BAYESIAN INFERENCE ON A MIXED-EFFECTS LOCATION-SCALE MODEL WITH NORMAL AND SKEWED ERROR DISTRIBUTIONS

by

Brian McGill

BA in Physics and Astronomy

University of Pennsylvania 2010

MS in Financial Engineering,

Temple University 2012

MA in Applied Statistics,

University of Pittsburgh 2014

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Statisitcs

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Brian McGill

It was defended on

June 12, 2017

and approved by

Satish Iyengar, PhD, Department of Statistics

Stewart Anderson, PhD, Department of Biostatistics

Yu Cheng, PhD, Department of Statistics

Zhao Ren, PhD, Department of Statistics

Meredith Wallace, PhD, Department of Psychiatry

Dissertation Director: Satish Iyengar, PhD, Department of Statistics

BAYESIAN INFERENCE ON A MIXED-EFFECTS LOCATION-SCALE MODEL WITH NORMAL AND SKEWED ERROR DISTRIBUTIONS

Brian McGill, PhD

University of Pittsburgh, 2017

In handling dependent data, mixed-effects models are commonly used. These models allow for each individual in the population to vary randomly about an overall population location. Most methods focus on modeling the mean structure and treat the resulting between- and within-subject variances as nuisance parameters. Hedeker has extended these models to allow for simultaneous modeling of both the mean and variance components, each with appropriate random effects. His work has focused on data with large amounts of repeated observations (30-50) from a one-week period. His Marginal Maximum Likelihood estimation approach provides unbiased estimates in those situations, but oftentimes fails to provide feasible results for these mixed-effects location-scale models in other situations. By implementing a Bayesian Markov chain Monte-Carlo I am able to fit these models in a more general setting that can include repeat observations collected over a two-year span. I have also adapted this model to utilize the skew-normal distribution which allows for skewed-error distributions. In applying these techniques to data from a bipolar clinical trial, I am able to explain how different treatments impact the resulting scores for depression and mania in both their mean and variance. These techniques lend themselves to addressing many research questions that would focus on stabilizing the mood in their subjects.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

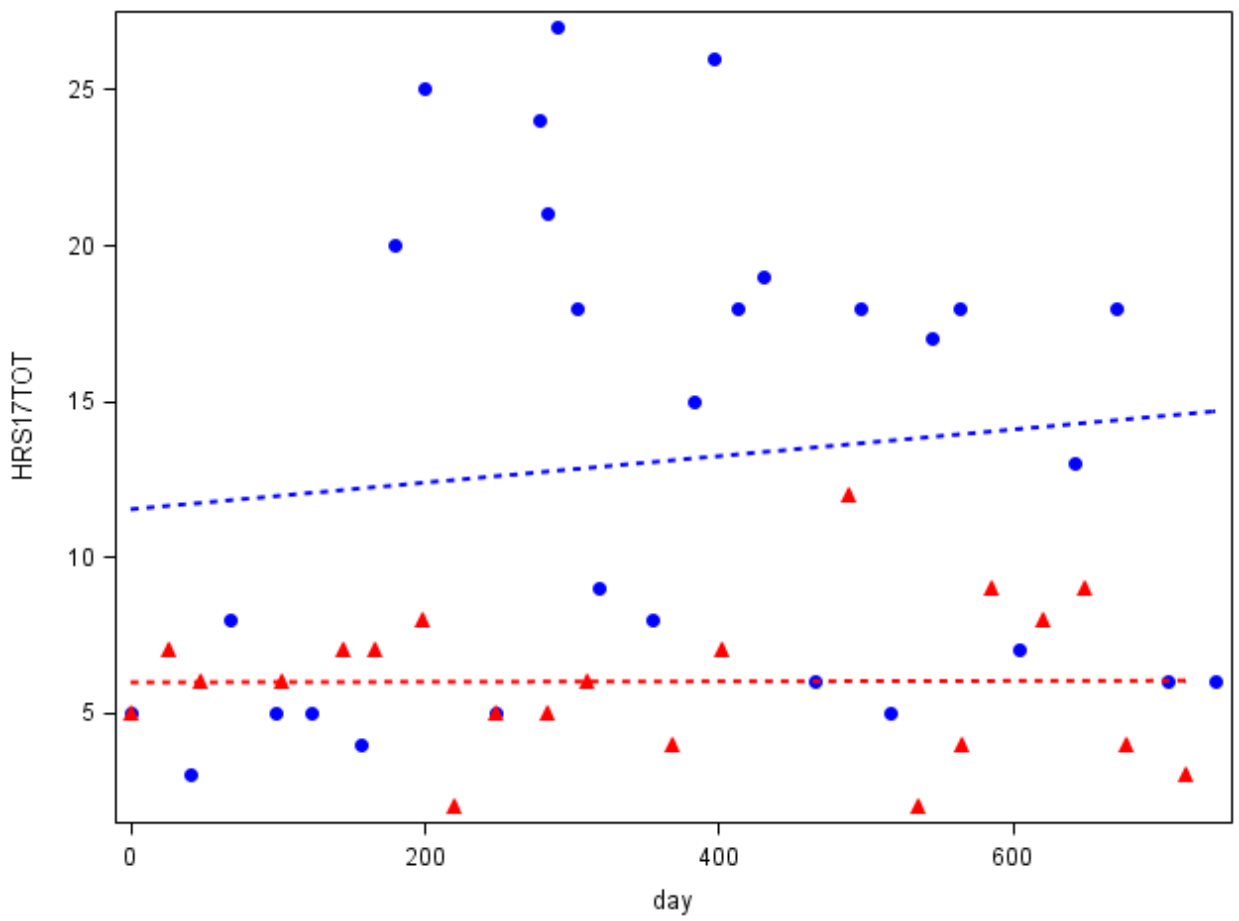
1.0 INTRODUCTION

Many applications in fields such as medicine and finance give rise to dependent data. Traditionally, individuals focus on modeling the mean structure of these types of data (e.g. comparing the mean trajectory of two treatments over time) and treat the within- and between-subject variability as nuisance parameters. However, Cleveland et al. (2000) showed using customer survey data that failure to deal with these nuisance parameters appropriately will lead to inefficient mean estimates. Cleveland introduced a method of dealing with random scale effects (subject specific variances) to account for this inefficiency. In addition there are many important questions that revolve around the variability of the observations as well as the mean. Hedeker et al. (2012) models both the mean and variance of dependent data in a study of mood disorders through the use of a mixed-effects location-scale model.

For data that arise from surveys, a common method of dealing with dependence due to repeated observations on the same individual is to use a mixed-effects location model. This model allows each individual in the population to vary randomly about an overall population location (mean) structure. It is widely acknowledged that survey rater data such as the ones in Cleveland et al. (2000) have not only varying locations (means), but varying scales (variances) as well (see Figure 1). It is uncommon however to model such scale effects by random effects due to the difficulties involved in model building and fitting (Clark et al., 1999).

The interest in modeling intraindividual variability in psychology stems from John Nesselroade of the University of Virginia, who in 1991 coined the phrase "measurement bursts" to describe the main feature of intensive longitudinal designs that allows for the collection of large amounts of data on individuals over a relatively short time period. This intensive data gathering method is also called ecological momentary assessment (EMA). In the inaugural issue of *Research in Human Development* (2004), Nesselroade explained the scientific and clinical advantages of studying

Figure 1: HRS17TOT trajectories for two individuals in the IRRI Group displaying both random location and scale effects (See Chapter 3 for data set description)



intraindividual variability, such as the prediction of a child's temperament in later years through the analysis of the intraindividual variability in infants' heart rates (Nesselroade and Ram, 2004).

Hedeker et al. (2012) introduced the mixed-effects location-scale (MELS) model in order to describe mood changes from an adolescent smoking study. In this study, similar to the work of Nesselroade, the focus shifts from examining not only the mean structure but the variability of the data or lability as well. With the MELS model it is now possible to analyze the effects that potential covariates have on the mood variability in an attempt to investigate different methods of stabilizing the lability.

Fitting complex multi-dimensional models such as the MELS model, frequentist estimation methods usually require the approximation of integrals. Accompanied with an optimization algorithm, the model fit either fails to converge, provides infeasible results, or provides suboptimal results for many data sets. The estimation method proposed by Hedeker et al. (2012) and based on Marginalized Maximum Likelihood, provides results for the adolescent smoking study, but may fail to provide results for other data sets. Research is needed in order to determine the best optimization method to use in fitting a MELS model with frquentist approaches. Along with the use of Markov chain Monte Carlo (MCMC), Bayesian techniques also allow for the estimation of complex multi-dimensional models without the use of an optimization algorithm that in many cases would not provide optimal, if any, results under frequentist approaches. I propose the use of a Bayesian Markov chain Monte Carlo using Gibbs Sampling in order to estimate the mixed-effects location-scale model and generalize its use to various fields of study.

A common assumption in fitting mixed-effects models is that the random effects (both the between- and within-subject) and the error distributions are normally distributed. While this assumption may be accurate in many situations, there are others where this assumption does not hold. One such scenario would be observations arising from a skewed distribution. Chen (2012) showed that when skewed distributions are fitted using normally distributed mixed-effects models the estimates are biased. A common method of eliminating this skewness is to transform the response variable; however, an unintended consequence of the transformation is the reduction of the variability in the transformed data. When the research question involves simultaneously modeling the mean and variance of the non-transformed data a skew distribution such as the skew-normal and skew-t (proposed by Chen) would address this issue.

Previous work by Hedeker and his colleagues have assumed that the within-subject errors are normally distributed. This assumption was used to fit an EMA data set where an average of 30 observations per subject were obtained over a one-week period. In this short time frame and for the data at hand the normal assumption was accurate. In a clinical trial conducted by the Department of Psychiatry at the University of Pittsburgh (Frank et al., 2015), bipolar patients currently in remission were studied to see if a new treatment would aid in lowering the patient's BMI even in the presence of potential weight inducing medicine. At baseline many patients had low manic and depressive symptom levels due to their remission; however, some individuals had

sub-threshold mania or depression symptoms. During the study many individuals would become manic or experience depressive episodes, resulting in skewed data such as the mania rating. This skewness violates the normality assumption of Hedeker's model and provides the motivation for my adaptation of the skew-normal distribution in the MELS model.

From the bipolar study the two response variables of interest are the Hamilton Rating Scale for Depression and the Young Mania Rating Scale. These variables are measured on the same individuals at the same time which would suggest that their underlying error distributions are closely related. Graphical results reveal normal error distributions for the Hamilton and skewed error distributions for the Young ratings. Azzalini (1985) proposed a skew-normal distribution that had strict inclusion of the normal density and allowed for a wide range of indices of skewness and kurtosis. In order to allow for comparable inferences between the two rating scales I am proposing a generalized form of the MELS model with skew-normal distributed errors in modeling the Young Mania Rating Scale. This distribution allows for the special case of the normal distribution, which the Hamilton distribution follows and has been used in previous applications of the mixed-effects location-scale model. The form of the skew-normal distribution is based on the one proposed by Chen (2012), but allows for random scale as well as random location effects. There are other skew distributions such as the class of skew-elliptical and skew-t distributions that may be used, which will be considered in later works.

In this dissertation I am proposing: 1) a new estimation technique (Bayesian MCMC) to use in estimating a MELS model; and 2) the use of a skew-normal distribution to model skewed data in a MELS model. In addition 3) the effects of sample size, both in number of subjects and number of observations per subject, and their impact on the detection of treatment effects in both the mean and variance structures will be explored.

This paper will be organized as follows. Chapter 2 will provide background information on a MELS model starting with a literature review followed by a description of the present estimation methods for the model. Chapter 3 will provide a detailed description of the bipolar data that will be examined. Chapter 4 will introduce the proposed Bayesian MCMC methodology and apply it to three data sets with normally distributed errors. Chapter 5 will introduce a skew-normal distribution, adapt the Bayesian MCMC approach to estimate the model, and apply it to the bipolar data. Chapter 6 assesses the impact of different sample sizes in estimating treatment effects for a

MELS model. Chapter 7 concludes the dissertation with a summary of my results and discusses future applications of the MELS model.

2.0 MIXED-EFFECTS LOCATION-SCALE MODEL

2.1 BACKGROUND

Longitudinal data (repeated observations on the same subject over time) is a classic example of dependent data. Extensive research has been conducted that has accounted for this dependency by explicitly modeling the covariance structure with certain forms, such as Toeplitz or autoregressive. The first authors to model this dependence using a mixed-effects location model were Laird and Ware (1982). This model (2.1) was an improvement over previous ones because it allowed for unbalanced data, e.g. repeat observations on patients with varying time intervals between them. This model also allowed for the explicit modeling and analysis of both the between- and within-subject variability. Mixed-effects models account for the dependence in the data by allowing for the mean structure to vary randomly, while assuming constant within-subject (WS) variability over time. The model

$$\boldsymbol{y_i} = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{Z_i}\boldsymbol{b_i} + \boldsymbol{e_i} \tag{2.1}$$

is a general case of a mixed-effects location model for an individual i with n_i observations. $\boldsymbol{y_i}$ is a $n_i \times 1$ vector with a fixed effects design matrix $\boldsymbol{X_i}$ (a $n_i \times p$ matrix). $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects parameters. $\boldsymbol{Z_i}$ is a $n_i \times q$ random effects design matrix and $\boldsymbol{b_i}$ is a $q \times 1$ vector of subject-specific random effects parameters. The random effects represent the between-subject variability and are often modeled as $\boldsymbol{b_i} \sim N(\boldsymbol{0}, \boldsymbol{D})$ and the within-subject errors assumed independent of the random effects are modeled as $\boldsymbol{e_i} \sim \text{N}(\boldsymbol{0}, \boldsymbol{R_i})$. Here $\boldsymbol{D}$ and $\boldsymbol{R_i}$ are the covariance matrices for the between- and within-subject random effects, respectively.

In these models the within-subject errors are often assumed to follow a distribution with con-

stant variance. However, recent literature has moved away from this assumption and allows for this variability to arise from a random process. One of the earliest known works to allow for a random process governing the within-subject variability dates back to Lindley (1971). In this work variances among different groups were allowed to differ and the model was estimated using a Bayesian approach where the variances followed a chi-squared distribution. The use of covariates in explaining variability in the presence of these random processes was used by Cox and Solomon (1986) in a blood pressure study. In this study, an overdispersion model (2.2) is presented to allow for different group variances as follows

$$Var(Y_{is}) = \sigma_i^2 \tag{2.2}$$

where i is the group and s is the observation within the ith group. Here the variances are independent unobserved values of a random variable T with a probability density function $h(t)$. Johnson (1997) introduced the concept of a random variance to multirater ordinal data which presented a method for measuring student performance that accounted for both the achievement level of students within a class and instructor-specific grade cutoffs.

These works have been extended to a more general class of models where both the between- and the within-subject variability are allowed to vary randomly. James et al. (1994) modeled the random within-subject variation in a nonlinear random effects model using an inverse gamma distribution for mitochondrial enzyme kinetic data. Chinchilli et al. (1995) used a mixed-effects model with within-subject variances derived from an inverse gamma distribution for serum cholesterol data. Lin et al. (1997) allowed for covariates to influence the within-subject variability in a mixed-effects model with variances from an inverse gamma distribution for menstrual diary data. Scarpa et al. (2008) extended the use of the previous models to economics in modeling the willingness of customers to pay for various vacation locations in the Alps. A Gumbel distribution is used to model the within-subject variability in this paper.

Over the last decade work in models with random scale effects has attempted to generalize their use into other models. Cleveland et al. (2000) developed a sequence of steps to fit a general class of models with normal errors through the use of unit regression. This work was extended upon by Shu (2008) who developed a stepwise model building process for mixed-effects models with random scale effects. This work allowed for different distributions to model the random effects

such as a t-distribution for the random location effects and inverse gamma and log normal for the random scale effects.

The most recent work into modeling mixed-effects models with random scale effects is by Hedeker. Hedeker proposes a mixed-effects location-scale model that allows for the inclusion of covariates in modeling the within-subject (WS) variability in addition to random scale effects. He has extended the model to allow for the random location and random scale effects to be correlated (Hedeker et al., 2008). Hedeker's mixed-effects location-scale (MELS) model takes the form for a measurement y of subject i ($i = 1, 2, \ldots, N$ subjects) on occasion j ($j = 1, 2, \ldots, n_i$):

$$y_{ij} = \boldsymbol{x_{ij}'}\boldsymbol{\beta} + \nu_i + \epsilon_{ij} \tag{2.3}$$

$\boldsymbol{x_{ij}}$ is a p x 1 vector of fixed-effect predictors and $\boldsymbol{\beta}$ is the corresponding p x 1 vector of fixed-effect coefficients. ν_i represents the random subject effect on the mean structure. Here the within-subject errors ϵ_{ij} are assumed to follow a normal distribution with mean 0 and variance $\sigma^2_{\epsilon_{ij}}$. To allow for covariates to influence the WS variability and random scale effects the variance is modeled using a log-linear form as:

$$\sigma^2_{\epsilon_{ij}} = \exp(\boldsymbol{w_{ij}'}\boldsymbol{\gamma} + \omega_i) \tag{2.4}$$

where $\boldsymbol{w_{ij}}$ is a vector of covariates and $\boldsymbol{\gamma}$ is a vector of coefficients affecting the WS variability. ω_i represents the random scale effect that allows for the subject-specific WS variability. The random location and scale effects can be modeled as:

$$\begin{pmatrix} \nu_i \\ \omega_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\nu_i} & \sigma_{\nu\omega} \\ \sigma_{\nu\omega} & \sigma^2_{\omega} \end{pmatrix} \right] \tag{2.5}$$

$$\sigma^2_{\nu_i} = \exp(\boldsymbol{u_i'}\boldsymbol{\tau}) \tag{2.6}$$

The random location and scale effects are allowed to be correlated with covariance $\sigma_{\gamma\omega}$. The between-subject variability can be modeled with influence by covariates $\boldsymbol{u_i'}$ as well. $\boldsymbol{\tau}$ is a vector of coefficients affecting the between-subject variability.

Further work by Hedeker has extended these models to account for ordinal data (Hedeker et al., 2006) (Hedeker et al., 2009) as well as a three-level MELS model that allows for the possibility of

systematic day-to-day variation (Li and Hedeker, 2012). Pugach et al. (2014) developed a bivariate MELS model to allow for the modeling of two continuous outcomes jointly, which was extended to a general class of multivariate longitudinal outcomes by Kapur et al. (2015).

2.2 CURRENT METHODS OF MODEL ESTIMATION

Presently, there are two primary methods of estimating mixed-effects model with random scale effects. One method based on unit regression was proposed by Cleveland et al. (2000) and expanded upon by Shu (2008). The other method used by Hedeker et al. (2008) is a maximum marginal likelihood (MML) estimation. The unit regression approach is a general top-down method that builds upon successive residual analyses and is most useful for models where the focus is solely on the estimation of location parameters. The MML procedure fits a specified parameterized model and allows for the WS variability to be modeled as dependent on specified covariates which at present Cleveland's approach does not allow for. A drawback to MML is that the procedure relies on an integral approximation and an optimization approach that in several cases may fail to converge or provide sub-optimal estimation results.

2.2.1 Unit Regression

The unit regression procedure (which refers to scaling the residual terms to obtain an error variance of 1) consists of a stepwise procedure that starts by fitting a fixed effects model then examines residuals to determine if any underlying structure remains. The mixed-effects model with random scale effects takes the form of

$$y_{ij}^{*} = \boldsymbol{x_{ij}}\boldsymbol{\beta} + \nu_i + \gamma_i \frac{\zeta_{ij}}{\sigma(\epsilon)} \tag{2.7}$$

where β are the fixed-effect parameters, ν_i are the random location parameters, γ_i^2 are the random scale effects with a specified distribution (e.g. inverse gamma) and $E(\gamma_i^2) = 1$, and ζ_{ij} is the error term assumed to be $i.i.d\ N(0, 1)$.

2.2.1.1 Fixed Effect Estimation

The first step is to estimate the fixed effect parameters using ordinary least squares regression (OLS) for the specified model. An example of such model would be:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 \tau_i + \beta_3 t_{ij} \times \tau_i + \epsilon_{ij} \tag{2.8}$$

The model in 2.8 takes the form that typical psychiatric data might follow where y_{ij} is the response variable (e.g. measure of a patient's mood) for the ith subject ($i = 1, 2, \ldots, N$) at the jth time point ($j = 1, 2, \ldots, n_i$). t_{ij} is the time point that may vary for each subject and τ_i is an indicator variable that is 1 for the new treatment and 0 for the control treatment. The fixed effects parameters β_0, β_1, β_2, and β_3 will be estimated using OLS.

2.2.1.2 Adjust for Fixed Scale Effects

Let r^*_{ij} be the residuals from the fitted model in 2.8.

$$r^*_{ij} = y_{ij} - \hat{\beta}_0 - \hat{\beta}_1 t_{ij} - \hat{\beta}_2 \tau_i - \hat{\beta}_3 t_{ij} \times \tau_i \tag{2.9}$$

Check the box plots for residuals against the fixed effects t_{ij} and τ_i. If a relation is found for say the treatment effect τ_i, then use a weighted least square estimate approach for the regression model

$$r^*_{ij} = \beta_{\tau_i} + e_{ij} \tag{2.10}$$

assuming constant variance within treatment groups to estimate $\hat{\sigma}^2_0(\epsilon)$ for the group with the control treatment and $\hat{\sigma}^2_1(\epsilon)$ for the group with the new treatment.

Define:

$$r_{ij} = \frac{r^*_{ij}}{\hat{\sigma}_{\tau_i}(\epsilon)}, \; x_{ij} = \frac{1}{\hat{\sigma}_{\tau_i}(\epsilon)} \tag{2.11}$$

This adjustment allows for the error distribution to have variance 1.

2.2.1.3 Unit Regression and Error Distribution Check

Using the notation defined in 2.11 to obtain N unit regressions:

$$r_{ij} = \nu_i x_{ij} + \tau_{ij} = \nu_i x_{ij} + \gamma_i \zeta_{ij} \tag{2.12}$$

where

$$\tau_{ij} = \gamma_i \zeta_{ij} \tag{2.13}$$

By assuming that the random scale effect γ_i satisfies $E(\gamma_i^2) = 1$, this allows:

$$E(\tau_{ij}) = 0, \quad Var(\tau_{ij}) = 1 \tag{2.14}$$

Using the assumptions in 2.14 we can fit 2.12 to obtain:

$$\hat{\tau}_{ij} = r_{ij} - \hat{\nu}_i x_{ij} = \gamma_i \hat{\zeta}_{ij} \tag{2.15}$$

The standardized residuals are:

$$\hat{\psi}_{ij} = \frac{\hat{\tau}_{ij}}{\sqrt{Var(\hat{\zeta}_{ij})}} = \gamma_i \frac{\hat{\zeta}_{ij}}{\sqrt{Var(\hat{\zeta}_{ij})}} \tag{2.16}$$

The residual variance is:

$$s_i^2 = \frac{\sum_{j=1}^{n_i} \tau_{ij}^2}{n_i - 1} = \gamma_i^2 \frac{\hat{\zeta}_{ij}^2}{n_i - 1} \tag{2.17}$$

The studentized residual is:

$$\hat{\phi}_{ij} = \frac{\hat{\psi}_{ij}}{s_i} \tag{2.18}$$

Using the standardized residuals we can check the normality distribution assumption of the error distribution. If it is not validated a new distribution assumption (such as a t-distribution) will be used and the process will start over with the new distribution.

2.2.1.4 Check for Presence of Random Scale Effects

Based on the error distribution found in 2.2.1.3 we can check if random scale effects are present:

$$H_0 : \gamma_i^2 = 1, \; i = 1, \ldots, N \tag{2.19}$$

If the null hypothesis is rejected, then random scale effects are present and the distribution will be determined. (See Cleveland et al. (2000) and Shu (2008) for further information). Based on

simulation results Shu (2008) found that when random scale effects are present, but ignored in the model the bias of the location estimates $\hat{\beta}$ are not affected; however, the efficiency is reduced.

2.2.2 Marginalized Maximum Likelihood

The estimation procedure used by Hedeker et al. (2008) is based on a variant of maximum likelihood proposed by Bock (1989) referred to as maximum marginal likelihood (MML). The unit regression estimation technique uses a top-down estimation approach, whereas the MML estimation begins by specifying the full model and then uses appropriate software (SAS PROC NLMIXED) to fit the model. Unlike the unit regression approach which has only been used for mixed-effects models with random scale effects such as those specified in equation 2.7, the MML technique allows for the estimation of the MELS model proposed by Hedeker (given in equations 2.20 and 2.21).

