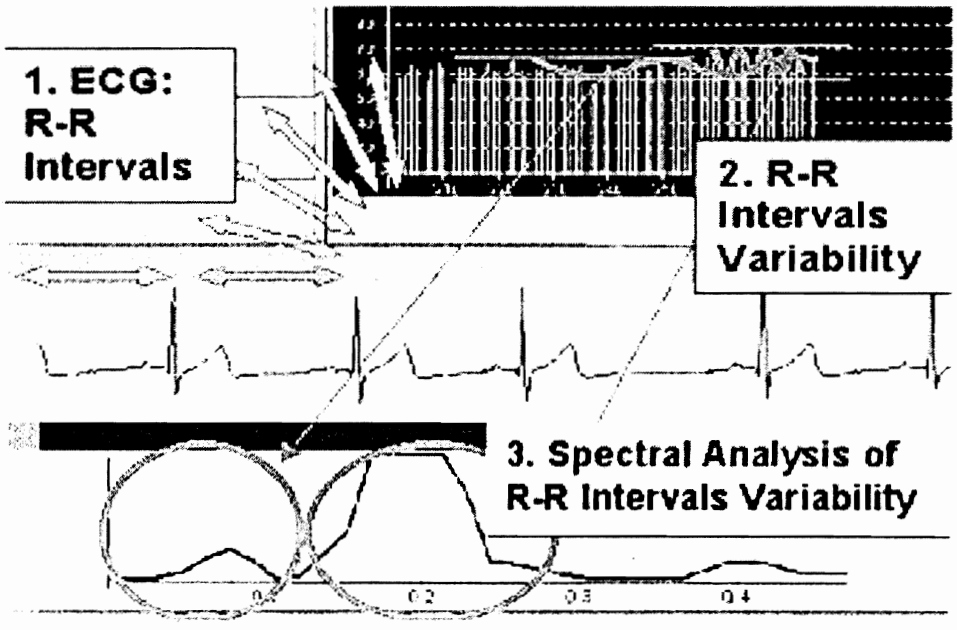
An alternative method for measuring HRV involves the use of frequency domain or power spectrum methods. Time domain methods make it difficult to obtain precise physiological data about changes in autonomic function (Pumprla, et al., 2002). This has led some researchers to focus on the cyclical changes in HRV. Frequency domain based methods are used to explore how power, or variance, is distributed as a function of frequency.

The heart rate is primarily influenced by two factors, "the intrinsic firing rate of the automatic (pacemaker) cells of the sinoatrial node" and "the modulating influences of the autonomic nervous system" (Őri et al., 1992). There are several modulating neuroregulatory influences on heart rate such as respiration, blood pressure, cardiac output, thermoregulation, and the renin-angiotensin system. These influences affect the

sinus node through the autonomic nervous system. The two main subsystems of the nervous system are the sympathetic and the parasympathetic systems, both of which supply the sinoatrial node. The sympathetic system, which enhances spontaneous firing rate, increases the heart rate, while the parasympathetic system exerts a counter inhibitory action, which slows the heart rate. The balance of these two subsystems is most likely the "principal determinant of the normal heart rate" (Őri et al., 1992). These two subsystems operate at different frequencies, thus variations in heart rate relating to each of the subsystems can be identified and quantified by use of frequency domain methods.

In order to perform a frequency domain analysis of HRV a series of consecutive RR-intervals is first determined. The intervals must be free of artifacts, such as missed or spurious R-waves. These consecutive time intervals can be treated as if they are equally spaced (van Ravenswaaij-Arts et al., 1993). A tachogram is then produced through the steps described below and demonstrated with Figure 2 (Pumprla et al., 2002).

Figure 2: Sample Tachogram



First the ECG signal is obtained and the RR-intervals are determined (middle portion of Figure 2). Next, the series of consecutive RR-intervals are plotted consecutively, with the length of the interval (in ms) on the vertical axis and interval number on the horizontal axis (upper portion of Figure 2). A curve passing through the peaks of the intervals is then drawn and represents the variation in heart rate over time. This signal (curve) can be analyzed using time series methods (lower portion of Figure 2).

Time series methods require that the signal be periodic and stationary (Penaz, 1978). A signal is periodic if it repeats itself exactly after some period or cycle, just as an ECG signal does. With HRV data the requirement is that the data be at least weakly stationary (Weber et al., 1992), in that only the first (mean) and second (variance) moments are required to be independent of time. Violations of this assumption can have serious consequences for statistical tests when pooling across time. The signal can be made stationary by using a detrending procedure such as subtracting a least-squares polynomial approximation from the original signal or by high-pass filtering (Penaz, 1978). The periodic and stationary signal can then be analyzed by using time series methods such as the Fast Fourier transform.

The Fast Fourier transform decomposes the power spectrum into its frequency components and quantifies them in terms of their relative intensity. For healthy adult subjects, the power spectrum of short term recording periods (2-5 minutes) consists of three main frequency components, high frequency (HF) bands ranging between 0.15 Hz and 0.4 Hz, low frequency (LF) bands ranging between 0.04 Hz and 0.15Hz, and very low frequency (VLF) bands ranging between 0 and 0.04 Hz. In longer recording periods

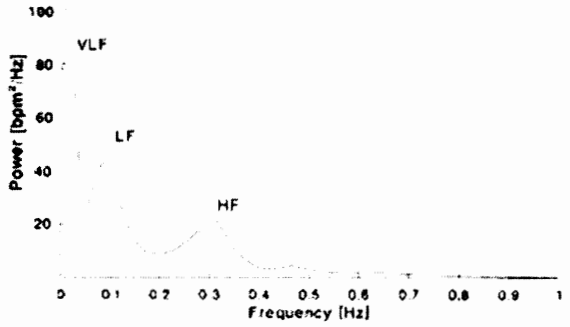
(24 hours) an ultra-low frequency (ULF) band has been identified in the range between 0 and 0.003 Hz. The VLF band is then represented in the range of 0.003 Hz and 0.04 Hz (Task Force, 1996). The band limits for these frequency components have been reported to vary somewhat from the bands described above, within different cohorts of subjects.

The HF component has been associated with parasympathetic activity and is represented primarily by variations in respiration. Cyclical changes in heart rate associated with respiration typically occur between 0.25 and 0.35 Hz in normal adults (Fallen 1988, Pumprla et al., 2002, Pagani et al., 1986). The LF component is associated with both parasympathetic and sympathetic activity (Stein et al., 1994 and Pumprla et al., 2002) and is thought to be primarily affected (due to changes in blood pressure) by cyclical variation in the baroreceptor system (Kamath et al., 1987, Sayers, 1973, and Pomeranz et al., 1985). This type of variation typically occurs around 0.1 Hz. The VLF and ULF bands are less understood, but have been associated with thermoregulatory processes (Fleisher et al., 1996 and Fallen et al., 1988), peripheral vasomotor (chemoreceptor) activity (Fallen et al., 1988, Ponikowski et al., 1997, and Francis et al., 2000), and the rennin-angiotensin system (Duprez et al., 1995, Taylor et al., 1998, and Akselrod et al., 1981).

The Fast Fourier transform (FFT) produces a power spectral curve decomposing the power spectrum into its frequency components. An example of the FFT for HRV data is shown in Figure 3 (van Ravenswaaij-Arts et al., 1993). The vertical axis is used to represent power, also termed variance, and the horizontal axis displays frequency in hertz (HZ). The units for power are typically in squared seconds per hertz, s^2/Hz, or

equivalently, using milliseconds, $ms^2 \times 10^{-6}$ Hz. The power units in Figure 3 are shown in squared beats per minute per hertz, BPM^2/Hz since the tachogram was produced with instantaneous heart rates rather than RR-intervals. The VLF, LF, and HF peaks have been identified in this Figure 3 to indicate the separate components of the power spectral density.

Figure 3: Sample Power Spectrum



The total area under the power spectral curve is defined as total power. The area under the curve within a particular band is defined as component power. The ratio of the power of the individual components to the total power, minus the VLF component, is termed normalized power. Expressing the HF and LF components in normalized units "emphasize[s] the controlled and balanced behaviors of the two branches of the autonomic nervous system" and minimizes "the effect on the values of LF and HF components of the change in the total power" (Task Force, 1996). Fractional power is defined as the percent of the total power expressed by the individual components (Öri et al., 1992). The ratio of LF to HF (LF/HF) has been used as a measure of sympathovagal balance or sympathetic modulation (Task Force, 1996).

The time and frequency domain measures taken from a 24 hour ECG recording are highly correlated ($r > 0.85$) (Kleiger et al., 1991). Total power and the square of SDNN are virtually equivalent since both are measures of the amount of total variability in the heart rate signal. The HF measures correlate strongly with RMSSD and p-NN50 while LF and VLF measures correlate strongly with SDNN index. ULF measures correlate well with SDNN and SDANN.

1.4. HRV Data Acquisition

In order to obtain interpretable measures of HRV, the ECG signal must be collected with accuracy. Today ECG signals are obtained by converting analog signals to digital signals with computers through a process termed digitization. The sampling rate is the most important issue concerning the acquisition of the ECG signal. There are a finite number of samples that can be obtained during any specified interval of time. Thus errors are introduced when the signal is ECG signal is sampled and digitized. Choice of the sampling frequency typically depends on the ECG bandwidth and on the type of analysis of interest. A high enough sampling rate must be chosen so that the frequencies of interest are identifiable, typically twice the highest frequency, while minimizing computer memory issues. The suggested sampling rate is between 500 and 1000 Hz (samples per second), providing a basic resolution of 1 – 2 milliseconds (Merri et al., 1990 and Riniolo and Porges, 1997).

Once the signal has been digitized, the RR-intervals can be identified. Besides issues regarding the sampling frequency, other errors can be introduced into the signal.

Noise in the signal may result in spurious peaks in the signal, making R-peak identification algorithms problematic. The algorithm may identify spurious peaks as R-peak, if the amplitudes are large enough, thus resulting in spurious RR-intervals. Interruptions in the signal, even if in small intervals of time, disrupt the continuity of the time series which can have implications on the assumptions in the analyses. In any ECG recording it is nearly impossible to obtain a signal free of noise and without interruptions. In order to account for this, much research has been conducted regarding preliminary processing of the RR-interval series.

Several types of RR-interval detection techniques are available ranging from simple smoothing or filtering procedures of the digitized data to more complex methods utilizing derivative or template matching techniques (Friesen et al, 1990). Some progress has also been made utilizing wavelet transformation techniques to reduce the noise in the ECG signals (May et al., 1997 and Gyaw and Ray, 1994). The progress with wavelet transforms is significant in that methods unitizing wavelet analyses do not require the signals to be stationary, where as methods such as the FFT do require this property. Comparison between the results obtained using FFT and wavelet transforms for analysis of HRV data have been compared and minimal differences noted (Pichot et al., 1999).

After RR-interval identification, the series must be examined for artifacts. Artifacts in the RR-interval series are typically due to missed or spurious beats. The number of artifacts that may be obtained in a series is highly dependent on the length of the signal. Thus, considerable attention has been given to obtaining methods for identifying artifacts (Malik et al., 1989, Berntson et al., 1990, and Weber et al., 1992).

When spurious beats occur, typically due to noise in the signal, no information regarding the signal is lost. The effect of a spurious beat results in two identified (artifact) RR-intervals, whose sum is the correct RR-interval. In contrast to spurious beats, when heart beats are missed information regarding the signal is lost. The result is an artifact RR-interval with an excessive length, being the combination of two or more RR-intervals. When the number of identified artifacts is minimal, the original ECG recorded can be examined to determine the correct RR-interval lengths surrounding the artifact RR-intervals. When the number of identified artifacts is large other methods have been suggested, such as interpolating the missing intervals with using surrounding intervals or splitting the excessively large intervals into two or more shorter intervals (Berntson et al, 1997). Without resolution of these artifacts, significant biases are likely to be obtained and be misleading (Berntson and Stowell, 1998, Malik and Camm, 1993, Ramanathan and Myers, 1996 and Xia et al., 1993).

While many of the methods discussed thus far are not employed in this dissertation, they are provided for completion and to illustrate the complexity of the nature of HRV data. Next, the data set used for applications in Chapter 4 is introduced. Then, in the remaining sections of this chapter, the modeling aspect of HRV is discussed.

1.5. Heart Rate Variability in the Loneliness Study

The data used for illustration in this dissertation is a subset of nine healthy adults selected randomly from 45 male and 44 female undergraduates who participated in a loneliness study at the General Research Center at the Ohio State University (Cacioppo et al., 2002).

This data subset will be referred to as the Loneliness data from this point forward. The same subset is used in Mandrekar's dissertation (2002) in part to illustrate the significance of methods involving normal-mixture distributions, rather than unimodal normal distributions, to estimate the parameters of the distribution for RR-interval data. Her methodology, however, requires each minute from each subject to be estimated with separate models. Although this can lead to decent estimates for the mixture proportion, component means, and component variances, it lacks the ability to use the information (data) from the other subjects and even other minutes for the same subjects to obtain more precise estimates. Random-effects models, on the other hand, allow the observations within and between subjects to take on specific variance-covariance structures, which allow the estimates of the model to be obtained using all of the data, across all subjects. The methodology proposed in this dissertation extends that suggested by Mandrekar by using random-effects methodology within an EM algorithm in order to estimate parameters for a model containing all the minutes from all of the subjects.

During the study, several cardiovascular and psychosocial measures were recorded on each of the subjects enrolled. ECG was recorded using the standard lead II configuration with Ag / AgCI disposable electrodes (Portrace 9113) and the signal quality was ensured after the attachment of the sensors and a 15 minute adaptation period had elapsed (Mandrekar, 2002). During the ECG recording period each subject performed six psychosocial stressor tasks and one orthostasis stressor task. The psychosocial stressor tasks included five speech related tasks and one verbal mental arithmetic task (assigned randomly to each subject). The mental arithmetic task (MA) had 4 minutes of baseline

and 4 minutes of task. The "Saab" (BS) speech task had 4 minutes of baseline, 2 minutes of preparation, and 2 minutes of delivery. The remaining 4 speech tasks, why I'm likeable (LS), ask for date (AS), describe way to school (WS), and describe inanimate objects in room (IS), each had 2 minutes of baseline, 2 minutes of preparation, and 2 minutes of delivery. The orthostasis stressor task (OS) had 8 minutes of task. Thus, a total of 48 minutes was collected for each subject. The BS task was always the last task performed with the OS task immediately before it. Table 1 displays the demographic data for each subject and the order in which the stressor tasks were preformed.

Table 1: Task Orders and Demographics

Subject	Task Order	Gender	Age
1	WS, MA, LS, AS, IS, OS, BS	Female	19
2	AS, IS, WS, LS, MA, OS, BS	Female	19
3	MA, LS, IS, WS, AS, OS, BS	Male	19
4	AS, WS, MA, IS, LS, OS, BS	Male	24
5	LS, WS, IS, AS, MA, OS, BS	Female	19
6	LS, AS, IS, WS, MA, OS, BS	Female	18
7	WS, LS, AS, MA, IS, OS, BS	Female	19
8	MA, AS, IS, LS, WS, OS, BS	Male	18
9	WS, IS, AS, MA, LS, OS, BS	Male	18

Each analog ECG signal, for every minute, and from every subject, was first acquired at 1000 Hz then decimated to 500 Hz. The R-peaks were identified using waveform matching templates and then a time/amplitude criterion. Once the R-peaks have been identified, the RR-interval series was calculated using the distance between successive R-peaks. After collection, the ECG data was carefully edited and checked for any artifacts. The algorithm used for artifact identification is described in Chapter 4.

Mandrekar (2002) fit a normal distribution to each minute of ECG data in the Loneliness study, among the various tasks and subjects, and noted that a single normal distribution provides a marginal fit to the RR-interval data in about 50% of the cases, and a two-component normal-mixture model seemed more appropriate. Her research proposed fitting a two-component normal-mixture model to each minute of data and showed promising results. However, as discussed earlier, her methodology lacks the ability to account for specific correlation inherent in the data. In the next section, the current progress towards modeling HRV is reviewed.

1.6. Modeling Heart Rate Variability

In section 1.5, it was suggested that the distributions of the minutes of RR-interval data, taken from a single subject appear to come from a two-component mixture distribution. To date, most methodology has modeled data from each subject independently of the other subjects in the studies. This is less desirable to a model containing all the subjects for several reasons. First, fitting several models requires a considerable amount of time when there are a large number of subjects in the data sets, or when there are several

recordings taken from each subject over a study period. In addition, the modeling within subjects and across several recordings assumes that these data are independent when there is an inherent within subject correlation structure. Furthermore, subjects are known to vary considerably with regards to HRV, a single model would be able to quantify this variability for researchers.

Although the past 45 years have seen a great progress in the measures of HRV, little has been made directly modeling heart rate data. In order to better understand the variations in heart rate, the underlying distribution of RR-interval data must be further explored.

1.6.1 Distributional Properties of Heart Rate Variability

HRV measures are typically assumed to follow a normal distribution. Several studies, however, have reported that RR-interval data do not necessarily follow a normal distribution (Nagaraja et al., 1995, Jennings et al., 1974, Hashida at al., 1973, Eckberg et al., 1983, and Riniolo and Porges, 2000). An incorrect assumption for the underlying distribution of the data can lead to invalid results in statistical analysis. Some progress has been made in determining an appropriate distribution for the data in order to accurately estimate model parameters. Hashida et al. (1973) commented on the irregularity in the shapes of the histograms of RR-interval data taken from different subjects in atrial fibrillation, or absolute arrhythmia. Studies before them (Jordan, 1954 and Horan and Kistler, 1961) had shown varying shapes ranging from a "normal distribution, unimodal and bimodal distribution tailing to the right, and less commonly

high plateau-like or long, low, flattened distributions" (Hashida et al., 1973). They determined that the most appropriate fit to the histograms of their interval data was an Erlangian distribution. Around the same time, Jennings et al. (1974) noted that RR-interval data failed to satisfy the normality assumption in 60% of their samples of young adults. In young healthy subjects, Eckberg (1983) observed that as controlled breathing rates decreased, the bimodal pattern in the RR-interval data became more apparent. Nagaraja (1995) examined the distribution of RR-intervals in 10 CHF patients and noted previously unrecognized bimodal patterns in HRV at nighttime. Mandrekar and Nagaraja (2005) showed that the absolute values of the successive differences in RR-interval data approximately follow Weibull distributions. The methods proposed in their work suggest that the use of a Weibull distribution may also improve artifact detection algorithms. All of these studies agree that the data is not normally distributed and many seem to suggest a bimodal pattern in both healthy and heart diseased subjects. It is not known at this time why the RR-intervals exhibit a bimodal pattern. However, some speculation and research has been made. Eckberg (1983) suggested that respiratory sinus arrhythmia (RSA) may be the physiological explanation for the bimodal pattern. Heart period shortens during the inspiratory phase and lengthens during the expiratory phase of a breathing cycle.

Often times, the most appropriate models for data with bimodal distributions are mixture models (not to be confused with mixed-effects models). This is due to the flexibility in the shape of the distribution across the range of values for its parameters. Nagaraja (1995) modeled RR-interval data with a normal-mixture distribution and Mandrekar (2002) describes in detail the methodology used to fit normal-mixture

distributions with two components to RR-interval data from 9 healthy young adults. It is not difficult to argue that data across subjects are independent. However, independence in the data within subjects needs to be further examined. In this dissertation, the methodology for modeling normal-mixtures used by Mandrekar is extended to include fitting appropriate correlation structures to the data. These extensions focus on combining group level information in order to obtain more efficient estimates of the parameters of the distribution.

In comparison to fixed-effects models, mixed-effects models are able to model data in which the observations are not independent. This is accomplished by allowing for a combination of fixed and random-effects in the model. The mixed-effects model is defined in detail in section 2.2. In order to distinguish between the term mixed, used in references to mixed-effects models, and the term mixture, used to describe the distribution, mixed-effects models will be referred to as random-effects models from this point forward. In this dissertation, random-effects models allow for a combination of fixed and random-effects.

1.6.2 Random-Effects under the assumption of Mixture Distributions

Although much progress has been made in the field of mixture distributions, as well as fitting mixed-effects models, little has been published concerning fitting random-effects models under the assumption of a mixture distribution. Belin and Rubin (1995) published a paper analyzing repeated-measures data on schizophrenic reaction times by modeling the response times for and within each schizophrenic individual as a two-

component mixture distribution. Around the same time, Xu and Hedeker (2001) explored mixtures in random-effects regression models while Verbeke and Lesaffre (1996) fit a linear mixed-effects model with heterogeneity in the random-effects population. Although it is tempting to use the modeling techniques suggested by these previously mention authors to fit our data, several significant issues need to be addressed. The methodology proposed by these authors is not clearly detailed in their work, making application difficult. Second, these methods assume that the random-effects follow a mixture distribution, and the resulting conditional density (conditioned on the random-effects) to be normally distributed (Verbeke et al., 2001). The response variables for their models then follow a complex type of mixture distribution. These require complicated methodology, involving a combination of the EM algorithm and the Fisher Scoring algorithm, to obtain estimates for the parameters of the model. For this research, the opposite is assumed. Here, the random-effects are assumed to be normally distributed, and the resulting conditional distribution (conditioned on the random effects) is then a normal-mixture distribution. We would like the errors to be distributed as a normal-mixture, while the random-effects remain normally distributed. This difference here is that the subjects are not assumed to be grouped, but rather the RR-intervals within the recordings from each subject. A model with these assumptions then leads to a much more elegant algorithm used for obtaining parameter estimates for the model as will be shown in Chapter 2.

In this dissertation, a method is proposed for modeling RR-interval data without ignoring the true nature of their distribution, yet still accounting for more appropriate

correlation structures between and within the subjects. Theoretical and methodological details for the proposed model will be outlined in chapter 2. In chapter 3, the model will be tested with several simulated data sets, and then used to fit the Loneliness data in Chapter 4. Future research involving the normal-mixture random-effects model will be discussed in Chapter 5.

2 Methodology

2.1. Introduction

In this chapter, a random-effects model is defined where the residuals are assumed to follow an M-component normal-mixture distribution and methods are outlined for maximum-likelihood estimation, using an Expectation-Maximization (EM) algorithm. In section 2.2, the normal random-effects model is defined and the notation used throughout this dissertation is introduced. The normal "random-effects" models, as defined in this section, are more commonly referred to as normal "mixed-effects" models, and contain both fixed and random-effects. In order to avoid confusion with the terms "mixed" and "mixture", the "normal-mixed model" is referred to as the "normal random-effects model" in this dissertation. In section 2.3, the two-component joint normal-mixture random-effects model is defined. The model is extended to include M-components in Section 2.4. Then, in section 2.5, the distributional properties of the M-component joint normal-mixture random-effects model are discussed. A brief review of estimation in normal random-effects models is covered in Sections 2.6. In section 2.7.1, maximum likelihood estimation, using the EM-algorithm, for the parameters of the M-component normal-mixture model is reviewed. The EM-algorithm is then extended, in section 2.7.2, to include fixed-effects in the M-component joint normal-mixture model. This will be accomplished by defining a set of weights for each component of the mixture, weighting the observations by these component weights, and then applying standard maximum-likelihood techniques to obtain parameter estimates of the M likelihoods, given the weighted observations. The weights introduced in this section provide a way of reducing

the problem of estimation in the M-component normal-mixture model to M separate estimations of joint normal models. The EM-algorithm is then further extended, in section 2.8, to incorporate random-effects into the M-component normal-mixture model by proportionately assigning the random-effects to the M components and fitting M separate random-effects models. The M estimates of the random-effects produced by these models are then combined to obtain the overall variance estimates of the random-effects. In section 2.9, methods for testing the significance of the model parameters are described. The chapter is concluded in section 2.10 with a general discussion of the M-component joint normal-mixture random-effects model, including its benefits and limitations.

2.2. Normal Random-Effects Model Definition

The normal random-effects model introduced by Laird and Ware (1982) can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.1}$$

where

$\mathbf{y}$ is the $N \times 1$ vector of observed response values,

$\boldsymbol{\alpha}$ is the $P \times 1$ vector of fixed-effects parameters,

$\mathbf{X}$ is the $N \times P$ observed design matrix corresponding to the fixed-effects,

$\boldsymbol{\beta}$ is the $Q \times 1$ vector of random-effects parameters,

$\mathbf{Z}$ is the $N \times Q$ observed design matrix corresponding to the random-effects, and

$\boldsymbol{\varepsilon}$ is the $N \times 1$ vector of residuals.

This model assumes that β has a Q-variate normal density with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{G}$. Using the symbol ~ for the phrase "is distributed as" and a subscript to indicate the dimensions of the multivariate normal density, this can be denoted with $\beta \sim N_Q(\mathbf{0}, \mathbf{G})$. The vector of random-effects, β, is assumed to be independent of the vector of residuals, ε, which is assumed to be follow an N-variate normal density with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{\Sigma}$. That is, $\varepsilon \sim N_N(\mathbf{0}, \mathbf{\Sigma})$. Under these assumptions, $\mathbf{y}$ then follows an N-variate normal density with mean vector $\mathbf{X}\alpha$ and variance-covariance matrix $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{\Sigma}$. That is, $\mathbf{y}$ is a vector of observations taken from $N_N(\mathbf{X}\alpha, \mathbf{V})$.

The model in equation (2.1) can be equivalently expressed as $\mathbf{y} = \mathbf{Z}\beta + \varepsilon$, by assuming the residuals have mean vector $\mathbf{X}\alpha$ rather than $\mathbf{0}$. That is, $\varepsilon \sim N_N(\mathbf{X}\alpha, \mathbf{\Sigma})$. This is the first step in redefining the normal random-effects model to incorporate the assumption of a mixture distribution for the residuals. In this chapter, the normal random-effects model, described here, is extended to a random-effects model in which the errors are assumed to have an M-component joint normal-mixture density. This will be done in the next section for the two-component case and then extended to include M-components in section 2.4.

2.3. Two-Component Joint Normal-Mixture Random-Effects Model Definition

A two-component joint normal-mixture random-effects model can be defined as

$$\mathbf{y} = \mathbf{Z}\beta + \varepsilon, \tag{2.2}$$

where $\mathbf{y}$, $\mathbf{Z}$, β, and ε are defined as in equation (2.1). Here, the random-effects considered are random subject-effects for a model containing S subjects. It is assumed that the vector of random subject effects, β, is $N_S(\mathbf{0},\mathbf{G})$. For this model definition, ε is assumed to follow a joint normal-mixture density with two $N \times 1$ component mean vectors, $\mathbf{X}\alpha_1$ and $\mathbf{X}\alpha_2$, two $N \times N$ component variance-covariance matrices, Σ_1 and Σ_2, and an $S \times 1$ vector of mixture proportion parameters, λ. Here, $\mathbf{X}$ is the fixed effect design matrix defined in(2.1), and the α_k are $P \times 1$ vectors of fixed-effect parameters. This will be denoted

$$\varepsilon \sim JMX(\mathbf{X}\alpha_1, \mathbf{X}\alpha_2, \Sigma_1, \Sigma_2, \lambda). \tag{2.3}$$

The vector of mixture proportion parameters λ contains the S subject proportions of the first normal distribution and the vector $(1-\lambda)$ contains the S subject proportions of the second normal distribution, in the joint mixture distribution. It is important to understand that the joint density of the residuals, ε, is a product (multiplication) of normal-mixture densities and not a mixture of multivariate normal densities. The joint density of the residuals will be explicitly defined within this section.

Suppose that N_j observations, $j = 1,...,S$, are taken from S subjects with $N = \sum_{j=1}^{S} N_j$. A two-component normal-mixture random-effects model can be specified for these data as

$$y_{ij} = \beta_1 z_{i1} + \beta_2 z_{i2} + ... + \beta_S z_{iS} + \varepsilon_{ij}, \tag{2.4}$$

where y_{ij} is the i^{th} observation, $i = 1,...,N_j$, from the j^{th} subject, β_j is the random-effects parameter for the j^{th} subject, z_{ij} is the corresponding design variable for the i^{th} observation on the j^{th} subject ($z_{ij} = 1$ for the j^{th} subject and $z_{ij} = 0$ for all $j' \neq j$), and ε_{ij} is the residual for the i^{th} observation from the j^{th} subject. In matrix notation, let $\mathbf{y}_j$ and $\boldsymbol{\varepsilon}_j$ be the $N_j \times 1$ vectors of observations and residuals, respectively, from the j^{th} subject. That is, $\mathbf{y}_j = \begin{pmatrix} y_{1j} & y_{2j} & \cdots & y_{N_j j} \end{pmatrix}'$ and $\boldsymbol{\varepsilon}_j = \begin{pmatrix} \varepsilon_{1j} & \varepsilon_{2j} & \cdots & \varepsilon_{N_j j} \end{pmatrix}'$.

Defining $\mathbf{Z}_j = \begin{pmatrix} z_{1j} & z_{2j} & \cdots & z_{N_j j} \end{pmatrix}'$, the two-component joint normal-mixture random-effects model in equation (2.4), for the j^{th} subject, can be expressed using matrix notation as

$$\mathbf{y}_j = \mathbf{Z}_j \beta_j + \boldsymbol{\varepsilon}_j. \tag{2.5}$$

This model assumes that the β_j are independent across j and follow identical normal distributions with mean 0 and variance σ_δ^2, independent of the vector of residuals $\boldsymbol{\varepsilon}_j$. Using the acronym "iid" for the phrase "independently and identically distributed", this can be expressed as β_j are iid $N\left(0, \sigma_\delta^2\right)$. The residuals, ε_{ij}, for subject j, are independent across i, however, they are assumed to follow a two-component normal-mixture distribution with component mean parameters, μ_{j1} and μ_{j2}, component variance parameters, σ_1^2 and σ_2^2, and mixture proportion parameter, λ_j. The vector of residuals, $\boldsymbol{\varepsilon}_j$, is assumed to follow a two-component joint normal-mixture density

determined by the product of the N_j univariate two-component normal-mixture densities for the ε_{ij}. That is, the two-component joint normal-mixture density of the vector of residuals, $\boldsymbol{\varepsilon}_j$, denoted by ϕ_{JMX}, is

$$\phi_{JMX}\left(\boldsymbol{\varepsilon}_j;\boldsymbol{\mu}_{j1},\boldsymbol{\mu}_{j2},\boldsymbol{\Sigma}_{j1},\boldsymbol{\Sigma}_{j2},\lambda_j\right)=\prod_{i=1}^{N_j} f\left(\varepsilon_{ij};\mu_{ij1},\mu_{ij2},\sigma_1^2,\sigma_2^2,\lambda_j\right), \qquad (2.6)$$

where the two-component univariate normal-mixture density for the i^{th} residual from the j^{th} subject is

$$f\left(\varepsilon_{ij};\mu_{ij1},\mu_{ij2},\sigma_1^2,\sigma_2^2,\lambda_j\right)=\lambda_j\phi\left(\varepsilon_{ij};\mu_{ij1},\sigma_1^2\right)+\left(1-\lambda_j\right)\phi\left(\varepsilon_{ij};\mu_{ij2},\sigma_2^2\right). \qquad (2.7)$$

Here, $\phi\left(\varepsilon_{ij};\mu_{ijk},\sigma_k^2\right)$ is standard notation for the univariate normal density with mean μ_{ijk} and variance σ_k^2, $k=1,2$, $\phi\left(\varepsilon_{ij};\mu_{ijk},\sigma_k^2\right)=\dfrac{1}{\sqrt{2\pi\sigma_k^2}}\exp\left\{\dfrac{-1}{2\sigma_k^2}\left(\varepsilon_{ij}-\mu_{ijk}\right)^2\right\}$.

The k^{th} component mean vectors, $\boldsymbol{\mu}_{jk}=\left(\mu_{1jk},...,\mu_{N_jjk}\right)'$, seen in the joint density in equation (2.6), are vectors containing the k^{th} component means of the N_j residuals from the j^{th} subject. When there are no fixed-effects in the model, the μ_{ijk} are equal across *i*, and the means can be expressed as $\mu_{ijk}=\mu_{jk},\forall i$. Thus, $\boldsymbol{\mu}_{jk}=\left(\mu_{jk},....,\mu_{jk}\right)'$. Extensions with fixed-effects will be derived by modeling the component means μ_{ijk}. This is accomplished by defining a vector of fixed-effects parameters,

$$\alpha_k = \begin{pmatrix} \alpha_{1k} & \alpha_{2k} & \cdots & \alpha_{Pk} \end{pmatrix}', \tag{2.8}$$

and a corresponding design matrix, $\mathbf{X}_j$, for the j^{th} subject,

$$\mathbf{X}_j = \begin{pmatrix} x_{1j1} & x_{1j2} & \cdots & x_{1jP} \\ x_{2j1} & x_{2j2} & \cdots & x_{2jP} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N_j j1} & x_{N_j j2} & \cdots & x_{N_j jP} \end{pmatrix}. \tag{2.9}$$

Here, α_{pk} is the p^{th} fixed-effects parameter, $p = 1,...,P$, for the k^{th} component, and x_{ijp} is the corresponding observed design variable of the p^{th} fixed-effects parameter for the i^{th} observation on the j^{th} subject. This model contains $(P-1)$ fixed-effects parameters, α_{2k}, α_{3k}, …, α_{Pk}, and one intercept parameter, α_{1k}, within each component. The k^{th} component mean, $k = 1,2$, for the i^{th} residual on the j^{th} subject is then

$$\mu_{ijk} = \alpha_{1k} x_{ij1} + \alpha_{2k} x_{ij2} + \cdots + \alpha_{Pk} x_{ijP}. \tag{2.10}$$

The component mean vectors in equation (2.6), for the j^{th} subject, are then expressed as $\mu_{j1} = \mathbf{X}_j \alpha_1$ and $\mu_{j2} = \mathbf{X}_j \alpha_2$. These component mean vector, $\mathbf{X}_j \alpha_k$, contain the means of the N_j residuals, for the k^{th} component, determined by the observed values of the fixed-effects.

The component variance-covariance matrices, Σ_{jk}, of the two-component joint normal-mixture density, in equation (2.6), are $N \times N$ diagonal matrices with the component variances along the diagonals and zeros on the off diagonals. That is,

$$\boldsymbol{\Sigma}_{jk} = diag\left(\sigma_k^2, \sigma_k^2, ..., \sigma_k^2\right) = \begin{pmatrix} \sigma_k^2 & 0 & \cdots & 0 \\ 0 & \sigma_k^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k^2 \end{pmatrix}. \tag{2.11}$$

The expression $diag(\ )$ will be used to simplify notation when denoting diagonal matrices. In this notation, the diagonal elements (or matrices) are listed in order within the parenthesis, and the off diagonal elements (or blocks) are zeros (or blocks of zeros). The structure in equation (2.11), with $\text{cov}\left(\varepsilon_{ij}, \varepsilon_{i`j}\right) = 0$, $\forall i \neq i`$, is indicative of the independence assumption imposed on the residuals within each component. The two-component joint normal-mixture distribution, introduced in equation (2.3), for the j^{th} subject, is denoted as

$$\boldsymbol{\varepsilon}_j \sim JMX\left(\mathbf{X}_j\boldsymbol{\alpha}_1, \mathbf{X}_j\boldsymbol{\alpha}_2, \boldsymbol{\Sigma}_{j1}, \boldsymbol{\Sigma}_{j2}, \lambda_j\right). \tag{2.12}$$

The moments of the residuals, $\boldsymbol{\varepsilon}_j$, are described below. Using the rules for calculating expected values, the expected value of the i^{th} residual on the j^{th} subject is found by

$$\begin{aligned} \text{E}\left(\varepsilon_{ij}\right) &= \int_{-\infty}^{\infty} \varepsilon_{ij} f\left(\varepsilon_{ij}; \mu_{ij1}, \mu_{ij2}, \sigma_1^2, \sigma_2^2, \lambda_j\right) d\varepsilon_{ij} \\ &= \lambda_j \int_{-\infty}^{\infty} \varepsilon_{ij} \phi\left(\varepsilon_{ij}; \mu_{ij1}, \sigma_1^2\right) d\varepsilon_{ij} + \left(1-\lambda_j\right) \int_{-\infty}^{\infty} \varepsilon_{ij} \phi\left(\varepsilon_{ij}; \mu_{ij2}, \sigma_2^2\right) d\varepsilon_{ij} \\ &= \lambda_j \mu_{ij1} + \left(1-\lambda_j\right)\mu_{ij2}. \end{aligned} \tag{2.13}$$

This is the mean of a two-component normal-mixture random variable. Since the mixture proportion, λ_j, within the j^{th} subject, is assumed to be equal across the observations, the expected value of the vector of residuals from the j^{th} subject, $\mathrm{E}(\boldsymbol{\varepsilon}_j)$, can be expressed as

$$\mathrm{E}(\boldsymbol{\varepsilon}_j) = \lambda_j \mathbf{X}_j \boldsymbol{\alpha}_1 + (1-\lambda_j)\mathbf{X}_j \boldsymbol{\alpha}_2. \tag{2.14}$$

By defining $N_j \times N_j$ diagonal matrices, $\boldsymbol{\Lambda}_j$, with the λ_j along the diagonals and zeros on the off diagonals, denoted by $\boldsymbol{\Lambda}_j = diag(\lambda_j, \lambda_j, ..., \lambda_j)$, the expected value of $\boldsymbol{\varepsilon}_j$, can be completely expressed in matrix notation as

$$\mathrm{E}(\boldsymbol{\varepsilon}_j) = \boldsymbol{\Lambda}_j \mathbf{X}_j \boldsymbol{\alpha}_1 + (\mathbf{I}_j - \boldsymbol{\Lambda}_j)\mathbf{X}_j \boldsymbol{\alpha}_2, \tag{2.15}$$

where $\mathbf{I}_j$ is an N_j-dimensional identity matrix. The notation used in equation (2.15) will be useful later in this section when a model with multiple subjects is defined. The mean vector, $\mathrm{E}(\boldsymbol{\varepsilon}_j)$, contains the N_j expected values of the N_j univariate normal-mixture densities. This will be referred to as the general form of a mean vector from a two-component joint normal-mixture density with component mean vectors, $\mathbf{X}_j \boldsymbol{\alpha}_1$ and $\mathbf{X}_j \boldsymbol{\alpha}_2$, component variance-covariance matrices, $\boldsymbol{\Sigma}_{j1}$ and $\boldsymbol{\Sigma}_{j2}$, and mixture proportions, λ_j, defining the diagonal elements of $\boldsymbol{\Lambda}_j$. The variance of the i^{th} residual on the j^{th} subject, ε_{ij}, by definition, is $\mathrm{var}(\varepsilon_{ij}) = \mathrm{E}(\varepsilon_{ij}^2) - \mathrm{E}(\varepsilon_{ij})^2$. The expected value

of ε_{ij} is calculated in equation (2.13). Simplifying, as in equation (2.13), and applying $\mathrm{E}\left(\varepsilon_{ij}^2\right) = \mu_{ijk}^2 + \sigma_k^2$ for the $N\left(\mu_{ijk}, \sigma_k^2\right)$ case, the expected value of ε_{ij}^2 is

$$E\left(\varepsilon_{ij}^2\right) = \lambda_j\left(\mu_{ij1}^2 + \sigma_1^2\right) + \left(1-\lambda_j\right)\left(\mu_{ij2}^2 + \sigma_2^2\right). \tag{2.16}$$

Plugging $\mathrm{E}\left(\varepsilon_{ij}\right)$ and $\mathrm{E}\left(\varepsilon_{ij}^2\right)$ into the variance equation yields:

$$\begin{aligned} \mathrm{var}\left(\varepsilon_{ij}\right) &= \lambda_j\left(\mu_{ij1}^2 + \sigma_1^2\right) + \left(1-\lambda_j\right)\left(\mu_{ij2}^2 + \sigma_2^2\right) - \left[\lambda_j\mu_{ij1} + \left(1-\lambda_j\right)\mu_{ij2}\right]^2 \\ &= \lambda_j\sigma_1^2 + \left(1-\lambda_j\right)\sigma_2^2 + \lambda_j\left(1-\lambda_j\right)\left(\mu_{ij1} - \mu_{ij2}\right)^2. \end{aligned} \tag{2.17}$$

The second line, in equation (2.17), expresses the variance of the ε_{ij} as a sum of the within-component variances and the between-subject variance, all of which are weighted by their corresponding mixture proportions. The variance-covariance matrix for $\boldsymbol{\varepsilon}_j$ is then an $N_j \times N_j$ dimensional diagonal matrix with the variances for the N_j residuals along the diagonal and zeros on the off diagonal. That is,

$$\boldsymbol{\Sigma}_j = \mathrm{var}\left(\boldsymbol{\varepsilon}_j\right) = diag\left[\mathrm{var}\left(\varepsilon_{1j}\right), \mathrm{var}\left(\varepsilon_{2j}\right), ..., \mathrm{var}\left(\varepsilon_{N_j j}\right)\right]. \tag{2.18}$$

As with the component variance-covariance matrices, the structures of the $\boldsymbol{\Sigma}_j$, with $\mathrm{cov}\left(\varepsilon_{ij}, \varepsilon_{i`j}\right) = 0, \forall i \neq i`$, are indicative of the independence assumption imposed on the residuals. The structure in (2.18) will be referred to as the general form of a variance-covariance matrix from a two-component joint normal-mixture density with component

mean vectors, $\mathbf{X}_j\alpha_1$ and $\mathbf{X}_j\alpha_2$, component variance-covariance matrices, $\mathbf{\Sigma}_{j1}$ and $\mathbf{\Sigma}_{j2}$, and mixture proportion, λ_j, defining the block diagonals of $\mathbf{\Lambda}_j$.

Next, the model in equation (2.5) is extended to include multiple subjects. For $N = \sum_{j=1}^{S} N_j$, the two-component joint normal-mixture random-effects model, for S subjects, is given by equation (2.2), where $\mathbf{y}$ and $\boldsymbol{\varepsilon}$ are the $N \times 1$ vectors of observations and residuals, respectively, from the j^{th} subject, $\boldsymbol{\beta}$ is the $S \times 1$ vector of random subject effects, and $\mathbf{Z}$ is the observed $N \times S$ design matrix corresponding to the random subject effects, with

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_S \end{pmatrix}, \ \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_S \end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_S \end{pmatrix}, \text{ and } \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_S \end{pmatrix}. \tag{2.19}$$

For this model, the random subject effects, $\boldsymbol{\beta}$, are assumed to be independent and follow an S-variate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix

$$\mathbf{G} = diag\left(\sigma_\delta^2, \sigma_\delta^2, ..., \sigma_\delta^2\right). \tag{2.20}$$

That is, $\boldsymbol{\beta} \sim \mathrm{N}_S(\mathbf{0}, \mathbf{G})$. For the variance-covariance matrix in equation (2.20), the variances along the diagonal are equal and the covariances are zero, since the β_j are assumed to be iid $N\left(0, \sigma_\delta^2\right)$.

The vector of residuals, $\boldsymbol{\varepsilon}$, is independent of the vector of random-effects and assumed to follow a two-component joint normal-mixture density determined by the

product of the N univariate two-component normal-mixture densities for the ε_{ij}. That is, the two-component joint normal-mixture density of the vector $\boldsymbol{\varepsilon}$ is

$$\phi_{JMX}(\boldsymbol{\varepsilon};\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2,\boldsymbol{\lambda}) = \prod_{j=1}^{S}\prod_{i=1}^{N_j} f\left(\varepsilon_{ij};\mu_{ij1},\mu_{ij2},\sigma_1^2,\sigma_2^2,\lambda_j\right), \tag{2.21}$$

where the two-component univariate normal-mixture densities are defined in equation (2.7). The density in equation (2.21) is the extension of equation (2.6) to S subjects. The $S \times 1$ vector of mixture proportions, $\boldsymbol{\lambda} = \begin{pmatrix}\lambda_1 & \lambda_2 & \cdots & \lambda_S\end{pmatrix}'$, in equation (2.21), contains the mixture proportions for the S subjects. A special case is when the mixture proportions across subjects are assumed equal. Then, $\lambda = \lambda_j, \forall j$, and

$\boldsymbol{\lambda} = \begin{pmatrix}\lambda & \lambda & \cdots & \lambda\end{pmatrix}'$.

The component mean vectors, $\boldsymbol{\mu}_k = \left(\boldsymbol{\mu}_{1k}',...,\boldsymbol{\mu}_{Sk}'\right)'$, seen in the joint density, in equation (2.6), are vectors containing the component means of the N residuals for all subjects. Extensions with fixed-effects are derived by modeling the component means μ_{ijk}, as in equation (2.10). This is accomplished by defining a vector of fixed-effects parameters, $\boldsymbol{\alpha}_k$, as in equation (2.8), and a corresponding design matrix,

$$\mathbf{X} = \begin{pmatrix}\mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_S\end{pmatrix}, \tag{2.22}$$

where the $\mathbf{X}_j$, $j = 1,...,S$, are defined in equation (2.9). Notice that the vector of fixed-effects is the same for each subject. The component mean vectors in equation (2.21) can then be expressed as $\mu_1 = \mathbf{X}\alpha_1$ and $\mu_2 = \mathbf{X}\alpha_2$. The component mean vector, $\mathbf{X}\alpha_k$, $k = 1,2$, contain the means of the N residuals for the k^{th} component, determined by the observed values of the fixed-effects.

The distribution in equation (2.21) is denoted as in equation (2.3) by $\boldsymbol{\varepsilon} \sim JMX(\mathbf{X}\alpha_1, \mathbf{X}\alpha_2, \Sigma_1, \Sigma_2, \lambda)$ and is the two-component joint normal-mixture density with component mean vectors, $\mathbf{X}\alpha_k$, component variance-covariance matrices, $\boldsymbol{\Sigma}_k$, and vector of mixture proportions λ, for $k = 1,2$. The component variance-covariance matrices, $\boldsymbol{\Sigma}_k$, are block-diagonal matrices with the subject component variance-covariance matrices blocked along the diagonals and blocks of zeros on the off diagonals. That is,

$$\boldsymbol{\Sigma}_k = diag(\boldsymbol{\Sigma}_{1k}, \boldsymbol{\Sigma}_{2k}, ..., \boldsymbol{\Sigma}_{Sk}). \tag{2.23}$$

Notice that the structure of the component variance-covariance matrices, given in equation (2.23), for the model containing all subjects, with $\text{cov}(\varepsilon_{ij}, \varepsilon_{ij`}) = 0, \forall j \neq j`$ and $\text{cov}(\varepsilon_{ij}, \varepsilon_{i`j}) = 0, \forall i \neq i`$, now takes into account the assumption of independence in the residuals between subjects, along with the assumption of independence in the residuals within subjects.

The expected value vector of $\boldsymbol{\varepsilon}$, $\text{E}(\boldsymbol{\varepsilon})$, is a vector containing the S expected value vectors, $\text{E}(\boldsymbol{\varepsilon}_j)$, for the S subjects, given by

$$\mathrm{E}(\boldsymbol{\varepsilon}) = \begin{pmatrix} \mathrm{E}(\boldsymbol{\varepsilon}_1) \\ \mathrm{E}(\boldsymbol{\varepsilon}_2) \\ \vdots \\ \mathrm{E}(\boldsymbol{\varepsilon}_S) \end{pmatrix} = \begin{pmatrix} \lambda_1 \mathbf{X}\boldsymbol{\alpha}_1 + (1-\lambda_1)\mathbf{X}\boldsymbol{\alpha}_2 \\ \lambda_2 \mathbf{X}\boldsymbol{\alpha}_1 + (1-\lambda_2)\mathbf{X}\boldsymbol{\alpha}_2 \\ \vdots \\ \lambda_S \mathbf{X}\boldsymbol{\alpha}_1 + (1-\lambda_S)\mathbf{X}\boldsymbol{\alpha}_2 \end{pmatrix}. \tag{2.24}$$

By defining an $N \times N$ block-diagonal matrix, $\boldsymbol{\Lambda}$, $\boldsymbol{\Lambda} = diag(\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, ..., \boldsymbol{\Lambda}_S)$, the expected value of $\boldsymbol{\varepsilon}$, can be completely expressed in matrix notation as

$$\mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{\Lambda}\mathbf{X}\boldsymbol{\alpha}_1 + (\mathbf{I} - \boldsymbol{\Lambda})\mathbf{X}\boldsymbol{\alpha}_2, \tag{2.25}$$

where $\mathbf{I}$ is an N-dimensional identity matrix. This is the form of a mean vector, for S subjects, from a two-component joint normal-mixture density with component mean vectors, $\mathbf{X}\boldsymbol{\alpha}_1$ and $\mathbf{X}\boldsymbol{\alpha}_2$, component variance-covariance matrices, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, and vector of mixture proportions, $\boldsymbol{\lambda}$, defining the diagonal blocks of $\boldsymbol{\Lambda}$. The variance-covariance matrix for $\boldsymbol{\varepsilon}$ is then a block-diagonal matrix with the subject variance-covariance matrices blocked along the diagonal and blocks of zeros on the off diagonal. That is,

$$\boldsymbol{\Sigma} = diag(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, ..., \boldsymbol{\Sigma}_S), \tag{2.26}$$

where $\boldsymbol{\Sigma}_j$, $j = 1,...,S$, is defined in equation (2.18). This is the form of a variance-covariance matrix, for S subjects, from a two-component joint normal-mixture density with component mean vectors, $\mathbf{X}\boldsymbol{\alpha}_1$ and $\mathbf{X}\boldsymbol{\alpha}_2$, component variance-covariance matrices, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, and vector of mixture proportions, $\boldsymbol{\lambda}$, defining the diagonal blocks of $\boldsymbol{\Lambda}$.

In the next section, the two-component joint normal-mixture random-effects model definition seen in this section is extended to include more than two components.

Although much of the extension is obvious, the complete model definition for M components will be derived for completion.