One of my contributions to the work performed on mixed-effects model with random scale effects will be to provide an alternate estimation approach to MML. Hedeker's MELS model contains the flexibility to model the WS variability using covariates that Cleveland's model does not. This allows for research questions that involve assessing the effects of covariates on both the mean and variance to be addressed with a single model. Due to this advantage my work will directly expand upon those of Hedeker and his colleagues. I will compare my model with Hedeker's approach.

The MML estimation technique uses the distributional assumptions of the random effects to integrate out the individual-specific attributes in order to obtain the marginal distribution of the response vector $\boldsymbol{y_i}$. The following model as detailed by Hedeker et al. (2008) takes the form:

$$y_{ij} = \boldsymbol{x}_{\boldsymbol{ij}}^{'}\boldsymbol{\beta} + \nu_i + \epsilon_{ij} \tag{2.20}$$

where

$$
\begin{aligned}
\epsilon_{ij} &\sim N(0, \sigma^2_{\epsilon_{ij}}) \\
\sigma^2_{\epsilon_{ij}} &= \exp(\boldsymbol{w}'_{\boldsymbol{ij}}\boldsymbol{\gamma} + \omega_i) \\
\begin{pmatrix} \nu_i \\ \omega_i \end{pmatrix} &\sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\nu_i} & \sigma_{\nu\omega} \\ \sigma_{\nu\omega} & \sigma^2_{\omega} \end{pmatrix} \right] \\
\sigma^2_{\nu_i} &= \exp(\boldsymbol{u}'_{\boldsymbol{i}}\boldsymbol{\tau})
\end{aligned}
\tag{2.21}
$$

In order to aid in estimation, the random effects will be expressed in standardized form using a Cholesky factorization (Bock, 1975):

$$
\begin{bmatrix} \nu_i \\ \omega_i \end{bmatrix} = \begin{bmatrix} s_{1i} & 0 \\ s_{2i} & s_{3i} \end{bmatrix} \begin{bmatrix} \theta_{1i} \\ \theta_{2i} \end{bmatrix} = \begin{bmatrix} \sigma_{\nu_i} & 0 \\ \sigma_{\nu\omega}/\sigma_{\nu_i} & \sqrt{\sigma^2_{\omega} - \sigma^2_{\nu\omega}/\sigma^2_{\nu_i}} \end{bmatrix} \begin{bmatrix} \theta_{1i} \\ \theta_{2i} \end{bmatrix} \tag{2.22}
$$

Using the Cholesky factorizaton we can rewrite 2.20 as

$$
y_{ij} = \boldsymbol{x}'_{\boldsymbol{ij}}\boldsymbol{\beta} + s_{1i}\theta_{1i} + \epsilon_{ij} \tag{2.23}
$$

where $s_{1i} = \sigma_{\nu_i} = (\exp(\boldsymbol{u}'_{\boldsymbol{i}}\boldsymbol{\tau}))^{\frac{1}{2}}$ and the WS and marginal variances are represented by

$$
\sigma^2_{\epsilon_{ij}} = \exp(\boldsymbol{w}'_{\boldsymbol{ij}}\boldsymbol{\gamma} + s_{2i}\theta_{1i} + s_{3i}\theta_{2i}) \tag{2.24}
$$

$$
V(y_{ij}) = \exp(\boldsymbol{u}'_{\boldsymbol{i}}\boldsymbol{\tau} + \exp(\boldsymbol{w}'_{\boldsymbol{ij}}\boldsymbol{\gamma} + \frac{1}{2}\sigma^2_{\omega}) \tag{2.25}
$$

Written in vector notation the model to be estimated for subject i is

$$
\boldsymbol{y_i} = \boldsymbol{X_i}\boldsymbol{\beta} + \mathbf{1_i}s_{1i}\theta_{1i} + exp\Big\{\frac{1}{2}(\boldsymbol{W_i}\boldsymbol{\gamma} + \mathbf{1_i}s_{2i}\theta_{1i} + \mathbf{1_i}s_{3i}\theta_{2i})\Big\} \tag{2.26}
$$

where $\boldsymbol{X_i}$ is the location design matrix, $\boldsymbol{W_i}$ is the scale design matrix and the standardized random effects θ_{1i} and θ_{2i} are standard normal and independent of each other. Using a conditional likelihood of the response variable $\boldsymbol{y_i}$ on its respective random effects, the marginal likelihood becomes

$$
h(\boldsymbol{y_i}) = \int_{\Theta} f(\boldsymbol{y_i}|\boldsymbol{\theta_i})g(\boldsymbol{\theta_i})d\boldsymbol{\theta_i} \tag{2.27}
$$

$f(\boldsymbol{y_i}|\boldsymbol{\theta_i})$ is the normal distribution of $\boldsymbol{y_i}$ and $g(\boldsymbol{\theta})$ is a standard bivariate normal density resulting from the standardization performed by the Cholesky decomposition. The log-likelihood to be maximized becomes

$$\log L = \sum_{i=1}^{n} \log h(\boldsymbol{y_i}) \tag{2.28}$$

The integral in 2.27 does not have an analytical expression so Gaussian quadrature will be used to approximate the integral. After an appropriate approximation is found an iterative optimization approach, such as Newton-Raphson and Trust Region, will be used to maximize the likelihood function in 2.28.

2.2.2.1 Estimation Procedure Due to the maximization of a nonlinear model over several dimensions, a detailed estimation approach must be used. This will maximize the chance of achieving the global maximum for the likelihood functions. Other possible problems that arise in fitting this model come from convergence issues with respect to approximating the integral in 2.27 and optimizing the equation in 2.28. These convergence difficulties may arise due to the lack of repeated observations.

Hedeker proposes the following steps to fit a MELS model using MML.

1. Fit the model in 2.20 as a mixed-effects location model with constant between- (BS) and within-subject (WS) variances.
2. Specify the model to be approximated in SAS PROC NLMIXED
3. Use the parameter estimates from step 1 as initial estimates for step 2. The log of the BS and WS variance estimates are used as the initial value of the intercept parameters for the fully-parameterized variance structures. For additional variance parameters the initial parameter estimate will be set as 0.
4. Run the model in SAS PROC NLMIXED using different optimization techniques (e.g. Quasi-Newton, Trust Region, Newton-Raphson, Newton-Raphson w/ Ridging, Nelder-Mead, Double - Dogleg, and Conjugate Gradient).

2.2.2.2 Simulation Results In order to assess the accuracy of the steps taken in 2.2.2.1, I conducted various simulations on data sets modeled to resemble a randomized treatment. In these simulations subjects are divided equally into a control and a treatment group, baseline measurements are recorded and follow up visits occur every half week. Half week follow up visits are used in order to closely match Hedeker's original model use. His original use involved observations all recorded within one week, which would be difficult and in most cases non-informative for randomized clinical trials. A half week interval provides a suitable balance (for modeling purposes) in order to extend Hedeker's original model use to clinical trials without seriously deviating from the original real world use case. There will be one response variable and two independent variables (week and treatment). The independent variables will impact the variance structure in select cases. The subjects will have both individual specific means and WS variances (i.e. random location and scale effects). The parameter values were chosen from fixed-effects models and descriptive statistics used to analyze the Hamilton Rating Scale for Depression (discussed in Chapter 3). Simulations were done using SAS/IML.

For the ith subject ($i = 1, \ldots, N$) at the jth time point ($j = 1, \ldots, n$) the first model considered is

$$y_{ij} = \beta_0 + \beta_1 trt_i + \beta_2 week_{ij} + \nu_i + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma^2_\epsilon)$$

$$\begin{pmatrix} \nu_i \\ \omega_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\nu_i} & \sigma_{\nu\omega} \\ \sigma_{\nu\omega} & \sigma^2_\omega \end{pmatrix} \right]$$

$$\sigma^2_{\nu_i} = \exp(\tau_0 + \tau_1 trt_i)$$

$$\sigma^2_\epsilon = \exp(\gamma_0 + \gamma_1 trt_i + \omega_i)$$

Using the built-in normal likelihood function for a simulation with 100 subjects and 33 observations per subject only resulted in convergence rates between 14% and 74% based on 50 simulations as shown in Table 1.

The optimization procedure found to provide the best fit was Newton-Raphson; however, this only shows that this approach converges most frequently not whether the estimates are solutions from the global maxima rather than a local one. As seen in Table 1 none of the optimization techniques converge 100% of the time. Hedeker and Pugach are able to obtain reliable convergence

Table 1: % Convergence based on 50 Simulations for different optimization techniques

Optimization Technique	% Converged
Quasi-Newton	54
Trust Region	64
Newton-Raphson	74
Nelder-Mead	68
Double-Dogleg	58
Conjugate Gradient	14

by using a user-specified likelihood function rather than the built-in version from SAS. By using a user-specified likelihood I am able to increase the convergence rate to 100% for as many as 1,000 simulated data sets using the Newton-Raphson approach.

To assess the accuracy of this approach in capturing the global maxima four different scenarios were run (results found in Tables 2 and 3) (Note: All analyses were done on a Samsung NP305E5A laptop).

1. 50 subjects, 17 observations per subject
2. 50 subjects, 33 observations per subject
3. 100 subjects, 17 observations per subject
4. 100 subjects, 33 observations per subject

Looking at the estimates and their biases reveals that the MML procedure is an efficient estimation method for the proposed model. The largest standardized bias in each scenario is for the estimate of the random scale variance σ^2_ω; however, this decreases by over 10% with an increase in the number of observations per subject for both 50 and 100 subjects. Looking at Table 3 reveals that the model has low power in detecting treatment effects in the variance at values close to 0. The power does increase as the number of subjects and/or observations per subject increase. Chapter 6 will look at the issue of power for a MELS model in greater detail (For SAS Code related to the fitting of the model see APPENDIX A).

In clinical data (such as the aforementioned bipolar data) a clinically significant question of interest would involve the effect different treatments have over time. To address this question

a treatment by time interaction can be added to the mean structure. Through the use of a MELS model this interaction can be added to the structure of the within-subject variance as well. A model that would address this question would take the form of

$$
\begin{aligned}
y_{ij} &= \beta_0 + \beta_1 trt_i + \beta_2 week_{ij} + \beta_3 trt_i \times week_{ij} + \nu_i + \epsilon_{ij} \\
\epsilon_{ij} &\sim N(0, \sigma^2_\epsilon) \\
\begin{pmatrix} \nu_i \\ \omega_i \end{pmatrix} &\sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\nu_i} & \sigma_{\nu\omega} \\ \sigma_{\nu\omega} & \sigma^2_\omega \end{pmatrix} \right] \\
\sigma^2_{\nu_i} &= \exp(\tau_0 + \tau_1 trt_i) \\
\sigma^2_\epsilon &= \exp(\gamma_0 + \gamma_1 trt_i + \gamma_2 week_{ij} + \gamma_3 trt_i \times week_{ij} + \omega_i)
\end{aligned}
\tag{2.29}
$$

With the addition of the time component to the within-subject variance and using the same procedure for the first specified model, out of 50 simulated data sets there were no feasible results, either from convergence issues or the inability to approximate the integral.

In some of Hedeker and his colleague's papers (Hedeker et al. (2008) and Pugach et al. (2014)) a continuous covariate has been used in the log linear model of the WS variance; however, due to the complexities involved in optimizing a model over several parameters the nature of the data plays a critical component in obtaining reasonable estimates. Further research could be performed into obtaining suitable optimization results using Hedeker's proposed MML technique. These could include reparameterizing the model and adapting an optimization technique that is more suitable for the nonlinear nature of the MELS models.

In order to move away from the use of optimization and the step-wise procedures that both Cleveland and Hedeker use in their model fitting, I propose the use of a Bayesian MCMC approach in estimating the model (see Chapter 4). This technique often referred to as one of "last resort" is more adept at estimating non-linear hierarchical models than frequentist methods (Gelman et al., 2014a). A comparison between Hedeker's frequentist and my Bayesian approach will be compared using an EMA and a psychiatric longitudinal data set found in Hedeker and Nordgren (2013).

Table 2: Simulation Results based on the User Specified Log Likelihood Estimation (Optimized with Newton-Raphson)

		50 Subjects, 17 observations per subject					50 Subjects, 33 observations per subject				
Parameters	True Value	EST	SE	BIAS	ST BIAS	RMSE	EST	SE	BIAS	ST BIAS	RMSE
β_0	6.9	6.908	.200	.008	4.21	.200	6.905	.198	.005	2.59	.199
β_1	.4	.396	.254	-.004	-1.76	.254	.398	.239	-.002	-.87	.239
β_2	-.2	-.200	.013	.000	3.49	.013	-.200	.005	.000	-9.26	.005
τ_0	-.1	-.197	.316	-.097	-30.69	.330	-.170	.298	-.070	-23.62	.306
τ_1	-.4	-.393	.427	.007	1.65	.427	-.412	.394	-.012	-3.09	.394
γ_0	-.2	-.197	.130	.003	2.44	.130	-.200	.115	.000	.42	.115
γ_1	.1	.101	.185	.001	.60	.185	.106	.171	.006	3.71	.171
$\sigma_{\nu\omega}$	.15	.142	.087	-.008	-8.94	.087	.145	.074	-.005	-7.09	.074
σ^2_ω	.3	.264	.084	-.036	-43.08	.092	.280	.071	-.020	-28.34	.073
# of simulations		1000					1000				
# of simulations converged		995					998				
# of minutes to converge		41.52					71.97				
		100 Subjects, 17 observations per subject					100 Subjects, 33 observations per subject				
Parameters	True Value	EST	SE	BIAS	ST BIAS	RMSE	EST	SE	BIAS	ST BIAS	RMSE
β_0	6.9	6.899	.143	-.001	-1	.143	6.892	.140	-.008	-5.38	.140
β_1	.4	.399	.183	-.001	-.42	.183	.405	.177	.005	2.94	.177
β_2	-.2	-.2	.009	.000	2.68	.009	-.200	.003	.000	-.34	.003
τ_0	-.1	-.146	.222	-.046	-20.66	.227	-.151	.213	-.051	-23.94	.219
τ_1	-.4	-.4	.303	.000	.07	.303	-.398	.291	.002	.82	.291
γ_0	-.2	-.199	.092	.001	.99	.092	-.206	.086	-.006	-6.38	.086
γ_1	.1	.098	.132	-.002	-1.34	.132	.105	.121	.005	4.11	.121
$\sigma_{\nu\omega}$	.15	.148	.061	-.002	-3.18	.061	.144	.053	-.006	-10.92	.053
σ^2_ω	.3	.276	.06	-.024	-39.95	.065	.286	.050	-.014	-28.85	.052
# of simulations		1000					1000				
# of simulations converged		1000					1000				
# of minutes to converge		74.23					155.18				

Table 3: Simulation Results based on Log Likelihood Estimation (Optimized with Newton-Raphson)

		50 Subjects, 17 observations per subject			50 Subjects, 33 observations per subject		
Parameters	True Value	95% Coverage	Avg Width	Power	95% Coverage	Avg Width	Power
β_0	6.9	.944	.789	1.000	.938	.774	1.000
β_1	.4	.943	.998	.344	.956	.987	.344
β_2	-.2	.954	.051	1.000	.947	.018	1.000
τ_0	-.1	.926	1.180	.114	.943	1.134	.932
τ_1	-.4	.945	1.646	.150	.956	1.553	.178
γ_0	-.2	.941	.503	.322	.950	.469	.388
γ_1	.1	.940	.711	.096	.945	.664	.104
$\sigma_{b\omega}$	.15	.936	.324	.396	.941	.298	.477
σ^2_ω	.3	.873	.315	.987	.897	.276	.995
# of simulations		1000			1000		
# of simulations converged		995			998		
# of minutes to converge		41.52			71.97		
		100 Subjects, 17 observations per subject			100 Subjects, 33 observations per subject		
Parameters	True Value	95% Coverage	Avg Width	Power	95% Coverage	Avg Width	Power
β_0	6.9	.937	.560	1.000	.948	.542	1.000
β_1	.4	.948	.707	.597	.952	.692	.629
β_2	-.2	.945	.035	1.000	.947	.013	1.000
τ_0	-.1	.929	.818	.117	.938	.789	.124
τ_1	-.4	.933	1.137	.285	.940	1.083	.303
γ_0	-.2	.951	.231	.565	.949	.220	.678
γ_1	.1	.956	.506	.128	.948	.468	.140
$\sigma_{b\omega}$	.15	.941	.231	.725	.950	.210	.802
σ^2_ω	.3	.869	.227	1.000	.914	.196	1.000
# of simulations		1000			1000		
# of simulations converged		1000			1000		
# of minutes to converge		74.23			155.18		

3.0 BIPOLAR DATA SET

3.1 APPLICATION TO CLINICAL DATA

Previous work on mixed-effects location-scale models has focused primarily on its application to EMA data sets. Another data collection process that would benefit from the model would be a clinical trial. Using this method researchers are able to design experiments to determine (in most cases) whether a treatment effect exists. With the use of a mixed-effects location-scale model it is now possible to determine such an effect in not only the mean, but the variance as well. A useful case arises in the study of bipolar.

Individuals suffering from bipolar I disorder spend varying amounts of time in one of three states: depression, mania and euthymia (neither depressed nor manic). Mania is characterized as a period with high levels of energy and euphoria. Depression is the longer lasting state and is characterized by a negative outlook (Anderson et al., 2012). Often times in measuring the bipolar characteristics of the individuals a questionnaire will be given which seeks to quantify the patient's mood and severity of their condition. Due to the oscillating nature of the disease the mood measurements may vary widely. Focusing solely on the mean of the response variable only reveals one part of the disorder. By jointly modeling the mean and variance through the use of the MELS model researchers are able to get a better understanding of the treatment effects for these individuals.

3.2 BIPOLAR DATA DESCRIPTION

In a clinical trial conducted by the Department of Psychiatry at the University of Pittsburgh (Frank et al., 2015), bipolar I patients (currently in remission) were investigated to see if a new treatment would aid in lowering the patient's BMI against the status quo. The typical treatment course involves psychiatric treatment by a study psychiatrist and the care received from the patient's primary care physician. As one of the leading causes of disability globally and with over $150 billion in medical costs the researchers, seeking to mitigate this bipolar epidemic, developed the Integrated Risk Reduction Intervention (IRRI) as a way to benefit bipolar patients by affecting their modifiable health risks, such as their sleep/wake cycle, amount of social interaction, nutritional intake, etc. In this study participants were assigned to either the IRRI or the Psychiatric Care with Medical Monitoring (PCMM or status quo). The subjects in both of the groups were monitored by a psychiatrist and provided with assessments and referrals when necessary. The main difference between the groups comes from those assigned with the IRRI receiving a healthy lifestyle behaviors program. This program provided participants with educational sessions designed to promote a healthy lifestyle (Frank et al. (2015) contains additional information on the study's setup).

In order to enter the study individuals had to have had a lifetime diagnosis of bipolar I disorder, which was currently in remission. Remission was determined through the use of the Hamilton Rating Scale for Depression (HRS17TOT), Young Mania Rating Scale (YOUNGTOT) and Clinical Global Impressions for Bipolar Disorder Scale (CGI-BP-S). Individuals were classified as being in remission if for four consecutive weeks before the start of the study their HRS17TOT score was $\leq$ 7, YOUNGTOT score was ≤ 7, and CGI-BP-S score was < 3 concurrently. Since the focus was on lowering an individual's BMI only those with a BMI score ≥ 25 were included in order to deal with those individuals most likely to benefit from such treatment. Participants entered the study between 12 November 2008 and 14 July 2011, with the final participant visit coming on 4 September 2013. 122 individuals were entered into the study (evenly split between the two groups). The study was designed to treat patients for two years with measurements made approximately every two months. As is common with most clinical trials dropouts and missed visits did occur; however, further investigation (not shown) concluded that these were likely missing at random.

3.3 HRS17TOT VARIABLE ANALYSIS

A characteristic of clinical trials is the presence of unbalanced data through dropouts and missed visits. In order to assess the overall shape and distribution the observations will be placed into eight different bins corresponding to the time (measured in days) in which they were measured.

Table 4: Number of Observations per Bin

Bin	Days Since Randomization	# of Observations
1	0 - 100	421
2	101 - 200	298
3	201 - 300	265
4	301 - 400	235
5	401 - 500	219
6	501 - 600	189
7	601 - 700	166
8	701 - 800	64

The first variable analyzed is the Hamilton-Rating Scale for Depression (HRS17TOT) which as previously mentioned was at a value ≤ 7 at the beginning of the study. Looking at Figures 2 and 4 one can see that the threshold value of 7 is exceeded by numerous participants. Furthermore, Figures 6 and 7 shows that over half of the participants exceeded that threshold for almost the entire length of the study. This indicates that over the course of the study many of the individuals became symptomatic.

Looking at Figures 6 and 7 reveals varying amounts of within-subject variability. Figures 8 and 9 show a histogram and box plot respectively which compares the spread of the WS variability among the two treatments. Both the range and interquartile range of variability is larger for the IRRI group than the PCMM which would indicate a less stable treatment. Furthermore a breakdown of variances over time reveals that the IRRI has a consistently higher variance than the PCMM (see Figure 10). The variance measurements in Figure 10 were obtained by taking the average variance of all observations in the time bin. Unlike the results in Figures 6 and 7 these results do not account for subject-specific mean levels. This indicates greater variability among the IRRI group. The drop in variability at the 8th bin corresponds with the low number of observations

at that time.

These figures reveal the extent to which the WS variability differs by both subject and treatment and a MELS model will be used to assess their impact on the mean and variability. In implementing the model an appropriate distribution must be specified. Hedeker's work often assumes a normal distribution and this will be used to model the HRS17TOT variable. As shown in the QQ-plots in Figures 3 and 5 there are some data points in each bin that tail off and deviate from the normal distribution assumption; however, most of the data in each bin follows the normal distribution. The fitting of a MELS model to the Hamilton Rating variable under the normal assumption will be performed in Section 4.4.