2.4. M-Component Joint Normal-Mixture Random-Effects Model Definition

In this section, the two-component joint normal-mixture random-effects model, introduced in section 2.3, is extended to a model containing M-components. The model is still defined as in equation (2.2), however, $\boldsymbol{\varepsilon}$ is assumed to have an M-component joint normal-mixture density with M component $N \times 1$ mean vectors $\mathbf{X}\boldsymbol{\alpha}_1$, $\mathbf{X}\boldsymbol{\alpha}_2$, ..., $\mathbf{X}\boldsymbol{\alpha}_M$, M component $N \times N$ variance-covariance matrices $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$, ..., $\boldsymbol{\Sigma}_M$, and $M-1$ component $S \times 1$ vectors of mixture proportion parameters, $\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2$, ..., $\boldsymbol{\lambda}_{M-1}$. That is,

$$\boldsymbol{\varepsilon} \sim \mathrm{JMX}\left(\mathbf{X}\boldsymbol{\alpha}_1,...,\mathbf{X}\boldsymbol{\alpha}_M,\boldsymbol{\Sigma}_1,...,\boldsymbol{\Sigma}_M,\boldsymbol{\lambda}_1,...,\boldsymbol{\lambda}_{M-1}\right). \tag{2.27}$$

The j^{th} element in the k^{th} vector of mixture proportions, λ_{jk}, $j = 1,...,S$, $k = 1,...,M-1$, contains the proportion, for the j^{th} subject, of the k^{th} normal distribution, in the joint normal-mixture distribution. It is assumed that $\sum_{k=1}^{M} \lambda_{jk} = 1$, $0 < \lambda_{jk} < 1$, thus, only $M-1$ of the M mixture proportions, for each subject, need to be estimated, since

$$\lambda_{jM} = 1 - \sum_{k=1}^{M-1} \lambda_{jk}.$$

The M-component joint normal-mixture random-effects model for the j^{th} subject is expressed in equation (2.5). The residuals, ε_{ij}, for the j^{th} subject are assumed to be independent across i, however, they are assumed to follow an M-component normal-

mixture density with component means, μ_{j1}, μ_{j2}, ..., μ_{jM}, component variances, σ_1^2, σ_2^2, ..., σ_M^2, and mixture proportions, λ_{j1}, λ_{j2}, ..., λ_{jM-1}. The vector of residuals, for the j^{th} subject, $\boldsymbol{\varepsilon}_j$, is then assumed to follow a density given by the product of the N_j univariate M-component normal-mixture densities for the ε_{ij}. That is, the M-component joint normal-mixture density of the vector $\boldsymbol{\varepsilon}_j$ is

$$\begin{aligned}&\phi_{JMX}\left(\boldsymbol{\varepsilon}_j;\boldsymbol{\mu}_{j1},...,\boldsymbol{\mu}_{jM},\boldsymbol{\Sigma}_{j1},...,\boldsymbol{\Sigma}_{jM},\lambda_{j1},...,\lambda_{jM-1}\right)\\&\quad=\prod_{i=1}^{N_j} f\left(\varepsilon_{ij};\mu_{ij1},...,\mu_{ijM},\sigma_1^2,...,\sigma_M^2,\lambda_{j1},...,\lambda_{jM-1}\right),\end{aligned} \tag{2.28}$$

where the M-component univariate normal-mixture density of the i^{th} residual from the j^{th} subject is given by

$$f\left(\varepsilon_{ij};\mu_{ij1},...,\mu_{ijM},\sigma_1^2,...,\sigma_M^2,\lambda_{j1},...,\lambda_{jM-1}\right)=\sum_{k=1}^{M}\lambda_{jk}\phi\left(\varepsilon_{ij};\mu_{ijk},\sigma_k^2\right). \tag{2.29}$$

The component mean vectors, $\boldsymbol{\mu}_{jk}=\left(\mu_{1jk},...,\mu_{N_j jk}\right)'$, seen in the joint density, in equation (2.28), are vectors containing the component means of the N_j residuals from the j^{th} subject. When there are no fixed-effects in the model, the μ_{ijk} are equal across i, and the means can be expressed as $\mu_{ijk}=\mu_{jk},\forall i$. Thus, $\boldsymbol{\mu}_{jk}=\left(\mu_{jk},...,\mu_{jk}\right)'$. Extending the model to include fixed-effects, the vectors $\boldsymbol{\alpha}_k$ and the corresponding design matrix, $\mathbf{X}_j$, for the j^{th} subject, are defined as in equations (2.8) and (2.9), respectively. The models for the k^{th} component mean, for the i^{th} residual on the j^{th} subject

are given in equation (2.10). The component mean vectors for the density in equation (2.28) are then expressed as $\boldsymbol{\mu}_{jk} = \mathbf{X}_j \boldsymbol{\alpha}_k$, $k = 1,...,M$.

The component variance-covariance matrices, $\boldsymbol{\Sigma}_{jk}$, of the M-component joint normal-mixture density in equation (2.28), are diagonal matrices with the component variances along the diagonals and zeros on the off diagonals, given by equation (2.11). The M-component joint normal-mixture density, for the j^{th} subject, of the vector of residuals, $\boldsymbol{\varepsilon}_j$, expressed in equation (2.28), is denoted as

$$\boldsymbol{\varepsilon}_j \sim JMX\left(\mathbf{X}_j\boldsymbol{\alpha}_1,...,\mathbf{X}_j\boldsymbol{\alpha}_M,\boldsymbol{\Sigma}_{j1},...,\boldsymbol{\Sigma}_{jM},\lambda_{j1},...,\lambda_{jM-1}\right). \tag{2.30}$$

The expected value of the i^{th} residual for the j^{th} subject is computed as in equation (2.13), and is given by

$$\mathrm{E}\left(\varepsilon_{ij}\right) = \sum_{k=1}^{M} \lambda_{jk}\mu_k. \tag{2.31}$$

The expected value of the vector of residuals from the j^{th} subject, $\boldsymbol{\varepsilon}_j$, can be expressed, using matrices, as

$$\mathrm{E}\left(\boldsymbol{\varepsilon}_j\right) = \sum_{k=1}^{M} \mathrm{E}\left(\boldsymbol{\varepsilon}_j\right) = \boldsymbol{\Lambda}_{jk}\mathbf{X}_j\boldsymbol{\alpha}_k, \tag{2.32}$$

where the $\boldsymbol{\Lambda}_{jk} = diag\left(\lambda_{jk},\lambda_{jk},...,\lambda_{jk}\right)$ are $N_j \times N_j$ diagonal matrices with λ_{jk} along the diagonals and zeros on the off diagonals. The mean vector, in equation (2.32), is the form of a mean vector from an M-component joint normal-mixture density with component mean vectors, $\mathbf{X}_j\boldsymbol{\alpha}_k$, component variance-covariance matrices, $\boldsymbol{\Sigma}_{jk}$, and

diagonal matrix Λ_{jk}, containing the k^{th} component mixture proportion for the j^{th} subject, λ_{jk}, $k = 1,...,M$. The variance-covariance matrix, Σ_j, for ε_j, is given by equation (2.18), where

$$\text{var}\left(\varepsilon_{ij}\right) = \sum_{k=1}^{M} \lambda_{jk}\left(\mu_{ijk}^2 + \sigma_k^2\right) - \left(\sum_{k=1}^{M} \lambda_{jk}\mu_{ijk}\right)^2 \tag{2.33}$$

denotes the variance of the i^{th} residual on the j^{th} subject.

The model in equation (2.2), containing S subjects, remains as it was defined in section 2.3, however, the vector of residuals, ε, is assumed to follow a distribution determined by the product of the N univariate M-component normal-mixture densities for the ε_{ij}. That is, the M-component joint normal-mixture density of the ε is

$$\begin{aligned}&\phi_{JMX}\left(\varepsilon;\mu_1,...,\mu_M,\Sigma_1,...,\Sigma_M,\lambda_1,...,\lambda_{M-1}\right)\\ &= \prod_{j=1}^{S}\prod_{i=1}^{N_j} f\left(\varepsilon_{ij};\mu_{ij1},...,\mu_{ijM},\sigma_1^2,...,\sigma_M^2,\lambda_{j1},...,\lambda_{jM-1}\right),\end{aligned} \tag{2.34}$$

where the M-component univariate normal-mixture densities are defined in equation (2.29).

The component mean vectors, $\mu_k = \left(\mu_{1k}',...,\mu_{Sk}'\right)'$, seen in the joint density in equation (2.34), are vectors containing the component means of the N residuals for all subjects. Extensions with fixed-effects are derived by modeling the component means μ_{ijk}, as in equation (2.10). This is accomplished by defining a vector of fixed-effects parameters, α_k, as in equation (2.8) and a corresponding design matrix, $\mathbf{X}$, as in

equation (2.22). The component mean vectors in equation (2.34) can then be expressed as $\mu_k = \mathbf{X}\alpha_k$, $k = 1,2,\cdots,M$. The component variance-covariance matrices, $\mathbf{\Sigma}_k$, of the M-component joint normal-mixture density in equation (2.34), are block diagonal matrices with the subject component variance-covariance matrices, $\mathbf{\Sigma}_{jk}$ blocked along the diagonal and blocks of zeros on the off diagonal, given in equation (2.23).

The expected value vector of $\boldsymbol{\varepsilon}$ is given by

$$\mathrm{E}(\boldsymbol{\varepsilon}) = \sum_{k=1}^{M} \Lambda_k \mathbf{X}\alpha_k, \tag{2.35}$$

where $\Lambda_k = diag(\Lambda_{1k}, \Lambda_{2k}, ..., \Lambda_{Sk})$ are $N \times N$ block-diagonal matrices with the Λ_{jk} blocked along the diagonals and blocks of zeros on the off diagonals. This is the form of a mean vector, for S subjects, from an M-component joint normal-mixture density with component mean vectors, $\mathbf{X}\alpha_k$, component variance-covariance matrices, $\mathbf{\Sigma}_k$, and diagonal matrix Λ_{jk}, containing the k^{th} component mixture proportion for the j^{th} subject, λ_{jk}, $k = 1,...,M$. The variance-covariance matrix, $\mathbf{\Sigma}$, for $\boldsymbol{\varepsilon}$, is a block diagonal matrix with the subject variance-covariance matrices, $\mathbf{\Sigma}_j$, blocked along the diagonal and blocks of zeros on the off diagonal, given by equation (2.26), where $\mathbf{\Sigma}_j$, $j = 1,...,S$, is defined in equation (2.18) and $\mathrm{var}(\varepsilon_{ij})$ is given in equation (2.33).

It the next section, the distribution of the observations, y_{ij}, is discussed. It will be shown that although the distribution of the vector of observations, $\mathbf{y}$, is intractable, its definition may be avoided for purposes of estimation and parameter testing.

2.5. Distributional Properties of the Joint Normal-Mixture Random-Effects Model

In section 2.3, the distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ were completely defined for two components. The density of $\boldsymbol{\varepsilon}$ was then extended in section 2.4 for the case of M-components. Now that the distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ have been completely specified, the distribution of $\mathbf{y}$ can be discussed. This distribution will provide the likelihood for all of the model parameters, given the observed data. Start by considering the distribution of $\mathbf{y}$ conditional on $\boldsymbol{\beta}$. Using the properties of conditional expectation and the fact that linear combinations of normal distributions are also normal distributions, the conditional distribution of $\mathbf{y}$, given $\boldsymbol{\beta}$, can be shown to be a joint normal-mixture density given by

$$\mathbf{y} \mid \boldsymbol{\beta} \sim JMX\left(\mathbf{Z}\boldsymbol{\beta} + \mathbf{X}\boldsymbol{\alpha}_1, \ldots, \mathbf{Z}\boldsymbol{\beta} + \mathbf{X}\boldsymbol{\alpha}_M, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_M, \lambda_1, \ldots, \lambda_{M-1}\right). \tag{2.36}$$

To see this, first notice that the expected value of $\mathbf{y} \mid \boldsymbol{\beta}$ is

$$\begin{aligned} E(\mathbf{y} \mid \boldsymbol{\beta}) &= E\left[(\mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \mid \boldsymbol{\beta}\right] = E(\mathbf{Z}\boldsymbol{\beta} \mid \boldsymbol{\beta}) + E(\boldsymbol{\varepsilon} \mid \boldsymbol{\beta}) \\ &= \mathbf{Z}\boldsymbol{\beta} + \sum_{k=1}^{M} \Lambda_k \mathbf{X}\boldsymbol{\alpha}_k \\ &= \sum_{k=1}^{M} \Lambda_k \mathbf{Z}\boldsymbol{\beta} + \sum_{k=1}^{M} \Lambda_k \mathbf{X}\boldsymbol{\alpha}_k \\ &= \sum_{k=1}^{M} \Lambda_k (\mathbf{Z}\boldsymbol{\beta} + \mathbf{X}\boldsymbol{\alpha}_k). \end{aligned} \tag{2.37}$$

This is the form of a mean vector, for S subjects, from an M-component joint normal-mixture density with component mean vectors $\mathbf{Z}\boldsymbol{\beta} + \mathbf{X}\boldsymbol{\alpha}_k$ and vector of mixture proportions, λ_k, defining the diagonal blocks of Λ_k, $k = 1,...,M$. The variance of $\mathbf{y} \mid \boldsymbol{\beta}$ is given by

$$\text{var}(\mathbf{y} \mid \boldsymbol{\beta}) = \text{var}\left[(\mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \mid \boldsymbol{\beta}\right] = \text{var}(\mathbf{Z}\boldsymbol{\beta} \mid \boldsymbol{\beta}) + \text{var}(\boldsymbol{\varepsilon} \mid \boldsymbol{\beta}) = \text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}. \qquad (2.38)$$

This is the form of a variance-covariance matrix, for S subjects, from an M-component joint normal-mixture density with component mean vectors $\mathbf{Z}\boldsymbol{\beta} + \mathbf{X}\boldsymbol{\alpha}_k$, component variance-covariance matrices $\boldsymbol{\Sigma}_k$, and vector of mixture proportions, λ_k, defining the block diagonals of Λ, $k = 1,...,M$. The distributional form of $\mathbf{y} \mid \boldsymbol{\beta}$ is the same as that of the residuals, an M-component joint normal-mixture density, since $\boldsymbol{\beta}$ is assumed to be a constant in the conditional model. The variance of $\boldsymbol{\varepsilon}$ and $\mathbf{y} \mid \boldsymbol{\beta}$ are equivalent, however the mean for $\mathbf{y} \mid \boldsymbol{\beta}$ includes the $\boldsymbol{\beta}$ where the mean for the $\boldsymbol{\varepsilon}$ does not. Thus, the distribution in equation (2.36) follows.

Now, consider the vector of observations, $\mathbf{y}$. The expected value of the vector of observations is found by taking the expected value, with respect to $\boldsymbol{\beta}$, of the conditional expectation of $\mathbf{y} \mid \boldsymbol{\beta}$. That is, $\text{E}(\mathbf{y}) = \text{E}\left[\text{E}(\mathbf{y}|\boldsymbol{\beta})\right]$. Using the fact that $\text{E}(\boldsymbol{\beta}) = \mathbf{0}$, by assumption, the expected value of $\mathbf{y}$ is

$$\begin{aligned} \mathrm{E}(\mathbf{y}) &= \mathrm{E}\left[\mathrm{E}(\mathbf{y}|\boldsymbol{\beta})\right] = \mathrm{E}\left(\mathbf{Z}\boldsymbol{\beta} + \sum_{k=1}^{M} \boldsymbol{\Lambda}_k \mathbf{X}\boldsymbol{\alpha}_k\right) \\ &= \mathrm{E}(\mathbf{Z}\boldsymbol{\beta}) + \sum_{k=1}^{M} \mathrm{E}(\boldsymbol{\Lambda}_k \mathbf{X}\boldsymbol{\alpha}_k) \\ &= \sum_{k=1}^{M} \boldsymbol{\Lambda}_k \mathbf{X}\boldsymbol{\alpha}_k. \end{aligned} \tag{2.39}$$

This is the form of a mean vector from an M-component joint normal-mixture density with component mean vectors, $\mathbf{X}\boldsymbol{\alpha}_1$ and $\mathbf{X}\boldsymbol{\alpha}_2$, and vector of mixture proportions, $\boldsymbol{\lambda}$, defining the block diagonals of $\boldsymbol{\Lambda}$. Next, the variance of $\mathbf{y}$ is found by taking the variance, with respect to $\boldsymbol{\beta}$, of the conditional expectation of $\mathbf{y}|\boldsymbol{\beta}$ and adding that to the expected value, with respect to $\boldsymbol{\beta}$, of the conditional variance of $\mathbf{y}|\boldsymbol{\beta}$. That is, $\mathrm{var}(\mathbf{y}) = \mathrm{var}\left[\mathrm{E}(\mathbf{y}|\boldsymbol{\beta})\right] + \mathrm{E}\left[\mathrm{var}(\mathbf{y}|\boldsymbol{\beta})\right]$. Then the variance of $\mathbf{y}$, which will be denoted by $\mathbf{V}$, is given by

$$\begin{aligned} \mathbf{V} = \mathrm{var}(\mathbf{y}) &= \mathrm{var}\left[\mathrm{E}(\mathbf{y}|\boldsymbol{\beta})\right] + \mathrm{E}\left[\mathrm{var}(\mathbf{y}|\boldsymbol{\beta})\right] \\ &= \mathrm{var}\left[\mathbf{Z}\boldsymbol{\beta} + \sum_{k=1}^{M} \boldsymbol{\Lambda}_k \mathbf{X}\boldsymbol{\alpha}k\right] + \mathrm{E}\left[\mathrm{var}(\boldsymbol{\varepsilon})\right] \\ &= \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathrm{var}(\boldsymbol{\varepsilon}) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \boldsymbol{\Sigma}, \end{aligned} \tag{2.40}$$

which is the form of a variance-covariance matrix, for S subjects, from an M-component joint normal-mixture density with component mean vectors, $\mathbf{X}\boldsymbol{\alpha}_k$, component variance-covariance matrices, $\mathbf{V}_k = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \boldsymbol{\Sigma}_k$, and vectors of mixture proportions, $\boldsymbol{\lambda}_k$, $k = 1,\ldots,M$. The structure of the variance-covariance matrices for the observations, $\mathbf{V}$, is indicative of the assumption that observations are correlated within subjects,

$$\text{var}\left(y_{ij}\right)=\sigma_{\delta}^{2}+\text{var}\left(\varepsilon_{ij}\right) \text{ and } \text{cov}\left(y_{ij},y_{i`j}\right)=\sigma_{\delta}^{2},\forall i\neq i`, \tag{2.41}$$

and independent across subjects,

$$\text{cov}\left(y_{ij},y_{ij`}\right)=0,\forall j\neq j`, \tag{2.42}$$

where $\text{var}\left(\varepsilon_{ij}\right)$ is defined in equation (2.33).

The form of the mean and variance for the vector of observations, $\mathbf{y}$, was described above. Next the form of the distribution will be discussed. Recall that the joint distribution of two vectors of random variables, $\mathbf{y}$ and β, can be expressed as the product of the conditional distribution of $\mathbf{y}$ given β and the marginal distribution of β, $f(\mathbf{y},\beta)=f(\mathbf{y}\mid\beta)f(\beta)$. The marginal density of $\mathbf{y}$ is then

$$f(\mathbf{y})=\int_{-\infty}^{\infty}f(\mathbf{y}\mid\beta)f(\beta)d\beta. \tag{2.43}$$

Since the expression for the conditional density,

$$f(\mathbf{y}\mid\beta)=\prod_{j=1}^{S}\prod_{i=1}^{N_j}\left[\sum_{k=1}^{M}\lambda_{jk}\phi\left(y_{ij};\mu_{ijk}+\beta_j,\sigma_k^2\right)\right], \tag{2.44}$$

involves a double product of a summation over the components, the integral in equation (2.43) can not be simplified by interchanging the integral with the double product and therefore does not have an explicit form. Determining the maximum likelihood estimates would involve maximizing this integral which requires numerical methods involving indefinite integrals. The marginal density of the i^{th} observation from the j^{th} subject in the

vector of observations, $\mathbf{y}$, however, is tractable and is shown below. The marginal density of y_{ij} with respect to β is

$$\begin{aligned} f(y_{ij}) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_{ij}, \beta_j) d\beta_1 \ldots d\beta_S \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[f(y_{ij} \mid \beta_j) f(\beta_j) \right] d\beta_1 \ldots d\beta_S . \end{aligned} \tag{2.45}$$

The densities of the random subject effects are iid $N\left(0, \sigma_\delta^2\right)$, denoted by $\phi\left(\beta_j; 0, \sigma_\delta^2\right)$. When the random subjects effects, β_j, are conditioned on, they are assumed to be constant values. Thus, the y_{ij} can be expressed as the sum of the residual and the random effect, $y_{ij} = \beta_j + \varepsilon_{ij}$. The conditional densities, $y_{ij} \mid \beta_j$, are each univariate normal-mixture densities with component means $\beta_j + \mu_{ijk}$, component variances σ_k^2, and mixture proportions λ_{jk}, $i = 1,\ldots,N_j$, $j = 1,\ldots,S$, $k = 1,\ldots,M$. The density of $y_{ij} \mid \beta_j$ is then expressed as

$$f\left(y_{ij} \mid \beta_j; \beta_j + \mu_{ij1}, \ldots, \beta_j + \mu_{ijM}, \sigma_1^2, \ldots, \sigma_M^2, \lambda_1, \ldots, \lambda_{M-1}\right), \tag{2.46}$$

where the vectors λ_k contain the k^{th} component mixture proportions for the S subjects. Then, the marginal density of the observations in equation (2.45) can be expressed as

$$
\begin{aligned}
f(y_{ij}) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[f(y_{ij} \mid \beta_j) f(\beta_j) \right] d\beta_1 \ldots d\beta \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \left[\sum_{k=1}^{M} \lambda_{jk} \phi\left(y_{ij}; \beta_j + \mu_{ijk}, \sigma_k^2\right) \right] \phi\left(\beta_j; 0, \sigma_\delta^2\right) \right\} d\beta_1 \ldots d\beta_S \qquad (2.47) \\
&= \sum_{k=1}^{M} \left\{ \lambda_{jk} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\phi\left(y_{ij}; \beta_j + \mu_{ijk}, \sigma_k^2\right) \phi\left(\beta_j; 0, \sigma_\delta^2\right) \right] d\beta_1 \ldots d\beta_S \right\}.
\end{aligned}
$$

The integrand in equation (2.47) is now expressed as a bivariate normal density. It can be shown that the density for the y_{ij}, is the same M-component normal-mixture density as that of the residuals, with the exception of the variance parameter. To see this, the integrand in equation (2.47) is simplified by separating the terms involving β_j from the terms not involving β_j:

$$
\begin{aligned}
f(y_{ij} \mid \beta_j) f(\beta_j) &= \phi\left(y_{ij}; \mu_{ijk} + \beta_j, \sigma_k^2\right) \phi\left(\beta_j; 0, \sigma_\delta^2\right) \\
&= \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left\{ \frac{-1}{2\sigma_k^2} \left[y_{ij} - \left(\mu_{ijk} + \beta_j\right) \right]^2 \right\} \frac{1}{\sqrt{2\pi\sigma_\delta^2}} \exp\left\{ \frac{-1}{2\sigma_\delta^2} \beta_j^2 \right\} \\
&= \frac{1}{\sqrt{2\pi\sigma_k^2}\sqrt{2\pi\sigma_\delta^2}} \exp\left\{ -\frac{1}{2\sigma_\delta^2} \beta_j^2 \right\} \times \qquad (2.48) \\
&\qquad \exp\left\{ -\frac{1}{2\sigma_k^2} \left[y_{ij}^2 - 2y_{ij}\mu_{ijk} - 2y_{ij}\beta_j + \mu_{ijk}^2 + 2\mu_{ijk}\beta_j + \beta_j^2 \right] \right\}
\end{aligned}
$$

$$= \frac{1}{\sqrt{2\pi\sigma_k^2}\sqrt{2\pi\sigma_\delta^2}} \exp\left\{-\frac{1}{2\sigma_k^2}\left[y_{ij}^2 - 2y_{ij}\mu_{ijk} + \mu_{ijk}^2\right]\right\} \times$$
$$\exp\left\{-\frac{1}{2}\left[\frac{\beta_j^2}{\sigma_k^2} + \frac{\beta_j^2}{\sigma_\delta^2} + \frac{2\mu_{ijk}\beta_j}{\sigma_k^2} - \frac{2y_{ij}\beta_j}{\sigma_k^2}\right]\right\}.$$

After further simplification, the result in equation (2.48) can be expressed

$$f\left(y_{ij} \mid \beta_j\right) f\left(\beta_j\right) = \frac{1}{\sqrt{2\pi\sigma_k^2}\sqrt{2\pi\sigma_\delta^2}} \exp\left\{-\frac{1}{2\sigma_k^2}\left(y_{ij} - \mu_{ijk}\right)^2\right\} \times$$
$$\exp\left\{-\frac{\left(\sigma_k^2 + \sigma_\delta^2\right)}{2\sigma_k^2\sigma_\delta^2}\left[\beta_j^2 - 2\beta_j \frac{\sigma_\delta^2\left(y_{ij} - \mu_{ijk}\right)}{\left(\sigma_k^2 + \sigma_\delta^2\right)}\right]\right\}. \tag{2.49}$$

After completing the square for the exponential term involving the β_j, the result, in equation (2.49), is then multiplied by a constant factor of one, $\frac{\left(\sigma_k^2 + \sigma_\delta^2\right)}{\left(\sigma_k^2 + \sigma_\delta^2\right)}$, to form an expression, for the term involving β_j, that resembles a normal probability density function (pdf):

$$f\left(y_{ij} \mid \beta_j\right) f\left(\beta_j\right) = \frac{1}{\sqrt{2\pi\sigma_k^2}\sqrt{2\pi\sigma_\delta^2}} \exp\left\{-\frac{1}{2\sigma_k^2}\left(y_{ij} - \mu_{ijk}\right)^2\right\} \times$$

$$\exp\left\{-\frac{\left(\sigma_k^2 + \sigma_\delta^2\right)}{2\sigma_k^2\sigma_\delta^2}\left[\beta_j - \left(\frac{\sigma_\delta^2\left(y_{ij} - \mu_{ijk}\right)}{\left(\sigma_k^2 + \sigma_\delta^2\right)}\right)\right]^2\right\} \times$$

$$\exp\left\{\frac{\sigma_\delta^2}{2\sigma_k^2\left(\sigma_k^2 + \sigma_\delta^2\right)}\left(y_{ij} - \mu_{ijk}\right)^2\right\}$$

$$= \left[\frac{1}{\sqrt{2\pi\left(\sigma_k^2 + \sigma_\delta^2\right)}} \exp\left\{-\frac{1}{2\left(\sigma_k^2 + \sigma_\delta^2\right)}\left(y_{ij} - \mu_{ijk}\right)^2\right\}\right] \times$$

$$\left[\sqrt{\frac{\left(\sigma_k^2 + \sigma_\delta^2\right)}{2\pi\sigma_k^2\sigma_\delta^2}} \exp\left\{-\frac{\left(\sigma_k^2 + \sigma_\delta^2\right)}{2\sigma_k^2\sigma_\delta^2}\left[\beta_j - \left(\frac{\sigma_\delta^2\left(y_{ij} - \mu_{ijk}\right)}{\left(\sigma_k^2 + \sigma_\delta^2\right)}\right)\right]^2\right\}\right] \quad (2.50)$$

The expression, in equation (2.50), for the factor on the left is free of β_j, thus it can be brought out of the integrand in equation (2.47), as the λ_{jk} term was. Now, the integral of the factor on the right is equal to one, since it is expressed as a normal density with mean $\frac{\sigma_\delta^2\left(y_{ij} - \mu_{ijk}\right)}{\left(\sigma_k^2 + \sigma_\delta^2\right)}$ and variance Thus, the marginal density for the y_{ij} is

$$
\begin{aligned}
f\left(y_{ij}\right) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[f\left(y_{ij} \mid \beta_j\right) f\left(\beta_j\right) \right] d\beta_1 \ldots d\beta \\
&= \sum_{k=1}^{M} \left\{ \lambda_{jk} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\phi\left(y_{ij}; \mu_{ijk} + \beta_j, \sigma_k^2\right) \phi\left(\beta_j; 0, \sigma_\delta^2\right) \right] d\beta_1 \ldots d\beta_S \right\} \quad (2.51) \\
&= \sum_{k=1}^{M} \left[\lambda_{jk} \frac{1}{\sqrt{2\pi\left(\sigma_k^2 + \sigma_\delta^2\right)}} \exp\left\{ -\frac{1}{2\left(\sigma_k^2 + \sigma_\delta^2\right)} \left(y_{ij} - \mu_{ijk}\right)^2 \right\} \right].
\end{aligned}
$$

The density in equation (2.51) is recognized as an M-component normal-mixture density with component means μ_{ijk}, component variances, $\left(\sigma_k^2 + \sigma_\delta^2\right)$, and mixture proportions, λ_{jk}. Thus, the forms of the densities for y_{ij} and ε_{ij} are equivalent with the same component means and mixture proportion parameters, but different variances parameters. It is important to understand here that the joint density of the vector of observations, $\mathbf{y}$, and joint density of the vector of residuals, $\boldsymbol{\varepsilon}$, do not share this property. Although the forms of the component means for the two vectors are equal, the variance-covariance structure for the vector of observations must now take into account the correlations on observations within the subjects. Since the observations are not independent within subjects, as the residuals are, the joint density of the vector of observations does not follow as it did with the residuals by taking the double product of the N univariate mixture-normal densities given in equation (2.51). It will be shown in section 2.8, that estimation of the parameters for the M-component joint normal-mixture random-effects model can be obtained using the marginal density of the vector of observations.

However, before this it is necessary to review the methods for maximum likelihood estimation in the normal random-effects models and in the normal-mixture model.

2.6. Estimation in the Normal Random-Effects Model

In this section, the methods used for estimation in the normal random-effects model, described in section 2.2, are reviewed. More specifically, in section 2.6.1, the general methods used for maximizing the likelihood of the data are given, while estimation of the fixed-effects parameters, random-effects parameters, and variance parameters are covered in more detail in sections 2.6.2, 2.6.3, and 2.6.4, respectively.

2.6.1 Methods of Estimation

Suppose a random sample of N observations is obtained from a normal random-effects model as defined in equation (2.1). The likelihood of the model parameters, given the vector of N observations, is given by

$$L = L(\boldsymbol{\alpha}, \boldsymbol{\gamma}; \mathbf{y}) = (2\pi)^{-N/2} |\mathbf{V}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})\right\}, \qquad (2.52)$$

where $\boldsymbol{\alpha}$ is a vector of fixed-effects parameters and $\boldsymbol{\gamma}$ is a vector containing all the variance parameters. The log-likelihood function is then written as

$$\begin{aligned} l = l(\boldsymbol{\alpha}, \boldsymbol{\gamma}; \mathbf{y}) &= -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}) \\ &= C - \frac{1}{2}\left\{\log|\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})\right\}, \end{aligned} \qquad (2.53)$$

where $C = -\frac{N}{2}\log(2\pi)$ is a constant that can be ignored in the maximization process. Estimates of the parameters for the model given in equation (2.1) are found by maximizing the log-likelihood given in equation (2.53) with respect to α and γ. There are several methods that can be used to maximize the log-likelihood function, which may lead to different estimates of the model parameters. One method, referred to as maximum likelihood, first maximizes the log-likelihood with respect to the variance parameters, while treating the fixed-effects parameters as constants. Once the variance parameters are obtained, the fixed-effects parameters are then estimated by maximizing the log-likelihood with respect to the fixed-effects while treating the variance parameters as constant. This approach can produce variance parameters that are biased downwards because they are based on the assumption that the fixed-effects parameters are known (Brown and Prescott, 1999).

A second approach is the residual maximum likelihood (REML) method. This approach starts by removing the fixed-effects parameters from the log-likelihood, defining it only in terms of the variance parameters. Then, a likelihood function based on the full residuals, $\mathbf{y} - \mathbf{X}\hat{\alpha}$, rather than the ordinary residuals, $\mathbf{y} - \mathbf{X}\hat{\alpha} - \mathbf{Z}\hat{\beta}$ must be determined. It has been shown that $\mathbf{y} - \mathbf{X}\hat{\alpha}$ and $\hat{\alpha}$ are independent (Diggle et al., 1994). Since the full residuals are a linear combination of $\mathbf{y}$, the joint likelihood for α and the variance parameters γ, can be expressed as a product of the likelihoods based on $\mathbf{y} - \mathbf{X}\hat{\alpha}$ and $\hat{\alpha}$:

$$L(\gamma, \alpha; \mathbf{y}) = L(\gamma; \mathbf{y} - \mathbf{X}\hat{\alpha}) L(\alpha; \hat{\alpha}, \gamma). \tag{2.54}$$

Thus,

$$L(\gamma;\mathbf{y}-\mathbf{X}\hat{\alpha})=\frac{L(\gamma,\alpha;\mathbf{y})}{L(\alpha;\hat{\alpha},\gamma)}. \tag{2.55}$$

From equation (2.52),

$$L(\gamma,\alpha;\mathbf{y})\propto|\mathbf{V}|^{-1}\exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\alpha)'\,\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\alpha)\right\}. \tag{2.56}$$

Furthermore, $\hat{\alpha}$ has a multivariate normal distribution with mean and variance given by the maximum likelihood estimated described in equations (2.61) and (2.62), respectively. Thus,

$$L(\alpha;\hat{\alpha},\gamma)\propto\left|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right|^{1/2}\exp\left\{\frac{1}{2}(\hat{\alpha}-\alpha)'\,\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\hat{\alpha}-\alpha)\right\}. \tag{2.57}$$

Then, taking the ratio of equations (2.56) and (2.57) yields the REML likelihood:

$$L(\gamma;\mathbf{y}-\mathbf{X}\hat{\alpha})\propto\left|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right|^{-1/2}|\mathbf{V}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\hat{\alpha})'\,\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\hat{\alpha})\right\}. \tag{2.58}$$

The REML log-likelihood is then given by

$$\log L(\gamma;\mathbf{y}-\mathbf{X}\hat{\alpha})=K-\frac{1}{2}\left\{\log|\mathbf{V}|-\log\left|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right|^{-1}+(\mathbf{y}-\mathbf{X}\hat{\alpha})'\,\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\hat{\alpha})\right\}. \tag{2.59}$$

In equation (2.59), notice that although $\hat{\alpha}$ appears in the expression for the log-likelihood, it only does so as a function of the variance parameters, since it is an estimate. The variance parameters are then estimated by maximizing the REML log-likelihood given in equation (2.59), with respect to the variance parameters. The resulting variance parameter estimates are unbiased because the REML likelihood uses the fact that the fixed-effects are parameters and not constants. The fixed-effects parameters are then

estimated by treating the variance parameters as fixed and maximizing the REML log-likelihood with respect to the fixed-effects parameters.

2.6.2 Estimation of Fixed-effects Parameters

In order to obtain the fixed-effects estimates, the log-likelihood (or REML log-likelihood) must be maximized by differentiating the log-likelihood (or REML log-likelihood), with respect to α, and setting the resulting expression equal to zero. That is,

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{y}-\mathbf{X}\alpha)=\mathbf{0} \tag{2.60}$$

The solutions to this equation are the maximum likelihood estimates for the fixed-effects parameters:

$$\hat{\alpha}=\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \tag{2.61}$$

The variance of $\hat{\alpha}$ is given by

$$\begin{aligned}\operatorname{var}(\hat{\alpha}) &= \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\operatorname{var}(\mathbf{y})\mathbf{V}^{-1}\mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1} \\ &= \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{V}\mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}. \\ &= \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\end{aligned} \tag{2.62}$$

The variance of $\hat{\alpha}$ will show some downward bias, although very small, because it is based on an assumption that $\mathbf{V}$ is known (Brown and Prescott, 1999).

2.6.3 Estimation of Random Effects Parameters

The vector of random-effects, β, is assumed to follow a normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{G}$, $\beta \sim N(\mathbf{0},\mathbf{G})$. The specific values of the random-effects are taken to be realizations of a sample from a distribution. The expected values of the random-effects are then zero. It is possible, however, to obtain "estimates" of the random-effects, β, referred to as the predicted values of the random-effects, by defining a joint likelihood in terms of α, β, and γ. To do this we take the product of the likelihoods for $\mathbf{y}\,|\,\beta$ and β,

$$L(\alpha,\beta,\gamma;\mathbf{y}) = L(\alpha,\gamma_{\mathbf{\Sigma}};\mathbf{y}\,|\,\beta)\,L(\gamma_{\mathbf{G}};\beta). \tag{2.63}$$

The likelihood for β is written as

$$L(\gamma_{\mathbf{G}};\beta) = (2\pi)^{-q/2}\,|\mathbf{G}|^{-1/2}\exp\{-\frac{1}{2}(\beta)'\,\mathbf{G}^{-1}(\beta)\}, \tag{2.64}$$

and the likelihood for $\mathbf{y}\,|\,\beta$ is given by

$$L(\alpha,\gamma_{\mathbf{\Sigma}};\mathbf{y}\,|\,\beta) \propto |\mathbf{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\alpha-\mathbf{Z}\beta)'\,\mathbf{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\alpha-\mathbf{Z}\beta)\}. \tag{2.65}$$

Thus, the joint likelihood for α, β, and γ can be written as

$$\begin{aligned} L(\alpha,\beta,\gamma;\mathbf{y}) \propto\ & |\mathbf{\Sigma}|^{-1/2}\,|\mathbf{G}|^{-1/2} \\ & \exp\left\{-\frac{1}{2}\left[(\mathbf{y}-\mathbf{X}\alpha-\mathbf{Z}\beta)'\,\mathbf{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\alpha-\mathbf{Z}\beta)+\beta'\mathbf{G}^{-1}\beta\right]\right\}, \end{aligned} \tag{2.66}$$

and the corresponding log-likelihood function can be expressed as

$$\begin{aligned} l(\alpha,\beta,\gamma;\mathbf{y}) \propto\ & -\frac{1}{2}\{\log|\mathbf{\Sigma}|+\log|\mathbf{G}|\}- \\ & \frac{1}{2}\left\{(\mathbf{y}-\mathbf{X}\alpha-\mathbf{Z}\beta)'\,\mathbf{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\alpha-\mathbf{Z}\beta)+\beta'\mathbf{G}^{-1}\beta\right\}. \end{aligned} \tag{2.67}$$

To determine the maximum likelihood solution for β, we take the derivative of the log-likelihood (or REML log-likelihood), with respect to β, and set the resulting expression equal to zero. That is,

$$\begin{aligned}\frac{\partial l(\alpha,\beta,\gamma;\mathbf{y})}{\partial\beta} &= \mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\alpha-\mathbf{Z}\beta)-\mathbf{G}^{-1}\beta \\ &= \mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\alpha)-\left(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\right)\beta.\end{aligned} \tag{2.68}$$

Setting this equation equal to zero and using the fact that $\mathbf{V}=\mathbf{ZGZ}'+\boldsymbol{\Sigma}$, yields the maximum likelihood solution for β,

$$\begin{aligned}\hat{\beta} &= \left(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\right)^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\alpha) \\ &= \left[\left(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\right)^{-1}\left(\boldsymbol{\Sigma}\mathbf{Z}'^{-1}\right)^{-1}\right](\mathbf{y}-\mathbf{X}\alpha) \\ &= \left[\left(\boldsymbol{\Sigma}\mathbf{Z}'^{-1}\right)\left(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\right)\right]^{-1}(\mathbf{y}-\mathbf{X}\alpha) \\ &= \left[\mathbf{Z}+\boldsymbol{\Sigma}\mathbf{Z}'^{-1}\mathbf{G}^{-1}\right]^{-1}(\mathbf{y}-\mathbf{X}\alpha) \\ &= \left[(\mathbf{V}-\boldsymbol{\Sigma})\mathbf{Z}'^{-1}\mathbf{G}^{-1}+\boldsymbol{\Sigma}\mathbf{Z}'^{-1}\mathbf{G}^{-1}\right]^{-1}(\mathbf{y}-\mathbf{X}\alpha) \\ &= \left[\mathbf{V}\mathbf{Z}'^{-1}\mathbf{G}^{-1}\right]^{-1}(\mathbf{y}-\mathbf{X}\alpha) \\ &= \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\alpha).\end{aligned} \tag{2.69}$$

The variance of $\hat{\beta}$ is given by

$$
\begin{aligned}
\operatorname{var}\left(\hat{\beta}\right) &= \mathbf{GZ'V}^{-1} \operatorname{var}\left(\mathbf{y}-\mathbf{X}\alpha\right)\mathbf{V}^{-1}\mathbf{ZG} \\
&= \mathbf{GZ'V}^{-1} \operatorname{var}\left[\mathbf{y}-\mathbf{X}\left(\mathbf{X'V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X'V}^{-1}\mathbf{y}\right]\mathbf{V}^{-1}\mathbf{ZG} \\
&= \mathbf{GZ'V}^{-1} \operatorname{var}\left\{\left[\mathbf{I}-\mathbf{X}\left(\mathbf{X'V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X'V}^{-1}\right]\mathbf{y}\right\}\mathbf{V}^{-1}\mathbf{ZG} \\
&= \mathbf{GZ'V}^{-1}\left[\mathbf{I}-\mathbf{X}\left(\mathbf{X'V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X'V}^{-1}\right]\mathbf{V}\left[\mathbf{I}-\mathbf{V}^{-1}\mathbf{X}\left(\mathbf{X'V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X'}\right]\mathbf{V}^{-1}\mathbf{ZG} \\
&= \left[\mathbf{GZ'}-\mathbf{GZ'V}^{-1}\mathbf{X}\left(\mathbf{X'V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X'}\right]\mathbf{V}^{-1}\left[\mathbf{ZG}-\mathbf{X}\left(\mathbf{X'V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X'V}^{-1}\mathbf{ZG}\right] \\
&= \mathbf{GZ'V}^{-1}-\mathbf{GZ'V}^{-1}\mathbf{X}\left(\mathbf{X'V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X'V}^{-1}\mathbf{ZG}.
\end{aligned}
\tag{2.70}
$$

The variance of $\hat{\beta}$ will be biased downwards due to the fact that $\mathbf{V}$ is assumed to be known (Brown and Prescott, 1999). However, this bias is small.

2.6.4 Estimation of Variance Parameters

The variance parameters are also obtained by maximizing the log-likelihood function. The derivatives of the log-likelihood functions with respect to the variance parameters, however, are nonlinear. Thus, an iterative approach, such as the Newton-Raphson algorithm, is often used to find the maximum likelihood solutions. Variance and covariance estimates of the variance parameters for any given structure can also be obtained by using large sample theory. These estimates are based on asymptotic theory and should be interpreted with caution.

2.7. Estimation in the Normal-Mixture Model

In this section, the maximum-likelihood methods used for estimation in the normal-mixture model are described. A mixture model arises when measurements collected belong to a series of classes, but whose individual class membership is unavailable to the researcher. If this membership were known, then a variable, indicating class, could be included as a fixed-effect in the model. Without this information, the definition of the likelihood equation is considerably more complex. After definition of the likelihood equation, the usual approach proceeds by iteratively solving the maximum likelihood equations using Newton-Raphson. This approach, however, is tedious, because it requires first and second derivatives of a likelihood involving matrices. Methods to obtain estimates of the parameters in the normal-mixture model could be simplified by using the Expectation-Maximization (EM) algorithm, which treats the membership information as missing and, hence, treats the data as incomplete. The algorithm iterates by completing a series of steps, called E-steps and M-steps, in order, until the process converges to the maximum likelihood estimates of the parameters. In the E-step, the "incomplete" data are estimated so that a "complete" data set can be defined. The M-step, then, maximizes the likelihood with respect to the parameters of the normal-mixture model, given the complete data. This maximization is often much simpler and current methods for estimation exist. In section 2.7.1, an EM-algorithm for fitting data that arise from an M-component normal-mixture distribution will be described. In section 2.7.2, specifics involving incorporating the fixed-effects into the model are detailed.

2.7.1 EM Algorithm for the *M*-Component Normal-Mixture Model

Suppose that a random sample of N observations is obtained from an M-component normal-mixture density, $i = 1, \cdots, N$,

$$f\left(\varepsilon_i; \mu_1, ..., \mu_M, \sigma_1^2, ..., \sigma_M^2, \lambda_1, ..., \lambda_{M-1}\right) = \sum_{k=1}^{M} \lambda_k \phi\left(\varepsilon_i; \mu_k, \sigma_k^2\right), \tag{2.71}$$

where $\phi\left(\varepsilon_i; \mu_k, \sigma_k^2\right)$ is standard notation for a normal density with mean μ_k and variance σ_k^2. The likelihood for the model parameters, given the N observation, is given by the joint density of the sample:

$$L\left(\mu_1, ..., \mu_M, \sigma_1^2, ..., \sigma_M^2, \lambda_1, ., ., ., \lambda_{M-1}; \boldsymbol{\varepsilon}\right) = \prod_{i=1}^{N} \sum_{k=1}^{M} \lambda_k \phi\left(\varepsilon_i; \mu_k, \sigma_k^2\right). \tag{2.72}$$

Notice that the observations here are denoted by ε_i rather than by y_i, in order to incorporate this notation into the model containing random-effects in section 2.8. To obtain the maximum likelihood estimates, the log-likelihood,

$$l\left(\lambda, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2; \boldsymbol{\varepsilon}\right) = \sum_{i=1}^{N} \log\left[\sum_{k=1}^{M} \lambda_k \phi\left(\varepsilon_i; \mu_k, \sigma_k^2\right)\right], \tag{2.73}$$

is maximized, with respect to the parameters λ_k, μ_k, and σ_k^2. The maximum likelihood estimates are found by taking the first partial derivatives of the log-likelihood in equation (2.73), with respect to the particular parameters of interest, setting them equal to zero,

and solving for the parameters. Under the constraint that $\lambda_M = 1 - \sum_{k=1}^{M-1} \lambda_k$, the first partial derivatives are given below in equations (2.74)- (2.76) (Hasselblad, 1966):

$$\frac{\partial l}{\partial \lambda_k} = \sum_{i=1}^{N} \left[\frac{\phi\left(\varepsilon_i; \mu_k, \sigma_k^2\right) - \phi\left(\varepsilon_i; \mu_M, \sigma_M^2\right)}{\sum_{m=1}^{M} \lambda_m \phi\left(\varepsilon_i; \mu_m, \sigma_m^2\right)} \right], k = 1, \ldots, M-1 \tag{2.74}$$

$$\frac{\partial l}{\partial \mu_k} = \sum_{i=1}^{N} \left[\frac{\lambda_k \phi\left(\varepsilon_i; \mu_k, \sigma_k^2\right) \left(\frac{\varepsilon_i - \mu_k}{\sigma_k^2} \right)}{\sum_{m=1}^{M} \lambda_m \phi\left(\varepsilon_i; \mu_m, \sigma_m^2\right)} \right], k = 1, \ldots, M \tag{2.75}$$

$$\frac{\partial l}{\partial \sigma_k^2} = \sum_{i=1}^{M} \left[\frac{\lambda_k \phi\left(\varepsilon_i; \mu_k, \sigma_k^2\right) \left(\frac{(\varepsilon_i - \mu_k)^2}{\sigma_k^2} - 1 \right)}{\sum_{m=1}^{M} \lambda_m \phi\left(\varepsilon_i; \mu_m, \sigma_m^2\right)} \right], k = 1, \ldots, M \tag{2.76}$$

As described earlier, the solutions to these equations are easily determined by using the EM-algorithm

In the EM-algorithm, the first step is to define the incomplete data (Dempster et al., 1977). After defining the incomplete data, each iteration of the algorithm involves an expectation step (E-Step) followed by a maximization step (M-Step). The algorithm continues until the process converges. For an M-component normal-mixture model, the

incomplete data are single indicator variables assigning each observation to one of the M-groups. This can be written as

$$c_{ik} = \begin{cases} 1, \text{ if } \varepsilon_i \in k^{th} \text{ group} \\ 0, \text{ otherwise} \end{cases}. \tag{2.77}$$

These indicator variables have distributions specified by their posterior probabilities, $\pi_{ik} = P\left(c_{ik} = 1 | \varepsilon_i\right)$. The complete data, then, are comprised of the observed data and the incomplete data, $\left(\varepsilon_i \quad c_{i1} \quad c_{i2} \quad \cdots \quad c_{iM}\right)$. Since the c_{ik} can be determined only through the posterior π_{ik}, the complete data can also be defined as the $\left(\varepsilon_i \quad \pi_{i1} \quad \pi_{i2} \quad \cdots \quad \pi_{iM}\right)$. In each E-Step, the complete data are obtained by the expected values of the c_{ik}, given the observed data. These estimates represent the probabilities of the i^{th} observation falling into the k^{th} group. The posterior probability is found by taking a weighted average over the M components. That is,

$$\mathrm{E}\left[c_{ik} | \varepsilon_i\right] = P\left(c_{ik} = 1 | \varepsilon_i\right) = \pi_{ik} = \frac{\lambda_k \phi\left(\varepsilon_i; \mu_k, \sigma_k^2\right)}{\sum_{k=1}^{M} \lambda_k \phi\left(\varepsilon_i; \mu_k, \sigma_k^2\right)}. \tag{2.78}$$

The M-step of the EM-algorithm, then, uses the complete data to estimate the parameters of the distribution using maximum likelihood techniques. The estimates in equations (2.78) are used to simplify the maximum likelihood estimates for the parameters. The maximum likelihood estimates of the parameters, given the complete data, for the M-component normal-mixture model, are given in equations (2.79) - (2.81). The details of

their derivations can be found in Appendix 7.1. The maximum likelihood estimate of the proportion of total observations falling into the k^{th} group, λ_k, is given by

$$\hat{\lambda}_k = \frac{\sum_{i=1}^{N} \pi_{ik}}{N}, k = 1,...,M-1. \tag{2.79}$$

The MLE of the proportion falling into the M^{th} group is then $\hat{\lambda}_M = 1 - \sum_{k=1}^{M-1} \hat{\lambda}_k$. The maximum likelihood estimate for the mean of the k^{th} group, μ_k, is

$$\hat{\mu}_k = \frac{\sum_{i=1}^{N} \varepsilon_i \pi_{ik}}{\sum_{i=1}^{N} \pi_{ik}}, k = 1,...,M, \tag{2.80}$$

and the maximum likelihood estimate for the variance of the k^{th} group, σ_k^2, is

$$\hat{\sigma}_k^2 = \frac{\sum_{i=1}^{N} \pi_{ik} (\varepsilon_i - \hat{\mu}_k)^2}{\sum_{i=1}^{N} \pi_{ik}}, k = 1,...,M. \tag{2.81}$$

Each iteration of the EM-algorithm involves computing equations (2.78) - (2.81), in order. Since λ_k, μ_k, and σ_k^2 are unknown for the first iteration, starting values need to be specified to determine the π_{ik}. In subsequent iterations, maximum likelihood estimates produced in the M-step will be substituted to determine the π_{ik}. The algorithm

continues until the process converges. These steps are summarized below, with superscripts in parenthesis to denote the iteration number.

First Iteration ($t = 1$):

- E -step: Starting values for $\lambda_k^{(0)}$, $\mu_k^{(0)}$, and $\sigma_k^{2(0)}$ are specified to compute $\pi_{ik}^{(1)}$ with equation (2.78).
- M -step: The maximum likelihood estimates for $\lambda_k^{(1)}$, $\mu_k^{(1)}$, and $\sigma_k^{2(1)}$ are computed in order with equations (2.79) - (2.81) using $\pi_{ik}^{(1)}$.

Iterations greater than 1 ($t > 1$)

- E -step: The estimates, $\pi_{ik}^{(t+1)}$, are computed with equation (2.78) using $\lambda_k^{(t)}$, $\mu_k^{(t)}$, and $\sigma_k^{2(t)}$.
- M -step: The maximum likelihood estimates for $\lambda_k^{(t+1)}$, $\mu_k^{(t+1)}$, and $\sigma_k^{2(t+1)}$ are computed in order with equations (2.79) - (2.81) using $\pi_{ik}^{(t+1)}$.

At the end of each of the iterations, a check for convergence is performed. Convergence can be checked in a number of ways. One such check is to determine the differences in the parameter estimates between the previous iteration and the current iteration. If the squared sum of the absolute values in the differences is smaller than some predefined tolerance level, then convergence has been achieved and the algorithm ends. Otherwise the algorithm continues with the next iteration. Another check is to use the differences in

the π_{ik} between iterations, however this may be too restrictive and can result in longer convergence times.

2.7.2 Extensions with Fixed-Effects

As shown in previous sections, extensions with fixed-effects are derived by modeling the component means for each observation, ε_i, $i = 1,...,N$ as

$$\mu_{ik} = \alpha_{1k}x_{i2} + \alpha_{2k}x_{i2} + \cdots + \alpha_{Pk}x_{iP}, \tag{2.82}$$

where α_{pk} is the p^{th} fixed-effect parameter, $p = 1,...,P$, in the k^{th} component, $k = 1,...,M$, and x_{ip} is the corresponding observed design variable of the p^{th} fixed-effects parameter for the i^{th} observation. Here, note that the design elements for each observation remain the same across components. Let $\boldsymbol{\varepsilon}$ be the $N \times 1$ vector of observed values, $\boldsymbol{\alpha}_k$ be the $P \times 1$ vector of fixed-effect parameters for the k^{th} component, and $\mathbf{X}$ be the $N \times P$ observed design matrix corresponding to the fixed-effects. In matrix notation,

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}, \ \boldsymbol{\alpha}_k = \begin{pmatrix} \alpha_{1k} \\ \alpha_{2k} \\ \vdots \\ \alpha_{Pk} \end{pmatrix}, \text{ and } \mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1P} \\ x_{21} & x_{22} & \cdots & x_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NP} \end{pmatrix}. \tag{2.83}$$

The vector of fixed-effects parameters, $\boldsymbol{\alpha}_k$, within a component, includes one intercept parameter, α_{1k}, and $P - 1$ fixed-effects parameters, α_{2k}, ..., α_{Pk}. The likelihood for the two-component normal-mixture model parameters, given the observed data, is

$$L\left(\mathbf{X}\alpha_1,...,\mathbf{X}\alpha_M,\sigma_1^2,...,\sigma_M^2,\lambda_1,...,\lambda_{M-1};\varepsilon\right)=\prod_{i=1}^{N}\sum_{k=1}^{2}\lambda_k\phi\left(\varepsilon_i;\mu_{ik},\sigma_k^2\right). \quad (2.84)$$

The maximum likelihood estimates of the parameters in (2.84) can be obtained by taking the first derivatives of the log of the likelihood above, with respect to each of the parameters, equating them to zero, and solving for the parameters. However, the likelihood, in equation (2.84), is expressed as a product of a sum, which does not simplify in the log form. Therefore, the maximum likelihood equations are non-linear and a Newton-Raphson type of iterative algorithm needs to be applied. As described in the previous section, use of an EM-algorithm could simplify the maximum likelihood estimation process considerably. It should be noted that the incomplete part of the data, defined by the z_{ik}, appear in the procedure only through their posterior probabilities, π_{ik}. Essentially, if one defines the posterior probabilities divided by the sums of the posterior probabilities, within a component, as component weights of the corresponding observations, then the component weighted observations follow an N-dimensional joint normal density. The means of the distributions for the component weighted observations are the same as the component means and the variances of these normal distribution are N times the component variances. To further establish this, it is shown that the maximum likelihood estimates of the parameters, based on the likelihoods given the component weighted data, are constant factors of those obtained in the previous section.

Let w_{ik} be defined as the weight for the i^{th} observation, within the k^{th} component, computed by weighting the incomplete data, π_{ik}, by the sum of the incomplete data within the k^{th} component. That is,

$$w_{ik} = \frac{\pi_{ik}}{\sum_{i=1}^{N} \pi_{ik}} = \frac{\pi_{ik}}{\lambda_k N}. \tag{2.85}$$

The denominator in the above equation is equal to the expected proportion of the total sample size for the k^{th} component. By definition, the w_{ik} sum to one across the observations within a component, $\sum_{i=1}^{N} w_{ik} = 1$. Using these weights, 2 component $N \times N$ diagonal weight matrices are constructed, with the component weights along the diagonals and zeros on the off-diagonals,

$$\mathbf{W}_k = diag\left(w_{1k}, w_{2k}, \cdots, w_{Nk}\right). \tag{2.86}$$

Next, 2 separate sets of weighted observations are produced by multiplying the observations by the square roots of the weight matrices defined for each of the components, $\boldsymbol{\varepsilon}_k^* = \mathbf{W}_k^{1/2}\boldsymbol{\varepsilon}$. These sets of weighted observations will be refereed to as the component weighted observations, since they are functions of weights determined by specific components. The likelihood for the model parameters, given the k^{th} component weighted observations is

$$L\left(\boldsymbol{\alpha}_k^*, \boldsymbol{\Sigma}_k^*; \boldsymbol{\varepsilon}_k^*\right) = \frac{\exp\left\{-\frac{1}{2\sigma_k^{2*}}\left(\boldsymbol{\varepsilon}_k^* - \mathbf{W}_k^{1/2}\mathbf{X}\boldsymbol{\alpha}_k^*\right)'\left(\boldsymbol{\varepsilon}_k^* - \mathbf{W}_k^{1/2}\mathbf{X}\boldsymbol{\alpha}_k^*\right)\right\}}{(2\pi)^{N/2}\left(\sigma_k^{2*}\right)^{N/2}}$$
$$= \frac{\exp\left\{-\frac{1}{2\sigma_k^{2*}}\left(\boldsymbol{\varepsilon} - \mathbf{X}\boldsymbol{\alpha}_k^*\right)'\mathbf{W}_k\left(\boldsymbol{\varepsilon} - \mathbf{X}\boldsymbol{\alpha}_k^*\right)\right\}}{(2\pi)^{N/2}\left(\sigma_k^{2*}\right)^{N/2}}. \tag{2.87}$$

The log-likelihood then is

$$\log L\left(\boldsymbol{\alpha}_k^*, \boldsymbol{\Sigma}_k^*; \boldsymbol{\varepsilon}_k^*\right) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log\left(\sigma_k^{2*}\right) - \frac{1}{2\sigma_k^{2*}}\left(\boldsymbol{\varepsilon} - \mathbf{X}\boldsymbol{\alpha}_k^*\right)'\mathbf{W}_k\left(\boldsymbol{\varepsilon} - \mathbf{X}\boldsymbol{\alpha}_k^*\right). \tag{2.88}$$

As usual, the maximum likelihood solutions for $\boldsymbol{\alpha}_k^*$ and σ_k^{2*} are found by taking the first derivatives of the log-likelihood in equation (2.88), with respect to each of the parameters, setting the resulting expressions equal to zero, and solving for the parameters. The maximum likelihood estimates for the parameters of the two component model are given in equations (2.89) and (2.91) below. The maximum likelihood estimates of the vectors of fixed-effects parameters for the likelihood, given the weighted observations defined by the k^{th} component, are

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_k^* &= \left(\mathbf{X}'\mathbf{W}_k^{1/2}\boldsymbol{\Sigma}_k^{*-1}\mathbf{W}_k^{1/2}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}_k^{1/2}\boldsymbol{\Sigma}_k^{*-1}\boldsymbol{\varepsilon}_k^* \\ &= \left(\mathbf{X}'\mathbf{W}_k^{1/2}\left(\sigma_k^{2*}\right)^{-1}\mathbf{W}_k^{1/2}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}_k^{1/2}\left(\sigma_k^{2*}\right)^{-1}\mathbf{W}_k^{1/2}\boldsymbol{\varepsilon} \\ &= \left(\mathbf{X}'\mathbf{W}_k\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}_k\boldsymbol{\varepsilon}.\end{aligned} \tag{2.89}$$

When there are no fixed-effects in the model (only the intercept parameter) these reduce to the same estimates derived in section 2.7.1, since $\sum_{i=1}^{N} w_{ik} = 1$,

$$\hat{\mu}_k{}^* = \sum_{i=1}^{N} \varepsilon_i w_{ik} = \left(\sum_{i=1}^{N} w_{ik} \right)^{-1} \sum_{i=1}^{N} \varepsilon_i w_{ik} = \frac{\sum_{i=1}^{N} \varepsilon_i \pi_{ik}}{\sum_{i=1}^{N} \pi_{ik}} = \hat{\mu}_k . \tag{2.90}$$

Therefore, the maximum likelihood estimates for the two component means, obtained by maximizing the likelihood based on the observed normal-mixture data, are the same as the maximum likelihood estimates for the means, obtained by maximizing the two likelihoods based on the k^{th} component weighted normal data. The maximum likelihood estimates of the variance parameters for the likelihood in equation (2.87), given the weighted observations defined by the k^{th} component, are

$$\begin{aligned} \hat{\sigma}_k^{2*} &= \frac{1}{N} \left(\boldsymbol{\varepsilon}_k^* - \mathbf{W}_k^{1/2} \mathbf{X} \boldsymbol{\alpha}_k{}^* \right)' \left(\boldsymbol{\varepsilon}_k^* - \mathbf{W}_k^{1/2} \mathbf{X} \boldsymbol{\alpha}_k{}^* \right) \\ &= \frac{1}{N} \left(\boldsymbol{\varepsilon} - \mathbf{X} \boldsymbol{\alpha}_k{}^* \right)' \mathbf{W}_k \left(\boldsymbol{\varepsilon} - \mathbf{X} \boldsymbol{\alpha}_k{}^* \right). \end{aligned} \tag{2.91}$$

When there are no fixed-effects in the model (only the intercept parameter) these reduces to N times the variance estimates derived in section 2.7.1,

$$N\hat{\sigma}_k^{2*} = \sum_{i=1}^{N} w_{ik} \left(\varepsilon_i - \hat{\mu}_k \right)^2 = \frac{\sum_{i=1}^{N} \pi_{ik} \left(\varepsilon_i - \hat{\mu}_k \right)^2}{\sum_{i=1}^{N} \pi_{ik}} = \hat{\sigma}_k^2 . \tag{2.92}$$

Therefore, the maximum likelihood estimates for the two component variances, obtained by maximizing the likelihood based on the observed normal-mixture data, are the N times the maximum likelihood estimates for the variances, obtained by maximizing the two likelihoods based on the k^{th} component weighted normal data. The steps of the EM-algorithm for maximum likelihood estimation of the parameters, including the fixed-effects, using weights are summarized below.

First Step ($t = 1$):

- E -step: Starting values for $\lambda_k^{(0)}$, $\mu_k^{(0)}$, and $\sigma_k^{2(0)}$ are specified to compute $\pi_{ik}{}^{(1)}$ with equation (2.78). Component weight matrices, $\mathbf{W}_k^{(1)}$, are constructed as in equation (2.86), where the $w_{ik}{}^{(1)}$ are computed with equation (2.85).
- M -step: The maximum likelihood estimates for the $\lambda_k^{(1)}$ are computed with equation (2.79). The maximum likelihood estimates for $\boldsymbol{\alpha}_k^{(1)} = \boldsymbol{\alpha}_k^{*(1)}$, and $\sigma_k^{2(1)} = N\sigma_k^{2*(1)}$ are computed with equations (2.89) and (2.91), respectively, from the normal likelihoods given the weighted data, $\boldsymbol{\varepsilon}^{*(1)} = \mathbf{W}_k^{1/2(1)}\boldsymbol{\varepsilon}$.

Iterations greater than 1 ($t > 1$)

- E -step: The estimates, $\pi_{ik}^{(t+1)}$, are computed with equation (2.78) using $\lambda_k^{(t)}$, $\mu_k^{(t)}$, and $\sigma_k^{2(t)}$. Component weight matrices, $\mathbf{W}_k^{(t+1)}$, are constructed as in equation (2.86), where the $w_{ik}^{(t+1)}$ are computed with equation (2.85).
- M -step: The maximum likelihood estimates for the $\lambda_k^{(t+1)}$ are computed with equation (2.79). The MLEs for $\boldsymbol{\alpha}_k^{(t+1)} = \boldsymbol{\alpha}_k^{*(t+1)}$, and $\sigma_k^{2(t+1)} = N\sigma_k^{2*(t+1)}$ are computed with equations (2.89) and (2.91) from the normal likelihoods given the weighted data, $\boldsymbol{\varepsilon}^{*(t+1)} = \mathbf{W}_k^{1/2(t+1)}\boldsymbol{\varepsilon}$.

As before the EM-algorithm continues iterating through these steps until a check for convergence has been satisfied. In the next section, the EM-algorithm with weights, described here, is further extended to include estimation of the random-effects that might be included in the model.

2.8. Estimation in Joint Normal-Mixture Random-Effects Model

In this chapter thus far, the normal random-effects model defined in section 2.2 was extended to allow for the residuals to follow an *M*-component joint normal-mixture random-effects distribution rather than a (unimodal) joint normal random-effects distribution. This was accomplished in sections 2.3 and 2.4 by redefining the random-effects model seen in equation (2.1) as $\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. For this new model, referred to as the *M*-component joint normal-mixture random-effects model, the vector of residuals are no

longer assumed to have a mean vector of zeros, but rather one determined by the component vectors of fixed-effects, α_k, and the design matrix, $\mathbf{X}$. They are further assumed to have an M-component joint normal-mixture distribution determined by the product of the M-component univariate normal-mixture densities for the individual residuals. This density was defined in equation (2.21). Fixed-effects were incorporated into the model definition by modeling the component means for the residuals and setting $\mu_k = \mathbf{X}\alpha_k$. The random-effects incorporated into the model are random subject effects. As usual for the random subject effects, the observations between subjects are assumed to be independent, while the observations within subjects are assumed to be correlated. Thus, the structure of the variance covariance matrix for the vector of random subjects effects is taken to be $\mathbf{G} = diag\left(\sigma_\delta^2, \sigma_\delta^2, ..., \sigma_\delta^2\right)$, where σ_δ^2 represents the between subjects varaince. The residuals are assumed to be independent, between and within subjects, with component variances given by σ_k^2, thus the structure of the component variance-covariance matrices for the vector of residuals is $\mathbf{\Sigma}_k = diag\left(\sigma_k^2, \sigma_k^2, ..., \sigma_k^2\right)$. In section 2.6 the standard maximum likelihood estimation methods used for normal random-effects models were reviewed. The usual methods for maximum likelihood estimation with the EM-algorithm were covered for the joint normal-mixture model without fixed or random-effects in section 2.7.1. Then, in the following subsection, 2.7.2, extensions with fixed-effects were derived by describing the EM-algorithm with weighted data. This was done so that an EM-algorithm can be applied on the vector of residuals from the model with random-effects. In this section, estimation in the *two-*

component joint normal-mixture random-effects model is described. For simplicity, details are given for the two-component ($M = 2$) case.

2.8.1 Maximum Likelihood Estimation

Recall, from section 2.3, that the two-component joint normal-mixture random-effects model defined for each subject is given by $\mathbf{y}_j = \mathbf{Z}_j\beta_j + \boldsymbol{\varepsilon}_j$ and the model containing all subjects is given by $\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. It is assumed that the vector of random-effects, $\boldsymbol{\beta}$, is $\mathrm{N}(\mathbf{0},\mathbf{G})$ and independent of the vector of residuals, $\boldsymbol{\varepsilon}$, which are assumed to follow a $\mathrm{JMX}(\mathbf{X}\boldsymbol{\alpha}_1,\mathbf{X}\boldsymbol{\alpha}_2,\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2,\boldsymbol{\lambda})$. The distribution of $\boldsymbol{\varepsilon}$ is the two-component joint normal-mixture density with component mean vectors, $\mathbf{X}\boldsymbol{\alpha}_k$, component variance-covariance matrices, $\boldsymbol{\Sigma}_k$, $k = 1,2$, and vectors of mixture proportion, $\boldsymbol{\lambda} = (\lambda_1,\lambda_2,...,\lambda_S)$, given by

$$\phi_{JMX}(\boldsymbol{\varepsilon};\mathbf{X}\boldsymbol{\alpha}_1,\mathbf{X}\boldsymbol{\alpha}_2,\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2,\boldsymbol{\lambda}) = \prod_{j=1}^{S}\prod_{i=1}^{N_j} f\left(\varepsilon_{ij};\mu_{ij1},\mu_{ij2},\sigma_1^2,\sigma_2^2,\lambda_j\right). \qquad (2.93)$$

where $f\left(\varepsilon_{ij};\mu_{ij1},\mu_{ij2},\sigma_1^2,\sigma_2^2,\lambda_j\right)$ is the univariate normal-mixture densities for the residuals defined in equation (2.7). For this model definition, the mixture proportion is allowed to vary for each subject. The special case where the mixture proportion is assumed equal for each subject is discussed in section 2.8.3. For the two-component case, only one of the component mixture proportions needs to be estimated for each subject, since $\lambda_{j1} = \lambda_j$ and $\lambda_{j2} = 1 - \lambda_j$, $j = 1...,S$. The residuals are assumed to be independent, between and within subjects, with equal variances within each of the

components. Thus, the variance-covariance matrices, $\boldsymbol{\Sigma}_k$, are diagonal matrices with the component variance along the diagonal and zeros on the off diagonal, given by $\boldsymbol{\Sigma}_k = diag\left(\sigma_k^2, \sigma_k^2, ..., \sigma_k^2\right)$. In section 2.5, it was shown that the density, and hence the likelihood, for vector of observations, $\mathbf{y}$, can not be derived by extending the arguments in section 2.7.1 used for the vector of residuals, $\boldsymbol{\varepsilon}$, since the observations within the subjects are now assumed to be correlated. The likelihood for the two-component joint normal-mixture random-effects model based on the observations is equal to the joint density (of the marginal densities) of the observations, given by

$$L\left(\alpha_1, \alpha_2, \sigma_1^2, \sigma_2^2, \sigma_\delta^2, \lambda; \mathbf{y}\right) = f(\mathbf{y}) = \int_{-\infty}^{\infty} f(\mathbf{y} \mid \beta) f(\beta) d\beta, \tag{2.94}$$

where the conditional density of the observations given the vector of random-effects is

$$\begin{aligned} f(\mathbf{y} \mid \beta) &= \prod_{j=1}^{S} \prod_{i=1}^{N_j} \left[\sum_{k=1}^{2} \lambda_{jk} \phi\left(y_{ij} \mid \beta_j; \mu_{ijk} + \beta_j, \sigma_k^2\right) \right] \\ &= \prod_{j=1}^{S} \prod_{i=1}^{N_j} \left[\sum_{k=1}^{2} \lambda_{jk} \phi\left(\varepsilon_{ij}; \mu_{ijk}, \sigma_k^2\right) \right], \end{aligned} \tag{2.95}$$

with $\varepsilon_{ij} = y_{ij} - \beta_j$. To estimate the parameters for the two-component joint normal-mixture random-effects model, the likelihood in equation (2.94) should be maximized. Due to its intractable form, the process of maximum likelihood estimation will be carried by using the EM-algorithm with weights and exploiting the fact that the conditional density in equation (2.95) is the two-component joint normal-mixture density. The weights defined in the E-step are used to compute two sets of weighted observations

whose joint densities are determined by the product of the univariate normal random-effects densities for the individual weighted observations. In the M-step, the likelihoods for the weighted observations are then used, along with the distribution of the random-effects, to obtain the overall maximum likelihoods estimates for the joint normal random-effects model, using standard random-effects model theory.

In the next section, it will be examined how these maximum likelihood estimates can lead to estimates of the component means and variances for the two-component joint normal-mixture random-effects model. The maximum likelihood estimate for the mixture proportions is determined by maximizing the conditional density in equation (2.95), with respect to λ_j, since they are completely specified within this density. It will be then be conjectured that the an estimate for the random-effects variance parameter of the two-component joint normal-mixture random-effects model can be determined by computing a weighted average of the two random-effects variance estimates produced from the two joint normal random-effects models. This process of maximum likelihood estimation, using the EM-algorithm with weights, is described in more detail in the following section.

2.8.2 EM-Algorithm for the Joint Normal-Mixture Random-Effects Model

Suppose that a random sample of N observations from S subjects, with N_j, $j = 1,...,S$, observations taken from the j^{th} subject, $N = \sum_{j=1}^{S} N_j$, are obtained from a two-component normal-mixture random-effects density, $i = 1,...,N$, as defined in section 2.3. The

observations are expressed as the sums of the predicted values for the j^{th} random subject-effects and the residuals of the i^{th} observation from the j^{th} subject, $y_{ij} = \beta_j + \varepsilon_{ij}$.

As mentioned in the previous section, the likelihood for the joint density of the observations is intractable. The process of estimation, however, can be carried out by using a weighted EM-algorithm. As usual, the EM-algorithm requires the incomplete data to be defined. For the two-component joint normal-mixture random-effects model, the incomplete data are defined as single indicator variables assigning each observation to one of the two components. This can be written, $k = 1, 2$, as

$$c_{ijk} = \begin{cases} 1, \text{ if } \varepsilon_i \in k^{th} \text{ group} \\ 0, \text{ otherwise} \end{cases}. \tag{2.96}$$

The posterior probabilities of the i^{th} observation from the j^{th} subject falling into the k^{th} component, conditioned on the j^{th} random-effect, is given by the expected values of the c_{ijk}. In other words, $E\left[c_{ijk} \middle| y_{ij}, \beta_j \right]$. This further simplifies to $E\left[c_{ijk} \middle| \varepsilon_{ij} \right]$, where the residuals are $\varepsilon_{ij} = y_{ij} - \beta_j$. This yields

$$E\left[c_{ijk} \middle| \varepsilon_{ij} \right] = P\left(c_{ijk} = 1 \middle| \varepsilon_{ij} \right) = \pi_{ijk} = \frac{\lambda_{jk} \phi\left(\varepsilon_{ij}; \mu_{ijk}, \sigma_k^2 \right)}{\sum_{m=1}^{M} \lambda_{jk} \phi\left(\varepsilon_{ij}; \mu_{ijk}, \sigma_k^2 \right)}. \tag{2.97}$$

For the two component case, the subscript denoting component can be dropped, since $\pi_{ij1} + \pi_{ij2} = 1$. Thus, the incomplete data for the first component can be represented with π_{ij} and the incomplete data for the second component can be represented with $1 - \pi_{ij}$. The incomplete data, defined in equation (2.97), are computed, within each E-

step, by using the estimates produced in the previous M-step. As usual, for the first iteration, starting values are specified for the parameters to determine the incomplete data. The estimates in equation (2.97) are used to obtain the component weights. These weights are designed in order to produce two sets of weighted observations that follow joint normal random-effects densities. The following describes the process for obtaining the weighted observations.

Let w_{ijk} be defined as the k^{th} component weight for the i^{th} observation from the j^{th} subject, conditioned on the j^{th} random-effect, computed by weighting the incomplete data, π_{ijk}, by the sum of the incomplete data within the j^{th} subject and the k^{th} component. That is,

$$w_{ijk} = \frac{\pi_{ijk}}{\sum_{i=1}^{N_j} \pi_{ijk}} = \frac{\pi_{ijk}}{\lambda_{jk} N_j}. \tag{2.98}$$

The denominator, in the above equation, is equal to the expected proportion of the sample size for the j^{th} subject, for the k^{th} component. By definition, within a component, the w_{ijk} sum to one across the observations for the j^{th} subject, $\sum_{i=1}^{N_j} w_{ijk} = 1$. The component weight matrices are then computed as follows. First, for each subject, the component weight matrices are defined as

$$\mathbf{W}_{jk} = diag\left(w_{1jk},...,w_{N_j jk}\right). \tag{2.99}$$

Next, the subject component weight matrices in equation (2.99) are combined to produce the overall component weight matrices,

$$\mathbf{W}_k = diag\left(\mathbf{W}_{1k},...,\mathbf{W}_{Sk}\right). \tag{2.100}$$

Then, two separate sets of weighted observations are obtained by first multiplying the residuals by the corresponding square roots of the weights and then adding the result to the predicted values for the j^{th} random subject effect, $y_{ijk} = \beta_j + \varepsilon_{ij}\sqrt{w_{ijk}}$. With matrices, the vectors of weighted observations are given by

$$\mathbf{Y}_k = \mathbf{Z}\boldsymbol{\beta} + \mathbf{W}_k^{1/2}\boldsymbol{\varepsilon}. \tag{2.101}$$

As argued in section 2.7.2, weighting the residuals is equivalent to assuming that the component weighted residuals follow an N-dimensional joint normal distribution, with mean vector $\mathbf{X}\boldsymbol{\alpha}_k^*$ and variance-covariance matrix $\boldsymbol{\Sigma}_k^* = \mathbf{W}_k^{1/2}\boldsymbol{\Sigma}_k\mathbf{W}_k^{1/2}$. Then, using random-effects model theory, the weighted observations, $\mathbf{Y}_k$, are assumed to follow an N-dimensional joint normal random-effects distribution with mean vector $\mathbf{X}\boldsymbol{\alpha}_k^*$ and variance-covariance matrix $\mathbf{V}_k^* = \mathbf{ZGZ}' + \mathbf{W}_k^{1/2}\boldsymbol{\Sigma}_k\mathbf{W}_k^{1/2}$.

In the M-step, the likelihoods, based on the component weighted observations, given by

$$L\left(\mathbf{X}\boldsymbol{\alpha}_k^*, \mathbf{V}_k^*; \mathbf{Y}_k\right) = \frac{\exp\left\{-\frac{1}{2}\left(\mathbf{Y}_k - \mathbf{X}\boldsymbol{\alpha}_k^*\right)' \mathbf{V}_k^{*-1}\left(\mathbf{Y}_k - \mathbf{X}\boldsymbol{\alpha}_k^*\right)\right\}}{(2\pi)^{N/2}\left|\mathbf{V}_k^*\right|^{1/2}}, \tag{2.102}$$

are maximized, using the methodology described in section 2.6, to obtain the maximum likelihood estimates of the parameters α_k^*, β_k^*, σ_k^{2*} and $\sigma_{\delta k}^{2\,*}$. From section 2.6.2, the maximum likelihood estimates for the vectors of fixed-effects parameters are given by

$$\hat{\alpha}_k^* = \left(\mathbf{X}'\mathbf{W}_k^{1/2}\hat{\mathbf{V}}_k^{*-1}\mathbf{W}_k^{1/2}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{W}_k^{1/2}\hat{\mathbf{V}}_k^{*-1}\mathbf{Y}_k. \tag{2.103}$$

From section 2.6.4, the methods for obtaining the maximum likelihood estimates for the variance parameters, σ_k^{2*} and $\sigma_{\delta k}^{2\,*}$, in the variance-covariance matrix $\mathbf{V}_k^*$ involve iterative types of methods. From section 2.6.3, the predicted values, based on the maximum likelihood estimates, for the vector of random subject-effects are given by

$$\hat{\beta}_k^* = \hat{\mathbf{G}}^*\mathbf{Z}'\hat{\mathbf{V}}_k^{*-1}\left(\mathbf{Y}_k - \mathbf{X}\hat{\alpha}_k^* \right). \tag{2.104}$$

The entire process for obtaining the maximum likelihood estimates described above involves using iterative methods within each M-step, such as the Newton-Raphson algorithm or the EM-algorithm. These can be carried out however with standard methodology for random-effects models included within many statistical software packages.

The relationships between the maximum likelihood estimates obtained by weighting the observations and the ones for the normal-mixture parameters for models not including random-effects were described in section 2.7.2. Using a similar argument, the estimates for the component means determined by the vectors of fixed-effects parameters of the two-component joint normal-mixture random-effects model, α_k, are

equivalent to the mean vectors obtained in equation (2.103) by using the likelihoods based on the component weighted observations, α_k^*. That is,

$$\hat{\alpha}_k = \hat{\alpha}_k^* = \left(\mathbf{X}'\mathbf{W}_k^{1/2}\hat{\mathbf{V}}_k^{*-1}\mathbf{W}_k^{1/2}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{W}_k^{1/2}\hat{\mathbf{V}}_k^{*-1}\mathbf{Y}_k . \qquad (2.105)$$

The estimates for the component variance-covariance matrices for the weighted residuals, $\boldsymbol{\Sigma}_k$, determined by the component variance parameters of the residuals for the two-component joint normal-mixture random-effects model, σ_k^2, are a constant factor of the maximum likelihood estimates for the variance-covariance matrices obtained by using the likelihoods based on the component weighted observations, $\boldsymbol{\Sigma}_k^*$. The need for a constant factor is a consequence of weighting the data. For a design (model), with balance across the subjects, that is, equal number of observations per subject, the factor is $N^* = \frac{N}{S} = N_j, \forall j$. For an unbalanced design across subjects, the factor is conjectured to be $N^* = \frac{N}{S} = \frac{1}{S}\sum_{j=1}^{S} N_j$. It is important to understand that the estimates for $\boldsymbol{\Sigma}_k^*$ will be biased downwards. This is a typical result of the maximum likelihood procedure with random-effects in the model. Using REML to estimate the parameters of the random-effects models within each M-step could result in parameters with less downward bias. Further discussion can be found in section 2.10.

The maximum likelihood estimates for the mixture proportions are determined by maximizing the likelihood for the observations, conditioned on the random-effects, with

respect to λ_j, since the λ_j are completely specified within the density of the residuals, and not within the density of the random-effects. Recall, this likelihood is

$$L(\mathbf{X}\alpha_1, \mathbf{X}\alpha_2, \Sigma_1, \Sigma_2, \lambda; \varepsilon) = \prod_{j=1}^{S}\prod_{i=1}^{N_j}\left[\lambda_j \phi\left(\varepsilon_{ij}; \mu_{ij1}, \sigma_1^2\right) + \left(1-\lambda_j\right)\phi\left(\varepsilon_{ij}; \mu_{ij2}, \sigma_2^2\right)\right]. \tag{2.106}$$

To obtain the maximum likelihood estimates for the vector of subject mixture proportions, λ, the derivative of the log of the likelihood in equation (2.106), with respect to λ_j, is set equal to zero, and the resulting expression is solved for λ_j. The first partial derivative with respect to λ_j, the mixture proportion for the j^{th} subject, $j = 1,...,S$, of the likelihood is

$$\frac{\partial l}{\partial \lambda_j} = \sum_{j=1}^{S}\sum_{i=1}^{N_j}\left[\frac{\phi\left(\varepsilon_{ij}; \mu_{ij1}, \sigma_1^2\right) - \phi\left(\varepsilon_{ij}; \mu_{ij2}, \sigma_2^2\right)}{\lambda_j \phi\left(\varepsilon_{ij}; \mu_{ij1}, \sigma_1^2\right) + \left(1-\lambda_j\right)\phi\left(\varepsilon_{ij}; \mu_{ij2}, \sigma_2^2\right)}\right]. \tag{2.107}$$

Substituting the π_{ij} into equation (2.107) yields,

$$\begin{aligned}\sum_{j=1}^{S}\sum_{i=1}^{N_j}\left[\frac{\pi_{ij}}{\lambda_j} - \frac{\left(1-\pi_{ij}\right)}{\left(1-\lambda_j\right)}\right] &= \sum_{j=1}^{S}\sum_{i=1}^{N_j}\left[\left(1-\lambda_j\right)\pi_{ij} - \lambda_j\left(1-\pi_{ij}\right)\right] \\ &= \sum_{j=1}^{S}\sum_{i=1}^{N_j}\left[\pi_{ij} - \lambda_j\right] \\ &= \sum_{j=1}^{S}\left[\left(\sum_{i=1}^{N_j}\pi_{ij}\right) - N_j\lambda_j\right].\end{aligned} \tag{2.108}$$

Setting equation (2.108) equal to zero,

$$\sum_{j=1}^{S} N_j \lambda_j = \sum_{j=1}^{S} \sum_{i=1}^{N_j} \pi_{ij}. \tag{2.109}$$

The expression in equation (2.109) is equivalent to

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{N_j} \pi_{ij}}{N_j}, j = 1,...,S, \tag{2.110}$$

giving the maximum likelihood estimate for the mixture proportion for the j^{th} subject. The maximum likelihood estimate for the vector of subject mixture proportions, $\boldsymbol{\lambda}$, is then

$$\hat{\boldsymbol{\lambda}} = \left(\hat{\lambda}_1,...,\hat{\lambda}_S\right). \tag{2.111}$$

Since one of the model assumptions for the two-component joint normal-mixture random-effects model is that the variance for the random-effects is equal across components, an overall estimate of σ_δ^2 must be determined. It is proposed that the random-effects variance estimates from the two normal random-effects models, $\hat{\sigma}_{\delta_k}^2$, may be combined by first determining a common β_j from the predicted values of the random-effects, β_{jk}, computed as

$$\hat{\beta}_j = \hat{\lambda}_j \hat{\beta}_{j1} + \left(1 - \hat{\lambda}_j\right)\hat{\beta}_{j2}, \tag{2.112}$$

then calculating the estimate for σ_δ^2 as

$$\hat{\sigma}_{\delta}^{2} = \frac{1}{(S-1)} \sum_{j=1}^{S} \left(\hat{\beta}_j - \bar{\beta}\right)^2, \tag{2.113}$$

with $\bar{\beta}$ denoting the mean of the $\hat{\beta}_j$. After estimating all of the parameters of the two-component joint normal-mixture random-effects model, the maximum likelihood estimates obtained in each M-step are then used in the subsequent E-steps to obtain new values for the incomplete data. The algorithm iterates until the process converges.

The steps for determining the maximum likelihood estimates for the two-component joint normal-mixture random-effects density using the EM algorithm with weights are summarized below for the t^{th} iteration.

First Step (t = 1):

- E -step: Starting values for $\lambda^{(0)}$, $\alpha_k^{(0)}$, $\sigma_k^{2(0)}$, and $\beta^{(0)}$ are specified. The initial $\sigma_{\delta}^{2(0)}$ is computed from $\beta^{(0)}$ with equation (2.113). The residual vector is then determined by $\varepsilon^{(0)} = \mathbf{Y} - \mathbf{Z}\beta^{(0)}$. The $\pi_{ij}^{(1)}$ are computed, using these starting values, with equation (2.97). The component weight matrices, $\mathbf{W}_k^{(1)}$, are constructed from the $\pi_{ij}^{(1)}$, with equations (2.98) - (2.100). Using these component weight matrices, two sets of weighted observations are created as, $\mathbf{Y}_k^{(1)} = \mathbf{Z}\beta^{(0)} + \mathbf{W}_k^{1/2(1)}\varepsilon^{(0)}$. The residuals are assumed to follow a two-component joint normal-mixture density, the weighted residuals are assumed to follow a joint normal density, and the weighted observations are assumed to follow a joint normal random-effects density.

- M -step: The estimates for the component means, component variances, and random-effects variance, are obtained by maximizing the two separate likelihoods, in equation (2.106), based on the two sets of weighted observations. The component vectors of fixed-effects parameters, $\alpha_k^{(1)}$, for the two-component joint normal-mixture random-effects model are equivalent to the vectors of fixed-effects parameters obtained by maximizing the separate likelihoods for the normal random-effects model in (2.102). The estimates for the component variance parameters, $\sigma_k^{2(1)}$, are $N^* = \frac{N}{S}$ times the variance estimates obtained by maximizing the likelihoods for the normal random-effects model in (2.102). The maximum likelihood estimate for the $\lambda^{(1)}$ are computed with equation (2.110) and (2.111) by maximizing the likelihood, in equation (2.106), for the two-component normal-mixture model parameters based on the residuals, $\varepsilon^{(0)}$. Finally, the predicted values for the two vectors of random subject-effects, $\beta_k^{(1)}$, are determined, using equation (2.104), by maximizing the likelihoods, in equation (2.102), for the joint normal random-effects model parameters, given the weighted observations. The overall vector of predicted random subject effects, $\beta^{(1)}$, is defined with equation (2.112), and the overall variance estimate for the random subject-effects, $\sigma_\delta^{2(1)}$, is given by equation (2.113).

Iterations greater than 1 ($t > 1$):

- E -step: The $\pi_{ij}^{(t+1)}$ are determined, with equation (2.97), using the values $\lambda^{(t)}, \alpha_k^{(t)}, \sigma_k^{2(t)}$ obtained in the t^{th} E-step. The component weight matrices, $\mathbf{W}_k^{(t+1)}$, are constructed from the $\pi_{ij}^{(t+1)}$, with equations (2.98) - (2.100). Using these component weight matrices, two sets of weighted observations are created as, $\mathbf{Y}_k^{(t+1)} = \mathbf{Z}\boldsymbol{\beta}^{(t)} + \mathbf{W}_k^{1/2(t+1)}\boldsymbol{\varepsilon}^{(t)}$.

- M -step: The estimates for the component means, component variances, and random-effects variance, are obtained by maximizing the two likelihoods, in equation (2.106), based on the two sets of weighted observations. The estimates for the component vectors of fixed-effects parameters, $\alpha_k^{(t+1)}$, for the two-component joint normal-mixture random-effects model are equivalent to the maximum likelihood estimates for the vectors of fixed-effects parameters obtained by maximizing the likelihoods for the normal random-effects model in (2.102). The estimates for the component variance parameters, $\sigma_k^{2(t+1)}$, are $N^* = \frac{N}{S}$ times the maximum likelihood estimates for the variances obtained by maximizing the likelihoods for the normal random-effects model in (2.102). The maximum likelihood estimate for the $\lambda^{(t+1)}$ are computed with equation (2.110) and (2.111) by maximizing the likelihood, in equation (2.106), for the two-component joint normal-mixture model parameters based on the residuals,

$\varepsilon^{(t)}$. Finally, the predicted values for the two vectors of random subject-effects, $\beta_k^{(t+1)}$, are determined, using equation (2.104), by maximizing the likelihoods, in equation (2.102), for the joint normal random-effects model parameters, given the weighted observations. The overall vector of predicted subject effects, $\beta^{(t+1)}$, is defined with equation (2.112), and the overall variance estimate for the random subject-effects, $\sigma_\delta^{2(t+1)}$, is given by equation (2.113).

As in the previous sections, a check for convergence is performed at the end of each M-step. One such check would be to compute the squared differences between the estimates of the parameters in the t^{th} iteration and t - 1 iteration. If the sum of these squared differences is smaller than a predetermined tolerance level then the process has converged.

2.8.3 A Special Case

A special case of the two-component joint normal-mixture random-effects model is when the mixture proportion is not assumed to vary across subjects. The algorithm described in the previous section can easily be modified to fit this special case. The following describes the necessary changes. First, the weights in equation (2.98) are defined as the incomplete data weighted by the sum of the incomplete data across all subjects,

$$w_{ijk} = \frac{\pi_{ijk}}{\sum_{j=1}^{S}\sum_{i=1}^{N_j}\pi_{ijk}} = \frac{\pi_{ijk}}{\lambda_k N}. \tag{2.114}$$

The denominator, in the equation above, is equal to the expected proportion of the total sample size, for all subjects, for the k^{th} component. Next, the maximum likelihood estimate for the overall mixture proportion is computed as

$$\hat{\lambda} = \frac{\sum_{j=1}^{S}\hat{\lambda}_j}{S} = \frac{\sum_{j=1}^{S}\sum_{i=1}^{N_j}\pi_{ij}}{N}. \tag{2.115}$$

The factor used to compute the estimates for the component variances from the maximum likelihood estimate produced from the likelihoods based on the weighted observations would become $N^* = N$. Finally, the maximum likelihood estimates for the random-effects variance, $\hat{\sigma}^2_{\delta_k}$, from the two normal random-effects models may be combined by first determining an overall vector of predicted values for the random-effects, β_j, from the predicted values of the random-effects, β_{jk}, computed as

$$\hat{\beta}_j = \hat{\lambda}\hat{\beta}_{j1} + \left(1-\hat{\lambda}\right)\hat{\beta}_{j2}, \tag{2.116}$$

then calculating the overall estimate for σ^2_{δ} as

$$\hat{\sigma}^2_{\delta} = \frac{1}{(S-1)}\sum_{j=1}^{S}\left(\hat{\beta}_j - \bar{\beta}\right)^2, \tag{2.117}$$

with $\bar{\beta}$ denoting the mean of the $\hat{\beta}_j$. An alternative estimator could be computed as

$$\hat{\sigma}_{\delta}^{2} = \hat{\lambda}\hat{\sigma}_{\delta_1}^{2} + \left(1-\hat{\lambda}\right)\hat{\sigma}_{\delta_2}^{2}. \qquad (2.118)$$

The EM-algorithm for this special case is demonstrated in chapter 3 with a simulation study. Then, in chapter 4, the EM-algorithm for the general case, where the mixture proportions are allowed to vary across subjects, is used to fit the heart rate data from the Loneliness study. In the next section, test for the parameters are developed using likelihood ratio based methods.

2.9. Significance Tests for the Model Parameters

Most significance testing for the model parameters can be carried out with likelihood ratio tests. A likelihood ratio test compares the likelihoods of models including and excluding the parameters of interest. It is well established that under the null hypothesis, the model parameter has no effect, then twice the difference in the logs of the two maximum likelihoods is asymptotically distributed as chi-squared with degree of freedom equal to the difference in the number of parameters in the two models, denoted by χ_{ν}^{2}. It is important to note that the likelihood ratio test can only compare nested model structures. In situations where the models are not nested, the significance of the model parameters can be determined by comparing the Akaike's information criteria (AIC) or the Schwarz's-Bayesian information criteria (SIC or BIC) (Brown and Prescott, 1999). All significance testing of the component model parameters, with the exception of the mixture proportions, for the M-component joint normal-mixture random-effects model described in this dissertation may be conducted by means of a likelihood ratio test. For tests concerning the mixture proportion, the types of test depends on the question of

interest. For example, testing if the mixture proportions are equal across subjects the model structures are nested and likelihood ratio tests can be performed. When testing if the mixture proportion is equal to 0 or 1, that is testing the number of components in the model, the model structures are not nested, thus AIC or BIC must be used. In this case, this would be a test of goodness of fit of the mixture distribution.

Recall from the previous section, the overall likelihood for the two-component joint normal-mixture random-effects model parameters, given the vector of observations, does not have an explicit form. The EM-algorithm proposed in section 2.8 did not require this definition for maximum likelihood estimation. The likelihood ratio tests using the likelihood for the model parameters, given the vector of observations, would involve indefinite integrals. Thus, methods using a conditional LRT, AIC, and BIC are proposed.

It can be argued that likelihood ratio tests for the fixed-effects may be carried out by means of the likelihoods based on the joint distribution of the residuals. This will be called a conditional likelihood ratio test because the distribution of the residuals assumes the vector of predicted random subject-effects to be fixed. The definition of this likelihood is

$$L(\mathbf{X}\boldsymbol{\alpha}_1, \mathbf{X}\boldsymbol{\alpha}_2, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\lambda}; \boldsymbol{\varepsilon}) = \prod_{j=1}^{S} \prod_{i=1}^{N_j} \left[\lambda_j \phi\left(\varepsilon_{ij}; \mu_{ij1}, \sigma_1^2\right) + \left(1-\lambda_j\right) \phi\left(\varepsilon_{ij}; \mu_{ij2}, \sigma_2^2\right) \right]. \quad (2.119)$$

This argument is made because of the form of the LRT statistic determined from the vector of observations. For example, suppose the vector $\boldsymbol{\theta}$ is defined as the as the vector

containing all of the model parameters, and the vector θ^* as the same vector minus the fixed-effects parameters to be tested for significance,

$$\theta = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_S \\ \alpha_1 \\ \alpha_2 \\ \sigma_1^2 \\ \sigma_2^2 \\ \sigma_\delta^2 \end{pmatrix} \qquad \theta^* = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_S \\ \alpha_1^* \\ \alpha_2^* \\ \sigma_1^2 \\ \sigma_2^2 \\ \sigma_\delta^2 \end{pmatrix}. \tag{2.120}$$

Then the LRT statistic is

$$-2\left[\log L\left(\hat{\theta};\mathbf{y}\right) - \log L\left(\hat{\theta}^*;\mathbf{y}\right)\right] = -2\log\left(\frac{L\left(\hat{\theta};\mathbf{y}\right)}{L\left(\hat{\theta}^*;\mathbf{y}\right)}\right) \tag{2.121}$$

Now, since the likelihoods in both the numerator and the denominator are integrals, over the random-effects parameter space, of the product of the conditional density, $f(\mathbf{y} \mid \beta)$, and the marginal density, $f(\beta)$, the likelihood ration statistic is defined

$$\frac{L(\theta;\mathbf{y})}{L\left(\theta^*;\mathbf{y}\right)} = \frac{\int_{-\infty}^{\infty} f\left(\mathbf{y} \mid \beta;\theta_{(\delta)}\right) f\left(\beta;\sigma_\delta^2\right) d\beta}{\int_{-\infty}^{\infty} f\left(\mathbf{y} \mid \beta;\theta^*\right) f(\beta) d\beta}. \tag{2.122}$$

For tests involving the fixed-effects parameters, the ratio of the conditional densities is proposed:

$$\frac{f\left(\mathbf{y} \mid \beta; \theta_{(\delta)}\right)}{f\left(\mathbf{y} \mid \beta; \theta^{*}\right)} \tag{2.123}$$

The properties of this conditional likelihood ratio statistic need to be further studied; this will be left for future research.

Tests involving the component variances for the two-component joint normal-mixture random-effects model are limited to tests comparing the component variance to each other and may also use the statistic proposed in equation (2.123). Future research into different possible error structures will lead to more types of tests involving the appropriate variance-covariance structure of the model.

The general procedure for testing random-effects parameters in a single component normal random-effects model involves determining a confidence interval on the ratio $\frac{\sigma_{\delta}^2}{\sigma^2 + \sigma_{\delta}^2}$ where σ^2 is the between-subject variance and $\sigma^2 + \sigma_{\delta}^2$ is the total variance. Then a confidence interval containing one would indicates that the variance parameter for the random-effects, σ_{δ}^2, is insignificant. This in turn indicates that the subjects have the same mean vector and the inclusion of random-effects in the model is unnecessary. Recall from section 2.3 that the expression for the variance of the residuals in the two-component joint normal-mixture random-effects model is given by

$$\operatorname{var}\left(\varepsilon_{ij}\right) = \lambda_j \sigma_1^2 + \left(1 - \lambda_j\right)\sigma_2^2 + \lambda_j\left(1 - \lambda_j\right)\left(\mu_{ij1} - \mu_{ij2}\right)^2. \tag{2.124}$$

Then, since the residuals are independent of the random-effects, the variance of the observations is given by

$$\text{var}\left(y_{ij}\right) = \lambda_j \sigma_1^2 + \left(1-\lambda_j\right)\sigma_2^2 + \lambda_j\left(1-\lambda_j\right)\left(\mu_{ij1} - \mu_{ij2}\right)^2 + \sigma_\delta^2. \qquad (2.125)$$

Then a test for the significance of σ_δ^2 is proposed by forming a confidence interval on the ratio of the variance of the residuals to the variance of the observations,

$$\frac{\text{var}\left(\varepsilon_{ij}\right)}{\text{var}\left(y_{ij}\right)}. \qquad (2.126)$$

As before, a confidence interval containing one indicates that the inclusion of random-effects is unnecessary. Forming this confidence interval could be complex, however, since obtaining the standard errors for the variances and covariances is difficult. An alternative method is suggested as follows. In order to test the significance of σ_δ^2 in the two-component joint normal-mixture random-effects model, it is suggested that simultaneous confidence intervals be produced for both components, $\frac{\sigma_k^2}{\sigma_k^2 + \sigma_\delta^2}$. Then if both confidence intervals include the value one, then it will be concluded that the random-effects variance parameters are insignificant. Because this is a simultaneous confidence interval, appropriate adjustments for the multiple comparisons should be made.