Figure 2: Distribution of HRS17TOT scores for Bins 1-4

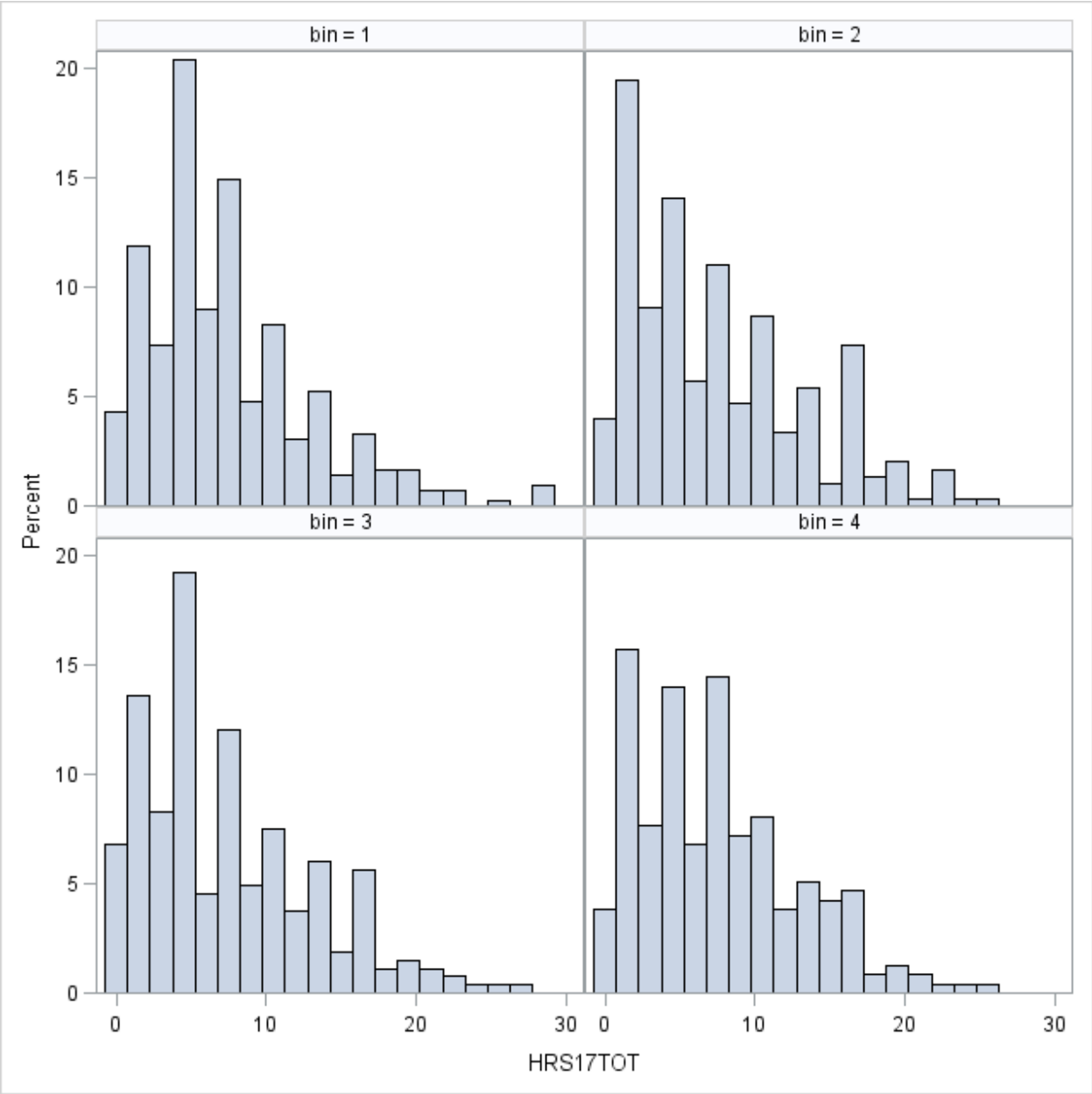
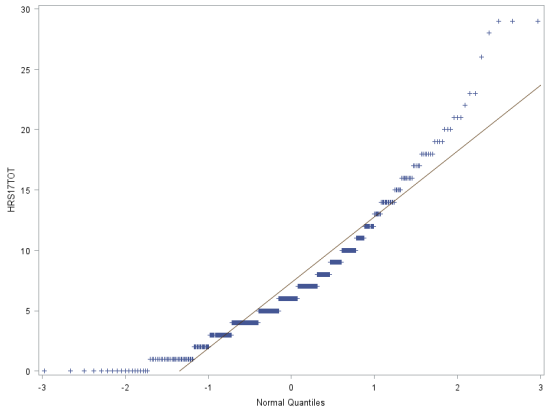
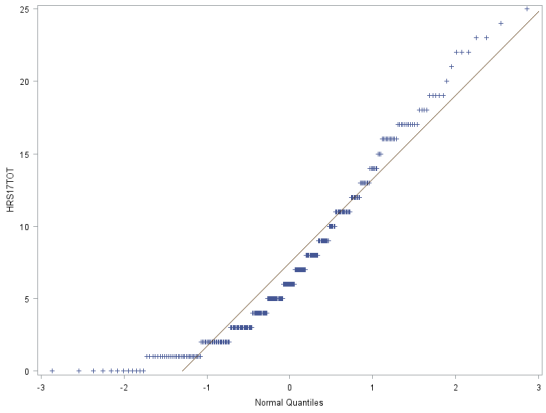


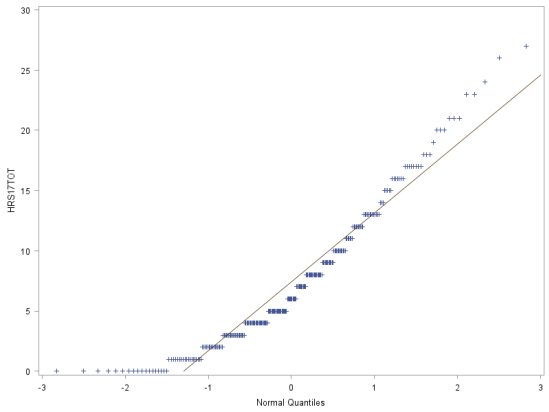
Figure 3: QQ-Plots of HRS17TOT Scores for Bins 1-4



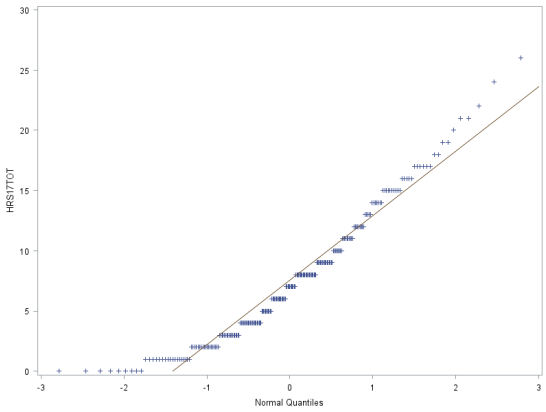
(a) Bin 1



(b) Bin 2



(c) Bin 3



(d) Bin 4

Figure 4: Distribution of HRS17TOT scores for Bins 5-8

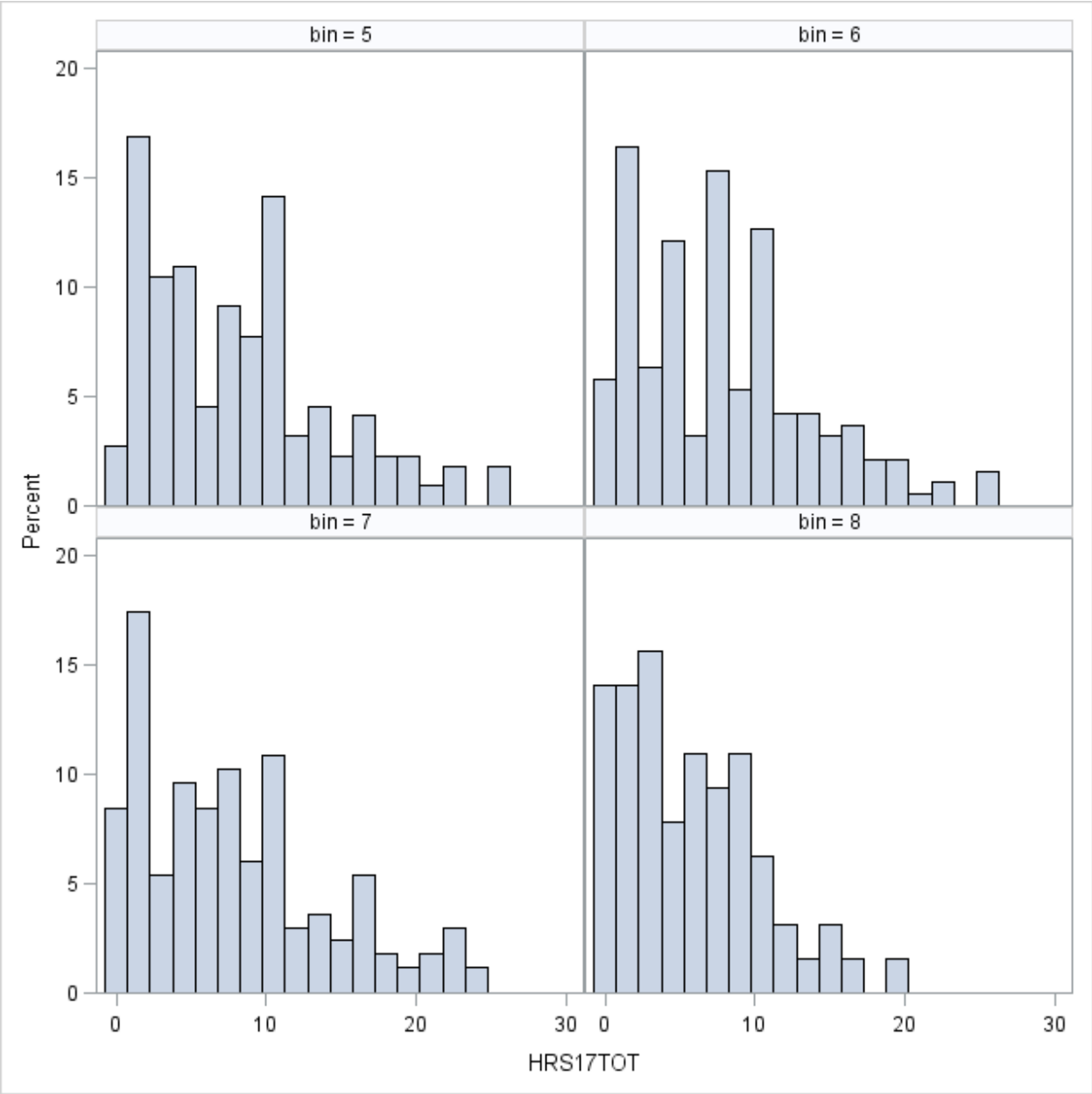
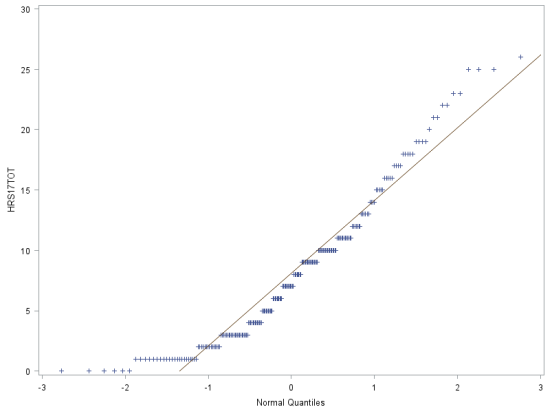
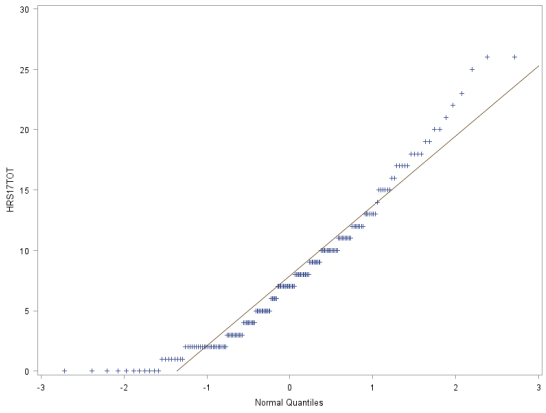


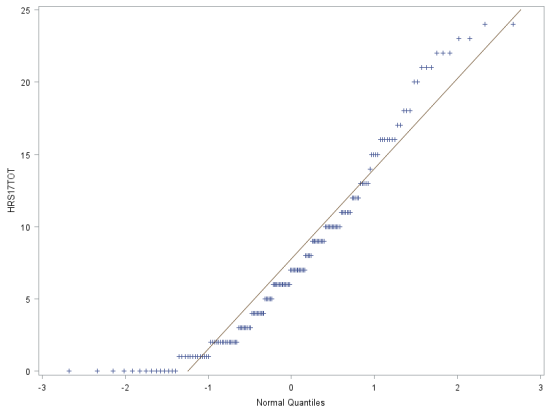
Figure 5: QQ-Plots of HRS17TOT Scores for Bins 5-8



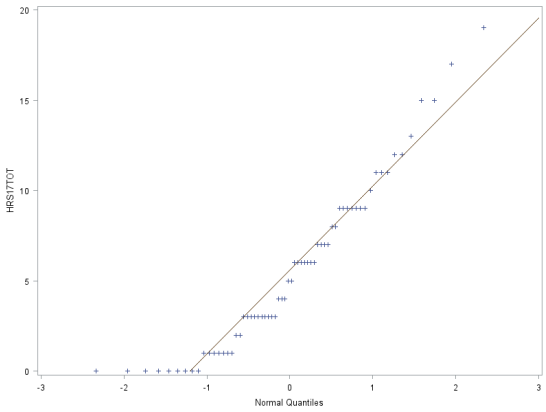
(a) Bin 5



(b) Bin 6

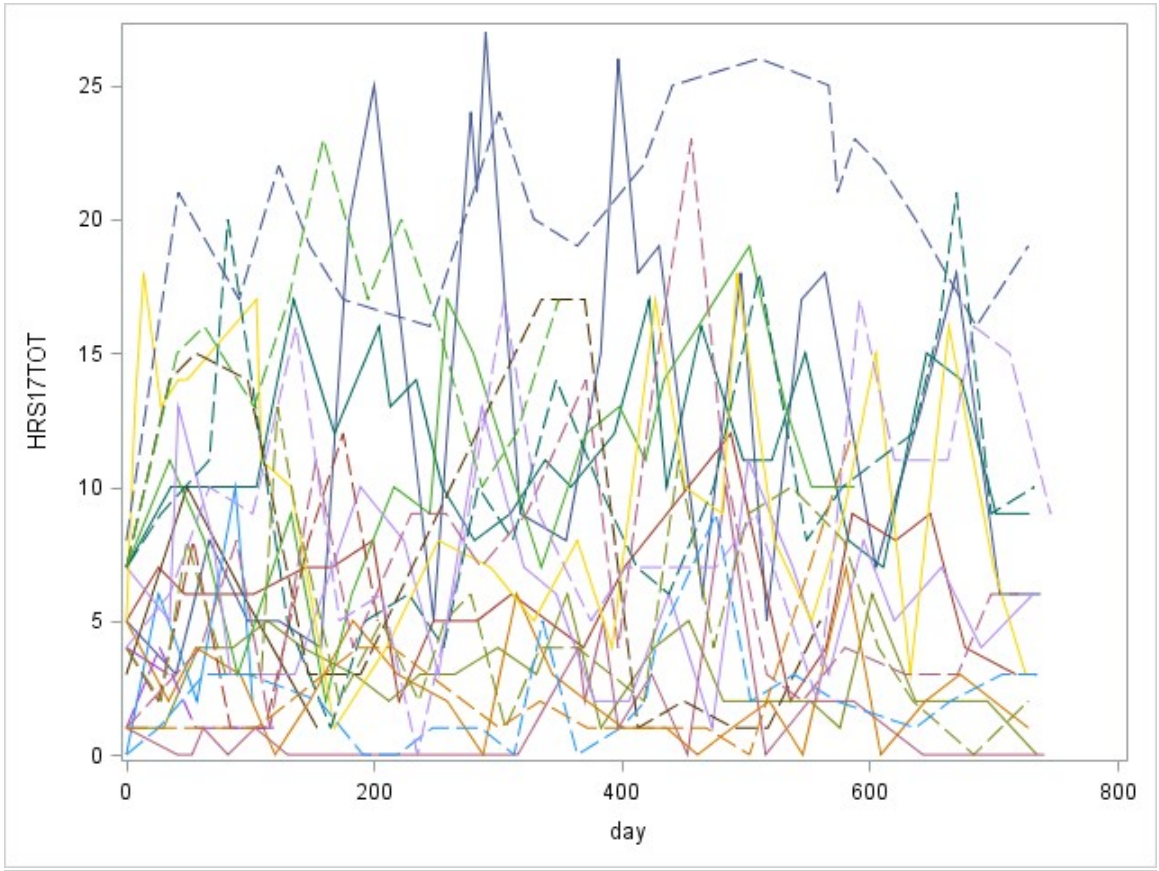


(c) Bin 7



(d) Bin 8

Figure 6: HRS17TOT Scores by Time for each IRRI subject



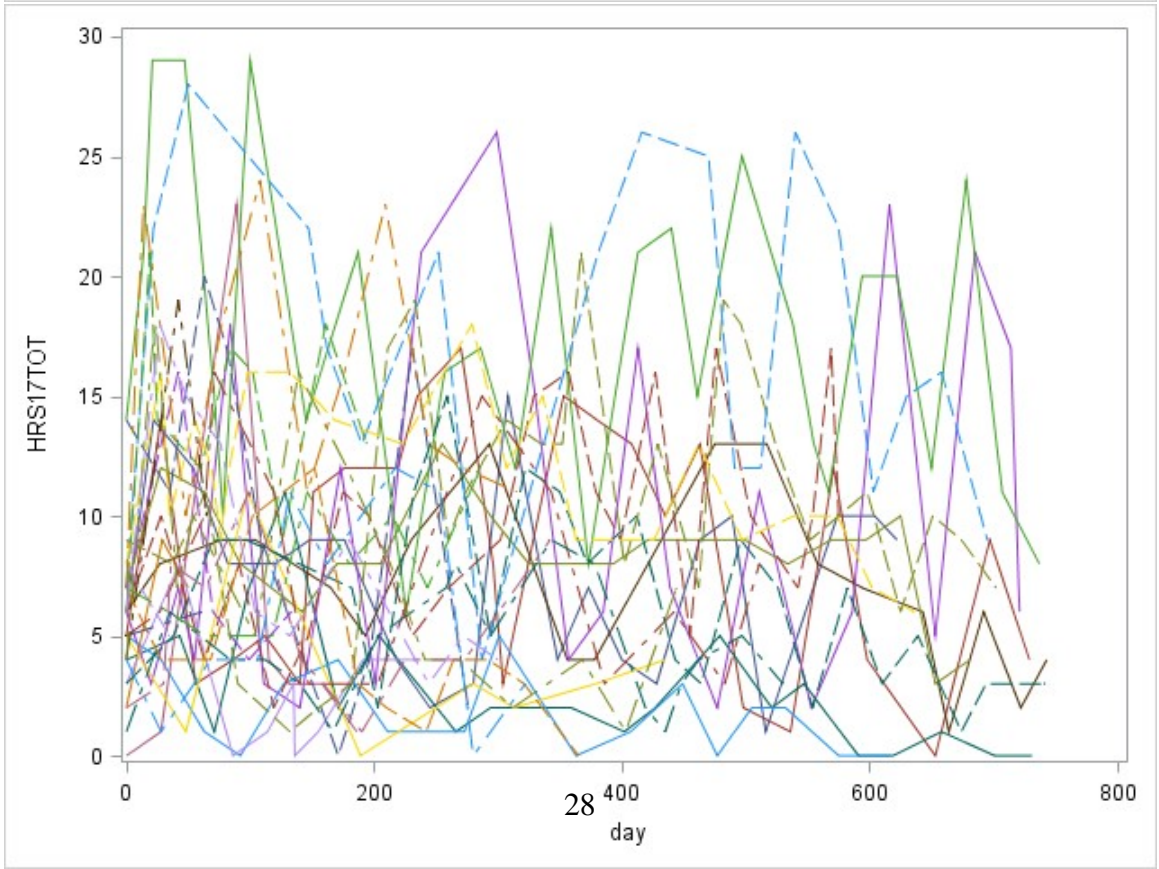
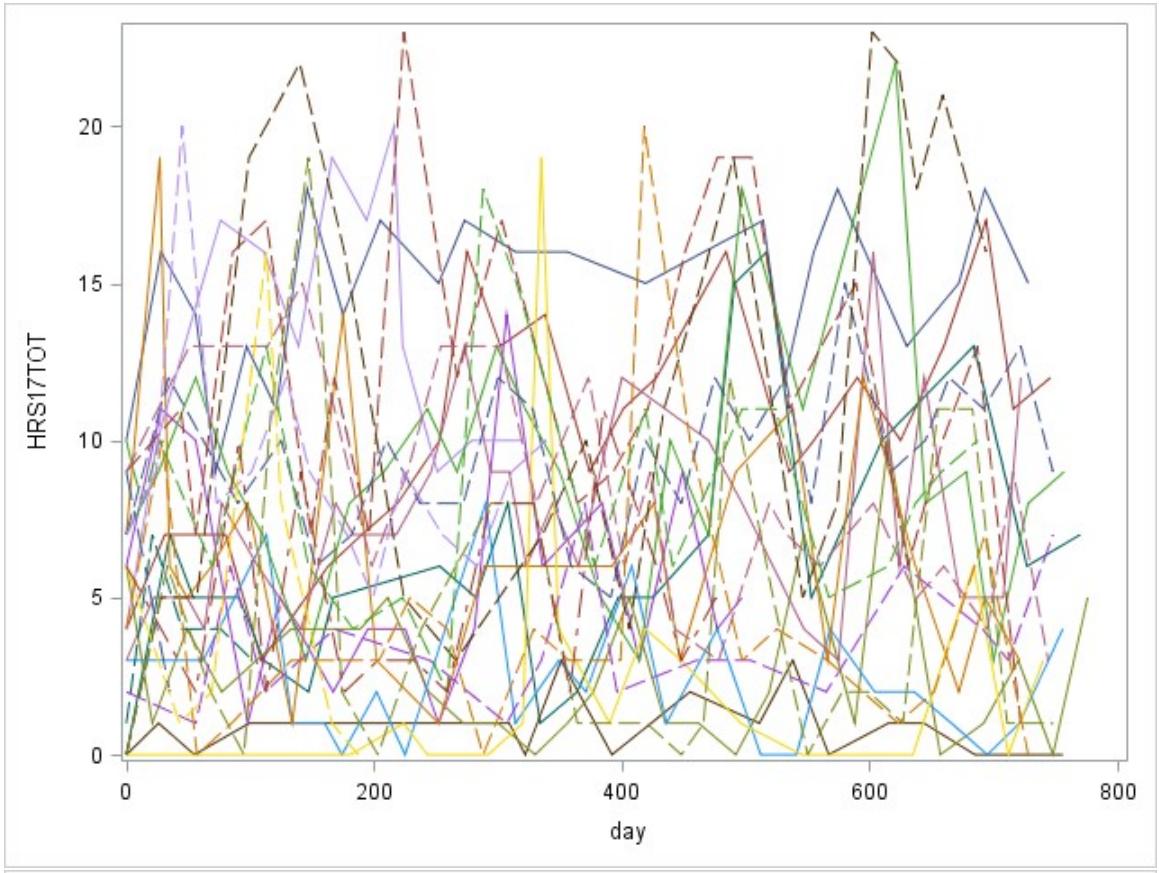


Figure 7: HRS17TOT Scores by Time for each PCMM subject



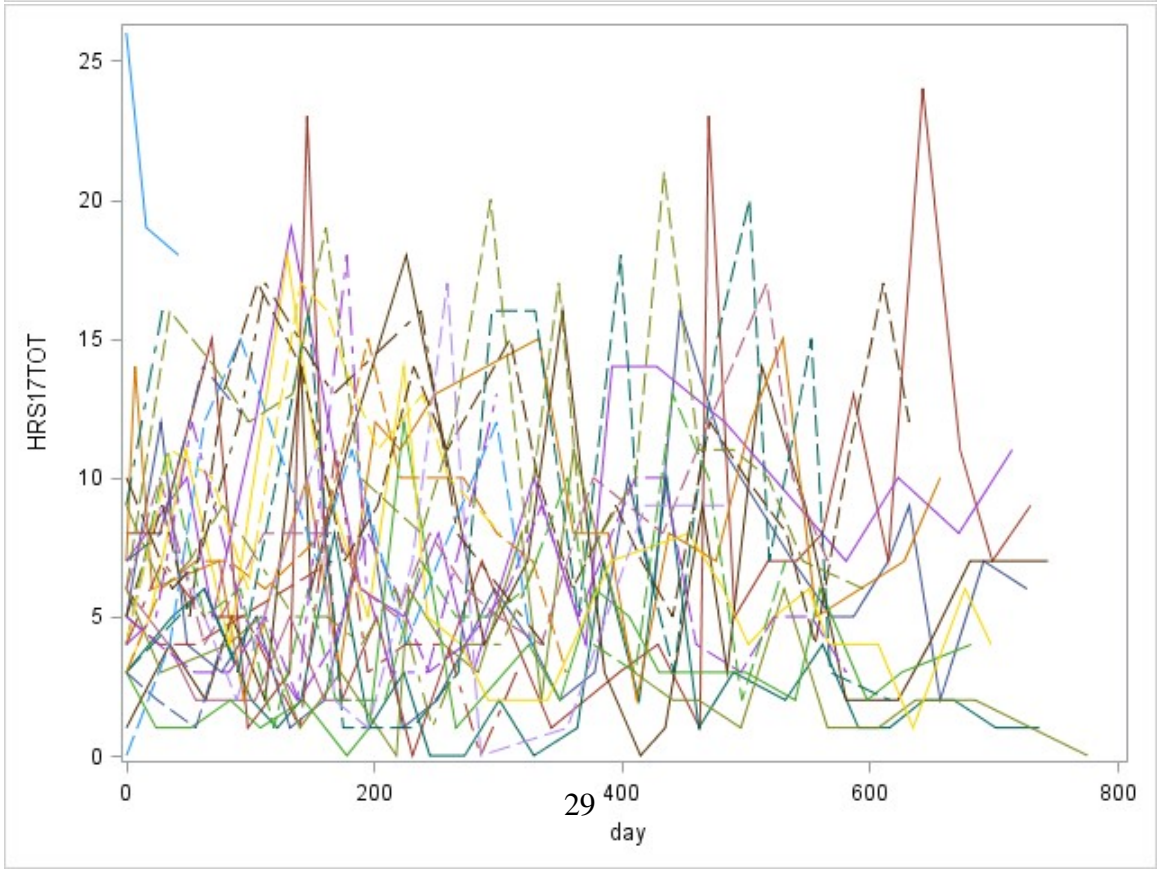


Figure 8: Histograms of Individual Standard Deviations by Treatment

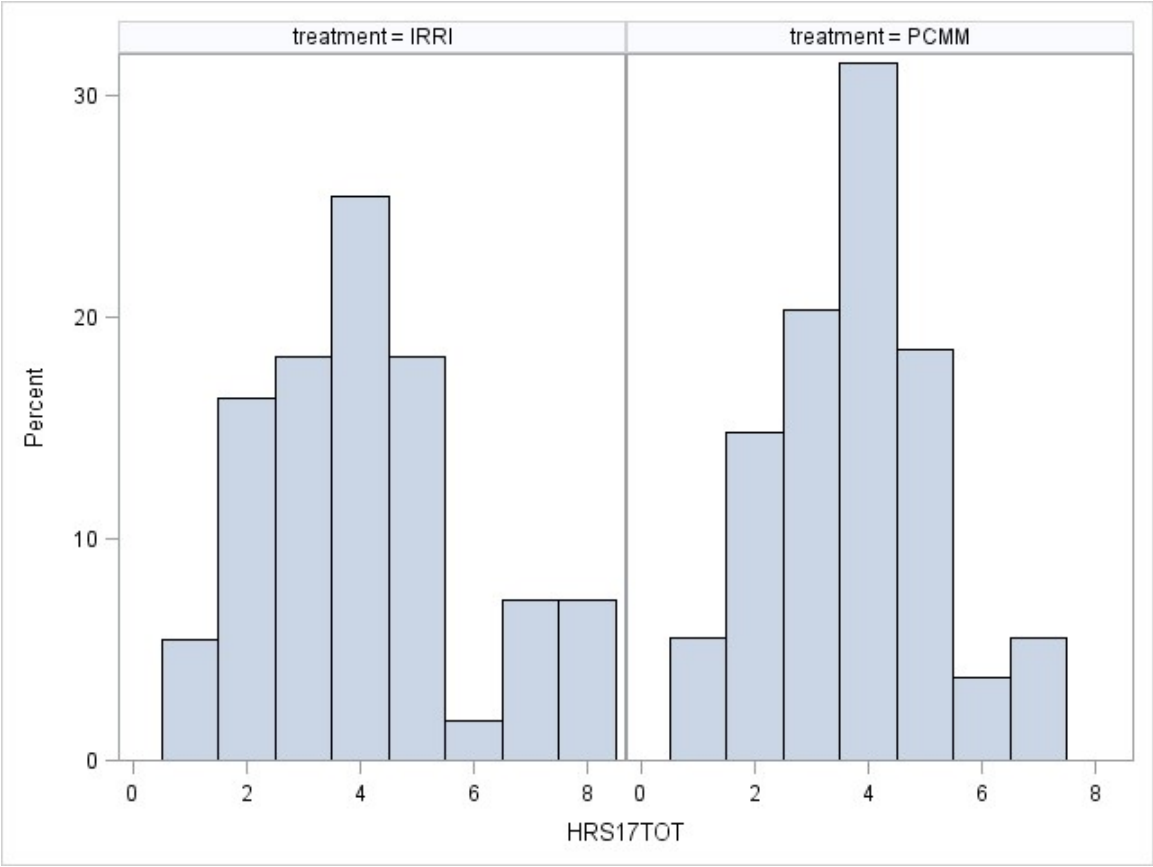


Figure 9: Box plots of Individual Standard Deviations by Treatment

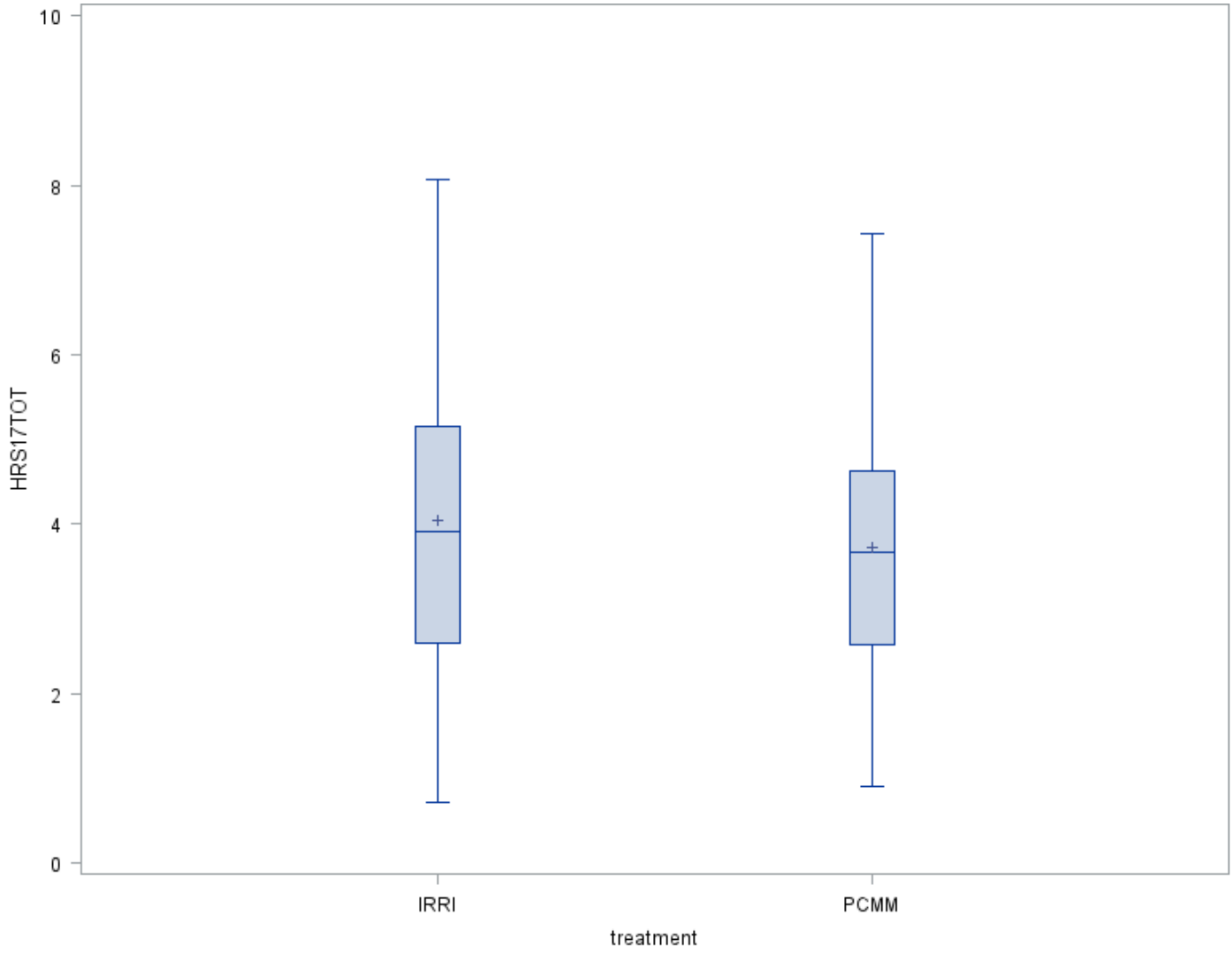
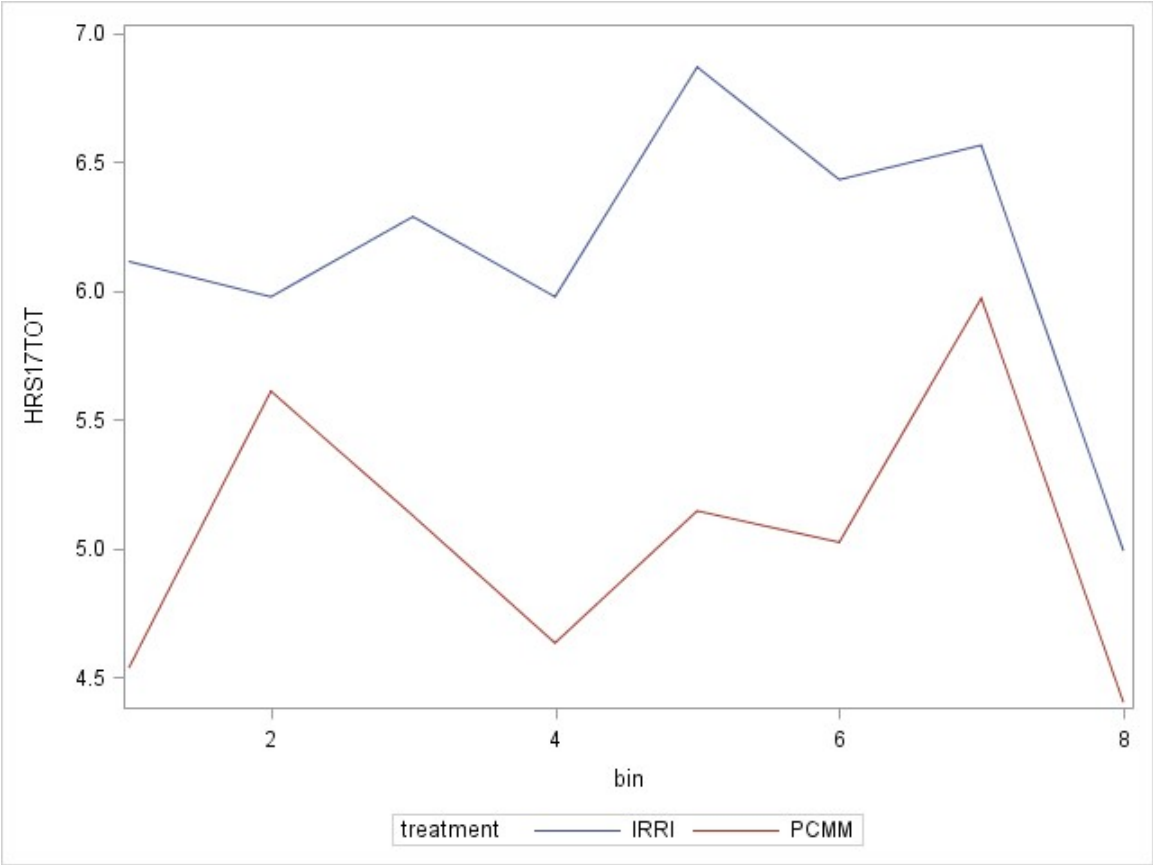


Figure 10: HRS17TOT Variability Over Time



3.4 YOUNGTOT VARIABLE ANALYSIS

The second variable analyzed is the Young Mania Rating Scale (YOUNGTOT). Similar methods used to analyze HRS17TOT are used to analyze YOUNGTOT. The main difference between the two variables is seen in Figures 11 and 13. Where the HRS17TOT distribution was roughly normal with slight right skew, the YOUNGTOT distribution is strongly positively skewed. Over time the amount of skewness changes and as evident in the 5th and 6th bins in Figure 13 the mean of the responses changes as well indicating recurrence of bipolar symptoms. Figures 12 and 14 are Normal QQ-plots which show strong deviation from the normal distribution. Looking at the time plots in Figures 15 and 16 shows different variances for the two intervention groups. The IRRI group contains a few possible outliers where scores are recorded around or greater than 30. Figures 17 and 18 reveal a greater spread among the standard deviations from the individuals in the IRRI group compared to those in the PCMM. Accompanied with the trajectory of the variances shown in Figure 19 these results seem to indicate that the IRRI group has a more volatile treatment similar to the findings from the HRS17TOT analysis. Unlike the HRS17TOT time does appear to have an effect on the variability.

A MELS model will be able to model the treatment and time effects on both the mean and variance structure; however, the issue of skewness must be addressed before its use. Classical techniques to eliminate skewness involve transforming the variables using the log, square root and other transformations (e.g. Box-Cox). For data with positive skew such as the YOUNGTOT variable the log-transformation is most commonly used. Besides removing skewness the transformation has (in most cases) the desirable trait of removing heterogeneity. Looking at the log-transformed Figures 20, 21, and 22 in comparison to the non-transformed Figures 15, 16, and 17 shows the removal of heterogeneity that was present in the data before the log-transformation (Figure 17 vs Figure 22). Before the transformation the PCMM revealed a more stable treatment than the IRRI; after the transformation this difference is not discernible.

If the primary goal is to assess whether there is a treatment effect present either in the variance or both the mean and variance then the underlying structure of the data must be preserved. Rather than transforming the data to account for skewness Arellano-Valle et al. (2007) and Chen (2012) have proposed the use of the skew-normal and skew-t distributions, respectively. These distribu-

tions preserve the original nature of the data and are able to account for the skewness. In order to model the YOUNGTOT variable with the MELS model I propose the use of the skew-normal distribution. Fitting of this model will be performed in Chapter 5.

Figure 11: Distribution of YOUNGTOT scores for Bins 1-4

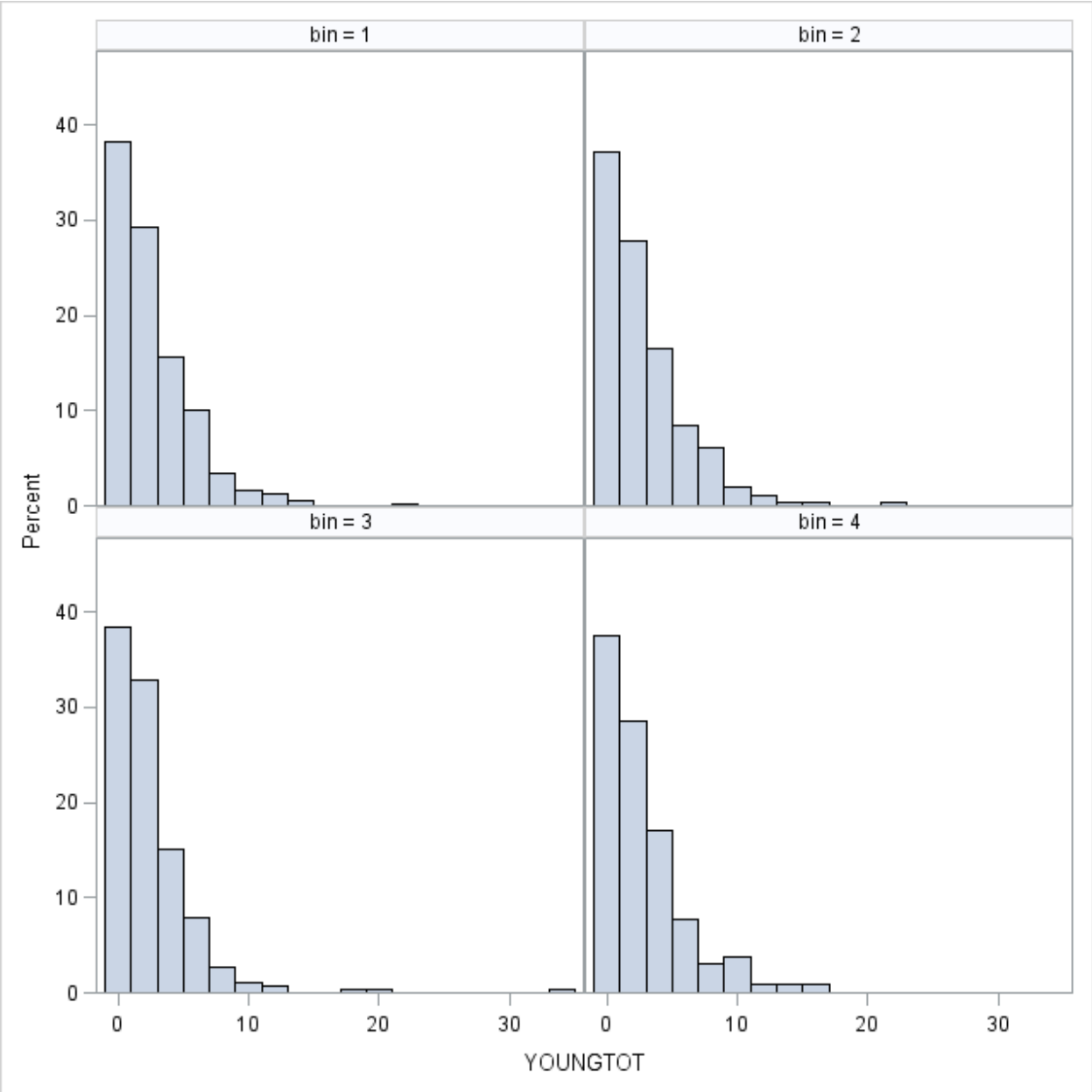
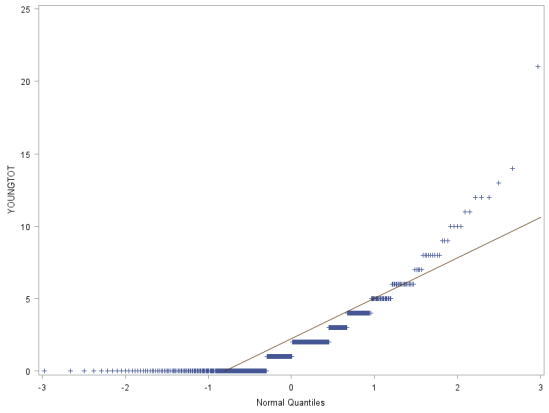
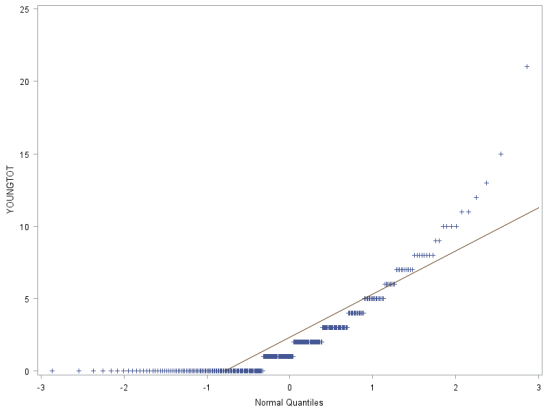


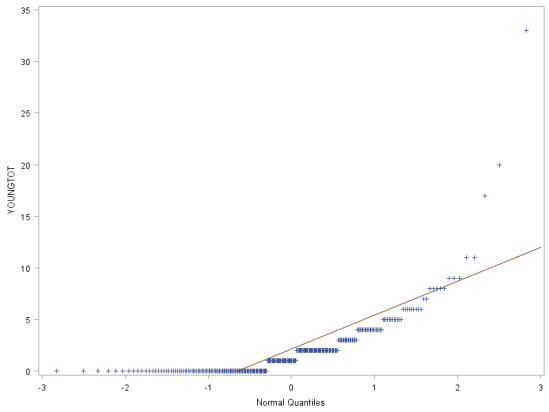
Figure 12: QQ-Plots of YOUNGTOT Scores for Bins 1-4



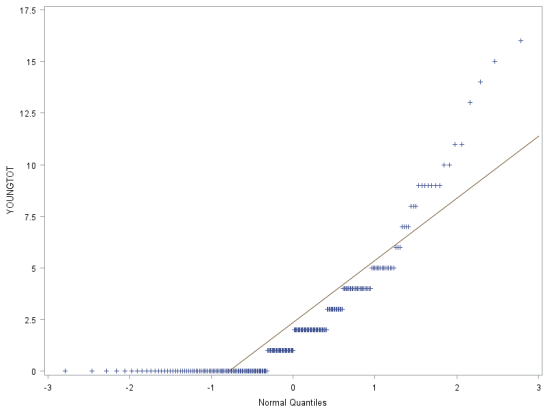
(a) Bin 1



(b) Bin 2



(c) Bin 3



(d) Bin 4

Figure 13: Distribution of YOUNGTOT scores for Bins 5-8

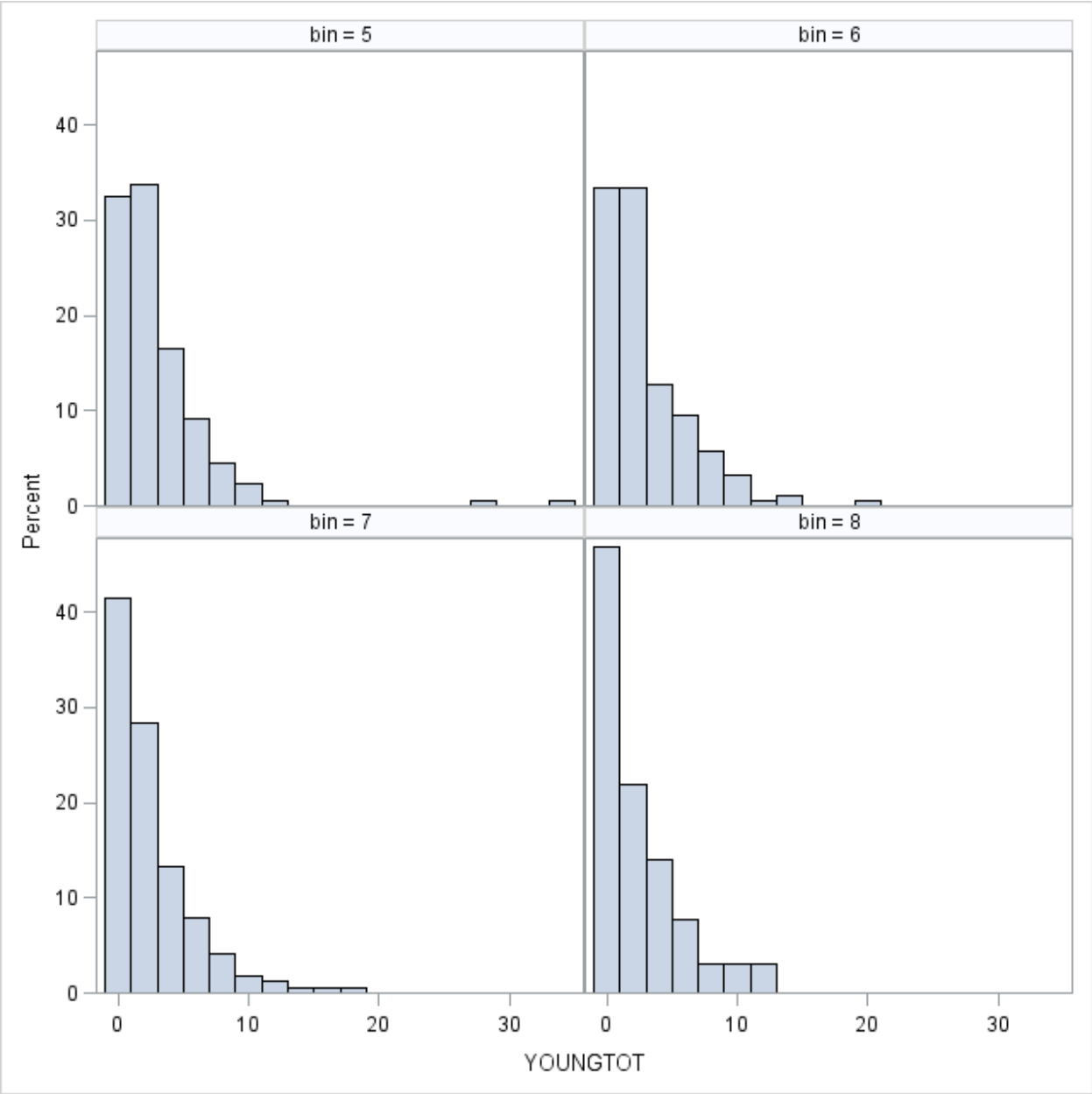
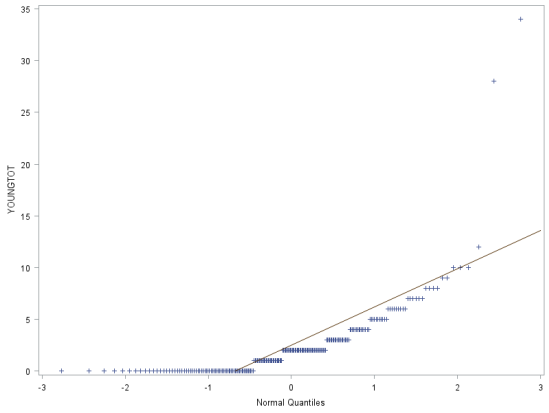
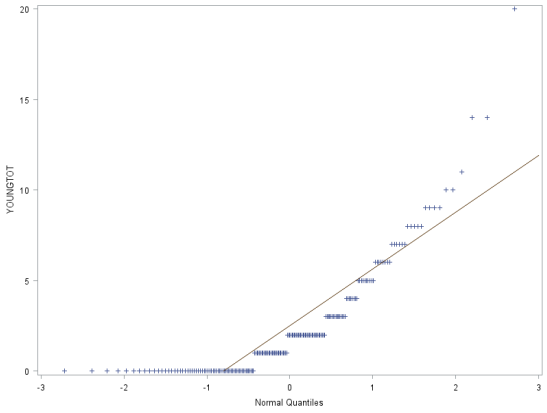