The methods for testing, suggested in this section, are demonstrated in chapters 3 with a simulation study and in chapter 4 with the Loneliness data.

2.10. Summary

In this chapter, the M-component joint normal-mixture random-effects model was defined. Since many of the definitions and methodology were borrowed from random-effects model theory as well as normal-mixture model theory, these two types of models and the methods used for estimating their model parameters were reviewed in sections 2.2, 2.6, and 2.7. The methodology used for estimation of the M-component joint normal-mixture random-effects model parameters was formulated by first extending the joint normal-mixture model to include fixed-effects by defining weights determined by the incomplete data. Next, the random-effects were incorporated by further extending the methodology to condition on the random-effects. Although estimation based on the conditional likelihood does introduce some uncertainty into the estimate of variability of the random subject effects, the effect on the remaining model parameters is negligible. In the next two chapters, this model will further be explored. First, however, it is necessary to discuss some of the benefits and limitations of the proposed methodology for estimation in the two-component joint normal-mixture random-effects model.

The most obvious benefit of the algorithm proposed in section 2.8 is its elegance and simplicity. The precise definition of the weights allows us to define weighted observations which are assumed to follow joint normal random-effects densities. The likelihoods, based on the weighted observations are standard and the resulting maximum likelihood solutions are explicitly defined and well-studied. The estimates produced by maximizing these likelihoods then directly lead to the component estimates for the two-

component joint normal-mixture random-effects model parameters through the EM theory. The algorithm is easily implemented in statistical software packages, such as SAS, because the existing random-effects estimation procedures can be utilized within the M-steps.

With regards to estimation procedures, this method is attractive since the alternative is to use Newton-Raphson, which involves indefinite integrals. This can be tedious and the implementation can be difficult. Xu and Hedeker proposed this method of estimation for their random-effects model which assumed the residuals to be normally distributed and the random-effects to be normal-mixture. Although thorough, their work is difficult to implement making it highly unattractive.

The extension for fitting normal-mixtures with fixed-effects is significant in that this is the first attempt at modeling the component means in the mixture density. The algorithm, which is a special case of the EM-algorithm for the two-component joint normal-mixture random-effects model, reduces to the usual estimates obtained for the mixture model when no fixed-effects are included. This proposed EM-algorithm with weights is again easily implemented and now offers researchers and statisticians more flexibility when modeling normal-mixture models with fixed-effects. The extension with random-effects then further allows researchers to include random-effects in their model. Although the random-effects discussed thus far have been subject effects, they could be visit effects for data taken from a single subject across several visits. The inclusion of both a visit and a subject effect is left for future research due to the canonical correlation structure of the variance-covariance matrix for the observations. It is anticipated that the

methodology will be extended by using some type of multivariate density and likelihood definition.

Although the methods proposed in this dissertation have many benefits, there are several limitations worth mentioning. The main limitation, of the proposed method for estimation, is that the algorithm does not maximize the overall likelihood, but rather the conditional likelihood of the observations conditioned on the random-effects. As suggested earlier, the estimation based on the conditional likelihood does introduce some uncertainty into the estimate of variability of the random subject effects, but the effect on the remaining model parameters is negligible. This effect is further explored in the simulation study in the next chapter. Furthermore, the conjectures regarding the overall estimate for the variance of the random-effects has not been proven.

A limit of the theory thus far is the assumption of independence in the residuals. This assumption is too restrictive. The attractiveness of the random-effects model is the ability to fit a wide variety of covariance structure to the residuals, along with the structures for the random-effects. This dissertation has limited the random-effects model with the assumption of independence in the residuals, allowing correlations to enter the model only through the random-effects. The heart rate interval data that inspired this dissertation most definitely violates this independence assumption. Although the applications using this model for these types of data are still unfinished, this dissertation begins the work towards the more appropriate model definition. Until future work into more complex structure is completed, it is suggested that time-series methodology be

applied before the EM-algorithm is used, to remove the lower frequency variations in the data.

In the following two chapters, the proposed method of estimation for the two-component joint normal-mixture random-effects model using the EM-algorithm with weights will be demonstrated. In chapter 3, a simulation study is performed to explore the various conditions under which this method may be utilized. Then, in Chapter 4, the algorithm is applied to the Loneliness data introduced in section 1.5.

3 Simulations

In this chapter, properties of the methodology for estimation proposed in Chapter 2 are studied using simulations. The simulation study is designed to provide an insight into the feasibility of the model, the magnitudes of the bias and standard errors in the estimation of the parameters, under different scenarios defined by the parameter values. The simulation of the data and the subsequent estimation using the EM algorithm proposed in section 2.8 will be performed with SAS v.9.1. The estimates obtained from the EM algorithm will then be examined through plots and factorial analyses using JMP 6.0. The data used within this chapter can be found on the data CD (see Appendix 7.4).

3.1. Simulating Two-Component Normal-Mixture Random Effects Models

In this section, the various factors used in the simulation study, along with the levels of the factors, are defined and the method used to simulate the two-component normal-mixture random-effects model is described. Although the methodology proposed in Chapter 2 is applicable in a wide range of research applications, in this dissertation, HRV data introduced in Chapter 1 is of interest. Therefore, the parameters for the simulated data sets created are derived from actual adult RR-interval (HRV) data. For simplicity, only a two-component normal-mixture distribution with one random subject effect will be considered. It is further assumed that the mixture proportion is equal across all subjects. Define the parameter vector θ to be the vector containing all of the parameters defining the two-component normal-mixture random-effects density. That is,

$$\theta = \left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_\delta^2, \lambda\right)' . \tag{3.1}$$

Then the parameter estimates obtained from the weighted EM-algorithm are examined for θ obtained over a variety of conditions. Primarily, the focus is in studying how changes in the mixture proportion, λ (Factor 1), the percentage overlap between the two components, POV (Factor 2), the magnitude of the between subject variability, σ_δ^2 (Factor 3), the total number of subjects, *S* (Factor 4), and the total number of observations per subject, N_j (Factor 5), affect the bias and mean squared error of the estimators. The levels of these factors are listed below.

- Factor 1: Th ree levels for the mixture proportion, λ, namely 0.5, 0.6, and 0.7.
- Factor 2: T wo levels for the percentage overlap, POV, 10% or 20%.
- Factor 3: T wo levels of the between subject variance factor, BSF, 0.5 or 1.5, indicating when the between subject variability, σ_δ^2, is defined as 0.5 or 1.5 times the total within subject variability,

$$\sigma^2 = \lambda\sigma_1^2 + (1-\lambda)\sigma_2^2 + \lambda(1-\lambda)(\mu_1 - \mu_2)^2 . \tag{3.2}$$

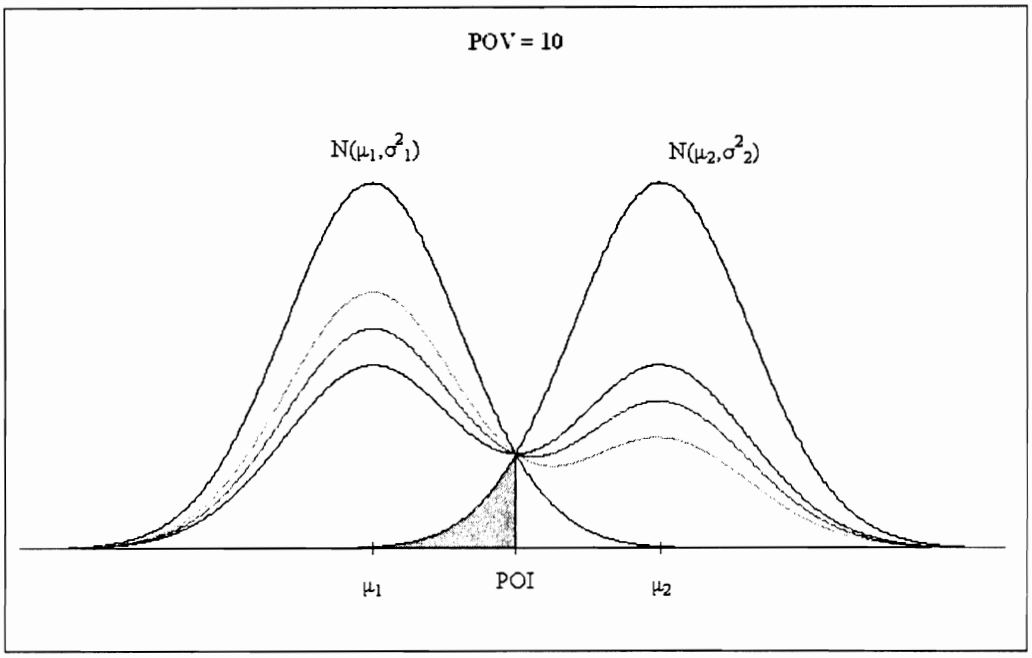
  This factor is described in more detail below.
- Factor 4: Th ree levels for the number of subjects, *S*, namely 5, 10, and 20.
- Factor 5: T wo levels for the number or observations within subject, N_j, namely 100 and 200. Here, N_j is set equal for all subjects to simplify the simulation process.

Twelve different conditions on the first three factors are used to simulate the data for the two-component normal-mixture random-effects density such that:

- The first component me an is fixed arbitrarily at $\mu_1 = 70$.
- The components vari ance, σ_1^2 and σ_2^2, are both fixed at 25.
- Dependin g on the percentage of overlap, 10% or 20%, the second component mean is set to $\mu_2 = 86.44$ or $\mu_2 = 82.82$, respectively, determined by the fixed values of μ_1, σ_1^2, and σ_2^2. This is described in more detail in the following paragraphs.

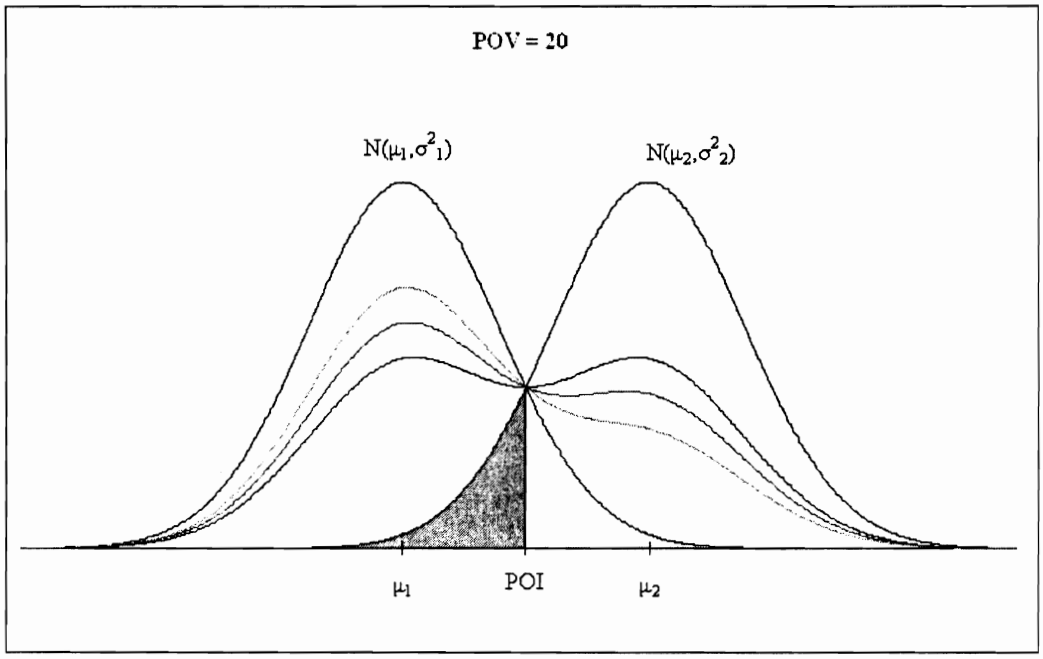
To understand the percent overlap factor, POV, consider the pictures in Figure 4 and Figure 5, using 10% and 20% as the POV, respectively.

Figure 4: 10% Overlap



Several distributions are displayed in each figure. The blue, red, and green lines represent two-component normal-mixture densities with equal component variances $\left(\sigma_1^2 = \sigma_2^2\right)$ and mixture proportions $\lambda = 0.5$, $\lambda = 0.6$, and $\lambda = 0.7$, respectively. The two black lines represent the two mixing normal densities. The light grey shaded region represents the area under the first normal density for values larger than the point of intersection (POI) of the two mixing normal densities. Similarly, the dark grey shaded area represents the area under the second normal density for values smaller than the POI.

Figure 5: 20% Overlap



Without loss of generality, in the simulation study the component variances will be assumed to be equal $\left(\sigma_1^2 = \sigma_2^2\right)$. In that case, POI = $(\mu_1 + \mu_2)/2$. If the component variances are not equal $\left(\sigma_1^2 \neq \sigma_2^2\right)$, then the POI would be a weighted average of the

two component means. When the mixture proportion is 0.5, the overall mixture mean, $\mu = \lambda\mu_1 + (1-\lambda)\mu_2$, coincides with the POI. As the mixture proportion increases from 0.5 this mean will approach the first component mean. Similarly, as the mixture proportion decreases from 0.5, the overall mixture mean will approach the second component mean. The overall mixture mean can be interpreted graphically as a point of inflection in the mixture density. There are two points of inflection in the two-component normal-mixture density that lie between the two component means. The first is determined by the overall mixture mean and the second is determined by $(1-\lambda)\mu_1 + \lambda\mu_2$.

The sum of the light and dark grey regions in either Figure 4 or Figure 5 is defined as the percent overlap (POV). In the simulations, the value of the second component mean is then determined by fixing the value of the first component mean and specifying a percentage of overlap between the two mixing normal distributions. Given μ_1 and σ_1^2, the area shaded in light grey is $1 - P\left(Z \le (\mu_2 - \mu_1)/2\sigma_1^2\right)$. Similarly, given μ_2 and σ_2^2, the area shaded in dark grey is $P\left(Z \le (\mu_1 - \mu_2)/2\sigma_2^2\right)$. Then, given the POV, μ_2 is obtained by solving

$$1 - P\left(Z \le (\mu_2 - \mu_1)/2\sigma_1^2\right) + P\left(Z \le (\mu_1 - \mu_2)/2\sigma_2^2\right) = POV. \qquad (3.3)$$

For the simulations, when $\mu_1 = 70$, $\sigma_1^2 = 25$, and $\sigma_2^2 = 25$ the values $\mu_2 = 86.44$ and $\mu_2 = 82.82$ satisfy the equation in (3.3) for 10% and 20% POV, respectively. As the

percentage of overlap increases, the distance between the two component means decreases.

The factor indicating the between subject variance factor, BSF, represents the two conditions explored for the amount of between subject variability. For example, when BSF = 0.5, the magnitude of the between subject variability, σ_δ^2, is half the magnitude of the total variability exhibited within the subjects, σ^2. Taking $\lambda = 0.5$, $\mu_1 = 70$, $\mu_2 = 86.44$, $\sigma_1^2 = 25$, and $\sigma_2^2 = 25$, the values $\sigma_\delta^2 = 46.28$ and $\sigma_\delta^2 = 138.85$ satisfies the equation in (3.2) for BSF = 0.5 and BSF = 1.5, respectively.

Now, using the five factors described earlier, 12 different parameter vectors are obtained in this fashion. The values obtained for the 12 different parameter vectors are used to simulate the data are summarized in Table 2.

Table 2: Simulation Parameters

$\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_\delta^2, \lambda\right)'$		
	10 % Overlap	20 % Overlap
	$\sigma_\delta^2 = (0.5)\sigma^2$	
$\lambda = 0.5$	$(70, 86.44, 25, 25, 46.28, 0.5)'$	$(70, 82.81, 25, 25, 33.01, 0.5)'$
$\lambda = 0.6$	$(70, 86.44, 25, 25, 44.93, 0.6)'$	$(70, 82.81, 25, 25, 32.19, 0.6)'$
$\lambda = 0.7$	$(70, 86.44, 25, 25, 40.88, 0.7)'$	$(70, 82.81, 25, 25, 29.73, 0.7)'$
	$\sigma_\delta^2 = (1.5)\sigma^2$	
$\lambda = 0.5$	$(70, 86.44, 25, 25, 138.85, 0.5)'$	$(70, 82.81, 25, 25, 99.04, 0.5)''$
$\lambda = 0.6$	$(70, 86.44, 25, 25, 134.80, 0.6)'$	$(70, 82.81, 25, 25, 96.57, 0.6)'$

$\lambda = 0.7$	$(70, 86.44, 25, 25, 122.64, 0.7)'$	$(70, 82.81, 25, 25, 89.19, 0.7)'$

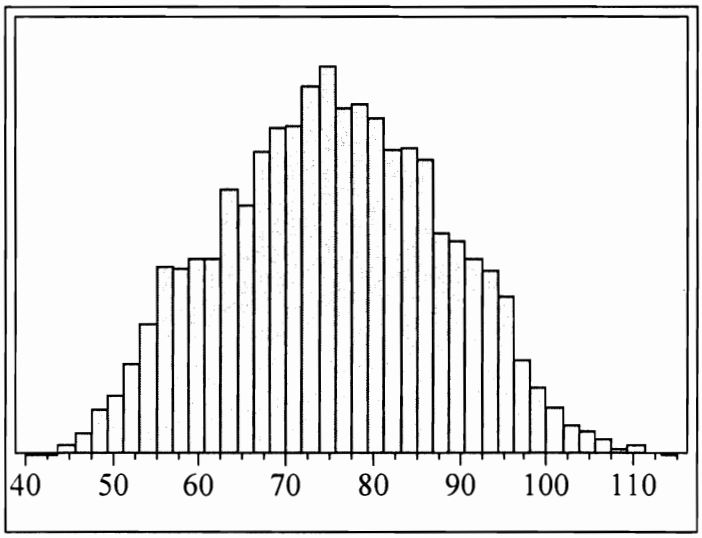
The five factors and various levels lead to a $3^2 \times 2^3$ factorial design with 72 factor combinations. For each of these 72 factor combinations, 100 data sets are then simulated. The methods used to simulate the data are briefly described below. A more thorough explanation can be found in Appendix 7.2 along with the SAS code used to generate the 7200 data sets.

For a specified parameter vector $\theta = \left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_\delta^2, \lambda\right)'$ and vector of simulation factors, $\mathbf{F} = \left(N_j, S\right)'$, a data set is simulated by first generating a random sample of u_{ij}, ε_{ij1}, and ε_{ij2}, $i = 1,...,N_j$, $j = 1,...,S$ from a uniform density, a normal density with mean μ_1 and variance σ_1^2, and a normal density with mean μ_2 and variance σ_2^2, respectively. In addition, to include a random subject effect, the values values, β_j, $j = 1,...,S$, are drawn from a normal density with mean zero and variance σ_δ^2. Then, the random normal-mixture deviates, ε_{ij}, with component means, μ_1 and μ_2, component variances, σ_1^2 and σ_2^2, and mixture proportion, λ, are generated as follows:

$$\varepsilon_{ij} = \begin{cases} \varepsilon_{ij1}, \text{ if } u_{ij} < \lambda \\ \varepsilon_{ij2}, \text{ if } u_{ij} \geq \lambda \end{cases}.$$

Finally, to obtain the observations, y_{ij}, from the two-component normal-mixture random-effects density, the deviates ε_{ij} are added to the random subject effects, β_j. (See Appendix 7.2 for SAS code and further details). An example of a histogram of the simulated data set using the parameter vector $\theta = (70, 86.44, 25, 25, 46.28, 0.5)'$, indicating 10% POV and 0.5 BSF, and the simulation factor vector $\mathbf{F} = (200, 20)'$ is displayed in Figure 6 (specifically the 24th simulated data set from the 100).

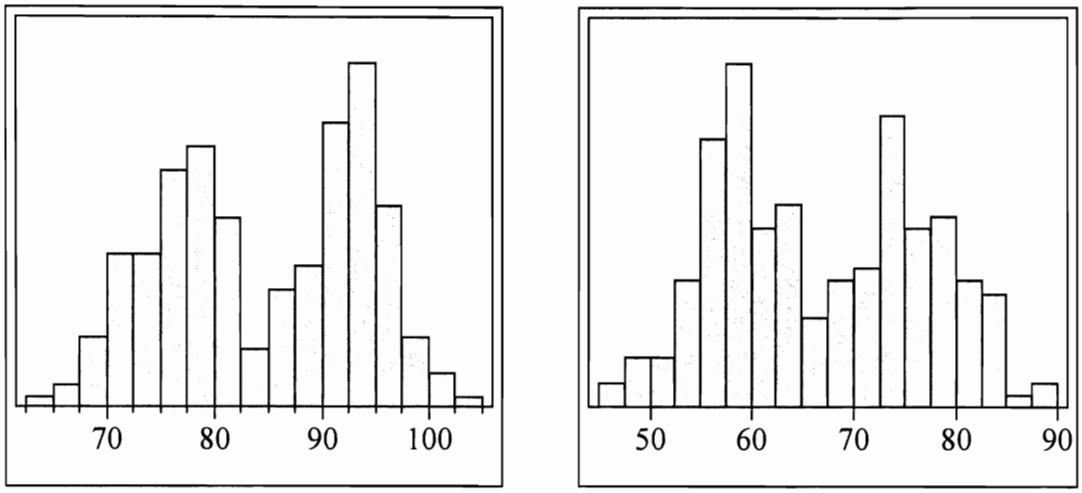
Figure 6: Simulated RRI Data - All Subjects



The histogram in Figure 6 is a good example of how the data can be deceiving when the subject effects are not taken into account. The histogram in Figure 6 contains an example of simulated data from all 20 subjects. A simple glance at the histogram could indicate these data do not come from a mixture of normal densities but instead from a single normal density. However, if the histograms are viewed by subject, a very different story is observed. Histograms of simulated RR-interval data for two separate subjects from the

same data set in Figure 6 are shown in Figure 7. Examination of the two histograms in Figure 7 indicates a mixture density for both subjects with different component means for each subject (indicative of a random subject effect).

Figure 7: Simulated RR-Interval Data – Subject 1 and Subject 17



After simulating the data sets for a particular combination of $\boldsymbol{\theta}$ and $\mathbf{F}$, the proposed EM-algorithm is applied to each of the 100 data sets and the parameter estimates are obtained. This is repeated for each of the 72 combinations in the $3^2 \times 2^3$ factorial design. The EM-algorithm, written in SAS, along with a thorough explanation of the code, can be found in Appendix 7.3. As explained in chapter 2, initial values have to be provided for the EM-algorithm. The following describes the methods used for determining initial values for each of the parameters.

For the mixture proportion, a logical choice for the starting values is 0.5. For the first and second component means, a simple method for obtaining reasonable starting values is to compute the 25% and 75% quartiles of the data, respectively. Starting values for the variance parameters could be obtained by fitting a single-component normal

random-effects model (with a random intercept for each subject). The estimate for the between subject variance provides a decent starting value for the between subject variance in the two-component normal-mixture random-effects model. Then, half of the estimate for the within subject variance could be used as the starting values for both the component variances. Since the EM-algorithm is quite robust in terms of initial values, the above suggestions are reasonable.

Each execution of the EM-algorithm produces a vector of parameter estimates, $\hat{\theta} = \left(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_\delta^2, \hat{\lambda}\right)'$ upon convergence. Appendix 7.4.1 contains a file with all 7200 vectors of parameter estimates. Cumulative sums and sums of squares for the elements of this vector are computed for each combination of simulation factors, over the 100 runs. For each combination, the mean and variance is stored (see Appendix 7.4.2). The bias and MSE of the 72 averaged estimates are then computed. The bias is defined as the difference between the mean (of the 100 simulations) and the true parameter value. The true parameter values are specified by the parameters in the vector θ, given in Table 2. The MSE is then defined as the sum of the squared bias and the variance (of the 100 simulations).

3.2. Presenting Results from the Simulation Study

For each of the parameters estimated by the EM-algorithm, a graph of the squared bias versus the mean squared error (MSE) of the 72 combinations will be plotted. From the graph, several conclusions are then drawn regarding the squared bias and the MSE of the particular parameter being examined. These graphs provide a visual interpretation of the

estimates across the varying conditions. The same scales for the squared bias and the MSE will be used for the two component means (and variances). This will aid in interpreting the estimates across the components. Since there are numerous factor combinations, in order to identify relevant factors, a (single replicate) full-factorial analysis is performed using absolute bias and MSE as the outcome variables. The $3^2 \times 2^3$ full-factorial design includes the overall mean (intercept), the 7 design effects listed above, the 19 two-way interaction effects, the 25 three-way interaction effects, the 16 four-way interaction effects, and the 4 five-way interaction effect, for a total of 72 effects.

Several transformations on the response variables (absolute bias and MSE) are considered for each of the 12 models in order to improve the fit of the models and to reduce violations in the normality and constant variance assumptions for the full-factorial model. A common method for determining the most appropriate transformation is the Box-Cox Procedure which essentially determines the best power transformation of the response variable given the data (Myers and Montgomery, pp 260-264). The (single-replicate) full-factorial model lacks the degrees of freedom to estimates the residual sums of squares necessary for the Box-Cox Procedure. Thus, a practical approach is to drop the 4 five-way interaction effects, assuming their effects to be negligible, and combining their mean squares to estimate the sums of squares for the residuals (Myers and Montgomery, pp 104). These reduced models are then used to determine the best transformations for the response variables. The Box-Cox Procedure, used in JMP, chooses the value λ_{BC} in the interval (-2, 2) which minimizes the residual sums of

squares. This is equivalent to performing a maximum-likelihood procedure to estimate λ_{BC}. Although any real value in the interval (-2, 2) could be estimated, the choices for λ_{BC} will be limited to 0, ½, 1, or 2, relating to the natural log, square root, identity, or the square transformations, respectively. The same transformation for the bias (or MSE) of the two component means (or variances) will be used so that conclusions may be drawn across the components. After determining the best transformation, the full-factorial is fit and the factor effects are examined.

The primary goal of this factorial analysis is to identify factor combinations that are most significantly different from the other factors. Since the saturated (single-replicate) design does not allow tests of significance, instead, a method attributed to Daniel (1959) is suggested. This method suggests plotting the estimates with a normal probability (quantile) plot. Since the effects are still expected to exhibit some assumption violations, the effect estimates are first orthogonalized, to ensure they are uncorrelated, and then standardized, to ensure equal variance. Negligible effects have a standard normal distribution and tend to fall along a straight line. Important effects, however, will exhibit large departures from this straight line.

The software JMP uses a reference cell parameterization when fitting the full-factorial model. This results in 72 combinations of the various factor levels. The reference cell for each factor is the highest level. These are listed in Table 3.

Table 3: Reference Cells

Factor effect	Reference cell
λ	0.7
POV	20
BSF	1.5

S	20
N_j	200

The estimates for the effects are then the estimates for the difference between the effect and the reference cell. For example, the estimate for the effect $\lambda(0.5) \times \text{POV}(10)$ represents the difference between the average response for the case where $\lambda = 0.5$ and POV = 10 and the case where $\lambda = 0.7$ and POV = 20. When interpreting the effects from a model with a reference cell parameterization, conclusions are drawn between the effect and its reference cell.

In section 3.4 the results are presented. This will be done by first presenting the results from the 12 full-factorial models for bias and MSE. The methods for obtaining the results from the full-factorial models will be demonstrated here with the model for the absolute bias of the (estimates of the) mixture proportion, λ. Then, a subsection for each of the six parameters is given. Within each subsection, the scatter plot for the particular parameter of interest is displayed. The graphs and the results from the factorial analyses are then discussed. The results presented in section 3.4 are then interpreted in section 3.5. First, however, a legend used for the scatter plots is described in the next section in detail.

3.3. Legends for the Squared Bias versus MSE Plots

The legend described in this section relates to the six scatter plots seen in section 3.4 (Figure 10, Figure 12, Figure 15, Figure 18, Figure 21, and Figure 24). Each plot represents a separate parameter in the vector $\hat{\theta} = \left(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_\delta^2, \hat{\lambda}\right)'$. For a

particular parameter, the squared bias versus the MSE is displayed for each of the 72 averages of the parameter estimate. Recall, the 72 averages represent the averaged estimates for the 100 runs from each of the 72 varying conditions described in section 3.1. Due to the amount of information contained in each graph, several colors, shapes, and sizes are used in order to distinguish the combinations and highlight their differences. Below is a thorough description of the legend used for these six graphs.

- Data simulated with a tru e mixture proportion of $\lambda = 0.5$ are shown in BLUE.
- Data simulated with a true mixture proportion of $\lambda = 0.6$ are shown in RED.
- Data simulated with a tru e mixture proportion of $\lambda = 0.7$ are shown in GREEN.
- Data simulated with 10% overlap and $\sigma_{\delta}^2 = \frac{1}{2}\sigma^2$ are shown with DIAMONDS.
- Data simulated with 10% overlap and $\sigma_{\delta}^2 = \frac{3}{2}\sigma^2$ are shown with SQUARES.
- Data simulated with 20% overlap and $\sigma_{\delta}^2 = \frac{1}{2}\sigma^2$ are shown with TRIANGLES.
- Data simulated with 20% overlap and $\sigma_{\delta}^2 = \frac{3}{2}\sigma^2$ are shown with CIRCLES.
- Data simulated with 100 observations per subje ct are shown with HOLLOW points.
- Data simulated with 200 observations per subje ct are shown with SOLID points.
- Data simulated with 5 subjects are sho wn with SMALL size points.
- Data simulate d with 10 subjects are shown with MEDIUM size points.
- Data simulated with 20 subjects are sho wn with LARGE size points.

The legend, used for all six plots, is displayed in four parts:

Legend 1 of 4

10% Overlap and BSF = 1/2

- λ = 0.5, S = 5, Nj = 100
- λ = 0.5, S = 5, Nj = 200
- λ = 0.5, S = 10, Nj = 100
- λ = 0.5, S = 10, Nj = 200
- λ = 0.5, S = 20, Nj = 100
- λ = 0.5, S = 20, Nj = 200
- λ = 0.6, S = 5, Nj = 100
- λ = 0.6, S = 5, Nj = 200
- λ = 0.6, S = 10, Nj = 100
- λ = 0.6, S = 10, Nj = 200
- λ = 0.6, S = 20, Nj = 100
- λ = 0.6, S = 20, Nj = 200
- λ = 0.7, S = 5, Nj = 100
- λ = 0.7, S = 5, Nj = 200
- λ = 0.7, S = 10, Nj = 100
- λ = 0.7, S = 10, Nj = 200
- λ = 0.7, S = 20, Nj = 100
- λ = 0.7, S = 20, Nj = 200

Legend 2 of 4

10% Overlap and BSF = 3/2

- λ = 0.5, S = 5, Nj = 100
- λ = 0.5, S = 5, Nj = 200
- λ = 0.5, S = 10, Nj = 100
- λ = 0.5, S = 10, Nj = 200
- λ = 0.5, S = 20, Nj = 100
- λ = 0.5, S = 20, Nj = 200
- λ = 0.6, S = 5, Nj = 100
- λ = 0.6, S = 5, Nj = 200
- λ = 0.6, S = 10, Nj = 100
- λ = 0.6, S = 10, Nj = 200
- λ = 0.6, S = 20, Nj = 100
- λ = 0.6, S = 20, Nj = 200
- λ = 0.7, S = 5, Nj = 100
- λ = 0.7, S = 5, Nj = 200
- λ = 0.7, S = 10, Nj = 100
- λ = 0.7, S = 10, Nj = 200
- λ = 0.7, S = 20, Nj = 100
- λ = 0.7, S = 20, Nj = 200

Legend 3 of 4

20% Overlap and BSF = 1/2

- λ = 0.5, S = 5, Nj = 100
- λ = 0.5, S = 5, Nj = 200
- λ = 0.5, S = 10, Nj = 100
- λ = 0.5, S = 10, Nj = 200
- λ = 0.5, S = 20, Nj = 100
- λ = 0.5, S = 20, Nj = 200
- λ = 0.6, S = 5, Nj = 100
- λ = 0.6, S = 5, Nj = 200
- λ = 0.6, S = 10, Nj = 100
- λ = 0.6, S = 10, Nj = 200
- λ = 0.6, S = 20, Nj = 100
- λ = 0.6, S = 20, Nj = 200
- λ = 0.7, S = 5, Nj = 100
- λ = 0.7, S = 5, Nj = 200
- λ = 0.7, S = 10, Nj = 100
- λ = 0.7, S = 10, Nj = 200
- λ = 0.7, S = 20, Nj = 100
- λ = 0.7, S = 20, Nj = 200

Legend 4 of 4

20% Overlap and BSF = 3/2

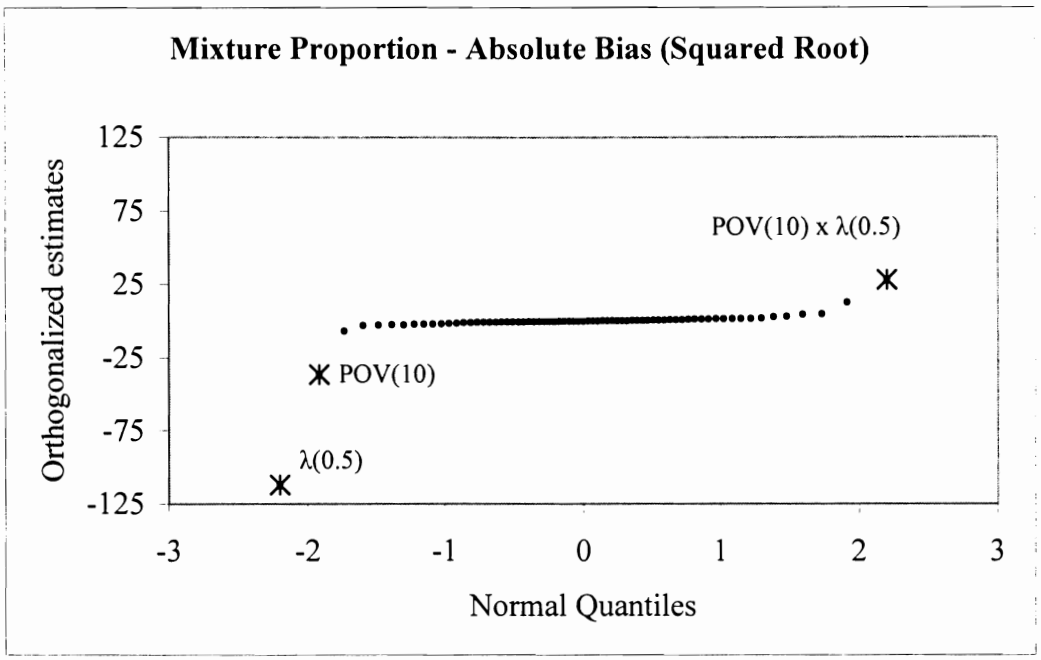
- λ = 0.5, S = 5, Nj = 100
- λ = 0.5, S = 5, Nj = 200
- λ = 0.5, S = 10, Nj = 100
- λ = 0.5, S = 10, Nj = 200
- λ = 0.5, S = 20, Nj = 100
- λ = 0.5, S = 20, Nj = 200
- λ = 0.6, S = 5, Nj = 100
- λ = 0.6, S = 5, Nj = 200
- λ = 0.6, S = 10, Nj = 100
- λ = 0.6, S = 10, Nj = 200
- λ = 0.6, S = 20, Nj = 100
- λ = 0.6, S = 20, Nj = 200
- λ = 0.7, S = 5, Nj = 100
- λ = 0.7, S = 5, Nj = 200
- λ = 0.7, S = 10, Nj = 100
- λ = 0.7, S = 10, Nj = 200
- λ = 0.7, S = 20, Nj = 100
- λ = 0.7, S = 20, Nj = 200

3.4. Results

In this section the results from the simulations are presented for the six parameters estimates in the vector $\theta = \left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_\delta^2, \lambda\right)'$. The estimates of the bias and MSE for all the factor combinations are tabulated. This table is primarily for reference and therefore will be consolidated on the data CD (see Appendix 7.4.2). For facilitating interpretations of the results, scatter plots and normal quantile plots of factor estimates from a full-factorial model are utilized. A separate sub-section for each of the parameters is given so that the all the information could be better organized.

Fitting the full-factorial model is now demonstrated for the bias of the estimate for the mixture proportion, $\hat{\lambda}$. The absolute value of the bias is fit as the response variable (in order to make use of transformations such as the log and square root). The Box-Cox procedure for the reduced factorial model indicates a square-root transformation on the response (absolute bias). The full-factorial model includes 72 combinations of the various levels of the 5 factors discussed in sections 3.1 and 3.2. Using the reference cell parameterization, 72 factor effects are estimated with the model. A normalizing transformation is then applied to the effect estimates to make them uncorrelated (orthogonal) and have equal variance (standardized). The normal quantiles of the normalized effect estimates are determined and then plotted against the normalized effect estimates and shown in the example Figure 8. The normal quantiles along with the untransformed, orthogonalized, and standardized effect estimates are found on the data CD (see Appendix 7.4.3).

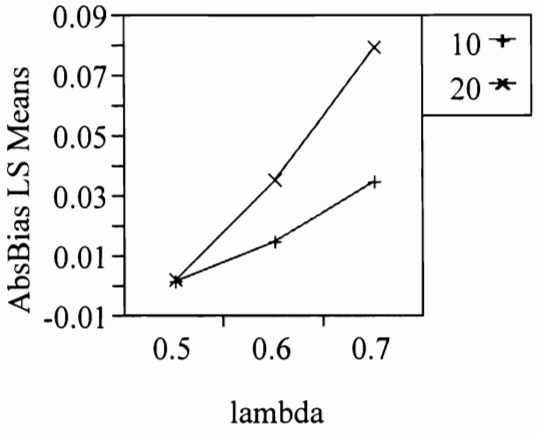
Figure 8: Normal Quantile Plot of Effect Estimates



From the normal-quantile plot in example Figure 8, the important effects that emerge from this analysis are the main effects, $\lambda(0.5)$ and POV(10), and the $\lambda(0.5)$ × POV(10) interaction effect.

The main effects do not have much meaning when they are involved in interactions, thus interpretations must be made through the interaction effect. The interaction effect $\lambda(0.5)$ × POV(10) indicates that the effect of decreasing the mixture proportion from 0.7 (the reference cell) to 0.5 and the POV from 20% (the reference cell) to 10% contributes to large decreases in the (square root of the) absolute bias. The least squares (LS) means of the response (square root of the absolute bias) for the various levels of the mixture proportion (lambda) are plotted in Figure 9 with separate lines indicating the levels of POV.

Figure 9: LS Means Plot (square root) for Bias $(\hat{\lambda})$



From the LS means plot in Figure 9, it is observed that the (square root of the) absolute bias increases as the mixture proportion increases from 0.5 regardless of the value for POV. However, the increase is more pronounced (nearly 3 fold) when the POV is 20% as opposed to 10%.

The results from the 12 full-factorial analyzes for the bias and MSE are displayed in Table 4 and Table 5, for each parameter, respectively. The transformations determined from the respective Box-Cox Procedures are also listed in these tables. In each table, a separate row is designated for each of the six parameters being estimated with the EM-algorithm in the two-component normal-mixture random-effects model. For example, the first row in the bias table (Table 4) refers to the full-factorial model with the absolute bias of the estimates for the mixture proportion, $\hat{\lambda}$, as the response, relating to Figure 8. These tables display the effects in the order of most importance to least importance. Since a reference cell parameterization is used, the level of the effect is displayed in parentheses.

Table 4: Full-Factorial Results for Bias

Bias		
Parameter	**Transformation**	**Significant Effects**
$\hat{\lambda}$	Square Root	$\lambda(0.5)$, POV(10), $\lambda(0.5)$ × POV(10)
$\hat{\mu}_1$	Square	$\lambda(0.5)$, POV(10)
$\hat{\mu}_2$	Square	$\lambda(0.5)$, POV(10), $\lambda(0.5)$ × POV(10)
$\hat{\sigma}_1^2$	Square	POV(10), $\lambda(0.5)$
$\hat{\sigma}_2^2$	Square	POV(10), $\lambda(0.5)$
$\hat{\sigma}_\delta^2$	Square-Root	S(5), BSF(0.5), BSF(0.5) × S(5) , S(10)

Table 5: Full-Factorial Results for MSE

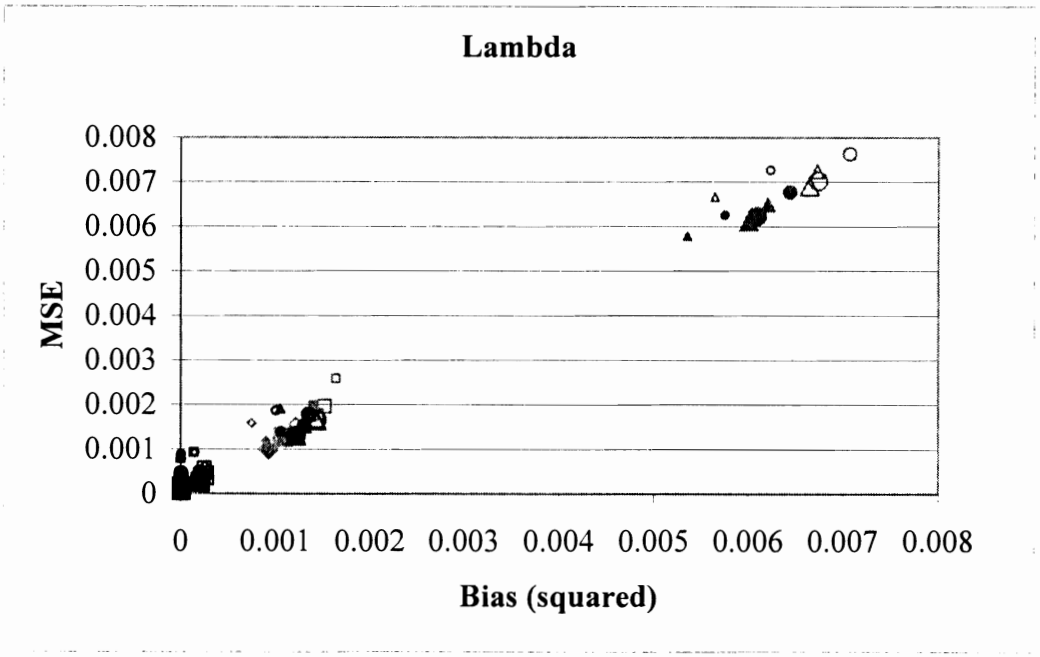
MSE		
Parameter	**Transformation**	**Significant Effects**
$\hat{\lambda}$	Log	$\lambda(0.5)$, POV(10),
$\hat{\mu}_1$	Log	S(5), BSF(0.5)
$\hat{\mu}_2$	Log	S(5), BSF(0.5), $\lambda(0.5)$
$\hat{\sigma}_1^2$	Log	POV(10), $\lambda(0.5)$
$\hat{\sigma}_2^2$	Log	POV(10), $\lambda(0.5)$
$\hat{\sigma}_\delta^2$	Log	BSF(0.5), S(5), POV(10)

The results given in Table 4 and Table 5 are discussed in conjunction with the scatter plots in the next six subsections.

3.4.1 Mixture Proportion $\left(\hat{\lambda}\right)$

The bias and MSE of the estimates for the mixture proportion, $\hat{\lambda}$, across a variety of conditions, are presented in Figure 10. The figure displays a scatter plot of the squared bias versus the MSE for each of the estimates of the mixture proportion across the 72 factor combinations.

Figure 10: Mixture Proportion $\left(\hat{\lambda}\right)$ – Bias^2 versus MSE



The scatter plot in Figure 10, suggests the following.

i.) The bias and MSE both decrease as λ decreases from 0.7 (green) to 0.6 (red) to 0.5 (blue).

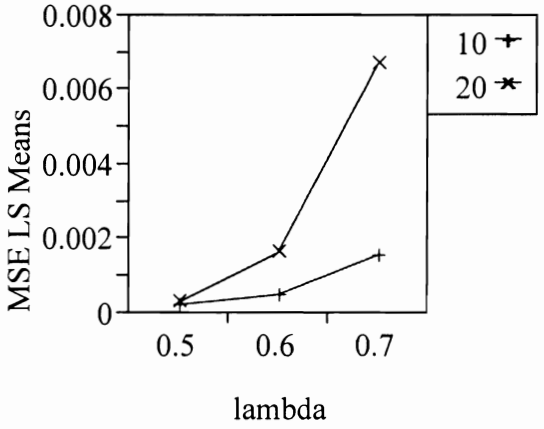
ii.) When λ is 0.6 (red) or 0.7 (green,) the bias and MSE both decrease as POV decreases from 20% (diamonds and squares) to 10% (triangles and circles), .

iii.) The remaining factors, BSF, S, and N_j do not seem to have a substantial effect on changes in either the bias or the MSE.

The full-factorial analysis for the (square root of the absolute) bias, discussed in depth at the beginning of section 3.4, indicates that the important effects contributing to large changes in the bias are $\lambda(0.5)$, POV(10), and $\lambda(0.5)$ × POV(10). The LS means plot in Figure 9 suggests that the (square root of the) absolute bias increases as the mixture proportion increases from 0.5 to 0.7 regardless of the value for POV. However, the increase is more pronounced when the POV is 20% as opposed to 10%.

The results from the full-factorial model for the (log of the) MSE, displayed in the first row of Table 5, suggest that the important effects contributing to large changes in the (log of the) MSE are $\lambda(0.5)$ and POV(10).

Figure 11: LS Means Plot for (log) MSE $(\hat{\lambda})$

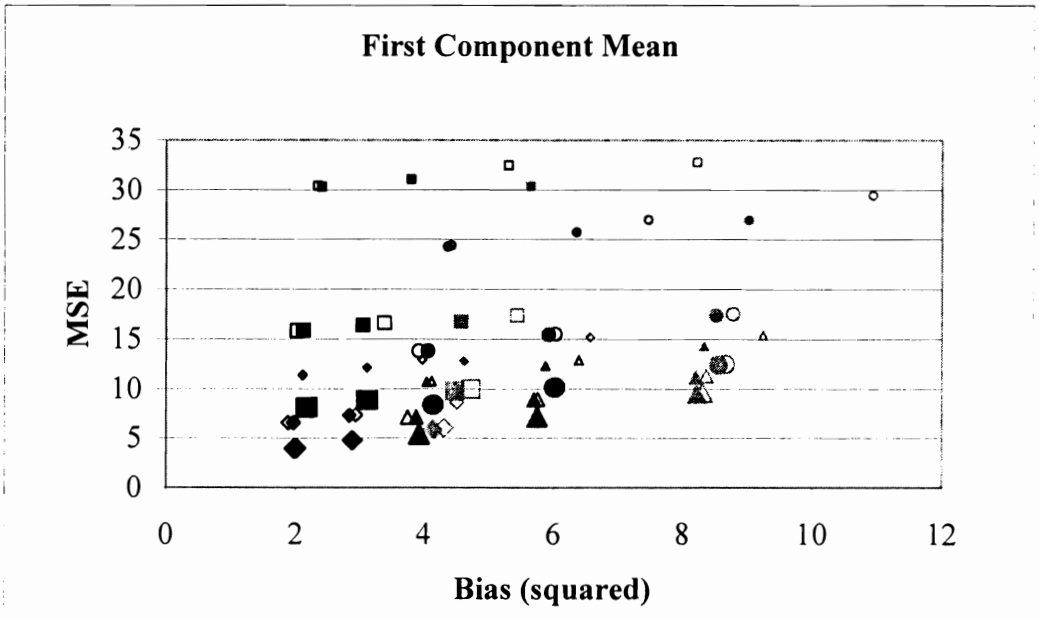


The least squares (LS) means of the (log of the) MSE for the mixture proportion (lambda) are plotted in Figure 11 with separate lines indicating the levels of POV. From this LS means plot, is observed that the MSE increases as the mixture proportion increases from 0.5 regardless of the value for POV. However, the increase is more pronounced when the POV is 20% as opposed to 10%.

3.4.2 First Component Mean $(\hat{\mu}_1)$

The bias and MSE of the estimates for the mixture proportion, $\hat{\mu}_1$, across a variety of conditions, are presented in Figure 12. The figure displays a scatter plot of the squared bias versus the MSE for each of the estimates of the first component mean across the 72 factor combinations.

Figure 12: First Component Mean $(\hat{\mu}_1)$ – Bias^2 versus MSE



The scatter plot in Figure 12 suggests the following.

i.) The (squared) bias decreases as λ decreases from 0.7 (green) to 0.6 (red) to 0.5 (blue).

ii.) The (squared) bias decreases as POV decreases from 20% (triangles and circles) to 10% (diamonds and squares).

iii.) The MSE decreases substantially as S is increased from 5 (small) to 10 (medium) and a bit more as S increases from 10 (medium) to 20 (large).

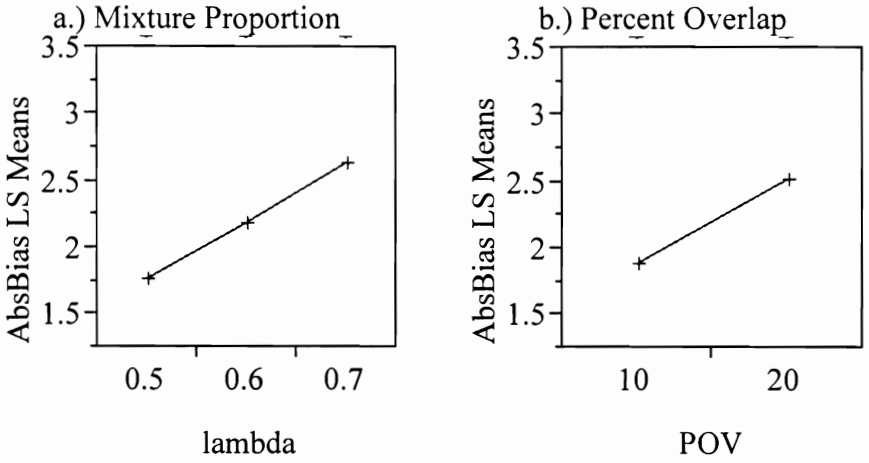
iv.) The MSE decreases as BSF decreases from 1.5 (diamonds and triangles) to 0.5 (squares and circles).

v.) The factors BSF, S, and N_j do not seem to have large effects on the (squared) bias.

vi.) The factors λ, POV, and N_j do not seem to affect the MSE.