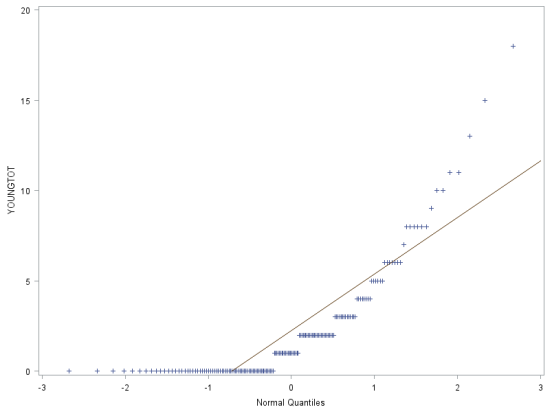
Figure 14: QQ-Plots of YOUNGTOT Scores for Bins 5-8



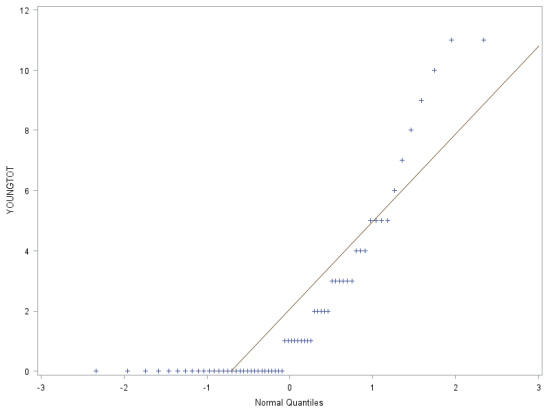
(a) Bin 1



(b) Bin 2

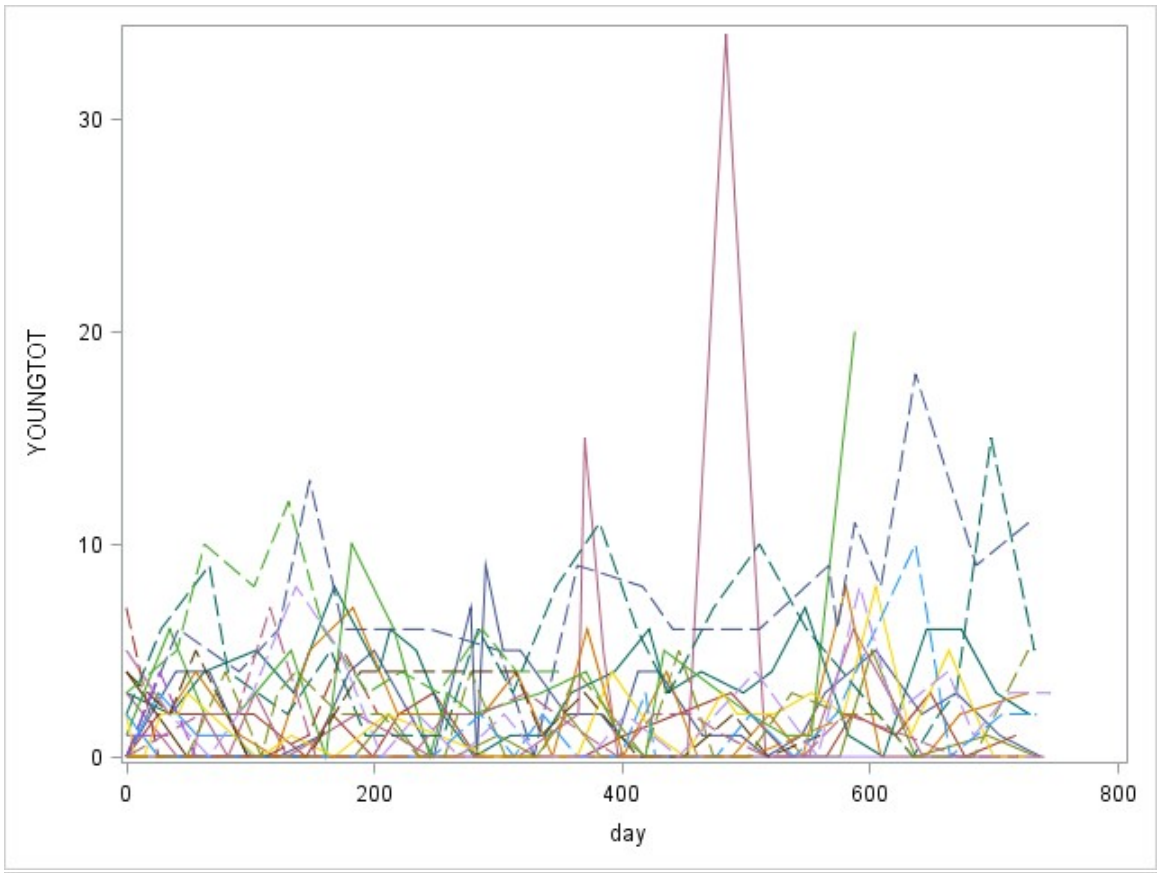


(c) Bin 3



(d) Bin 4

Figure 15: YOUNGTOT Scores by Time for each IRRI subject



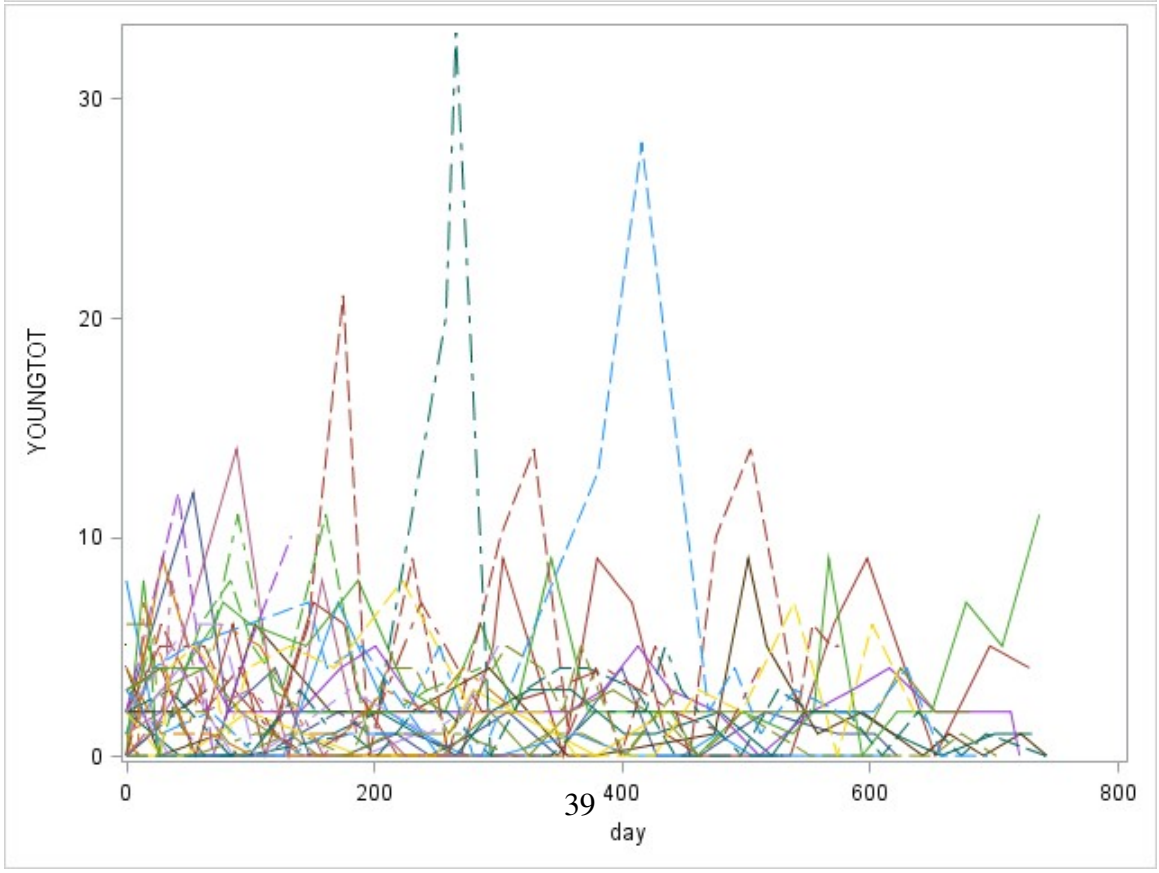
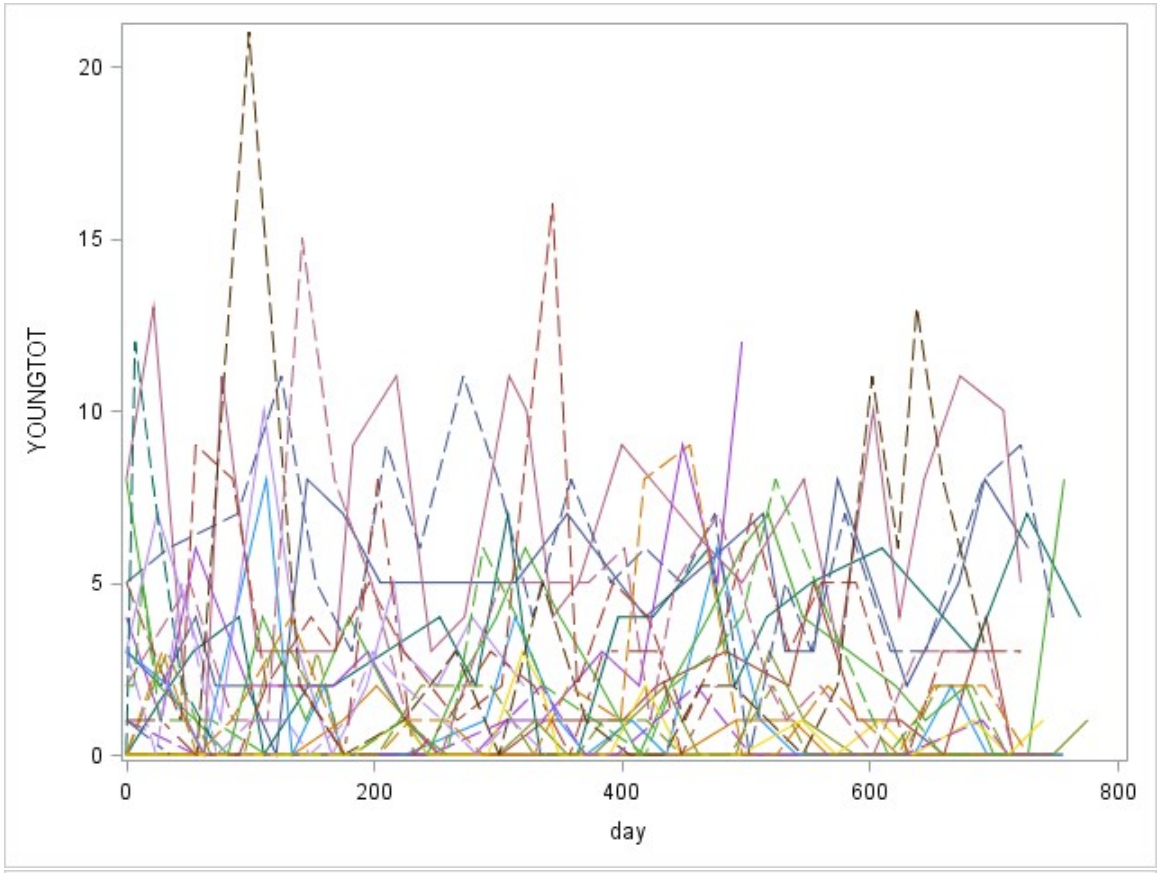