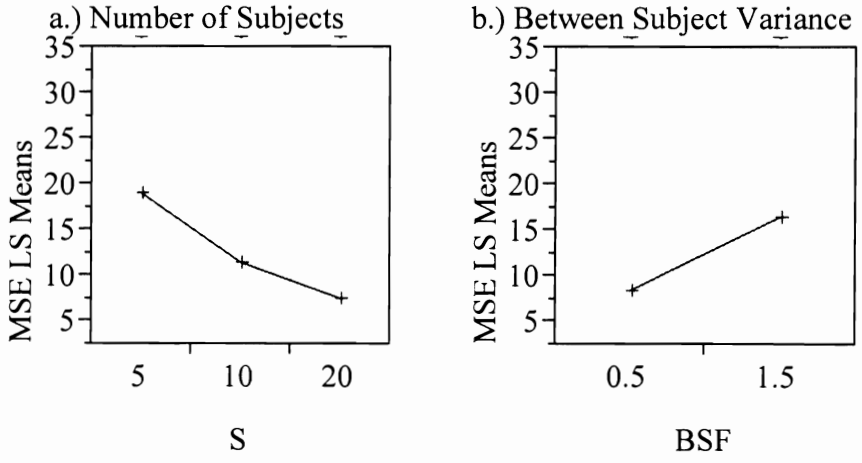
The results from the full-factorial model for the (square of the) absolute bias, displayed in the second row of Table 4, suggest that the important effects contributing to large changes in the (square of the) absolute bias are $\lambda(0.5)$ and POV(10). LS means plots of the (square of the) absolute bias for the various levels, namely, the mixture proportion (lambda) and the percentage of overlap are shown in Figure 13, panels a.), and b.), respectively. From these LS means plots in, it is observed that the (square of the) absolute bias decreases as the mixture proportion decreases from 0.7 to 0.5 or the percent of overlap decreases from 20% to 10%.

Figure 13: LS Means plot for (square) Bias $(\hat{\mu}_1)$



The results from the full-factorial model for the (log of the) MSE, displayed in the second row of Table 5, indicate that the important effects contributing to large changes in (the log of) the MSE are *S*(5) and BSF(0.5). LS means plots of the (log of the) MSE for the various levels, namely, the number of subjects and the between subject variance factor are shown in Figure 14, panels a.) and b.), respectively.

Figure 14: LS Means Plot for (log) MSE $(\hat{\mu}_1)$

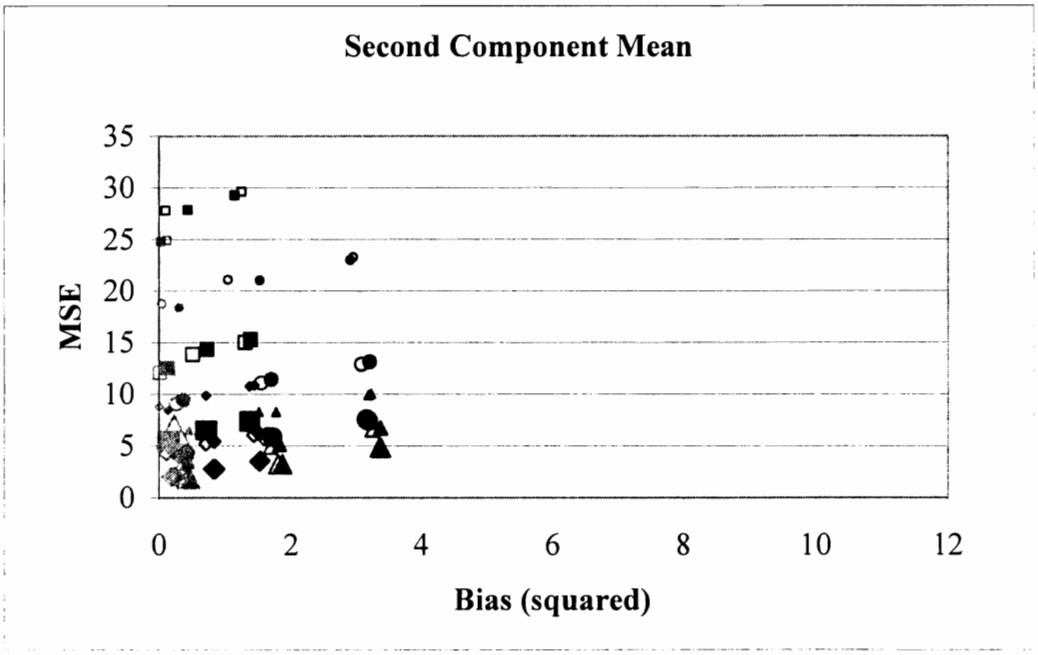


From the LS means plots in Figure 14, it is observed that the (log of the) MSE decreases as the number of subjects increases from 5 to 20 or the between subject variance factor decreases from 1.5 to 0.5.

3.4.3 Second Component Mean $(\hat{\mu}_2)$

The bias and MSE of the estimates for the mixture proportion, $\hat{\mu}_2$, across a variety of conditions, are presented in Figure 15. The figure displays a scatter plot of the squared bias versus the MSE for each of the estimates for the second component mean across the 72 factor combinations.

Figure 15: Second Component Mean $(\hat{\mu}_2)$ – Bias^2 versus MSE



The scatter plot in Figure 15 suggests the following.

i.) The (squared) bias increases as λ decreases from 0.7 (green) to 0.6 (red) to 0.5 (blue).

ii.) The (squared) bias increases as POV decreases from 20% (triangles and circles) to 10% (diamonds and squares). These decreases become more pronounced as λ decreases from 0.7 (green) to 0.5 (blue).

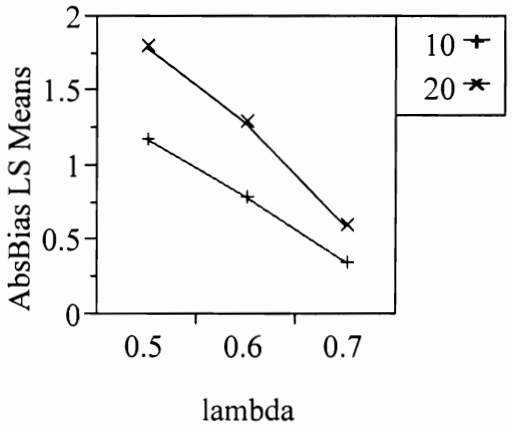
iii.) The MSE decreases as S is increased from 5 (small) to 10 (medium) to 20 (large).

iv.) The MSE decreases as BSF decreases from 1.5 (diamonds and triangles) to 0.5 (squares and circles).

v.) The MSE decreases as λ increases from 0.5 (blue) to 0.6 (red) to 0.7 (green).

The results from the full-factorial model for the (square of the) bias, displayed in the third row of Table 4, suggest that the important effects contributing to large changes in the (square of the) bias are $\lambda(0.5)$, POV(10), and $\lambda(0.5) \times$ POV(10).

Figure 16: LS Means Plot for (square) Bias $(\hat{\mu}_2)$

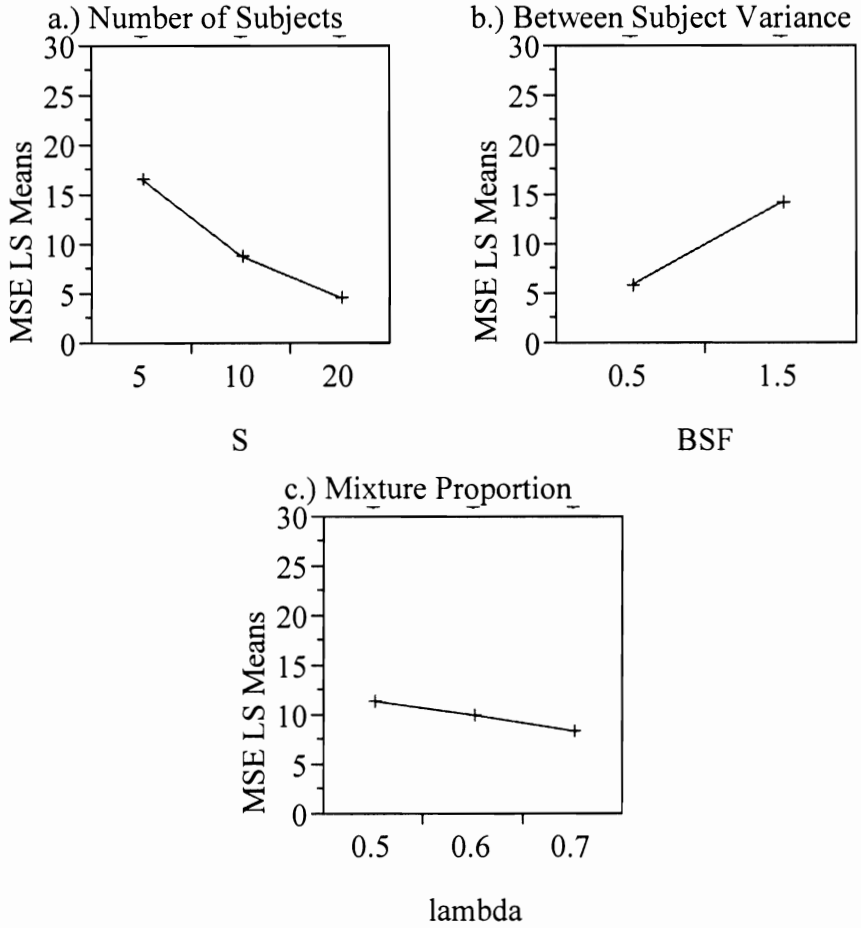


The least squares (LS) means of the response (square of the) bias for the various levels of the mixture proportion (lambda) are plotted in Figure 16 with separate lines indicating the

levels of POV. From the LS means plot in Figure 16, it is observed that the (square of the) bias decreases as the mixture proportion increases from 0.5 to 0.7, regardless of the value for POV. However, the decrease is more pronounced when the POV is 20% as opposed to 10%.

The results from the full-factorial model for the (log of the) MSE, displayed in the third row of Table 5, indicate that the important effects contributing to large changes in (the log of) the MSE are $S(5)$, BSF(0.5), and $\lambda(0.5)$.

Figure 17: LS Means Plot for (log) MSE $(\hat{\mu}_2)$

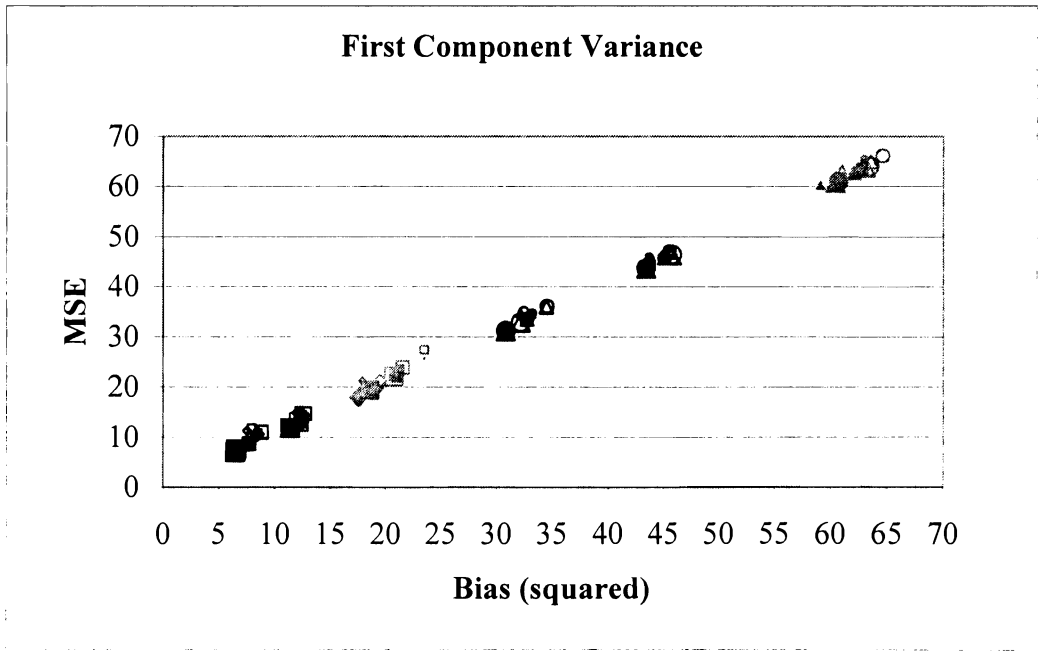


The LS means plots of the (log of the) MSE for the various levels, namely, the number of subjects, the between subject variance factor, and the mixture proportion, are shown in Figure 17, panels a.), b.), and c.), respectively. From these LS means plots, it is observed that the (log of the) MSE decreases as the number of subjects increases from 5 to 20, the between subject variance factor decreases from 1.5 to 0.5, or the mixture proportion increases from 0.5 to 0.7.

3.4.4 First Component Variance $\left(\hat{\sigma}_1^2\right)$

The bias and MSE of the estimates for the mixture proportion, $\hat{\sigma}_1^2$, across a variety of conditions, are presented in Figure 18.

Figure 18: First Component Variance $\left(\hat{\sigma}_1^2\right)$ – Bias^2 versus MSE



The figure displays a scatter plot of the squared bias versus the MSE for each of the estimates of the first component variance across the 72 factor combinations. The scatter plot in Figure 18 suggests the following.

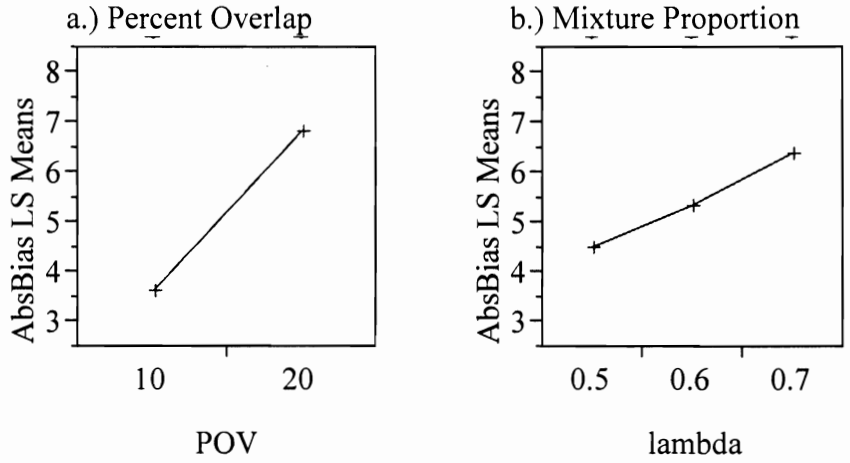
i.) The squared bias and the MSE both decrease as POV decreases from 20% (triangles and circles) to 10% (diamonds and squares).

ii.) The (squared) bias and the MSE both decrease as λ decreases from 0.7 (green) to 0.6 (red) to 0.5 (blue). These decreases are more pronounced when POV is 20% (triangles and circles) as opposed to 10% (diamonds and squares).

iii.) The factors BSF, S, and N_j do not seem to affect the (squared) bias or the MSE.

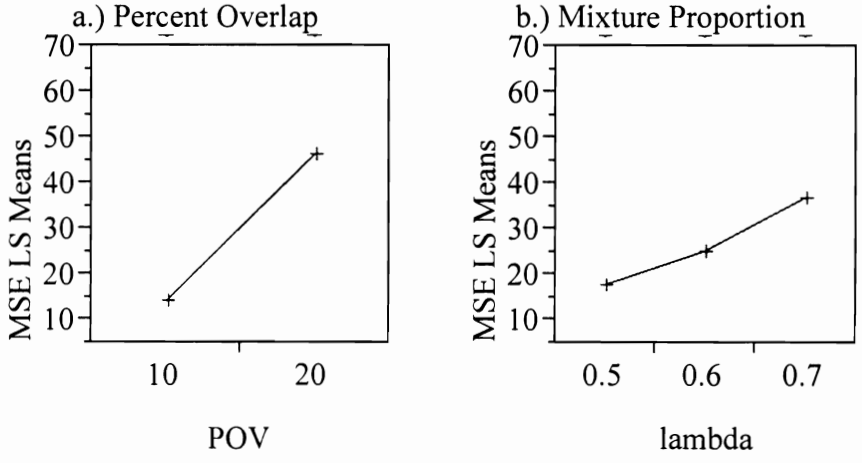
The results from the full-factorial model for the (square of the) bias, displayed in the fourth row of Table 4, suggest that the important effects contributing to large changes in the (square of the) bias are POV(10) and $\lambda(0.5)$. LS means plots of the (square of the) bias for the various levels, namely, the percentage of overlap and the mixture proportion are shown in Figure 19, panels a.) and b.), respectively. From these LS means plots, it is observed that the (square of the) bias increases as the percentage of overlap increases from 10% to 20% or the mixture proportion increases from 0.5 to 0.7.

Figure 19: LS Means Plots for (square) Bias $\left(\hat{\sigma}_1^2\right)$



The results from the full-factorial model for the (log of the) MSE, displayed in the fourth row of Table 5, suggest that the important effects contributing to large changes in the (log of the) MSE are POV(10) and $\lambda(0.5)$. LS means plots of the (log of the) MSE for the various levels, namely, the percentage of overlap and the mixture proportion (lambda) are shown in Figure 20, panels a.) and b.), respectively.

Figure 20: LS Means Plot for (log) MSE $\left(\hat{\sigma}_1^2\right)$

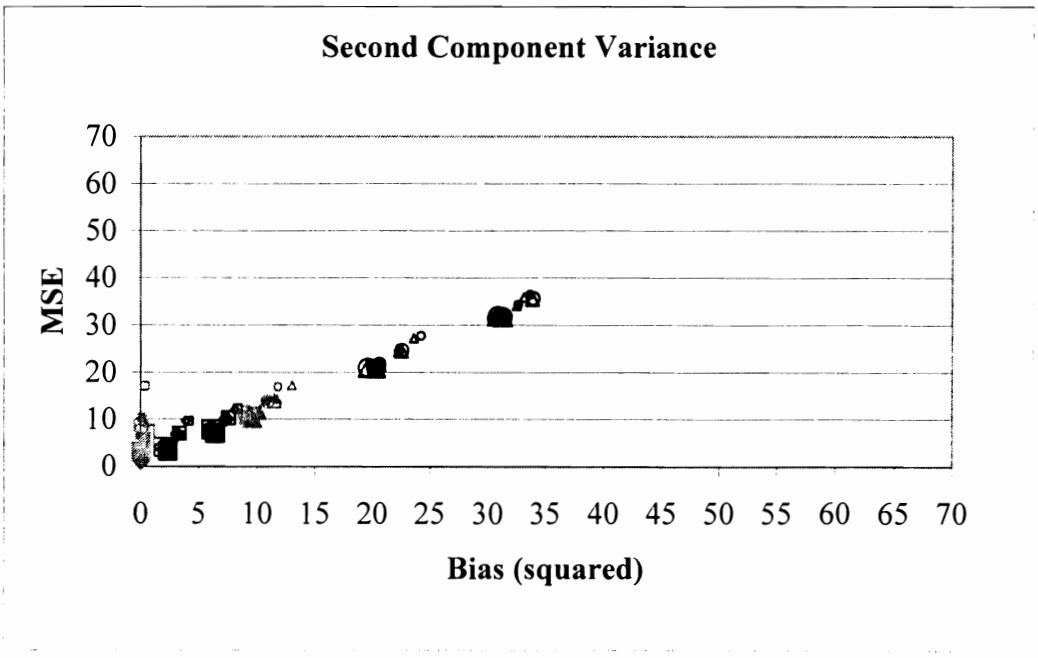


From these LS means plots, it is observed that the (log of the) MSE increases as the percentage of overlap increases from 10% to 20% or the mixture proportion increases from 0.5 to 0.7.

3.4.5 Second Component Variance $\left(\hat{\sigma}_2^2\right)$

The bias and MSE of the estimates for the mixture proportion, $\hat{\sigma}_2^2$, across a variety of conditions, are presented in Figure 21. The figure displays a scatter plot of the squared bias versus the MSE for each of the estimates of the second component variance across the 72 factor combinations.

Figure 21: Second Component Variance $\left(\hat{\sigma}_2^2\right)$ – Bias^2 versus MSE



The scatter plot in Figure 21 suggests the following.

i.) The squared bias and the MSE both decrease as POV decreases from 20% (triangles and circles) to 10% (diamonds and squares).

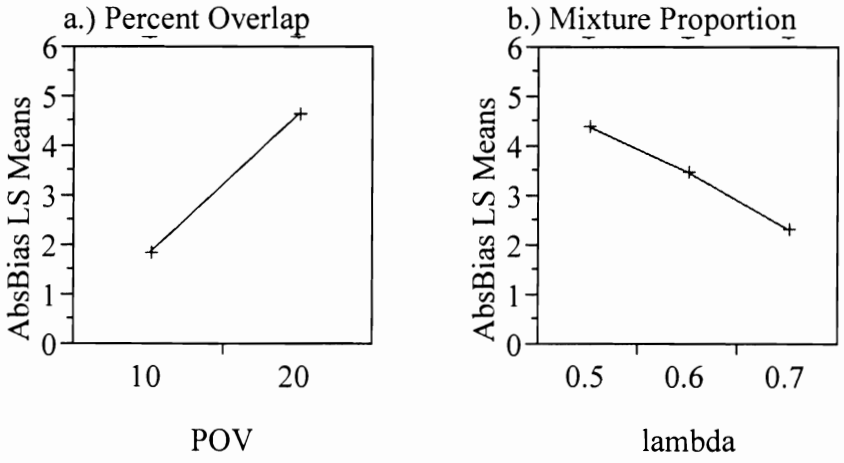
ii.) The (squared) bias and the MSE both decrease as λ increases from 0.5 (blue) to 0.6 (red) to 0.7 (green). These decreases are more pronounced when POV is 20% (triangles and circles) as opposed to 10% (diamonds and squares).

iii.) The (squared) bias and the MSE both seem to show minor decreases as S increases from 5 (small) to 10 (medium) to 20 (large).

iv.) The factors BSF and N_j do not seem to affect the (squared) bias or the MSE.

The results from the full-factorial model for the (square of the) bias, displayed in the fifth row of Table 4, suggest that the important effects contributing to large changes in the (square of the) bias are POV(10) and $\lambda(0.5)$. LS means plots of the (square of the) bias for the various levels, namely, the percentage of overlap and the mixture proportion are shown in Figure 22, panels a.) and b.), respectively.

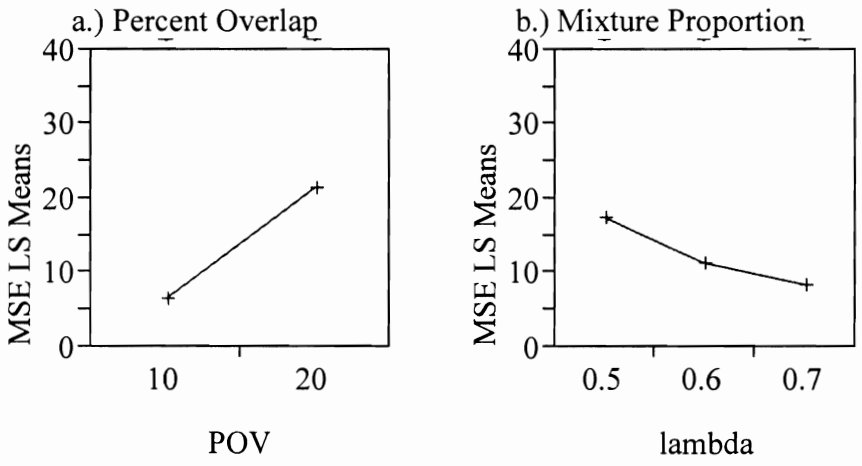
Figure 22: LS Means Plots for (square) Bias $\left(\hat{\sigma}_2^2\right)$



From the LS means plots, it is observed that the (square of the) bias decreases as the percentage of overlap decreases from 20% to 10% or the mixture proportion increases from 0.5 to 0.7.

The results from the full-factorial model for the (log of the) MSE, displayed in the fifth row of Table 5, suggest that the important effects contributing to large changes in the (log of the) MSE are POV(10) and $\lambda(0.5)$. The LS means plot of the (log of the) MSE for the various levels, namely the percentage of overlap and the mixture proportion are shown in Figure 23, panels a.) and b.), respectively.

Figure 23: LS Means Plots for (log) MSE $\left(\hat{\sigma}_2^2\right)$

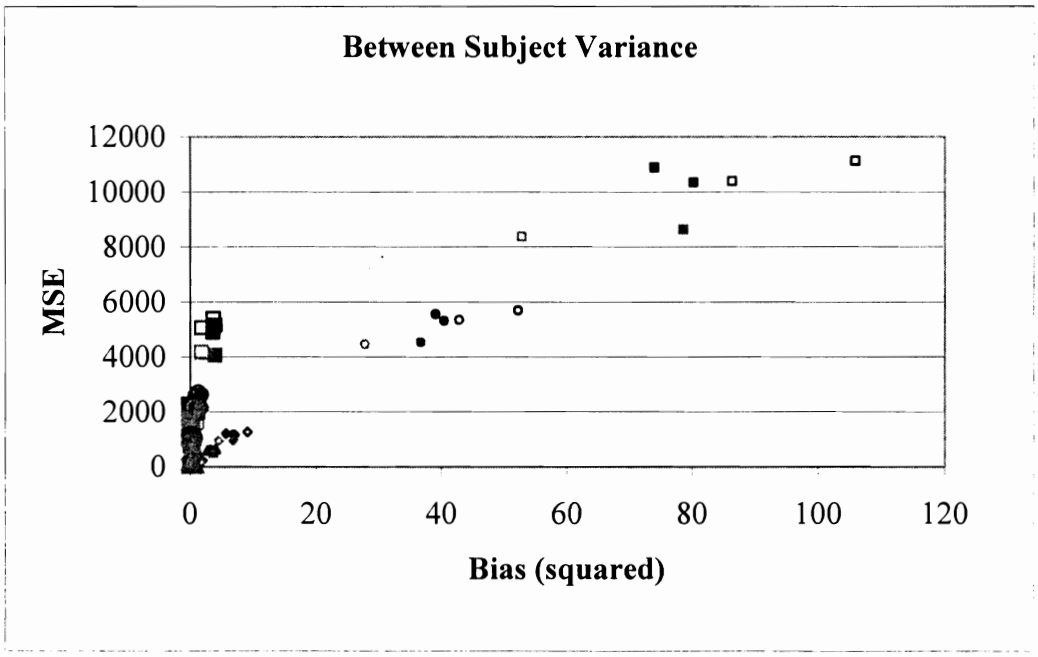


From the LS means plot in Figure 23 it is observed that the (log of the) MSE decreases as the POV decreases from 20% to 10% or the mixture proportion increases from 0.5 to 0.7.

3.4.6 Between Subject Variance $\left(\hat{\sigma}_{\delta}^{2}\right)$

The bias and MSE of the estimates for the mixture proportion, $\hat{\sigma}_{\delta}^{2}$, across a variety of conditions, are presented in Figure 24. The figure displays a scatter plot of the squared bias versus the MSE for each of the estimates for the between subject variance across the 72 factor combinations.

Figure 24: Between Subject Variance $\left(\hat{\sigma}_{\delta}^{2}\right)$ – Bias^2 versus MSE



The scatter plot in Figure 24 suggests the following.

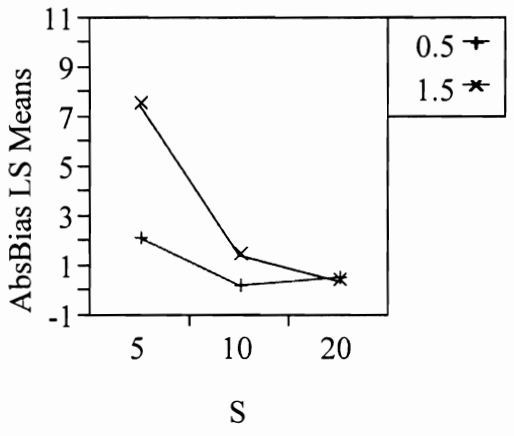
i.) The MSE and (squared) bias both decrease substantially as S increases from 5 (small) to 10 (medium). The MSE continues to decrease as S increases from 10 (medium) to 20 (large).

ii.) Large decreases in the (squared) bias and the MSE are associated with decreases in the BSF, from 1.5 (diamonds and triangles) to 0.5 (squares and circles), especially when S is 5 (small).

iii.) The remaining factors POV, λ, and N_j do not seem to be associated with changes in the MSE or (squared) bias.

The results from the full-factorial model for the (square root of the) bias, displayed in the sixth row of Table 4, suggest that the important effects contributing to large changes in the (square root of the) bias are $S(5)$, BSF(0.5), BSF(0.5) × $S(5)$, and $S(10)$. The least squares (LS) means of the response (square root of the bias) for the various levels for the number of subjects are plotted in Figure 25 with separate lines indicating the levels of BSF.

Figure 25: LS Means Plots for (square root) Bias $\left(\hat{\sigma}_{\delta}^{2}\right)$

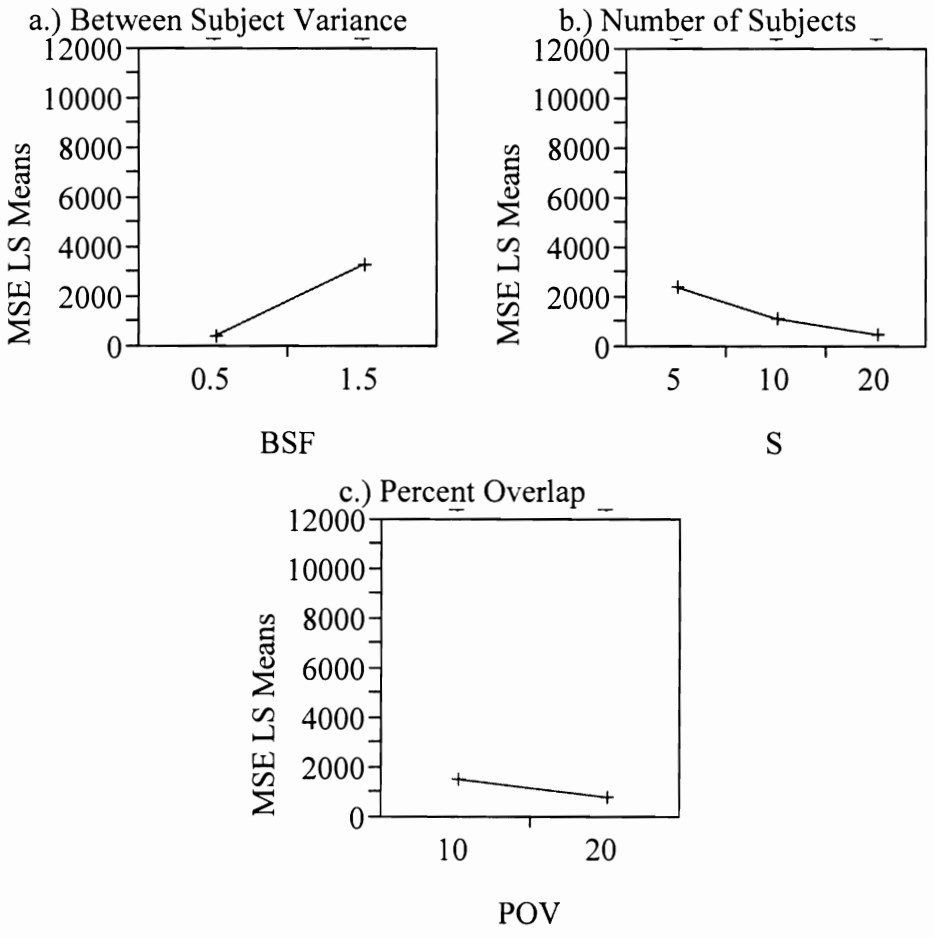


From the LS means plot in Figure 25 it is observed that the (square root of the) bias decreases as the number of subjects increases from 5 to 10, regardless of the value of the BSF. However, this decrease is more pronounced when the BSF is 1.5. The (square root

of the) bias continues to decrease when BSF is 1.5 as the number of subjects increases from 10 to 20, however small increases in the (square root of the) bias are seen when BSF is 0.5.

The results from the full-factorial model for the (log of the) MSE, displayed in the sixth row of Table 5, suggest that the important effects contributing to large changes in the (log of the) MSE are BSF(0.5), *S*(5), and POV(10).

Figure 26: LS Means Plots for (log) MSE $\left(\hat{\sigma}_{\delta}^{2}\right)$



LS means plots of for the various levels, namely the between subject variance factor, the number of subjects, and the percentage of overlap are shown in the plots Figure 26, panels a.), b.) and c.), respectively. From these LS means plot it is observed that the (log of the) MSE decreases as the between subject variance factor decreases from 1.5 to 0.5, the number of subjects increases from 5 to 10, or the percentage of overlap increases from 10% to 20%. The MSE continues to decrease as the number of subjects is increased from 10 to 20, however the effect S(10) did not stand out as important on the normal probability plot.

3.5. Discussion and summary of the Simulation Study

In summary, the simulation results show, the estimation of the mixture proportion is the least sensitive to the simulation factors considered, while the between subject variance is the most sensitive. In what follows, the results for each of the parameters are further discussed.

For the mixture proportion, λ, the range of values for both the MSE and the bias squared are narrow (0 – 0.009). There is a high correlations between the MSE and the squared bias and therefore the variances of the estimates of λ are small across all factor combinations (variance = MSE – bias squared). The bias and MSE decrease as the magnitude of the mixture proportion approaches 0.5. This is expected since the variance of any Bernoulli event is minimized when the probability of the event is 0.5. The decrease is more pronounced as the percentage of overlap between the two mixing distribution decreases. In other words, the estimation of the mixture proportion is more

precise as the two missing distributions become more distinct, which also conforms to intuition.

For the component means, μ_1 and μ_2, the range of values for the bias squared, (0 – 12) and (0 – 4), respectively, are larger than seen with the mixture proportion, however they are considerably smaller than exhibited by the variance parameters. The first component means are consistently underestimated, while the second component means are almost always over estimated. These leads to a much smaller bias in the overall mean of the mixture distribution. Also, the bias in the second component mean is consistently smaller than the bias in the first component mean. A noteworthy trend is observed between the two components. While increases in the mixture proportion factor result in increases in the bias in the first component, the opposite occurs in the second component. Here, decreases in the mixture proportion factor are associated with increases in the bias. This could be due to the fact that when a smaller proportion is mixed, the spread of that component is reduced (see Figure 4 and Figure 5). Decreases in the percentage of overlap are associated with decreases in the bias across both component means. That is, the estimation of the component means is more precise as the two mixing densities become more distinct. For both components, however, the mixture proportion and the percentage of overlap are the only important factors identified as contributing to large changes in the bias of the estimates for the component means.

The ranges of the MSE for the component means are similar, about 0 to 35, indicating that the variance in the estimates for the two components is comparable across the component means. The important factors contributing to changes in the MSE of the

estimates for the component means included the number of subjects and the between subject variance factor. Both component means exhibit decreases in the MSE of their estimates when the between subject variance factor is decreased or the number of subjects is increases. Decreases in the MSE for the estimates of the second component mean are also associated with the increases in the mixture proportion.

A similar high correlation between the squared bias and MSE, as seen when estimating the mixture proportion, is seen for the component variances, indicative of small variances for these estimates. Furthermore, as with the component means, the first component shows much greater magnitudes of bias than the second component, especially as the mixture proportion moves away from 0.5. The most apparent effects on MSE and bias, for both component variances, are due to the percentage of overlap and the mixture proportion. Interestingly, while increases in the mixture proportion factor result in increases in the bias and MSE in the first component, the opposite occurs in the second component. Here, decreases in the mixture proportion factor are associated with increases in the bias and MSE. Decreases in the percentage of overlap result in decreases in the bias and MSE across when estimating either of the component variances.

Much of the bias (and hence the MSE) observed with the component variances are due to the weighted EM-algorithm underestimating the component variances. This, however, is a typical result of the maximum likelihood procedure. The estimate for the variance parameter in the maximum likelihood procedure in any random-effects model does not account for the uncertainty in the estimation of the means and therefore suffers from large bias, and by definition is a biased estimator. An REML approach in the

normal random-effects model is often used to reduce the bias. One could consider such an approach for the mixture models as well. Further discussion of this is provided in chapter 5.

The between subject variance is estimated with the worst precision. The range of the squared bias and MSE are about (0,120) and (0,12000), respectively. The considerable size of the MSEs are due to the large variances for the between subject variance estimates. The number of subjects and the between subject variance factors are the most important contributors to changes in the MSE and the bias. Both the MSE and the bias show considerable decreases when the number of subjects is increased and when the between subject variance factor is decreased. Increases in the number of subjects, from just 5 to 10, result in estimates with squared bias of less than 5 units. The effect on the MSE is just as dramatic. It is also noted that the MSE decreases as the percentage of overlap increases. This could be due to the fact that as the percentage of overlap increases the total spread in the mixture decreases.

As a whole, the EM-algorithm proposed in section 2.8.2, estimates very well. The mixture proportion is estimated with the most accuracy and the between subject variance with the worst accuracy. The percentage of overlap and the mixture proportion are the dominating factors in changes in the bias when estimating the mixture proportion and the component means and variances. The MSEs for all parameter estimates are affected by a variety of factors described earlier. Precision in one component usually means a reduction of accuracy in the other component when estimating either the component means or variances. Small increases in the number of subjects result in dramatic

decreases in both the bias and MSE for the estimates of the between subject variance. Increasing the number of observations per subject from 100 to 200 did not have a large effect on the MSE or bias for any of the parameter estimates.

When dealing with HRV data, the bimodal nature of the data tends to occur at high frequencies (in the power spectrum). Thus careful attention must be given to the long term effects in the data due to lower frequencies trends in the data. These lower frequencies will dominate the means in the data and the bimodal nature may be lost or become misleading. Several detrending methods are suggested in section 1.4. The next chapter focuses on applying the weighted EM-algorithm to the HRV data from the Loneliness study introduced in section 1.5. The data encountered in this study is a good example of the necessary detrending of the data.

4 Data Analysis

4.1. Introduction

In this chapter, the Loneliness data introduced in section 1.5 is analyzed using the methodology in section 2.8.2. Although this study is not the initial motivation for this research, it contains the same qualities as the motivational studies. Reasons for using the data from the Loneliness study rather than the data from the motivational studies is discussed at the conclusion of this chapter.

The Loneliness data consists of 9 subjects randomly chosen from 89 (51% Male) undergraduates who participated in a loneliness study at the General Research Center at the Ohio State University (Cacioppo et al, 2002). The data collected include ECG signal, age, and gender. The sample of 9 subjects has a mean age of 19.22 years (S.D. = 1.86) and consists of 4 males (44%). During each ECG recording period, the subjects performed a series of six psychological stressor tasks and one orthostasis stressor task. ECG data acquired during the minutes from the five speech stressor tasks, the "Saab" speech stressor task (BS), ask for a date (AS), describe inanimate objects in room (IS), why I'm likeable (LS), and describe ways to school (WS), will be considered in this analysis. A series of RR-intervals was obtained for each of these five speech stressor tasks during three recording periods, a baseline period, a preparation period, and a delivery period. There were two minutes of baseline ECG data collected before each of the AS, IS, LS, and WS speech stressor tasks and four minutes of baseline ECG data collected before the BS task. After the baseline period, ECG signals were recorded on

the subjects during the two minutes allotted to prepare their speeches, and then again during the two minutes allotted to deliver their speeches. The combination of speech stressor tasks and recording periods, used in this analysis, along with the number of minutes recorded for each combination are summarized in Table 6.

Table 6: ECG Minutes in the Loneliness Study

Task	Number of minutes		
	Period		
	Baseline	Preparation	Delivery
BS	4	2	2
AS	2	2	2
IS	2	2	2
LS	2	2	2
WS	2	2	2

The verbal mental arithmetic task and the orthostasis task are not considered in this analysis since the ECG signal was not consistently recorded during the baseline, preparation, and delivery periods for these tasks. A total of 32 minutes of ECG data is then used for the data analysis seen in this chapter.

After data collection, the ECG signals acquired at 1000 Hertz (Hz) were decimated to 500 Hz. The R-peaks were then identified using waveform matching templates and then a time/amplitude criterion (Mandrekar 2002). After identifying the R-peaks, the RR-interval series were calculated using the distance between successive R-peaks. This "raw" (unedited) form of the RR-interval data was then obtained from Mandrekar to be used for analysis in this dissertation and can be found on the data CD (Appendix 7.4.4). A sample portion of the data can be seen in Appendix 7.5.

4.2. Data Preparation

In order to use this "raw" RR-interval data for analysis (Appendix 7.4.4), a series of steps should be performed. First, artifacts in the series, such as missed or spurious heart beats, need to be identified. The series of RR-intervals from each subject were manually examined by task and period to identify possible artifacts. An interval was considered an artifact if it was "large" or "small" in relation to all the other intervals for that subject during the specified task and period combination, and in relation to the intervals surrounding the "suspicious" interval. The total number of RR-intervals measured on each subject, across all combinations of tasks and periods, are displayed in Table 7.

Table 7: RR-intervals – Useable versus Artifact

RR-intervals				
Subject	**Useable**	**Artifacts**	**Total**	**% Artifact**
1	2101	74	2175	3.40
2	2354	72	2426	2.97
3	2163	102	2265	4.50
4	1982	73	2055	3.55
5	2266	67	2333	2.87
6	2176	60	2236	2.68
7	2060	64	2124	3.01
8	2432	33	2465	1.34
9	2294	43	2337	1.84
Total	19828	588	20416	2.88

On average 2.88 percent of the data from each subject was identified as artifacts, with as much as 4.50% identified for subject 3 and as little as 1.34% identified for subject 8. The 588 intervals identified as artifacts were not used in subsequent analyses.

After excluding artifacts, the data must be detrended in order to remove the ultra-low frequency trends in the data to reduce some of the non-stationarity in the data

(Litvack et al., 1995). This is accomplished by fitting first-order polynomials to the series of RR-intervals obtained from each subject for each combination of task and period. This results in (9 × 3 × 5) 135 separate first-order fitted models. Each fit produces an estimate for the intercept and slope, along with a set of corresponding residuals with mean zero. The set of residuals from each fit are added to the estimated intercept to produce a series of detrended RR-interval data that is still centered around its original mean (intercept). The slopes and intercepts specified by the fits for the first subject are displayed in Table 8.

Table 8: Slopes and Intercepts for Detrending – Subject 1

Intercept Slope		**Task**				
Subject	**Period**	BS	AS	IS	LS	WS
1	Baseline	924.74 -0.10	947.39 0.01	939.48 -0.12	940.11 -0.17	978.01 -0.21
	Preparation	840.59 0.15	913.75 -0.13	927.29 -0.32	868.92 -0.11	899.11 0.70
	Delivery	732.31 0.27	713.22 0.88	775.05 0.77	711.73 0.78	783.14 0.53

The slopes and intercepts obtained for all 9 subjects can be found in Appendix 7.6. From Table 8, observe that the average RR-interval length in milliseconds (ms) for the first subject tends to decreases from baseline to preparation to delivery, consistently across all tasks. This trend is observed in nearly all of the subjects. That is, there appears to be an increase in the heart rate from baseline to preparation to delivery.

Finally, in preparation for the analysis, any continuous covariates to be included in the models should be centered, so that interpretations of the effects are made with reference to the averages. For these data, the only continuous covariate is the age

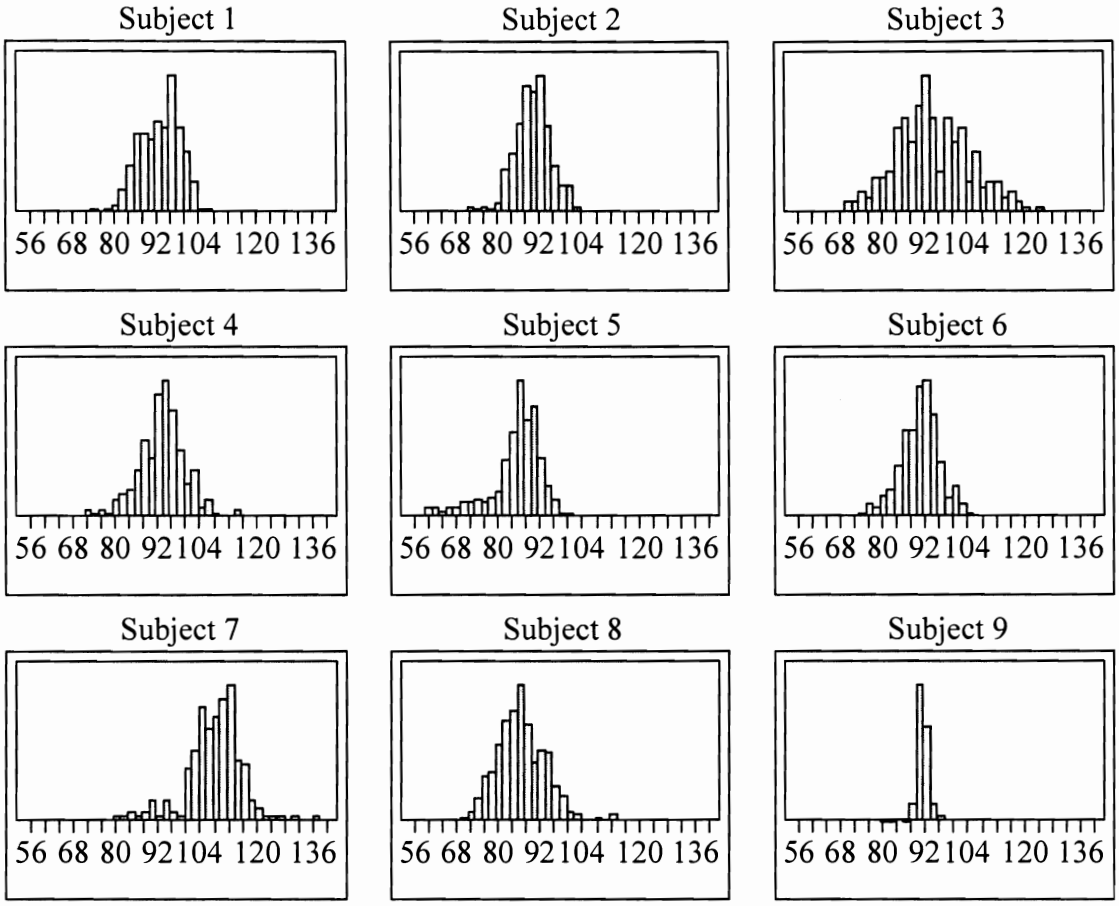
variable (mean age 19.22). This mean age was then subtracted from each subject's age to produce the centered age variable to be used in the analysis. After centering the continuous covariates, the RR-interval series was rescaled by a factor of 10 (RR-interval / 10) for convenience (in order to reduce the magnitude of the results in the subsequent analysis). This detrended, artifact-free, centered, and rescaled data is what is used for analysis in this chapter and can be found on the data CD (Appendix 7.4.5).

4.3. The Analysis Data Set

Most methods used for fitting RR-interval data assume that the data arise from a normal distribution, although it is widely recognized that this assumption is false. Several researchers have commented on the irregularity in the shapes of the histograms of RR-interval data (Hashida et al., 1973, Jennings et al., 1974, Eckberg, 1983, Nagaraja, 1995, and Mandrekar, 2005) and most suggest a bimodal pattern in both healthy and heart diseased adult subjects. Histograms of the 4 baseline minutes from each subject during the BS task are show in Figure 27. From these 9 histograms, it is clear that there are large departures from normality. It is further apparent, in terms of the central tendency, that there is great deal of variability and a suggestion of bimodality, both of which need to be incorporated into the analyses. For example, subject 3 exhibits a large amount of variability during the BS baseline period while subject 9 exhibits very little and subject 7 has larger component means than subject 5. In the analysis the subject to subject variability is incorporated though a random-effects term. The bimodality is incorporated

through the assumptions of two-component normal-mixture densities for the residuals as described in Chapter 2.

Figure 27: Histograms of RR-Intervals by Subject (BS baseline)



The mixture distributions from subject to subject will initially be allowed to have different mixing proportions as suggested by Figure 27. For example, subject 4 seems to have a mixture proportion smaller than 0.5 for the first component, while subject 7 indicates a mixture proportion greater than 0.5 for the first component. The current methods for fitting RR-interval data that allow for normal-mixture distributions do not accommodate variation between the subjects. Instead, separate models are fit for each

subject. In addition to modeling the fixed-effects, the methodology provided in this dissertation allows for the subject to subject variability to be taken into account by including a random effect (a random intercept for each subject) in the model while assuming that the observations between subjects are independent. In the next section, the methodology used to fit the series of RR-interval data from the 9 subjects in this study is described.

4.4. The Model

The edited series of RR-interval data obtained using the methods described in section 4.2 are modeled using a two-component normal-mixture random-effects model as defined in section 2.3, given by equation (2.2), $\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Interest is in modeling the series of RR-interval data across the 9 subjects while accounting for variations due to the fixed-effects and random-effects. Here, the $N \times 1$ vector, $\mathbf{y} = (\mathbf{y}_1 \quad \mathbf{y}_2 \quad \cdots \quad \mathbf{y}_9)'$, contains the series of RR-interval data from all 9 subjects across all 15 task by period observation periods. The vector $\mathbf{y}_j, j = 1, \ldots, 9$, is defined as the $N_j \times 1$ vector of observed RR-intervals from the j^{th} subject, $N = \sum_{j=1}^{9} N_j$. The series of observations within the vector $\mathbf{y}_j$ are grouped by task and period in the order in which they were performed by the subjects.

In this analysis, the only random-effect considered is a random subject effect. Define β_j as the random-effects parameter for the j^{th} subject and z_{ij} as the corresponding design variable for the i^{th} observation on the j^{th} subject ($z_{ij} = 1$ for the j^{th}

subject and $z_{ij} = 0$ otherwise). Then, the $N \times 9$ design matrix for the random-effects is expressed as

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_9 \end{pmatrix}$$

where $\mathbf{Z}_j = \begin{pmatrix} z_{1j} & z_{2j} & \cdots & z_{Nj} \end{pmatrix}'$ is the $N_j \times 1$ random-effects design matrix for the j^{th} subject and reduces to a vector of ones. The random subject effects, β_j, are assumed to be iid and come from a $N\left(0, \sigma_\delta^2\right)$, independent of the residuals. The vector of subject effects, β, then follows a $N_9\left(\mathbf{0}, \mathbf{G}\right)$, where

$$\mathbf{G} = \begin{bmatrix} \sigma_\delta^2 & 0 & \cdots & 0 \\ 0 & \sigma_\delta^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_\delta^2 \end{bmatrix}.$$

Here, the parameter σ_δ^2 describes the variations in the mean RR-intervals due to the random subject effects.

Within each subject, the residuals, ε_{ij}, are assumed to be iid and come from a two-component normal-mixture distribution, with component means, μ_{ijk}, component variances, σ_k^2, and mixture proportion, λ_j, for $k = 1, 2$. For this model, the mixture proportion, λ_j, is allowed to vary for each subject. The vector of residuals for each

subject, $\boldsymbol{\varepsilon}_j$, are assumed to follow a distribution given by the product of the N_j univariate normal-mixture densities for the ε_{ij}, given in equation (2.6) and expressed as

$$\phi_{JMX}\left(\boldsymbol{\varepsilon}_j;\boldsymbol{\mu}_{j1},\boldsymbol{\mu}_{j2},\boldsymbol{\Sigma}_{j1},\boldsymbol{\Sigma}_{j2},\lambda_j\right)=\prod_{i=1}^{N_j} f\left(\varepsilon_{ij};\mu_{ij1},\mu_{ij2},\sigma_1^2,\sigma_2^2,\lambda_j\right).$$

The component variance-covariance structures, for each subject, of the vector of residuals, $\boldsymbol{\varepsilon}_j$, are independent and given by

$$\boldsymbol{\Sigma}_{jk}=\begin{bmatrix}\sigma_k^2 & 0 & \cdots & 0\\ 0 & \sigma_k^2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma_k^2\end{bmatrix}.$$

Here, the assumption of independence in the residuals from a single subject may be weak since the observed data are in fact time-series data, and most likely have an inherent auto-regressive structure. With the inclusion of the random subject effects and assuming the independence structure for $\mathbf{G}$, the vector of observations from the j^{th} subject, $\mathbf{Y}_j$, has component variance covariance structure $\mathbf{V}_{jk}$ given by

$$\mathbf{V}_{jk}=\mathbf{Z}_j\mathbf{G}\mathbf{Z}'_j+\boldsymbol{\Sigma}_{jk}=\begin{bmatrix}\sigma_\delta^2+\sigma_k^2 & \sigma_\delta^2 & \cdots & \sigma_\delta^2\\ \sigma_\delta^2 & \sigma_\delta^2+\sigma_k^2 & \cdots & \sigma_\delta^2\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_\delta^2 & \sigma_\delta^2 & \cdots & \sigma_\delta^2+\sigma_k^2\end{bmatrix}.$$

That is, observations from the same subject have variance $\sigma_\delta^2 + \sigma_k^2$ and covariance σ_δ^2. Although the random-effects assumption leads to a compound symmetric structure for the correlations among the observations within a subject, other structures, such as the autoregressive (1) or more complex time-series structures may be more appropriate. The methodology proposed in Chapter 2 does not allow for structures other than the compound symmetric, which results from the random-effects assumption. Further discussion on future research incorporating time-series structures is provided in chapter 5.

The vector of residuals, for all subjects, $\boldsymbol{\varepsilon}$, is assumed to follow a distribution given by the product of the N univariate two-component normal-mixture densities for the ε_{ij}, given in equation (2.21) and expressed as

$$\phi_{JMX}\left(\boldsymbol{\varepsilon};\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2,\boldsymbol{\lambda}\right) = \prod_{j=1}^{S}\prod_{i=1}^{N_j} f\left(\varepsilon_{ij};\mu_{ij1},\mu_{ij2},\sigma_1^2,\sigma_2^2,\lambda_j\right).$$

The component variance-covariance structures, for all subjects, of the vector of residuals, $\boldsymbol{\varepsilon}$, is block-diagonal matrix given by

$$\boldsymbol{\Sigma}_k = \begin{bmatrix} \boldsymbol{\Sigma}_{1k} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{2k} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_{9k} \end{bmatrix}.$$

With the inclusion of the random subject effects and assuming the independence structure for $\mathbf{G}$, the vector of observations for all subjects, $\mathbf{Y}$, has component variance covariance structure $\mathbf{V}_k$ given by

$$\mathbf{V}_k = \mathbf{ZGZ}' + \mathbf{\Sigma}_k = \begin{bmatrix} \mathbf{V}_{1k} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{2k} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V}_{9k} \end{bmatrix}.$$

That is, observations from the same subject have variance $\sigma_\delta^2 + \sigma_k^2$ and covariance σ_δ^2, and observations from different subjects have variance $\sigma_\delta^2 + \sigma_k^2$ and covariance 0.

Fixed-effects are incorporated into the model by modeling the component means. The models for the component means, incorporating the levels of the fixed effect, are given by

$$\mu_{ijk} = \sum_{p=1}^{P} \alpha_{kp} x_{ijp}.$$

Here P is the total number of estimable fixed-effects parameters in the model. For the analysis data, the following fixed-effects are considered for the full model.

- Demog raphic Variables: age and gender
- Design Variables: task, period, and the task × period interaction

Thus, the over parameterized (singular) model includes 27 fixed-effects for each component: 1 intercept effect, 1 age effect, 2 gender effects, 5 task effects, 3 period effects, and 15 task × period interaction effects. This model is not full rank and thus 10 of the fixed-effects in each component are not estimable with this parameterization. The full-rank (nonsingular) model is obtained with reference cell parameterization. This model then has $P = 17$ estimable fixed-effects in each component: 1 intercept effect, 1 age effect, 1 gender effects, 4 task effects, 2 period effects, and 8 task × period interaction

effects. Using the reference cell parameterization, the component vectors of fixed-effects parameters, α_k, $k = 1,2$, are defined as $\alpha_k = (\alpha_{k1} \quad \alpha_{k2} \quad \cdots \quad \alpha_{k17})'$. Next, define the $N_j \times 17$ fixed-effects design matrix for the j^{th} as

$$\mathbf{X}_j = \begin{pmatrix} x_{1j1} & x_{1j2} & \cdots & x_{1jP} \\ x_{2j1} & x_{2j2} & \cdots & x_{2jP} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N_j j1} & x_{N_j j2} & \cdots & x_{N_j j1} \end{pmatrix}$$

where x_{ijp} denotes the value of the p^{th} fixed-effects value for the i^{th} observation on the j^{th} subject. Then, the fixed-effects design matrix for all subjects is defined as $\mathbf{X} = (\mathbf{X}_1 \quad \mathbf{X}_2 \quad \cdots \quad \mathbf{X}_9)'$. The following is used to define the values of x_{ijp}.

- I ntercept (p = 1) $x_{ij1} = 1$ for all observations on all subjects.
- Ag e (p = 2) x_{ij2} = Age of j^{th} subject – 19.22.
- Gender (p = 3) $x_{ij3} = 1$ if j^{th} subject is Female, $x_{ij3} = 0$ if j^{th} subject is Male.
- AS Task (p = 4) $x_{ij4} = 1$ if i^{th} RR-interval from the j^{th} subject is from the AS task, $x_{ij4} = 0$ otherwise.
- BS Task (p = 5) $x_{ij5} = 1$ if i^{th} RR-interval from the j^{th} subject is from the BS task, $x_{ij5} = 0$ otherwise.
- I S Task (p = 6) $x_{ij6} = 1$ if i^{th} RR-interval from the j^{th} subject is from the IS task, $x_{ij6} = 0$ otherwise.

- L S Task (p = 7) $x_{ij7} = 1$ if i^{th} RR-interval from the j^{th} subject is from the LS task, $x_{ij7} = 0$ otherwise.
- Baseline (p = 8) $x_{ij8} = 1$ if i^{th} RR-interval from the j^{th} subject is from the baseline period, $x_{ij8} = 0$ otherwise.
- Deliver y (p = 9) $x_{ij9} = 1$ if i^{th} RR-interval from the j^{th} subject is from the delivery period, $x_{ij9} = 0$ otherwise.
- AS × Baseline (p = 10) $x_{ij10} = 1$ if i^{th} RR-interval from the j^{th} subject is from the AS task and the baseline period, $x_{ij10} = 0$ otherwise.
- AS × Delivery (p = 11) $x_{ij11} = 1$ if i^{th} RR-interval from the j^{th} subject is from the AS task and the delivery period, $x_{ij11} = 0$ otherwise.
- BS × Baseline (p = 12) $x_{ij12} = 1$ if i^{th} RR-interval from the j^{th} subject is from the BS task and the baseline period, $x_{ij12} = 0$ otherwise.
- BS × Delivery (p = 13) $x_{ij13} = 1$ if i^{th} RR-interval from the j^{th} subject is from the BS task and the delivery period, $x_{ij13} = 0$ otherwise.
- I S × Baseline (p = 14) $x_{ij14} = 1$ if i^{th} RR-interval from the j^{th} subject is from the IS task and the baseline period, $x_{ij14} = 0$ otherwise.
- I S × Delivery (p = 15) $x_{ij15} = 1$ if i^{th} RR-interval from the j^{th} subject is from the IS task and the delivery period, $x_{ij15} = 0$ otherwise.

- L S × Baseline (p = 16) $x_{ij16} = 1$ if i^{th} RR-interval from the j^{th} subject is from the IS task and the baseline period, $x_{ij16} = 0$ otherwise.
- L S × Delivery (p = 17) $x_{ij17} = 1$ if i^{th} RR-interval from the j^{th} subject is from the IS task and the delivery period, $x_{ij17} = 0$ otherwise.

Due to the reference cell parameterization, the estimate for the k^{th} component intercept parameter, α_{k1}, represents the mean RR-interval length in the k^{th} component for a male subject with an average of 19.22 during the preparation period of the WS task. Then, the estimate for the age parameter, α_{k2}, represents the average increase in the k^{th} component intercept parameter each year increase in age from the average age of 19.22. The estimate for the gender parameter, α_{k3}, represents the average increase in the k^{th} component intercept parameter due to being a female rather than a male. The estimates for the task parameters, α_{k4}, α_{k5}, α_{k6}, and α_{k7}, each represents the average increases in the k^{th} component intercept parameter due to the tasks AS, BS, IS, and LS, respectively, rather than the WS task. The estimates for the period parameters, α_{k8}, and α_{k9}, each represents the average increases in the k^{th} component intercept parameter due to the baseline and delivery periods, respectively, rather than the preparation period. The estimates for the task × period parameters, α_{k10}, and α_{k11}, each represents the average increases in the k^{th} component intercept parameter due to the baseline and delivery periods of the AS, respectively, rather than the preparation period of the WS task. The estimates for the task × period parameters, α_{k12}, and α_{k13}, each represents the average

increases in the k^{th} component intercept parameter due to the baseline and delivery periods of the BS, respectively, rather than the preparation period of the WS task. The estimates for the task × period parameters, α_{k14}, and α_{k15}, each represents the average increases in the k^{th} component intercept parameter due to the baseline and delivery periods of the IS, respectively, rather than the preparation period of the WS task. The estimates for the task × period parameters, α_{k16}, and α_{k17}, each represents the average increases in the k^{th} component intercept parameter due to the baseline and delivery periods of the LS, respectively, rather than the preparation period of the WS task.

4.5. The EM-Algorithm for the Data

The EM-algorithm described in Chapter 2 is implemented for the analysis using SAS v.9. The SAS code to analyze the data in this chapter can be found on the data CD (see Appendix 7.7). Some aspects of the algorithm that are specific to the data and to SAS are described in this section. Within the M-step, in the EM-algorithm, two proc mixed procedures are fit to estimate the fixed-effects parameters and the variance-covariance parameters. Each model fits a random-effects model to a set of weighted observations. These weighted observations are the sum of the subject effects and the weighted residuals. A weight statement in the proc mixed procedure in conjunction with the original observations to produce these weighted observations is not correct since the random-effects are not weighted. Instead, the weights are applied to the residuals and the result is then added to the predicted subject effects to produce the weighted observations. Then the proc mixed procedure fits the weighted model using the weighted observations

as the response variables. To do this however, the design matrix must also account for these weights with the following adjustments to the code. First, the default intercept of ones used in the procedure is removed (noint option in the model statement) and a new intercept, defined as the square root weight vector) is incorporated into the model statement. Fixed-effects are then added to the model by interacting each effect with the new intercept parameter (that is add the age effect by adding the age*intercept effect), in order to weight these effects. The implication of removing the intercept parameter from the model adds one degree of freedom. However, the effect of interacting the intercept effect with all the other fixed-effects replaces that degree of freedom. Finally, since this is an unbalanced design, the Satterthwaite method for determining the denominator degrees of freedom is used (ddfm = satterth option in the model statement).

4.6. Research Questions

Using the Loneliness data set the several research questions are of interest. Primarily, interest is in how the fixed-effects affect mean changes in the RR-intervals. A mean change is considered significant when both component means are significantly affected by the fixed-effect of interest. The following research questions are focused on in this dissertation.

1.) Do the demographic variables, age and gender, have an effect on mean changes in RR-intervals?

2.) Is the mixture proportion equal for each subject?

3.) Do task and period have an effect on mean changes in RR-intervals? If so, is the period effect consistent across the tasks?

4.) Is there a significant amount of variability in the RR-intervals due to the subject effects?

5.) Are the estimates of the within subject variance parameters significantly different across the components?

These research questions will be translated into statistical hypotheses and likelihood ratio tests will be provided.

4.7. Methods for Model Selection

In this section, the steps used for selecting the model for the analysis of the Loneliness data set are described. The full-model is defined as the model containing the following.

- Fixed -effects parameters for age, gender, task, period, and task × period
- An independent va riance-covariance structure for the component densities, $\mathbf{\Sigma}_{1j} = \sigma_1^2 \mathbf{I}_{N_j}$ and $\mathbf{\Sigma}_{2j} = \sigma_2^2 \mathbf{I}_{N_j}$. Here, σ_k^2 describes the within-subject component variance of the residuals.
- One random subje ct effect with a random-effects variance, $\mathbf{G} = \sigma_\delta^2 \mathbf{I}_j$.
- A separ ate mixture proportion for each subject λ_j.

First, the full model will be fit with the EM-algorithm proposed in section 2.8.2, using a macro written in SAS. The full model will be reduced to the final model through a series of steps. First, tests of the fixed-effects parameters will be performed to determine their

significance in the model. Next, the assumption for a varying mixture proportion will be examined. If the test indicates that a single mixture proportion is appropriate, then the fixed-effects will be re-examined for significance in the model. These steps used to obtain the final model are described in more detail below.

After fitting the full-model, first the fixed-effects are considered for removal. The fixed-effects are tested using a likelihood ratio test. Here, the likelihood ratio test (LRT) is used to simultaneously test if a fixed-effect significantly affects mean changes in the RR-intervals in both components. By definition, the likelihood of the data increases as the number of model parameters increases. The LRT then determines if this increase is significant. If the LRT statistic is not significant (at a significance level of 0.05) then the effect is removed from the model (since the inclusion of the additional effect would not significantly increase the likelihood). The LRT compares the likelihoods of the full model and the reduced model using a chi-squared test with degrees of freedom, r, equal to the difference in the number of parameters in the two models. For example, to test if age should be removed from the model, the reduced model would contain 2 fewer parameters than the full model since there is an age effect estimate for each component. Define the vector $\hat{\theta}$ as the vector of parameter estimates obtained from the full model and $\hat{\theta}^*$ as the vector of parameter estimates for the reduced model. Here, both $\hat{\theta}$ and $\hat{\theta}^*$ contain the fixed-effects parameter estimates, as well as estimates for the variance parameters and the mixture proportion parameters. Recall, that LRTs in the two-component normal-mixture random-effects case involves comparing likelihoods conditioned on the random-effects, that is comparing the likelihoods of the model

parameters given the residuals, $\varepsilon_{ij} = y_{ij} - \beta_j$, instead of the observations, y_{ij}. The conditional likelihood of the full model is

$$\left[L\left(\hat{\theta},\mathbf{y}|\beta\right)\right]=\left[L\left(\hat{\theta},\varepsilon\right)\right]=\prod_{j=1}^{9}\prod_{i=1}^{N_j}\left[\hat{\lambda}_j\phi\left(\varepsilon_{ij};\hat{\mu}_{ij1},\hat{\sigma}_1^2\right)+\left(1-\hat{\lambda}_j\right)\phi\left(\varepsilon_{ij};\hat{\mu}_{ij2},\hat{\sigma}_2^2\right)\right].$$

The likelihood for the reduced model is obtained in the same fashion using the vector $\hat{\theta}^*$. The log of the likelihood for the full model is then defined as

$$\log\left[L\left(\hat{\theta},\varepsilon\right)\right]=\sum_{j=1}^{9}\sum_{i=1}^{N_j}\log\left[\hat{\lambda}_j\phi\left(\varepsilon_{ij};\hat{\mu}_{ij1},\hat{\sigma}_1^2\right)+\left(1-\hat{\lambda}_j\right)\phi\left(\varepsilon_{ij};\hat{\mu}_{ij2},\hat{\sigma}_2^2\right)\right].$$

Then the LRT testing the significance of an effect on mean changes in RR-intervals in both components is defined as

$$-2\left[\log L\left(\hat{\theta};\varepsilon\right)-\log L\left(\hat{\theta}^*;\varepsilon\right)\right]=-2\log\left(\frac{L\left(\hat{\theta};\varepsilon\right)}{L\left(\hat{\theta}^*;\varepsilon\right)}\right)\sim\chi_r^2.$$

If the likelihood ratio statistic is significantly larger than χ_r^2, using a significance level of 0.05, then the additional effects in the full model do not significantly improve the fit of the data, and the effects are removed from the model.

After considering the fixed-effects for removal, call the reduced model obtained Model A. Next, the mixture proportions for the nine subjects are examined. The full model, and hence Model A, allow the mixture proportion for each subject to vary. To test if the mixture proportions are equal across subjects, a LRT is performed. Here, $\hat{\theta}$ is defined as the parameter vector for Model A containing the nine mixture proportions, and

$\hat{\theta}^*$ as the parameter vector for the model containing one mixture proportion for all subjects. The difference between the number of parameters in the two models is $r = 8$. If the LRT is significant compared to a χ^2_8, then the model with varying mixture proportions (for all nine subjects) fits significantly better than the model with a single mixture proportion (for the nine subjects). If the LRT is not significant, then there is not a significant difference in the fit of the models and the model can be reduced to one containing a single mixture proportion.

After considering the mixture proportions, call the reduced model obtained Model B. At this point, the fixed-effects parameters remaining in Model B are re-examined for significance since reducing the model to contain a single mixture proportion may have an effect on the results of marginally significant fixed-effects parameters. The LRTs used to obtain Model A are utilized again with Model B and the resulting model is called the final model.

4.8. Methods for Presenting the Results

The results for the final model obtained through the model selection procedure described in section 4.7 will be presented with various tables and graphs. The first part of the results section will contain information regarding the model selection procedure. This includes tables of the LRTs for the fixed-effects parameters and the mixture proportions. The remaining part of the results section then will present the results for the final model. This will be accomplished by tabulating the fixed-effects parameter estimates from each component with their corresponding standard errors. The within subject component

variance parameters and the between subject variance parameter will also be tabulated with their standard errors and 95% Wald type confidence intervals.

If the LRT for the mixture proportions is not significant and the final model contains a single mixture proportion, then the analysis results will be presented in the following manner. First, the predicted values (LS means) for the two components are tabulated by the levels of the significant categorical fixed-effects at the average levels of the significant continuous variables. Using the information contained in these tables, LS means plots will be provided to aid in interpretation of the parameter estimates. Next, the overall mixture predicted values are determined as a weighted average of the component predicted values determined by the estimated mixture proportion. LS means plots of the overall mixture means will also be provided for interpretation.

If the LRT for the mixture proportions is significant and the final model contains varying mixture proportions for each subject, then the analysis results will be presented in the following manner. First, the predicted values (LS means) for the two components are tabulated by the levels of the significant categorical fixed-effects at the average levels of the significant continuous variables. Using the information contained in these tables, LS means plots will be provided to aid in interpretation of the parameter estimates. Next, the overall mixture predicted values are determined for each subject as a weighted average of the component predicted values determined by the estimated subject mixture proportions. LS means plots of the overall mixture means for each subject will also be provided for interpretation.

Finally, a diagnostic check of the independence assumption for the residuals within the subjects will be performed. Here the residuals are computed as the difference between the observed RR-interval and the predicted subject effects, $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\beta}}$. Then a diagnostic plot of ε_{ij} versus $\varepsilon_{i-1\,j}$, for each combination of subject, task, and period, displaying no discernable trend would indicate that the assumption of independence for the residuals within each combination has been met.

4.9. Results

Let R1a be the reduced model with age effects removed, R2a be the reduced model with gender effects removed, and R3a be the reduced model with the task × period interaction effects removed. The -2 log Likelihood values for the full model and the reduced models R1a, R2a, and R3a are displayed in the third column of Table 9. Likelihood ratio tests for each of the parameters, age, gender, and task × period, are computed by determining the difference in the -2 log likelihoods between the full model and each of the reduced models, R1a, R2a, and R3a, respectively.

Table 9: Likelihoods: Full Model, R1a, R2a, and R3a

Effect Removed	Model (# parameters)	- 2 Log Likelihood	Difference	DF	p-value
-	Full (46)	146769.87	-	-	-
Age	R1a (44)	146778.18	8.31	2	0.0157
Gender	R2a (44)	146941.63	171.76	2	<0.0001
Task × Period	R3a (30)	148016.50	1246.63	16	<0.0001

These differences, shown in Table 9, approximately follow chi-squared distributions each with degrees of freedom (DF) equal to the difference in the number of parameters in the

two models. The degrees of freedom used for the LRTs of age, gender, and task × period along with corresponding p-values are summarized in Table 9. The LRTs for each of the fixed-effects indicate that all three effects, age, gender, and task × period, are significant and should remain in the model. Thus, Model A is equivalent to the full model.

Next, the assumption for varying mixture proportions is examined. The estimates obtained from Model A of the mixture proportions for each subject are listed in Table 10. The mean of the nine estimated subject mixture proportions in Table 10 is

$$\bar{\hat{\lambda}} = \sum_{j=1}^{9} \hat{\lambda}_j = 0.4832 \text{ and the variance is } \sigma^2_{\hat{\lambda}_j} = \frac{1}{9-1}\sum_{j=1}^{9}\left(\hat{\lambda}_j - \bar{\hat{\lambda}}\right)^2 = 0.0030.$$

Table 10: Estimated Subject Miture Proportions $\left(\hat{\lambda}_j\right)$ Full Model

Subject	$\hat{\lambda}_j$	SE($\hat{\lambda}_j$)
1	0.4878	0.0093
2	0.4773	0.0084
3	0.5333	0.0096
4	0.4983	0.0094
5	0.4418	0.0082
6	0.4446	0.0090
7	0.5219	0.1022
8	0.5628	0.0086
9	0.3809	0.0091

A model fit with a single mixture proportion for all nine subjects yields a -2 log likelihood of 146772.50. Here, the estimate of the (overall) mixture proportion is $\hat{\lambda} = 0.4796$ (SE = 0.0030). This reduced model, compared to the full model (including all fixed-effects), has a chi-squared value of 2.63 with 8 degrees of freedom, p = 0.9554. This indicates that a model with a varying mixture proportions does not fit significantly

better than a model with a single mixture proportion. Thus, Model A is reduced to Model B by including only a single mixture proportion.

The significance of the fixed-effects parameters are re-examined again using likelihood ration tests. Let R1b be the reduced model with age effects removed, R2b be the reduced model with gender effects removed, and R3b be the reduced model with the task × period interaction effects removed. The -2 log Likelihood values for Model B and the reduced models R1b, R2b, and R3b are displayed in the third column of Table 11. Likelihood ratio tests for each of the parameters, age, gender, and task × period, are computed by determining the difference in the -2 log likelihoods between Model B and each of the reduced models, R1b, R2b, and R3b, respectively.

Table 11: Likelihoods: Full Model, R1b, R2b, and R3b

Effect Removed	Model (# parameters)	- 2 Log Likelihood	Difference	DF	p-value
-	B (38)	146772.50	-	-	-
Age	R1b (36)	146780.13	7.63	2	0.0220
Gender	R2b (36)	146922.06	149.56	2	<0.0001
Task × Period	R3b (22)	148038.12	1265.62	16	<0.0001

These differences, shown in Table 11, approximately follow chi-squared distributions with degrees of freedom equal to the difference in the number of parameters in the two models. The degrees of freedom (DF) used for the LRTs for age, gender, and task × period along with corresponding p-values are summarized in Table 11. The LRTs for each of the fixed-effects, age, gender, and task × period indicate that all three effects significantly affect changes in RR-intervals in both components and should remain in the model. Thus, the final model is equivalent to Model B. Next, the results of the final model are presented.

The estimates and standard errors (SE) for the fixed-effects parameters of the final model are tabulated for the first component, $\hat{\alpha}_{p1}$, and the second component, $\hat{\alpha}_{p2}$, in Table 12.

Table 12: Estimates and SEs of the Fixed-Effects Parameters

Fixed-Effects Parameters	**First Component** $\hat{\alpha}_{p1}$ (SE)	**Second Component** $\hat{\alpha}_{p2}$ (SE)	**Overall** $\hat{\alpha}_p$ (SE)
Intercept	68.0803 (0.1743)	89.9328 (0.1494)	79.4523
Age	1.2311 (0.0366)	1.0969 (0.0396)	1.1613
Female	2.6716 (0.1212)	1.6019 (0.1315)	2.1149
AS	-0.2388 (0.2095)	-2.5009 (0.1792)	-1.4160
BS	9.0418 (0.2086)	3.9175 (0.1942)	6.3751
IS	8.1623 (0.2108)	2.4166 (0.1835)	5.1722
LS	5.5512 (0.2040)	-0.6292 (0.1889)	2.3349
Baseline	10.5379 (0.2093)	7.9418 (0.1913)	6.0645
Delivery	-0.8538 (0.2009)	-5.3442 (0.1869)	-3.1906
AS × Baseline	-1.6546 (0.2949)	-0.9145 (0.2645)	-1.2695
AS × Delivery	-4.3598 (0.2739)	-2.0977 (0.2731)	-3.1826
BS × Baseline	-5.8747 (0.2693)	-3.3277 (0.2628)	-4.5492
BS × Delivery	-10.5592 (0.2726)	-3.0973 (0.2883)	-6.6760
IS × Baseline	-9.6416 (0.2853)	-4.6925 (0.2737)	-7.0661
IS × Delivery	-5.6413 (0.2786)	-0.6369 (0.2711)	-3.0370
LS × Baseline	-5.9216 (0.2824)	-1.6392 (0.2756)	-3.6930
LS × Delivery	-8.9997 (0.2738)	-0.7912 (0.2740)	-4.7280

Recall that the intercept effects describe the component average RR-interval length for a male with an average age of 19.22 during the WS task and preparation period. Also recall that the original RR-interval data have been rescaled by a factor of 1/10 and thus are reported in centiseconds rather than milliseconds. The overall estimates of the fixed-effects parameters, $\hat{\alpha}_p$, are determined as weighted sums of the component estimates, $\hat{\lambda}\hat{\alpha}_{1p} + \left(1-\hat{\lambda}\right)\hat{\alpha}_{2p}$, where $\hat{\lambda} = 0.4796$ is the estimated mixture proportion.

The estimates and standard errors for the variance parameters are given in Table 13, along with 95% Wald confidence intervals.

Table 13: Estimates of Variances Parameters $\left(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_\delta^2\right)$

Variance Parameter	Estimate	SE	95% Confidence Interval
$\hat{\sigma}_1^2$	22.6284	0.2273	(22.1894, 23.0806)
$\hat{\sigma}_2^2$	24.1271	0.2424	(23.6591, 24.6093)
$\hat{\sigma}_\delta^2$	11.1801	4.9689	(5.4781, 34.1265)

The predicted values (LS means) and standard errors for the first and second components, at the mean age of 19.22, are tabulated in Table 14 and Table 15, respectively.

Table 14: First Component Means LS (Predicted Values) and SEs

Task	Period		
	Baseline	Preparation	Delivery
	Males		
AS	77.9559 (0.7159)	69.0726 (0.1596)	63.8590 (0.1483)
BS	83.0164 (0.1325)	78.3532 (0.1607)	66.9402 (0.1493)
IS	78.3700 (0.1559)	77.4737 (0.1638)	70.9786 (0.1545)
LS	79.4789 (0.1632)	74.8626 (0.1574)	65.0091 (0.1584)
WS	79.8493 (0.1558)	69.3114 (0.1765)	68.4576 (0.1566)
	Females		
AS	80.6275 (0.1861)	71.7442 (0.1646)	66.5306 (0.1477)
BS	85.6880 (0.1344)	81.0248 (0.1614)	69.6118 (0.1477)
IS	81.0416 (0.1639)	80.1453 (0.1697)	73.6502 (0.1590)
LS	82.1505 (0.1697)	77.5342 (0.1560)	67.6807 (0.1549)
WS	82.5209 (0.1631)	71.9830 (0.1817)	71.1292 (0.1590)

In these tables, the LS means are displayed by gender, across all combinations of task and period, with standard errors reported in parenthesis.

Table 15: Second Component LS Means (Predicted Values) and SEs

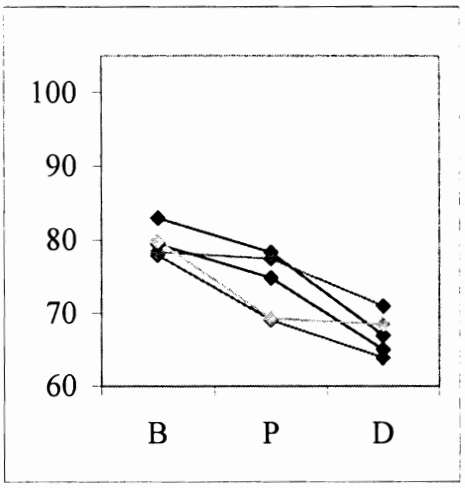
Task	Period		
	Baseline	Preparation	Delivery
	Males		
AS	95.5561 (0.1591)	88.5288 (0.1603	81.0869 (0.1760)
BS	99.5613 (0.1365)	94.9472 (0.1770)	86.5057 (0.1781)
IS	96.6956 (0.1724)	93.4463 (0.1643)	87.4652 (0.1683)
LS	96.7031 (0.1654)	90.4005 (0.1691)	84.2651 (0.1646)
WS	98.9715 (0.1759)	91.0297 (0.1521)	85.6855 (0.1672)
	Females		
AS	97.1580 (0.1589)	90.1307 (0.1615)	82.6888 (0.1842)
BS	101.1632 (0.1407)	96.5491 (0.1818)	88.1076 (0.1887)
IS	98.2975 (0.1723)	95.0482 (0.1654)	89.0671 (0.1723)
LS	98.3050 (0.1675)	92.0024 (0.1783)	85.8670 (0.1727)
WS	100.5734 (0.1750)	92.6316 (0.1557)	87.2874 (0.1725)

The predicted LS means for the first component, shown in Table 14, at an average age of 19.22, are plotted separately in Figure 28 for males, panel a.), and females, panel b.), across each period, baseline (B), preparation (P), and delivery (D), with separate lines for each for the five tasks. The predicted LS means for the second component, shown in Table 15, at an average age of 19.22 are plotted separately in Figure 28 for males, panel c.), and females, panel d.), across each period, baseline (B), preparation (P), and delivery (D), with a separate line for each task.

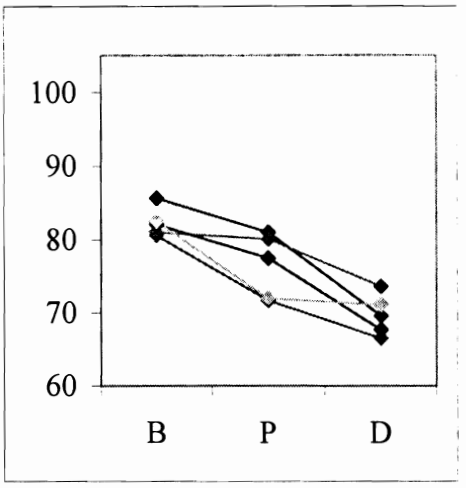
The overall predicted values, computed at an average age of 19.22, are equivalent across subject, since the final model contains a single mixture proportion for the nine subjects. The overall predicted value (overall mixture LS means) are computed as $\hat{\mu}_{fuv} = \hat{\lambda}\hat{\mu}_{1\,fuv} + \left(1-\hat{\lambda}\right)\hat{\mu}_{2\,fuv}$, where gender is denoted by f, task is denoted by u, and period is denoted by v.

Figure 28: Component LS Mean Plots

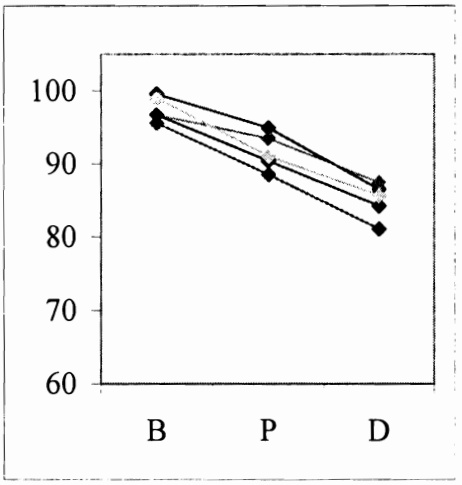
a.) Males – First Component



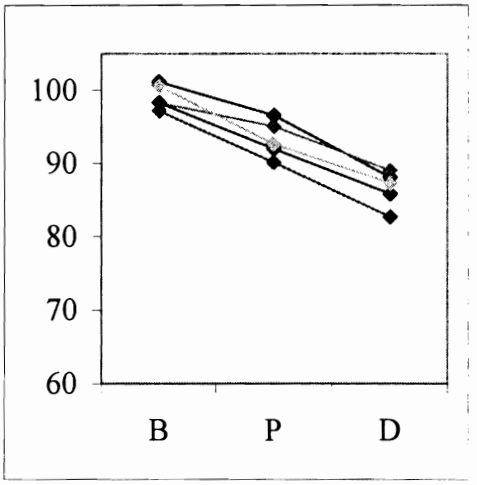
b.) Female- First Component

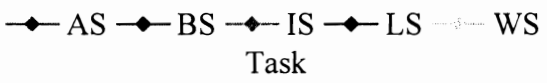


c.) Males – Second Component



d.) Females – Second Component





Task

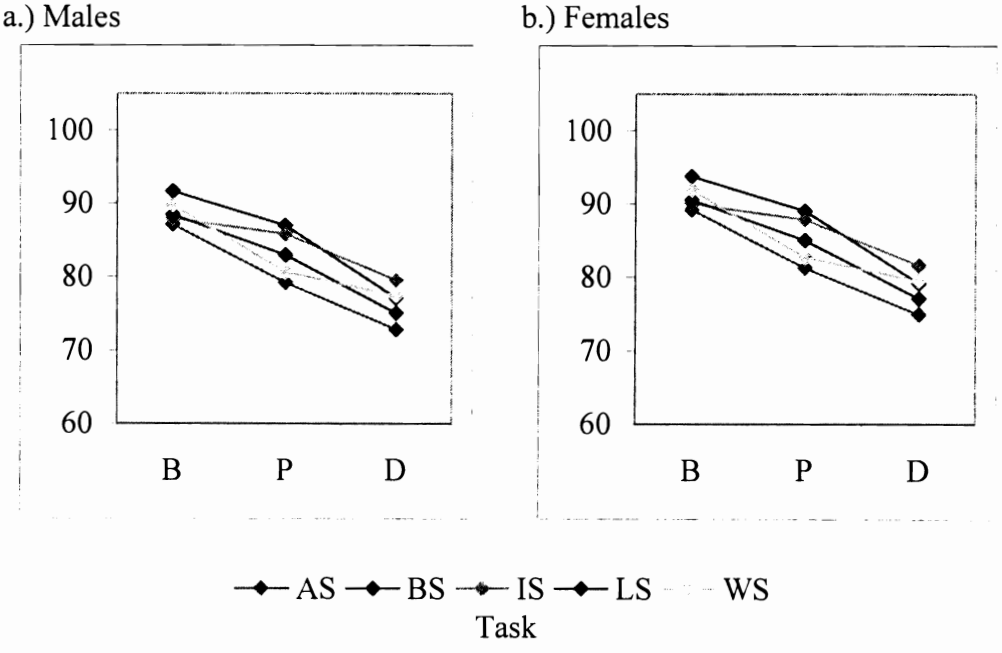
Using $\hat{\lambda} = 0.4796$, the overall mixture LS means are displayed in Table 16 by gender, across all combinations of task and period, with standard errors reported in parenthesis.

Table 16: Overall Mixture LS Means (Predicted Values)

Task	Period		
	Baseline	Preparation	Delivery
	Males		
AS	87.1156	79.1983	72.8250
BS	91.6269	86.9893	77.1227
IS	87.9073	85.7864	79.5588
LS	88.4430	82.9490	75.0306
WS	89.8011	80.6143	77.4236
	Females		
AS	89.2305	81.3132	74.9399
BS	93.7418	89.1042	79.2376
IS	90.0222	87.9013	81.6737
LS	90.5578	85.0639	77.1455
WS	91.9160	82.7292	79.5385

The overall mixture LS means, shown in Table 16, at an average age of 19.22, are plotted separately in Figure 29 for males, panel a.), and females, panel b.), across each period, baseline (B), preparation (P), and delivery (D), with a separate line given for each of the five tasks.

Figure 29: Overall Mixture Mean Plots



Next, the process for determining each subjects predicted RR-intervals density, using the predicted subject effects is demonstrated for the baseline period of the BS task. First, the component LS means in Table 14 and Table 15, during the baseline period of the BS task are computed for each subject, by incorporating their actual age rather than the average age of 19.22. To do this, first the LS means for the baseline period of the BS task for each gender are adjusted by removing the age effect. The estimated age effects for the first and second components are 1.2311 and 1.0969, respectively. The adjusted component LS means are obtained by subtracting the component age effects from the LS means in Table 14 and Table 15. The adjusted component LS means are presented in Table 17.

Table 17: LS Means for Baseline Period BS Task with Age Effect Removed

Gender	Component 1	Component 2
Males	81.7853	98.4644
Females	84.4569	100.0663

Now, age differences for each subject are computed as the subject's age minus the mean age of 19.22. Then, for each subject, and depending on their gender, the component means in Table 17 are added to the product of the subject's age difference with the component age effect to produce the component means accounting for each subject's age effect. The mixture mean is then computed as $\hat{\mu}_{juv} = \hat{\lambda}\hat{\mu}_{j1uv} + \left(1-\hat{\lambda}\right)\hat{\mu}_{j2uv}$, where subject is denoted by j, task is denoted by u, and period is denoted by v. The results of these computations are displayed in Table 18.

Table 18: LS Means for Baseline Period BS Task with Subject's Age Effects

Subject	Age	Age difference	Gender	Component 1 LS Means	Component 2 LS Means	Mixture LS Means
1	19	-0.2222	Female	84.1833	99.8226	92.3220
2	19	-0.2222	Female	84.1833	99.8226	92.3220
3	19	-0.2222	Male	81.5117	98.2207	90.2071
4	24	4.7778	Male	87.6672	103.7052	96.0134
5	19	-0.2222	Female	84.1833	99.8226	92.3220
6	18	-1.2222	Female	82.9522	98.7257	91.1607
7	19	-0.2222	Female	84.1833	99.8226	92.3220
8	18	-1.2222	Male	80.2806	97.1238	89.0458
9	18	-1.2222	Male	80.2806	97.1238	89.0458

Next, the predicted values for each subject are determined as the age adjusted component means and mixture means displayed in Table 18 shifted by the predicted random subject effects. The predicted values incorporating the random subject effects for the two components and the overall mixture mean during the baseline minutes of the BS task are displayed in Table 19.

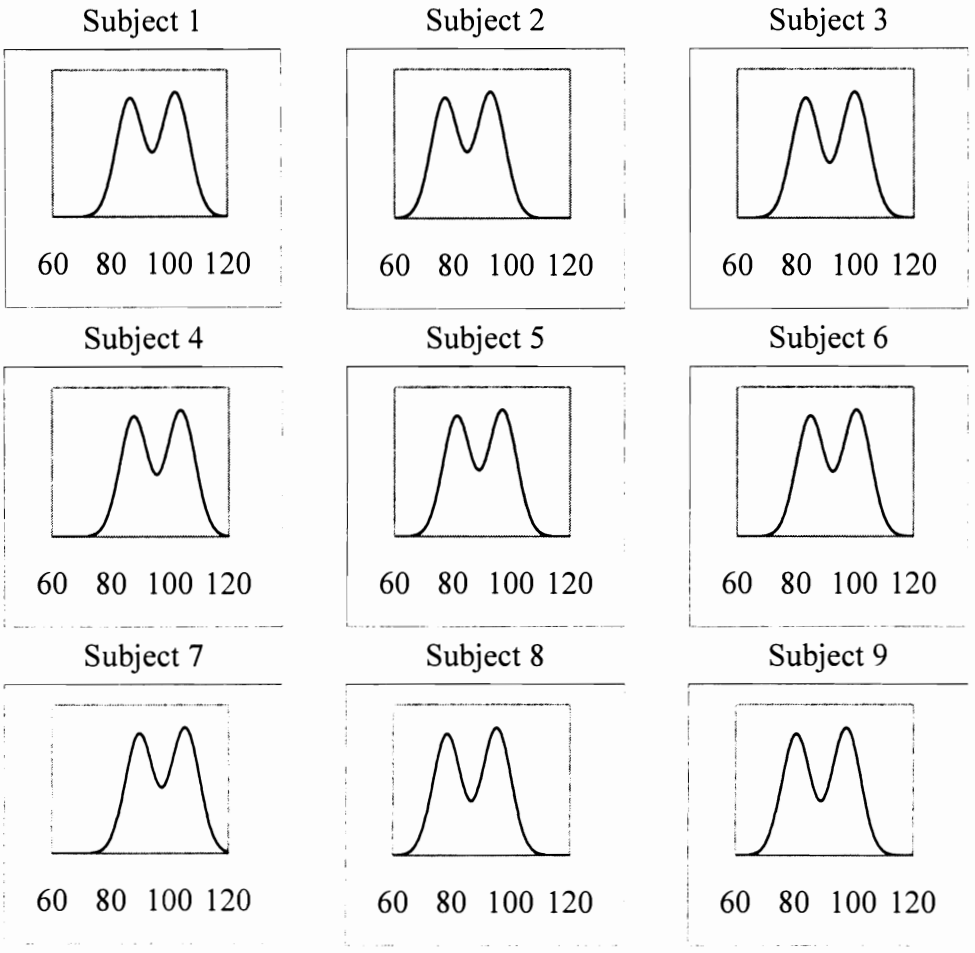
Table 19: LS Means for Baseline Period BS Task with Random Subject Effects

Subject	Subject Effects	Component 1 LS Means	Component 2 LS Means	Mixture LS Means
1	2.2842	86.4675	102.1068	94.6062
2	-6.9379	77.2454	92.8847	85.3841
3	1.6497	83.1614	99.8704	91.8568
4	-0.0186	87.6486	103.6866	95.9948
5	-2.8686	81.3147	96.9540	89.4534
6	1.8793	84.8315	100.6050	93.0400
7	5.3793	89.5626	105.2019	97.7013
8	-1.9570	78.3236	95.1668	87.0888
9	0.1986	80.4792	97.3224	89.2444

Then, using the component means in Table 19, the mixture densities for the nine subjects can be determined. This is accomplished by plotting the two-component normal-mixture

densities, $f\left(\hat{\mu}_{j1},\hat{\mu}_{j2},\hat{\sigma}_1^2,\hat{\sigma}_2^2,\hat{\lambda}\right)=\hat{\lambda}f\left(\hat{\mu}_{j1},\hat{\sigma}_1^2\right)+\left(1-\hat{\lambda}\right)\left(\hat{\mu}_{j2},\hat{\sigma}_2^2\right)$, for each of the nine subjects, where $f\left(\hat{\mu}_{jk},\hat{\sigma}_k^2\right)$ is standard notation for the normal density with mean $\hat{\mu}_{jk}$ and variance $\hat{\sigma}_k^2$. Here, $\hat{\mu}_{jk}$ denotes the k^{th} component LS mean (predicted value) for the j^{th} subject. Using the estimates for the mixture proportion, $\hat{\lambda} = 0.4796$, and component variance parameters, $\hat{\sigma}_1^2 = 22.6284$ and $\hat{\sigma}_2^2 = 24.1271$, the nine estimated subject densities are displayed in Figure 30.

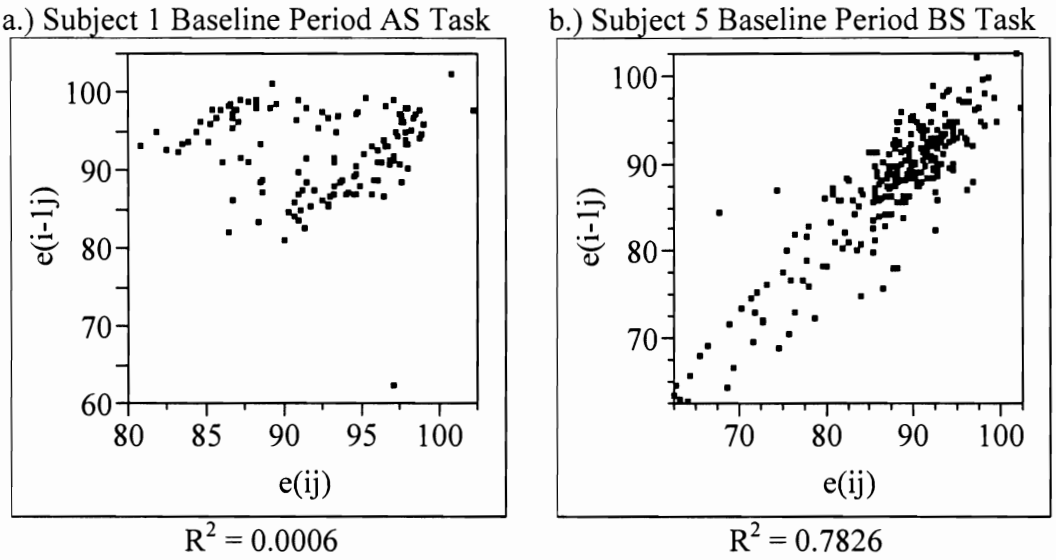
Figure 30: Predicted mixture distributions by subject



This same process for determining each subjects predicted RR-interval density could be repeated for each combination of task and period.

In order to assess the independence assumption of the residuals within each subject a simple diagnostic analysis is performed by examining the lag in the residuals. Here the residuals are computed as the difference between the observed RR-interval and the predicted subject effects, $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\beta}}$. Recall, the vector of residuals, $\boldsymbol{\varepsilon}$, then predicted to follow a two-component normal-mixture density with component means $\mathbf{X}\hat{\boldsymbol{\alpha}}_1$ and $\mathbf{X}\hat{\boldsymbol{\alpha}}_2$, component variances $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$, and mixture proportion $\hat{\lambda}$. Since the assumption of independence is on a subject's series of RR-interval data within a specific combination of task and period, regression fits of ε_{ij} versus ε_{ij-1} are conducted by subject, task, and period ($9 \times 5 \times 3 = 135$ fits). The range of the slopes and the R-squares for the 135 fits are (-0.0278, 0.9869) and (0.0006, 0.7826), respectively. Plots for the combinations with the smallest and largest correlations are presented in Figure 31. Subject 1, during the baseline period of the BS task, exhibited the least amount of correlation within the series of RR-interval data, panel a.). Subject 5, during the baseline period of the BS task, exhibited the most correlation, panel b.). The wide ranges in the slopes and R-squares of the fits indicate that the independence assumption in the residuals, and hence the compound symmetric structure for the observations may not be sufficient, and an autoregressive correlation structure is likely.

Figure 31: Diagnostic Plots of ε_{ij} versus $\varepsilon_{i-1\,j}$



4.10. Conclusions

In this section, the results of the fit for the RR-interval data from the loneliness study using a two-component normal-mixture random-effects density are interpreted. The initial fit with all fixed-effects parameters and a varying mixture proportion for each subject indicated that the fixed-effects, age, gender, task, period, and task × period, are all significant effects in the model (Table 9). That is to say, each of these fixed-effects significantly affects changes in RR-intervals in both components.