Figure 16: YOUNGTOT Scores by Time for each PCMM subject



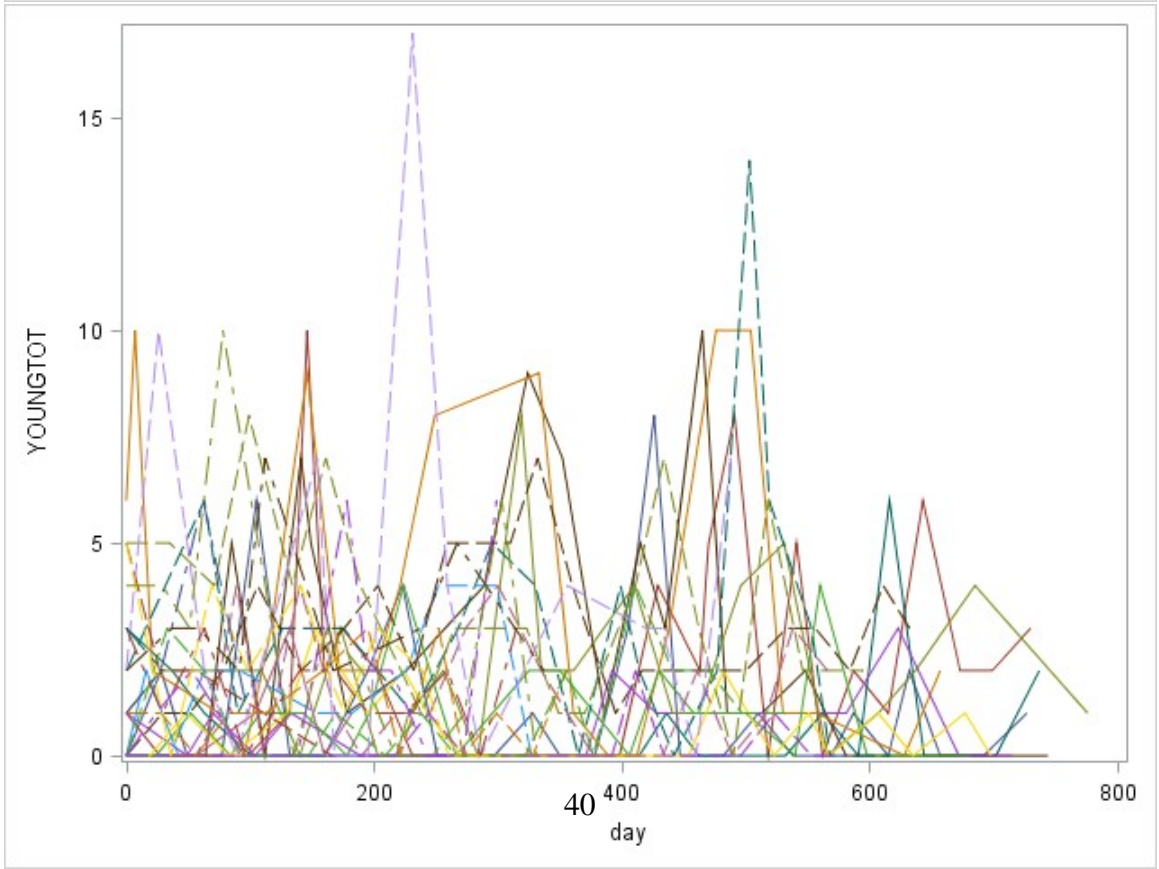


Figure 17: Histograms of Individual Standard Deviations by Treatment

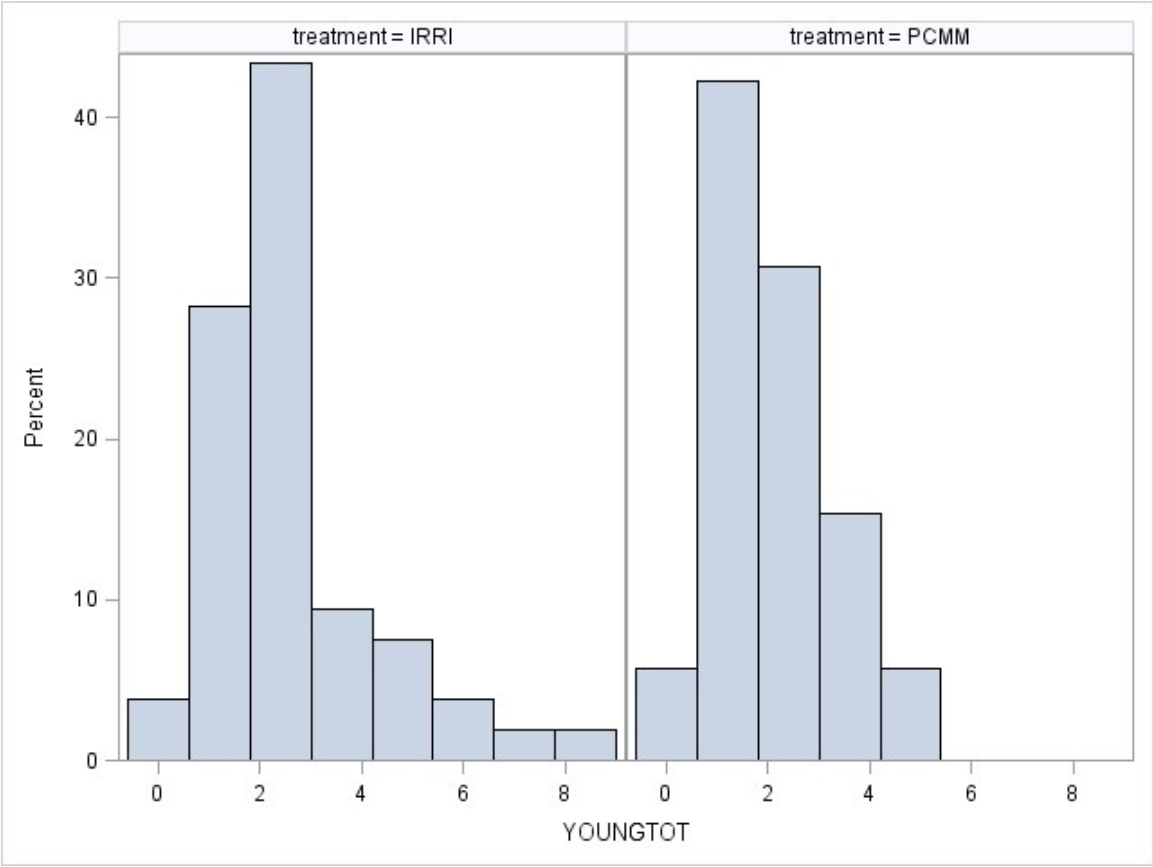


Figure 18: Box plots of Individual Standard Deviations by Treatment

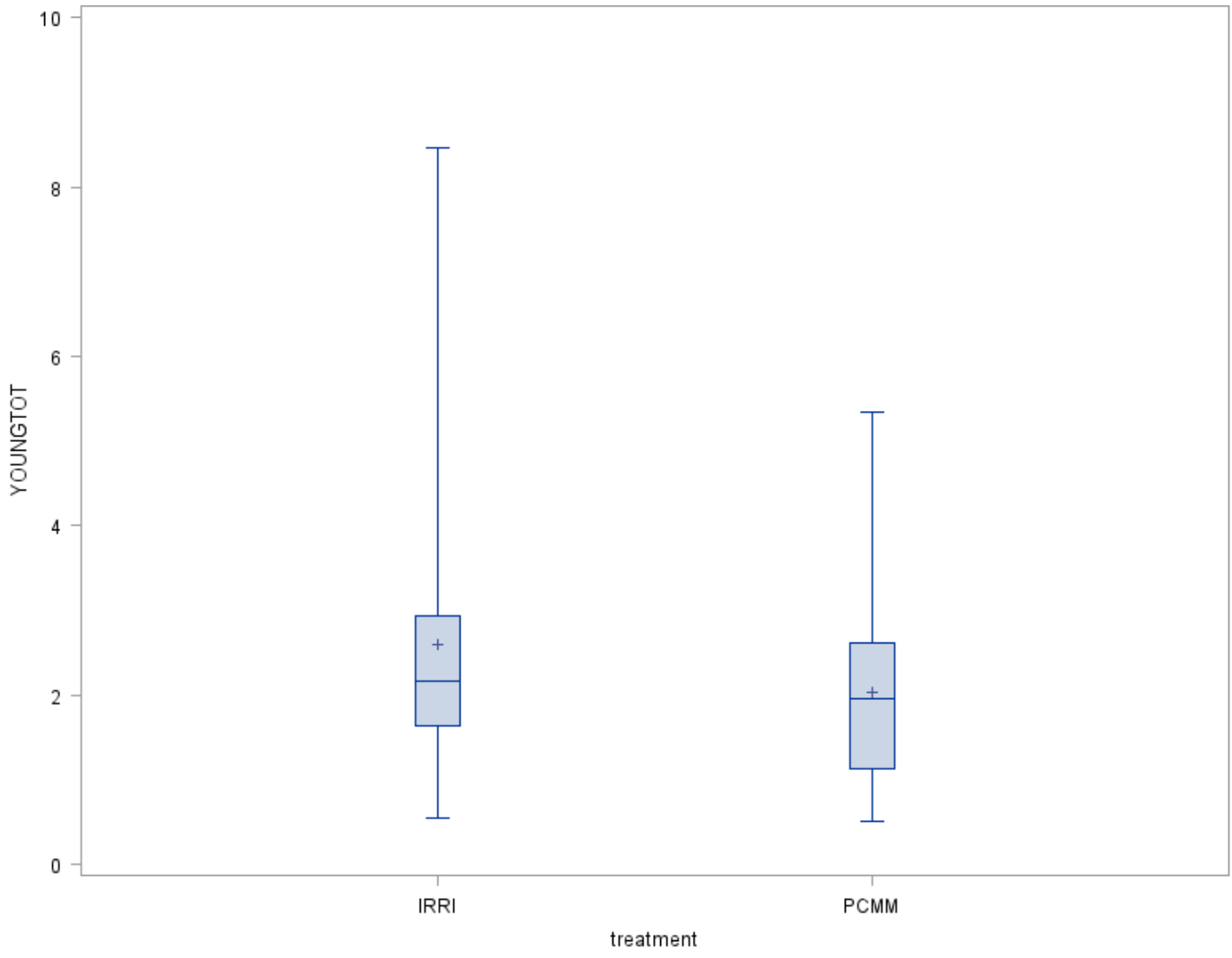


Figure 19: YOUNGTOT Variability Over Time

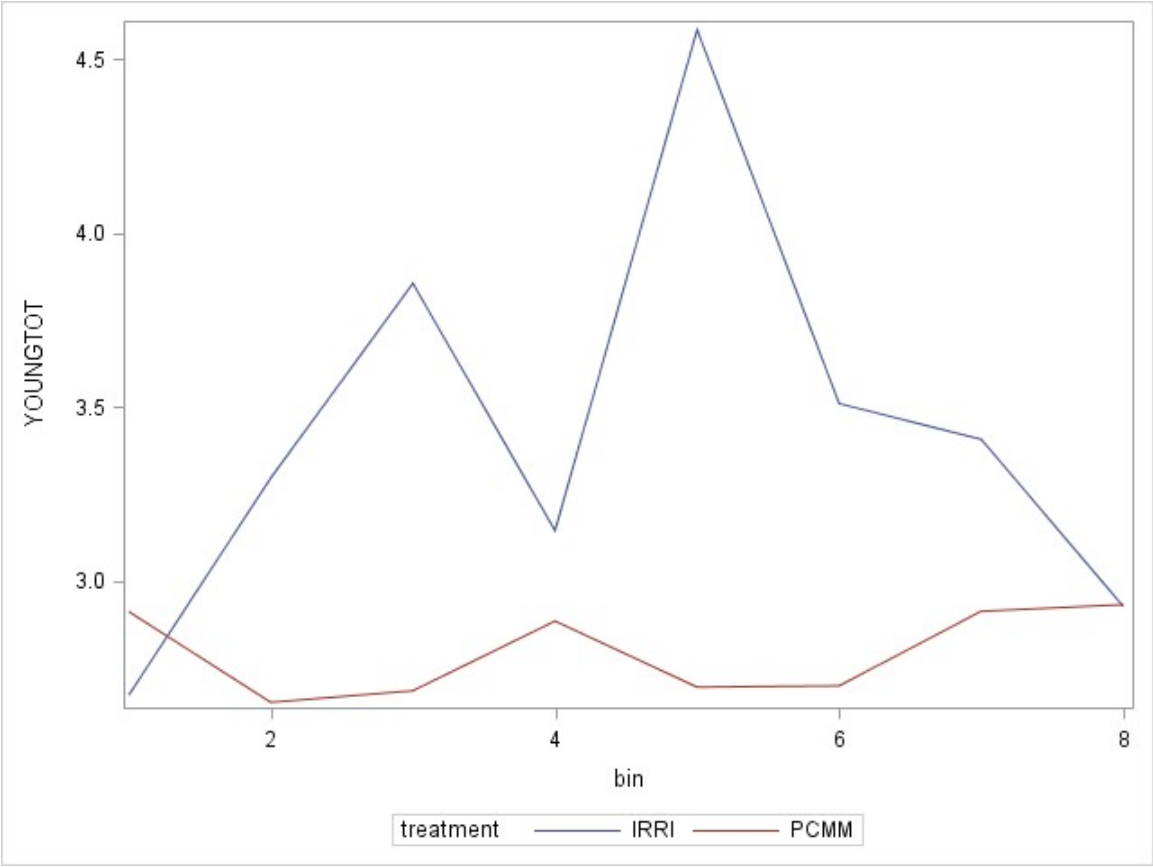
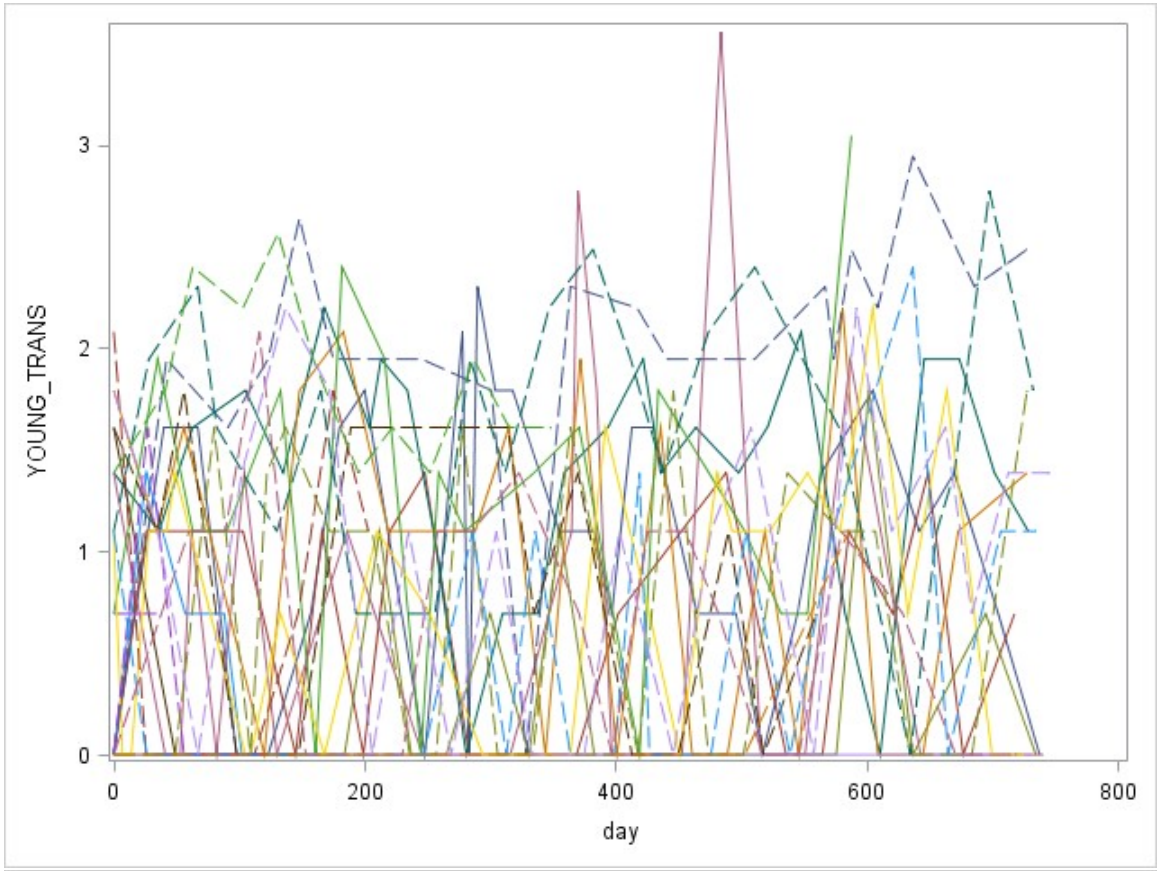


Figure 20: Distribution of YOUNGTOT scores (log-transformed) over Time for IRRI



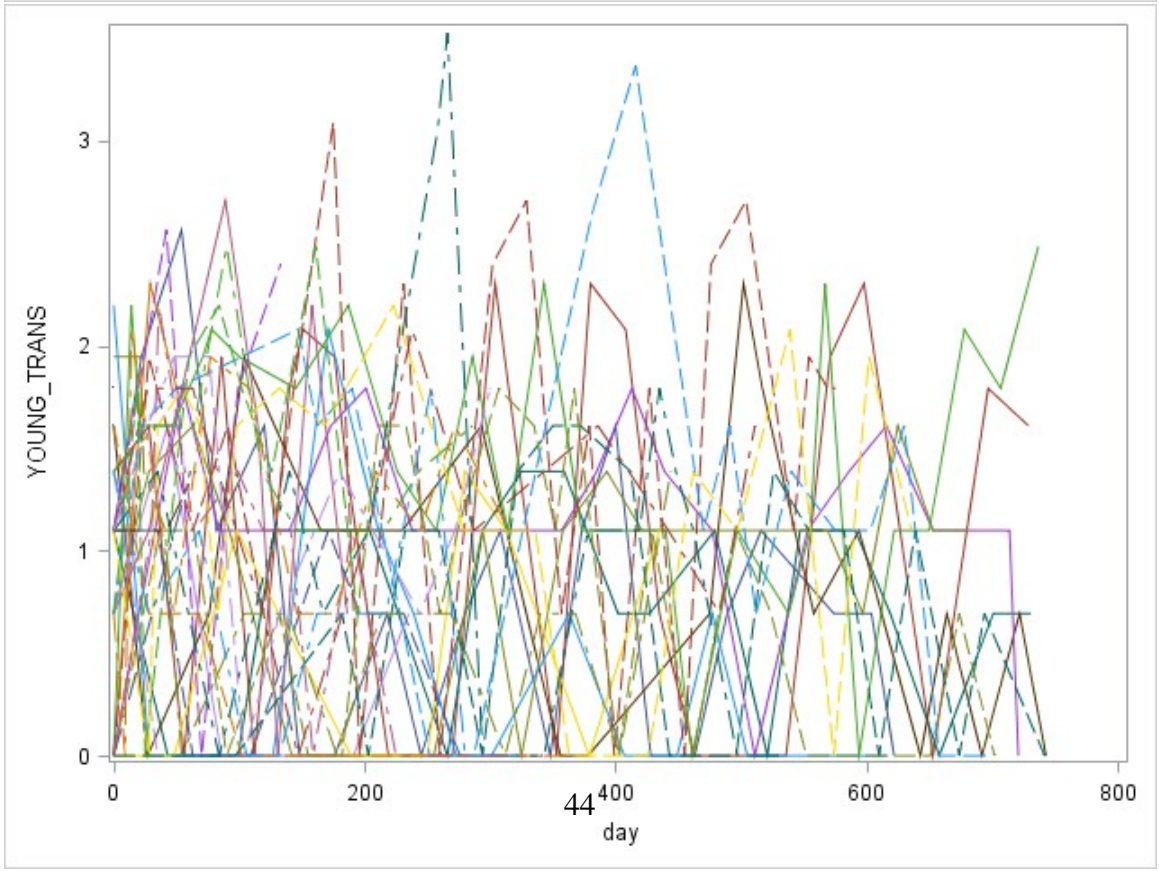
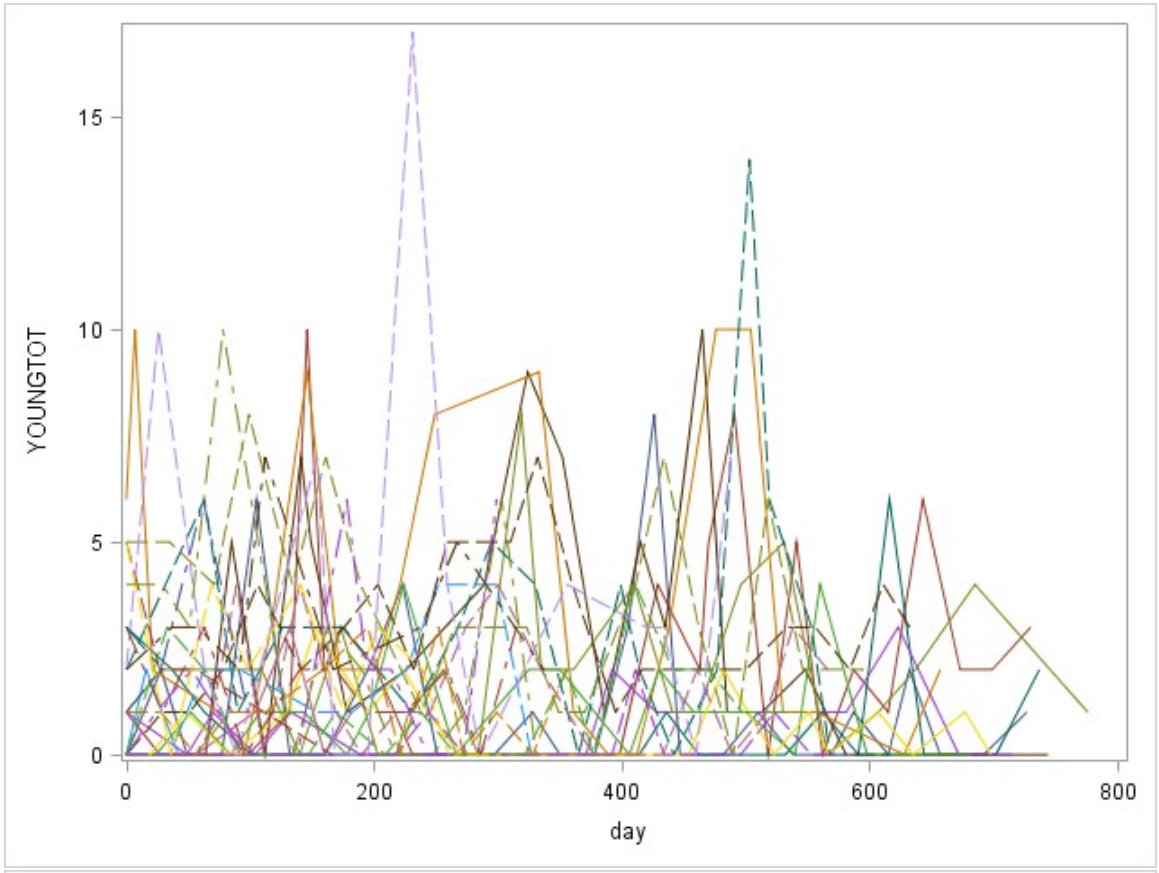


Figure 21: Distribution of YOUNGTOT scores (log-transformed) over Time for PCMM



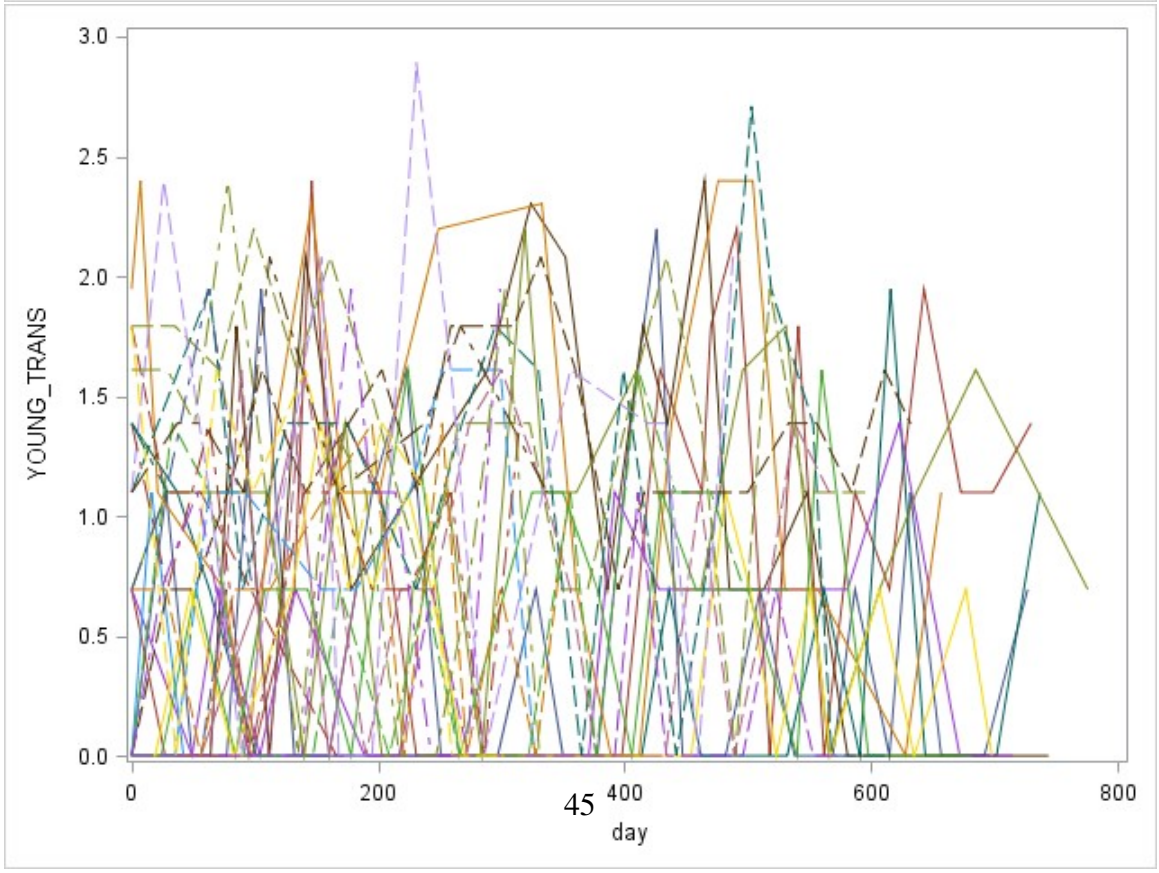
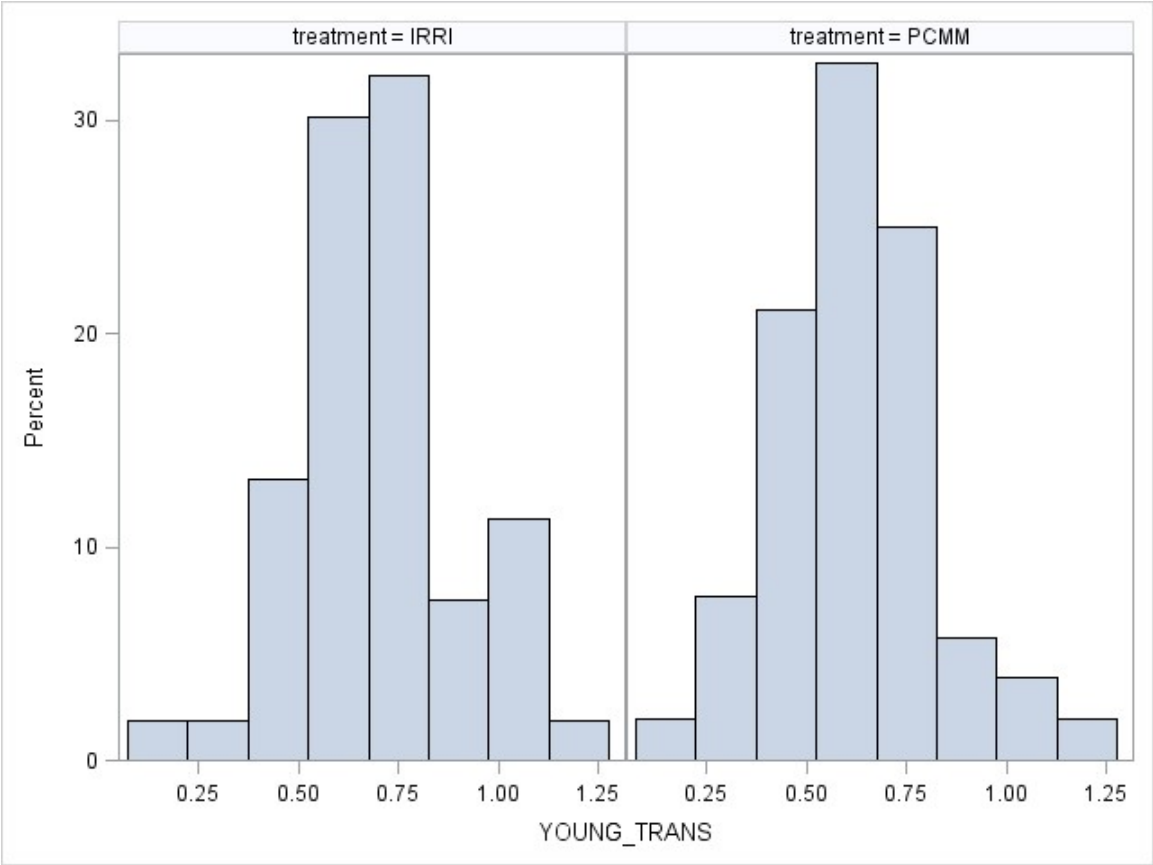


Figure 22: Histograms of Individual Standard Deviations by Treatment (log-transformed)



3.5 CLINICAL QUESTIONS ADDRESSED

Using standard mixed-effects models to address the clinical implications that arise from data sets such as the bipolar data, we are only able to deal with questions that involve the mean structure. Other statistical techniques may be used to deal with questions involving the variance structures, but few are able to address questions involving both the mean and variance simultaneously. One of the main benefits of the MELS model is that both the mean and variance parameters are able to be estimated simultaneously.

As seen in Figures 8 and 17 the variances of depression and mania symptoms within the subjects vary greatly. This suggests a random scale effect, which allows individuals to have their own within-subject variance. If this random effect is ignored then the parameter estimates will be inefficient (Shu, 2008). In addition, using a log linear model for the variance allows the effects of time and the intervention to be investigated. A MELS model allows researchers to assess the impact of various treatments on the mean and variance structure simultaneously. An example would be two individuals having the same mean score, but different variances. The subject with higher variability would be considered less stable and more symptomatic than the other subject. Moreover researchers can assess whether the proposed intervention (IRRI) results in a different trajectory (with respect to either the mean or variance) than the status quo (PCMM). Further work can be done to include additional covariates to assess their impact on the mean and variance structure.

A primary aim of Frank et al. (2015) was to assess whether individuals in IRRI and PCMM had different levels of variability over time, as high variability is an indicator of a less stable treatment course. Due to variable skewness and difficulties with model convergence the MELS model proposed by Hedeker cannot address these issues. I propose a skew-normal distribution to account for the skewness (this distribution allows for the original data structure of both the mean and variance to be preserved) and the use of Bayesian Markov chain Monte Carlo (MCMC) instead of Hedeker's MML approach to deal with the model convergence issues.

4.0 BAYESIAN MARKOV CHAIN MONTE-CARLO ESTIMATION FOR MIXED-EFFECTS LOCATION-SCALE MODELS WITH NORMAL ERRORS

The simulation results for the model estimated in Section 2.2.2.2 appear to provide adequate results for the specified model, where the WS variances are modeled using treatment only. In the bipolar clinical data (described in Chapter 3) one of the main clinical questions would involve a treatment by time interaction. Using the estimation procedure from 2.2.2.2 to fit the aforementioned model with the addition of a first-order time effect and second-order time and treatment effect to the WS variance fails to provide any reasonable estimates (either due to failure in the Gaussian quadrature step or the multidimensional optimization step). Pugach et al. (2014) were able to obtain convergence using continuous covariates in the WS variability using data from an adolescent smoking study. While this demonstrates the capability of Hedeker's technique to fit a mixed-effects location-scale (MELS) model with continuous covariates it also highlights the importance that the nature of the data has on the model fit especially as it pertains to the optimization process.

While Hedeker has developed an established estimation technique for a MELS model due to the lack in applications to different data sets from various fields a more general estimation approach is needed. Two methods of accomplishing this would involve the development of additional optimization methods and the development of a Bayesian approach. The benefit of the optimization approach would be the use of the well-established MML technique. However, drawbacks would consist of adapting advanced optimization techniques and devising a method to select the appropriate optimization for the data at hand. The Bayesian approach, often considered one of "last resort", has gained increasing popularity over the past twenty years as a way for modeling systems due to the ever growing computational infrastructure. Its benefits consist of the use of probability to estimate models and evidence that these approaches provide more accurate models

in comparison to non-Bayesian techniques especially for highly parameterized models (Gelman et al., 2014a). Drawbacks consist of the lack of a viable method for model checking and in some cases reliance on Markov chain Monte Carlo (MCMC) which may lead to approximation error due to its computationally intensive nature.

In this section I propose a Bayesian MCMC estimation technique (implementable using OPENBUGS software, which can also be called via R) and apply it to three different data sets. The first data set consists of the bipolar data discussed in Chapter 3 with the HRS17TOT serving as the response variable. The last two data sets are sample data sets from Hedeker and Nordgren (2013) with one resembling a longitudinal data set and the second resembling an intensive longitudinal data set from an EMA study.

4.1 BAYESIAN MARKOV CHAIN MONTE CARLO

4.1.1 Bayesian Analysis

Bayesian techniques differ from frequentist ones in that Bayesians believe parameters arise from probability distributions; whereas, frequentists take parameters as fixed which are estimated through optimization. When performing Bayesian analysis the first step is selecting the prior distributions for the parameters Θ. This may be done through the use of prior knowledge or without. If performed without prior knowledge prior distributions $p(\Theta)$ are often assumed to have large variances which enables a greater parameter space to be explored and to be conjugate priors which simplifies the derivation of the posterior distributions. The belief in Bayesian analysis is that by conditioning on the observations in the data set the posterior distributions will accurately represent the true distribution of the parameters for the specified data. This is performed as follows:

1. Set up joint probability $\quad p(y, \Theta) = p(y|\Theta)p(\Theta)$
2. Condition on data $\qquad p(\Theta|y) = \frac{p(y,\Theta)}{p(y)}$

Here $p(y|\Theta)$ represents the likelihood function and $p(y)$ is the normalizing constant. For basic models where appropriate conjugate priors are chosen, the posterior distribution $p(\Theta|y)$ takes on a

closed-form solution without the need for $p(y)$ to be computed numerically. However, in any non-linear model such as the MELS model computational difficulties arise when integrating $p(y, \Theta)$. A procedure popularized by Gelfand and Smith (1990) known as Markov chain Monte Carlo is used to bypass the integration.

4.1.2 Bayesian MCMC with Gibbs Sampling Procedure

In estimating the parameters of a MELS model using a Bayesian approach it is necessary to use a MCMC approach. This approach makes use of a Markov Chain which under specific conditions has a stationary condition. Let $p_{ij}^{(n)}$ represent the transition probabilities of a Markov Chain going from state i to j in n steps. If the Markov Chain is irreducible, aperiodic, and invariant, then as $n \to \infty$, $p_{ij}^{(n)} \to \pi$ where π represents a stationary distribution. The use of a Markov Chain in Bayesian inference uses the prior distributions and the initial data likelihood to determine the transition probabilities. Under the above conditions for stationarity, a stationary distribution will be determined which corresponds to the posterior distribution for the parameters after several iterations. Gibbs sampling will be used which consists of specifying transition probabilities for the parameters, sampling a new value then updating the value for use in the distribution of the next parameter.

The following steps will be used to determine the posterior distribution for the parameters (Geman and Geman (1984) and Gelfand and Smith (1990)):

1. Start with initial state for parameters $\boldsymbol{\Theta^{(0)}} = (\boldsymbol{\beta^{(0)}}, \boldsymbol{\tau^{(0)}}, \boldsymbol{\gamma^{(0)}}, \sigma_\omega^{2(0)})$
2. Define transition probabilities for parameters
   a. $\beta_k^{(t+1)} \sim \pi(\beta_k | \boldsymbol{y}, \beta_0^{(t+1)}, \ldots, \beta_{k-1}^{(t+1)}, \beta_k^{(t)}, \beta_{k+1}^{(t)}, \ldots, \beta_p^{(t)}, \boldsymbol{\tau^{(t)}}, \boldsymbol{\gamma^{(t)}}, \sigma_\omega^{2(t)})$
   b. $\tau_l^{(t+1)} \sim \pi(\tau_l | \boldsymbol{y}, \boldsymbol{\beta^{(t+1)}}, \tau_0^{(t+1)}, \ldots, \tau_{l-1}^{(t+1)}, \tau_l^{(t)}, \tau_{l+1}^{(t)}, \ldots, \tau_l^{(t)}, \boldsymbol{\gamma^{(t)}}, \sigma_\omega^{2(t)})$
   c. $\gamma_m^{(t+1)} \sim \pi(\gamma_m | \boldsymbol{y}, \boldsymbol{\beta^{(t+1)}}, \boldsymbol{\tau^{(t+1)}}, \gamma_0^{(t+1)}, \ldots, \gamma_{m-1}^{(t+1)}, \gamma_m^{(t)}, \gamma_{m+1}^{(t)}, \ldots, \gamma_m^{(t)}, \sigma_\omega^{2(t)})$
   d. $\sigma_\omega^{2(t+1)} \sim \pi(\sigma_\omega^2 | \boldsymbol{y}, \boldsymbol{\beta^{(t+1)}}, \boldsymbol{\tau^{(t+1)}}, \boldsymbol{\gamma^{(t+1)}}, \sigma_\omega^{2(t)})$
3. Repeat step 2 B times until a stationary distribution is found according to specified diagnostic criteria.

In conducting step 2, certain distributions π will not have a distribution that is easy to sample from directly; therefore, a Metropolis-Hastings algorithm will be used for sampling (Metropolis

et al. (1953) and Hastings (1970)). This method is able to draw samples from any probability distribution π by using a target function $f(x)$ that is proportional to π. The use of a distribution proportional to π eliminates the requirement of having to calculate a normalizing constant which in many cases might be numerically taxing. In addition a proposed distribution $Q(x'|x^{(t)})$ is used to generate samples from π. Depending on the parameters being estimated a symmetric distribution such as the Normal or Uniform is suggested in order to obtain a stationary distribution quicker. The proposed distribution used in this paper will be a normal distribution. The iteration steps involved in the Metropolis-Hastings Algorithm are:

1. Using an initial value, say $x^{(t)}$, generate a sample x' from the proposed distribution $Q(x'|x^{(t)})$
2. Calculate the acceptance ratio using the target function $f(x)$ as $a = \frac{f(x')}{f(x^{(t)})}$ which is $\frac{\pi(x')}{\pi(x^{(t)})}$ due to the proportional assumption
3. If $a \geq 1$ then accept state and $x' = x^{(t+1)}$, else if $a < 1$ accept state with probability a, which implies reject state with probability $1 - a$ and set $x^{(t+1)} = x^{(t)}$

4.2 MIXED-EFFECTS LOCATION-SCALE MODEL WITH NORMALLY DISTRIBUTED ERRORS

In order to fit the model in a Bayesian setup the following hierarchical set-up for the MELS model will be used for subject i with n_i observations:

$$\boldsymbol{Y_i}|\boldsymbol{\nu_i}, \boldsymbol{\beta}, \boldsymbol{\psi_i}, \boldsymbol{\omega_i} \overset{ind.}{\sim} N_{n_i}(\boldsymbol{X_i\beta} + \boldsymbol{1_{n_i}}\nu_i, \boldsymbol{\psi_i}) \tag{4.1}$$

$$\nu_i|\sigma_\nu^2 \overset{ind.}{\sim} N(0, \sigma_\nu^2), \quad i = 1, \ldots, N \tag{4.2}$$

where the variances are modeled as follows

$$\sigma_\nu^2 = \exp(\boldsymbol{u_i\tau})$$
$$\boldsymbol{\psi_i} = Diag(\sigma^2_{\epsilon_{i1}}, \ldots, \sigma^2_{\epsilon_{in_i}})$$
$$\sigma^2_{\epsilon_{ij}} = \exp(\boldsymbol{w_{ij}\gamma} + \omega_i)$$
$$\omega_i \sim N(0, \sigma^2_\omega)$$

Here the random effects ν_i (location) and ω_i (scale) are assumed to be independent. $\boldsymbol{X_i}$ is a $n_i \times p$ design matrix of the fixed effects and $\boldsymbol{w_{ij}}$ is a $1 \times q$ vector of covariates influencing the WS variability, which may vary over both subject and observation. $\boldsymbol{u_i}$ is a $1 \times s$ vector of covariates affecting the BS variability, which only varies by subject. $\mathbf{1_{n_i}}$ is a $n_i \times 1$ vector of 1's and $Diag$ is a $n_i \times n_i$ diagonal matrix.

The conditional pdf on the sample $\boldsymbol{y} = (\boldsymbol{y_1^T}, \ldots, \boldsymbol{y_N^T})$ is

$$f(\boldsymbol{y}|\boldsymbol{\nu}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\omega}) = \prod_{i=1}^{N} \phi_{n_i}(\boldsymbol{y_i}|\boldsymbol{X_i\beta} + \mathbf{1}_{\boldsymbol{n_i}}\boldsymbol{\nu_i}, \boldsymbol{\psi_i}) \tag{4.3}$$

where $\phi_{n_i}(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the pdf of a n_i dimensional multivariate normal random variable $\boldsymbol{y}$ with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$. The common estimation technique is to maximize the marginal likelihood of the response vector $\boldsymbol{y}$. Using a theorem from Arellano-Valle et al. (2007) I derived the following lemma (proof provided in APPENDIX).

Lemma 1 *Let* $\boldsymbol{Y_i} = \boldsymbol{X_i\beta} + \mathbf{1}_{\boldsymbol{n_i}}\nu_i + \boldsymbol{\epsilon_i}$*, where* $\nu_i \sim N(0, \sigma_\nu^2)$ *and* $\boldsymbol{\epsilon_i} \sim N(0, \boldsymbol{\psi_i})$ *are independent. The variances are denoted by* $\sigma_\nu^2 = \sigma_\nu^2(\tau)$ *and* $\boldsymbol{\psi_i} = \boldsymbol{\psi_i}(\boldsymbol{\gamma}, \omega)$ *with* $\omega \sim N(0, \sigma_\omega^2)$*. Then the marginal distribution of* $\boldsymbol{Y_i}$ *is*

$$f_{\boldsymbol{Y_i}}(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \sigma_\omega^2) = \int_{\mathbb{R}} \phi_n(\boldsymbol{y}|\boldsymbol{X\beta}, \boldsymbol{\psi_i} + \sigma_\nu^2\boldsymbol{J_n})\phi(\omega|0, \sigma_\omega^2)d\omega \tag{4.4}$$

The result from equation 4.4 in Lemma 1 cannot be simplified any further due to the nonlinear nature in which the random scale effect enters the model. Therefore when implementing MML integral approximations are used. The idea behind marginal maximum likelihood is to integrate out the "nuisance" parameters which variance terms typically are treated as. However, when using a MELS model the parameters associated with the variances are of importance as well as the mean

parameters. In order to obtain appropriate inferences on all of the parameters Bayesian framework is required (Verbeke and Molenberghs, 2000).

In order to estimate the MELS model defined in equations 4.1 and 4.2 using Bayesian MCMC, prior parameter distributions must be assigned. One case of specified priors will be:

$$\begin{aligned}
&\beta_k \sim N(\beta_{k0}, \sigma^2_{k0}) \quad k = 1, \ldots, p \\
&\tau_l \sim N(\tau_{l0}, \sigma^2_{l0}) \quad l = 1, \ldots, r \\
&\gamma_m \sim N(\gamma_{m0}, \sigma^2_{m0}) \quad m = 1, \ldots, q \\
&\sigma^2_\omega \sim IG(a_0, b_0)
\end{aligned} \tag{4.5}$$

Here N refers to the normal distribution and IG refers to an inverse gamma distribution. The initial values of the hyperparameters will be chosen assuming no prior knowledge. Using the prior distributions from 4.5 and the marginal pdf in 4.4 the joint posterior distribution of all of the parameters given the observed sample $\boldsymbol{y}$ is

$$\begin{aligned}
\pi(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \sigma^2_\omega | \boldsymbol{y}) \propto & \prod_{i=1}^{N} \int_{\mathbb{R}} \phi_n(\boldsymbol{y_i} | \boldsymbol{X\beta}, \boldsymbol{\psi_i} + \sigma^2_\nu \boldsymbol{J_n}) \phi(\omega | 0, \sigma^2_\omega) d\omega \\
& \times \prod_{k=1}^{p} \phi(\beta_k | \beta_{k0}, \sigma^2_{k0}) \times \prod_{l=1}^{r} \phi(\tau_l | \tau_{l0}, \sigma^2_{l0}) \times \prod_{m=1}^{q} \phi(\gamma_m | \gamma_{m0}, \sigma^2_{m0}) \\
& \times \frac{b_0^{a_0}}{\Gamma(a_0)} \sigma_\omega^{2(-a_0-1)} \exp\left(\frac{-b_0}{\sigma^2_\omega}\right)
\end{aligned} \tag{4.6}$$

As is characteristic of a MELS model the resulting posterior distributions such as 4.6 do not have closed form solutions therefore a Bayesian MCMC with Gibbs sampling and the Metropolis-Hastings Algorithm will be used to determine the posterior distribution.

4.3 SIMULATION RESULTS

In order to assess the accuracy of the proposed model, simulations will be performed using a simulated data set that resembles the bipolar data set. 100 subjects will be randomly assigned to two interventions: treatment and control. Measurements will be taken over a span of 108 weeks (approximately 2 years) spaced 9 weeks apart (approximately 2 months). The specified model will be

$$\begin{aligned} \boldsymbol{Y_i}|\boldsymbol{\nu_i}, \boldsymbol{\beta}, \boldsymbol{\psi_i}, \boldsymbol{\omega_i} &\overset{ind.}{\sim} N_n(\beta_0 + \beta_1 trt_i + \beta_2 week_j^* + \beta_3 trt_i \times week_j^* + 1_n \nu_i, \boldsymbol{\psi_i}) \\ \nu_i|\sigma_\nu^2 &\overset{ind.}{\sim} N(0, \sigma_\nu^2), \quad i = 1, \ldots, N \end{aligned} \tag{4.7}$$

where the variances are modeled as follows

$$\begin{aligned} \sigma_\nu^2 &= \exp(\tau_0 + \tau_1 \times trt_i) \\ \boldsymbol{\psi_i} &= Diag(\sigma_{\epsilon_{i1}}^2, \ldots, \sigma_{\epsilon_{in}}^2) \\ \sigma_{\epsilon_{ij}}^2 &= \exp(\gamma_0 + \gamma_1 trt_i + \gamma_2 week_j^* + \gamma_3 trt_i \times week_j^* + \omega_i) \\ \omega_i &\sim N(0, \sigma_\omega^2) \end{aligned}$$

The week variable is transformed to $week_j^* = week_j/100$ to aid in model estimation.

As is characteristic of Bayesian statistics a prior distribution must be specified for each parameter. The choice of distribution will affect the estimates and implementation speed, so two different scenarios with different priors will be assessed. Typically normal priors are assumed; however, in many models such as the MELS model using the normal for all priors (especially for those associated with the log-linear parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\tau}$) may result in erroneous results, such as the distributions diverging. This happens most often in cases where the prior distributions have large variances. A distribution such as a uniform would be able to limit the ranges that the values that the parameters take.

The first scenario specifies the prior for the $\boldsymbol{\beta}, \boldsymbol{\tau}$, and $\boldsymbol{\gamma}$ parameters as normal and the prior for σ_ω^2 as an inverse gamma. In setting the hyperparameters the exponential nature of the WS variance must be taken into consideration as values that are too large may result in the model diverging.

Therefore reasonable hyperparameters must be set in order to avoid this issue. They are specified as follows

$$\begin{aligned}
&\beta_k \sim N(0, 100) \quad k = 0, 1, 2, 3 \\
&\tau_l \sim N(0, 100) \quad l = 0, 1 \\
&\gamma_m \sim N(0, 2) \quad m = 0, 1, 2, 3 \\
&\sigma^2_\omega \sim IG(10, 10)
\end{aligned} \tag{4.8}$$

Th hyperparameter values were chosen through trial and error. Using large values of the hyperparameters is desired to have non-informative priors; however, this was not possible for all of the prior distributions. The variance hyperparameter for $\boldsymbol{\tau}$ was particularly troublesome and the model fit would diverge for any hyperparameter variance values greater than 2.

Due to the nature of the Metropolis-Hastings Algorithm consecutive draws are highly correlated. In order to make appropriate inference a Bayesian draw consisting of independent samples is required. This correlation diminishes the further the draws are from each other. A thinning procedure that samples every 10th draw will be used to diminish this dependency. A sample size of 5,000 will be used after a burn-in period of 50,000 iterations. Under these conditions one run for scenario 1 takes approximately 20 minutes on a Samsung NP305E5A laptop.

The 2nd scenario differs from scenario 1 by changing the prior distributions for the parameters in the log-linear model from normal to uniform. The idea behind the use of the normal prior is to use a symmetric conjugate prior. In most cases this works especially for mean parameters; however, due to the log-linear nature of the variance the normal distribution is not a conjugate prior for this model. Another symmetric prior distribution will be used and the results will be compared to that of scenario 1. Hyperparameters are set using the same concept as previously described for scenario 1. The specified priors are

$$\begin{aligned}
&\beta_k \sim N(0, 100) \quad k = 0, 1, 2, 3 \\
&\tau_l \sim U(-5, 5) \quad l = 0, 1 \\
&\gamma_m \sim U(-5, 5) \quad m = 0, 1, 2, 3 \\
&\sigma_\omega^2 \sim IG(10, 10)
\end{aligned} \tag{4.9}$$

An added benefit of using the more mathematically simple uniform distribution is that the number of iterations required for the model to converge is half as much as scenario 1. For scenario 2 the burn-in period consists of 25,000 iterations and inference is based on a sample size of 2,500 obtained by selecting every 10th sample. One run takes approximately 10 minutes on a Samsung NP305E5A laptop in comparison to the 20 minutes from the previous scenario. Table 5 provides a comparison of the two scenarios based on 40 simulations and using the mean of the sample size as the Bayesian estimate.

Comparing the results from the two scenarios reveals relatively similar results. The coverage probabilities are approximately the same; however, the second scenario provides higher values in five out of the eleven parameters whereas the first outperforms in only three instances. Looking at the standard biases (obtained by dividing the bias by the standard error of the Bayes' estimates) shows that two of the estimates fall outside of the $\pm$ 20 range for the first scenario; whereas, none of the estimates fall outside for the second scenario. These inferences are based on 40 simulations; therefore, with relatively similar results for the two scenarios one would reasonably expect that as the number of simulations increased the differences between the two would become even smaller. Despite the minor computational differences between the two scenarios one major benefit of the second scenario is the amount of computation time required (10 min/run compared to 20 min/run for the first). This is a result of the Bayesian draw for the second scenario requiring half as many iterations to satisfy the stationary distributions which can be visually inspected. The graphs from these results are not shown; however, they resemble those provided in Figures 23, 24 and 25. The number of iterations required for the distributions to resemble those in the aforementioned figures are half as much as those required for scenario 1. Due to this benefit the prior distributions specified in the second scenario will be used for the remaining models.

Table 5: Simulation Results between Scenario 1 and 2

		Scenario 1					
Parameters	True Value	Estimate	SE	Bias	St. Bias	Cov. Prob. (%)	Reject H_0 (%)
β_0	2	1.980	.210	-.020	-9.589	95	100
β_1	.5	.512	.323	.012	3.730	92.5	40
β_2	.2	.212	.119	.012	9.934	95	45
β_3	-.2	-.212	.150	-.012	-7.944	95	20
τ_0	.7	.680	.203	-.020	-9.856	95	92.5
τ_1	-.1	-.069	.304	.031	10.230	95	5
γ_0	-1.4	-1.395	.121	.005	4.059	95	100
γ_1	.2	.158	.187	-.042	-22.574	95	12.5
γ_2	2.5	2.468	.170	-.032	-18.856	97.5	100
γ_3	-.5	-.387	.222	.113	50.822	97.5	22.5
σ_ω	.6	.613	.105	.013	12.454	90	100
		Scenario 2					
Parameters	True Value	Estimate	SE	Bias	St. Bias	Cov. Prob. (%)	Reject H_0 (%)
β_0	2	1.980	.183	-.020	-10.984	100	100
β_1	.5	.542	.256	.042	16.304	97.5	50
β_2	.2	.204	.102	.004	4.373	97.5	30
β_3	-.2	-.213	.163	-.013	-7.783	95	22.5
τ_0	.7	.706	.195	.006	2.980	95	92.5
τ_1	-.1	-.139	.272	-.039	-14.251	97.5	5
γ_0	-1.4	-1.421	.165	-.021	-12.799	90	100
γ_1	.2	.187	.227	-.013	-5.621	95	20
γ_2	2.5	2.490	.234	-.010	-4.259	92.5	100
γ_3	-.5	-.465	.327	.035	10.798	92.5	47.5
σ_ω	.6	.600	.089	.000	.221	97.5	100

Simulation results were expanded to 100 in order to better assess the MCMC's accuracy. These results (shown in Table 6) reveal an overall improved model fit when compared with the original 40 simulations. The mean estimates of the parameters have little to no bias and all of the coverage probabilities are between 89% and 98%. As previously mentioned when examining data sets similar to the bipolar data the parameters of interest are often the time and treatment interaction terms which are represented by β_3 and γ_3. The "Reject H_0" column represents the % of simulations

whose 95% interval do not contain 0. The % of intervals that do not contain 0 are 22 and 41 respectively for the given true values. This reveals that the proposed technique does not do a good job in detecting the treatment effects under the given conditions for the data set. These conditions include the number of subjects, number of observations per subject and the effect size with a focus on the treatment by time interaction in the within-subject variance. Methods to determine an appropriate sample size for detecting the treatment by time effect will be discussed in Chapter 6.

Table 6: Simulation Results based on 100 simulations

Parameters	True Value	Estimate	SE	Bias	St. Bias	Cov. Prob. (%)	Reject H_0 (%)
β_0	2	1.989	.216	-.011	-5.272	93	100
β_1	.5	.523	.309	.023	7.57	89	45
β_2	.2	.199	.112	-.001	-.639	95	34
β_3	-.2	-.202	.156	-.002	-1.325	98	22
τ_0	.7	.698	.208	-.002	-1.106	94	91
τ_1	-.1	-.131	.297	-.031	-10.483	95	6
γ_0	-1.4	-1.419	.159	-.019	-11.920	90	100
γ_1	.2	.220	.226	.020	8.900	89	24
γ_2	2.5	2.495	.219	-.005	-2.097	93	100
γ_3	-.5	-.494	.314	.006	2.066	93	41
σ_ω	.6	.595	.097	-.005	-4.968	90	100

4.4 APPLICATION TO BIPOLAR DATA

Based on the results from Section 4.3 the proposed estimation technique provides estimates with low bias and appropriate confidence intervals. Using the prior distributions specified in 4.9 and the model formulation in 4.7 the MELS model will be fit to analyze the HRS17TOT variable (discussed in Section 3.3). The Bayesian draw will consist of a 25,000 burn-in period and inference will be based on 2,500 samples obtained by sampling every 10th sample. The results are provided in Table 7 and diagnostic graphs are provided in Figures 23, 24 and 25.

Table 7: Estimation Results for HRS17TOT

Parameters	Mean	SD	2.5p	MED	97.5p
β_0	6.308	.536	5.235	6.312	7.292
β_1	1.119	.868	-.585	1.150	2.758
β_2	.801	.393	.034	.813	1.549
β_3	-1.406	.634	-2.617	-1.428	-.178
τ_0	2.483	.224	2.057	2.480	2.910
τ_1	.423	.300	-.143	.417	1.007
γ_0	2.475	.119	2.246	2.476	2.698
γ_1	.326	.164	-.010	.327	.632
γ_2	-.103	.219	-.518	-.108	.332
γ_3	-.048	.310	-.633	-.039	.501
σ_ω	.837	.077	.701	.836	.987