Next, the assumption for a varying mixture proportion was examined. The estimated mixture proportions for the nine subjects were displayed in Table 10. A likelihood ratio test determined that the model with varying mixture proportions did not fit significantly better than a model with a single mixture proportion. Thus, a model with a single mixture proportion was chosen as the final model. This choice seems appropriate since the range of the estimates of the mixture proportions for the nine

subjects was small (0.3809 – 0.5628). The estimated mixture proportion for all nine subjects is $\hat{\lambda} = 0.4796$, indicating that the first and second components have mixture percents of 47.96 and 52.04. Standard errors for the mixture proportions estimates were not provided, however the process for determining them will be presented in Chapter 5.

The fixed-effects were re-examined again after reducing the model to a single mixture proportion. The likelihood ratio tests indicated that the fixed-effects, age, gender, and task × period, all remain significant in a model with a single mixture proportion (Table 11).

The fixed-effects parameter estimates for the final model were tabulated along with overall mixture fixed-effects parameter estimates (Table 12). The estimates for the overall mixture fixed-effects were determined using a weighted average of the two component parameter estimates, determined by the estimated mixture proportion, $\hat{\lambda} = 0.4796$. Standard errors for the estimates for the overall mixture fixed-effects were not provided and will be discussed in Chapter 5.

In order to best interpret the effect of the age, gender, task, and period on changes in RR-intervals, the component predicted values (LS means) are computed for males and females across the various combinations of task and period (Table 14 and Table 15) using a mean age of 19.22. These LS means were then plotted to visually display the effect of gender, task, and type (Figure 28).

The significance of the gender effect based on a likelihood ratio test ($p < 0.0001$) indicates that there is a significant difference in mean RR-intervals between males and females (regardless of component). From the plots in Figure 28 it is apparent that

females exhibit significantly greater mean RR-interval length, and hence slower heart rates, than males. The estimates of the gender effect differences (Females – Males) in the first and second components are 2.6716 and 1.6019 respectively. Thus, on average, RR-intervals for females in the first component are 2.6717 centiseconds (26.717 milliseconds) longer than RR-intervals for males, and 1.6019 centiseconds (16.016 milliseconds) longer in the second component. The overall mixture age effects is 2.1149 indicating that, on average, females exhibit 2.1149 centiseconds (21.149 milliseconds) longer RR-intervals than Males.

The significance of the age effect based on a likelihood ratio test ($p = 0.0220$) indicates that there are significant differences in mean RR-intervals due to age. The estimates for the age effects in the first and second components are 1.2311 and 1.0969, respectively, and both positive. Thus, the model for these data indicates that, for each year increase in age, RR-intervals, on average, increase in the first component 1.2311 centiseconds (12.311 milliseconds) and 1.0969 centiseconds (10.969 milliseconds) in the second component. It should be noted however, that there is not a large variability in the ages of the subjects. Of the nine subjects in the study, three are age 18, five are age 19, and one is age 24.

The analysis suggested that task and period are significant fixed-effects in the model through an interaction. The significance of the task × period interaction effect based on a likelihood ratio test ($p < 0.0001$) indicates that there are significant differences in component mean RR-intervals due to period, and these effects are not consistent across task (or vice versa). The plots of the means (by gender) for each task across the periods

in Figure 28 indicate that mean RR-intervals tend to decrease from baseline to preparation to delivery, regardless of task. The effect of the significant task × period interaction can be further interpreted by examining the average component difference in the predicted RR-intervals between the periods, across the various tasks. The average changes in RR-intervals between baseline and preparation (B – P), preparation and delivery (P – D), and baseline and delivery (B – D), for each task, by component, are given in Table 20.

Table 20: Task × Period RR-Interval Changes by Component

Task	Component 1 Differences			Component 2 Differences		
	B – P	P – D	B – D	B – P	P – D	B – D
AS	8.8833	5.2136	14.0969	7.0273	7.4419	14.4692
BS	4.6632	11.4130	16.0762	4.6141	8.4415	13.0556
IS	0.8963	9.8535	7.3914	3.2493	5.9811	9.2304
LS	4.6163	9.8535	14.4698	6.3026	6.1354	12.4380
WS	9.5379	0.8538	11.3917	7.9418	5.3442	13.2860

From Table 20 it is apparent, for both components, that the IS task exhibits the smallest decreases in average RR-intervals between baseline to preparation and between baseline and delivery. To determine which period differences are significantly different among the five tasks appropriate contrasts should be performed with F-tests. These will be discussed more in Chapter 5.

The estimates for the variance parameters are shown in Table 13. The estimate for the first component variance parameter is smaller than the estimate for the second component variance parameter. The confidence intervals for the two component variance parameters, $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$, do not overlap indicating that $\hat{\sigma}_1^2$ is significantly smaller than $\hat{\sigma}_2^2$. The 95% confidence interval on the estimate of the between-subject variance

parameter, $\hat{\sigma}_{\delta}^{2}$, does not include zero. Thus, indicating that there is a significant amount of variability in the RR-intervals between subjects.

The diagnostic analysis of the model indicated a more complex time-series type of correlation structure for the residuals within a subject. Several options are available and will be further discussed in Chapter 5.

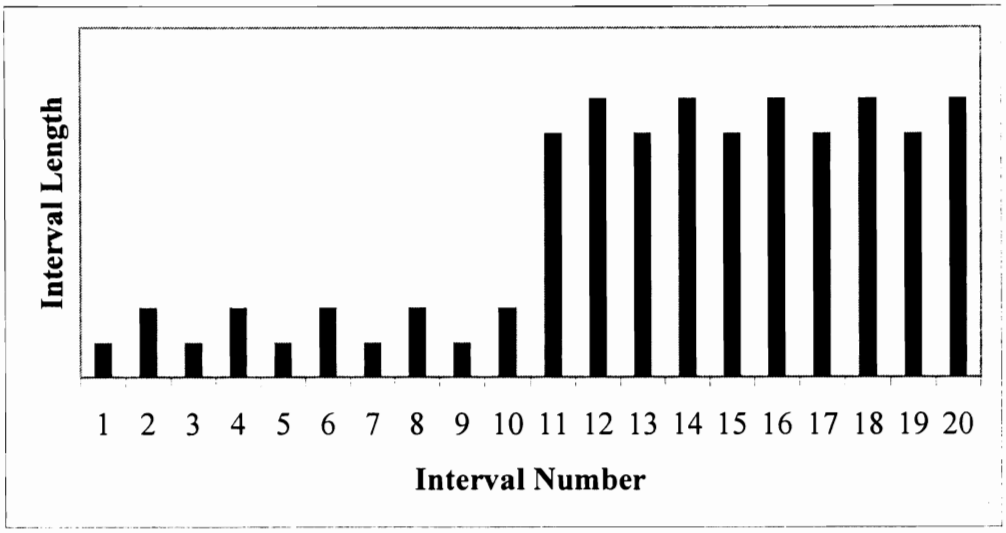
4.11. Motivational Data

There are two data sets, discussed briefly in 1.2, which provided the motivation for this research. Illustration of the proposed methodology however used a different data set. In this section the reasons for not using these data sets for illustration are discussed.

The first motivational data set came from a sample of sedated adult hospital inpatients selected randomly for a study conducted by the School of Nursing at Virginia Commonwealth University (Grap et al., 2006). During the study, each patient's ECG signal was recorded for periods of up to 24 hours. The issues with this data set primarily involve the length of the recording period. Ideally, short recording periods should be used for the type of models described in this research. The algorithm is used to identify the bimodal pattern within the high-frequency variations in the RR-intervals. Increasing the recording length introduces lower frequency variations in the data, which would overpower the high-frequency variations in the data. Consider the (simplified) example in Figure 32. In this example, 20 RR-intervals are displayed with interval length plotted on the vertical axis and interval number on the horizontal axis. Here, the variations seen within the first 10 intervals and seen within the last 10 intervals represent the high-

frequency variations in the RR-intervals. The variations seen between the first 10 intervals and the last 10 intervals represent the introduction of the low frequency variations in the RR-intervals.

Figure 32: Lower Frequency Variations



The EM-algorithm, used with this example, would incorrectly identify the first 10 intervals as members of the first component and the last 10 intervals as members of the second component. The algorithm should identify the odd numbered intervals as members of the first component and the even numbered intervals as members of the second component. Thus, in order to use long recording periods, methods should be implemented to isolate the high-frequency variations in the data. This could be accomplished by passing the RR-interval series though a band pass filter with upper and lower bounds equal to the bounds for the high frequency variations. While some may suggest a simple alternative such as taking a sample from the entire 24 hour recording period, this method should be considered carefully. A consecutive series of RR-interval

data should be taken from each subject, while under the same circumstances, for example from the same time of day.

Another issue arising with long recording periods is the amount of data. In the loneliness data, a total of 32 minutes were examined resulting 19828 RR-intervals (not identified as artifacts). As the recording length increases, computations performed within the EM-algorithm would become lengthy and convergence time will increase.

The second motivational data set came from a study also conducted by the School of Nursing at Virginia Commonwealth University (Pickler et al., 2006, Pickler et al., 2006, Pickler et al., 2005). In contrast to the first study described above, this study examined a random sample of preterm infants whose ECG signals were recorded during several bottle feeding sessions. These recordings typically lasted about 15 minutes and included ECG data from before, during, and after the bottle feeding sessions. While the length of the recording does introduce the lower frequency variations in the RR-intervals discussed previously other issues arose with the particular data. The nature of the bimodal distribution within the high-frequency ranges is still subjective. Many have attributed the bimodality to respiration. Eckberg (1983) observed that the bimodal pattern became more obvious as the controlled breathing rate in healthy young subjects decreased.

While there were issues encountered with both data sets, there is still interest in modeling these data with the methodology proposed in this research. For the adult data, the subject's breathing patterns were also recorded during the 24 hour recording periods. By correlating the breathing rates with the heart rates, there is potential to gain much

information regarding the plausible bimodal nature of the data. As previously discussed, inspiration has been attributed with an increase in heart rate and expiration a decrease in heart rate. The methodology proposed in the research may be extended to estimate the group membership with the estimates of the incomplete data. The data from the breathing signal would provide information regarding the actual status of the RR-interval (inspiration/expiration). A comparison between the estimated group membership and the actual group membership could indicate the accuracy of the relationship between breathing and bimodal nature of the high-frequency variations in the RR-intervals.

5 Conclusions and Future Research

In this dissertation a method for fitting a random-effects model under the assumption of a normal-mixture distribution is described. The need for such a model was motivated though data sets that typically exhibit bimodality. An example of such data is heart rate variability (HRV) data, particularly series of RR-intervals. In Chapter 1 the existing methodology for such HRV data were discussed in detail. The time and frequency domain methods discussed in section 1.3 is the current paradigm for summarizing HRV data. These methods are limited, providing only descriptive measures regarding the variability in the heart rate (such as the average RR-interval length, the standard deviation of all RR-intervals over the entire recording period, or the frequencies at which the heart rate varies). To further understand HRV, several researchers have explored modeling the variations in the heart rate using the RR-interval length. A brief review of the progress in the modeling aspect of the data was provided in section 1.6. One of the most important aspects seems to be that the distributions of RR-intervals indicate a normal-mixture density. Limitations of the current methodology for modeling RR-interval data under a normal-mixture density assumption were discussed. To address some of the limitations, in this dissertation, a random-effects model under a normal-mixture assumption was proposed. By using a random-effects model and incorporating a random (subject) effect, it was possible to incorporate within subject correlations across the time series. To obtain maximum likelihood estimated of the parameters in the model an EM-algorithm was suggested. In this EM-algorithm, by defining the missing data judiciously, existing software such as the PROC MIXED in SAS could be implemented.

In Chapter 3 a simulation study was performed in order to explore the properties of the estimates though bias and MSE. This study showed that the mixture proportion was estimated with the best accuracy, while the parameter estimate for the between-subject variance was estimated with the worst precision. The study further showed that increases in the number of subjects dramatically improve accuracy by reducing the bias and the MSE. The convergence of the EM-algorithm was found to be insensitive to the initial values. A SAS macro was written so that implementation of the method for other data sets is feasible with minimal adjustments.

5.1. Variance-Covariance Structure for the Time Series

The likelihood for the fixed-effects parameters was written as a product of normal-mixtures under the assumption of independence among the observations within subjects. Although the inclusion of a random subject effect in the model imposes a compound symmetric correlation structure for the observations within subjects across the time points (RR-intervals), this may not be adequate. A more complex time series modeling may be warranted in some data sets. To accommodate for this, the first approach may be to treat the residuals as multivariate normal-mixtures. However, this is likely to lead to complications in computations and the gain may not be significant enough to warrant this. For example, with HRV data there is an abundance of RR-interval data taken on each subject in just a few minutes of ECG recording. In particular, the Loneliness data analyzed in Chapter 4 consisted of a total of 19828 observations. In this case, a multivariate definition of the likelihood would then require computations using matrices

with dimensions as large 19828, thus impacting not only the computational time, but also introducing issues of computer memory. The second approach could be to remove the correlation structure from the data before fitting the model. This could be accomplished through some sort of white noise filtering process or what is known as pre-whitening (Box and Jenkins). Further examinations of both of these extensions are being considered and will be presented in future research.

5.2. Testing the Equality of the Component Variances

In Chapter 4 the parameter estimates of component variances were reported with standard errors and 95% confidence intervals. A test to determine if the parameter estimates were significantly different was conducted by comparing the two confidence intervals. In this case, the two confidence intervals did not overlap, thus indicating the parameter estimates to be significantly different. While this type of test might be useful, an alternative might be considered. A natural test of the equality of the component variances would be to conduct a likelihood ratio test. Define the full model as the model containing two component variance parameters, and the reduced model as the model containing a single variance parameter, equal for both components. Then a LRT comparing the difference in the -2 log likelihood values between the full and reduced model to a chi-squared with 1 degree of freedom would test the equality in the component variance parameters. Large difference would then indicate that the model with separate component variance parameters fits significantly better than the model with a single component variance parameter.

With the implementation of the EM-algorithm as described in this research this alternative is not possible. Here, the parameter estimates for the component variances are estimated independently with two separate random-effects models on observations computed through weighted observations. While this provides simplicity in the computations of the estimates, the ability to constrain the estimates across the two components (two separate models) is not possible. Further examination of this problem will be considered in subsequent research.

5.3. Estimating the Variance of the Parameter Estimates

In the analysis of the Loneliness data in Chapter 4 standard errors for some of the parameter estimates were reported. The methods used to obtain them and suggestions for alternative methods are presented here. The standard errors for the component fixed-effects parameters and the component variance parameters were determined by the fits of the two separate random-effects models. In SAS, the standard errors for the covariance parameters are determined as the square roots of the diagonal elements of the inverse of the observed Fisher information matrix. The Fisher information matrix is equal to twice the negative of the Hessian matrix, which consists of the second derivatives of the log of the likelihood with respect to the covariance parameters. Using the EM-algorithm presented here, the standard errors produced by the fits of the separate random-effects models are determined using likelihoods, which are not dependent on the mixture proportion parameter, λ. In order to obtain the appropriate standard errors for the parameter estimates, the overall likelihood should be used to determine the Fisher

information matrix. While computationally intense, these are important issues and will be provided in future work.

With regards to the suggested estimate for the between subject variance parameter, σ_δ^2, an estimate for the standard error was provided in the results of the analysis in Chapter 4. For the sample of $S = 9$ subjects, this was determined by using the fact that the ratio of $(S-1)$ times sample variance to the population variance is approximately distributed as chi-squared with $(S-1)$ degrees of freedom. The formula then used to compute the standard error for σ_δ^2 was

$$SE\left(\hat{\sigma}_\delta^2\right) = \sqrt{\frac{2\left(\hat{\sigma}_\delta^2\right)^2 (S-1)}{S^2}}.$$

The appropriateness of this estimator needs to be further explored.

Estimates for the standard errors of the estimates for the single mixture proportions, $\hat{\lambda}$, and the subject varying mixture proportions $\hat{\lambda}_j$, were also provided in the results of the analysis in Chapter 4. When the maximum likelihood estimates for parameters are difficult to obtain, a common method for estimation is the method of moments. For the model with subject varying mixture proportions, the estimate for the mixture proportions from equation (2.110) is given by

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{N_j} \pi_{ij}}{N_j}.$$

Here, the π_{ij} are the posterior probabilities of the i^{th} observation from the j^{th} subject falling into the first component shown in equation (2.97) and given by

$$\pi_{ij} = \frac{\lambda_j \phi\left(\varepsilon_{ij}; \mu_{ij1}, \sigma_1^2\right)}{\lambda_j \phi\left(\varepsilon_{ij}; \mu_{ij1}, \sigma_1^2\right) + \left(1-\lambda_j\right)\phi\left(\varepsilon_{ij}; \mu_{ij2}, \sigma_2^2\right)}.$$

The method of moments is demonstrated here for the model with varying mixture proportions. For a random sample, π_{1j}, π_{2j}, ..., $\pi_{N_1 j}$, the expected value of the first sample moment is given by

$$\bar{\lambda}_j = E\left[\frac{1}{N_j}\sum_{i=1}^{N_j}\pi_{ij}\right]$$

and is equivalent to the maximum likelihood estimate for the mixture proportion, $\hat{\lambda}_1$.

The variance of the first sample moment then is given by

$$\begin{aligned}\hat{\sigma}^2_{\lambda_j} &= \operatorname{var}\left(\frac{1}{N_j}\sum_{i=1}^{N_j}\pi_{ij}\right) = \left(\frac{1}{N_j}\right)^2 \sum_{i=1}^{N_j}\operatorname{var}\left(\pi_{ij}\right) \\ &= \left(\frac{1}{N_j}\right)^2 \sum_{i=1}^{N_j}\left[E\left(\pi_{ij}^2\right) - E\left(\pi_{ij}\right)^2\right] \\ &= \frac{1}{N_j}\sigma^2_{\lambda_j}.\end{aligned}$$

Here, $\sigma^2_{\lambda_1}$ is determined by

$$\sigma^2_{\lambda_1} = \frac{1}{N_1}\sum_{i=1}^{N_1}\left(\pi_{i1} - \bar{\lambda}_1\right)^2.$$

Thus, the method of moments estimate for the variance of the subject mixture proportion reduces to

$$\hat{\sigma}^2_{\lambda_1} = \frac{1}{N_1}\sigma^2_{\lambda_1} = \left(\frac{1}{N_1}\right)^2 \sum_{i=1}^{N_1} \left(\pi_{i1} - \bar{\lambda}_1\right)^2.$$

5.4. Goodness of Fit Tests

The work in this dissertation did not provide methodology for testing the number of model components. Most of the model testing in this research consisted of likelihood ratio tests. A condition of the LRT, however, is for the reduced model to be nested in the full model, essentially making the reduced model a special case of the full model. There is a natural desire to question if a model with a two-component normal-mixture density assumed for the residuals fits significantly better than a model with a normal density. The single-component model however is not a special case of the two-component mixture model thus LRTs are not applicable. Instead a goodness of fit test should be designed. Without a test of this sort, careful consideration should be given to the number of components chosen for the model, taking biological research into account.

5.5. Unconditional Likelihood Ratio Tests

The issues with the likelihood ratio tests involving the conditional likelihoods rather than the unconditional likelihoods were discussed in section 2.9. The advantage of the EM-algorithm is that it does not require specification of the full likelihood. The full likelihood for all the model parameters, given the vector of observations, would involve

indefinite integrals. This computationally intense process may not provide much gain in terms for the hypothesis tests, but this needs to be established.

5.6. Newton-Raphson Algorithm

An alternative to the EM-algorithm would be to use the Newton-Raphson algorithm. The Newton-Raphson algorithm obtains the parameter estimates by iteratively solving the maximum likelihood equations. The algorithm is unattractive in that it involves computations of first and second derivatives of the likelihood involving matrices. The EM-algorithm was thus chosen for its simplicity. The computation of variances for the parameter estimates discussed in section 5.3 however requires the same sorts of computation and would be a natural product of the Newton-Raphson algorithm. The Newton-Raphson may also be slow to converge and may not always converge to global maxima. Furthermore, the Newton-Raphson may be more sensitive to initial values. Research involving fitting the normal-mixture random-effects density with the Newton-Raphson will be explored in future work in order to lend comparison to the weighted EM-algorithm presented in this research and to provide alternatives to some of the limitations resulting from the EM-algorithm.

5.7. Other Limitations

Although modeling the mean of the normal-mixture density, which essentially models the link function in a generalized linear model (GLM), is reasonable, in some situations it may be useful to model each component mean. This would require different fixed-effects

design matrices. Such an extension can easily be made to the model proposed in this research. However, testing and model selection issues need to be further explored.

6 List of References

List of References

Akselrod, S., Gordon, D., Ubel, F. A., Shannon, N. C., Barger A. C., and Cohen, R. J. (1981). Power spectrum analysis of heart rate fluctuations: a quantitative probe beat-to-beat cardiovascular control. *Science* 213, 210-222.

Belin, T. R., and Rubin, D. B. (1995). The analysis of repeated-measures data on schizophrenic reaction times using mixture models. *Statistics in Medicine* 30; 14(8), 747 – 768.

Berntson, G. G., Quigley, K. S., Jang J. F., and Boysen, S. T. (1990). An approach to artifact identification: application to heart period data. *Psychophysiology* 27, 586 – 598.

Berntson, G. G., Bigger, J. T., Eckberg, D. L., Grossman, P., Kaufman, P. G., Malik, M., Nagaraja, H. N., Porges, S. W., Saul, J. P., Stone, P. H., and van Der Molen, M. W. (1997). Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology* 34, 623 – 648.

Berntson, G. G., and Stowell, J. R. (1998). ECG artifacts and heart period variability: don't miss a beat. *Psychophysiology* 35(1), 127 – 132.

Brown, H. and Prescot, R. (1999). *Applied Mixed Models in Medicine*. John Wiley & Sons Ltd.

Cacioppo, J. T., Hawkley, L .C. Crawford, L. E., Ernst, J. M., Burleson, J. M., Kowalewski, R. B., Malarakey, W. B., Van Cauter, E. and Bernston, G. G. (2002). Loneliness and health: Potential mechanisms. *Psychosomatic Medicine* 64: 407 – 417.

Daniel, C. (1959). Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics* 1(4) 311 – 341.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, B, 39, 1 – 38.

Diggle, P. Liang, K., and Zeger, S. (1994). *Analysis of Longitudinal Data*. Oxford Science Publications.

Duprez, D., De-Buyzere, M., Rietzschel, E., Rimbout, S., Kaufman, J. M., Van Hoecke, M. J., and Clement, D. L. (1995) Renin-angiotensin-aldosterone system, RR-interval and blood pressure variability during postural changes after myocardial infarction. *European Heart Journal* 16(8) 1050 – 1056.

Eckberg, D. L. (1983). Human sinus arrhythmia as an index of vagal cardiac outflow. *Journal of Applied Physiology* 54, 961 – 966.

Fallen, E. L., Kamath, M. V., and Ghista, D. N. (1988). Power spectrum of heart rate variability: a non-invasive test of integrated neurocardiac function. *Clinical and Investigative Medicine* 11(5), 331 - 340.

Fleisher, L. A., Frank, S. M., Sessler, D. I., Cheng, C., Matsukawa, T., and Vannier C.A. (1996). Thermoregulation and heart rate variability. *Clinical Science* 90(2) 97 – 103.

Francis, D. P., Davies, L. C., Willson, K., Ponikowski, P. Coats, A. J. S., and Piepoli, M. (2000). Very-low-frequency oscillations in heart rate and blood pressure in periodic breathing: role of the cardiovascular limb of the hypoxic chemoreflex. *Clinical Science* 99, 125 – 132.

Friesen, G. M., Jannett, T. C., Jadallah, M. A., Yates, S. L., Quint, S. R., and Nagle, H. T. (1990). A comparison of the noise sensitivity of nine QRS detection algorithms. IEEE Transactions on Biomedical Engineering 37(1), 85 – 98.

Graham, F. K. and Jackson, J. C. (1970). Arousal systems and infant heart rate responses. Advances in Child Development and Behavior 5, 59 – 117.

Grap, M. J., Wetzel, P. A., Best, A. M. (2006). Personal communications.

Gyaw, T. A. and Ray, S. R. (1994). The wavelet transform as a tool for recognition of biosignals. *Biomedical Sciences Instrumentation* 30, 63 – 68.

Hashida E., Inoue, T., Yoshitani, N., and Tasaki, T. (1973). A probabilistic study of the irregularity of R-R interval durations in atrial fibrillation – a preliminary report. *Japanese Circulation Journal* 37, 1423 – 1426.

Hasselblad, V. (1966). Estimation of parameters for a mixture of normal distributions. Technometrics 8(3), 431 – 444.

Horan, L. G. and Kistler, J. C. (1961). Study of ventricular response in atrial fibrillation. *Circulation Research* 9, 305 – 311.

Jenkins, D. and Gerred, S. 1996, ECG Library, viewed February 2006, < http://www.ecglibrary.com/norm.html>.

Jennings, R. J., Stringfellow, J. C., and Graham, M. (1974). A comparison of the statistical distribution of beat-by-beat heart rate and heart period. *Psychophysiology* 11, 207-210.

Jordan, H. (1954). Die zeitliched schwankungen der herzschlagintervalle bei arrhythmie. *Basic Research in Cardiology* 21, 40 – 49.

Kamath, M. V., Ghista D. N., Fallen, E. L. (1987). Heart rate variability power spectrogram as a potential noninvasive signature of cardiac regulatory system response, mechanisms, and disorders. *Hearts and Vessels* 3(1), 33-41.

Khachaturian, Z. S., Kerr, J., Kruger, R. and Schachter, J. (1972). A methodological note: Comparisons between period and rate data in studies of cardiac function. *Psychophysiology* 9, 539-545.

Kleiger R.E., Miller J.P., Bigger J.T. Jr, Moss A.J., and the Multicenter Post-Infarction Research Group (1987). Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *American Journal of Cardiology* 59:256-262

Kleiger, R. E., Bigger, J. T., Bosner, M. S., Chung, J. R., Rolnitzky, L. M., Steinman, R., and Fleiss, J. L. (1991). Stability over time of variables measuring heart rate variability in normal subjects. *American Journal of Cardiology* 68, 626-630.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* 38, 963-974

Litvack, D. A, Oberlander, T. F, Carney L. H, and Saul J. P. (1995). Time and frequency domain methods for heart rate variability analysis: A methodological comparison. *Psychophysiology* 32, 492 – 504.

Malik, M., Farrell, T., Cripps, T., and Camm, A. J., (1989). Prognostic value of heart rate variability after myocardial infarction. A comparison of different data-processing methods. *Medical and Biological Engineering and Computing* 27, 603 – 611.

Malik, M. and Camm, A. J. (1993). Components of heart rate variability – what they really mean and what we really measure. The American Journal of Cardiology 72(11), 821 – 822.

Malpas S. C. and Maling T. J. B. (1990). Heart rate variability and cardiac autonomic function in diabetes. *Diabetes* 29, 1177-1181.

Mandrekar, S. J., Nagaraja, H. N., and Berntson, G. G. (2005). Statistical modeling of the differences between successive R-R intervals. *Statistics in Medicine* 24, 437-451.

Mandrekar, S. J. (2002). Modeling of heart period data and a study of the associated spectral measures. Dissertation. The Ohio State University.

May, C., Hubing, N., and Hahn, A. W. (1997). Wavelet transforms for electrocardiogram processing. *Biomedical Sciences Instrumentation* 33, 1 – 6.

Merri, M. Farden, D. C., Mottley, J. G., and Titlebaum, E.L. (1990). Sampling frequency of the electrocardiogram for spectral analysis of the heart rate variability. *IEEE Transactions on Biomedical Engineering* 37(1) 99 – 106.

Myers, R. H., and Montgomery, D. C. (1995). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. John Wiley and Sons, Inc. pp 104-109, 260-264.

Nagaraja, H. N., Sampath, S., and Binkley, P. F. (1995). Statistical modeling of heart period data. *American Statistical Association* 1995 *Proceedings* (Biometric Section) 361 – 364.

Odemuyiwa O., Malik M., Farrell T., Bashir Y., Poloniecki J., Camm J. (1991). Comparison of the predictive characteristics of heart rate variability index and left ventricular ejection fraction for all-cause mortality, arrhythmic events and sudden death after acute myocardial infarction. *American Journal of Cardiology* 68:434-439.

Öri, Z., Monir, G. M., Weiss, J., Sayhouni, X., and Singer, D. H. (1992). Heart Rate Variability: Frequency Domain Analysis. *Ambulatory Electrocardiography* 10, (3), 499-537.

Pagani, M., Lombardi F., Guzzetti, S., Rimoldi, O., Furlan, R., Pizzinelli, P., Sandrone, G., Malfatto, G., Dell'Orto, S., and Piccaluga M. (1986) Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympathovagal innervation in man and conscious dog. *Circulation Research* 59, 178-193

Penaz, J. (1978). Mayer waves: History and methodology. *Automedica* 2, 135-141.

Pichot, V., Gaspoz, J. M., Molliex, S., Antoniadis, A., Busso, T., Roche, F., Costes, F., Quintin, L., Lacour, J. R., and Barthelemy J. C. (1999). Wavelet transform to

quantify heart rate variability and to assess its instantaneous changes. *Journal of Applied Physiology* 86(3) 1081 – 1091.

Pickler, R. H., Best, A. M., Reyna, B. A., Wetzel, P. A., and Gutcher, G. R. (2006). Changes in nutritive sucking in preterm infants. *Journal of Perinatalogy* 26, 693-699.

Pickler, R. H., Chiaranai, C., and Reyna, B. A. (2006). Relationship of the first suck burst to feeding outcomes in preterm infants. *Journal of Perinatal and Neonatal Nursing* 20, 155-160.

Pickler, R. H., Best, A. M., Reyna, B. A., Wetzel, P. A., and Gutcher, G. R. (2005). Prediction of feeding performance in preterm infants. *Newborn and Infant Nursing Reviews* 5, 116-123.

Pomeranz, B., Macaulay, R. J., Caudill, M. A., Kutz, I., Adam, D., Gordon, D., Kilborn, K. M., Barger, A. C, Shannon, D. C., Cohen, R. J., and Benson H. (1985). Assessment of autonomic function in humans by heart rate spectral analysis. *American Journal of Physiology: Heart and Circulatory Physiology* 272(1) H151 – H153

Ponikowski, P., Chua, T. P., Piepoli, M., Amadi, A. A., Harrington, D., Webb-Peploe, K., Volterrani, M., Colombo, R., Mazzuero, G., Giordano, A., and Coats, A. J. (1997). Chemoreceptor dependence of very low frequency rhythms in advanced chronic heart failure. *American Journal of Physiology: Heart and Circulatory Physiology* 272(1) H438 – 447.

Pumprla, J., Howorka, K., Groves, D., Chester, M., and Nolan, J. (2002). Functional assessment of heart rate variability: physiological basis and practical applications. *International Journal of Cardiology* 84, 1-14.

Ramanathan, A. and Myers, G. A. (1996). Data preprocessing in spectral analysis of heart rate variability. Handling trends, ectopy, and electrical noise. *Journal of Electrocardiology* 29(1), 45 – 47.

Riniolo, T. and Porges, S. W. (1997). Inferential and descriptive influences on measures of respiratory sinus arrhythmia: sampling rate, R-wave trigger accuracy, and variance estimates. *Psychophysiology* 34(5), 613 – 621.

Riniolo, T. and Porges, S. W. (2000). Evaluating group distributional characteristics: why psychophysiologists should be interested in qualitative departures from the normal distribution. *Psychophysiology* 37(5), 21 – 28.

Ribbert, L. S., Snijders, R. J., Nicolaides, K. H., and Visser G. H. (1991). Relation of fetal blood gasses and data from computer-assisted analysis of fetal heart rate patterns in small for gestation fetuses. *British Journal of Obstetrics and Gynaecology* 98(8), 820-823.

Sayers, B. McA., (1973) Analysis of heart rate variability. *Ergonomics* 16(1), 17 – 32.

Smith, J. H., Anand, K. J., Cotes, P. M., Dawes, G. S., Harkness, R. A., Howlett, T. A., Rees, L. H., and Redman, C. W. (1988). Antenatal fetal heart rate variation in relation to the respiratory and metabolic status of the comprised human fetus. *British Journal of Obstetrics and Gynaecology* 95(10), 980-989.

Stein, P. K., Bosner, M. S., Kleiger, R. E., and Cogner, B. M. (1994). Heart rate variability: a measure of cardiac autonomic tone. *American Heart Journal* 127, 1376-1381.

Task Force of the European Society for Cardiology and the North American Society of Pacing and Electrophysiology (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17, 354-381.

Taylor, J. A., Carr, D. L., Myers, C. W., and Eckberg, D. L. (1998). Mechanisms underlying very-low-frequency RR-interval oscillations in humans. *Circulation* 98, 547 – 555.

Thaler, I., Timor-Tritsch, I., and Blumenfeld, Z. (1985). The significance of increased heart rate variability. *Acta Obstetricia et Gynecologica Scandinavica* 64(1), 47-50.

van Ravenswaaij-Arts, C. M. A., Kollee, L. A. A., Hopman, J. C. W., Stoelinga, G. B. A., and van Geijn, H. P. (1993). Review: Heart Rate Variability. *Annals of Internal Medicine* 118(16), 436-447.

Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. Journal of the American Statistical Association 91, 217 – 221.

Verbeke, G., Speissens, B., and Lesaffre, E. (2001). Conditional Linear Mixed Models. *The American Statistician* 55, 25-34.

Weber E. J. M, Weber, P. C. M, van Der Molen, M. W. (1992) A nonstationarity test for the spectral analysis of physiological time series with an application to respiratory sinus arrhythmia. Psychophysiology, 29(1), 55 – 65.

Xia, R., Odemuyiwa, O. Gill, J., Malik, M., and Camm, A. J. (1993). Influence of recognition errors of computerized analysis of 24-hour electrocardiograms on the measurement of spectral components of heart rate variability. International Journal of Bio-Medical Computing 32(3 – 4), 223 – 235.

Xu, W. (1995). Mixtures in random-effects regression models; Un-published dissertation.

Xu, W. and Hedeker, D. (2001). A random-effects mixture model for classifying treatments in longitudinal clinical trials. *Journal of Biopharmaceutical Statistics*, 11(4), 253 – 273.

7 Appendices

7.1. Derivation of MLE Solutions for the M-component Normal-Mixture Model

Notation:

ε_i = the observation indexed by i

M = the number of mixture components

N = the number of observations

π_{ik} = the posterior probability of the i^{th} observation falling into the k^{th} group

λ_k = the mixture proportion for the k^{th} group

μ_k = the k^{th} group mean

σ_k^2 = the k^{th} group variance

Definitions:

The M-component normal-mixture model density:

$$f\left(\varepsilon_i;\mu_1,...,\mu_M,\sigma_1^2,...,\sigma_M^2,\lambda_1,...,\lambda_{M-1}\right)=\sum_{k=1}^{M}\lambda_k\phi\left(\varepsilon_i;\mu_k,\sigma_k^2\right).$$

The likelihood for the parameters of the model, given the N observations:

$$L\left(\boldsymbol{\varepsilon};\mu_1,...,\mu_M,\sigma_1^2,...,\sigma_M^2,\lambda_1.,,,.\lambda_{M-1}\right)=\prod_{i=1}^{N}\sum_{k=1}^{M}\lambda_k\phi\left(\varepsilon_i;\mu_k,\sigma_k^2\right).$$

The log-likelihood, under the restriction that $\lambda_M=1-\sum_{k=1}^{M-1}\lambda_k$:

$$l\left(\varepsilon;\lambda,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right)=\sum_{i=1}^{N}\log\left[\sum_{k=1}^{M}\lambda_k\phi\left(\varepsilon_i;\mu_k,\sigma_k^2\right)\right].$$

The posterior probability estimate, π_{ik}, is found by taking a weighted average of the M component densities. The formula for the π_{ik} is

$$\pi_{ik}=\frac{\lambda_k\phi\left(\varepsilon_i;\mu_k,\sigma_k^2\right)}{\sum_{m=1}^{M}\lambda_k\phi\left(\varepsilon_i;\mu_k,\sigma_k^2\right)}.$$

Derivations:

The MLEs are calculated in the usual manner, by taking the first derivative of the log-likelihood, setting them equal to zero, and solving. The derivations are shown below.

The MLE of λ_k is found be taking the first derivative of l, with respect to λ_k, setting it equal to zero, and solving the resulting expression for λ_k under the constraint

$$\lambda_M=1-\sum_{k=1}^{M-1}\lambda_k.$$

The first derivative of l, with respect to λ_k, set equal to zero is

$$\frac{\partial l}{\partial\lambda_k}=\sum_{i=1}^{N}\left[\frac{\phi\left(\varepsilon_i;\mu_k,\sigma_k^2\right)-\phi\left(\varepsilon_i;\mu_M,\sigma_M^2\right)}{\sum_{m=1}^{M}\lambda_m\phi\left(\varepsilon_i;\mu_m,\sigma_m^2\right)}\right]=0.$$

Substituting for π_{ik} yields:

$$\sum_{i=1}^{N}\left[\frac{\pi_{ik}}{\lambda_k}-\frac{\pi_{iM}}{\lambda_M}\right]=\sum_{i=1}^{N}\left[\pi_{ik}\left(1-\sum_{k=1}^{M-1}\lambda_k\right)-\pi_{iM}\lambda_k\right]=0.$$

Using a • to indicate summation over that index, yields:

$$\pi_{\bullet k}\left(1-\sum_{k=1}^{M-1}\lambda_k\right)-\pi_{\bullet M}\lambda_k=0.$$

Summing over the M – 1 equations and solving gives

$$\sum_{r=1}^{M-1}\left[\pi_{\bullet r}\left(1-\sum_{k=1}^{M-1}\lambda_k\right)-\pi_{\bullet M}\lambda_r\right]=\sum_{r=1}^{M-1}\left[\pi_{\bullet r}\left(1-\sum_{k=1}^{M-1}\lambda_k\right)\right]-\sum_{r=1}^{M-1}\pi_{\bullet M}\lambda_r=0$$

$$\Leftrightarrow\left(1-\sum_{k=1}^{M-1}\lambda_k\right)\sum_{r=1}^{M-1}\pi_{\bullet r}=\pi_{\bullet M}\sum_{r=1}^{M-1}\lambda_r$$

$$\Leftrightarrow\left(1-\sum_{k=1}^{M-1}\lambda_k\right)\sum_{r=1}^{M-1}\pi_{\bullet r}+\pi_{\bullet M}\left(1-\sum_{k=1}^{M-1}\lambda_k\right)=\pi_{\bullet M}\sum_{r=1}^{M-1}\lambda_r+\pi_{\bullet M}\left(1-\sum_{r=1}^{M-1}\lambda_r\right)$$

$$\Leftrightarrow\left(1-\sum_{k=1}^{M-1}\lambda_k\right)\sum_{r=1}^{M}\pi_{\bullet r}=\pi_{\bullet M}$$

$$\Leftrightarrow\lambda_M=\frac{\pi_{\bullet M}}{N}$$

$$\Leftrightarrow 1-\lambda_M=\frac{\sum_{k=1}^{M}\pi_{\bullet k}}{N}.$$

This is equivalent to

$$\lambda_k=\frac{\pi_{\bullet k}}{N}=\frac{\sum_{i=1}^{N}\pi_{ik}}{N}.$$

The MLE of μ_k is found be taking the first derivative of l, with respect to μ_k, setting it equal to zero, and solving the resulting expression for μ_k.

$$\frac{\partial l}{\partial \mu_k} = \sum_{i=1}^{N} \left[\frac{\lambda_k \phi\left(\varepsilon_i; \mu_k, \sigma_k^2\right)\left(\frac{\varepsilon_i - \mu_k}{\sigma_k^2}\right)}{\sum_{m=1}^{M} \lambda_m \phi\left(\varepsilon_i; \mu_m, \sigma_m^2\right)} \right] = 0.$$

Substituting for π_{ik} and solving yields:

$$\sum_{i=1}^{N} \frac{\pi_{ik}}{\lambda_k}\left(\frac{\varepsilon_i - \mu_k}{\sigma_k^2}\right) = 0$$

$$\Leftrightarrow \sum_{i=1}^{N} \pi_{ik}\varepsilon_i = \sum_{i=1}^{N} \pi_{ik}\mu_k.$$

Therefore,

$$\mu_k = \frac{\sum_{i=1}^{N} \varepsilon_i \pi_{ik}}{\sum_{i=1}^{N} \pi_{ik}}.$$

The MLE of σ_k^2 is found be taking the first derivative of l, with respect to σ_k^2, setting it equal to zero, and solving the resulting expression for σ_k^2.

$$\frac{\partial l}{\partial \sigma_k^2} = \sum_{i=1}^{M} \left[\frac{\lambda_k \phi\left(\varepsilon_i; \mu_k, \sigma_k^2\right) \left(\frac{(\varepsilon_i - \mu_k)^2}{\sigma_k^2} - 1 \right)}{\sum_{m=1}^{M} \lambda_m \phi\left(\varepsilon_i; \mu_m, \sigma_m^2\right)} \right] = 0.$$

Substituting for π_{ik} and solving yields:

$$\sum_{i=1}^{N} \frac{\pi_{ik}}{\lambda_k} \left(\frac{(\varepsilon_i - \mu_k)^2}{\sigma_k^2} - 1 \right) = \sum_{i=1}^{N} \frac{\pi_{ik} (\varepsilon_i - \mu_k)^2 - \pi_{ik} \sigma_k^2}{\lambda_k \sigma_k^2} = 0$$

$$\Leftrightarrow \sum_{i=1}^{N} \pi_{ik} \sigma_k^2 = \sum_{i=1}^{N} \pi_{ik} (\varepsilon_i - \mu_k)^2.$$

Therefore,

$$\sigma_k^2 = \frac{\sum_{i=1}^{N} \pi_{ik} (\varepsilon_i - \mu_k)^2}{\sum_{i=1}^{N} \pi_{ik}}.$$

7.2. Details for Data Simulation

The process of simulating the data for the two-component normal-mixture random-effects model is implemented by using a macro written in SAS v.9.0. This macro is shown below.

Definition of macro variables:

Simulate –name of the macro

nsim – total number of simulations

nsub – total number of subjects

nobs – total number of observations per subject

mu1 – first component mean

mu2 – second component mean

sig1 – first component variance

sig2 – second component variance

sigdelta – variance of the random subject effects

lambda – mixture proportion

s – simulation number

Code for Simulation macro:

```
%macro Simulate(nsim,nsub,nobs,mu1,mu2,sig1,sig2,sigdelta,lambda);
%do s = 1 %to &nsim; *Simulation do loop;
   proc iml;
      bb = j(&nsub,1,196*&s); *Creates nsub*1 vector containing the sim seed;
      b = sqrt(&sigdelta)#normal(bb); Creates nsub*1 vector containing N(0,sigdelta):
      do j = 1 to &nsub;        *subject do loop:
         esub = j(&nobs, 1, &
         ee1 = j(&nobs,1, 728*j*&s); *Creates nobs*1 vector containing 1st sub seed:
         ee2 = j(&nobs,1, 274*j*&s); *Creates nobs*1 vector containing  2nd sub seed;
```

```
        uu = j(&nobs,1, 632*j*&s); *Creates nobs*1 vector containing 3rd sub
seed;
        e1 = j(&nobs,1,&mu1)+sqrt(&sig1)#normal(ee1); * nobs*1 vec of N(mu1,sig1);
        e2 = j(&nobs,1,&mu2)+sqrt(&sig2)#normal(ee2); * nobs*1 vec of N(mu2,sig2);
        u = uniform(uu); * nobs*1 vector of uniform random values:;
        obsnumj = j(&nobs, 1, 0);
        do i = 1 to &nobs;
           obsnumj[i] = i; *Assigning order to observations;
           if u[i] < &lambda then esub[i] = e1[i]; *The residual is mix of the 2 normals;
           else esub[i] = e2[i];
        end;  *ends i (RR interval) loop;
        bsub = b[j]; *print bsub;
        ysub = esub + j(&nobs, 1, bsub); *The observation is residual + random effect;
        sub_id = j(&nobs, 1, j);
        sim_id = j(&nobs, 1, &s);
        if j = 1 then do;
           y = ysub;
           subid = sub_id;
           simid = sim_id;
           obsnum = obsnumj;
        end;
        else do;
           y = y//ysub; *Creates the vector of observations;
           subid = subid//sub_id; *Creates the vector of subject identifiers;
           simid = simid//sim_id; *Creates the vector of simulation identifiers:
           obsnum = obsnum//obsnumj; *Creates the vector of observation identifiers;
        end;
     end;     *ends j (subject) loop;
     Asim = y || subid || simid ||obsnum; *Creates the simulation data matrix;
     names = {HRI sub sim obsnum}; *Creates the column names for data set;
     create Y_&s from Asim [colname = names];   *Creates simulation data set;
     append from Asim;
     quit;     *quits the iml proc;
  %end;     *ends s (sim) loop;
%mend Simulate;
```

The macro above can be used to simulate *nsim* data sets each having *nsub* subjects with *nobs* observations drawn on each subject. The first and second component means for the residuals are specified with *mu1* and *mu2*, respectively. The first and second component variances for the residuals are specified with *sig1* and *sig2*, respectively. The variance of

the random subject effects (between subject variance) is specified with *sigdelta* and the mixture proportion for the residuals is given by *lambda*. The following statement is used in SAS to call the *Simulate* macro for 100 simulations, 20 subjects, 200 observations per subject, component means 70 and 86.44, component variance 25 and 25, between subject variance 46.28, and mixture proportion 0.5:

%*Simulate* (100, 20, 200, 70, 86.44, 25, 25, 46.28, .5);

The following describes the details of the *Simulate* macro. The macro starts with a do-loop, called the simulation do-loop, iterating from 1 to *nsim*. This enables *nsim* data sets to be simulated independently with the same model parameters. Next, the IML procedure is executed so that all calculations may be performed with vectors and matrices. The following describes the IML code used to create each of the *nsim* simulated data sets.

First, an *nsub* × 1 seed vector *bb* is created containing a seed. The elements of the vector *bb* are all equal to the same seed. (Please note that this seed must be multiplied by the simulation number so that each iteration of the simulation do-loop creates a unique seed). Next, an *nsub* × 1 vector *b* is created. The elements of this vector are independent and randomly drawn, using the seed vector *bb*, from a normal density with mean 0 and variance *sigdelta*. This vector, *b*, is a then a vector containing the *nsub* random subject effects. After determining the vector of random subject effects, a second do-loop, called the subject do-loop, is executed iterating from 1 to *nsub*. This do-loop, which iterates though the subjects, is used to determine the residuals for each subject. The following

describes this process. Please note that this code is designed so that the same number of observations is drawn from each subject.

First, an *nobs* × 1 vector, called *esub*, is created. Initially, the elements of this vector aer zeros. Later, the values will be changed to contain the randomly drawn values of the observations and residuals for the subject. Next, three *nobs* × 1 seed vectors are created, *ee1*, *ee2*, and *uu*. Each vector contains a different seed value, however, the elements within each vector are equal. (Please note these three seeds must be multiplied by the subject number so that each iteration of the subject do-loop creates unique seeds). Next, the *ee1* and *ee2* seed vectors are used to create two *nobs* × 1 vectors, *e1* and *e2*, respectively, of randomly drawn normal values. The vector *e1* contains *nobs* residuals independently and randomly drawn from a normal density with mean *mu1* and variance *sig1*. The vector *e2* contains *nobs* residuals independently and randomly drawn from a normal density with mean *mu2* and variance *sig2*. The seed vector *uu* is then used to determine an *nobs* × 1 vector, *u*, whose values are independently and randomly drawn from a uniform density. The uniform values (elements) of *u*, and used to create a vector of two-component normal-mixture values determined from the normal vectors *e1* and *e2*. This will be done with a do-loop, called the observation do-loop, iterating from 1 to *nobs*. Before the observation do-loop, a vector, *obsnumj*, must be defined. The *nobs* × 1 vector *obsnumj* is defined as a vector of zeros. Upon completion of the observation do-loop, this vector will contain values ordering the residuals. The following describes the observation do-loop used to produce the two-component normal-mixture values for the residuals.

The observation do-loop iterates through the *nobs* observations for each subject. Now for clarity, define the following:

i = the current iteration number in the observation do-loop

$obsnumj[i]$ = the i^{th} element in the vector *obsnumj*

$u[i]$ = the i^{th} element in the vector u

$esub[i]$ = the i^{th} element in the vector *esub*

$e1[i]$ = the i^{th} number in the vector *e1*

$e2[i]$ = the i^{th} number in the vector *e2*

First, $obsnumj[i]$ is set equal to the iteration number, i. Next, if, $u[i] < lambda$, then $esub[i]$ is set equal to $e1[i]$, otherwise it is set equal to $e2[i]$. Then the observation loop is concluded. At the end of the observation do-loop, the vector *obsnumj* contains values from 1 to *nobs*, in order. This vector is used …(I am not really sure if I even need this vector…I will come back to this later). The vector *esub*, for the particular subject in the subject do-loop, contains *nobs* values independently and randomly drawn from a two-component normal-mixture density. Next the random subject effects are added to the residuals to produce the vector of observations whose elements are independently and randomly drawn from a two-component normal-mixture random-effects density. The following describes this process.

Define j as the iteration number in the subject do-loop. Then an *nobs* × 1 vector, *bsub*, is created with elements equal to the j^{th} random subject effect determined by the j^{th} value in the vector b. The *nobs* × 1 vector, *ysub*, then is determined as the vector *esub* added to the vector *bsub*. This vector, *ysub*, then contains *nobs* values independently and

randomly drawn from a two-component normal-mixture random-effects density. Next, two vectors are created used to identify the subject and simulation numbers, the data was simulated from. The vectors *sub_id* and *sim_id* are *nobs* × 1 vectors whose elements are equal to the iteration numbers for the current subject and simulation do-loops, respectively. Before concluding the subject do-loop, a set of if-then-else statements are used to concatenate each of the vectors ysub, sub_id, sim_id, and obsnumj, with these vectors from the previous iterations. The following describes these statements.

For the first iteration of the subject do-loop, the following *nobs* × 1 vectors are defined:

y = *ysub*

subid = *sub_id*

simid = *sim_id*

obsnum = *obsnumj*

These are the vectors of observations, subject identifiers, simulation identifiers, and observation identifiers for the first subject. The remaining iterations then define the following:

y = *y* stacked on top of *ysub*

subid = *subid* stacked on top of *sub_id*

simid = *simid* stacked on top of *sim_id*

obsnum = *obsnum* stacked on top of *obsnumj*

Then when the subject do-loop is concluded, this results in four (*nsub* × *nobs*) × 1 vectors, *y*, *subid*, *simid*, and *obsnumj*, containing data from all subjects.

Now, before concluding the simulation do-loop, a SAS data set is created containing the data from these four vectors. This is accomplished by first defining an (*nsub* × *nobs*) × 1 matrix, *Asim*, whose columns contain the four vectors *y*, *subid*, *simid*, and *obsnumj*. The nest statement names the four columns *HRI*, *sub*, *sim*, and *obsnum*, respectively. Then the final two statements, before the IML procedure is ended, actually create a SAS data set called Y_&s (where &s is the current iteration number of the simulation do-loop) from the matrix *Asim*. The simulation do-loop and the macro are then ended. Once the macro is run, nsim data sets, named Y_1 – Y_100, are created with the specified parameters in the *Simulate* macro call statement. This process of obtaining 100 different data sets is repeated for the 72 different parameter vectors, thus obtaining a total of 7200 simulated data sets.

7.3. Details for the EM-Algorithm for the Simulated Data

The process of simulating the data for the two-component normal-mixture random-effects model is implemented by using a macro written in SAS v.9.0. This macro is shown below.

Definition of macro variables:

EMJMX – name of the macro

nsim – total number of simulations

nsub – total number of subjects

lambdaSV – initial value for the mixture proportion

name – output data set name extension

Code for Simulation macro:

```
%macro EMJMX (nsim, nsub, lambdaSV, name);
%do sim = 1 %to &nsim;*Simulation do–loop;
   data simdata;
      set Y_&sim;
      int = 1;
   run;
/***************DETERMINE THE STARTING VALUES***************/
ods select none;
   proc univariate data = simdata;
      var HRI;
      output out = quan Q1 = q_25 Q3 = q_75;
   run;
ods select all;
   proc mixed data = simdata method= ml noclprint noinfo noitprint;
      class Sub;
      model HRI = / solution outp = resuni0 outpm = intuni0;
      random sub / type = vc solution;
      ods output solutionr = beta0;
      ods output covparms = cpuni0;
      ods select solutionf solutionr covparms;
```

```
run;
/***************E-STEP FOR THE FIRST ITERATION***************/
proc iml;
   *Read in the necessary data sets;
   use quan var{q_25 q_75};          read all var {q_25 q_75} into muSV;
   use cpuni0 var {estimate};        read all var {estimate} into varSV;
   use beta0 var {estimate};         read all var {estimate} into betaSV;
   use resuni0 var {HRI sub pred};   read all var {HRI sub pred} into resSV;
   use intuni0 var {pred};           read all var {pred} into intSV;
   *Assign the starting values and creates IML variables;
   nobs = nrow(resSV);
   nsub = max(resSV[,2]); print nobs nsub;
   lambda = &lambdaSV;
   mu1 = muSV[1,1];
   mu2 = muSV[1,2];
   sigma1 = varSV[2,1]/2;
   sigma2 = varSV[2,1]/2;
   sigdel = varSV[1,1];
   *Create the parameter vector and initial difference;
   pold = lambda || mu1 || mu2|| sigma1 || sigma2 || sigdel;
   nameP = {lambda mu1 mu2 sigma1 sigma2 sigdel};
   print "The starting values parameter vector is:";
   print nameP;
   print pold;
   create parmsold from pold [colname = nameP];
   append from pold;
   thresh = j(1,1,100000);
   create oldthresh from thresh [colname = "OLDdiff"];
   append from thresh;
   beta = betaSV[,1];
   Zbeta = resSV[,3] - intSV[,1]; *print beta Zbeta;
   *Determine the vector residuals as the the observations - their subject effects;
    y = resSV[,1]; *observations;
    e = y - Zbeta;  *residuals
    *Determine the vector containing each residuals density value under the specified
component:
    pi = 3.141592654; *pi constant:
    phi1 = (1/sqrt(2#pi#(sigma1)))#exp(-((e-j(nobs,1,mu1))##2)/(2#sigma1));
    phi2 = (1/sqrt(2#pi#(sigma2)))#exp(-((e-j(nobs,1,mu2))##2)/(2#sigma2));
    *Determine the vector containing each residuals probability of being in a
particular component:
    pi1 =(lambda#phi1)/(lambda#phi1 + (1-lambda)#phi2);
    pi2 =((1-lambda)#phi2)/(lambda#phi1 + (1-lambda)#phi2);
```

```
/***************If Lambda is different across subjects***************/
*Still need to create my pi1 and pi2 with lambdajs;
/*      *Determine each residuals weight for subjects with different lambdas;
        sumpi1 = j(nsub,1,0); sumpi2 = j(nsub,1,0);
        wt1 = j(nobs,1,0);              wt2 = j(nobs,1,0);
        do j = 1 to nsub;
                subpi1 = j(nobs,1,0);  subpi2 = j(nobs,1,0);
                do i = 1 to nobs;
                        if resSV[i,2]=j then do;        subpi1[i] = pi1[i];
        subpi2[i]=pi2[i];       end;
                end;
                sumpi1[j] = subpi1[+];          sumpi2[j] = subpi2[+];
                do i = 1 to nobs;
                        if resSV[i,2]=j then do; wt1[i]=pi1[i]/sumpi1[j];
        wt2[i]=pi2[i]/sumpi2[j];        end;
                end;
        end; */
/*********************If Lambda is equal across subjects*********************/
        pi1 =(lambda#phi1)/(lambda#phi1 + (1-lambda)#phi2);
        pi2 =((1-lambda)#phi2)/(lambda#phi1 + (1-lambda)#phi2);
        wt1 = pi1/pi1[+];       wt2 = pi2/pi2[+];
/*************************************************************************
***/

        *print sumpi1 sumpi2;
        *print phi1 phi2 pi1 pi2 wt1 wt2;
        sumwt1 = wt1[+];      sumwt2 = wt2[+]; print sumwt1 sumwt2;
        *Determine the two sets of weighted observations;
        Y1 = Zbeta + (wt1##(1/2)#e);Y2 = Zbeta + (wt2##(1/2)#e);
        wt1int = wt1##(1/2);  wt2int = wt2##(1/2);
        *print y y1 y2 wt1 wt2 wt1int wt2int;
        A = resSV[,2] || y || y1 || y2 || wt1int || wt2int || pi1 || pi2;
        namesA = {sub HRI HRI1 HRI2 wt1int wt2int pi1 pi2};
        create BB from A [colname=namesA]; append from A;
quit;
%let check = %eval(0);
%let count = 0;
*Starts the EM loop;
%do %until (&check = 1);
        %let count = %eval(&count+1);
/*********************M-STEP FOR THE FIRST ITERATION*****************/
        data compdata;          set BB;run;
        data newparmsold;       set parmsold;           run;
        data thresh_old;        set oldthresh;          run;
```

```
proc mixed data = compdata method= ml;* noclprint noinfo noitprint;
        class Sub;
        model HRI1 =wt1int/ noint solution outp = res_1 outpm = int_1;
        random sub / type = vc solution;
        ods output solutionf = mu_1; ods output solutionr = beta_1;
        ods output covparms = cp_1;
        *ods select solutionf solutionr covparms;
run;
proc mixed data = compdata method= ml;* noclprint noinfo noitprint;
        class Sub;
        model HRI2 =wt2int/ noint solution outp = res_2 outpm = int_2;
        random sub / type = vc solution;
        ods output solutionf = mu_2; ods output solutionr = beta_2;
        ods output covparms = cp_2;
        *ods select solutionf solutionr covparms;
run;
proc iml;
        use compdata var {sub pi1 pi2 HRI};read all var {sub pi1 pi2 HRI} into
AA;
        use beta_1 var {estimate};          read all var {estimate} into beta1;
        use beta_2 var {estimate};          read all var {estimate} into beta2;
        use mu_1 var {estimate};            read all var {estimate} into mu1;
        use mu_2 var {estimate};            read all var {estimate} into mu2;
        use cp_1 var {estimate};            read all var {estimate} into var1;
        use cp_2 var {estimate};            read all var {estimate} into var2;
        use newparmsold var {lambda mu1 mu2 sigma1 sigma2 sigdel};
        read all var {lambda mu1 mu2 sigma1 sigma2 sigdel} into pold;
        use thresh_old var {OLDdiff};       read all var{OLDdiff} into
oldthresh;
        nsub = max(AA[,1]); nobs = nrow(AA);
        lambda = AA[:,2];
        lambdavec = j(nsub,1,0);
        do j = 1 to nsub;
                sumpi1 = 0; ct = 0;
                do i = 1 to nobs;
                        if AA[i,1] = j then do; sumpi1 = sumpi1 + AA[i,2]; ct = ct
+ 1; end;
                end;
                *lambdavec[j] = sumpi1/ct; *If subjects lambda assumed to be
different;
                lambdavec[j]=lambda; *If subjects lambda assumed to be equal;
        end;
        print lambdavec;
```

```
beta = lambdavec#beta1 + (j(nsub,1,1)-lambdavec)#beta2;
print beta1 beta2 beta;
sigdel = (beta-beta[:])`*(beta-beta[:])/(nsub-1);                print sigdel;
sigdel1 = (beta1-beta1[:])`*(beta1-beta1[:])/(nsub); print sigdel1;
sigdel2 = (beta2-beta2[:])`*(beta2-beta2[:])/(nsub); print sigdel2;

*If subject lambdas are different;
/*      sigma1 = (nobs/nsub)*var1[2,1];      sigma2 = (nobs/nsub)*var2[2,1];
*/

*If subject lambdas are equal;
sigma1 = nobs*var1[2,1];      sigma2 = nobs*var2[2,1];

pnew = lambda || mu1 || mu2|| sigma1 || sigma2 || sigdel;
nameP = {lambda mu1 mu2 sigma1 sigma2 sigdel};
print pold;
print pnew;
create parmsold from pnew [colname = nameP];              append from
pnew;
pdiff = pnew - pold; newthresh = pdiff[##];
print pdiff;
print newthresh;
create oldthresh from newthresh [colname = "OLDdiff"];    append from
newthresh;
threshdiff = oldthresh-newthresh; print threshdiff;
create Thresh1 from threshdiff [colname = "diff"];        append from
threshdiff;

y = AA[,4];    *observations;
Zbeta = j(nobs,1,0);
do j = 1 to nsub;       do i = 1 to nobs;
                        if AA[i,1] = j then Zbeta[i] = beta[j];
end;    end;
e = y - Zbeta;  *residuals;

*print y Zbeta e;
*Determine the vector contining each residuals density value under the
specified component;
pi = 3.141592654; *pi constant;
phi1 = (1/sqrt(2#pi#(sigma1)))#exp(-((e-j(nobs,1,mu1))##2)/(2#sigma1));
phi2 = (1/sqrt(2#pi#(sigma2)))#exp(-((e-j(nobs,1,mu2))##2)/(2#sigma2));

/**************If Lambda is differect across subjects*************/
*Determine each residuals weight;
```

```
/*      sumpi1 = j(nsub,1,0); sumpi2 = j(nsub,1,0);
        wt1 = j(nobs,1,0);              wt2 = j(nobs,1,0);
        do j = 1 to nsub;
                subpi1 = j(nobs,1,0);  subpi2 = j(nobs,1,0);
                do i = 1 to nobs;
                        if AA[i,1]=j then do:  subpi1[i] = pi1[i];
subpi2[i]=pi2[i];       end;
                end;
                sumpi1[j] = subpi1[+];          sumpi2[j] = subpi2[+];
                do i = 1 to nobs;
                        if AA[i,1]=j then do; wt1[i]=pi1[i]/sumpi1[j];
wt2[i]=pi2[i]/sumpi2[j];        end;
                end;
        end;    */

        /************If Lambda is equal across subjects************/
        *Determine the vector containing each residuals probability of being in a
particular component;
        pi1 =(lambda#phi1)/(lambda#phi1 + (1-lambda)#phi2);
        pi2 =((1-lambda)#phi2)/(lambda#phi1 + (1-lambda)#phi2);
        *Determine each residuals weight;
        wt1 = pi1/pi1[+];       wt2 = pi2/pi2[+];

/*****************************************************************/

        *print sumpi1 sumpi2;
        *print phi1 phi2 pi1 pi2 wt1 wt2;
        sumwt1 = wt1[+];        sumwt2 = wt2[+]; print sumwt1 sumwt2;
        *Determine the two sets of weighted observations;
        Y1 = Zbeta + (wt1##(1/2)#e);Y2 = Zbeta + (wt2##(1/2)#e);
        wt1int = wt1##(1/2);  wt2int = wt2##(1/2);
        *print y y1 y2 wt1 wt2 wt1int wt2int;
        A = AA[,1] || y || y1 || y2 || wt1int || wt2int || pi1 || pi2;
        namesA = {sub HRI HRI1 HRI2 wt1int wt2int pi1 pi2};
        create BB from A [colname=namesA]; append from A;
        print &count;
quit;
data Thresh;    set Thresh1;    if diff < 0.0001 then call symput('check', 1);    run;
%if &count=1 %then %do;
        data converge; iter = &count; set parmsold;  run;
%end;
%else %do;
        data converge_;         iter = &count; set parmsold;  run;
        data converge; set converge converge_;                          run;
```

```
        %end;
        %end; *ends EM loop:
proc print data = converge;    run;
%if &sim = 1 %then %do;
        data EMvalues;        set parmsold;  sim = &sim; TotIter = &count;
        run;
%end;
%else %do;
        data EMvalues_;       set parmsold;  sim = &sim;          TotIter = &count;
        run;
        data EMvalues;        set EMvalues EMvalues_;                  run;
%end;
%end; *Ends the sim do loop;
proc print data = EMvalues;
run;
data SIMS.EMvalues&name;
        set EMvalues;
run;
proc means data = SIMS.EMvalues&name n mean std stderr;
run;
%MEND;      *Ends the EMJMX Macro:
```

The *EMJMX* macro shown above is used to estimate the parameters of the two-component normal-mixture random-effects density for the 100 simulated data sets produced by the *Simulate* macro described in Appendix 7.2. Suppose the *Simulate* macro is used to create 100 simulated data sets with 20 subjects, 200 observations per subject, component means 70 and 86.44, component variance 25 and 25, between subject variance 46.28, and mixture proportion 0.5. Then the EMJMX macro is called with the following statement:

```
%EMJMX(100, 20, .5, _1_20_200);
```

For this call statement, the total number of simulated data sets (100), number of subjects (20), and an initial value for the mixture proportion (0.5) must be specified. In addition, a label for the output data must be specified. Any label may be used, however the label shown in the call statement above, _1_20_100, specifies that the EM algorithm is estimating model parameters for simulated data determined by the first simulation vector in Table 2, with 20 subjects and 200 observations per subject. The 100 input data sets all have four columns (variables), *HRI*, *sub*, *sim*, and *obsnum*. The *HRI* column contains the simulated (heart rate interval) two-component normal-mixture random-effects data. The *sub*, *sim*, and *obsnum* columns indicate the corresponding subject, simulation, and observation number indicators. The following describes the EMJMX macro in detail.

The macro begins with the execution of a do-loop, called the simulation do-loop, iterating from 1 to *nsim*. Next, a data step is used to rename simulated data set with a general name, *simdata*, to be used throughout the macro. The data step also creates an intercept variable, called *int*, to be used in the model.

Now, before the actual EM-algorithm starts iterating, initial values for the parameters must be determined. The methodology for determining these initial values was described in section 3.1 and is implemented here. The initial value for the mixture proportion is specified as 0.5 in the call statement for the EMJMX macro. The univariate procedure seen following the data step described above determines the 25% and 75% quartiles of the *HRI* data in the data set *simdata*. The output statement creates a dataset called *quan* containing these values. The "ods select none" and "ods select all"

statements seen immediately preceding and following the univariate procedure, respectively, are used to minimize the amount of output displayed by SAS. Next, a (unimodal) normal random-effects model is fit to determine the initial values for the component variances and the between subject variance. This model is fit with the proc mixed procedure. The following describes this procedure.

The input data set is *simdata*. The estimation procedure should be set to the maximum likelihood (ml) method rather then the (default) reduced maximum likelihood method (reml). (This is not really necessary in this step – however I am trying to remain consistent though out the macro). The variable *sub* in the data set *simdata* should be indicated as a class variable. The model statement is defined with *HRI* as the response and no model effects (model HRI = /). The options for the model statement must include outp and outpm. These options are used to create two data sets, *resuni0* and *intuni0*, containing the two types of residuals determined by the fit model. These will be explained in more detail later. The random statement is used to include a random intercept for each subject (random sub /). The option type = vc must be used to indicate that the subject effects are independent and identical normal random variables with mean 0 and variance σ_{δ}^{2}. The solution options in both the model and random statements are used to display the model intercept and the predicted subject effects in the output window. The options for the proc mixed procedure, noclprint, noinfo, and noitprint are used to minimize the output displayed by SAS. The two ods output statements are used to create two data sets, *beta0* and *cpuni0*. The data set *beta0* contains the *nsub* values of the predicted subject effects. The data set *cpuni0* contains the estimates for the between

and within subject variance estimated by the model. The ods select statement used within the proc mixed procedure is also used to minimize the output displayed by SAS. The initial value for the between subject variance parameter will be set equal to the estimate for the between subject variance obtained from the mixed model described above (contained in cpuni0). The initial values for the component variances are taken to be equal and set at half of the value of the estimate for the within subject variance obtained from the mixed model described above (contained in cpuni0).

Now that starting values have been determined, the EM-algorithm may be used to estimate the model parameters for the two-component normal-mixture random-effects model. As described in 2.8.2, each iteration of the EM-algorithm consists of an E-step followed by an M-step. Then, if certain convergence criteria are met the algorithm ends, otherwise the algorithm continues with the next iteration. After the proc mixed procedure, the macro calls the IML procedure. This procedure executes the first E-step of the EM-algorithm. The following describes this procedure.

After calling the IML procedure several variables from various datasets are read into to IML and stored into matrices or vectors. These are listed below.

Dataset read	Variables Read	IML matrix/vector name
quan	q_25, q_75	muSV
cpuni0	estimate	varSV
beta0	estimate	betaSV
resuni0	HRI, sub, pred	resSV
intuni0	pred	intSV

Next, several (IML) variables are defined. The IML variable *nobs* is equal to the total number of observations in the data set, across all subjects, determined as the number of rows in the matrix *resSV*. (Please note that this IML variable is different than the

macro variable *nobs* seen in the simulation macro, which is equal to the number of observations per subject). The second column in the matrix *resSV* contains the subject identifiers. The variable *nsub* is taken to be the maximum value of this column, which is equal to the total number of subjects (since the subjects were numbered consecutively starting with 1 when the data was simulated).

Now, the initial values can be defined with IML variables. The variable *lambda* is set equal to the macro variable *lambdaSV*, defined in the call statement for the macro as 0.5. The variables for the initial values for the component means, *mu1* and *mu2*, are defined as the 25% and 75% quartiles, respectively, determined earlier with the univariate procedure, stored in the first and second elements, respectively, in the vector *muSV*. The variables for the initial values for the component variances, *sigma1* and *sigma2*, are both defined as the half the within subject variance estimate determined earlier by the proc mixed procedure, stored in the second element of the vector *varSV*. The variable for the initial value for the between subject variance, *sigdel*, is defined as the between subject variance estimate determined earlier by the proc mixed procedure, stored in the first element of the vector *varSV*. Next, a (row) vector, *pold*, is defined containing the initial values for the parameters. The order of the elements is set to be *lambda*, *mu1*, *mu2*, *sigma1*, *sigma2*, and *sigdel*. Names are then given to the 6 columns in the row vector *pold*, taken to be the same names as the IML variables used to create the vector, with the name vector *nameP*. After printing *pold* and *nameP* to the output window, these vectors are used to create a dataset, *paramsold*. Next, an arbitrarily large value is stored into the

variable *thresh* (100,000) and a dataset, *oldthresh* is created from it. This variable will be discussed more later in this description.

Next, the residuals for the two-component normal-mixture random effect model, which are used to compute the incomplete data, are defined. Recall that the i^{th} observation from the j^{th} subject is defined as the sum of the predicted value of the random effect for the j^{th} subject and the i^{th} residuals from the j^{th} subject, $y_{ij} = \beta_j + \varepsilon_{ij}$. In order to compute the incomplete data, the residuals must be determined. The *nobs* × 1 vector of predicted values determined by the proc mixed procedure above are used. This vector, called *Zbeta*, is computed from the two sets of predicted values computed by the proc mixed procedure above, and stored in the matrices *resSV* and *intSV*. The predicted values outputted to the data set *intuni0*, stored in the matrix *intSV* are defined as $\hat{\mathbf{Y}}_I = \mathbf{X}\hat{\boldsymbol{\alpha}}$. The predicted values outputted to the data set *resuni0*, stored in the matrix *resSV* are defined as $\hat{\mathbf{Y}}_R = \mathbf{X}\hat{\boldsymbol{\alpha}} + \mathbf{Z}\hat{\boldsymbol{\beta}}$. The difference, then, $\hat{\mathbf{Y}}_R - \hat{\mathbf{Y}}_I = \mathbf{Z}\hat{\boldsymbol{\beta}}$ (*Zbeta*) yields the *nobs* × 1 vector of predicted subject effects. A vector, *y*, is created containing the observations, stored in the first column of the matrix *resSV*. The vector of residuals, *e*, (for the two-component normal-mixture random effect model in the first iteration) is then defined as the vector of observations, *y*, minus the vector of predicted values for the random subject effects (*Zbeta*). After defining the constant value $\pi = 3.14549\cdots$ with the variable pi, the incomplete data can be computed. This process is described below.

Recall that the incomplete data are defined as

$$\pi_{ijk} = \frac{\lambda_{jk}\phi\left(\varepsilon_{ij};\mu_{ijk},\sigma_k^2\right)}{\sum_{m=1}^{M}\lambda_{jk}\phi\left(\varepsilon_{ij};\mu_{ijk},\sigma_k^2\right)}.$$

First, two vectors, *phi1* and *phi2*, are created containing the values $\phi\left(\varepsilon_{ij};\mu_{ij1},\sigma_1^2\right)$ and $\phi\left(\varepsilon_{ij};\mu_{ij2},\sigma_2^2\right)$, respectively, for each of the two components, using the vector of residuals and the (component) parameters determined by the initial values. For the component means, this first E-step assumes that the component means are equal across observation and subject levels (differing only across components). Then two vectors, *pi1* and *pi2*, containing the incomplete data for each of the two components are computed using the two vectors *phi1* and *phi2*, respectively, and the initial value for the mixture proportion. Next, the observations are weighted before the E-step concludes and the M-step initiates. This process is described below.

Recall that the weights for the two-component normal-mixture random-effects model, where the mixture proportion is assumed equal across subjects, are defined as

$$w_{ijk} = \frac{\pi_{ijk}}{\sum_{j=1}^{S}\sum_{i=1}^{N_j}\pi_{ijk}}.$$

This equation is used to compute the two vectors of weights, *wt1* and *wt2*, for the two components. Next, the two component vectors of weighted observations, Y1 and Y2, are created with the formula $\mathbf{Y}_k = \mathbf{Z}\boldsymbol{\beta} + \mathbf{W}_k^{1/2}\boldsymbol{\varepsilon}$. Here, the *nobs* × 1 vector of predicted subject effects, $\mathbf{Z}\boldsymbol{\beta}$, are stored in *Zbeta*, the *nobs* × 1 vector of residuals, $\boldsymbol{\varepsilon}$, are stored in

the vector e, and the component vectors of weights, $\mathbf{W}_k$, are stored in *wt1* and *wt2*. Next, two intercept vectors are created, *wt1int* and *wt2int*, with elements equal to the square roots of the elements of the weight vectors, *wt1* and *wt2*, respectively. These vectors will be used as the intercepts in the proc mixed procedures performed in the subsequent M-step.

Before the conclusion of the first E-step a data set is created containing the relevant information for the M-step. This is done by first creating a matrix, *A*, with the following data contained in its columns.

- The subject identifier v ector, stored in the second column of the matrix *resSV*.
- The observation ve ctor, *y*.
- The first component w eighted observation vector, *Y1*.
- The second component weighted observation vector, *Y2*.
- The first component w eighted intercept vector, *wt1int*.
- The second component weighted intercept vector, *wt2int*.
- The first component inco mplete data vector, *pi1*.
- The second component incomplete vector, *pi2*.

The names of these columns (sub, HRI, HRI1, HRI2, wt1int, wt2int pi1, and pi2) are stored in the vector *namesA*. Then, the data set *BB* is created using the matrix *A* and the vector *namesA*. This concluded the IML procedure.

Next, global macro statements are defined to set up the do-until-loop used for the iterations of the EM-algorithm. The macro variables *check* and *count* are initiated with the value 0. The do-until-loop is then set up to iterate through the EM-algorithm until the variable *check* becomes equal to 1. A check at the conclusion of each M-step will determine when this variable changes. The do-until-loop used for the iterations (all subsequent E and M-steps) of the EM-algorithm is described below.

First, the macro variable *count* is incremented by an integer value of 1. This variable is used to keep track of the number of iterations in the EM-algorithm. The three data sets, *BB*, *paramsold*, and *oldthresh*, are then renamed as *compdata*, *newparamsold*, *thresh_old*, respectively. (This is necessary for the data to be read into ILM procedures performed later). The M-step then begins with the execution of two proc mixed procedures using the data set *compdata*. Both procedures must indicate sub as a class variable and the maximum likelihood (ml) procedure for estimation. Furthermore, both models use a random statement to include a random intercept for each subject (random sub /). The option type = vc must be used to indicate that the subject effects are independent and identical normal random variables with mean 0 and variance σ_{δ}^{2}. The first proc mixed procedure uses the first component weighted observations, HRI1, as the response variable and the weighted intercept, wt1int, as the only fixed effect. Similarly, the second proc mixed procedure uses the second component weighted observations, HRI1, as the response variable and the weighted intercept, wt2int, as the only fixed effect. The option is then given, in both procedures, to exclude the default intercept included by SAS. The inclusion of this "user-defined" weighted intercept results in the two procedures appropriately estimating the two component means. The model and random statements in both procedures also include the solution option, used to display relevant information regarding the component means and the predicted values of the random-effects to the output window (also must be used to produce the output data sets). The two proc mixed procedures create several output data sets described below.

- mu_1 (from the first proc edure) fixed-effects solution vector.
- mu_2 (from the second p rocedure) fixed-effects solution vector.

- cp_1 (from the first procedure) estimated covariance parameters.
- cp_2 (from the s econd procedure) estimated covariance parameters.
- beta_1 (from the first procedure) random-effects solution vector.
- beta_2 (from the s econd procedure) random-effects solution vector.

Next, the IML procedure is executed in order to perform the necessary matrix computations.

First, the variables *sub*, *pi1*, *pi2*, and *HRI*, from the *compdata* data set, are read into IML and stored into the matrix *AA*. Next, the *estimate* variable from each of the six data sets list above is read into IML and stored into the vectors *mu1*, *mu2*, *var1*, *var2*, *beta1*, and *beta2*, respectively. The parameter estimates in the data set newparamsold are read into IML and stored into the row vector pold. The data set thresh_old is read into IML and stored into the vector oldthresh.

The IML variable *nsub* is defined and taken to be the maximum value of the first column in the matrix *AA*, which is equal to the total number of subjects (since the subjects were numbered consecutively starting with 1 when the data was simulated). The IML variable *nobs* is defined and is equal to the total number of observations in the data set, across all subjects, determined as the number of rows in the matrix *AA*. Next, the parameters for the two-component normal-mixture random-effects model are estimated. This process is described below.

The mixture proportion, common to all subjects, is estimated with

$$\hat{\lambda} = \frac{\sum_{j=1}^{S} \hat{\lambda}_j}{S} = \frac{\sum_{j=1}^{S} \sum_{i=1}^{N_j} \pi_{ij}}{N}.$$

Then the estimate for the mixture proportion, lambda, is computed by averaging the elements in the second column of *AA*, which contains the elements from *pi1*. The component means were computed by the proc mixed procedure are stored in the in the variables *mu1* and *mu2*. The within subject variance parameter estimates determined by the proc mixed procedure, stored in the second element of the vectors *var1* and *var2*, respectively, must be multiplied by the total number of observation, *nobs*. Their results are then stored with the variable name *sigma1* and *sigma2*. The between subject variance estimate is determined from the vector of predicted values for the random subject effects. Thus, vector of predicted values, *beta*, is determined, common to both components using the formula $\hat{\beta}_j = \hat{\lambda}\hat{\beta}_{j1} + \left(1-\hat{\lambda}\right)\hat{\beta}_{j2}$. This is achieved by multiplying the vector of predicted values of the random subject effects from the first mixed procedure, *beta1*, by *lambda* and multiplying the predicted values of the random subject effects from the second mixed procedure, *beta2*, by 1 – *lambda*. The sum of these two results is then stored in the vector *beta*. The estimate for the between subject variance parameter, *sigdel*, is then determined using the formula

$$\hat{\sigma}_{\delta}^{2} = \frac{1}{(S-1)} \sum_{j=1}^{S} \left(\hat{\beta}_j - \bar{\beta}\right)^2$$

with $\bar{\beta}$ denoting the mean of the $\hat{\beta}_j$ and S the total number of subjects, *nsub*.

The six parameter estimates, lambda, mu1, mu2, sigma1, sigma2, and sigdel, are then stored in the row vector *pnew* and their names are stored in the row vector *nameP*. The row vector containing the previous iterations' (initial values if in the first iteration) parameter estimates, *pold*, is printed to the output window along with the new parameter

estimates in the row vector *pnew*. Next, the *pnew* row vector and its corresponding name vector, *nameP*, are used to create a dataset called *paramsold* (this replaces the old data set). The vectors pold and pnew are shown below for clarity.

$$pold = \theta^{(t)} = \begin{pmatrix} \lambda^{(t)} & \mu_1^{(t)} & \mu_2^{(t)} & \sigma_1^{2(t)} & \sigma_2^{2(t)} & \sigma_\delta^{2(t)} \end{pmatrix}$$

$$pnew = \theta^{(t+1)} = \begin{pmatrix} \lambda^{(t+1)} & \mu_1^{(t+1)} & \mu_2^{(t+1)} & \sigma_1^{2(t+1)} & \sigma_2^{2(t+1)} & \sigma_\delta^{2(t+1)} \end{pmatrix}$$

After determining the new parameter estimates, they are then compared to the old parameter estimates. The vector, *pdiff*, is created containing the differences between each of the new and old parameter estimates (this is an element wise subtraction):

$$pdiff = \theta^{(t+1)} - \theta^{(t)} = \begin{pmatrix} \lambda^{(t+1)} - \lambda^{(t)} \\ \mu_1^{(t+1)} - \mu_1^{(t)} \\ \mu_2^{(t+1)} - \mu_2^{(t)} \\ \sigma_1^{2(t+1)} - \sigma_1^{2(t)} \\ \sigma_2^{2(t+1)} - \sigma_2^{2(t)} \\ \sigma_\delta^{2(t+1)} - \sigma_\delta^{2(t)} \end{pmatrix}'.$$

The variable *newthresh* is then the sums of these squared differences. This variable is then stored into an output data set, called *oldthresh*, and quantifies the difference between the current iteration's parameter estimates and the previous iteration's parameter estimates. Convergence, however, is satisfied when the difference, *newthresh*, determined in the current iteration is larger than the difference, *oldthresh*, determined in the previous iteration. (For the first iteration the arbitrarily large value 100,000 was used for *oldthresh*). This variable *threshdiff* is then created containing this difference. Then

this variable is used to create the dataset *Thresh1*, which will be used later to actually check for convergence. This concludes the M-step. The next iteration's E-step is then executed within the same IML procedure as the previous iteration's M-step. (This results in an additional E-step always being performed at the conclusion of the EM-algorithm).

The E-step then begins with the *nobs* × 1 vector *y* being defined as the vector of observations, stored in the fourth column of the matrix *AA*. The vector *Zbeta* is then defined as the *nobs* × 1 vector of random-effects taken from the elements in the common vector *beta*. (The first N_j elements are equal to β_1, the next N_j elements are equal to β_2, and so on though β_S). The *nobs* × 1 vector of residuals, *e*, is then computed as the vector of observations minus the vector Zbeta. After defining the constant value $\pi = 3.14549\cdots$ with the variable *pi*, the incomplete data can be computed. This process is described below.

First, two vectors, *phi1* and *phi2*, are created containing the values $\phi\left(\varepsilon_{ij};\mu_{ij1},\sigma_1^2\right)$ and $\phi\left(\varepsilon_{ij};\mu_{ij2},\sigma_2^2\right)$, respectively, for each of the two components, using the vector of residuals and the parameter estimates determined in the previous M-step and stored in the row vector *pnew*. Then two vectors, *pi1* and *pi2*, containing the incomplete data for each of the two components are computed using the two vectors *phi1* and *phi2*, respectively, and the estimate for the mixture proportion. Next, the observations are weighted before the E-step concludes and the next M-step initiates. This process is described below.

The two vectors of weights, *wt1* and *wt2*, for the two components are computed as in the first E-step. Next, the two component vectors of weighted observations, *Y1* and

Y2, are created with the formula $\mathbf{Y}_k = \mathbf{Z}\boldsymbol{\beta} + \mathbf{W}_k^{1/2}\boldsymbol{\varepsilon}$. Here, the *nobs* × 1 vector of predicted subject effects, $\mathbf{Z}\boldsymbol{\beta}$, are stored in *Zbeta*, the *nobs* × 1 vector of residuals, $\boldsymbol{\varepsilon}$, are stored in the vector *e*, and the component vectors of weights, $\mathbf{W}_k$, are stored in *wt1* and *wt2*. Next, two intercept vectors are created, *wt1int* and *wt2int*, with elements equal to the square roots of the elements of the weight vectors, *wt1* and *wt2*, respectively. These vectors will be used as the intercepts in the proc mixed procedures performed in the subsequent M-step.

Before the conclusion of the E-step a data set is created containing the relevant information for the subsequent M-step. This is done by first creating a matrix, *A*, with the following data contained in its columns.

- The subject identifier vector, stored in the first column of the matrix *AA*.
- The observation ve ctor, *y*.
- The first component w eighted observation vector, *Y1*.
- The second component weighted observation vector, *Y2*.
- The first component w eighted intercept vector, *wt1int*.
- Th e second component weighted intercept vector, *wt2int*.
- The first component inco mplete data vector, *pi1*.
- The second component incomplete vector, *pi2*.

The names of these columns (sub, HRI, HRI1, HRI2, wt1int, wt2int pi1, and pi2) are stored in the vector *namesA*. Then, the data set *BB* is created using the matrix *A* and the vector *namesA*. This concluded the IML procedure.

Next, the check for convergence is performed before the EM loop is concluded. Here a data set, *Thresh*, is created from the previous data set created, *Thresh1*. Then an if statement is used to determine if the difference variable has fallen below 0.0001. If the difference has gotten this small then the variable check takes on the value of 1 (indicating

convergence), otherwise the value remains 0 and the iterations through the EM algorithm continue. The %if %then %do %else %do loop sequence that follows creates a data set, *converge*, with the parameter estimates from each iteration, along with the corresponding iteration number. The EM loop is then concluded.

After the process has converged, the *converge* data set is printed with a proc print procedure. Next, before the simulation loop is concluded, the %if %then %do %else %do loop sequence that follows creates a data set, *EMvalues*, with the final (converged) parameter estimates from each simulation, along with the corresponding simulation number. The simulation loop is then concluded.

After the simulation loop has been concluded, the EMvalues data set is printed with a proc print procedure and the final data set is stored in the library *SIMS* with the name *EMvalues_1_20_200* since the value _1_20_200 was stored in the macro variable name when the macro was called. This macro variable is primarily used so that the data sets created with each call of the EMJMX macro can be labeled uniquely. The means procedure is used to obtain the mean, standard deviations and standard errors across the 100 replications with the same parameter vector. These are used to compute the bias squared and MSE used in section 3.4. The EMJMX macro is then concluded.

7.4. The Data CD

The data CD contained with this dissertation contains several data sets used throughout this dissertation. The following subsections describe each file in depth.

7.4.1 Estimates.JMP

Column Names

LAMBDA
MU1
MU2
SIGMA1
SIGMA2
SIMDEL
Sim
TotIter
PV_obs
PV
Nsub
Nobs
Num

Description of Data

This data set contains 7200 rows, 100 for each of the 72 combinations of simulation parameter and factor vectors. The columns LAMBDA, MU1, MU2, SIGMA1, SIGMA2, and SIGDEL contain the parameter estimates determined by the EM-algorithm. The Sim column indicates the simulation number (1-100) within each of the 72 combinations. TotIter indicates the number of iterations necessary for convergence with the EM-algorithm. The PV column indicates the parameter vector number (1-12) from which the data was simulated. The nsub and nobs columns indicate the number of subjects and observations per subject used to simulate the data. The PV_obs column is a name column formed by concatenating the PV, Nsub, and Nobs

columns. The Num column then gives indicators (1.1 – 1.6, 1.2 – 2.6,…, 12.1 – 12.6) specifying the PV value (parameter vector number) then the following after the dot (.):

1 – 5 subject, 100 observations
2 – 5 subjects, 200 observations
3 – 10 subjects, 100 observations
4 – 10 subject, 200 observations
5 – 20 subjects, 100 observations
6 – 20 subjects, 200 observations

This data set is used to compute 72 means and variances for each combination using the 100 runs (rows) for each combination. These averages are found in the dataset AvgEstimates.JMP.

7.4.2 AvgEstimates.JMP

Column Names

Parameter
POV
BSF
Lambda
S
Nj
True
Estimate
SD
Bias(SQ)
MSE
AbsBias

Description of Data

This data set contains the estimates for the simulated data obtained from the EM-algorithm. The Parameter column indicates which parameter is being estimated. The POV, BSF, Lambda, S, and Nj columns refer to the levels of the five factors describe in

section 3.1. The True column displays the actual parameter values requested for simulations. The Estimate and SD columns then display the average and standard deviation, respectively, of the 100 estimates for the parameter under the specified factor levels. The Bias(SQ), MSE, and AbsBias columns then display the computed squared bias, MSE, and absolute bias of the estimates, respectively.

7.4.3 EffectEstimates.JMP

Column Names

Parameter
Outcome(Transfm)
Term
Original
Orthogonalized
Standardized
Quantiles

Description of Data

This data set contains the effect estimates from the 12 full-factorial models fitting absolute bias and MSE. The Parameter column indicates which parameter the absolute bias/MSE are computed from. The Outcome(Transfm) variable indicates which response variable was used (absolute bias or MSE) for the model and the transformation of the response (none, log, square root, or square). The Term column indicates the name of the effect, including the particular level of the factors. The Original column lists the estimates of the effects for each combination of parameters and outcomes. The Orthogonalized and Standardized columns then list the orthogonalized (uncorrelated) and

standardized (equal variance) effect estimates. The Quantiles column lists the normal quantiles of the standardized estimates.

7.4.4 Lone_Raw.JMP

These two data sets contain the raw (edited) RR-intervals from the Loneliness study as received from Mandrekar. The names of the columns in the JMP file are listed below.

JMP column names

Sub

gender

Age

task

type

Tasktyp

RRI(ORIG)

Minute

Description of Data

This is the unedited version of the Loneliness data used for analysis in chapter 4. Each row represents a separate RR-interval. The subject variable is a subject identifying variable. The gender variable (Male/Female) indicates the gender of the subject. The Age variable indicates the age of the variable. The task and type variables indicate the tasks and periods, respectively, from which the RR-interval was collected. The tasktyp variable identifies the task and periods with a single value. The RRI(ORIG) variable

indicates the value (in milliseconds) obtained. The minute variable indicates the minute from which the RR-interval was measured.

7.4.5 Lone_analysis.SAS and Lone_analysis .JMP

These two data sets contain the detrended, artifact-free, centered, and rescaled data used for analysis in chapter 4. Both a JMP and a SAS file have been provided.

JMP (SAS) column names

sub (Sub)

gender (gender)

Age 2 (Age_2)

task (task)

type (type)

tasktyp (tasktyp)

RRI_ORIG (not in SAS data set)

Artifact (Artifact)

RRnumA (RRnumA)

RRnumB (RRnumB)

RRnumC (RRnumC)

Residuals HRI By Sub By tasktyp (not in SAS data set)

Predicted HRI By Sub By Tasktyp (not in SAS data set)

Intercept HRI By Sub By Tasktyp (not in SAS data set)

RRI_new (RRI)

Description of Data

Each row represents a separate RR-interval. The subject variable is a variable indicating the subject identifier. The Age 2 variable indicates the centered age (difference from mean age) value for the subject. The task and type variables indicate the tasks and periods, respectively, from which the RR-interval was collected. The tasktyp variable identifies the task and periods with a single value. The RRI_ORIG variable (not in the SAS file) contains the original measurements of the RR-intervals. The Artifact variable indicates which intervals are artifacts (1 if artifact, 0 otherwise). The RRnumA, RRnumB, and RRnumC are numbering variables. RRnumA numbers the RR-intervals within each minute of recording. RRnumB numbers the RR-intervals within each task type combination. RRnumC numbers the RR-intervals within each subject. The Residuals HRI By Sub By tasktyp, Predicted HRI By Sub By Tasktyp, Intercept HRI By Sub By Tasktyp variables in the JMP data set (and not found in the SAS data set) are used to detrend the data. The detrended RR-intervals are then stored in the variable RR_new (RRI in the SAS data set).

7.5. Loneliness Data

The first baseline minute of Loneliness data during the WS task obtained from the first subject is displayed in Table 21.

Table 21: Subject 1 - First Baseline Minute of Data for the WS Task

HRI	Subject	Gender	Age	Minute	Task	Period
450.00	202	Female	19	1	WS	base
954.00	202	Female	19	1	WS	base
930.00	202	Female	19	1	WS	base

980.00	202	Female	19	1	WS	base
990.00	202	Female	19	1	WS	base
910.00	202	Female	19	1	WS	base
980.00	202	Female	19	1	WS	base
1026.00	202	Female	19	1	WS	base
978.00	202	Female	19	1	WS	base
906.00	202	Female	19	1	WS	base
1000.00	202	Female	19	1	WS	base
1038.00	202	Female	19	1	WS	base
970.00	202	Female	19	1	WS	base
922.00	202	Female	19	1	WS	base
1008.00	202	Female	19	1	WS	base
1054.00	202	Female	19	1	WS	base
974.00	202	Female	19	1	WS	base
918.00	202	Female	19	1	WS	base
984.00	202	Female	19	1	WS	base
1018.00	202	Female	19	1	WS	base
974.00	202	Female	19	1	WS	base
898.00	202	Female	19	1	WS	base
980.00	202	Female	19	1	WS	base
1054.00	202	Female	19	1	WS	base
1030.00	202	Female	19	1	WS	base
908.00	202	Female	19	1	WS	base
970.00	202	Female	19	1	WS	base
1038.00	202	Female	19	1	WS	base
1014.00	202	Female	19	1	WS	base
916.00	202	Female	19	1	WS	base
1006.00	202	Female	19	1	WS	base
1040.00	202	Female	19	1	WS	base
992.00	202	Female	19	1	WS	base
886.00	202	Female	19	1	WS	base
960.00	202	Female	19	1	WS	base
1006.00	202	Female	19	1	WS	base
950.00	202	Female	19	1	WS	base
922.00	202	Female	19	1	WS	base
972.00	202	Female	19	1	WS	base
1010.00	202	Female	19	1	WS	base
926.00	202	Female	19	1	WS	base
958.00	202	Female	19	1	WS	base
1042.00	202	Female	19	1	WS	base
1018.00	202	Female	19	1	WS	base
910.00	202	Female	19	1	WS	base
932.00	202	Female	19	1	WS	base

992.00	202	Female	19	1	WS	base
1008.00	202	Female	19	1	WS	base
902.00	202	Female	19	1	WS	base
946.00	202	Female	19	1	WS	base
1014.00	202	Female	19	1	WS	base
976.00	202	Female	19	1	WS	base
872.00	202	Female	19	1	WS	base
938.00	202	Female	19	1	WS	base
1020.00	202	Female	19	1	WS	base
996.00	202	Female	19	1	WS	base
908.00	202	Female	19	1	WS	base
988.00	202	Female	19	1	WS	base
1026.00	202	Female	19	1	WS	base
970.00	202	Female	19	1	WS	base
872.00	202	Female	19	1	WS	base
940.00	202	Female	19	1	WS	base
940.00	202	Female	19	1	WS	base

7.6. Linear Detrending

This appendix contains the estimated slopes and intercepts for the fitted first-order polynomials for the nine subjects in the Loneliness study used in Chapter 4, across all combinations of task and period.

Table 22: Slope and Intercepts for Detrending

Slope Intercept		**Task**				
Subject	**Period**	BS	AS	IS	LS	WS
1	Baseline	924.74 -0.10	947.39 0.01	939.48 -0.12	940.11 -0.17	978.01 -0.21
	Preparation	840.59 0.15	913.75 -0.13	927.29 -0.32	868.92 -0.11	899.11 0.70
	Delivery	732.31 0.27	713.22 0.88	775.05 0.77	711.73 0.78	783.14 0.53
2	Baseline	897.53 -0.17	827.22 -0.28	800.37 0.13	807.22 0.21	799.89 0.18
	Preparation	769.31 0.55	627.01 0.80	783.58 0.07	774.29 0.21	773.71 0.19
	Delivery	726.36 0.50	625.86 0.88	674.34 0.29	766.88 -0.10	738.93 0.40
3	Baseline	938.28 -0.04	774.19 0.59	824.15 -0.40	782.75 0.23	813.91 -0.01
	Preparation	921.32 -0.29	897.65 0.44	845.20 0.35	863.06 0.39	869.71 -0.04
	Delivery	797.21 -0.27	814.26 -0.41	768.77 -0.11	741.47 0.05	776.49 -0.40
4	Baseline	927.85 0.27	933.23 0.34	846.37 0.55	980.87 -0.75	868.50 0.57
	Preparation	855.72 0.74	781.37 -0.03	928.04 -0.08	853.25 0.75	854.38 0.14
	Delivery	917.11 -0.86	721.73 1.03	870.84 0.42	828.46 0.60	843.73 0.54
5	Baseline	848.18 0.27	857.94 -0.15	853.58 -0.01	865.74 0.19	838.85 -0.14
	Preparation	839.10 -0.12	810.25 -0.07	792.38 0.12	798.34 -0.15	839.82 -0.07
	Delivery	712.79 0.43	764.00 0.17	773.14 -0.09	754.65 0.50	765.41 0.36

Table 22 (continued): Slope and Intercepts for Detrending

Slope Intercept		Task				
Subject	**Period**	BS	AS	IS	LS	WS
6	Baseline	901.25 0.14	910.60 -0.14	942.12 0.14	906.90 -0.001	979.59 -0.22
	Preparation	900.73 -0.22	820.21 0.35	892.86 0.32	815.05 0.03	885.34 -0.04
	Delivery	730.97 0.43	693.12 0.49	838.32 0.28	661.56 0.80	795.35 0.27
7	Baseline	1071.91 0.09	1015.24 0.91	987.45 1.11	988.42 0.58	1073.71 0.30
	Preparation	980.27 -0.91	907.40 -0.51	991.36 -0.06	802.83 0.50	743.76 0.17
	Delivery	717.00 0.34	728.07 0.74	877.04 -0.03	722.63 0.43	726.75 0.54
8	Baseline	861.36 -0.22	780.49 0.46	776.63 0.03	784.93 0.05	810.79 0.11
	Preparation	774.87 0.91	732.67 0.24	802.99 -0.21	789.66 0.07	890.32 -1.00
	Delivery	742.61 0.45	669.98 0.52	758.77 0.04	761.11 -0.09	776.44 -0.41
9	Baseline	907.40 0.01	925.18 -0.15	942.12 -0.24	926.39 -0.41	926.79 -0.01
	Preparation	890.63 -0.20	744.74 0.26	793.10 0.60	841.38 0.04	677.99 0.76
	Delivery	683.96 0.73	692.21 0.63	732.41 0.80	729.11 0.38	688.25 0.73

7.7. EM code for the Loneliness Data

The code titled LONEcodeFINAL can be found on the data CD.

VITA

Jessica Lauren McKinney Ketchum was born April 20, 1978 in Farmington, New Mexico. She graduated from the Baylor School in Chattanooga, Tennessee in 1996. She received her bachelors of Science in Mathematics from New Mexico State University in May of 2001. At NMSU she was recognized for her academic achievements receiving the Crimson Scholar Award, as well as scholarships from the Departments of Mathematics and Engineering. In August of 2001, she moved to Richmond, Virginia to begin her graduate studies towards a Ph.D. in Biostatistics at Virginia Commonwealth University. She spent much of her graduate career consulting with the School of Nursing and the Department of Pathology within the School of Medicine. She has received the Charles C. Clayton and Karl E. Peace awards while at VCU. In addition, she has been a teaching assistant for a variety of graduate level Biostatistics courses. She currently is employed in a tenure track position with the Department of Biostatistics at Virginia Commonwealth University.