Figure 23: Convergence of Location Mean β Parameters

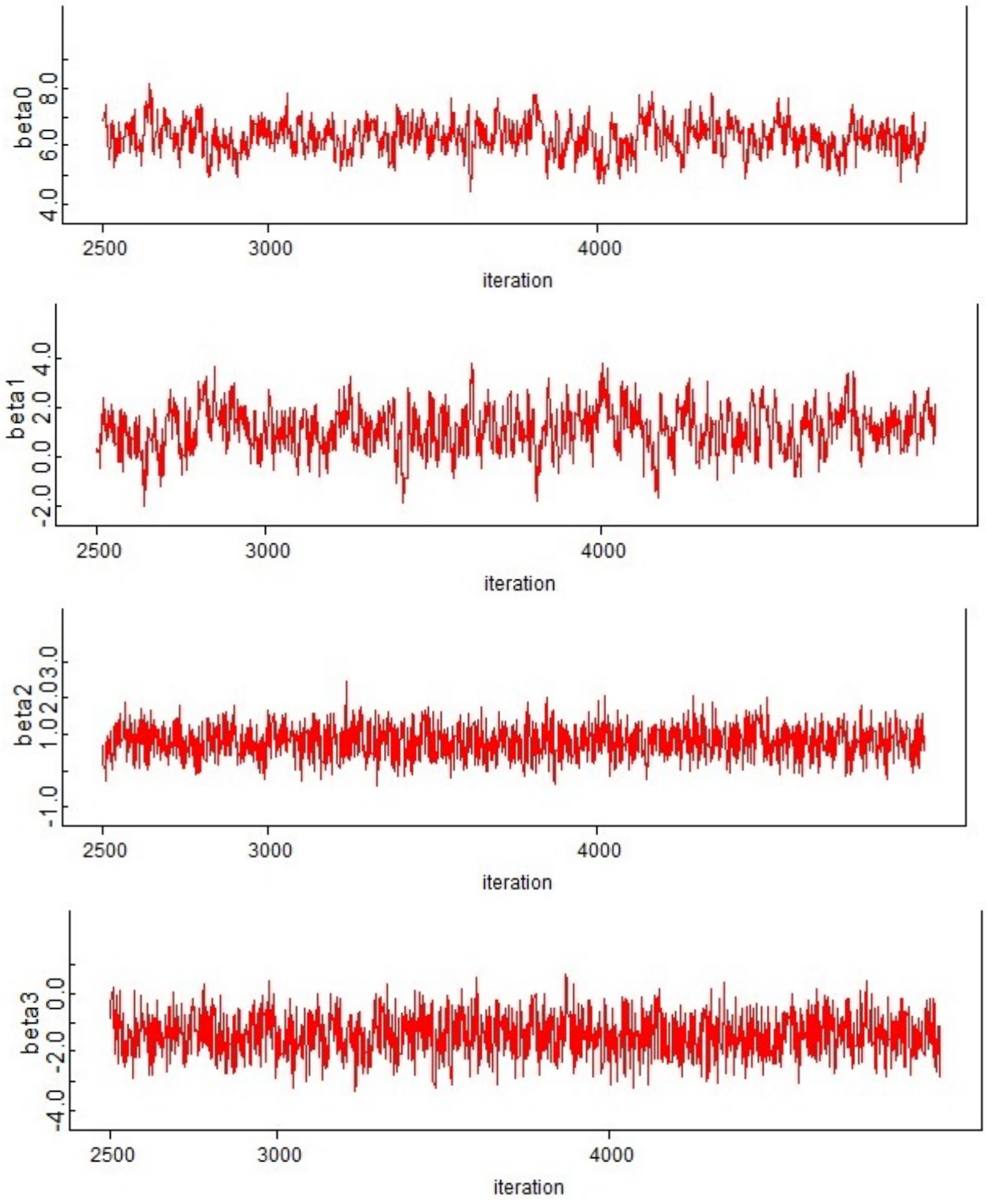


Figure 24: Convergence of Scale τ and σ_ω^2 Parameters

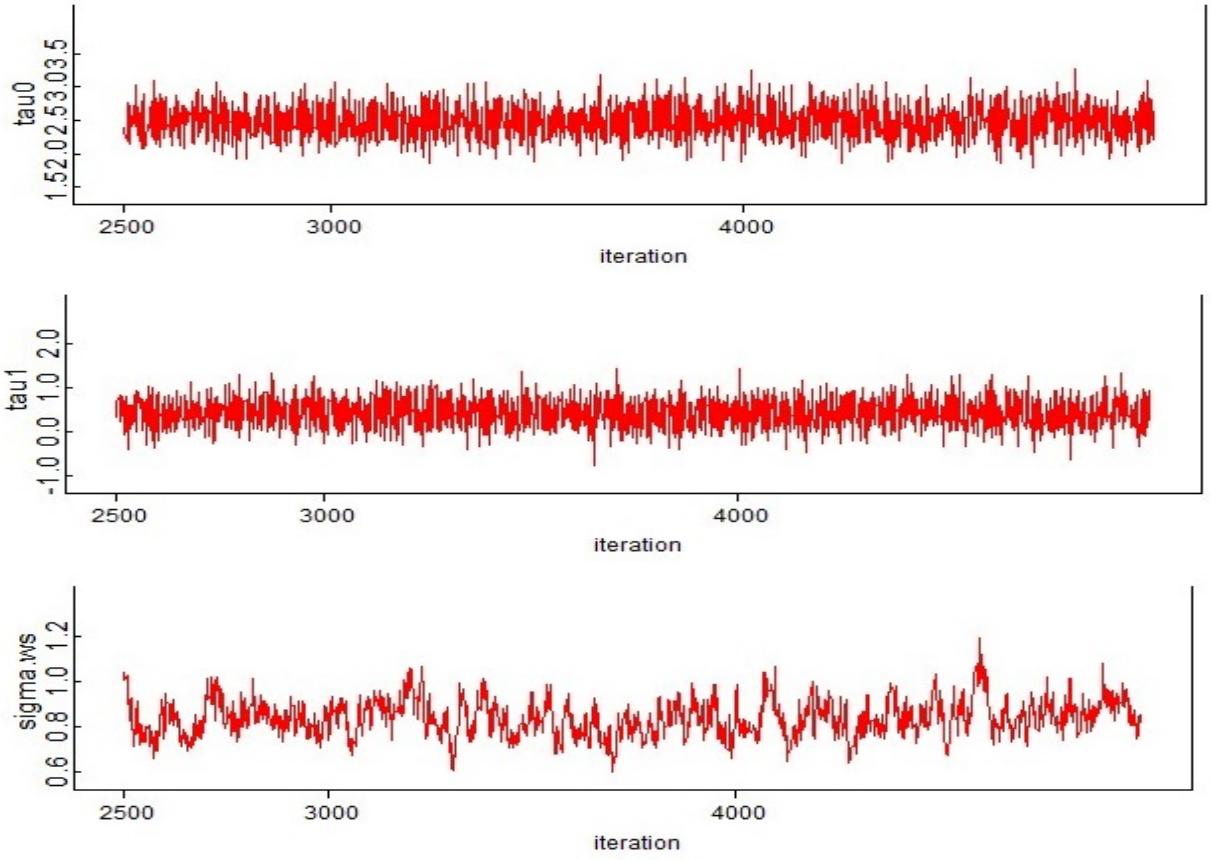
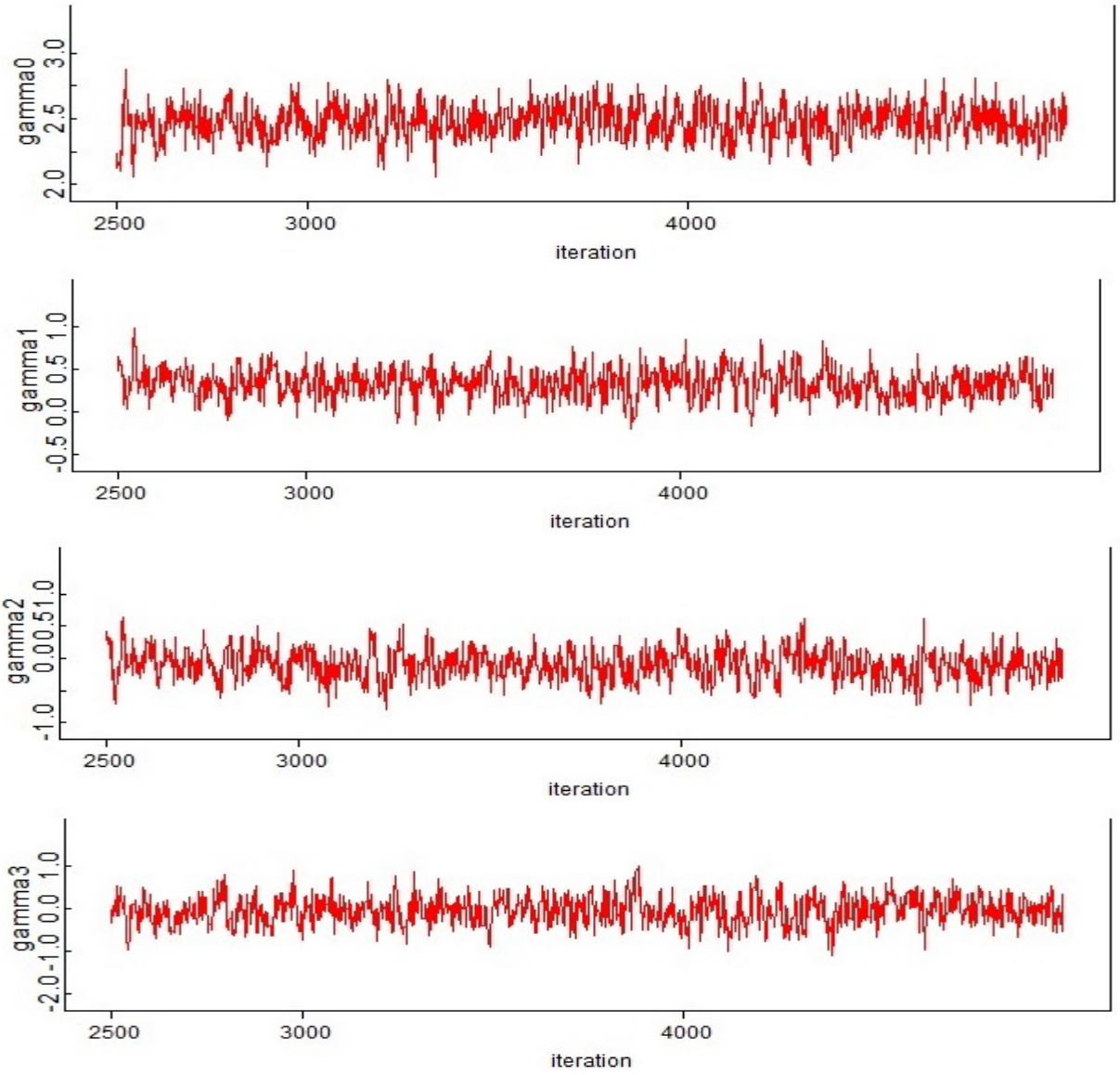


Figure 25: Convergence of Scale γ Parameters



Based on the results in Table 7 there appears to be a significant impact by the treatment-by-time location interaction term (β_3), but not by the treatment-by-time within-subject variance interaction term (γ_3) on the HRS17TOT variable. Analyzing the impact of time on the patients receiving the PCMM (β_2) reveals a significant impact on their HRS17TOT ratings. Over the course of the study these individuals on average have increasing scores over time which suggests a recurrence of their bipolar symptoms. When looking at the time-by-treatment interaction term the corresponding estimates are negative indicating a decrease in HRS17TOT scores over time. This suggests that those receiving the IRRI are having their depressive scores drop over time on average indicating fewer bipolar 1 symptoms. Further work into this phenomenon needs to be performed; however, one speculation would be the added psychological benefits from the new intervention.

The benefit of a MELS model is the capability of jointly modeling the location and scale effects. The results seem to indicate a marginal impact on variance by treatment (γ_1), but no impact due to time (γ_2 and γ_3). This supports the preliminary analysis shown in Figures 8 and 10 which suggested that the patients in IRRI had in general higher variability than those on PCMM. The results appear to indicate that IRRI may result in lower levels of bipolar I symptoms over time, but does so in a more volatile environment. No time effect was detected in the WS variance; however, this may be due to the low sample size preventing the effect from being detected.

4.5 COMPARISON WITH MML

The Bayesian MCMC approach has an advantage over the MML approach in that it can model data sets that the MML fails to do. One disadvantage is that it can be time consuming. Considering the advantages and disadvantages of each method it is necessary to compare their performance on data sets that can be fit with either method. The two data sets used were obtained from Hedeker and Nordgren (2013) and consist of an EMA data set (POSMOOD) and a longitudinal data set (REISBY).

4.5.1 Analysis of POSMOOD Data Set

The POSMOOD Data Set comes from a longitudinal, natural history study of adolescent smoking (Mermelstein et al., 2002). The study gave 515 students in either 8th or 10th grade a hand held computer for 7 consecutive days. They were prompted at random times throughout the day resulting in 17,514 total observations. The dependent variable is "posmood", which is a measure of the subject's positive mood and the explanatory variables are "alone" and "genderf". The model fit to the data is

$$\begin{aligned} \boldsymbol{Y_i}|\boldsymbol{\nu_i}, \boldsymbol{\beta}, \boldsymbol{\psi_i}, \boldsymbol{\omega_i} &\overset{ind.}{\sim} N_n(\beta_0 + \beta_1 alone_{ij} + \beta_2 genderf_i + 1_n\nu_i, \boldsymbol{\psi_i}) \\ \nu_i|\sigma^2_{\nu_i} &\overset{ind.}{\sim} N(0, \sigma^2_{\nu_i}), \quad i = 1, \ldots, N \end{aligned} \tag{4.10}$$

where the variances are modeled as follows

$$\begin{aligned} \sigma^2_{\nu_i} &= \exp(\tau_0 + \tau_1 genderf_i) \\ \boldsymbol{\psi_i} &= Diag(\sigma^2_{\epsilon_{i1}}, \ldots, \sigma^2_{\epsilon_{in}}) \\ \sigma^2_{\epsilon_{ij}} &= \exp(\gamma_0 + \gamma_1 alone_{ij} + \gamma_2 genderf_i + \omega_i) \\ \omega_i &\sim N(0, \sigma^2_\omega) \end{aligned}$$

Table 8: MML Estimation Results

Parameters	Estimate	SE	95% Confidence Limits	
β_0	7.015	.083	6.851	7.179
β_1	-.354	.0243	-.402	-.306
β_2	-.186	.110	-.401	.0292
τ_0	.383	.099	.189	.577
τ_1	-.064	.133	-.325	.196
γ_0	.757	.047	.665	.850
γ_1	.087	.025	.038	.135
γ_2	.221	.061	.102	.340
σ_ω	.627	.041	.579	.671

The MML estimation was performed in SAS using the method described in Section 2.2.2.1. The Bayesian results were obtained using the priors specified in 4.9. A 25,000 iteration burn-in period was used and inference was made on a sample of 2,500 obtained by selecting every 10th

Table 9: Bayesian MCMC Estimation Results

Parameters	Mean	SD	2.5p	MED	97.5p
β_0	7.184	.095	7.009	7.180	7.356
β_1	-.419	.025	-.047	-.419	-.365
β_2	-.159	.126	-.398	-.157	.074
τ_0	.532	.095	.356	.531	.723
τ_1	-.026	.128	-.276	-.024	.213
γ_0	.611	.028	.552	.611	.668
γ_1	.121	.029	.059	.121	.179
γ_2	.241	.028	.182	.240	.304
σ_ω	.796	.022	.750	.795	.845

sample. The MML approach took less than a minute to obtain an optimal solution; whereas, the Bayesian approach took $\approx$ 159 minutes.

Examining the estimation results presented in Tables 8 and 9 reveals comparable results between the two estimation techniques. The two methods have the same results in detecting nonzero parameter values. Both the β_2 and τ_1 parameters are not significant and this is shown by both estimation methods. In addition the 95% coverage intervals for the Bayesian MCMC technique all contain the MML estimate with the exception of σ_ω. The 95% confidence limits from the MML contain the mean value from the Bayesian MCMC technique for 5 out of the 9 parameters although from an absolute value perspective the mean estimate does not differ significantly nor does it provide a counter-intuitive result in terms of detecting a non-zero value. The only parameter that appears to be significantly different in the two estimation techniques is the random scale value. The Bayesian MCMC technique provides a larger individual subject variance in comparison to the MML approach with the upper 95% CL being .671 for MML in comparison to .750 for the lower 95% CL for the Bayesian approach. This result is expected as in Section 2.2.2.2 the estimate of σ_ω is consistently underestimated for the MELS model whereas the results from the Bayesian MCMC in Section 4.3 provide an unbiased estimate of σ_ω.

The diagnostic plots for the Bayesian MCMC (not shown) do not reveal any unusual patterns and show convergence for all parameters. The main difference between the two methods is the

amount of time it takes for the models to converge. This time difference is likely due to the amount of observations in the data set (17,514). This large number of observations drastically slowed the time it took to converge for the Bayesian MCMC approach, while it had no significant impact on the MML approach. Based on the amount of time to converge as well as both methods having comparable conclusions the MML would be the preferred method for the POSMOOD data set.

Table 10: Hedeker's Estimation Results

Parameters	Estimate	SE	95% Confidence Limits	
β_0	6.990	.081	6.831	7.149
β_1	-.370	.025	-.419	-.321
β_2	-.150	.109	-.364	.064
γ_0	.763	.047	.670	.856
γ_1	.081	.025	.032	.129
γ_2	.216	.061	.096	.336

The POSMOOD data set was used by Hedeker to demonstrate the use of the MIXREGLS program. This program written in FORTRAN is designed to implement a MELS model using maximum likelihood with the EM algorithm and a Newton-Raphson solution. It provides an alternative approach to the MML which is implemented in SAS. The results from MIXREGLS shown in Table 10 match closely with the two previously discussed approaches. Hedeker uses a slightly different model formulation for the between-subject structure and the random-scale effects, but this does not have any impact on the conclusions drawn from the mean and within-subject parameters.

Regardless of which method is used the results obtained provide the same interpretation for the data. Individuals who are alone tend to have lower positive mood scores and being female appears to indicate lower scores as well, although this effect is marginally significant in all three models. Female responses also tend to be less stable than males and being alone provides for greater variability in scores.

4.5.2 Analysis of REISBY Data Set

The REISBY Data Set (Reisby et al., 1977) is a psychiatric longitudinal data set of 66 depressed inpatients. The subjects were diagnosed with either endogenous or non-endogenous depression

and rated with the Hamilton rating (response variable). The study lasted 6 weeks and includes the type of depression (endog) and week (0-5) as the explanatory variables. There were 66 individuals in the study accounting for 396 observations (missing Hamilton Depression scores were present). The model fit to the data is

$$\begin{aligned} \boldsymbol{Y_i}|\boldsymbol{\nu_i}, \boldsymbol{\beta}, \boldsymbol{\psi_i}, \boldsymbol{\omega_i} &\overset{ind.}{\sim} N_n(\beta_0 + \beta_1 week_j + \beta_2 endog_i + \beta_3 week_j \times endog_i 1_n \nu_i, \boldsymbol{\psi_i}) \\ \nu_i|\sigma_\nu^2 &\overset{ind.}{\sim} N(0, \sigma_\nu^2), \quad i = 1, \ldots, N \end{aligned} \tag{4.11}$$

where the variances are modeled as follows

$$\begin{aligned} \sigma_\nu^2 &= \exp(\tau_0 + \tau_1 endog_i) \\ \boldsymbol{\psi_i} &= Diag(\sigma_{\epsilon_{i1}}^2, \ldots, \sigma_{\epsilon_{in}}^2) \\ \sigma_{\epsilon_{ij}}^2 &= \exp(\gamma_0 + \gamma_1 week_j + \gamma_2 endog_i + \omega_i) \\ \omega_i &\sim N(0, \sigma_\omega^2) \end{aligned}$$

Table 11: MML Estimation Results

Parameters	Estimate	SE	95% Confidence Limits	
β_0	22.251	.715	20.822	23.679
β_1	-2.265	.185	-2.635	-1.895
β_2	1.863	1.072	-.278	4.004
β_3	-.014	.272	-.557	.528
τ_0	2.169	.350	1.469	2.869
τ_1	.512	.450	-.387	1.412
γ_0	2.123	.227	1.670	2.577
γ_1	.185	.062	.062	.309
γ_2	.297	.232	-.166	.760
σ_ω	.605	.236	.3	.8

Table 12: Bayesian MCMC Estimation Results

Parameters	Mean	SD	2.5p	MED	97.5p
β_0	22.410	.755	20.830	22.430	23.910
β_1	-2.361	.187	-2.721	-2.361	-1.991
β_2	1.836	1.130	-.304	1.834	4.065
β_3	-.003	.267	-.532	.002	.516
τ_0	2.305	.368	1.597	2.301	3.026
τ_1	.565	.460	-.340	.563	1.457
γ_0	2.238	.216	1.812	2.237	2.661
γ_1	.175	.070	.036	.176	.308
γ_2	.221	.192	-.166	.225	.578
σ_ω	.557	.156	.303	.547	.871

The MML and Bayesian estimation procedures were the same ones used in the POSMOOD data set (see Section 4.5.1) and similar to the results in Tables 11 and 12 for the MML and Bayesian techniques are comparable. They have the same performance in detecting non-zero parameter values. In contrast to the POSMOOD data set the estimates from the REISBY data set all fall within the other method's 95% confidence interval. Comparing the MML's estimate with the mean of the Bayesian approach do not reveal any pattern in terms of one technique having consistently higher or lower values than the other one. In terms of choosing which method to use there are no major advantages nor disadvantages. The results are comparable and the convergence time for both is less than one minute. The comparable results for the Bayesian MCMC to the well-tested MML approach supports the use of this methodology and provides confidence to the accuracy of the model in its application to additional data sets.

Similar to the POSMOOD data set the REISBY data set was used by Hedeker to demonstrate the use of the MIXREGLS program. The results from MIXREGLS shown in Table 13 match closely with the two previously discussed approaches. Hedeker uses a slightly different model formulation for the random-scale effects, but this does not have any impact on the conclusions drawn from the mean and between/within-subject parameters.

The results for each method all provided the same conclusions. Over time (β_1) the subject's

Table 13: Hedeker's Estimation Results

Parameters	Estimate	SE	95% Confidence Limits	
β_0	22.378	.723	20.960	23.796
β_1	-2.295	.188	-2.663	-1.927
β_2	1.879	1.076	-.231	3.989
β_3	-.029	.268	-.749	.692
τ_0	2.198	.354	1.503	2.893
τ_1	.507	.458	-.391	1.405
γ_0	2.088	.236	1.624	2.551
γ_1	.192	.063	.069	.315
γ_2	.288	.245	-.193	.769

depression score would decrease on average regardless of the type of depression (β_2 and β_3). However, their scores would become more volatile over time (γ_1). Having endogenous depression may also result in more volatile responses, but this finding was marginal.

4.6 DISCUSSION

This chapter has investigated the use of a Bayesian MCMC approach in estimating the parameters of a MELS model. Hedeker's MML approach (implemented in SAS) is the prevailing estimation technique for the model and has been shown to provide unbiased estimates (see Section 2.2.2.2). However, due to the multidimensional and non-linear nature of the model there are numerous data sets and model forms where the MML approach fails to provide reasonable results. To address this issue I proposed the use of a Bayesian MCMC estimation technique. I was successfully able to use this technique to model a bipolar I data set that the MML technique was unable to. In Section 4.5 I compared the results from two data sets fitted with the Bayesian MCMC and MML approach. The results from both of the data sets revealed comparable estimates and similar clinical findings.

The Bayesian MCMC technique was used to model a bipolar data set (Section 4.4) to assess whether a new intervention had any impact on the recurrence of symptoms in both the mean and

variance of the individuals. The results showed a significant impact on the mean and variance of the HRS17TOT variable. Over time individuals began to develop more depressive symptoms, but individuals under the new intervention had this impact lessened and in some cases their Hamilton scores decreased over time. The major benefit of the MELS model is being able to jointly model the mean and variance. With this model we are able to identify a potential impact on the variance by the IRRI. Over time subjects receiving IRRI have lower depression scores in comparison to PCMM, but do so in a less stable environment (higher variability).

The major advantage that the Bayesian MCMC approach has over the MML is its applicability to a greater variety of data sets. Optimization over multidimensions fails to converge in many cases. Bayesian MCMC moves away from the reliability on optimization in order to obtain better model convergence. However, a disadvantage of this approach is the amount of computational time required. In modeling the POSMOOD data set the MML fit time was less than 30 seconds in comparison the Bayesian approach took $\approx$ 159 minutes. My recommendation would be to use the MML approach at first due to its smaller computational cost and use the Bayesian MCMC approach if necessary.

5.0 BAYESIAN MCMC ESTIMATION FOR MIXED-EFFECTS LOCATION-SCALE MODELS WITH SKEW-NORMAL ERRORS

In fitting a MELS model, Hedeker has assumed that the errors are normally distributed. For many mixed-effects models this is a reasonable assumption. Hedeker's application to an adolescent smoking EMA data set studies individuals over a short time frame; therefore, the normal and time-invariant assumption should be accurate. In the bipolar data set (discussed in Chapter 3) the Young Mania Rating (YOUNGTOT) scores are positively skewed and time-dependent. If the MELS model were to be fit to this data under the normal assumption then the results would have considerable bias. A common method of dealing with skewness is to transform the data. For positively skewed data log transformations remove this skewness; however, it also eliminates any heterogeneity. For skewed data sets where the focus lies on the mean parameters this would be an appropriate action. With the YOUNGTOT variable the clinical questions involve both the mean and variance. Through the use of the log transformation, questions involving treatment effects on the mean and variance of the data would not be addressed with a single model. Rather than use a log-transformation I propose the use of a skew-normal distribution in modeling the YOUNGTOT data. This distribution preserves the variability of the data and can be implemented with the MELS model in order to assess how the interventions affect the patient's mania scales in both their mean and variance.

The skew-normal distribution is not the only distribution that would be appropriate for skewed data. There is a whole family of skew-elliptical distributions such as the skew-t that are also available. I am proposing the use of the skew-normal distribution as a starting point in the expansion of the MELS model to other error distributions. The model has previously only been used with a normal error distribution, so by extending its use to a skew-normal which the normal is a special case of, I am allowing for a more flexible modeling approach.

5.1 MIXED-EFFECTS LOCATION-SCALE MODEL WITH SKEW-NORMAL ERRORS

The proposed MELS model with skew-normal errors in a hierarchical set-up for subject i with n_i observations is:

$$\boldsymbol{Y_i}|\boldsymbol{\nu_i}, \boldsymbol{\beta}, \boldsymbol{\psi_i}, \boldsymbol{\omega_i} \overset{ind.}{\sim} SN_{n_i}(\boldsymbol{X_i\beta} + \boldsymbol{1_{n_i}}\nu_i, \boldsymbol{\psi_i}, \boldsymbol{\Delta}) \tag{5.1}$$

$$\nu_i|\sigma^2_\nu \overset{ind.}{\sim} N(0, \sigma^2_\nu), \quad i = 1, \ldots, N \tag{5.2}$$

where the variances are modeled as follows

$$\begin{aligned}
\sigma^2_\nu &= \exp(\boldsymbol{u_i\tau}) \\
\boldsymbol{\psi_i} &= Diag(\sigma^2_{\epsilon_{i1}}, \ldots, \sigma^2_{\epsilon_{in_i}}) \\
\sigma^2_{\epsilon_{ij}} &= \exp(\boldsymbol{w_{ij}\gamma} + \omega_i) \\
\omega_i &\sim N(0, \sigma^2_\omega)
\end{aligned}$$

The same notation from Section 4.2 is used with the addition of $SN_{n_i}(\boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{\Delta})$ which represents a n_i dimensional skew-normal distribution with location parameter $\boldsymbol{\mu}$, variance parameter $\boldsymbol{\psi}$, and skewness parameter $\boldsymbol{\Delta}$. In the above model $\boldsymbol{\Delta} = Diag(\delta_1, \ldots, \delta_n)$, where δ_j is the skewness parameter for the jth time point.

5.1.1 Skew-Normal Distributions

The skew-normal distributions are part of a broad class of skew-elliptical distributions. Their usefulness stems from the desire to accommodate well-known and theoretically sound distributions in describing various data sets that arise in practical implementations. It was based on the idea to create a broad set of distributions that are flexible to the presence of skewness and easily implementable, which is obtained by transforming the error distributions rather than the data (Sahu et al., 2003).

In choosing which skew distribution to model, the nature of the data must be considered. In Section 4.4 the HRS17TOT was analyzed under the assumption of a normal error distribution

(supported through graphical analysis). Skew-elliptical distributions are a broad class of distributions, which include the skew-normal, that allows for a skewed, unimodal density with heavy tails. Graphical analysis of the YOUNGTOT reveals a unimodal, skewed distribution, but does not provide evidence for heavy tails. The presence of heavy tails would suggest a skew-student distribution; however without their presence I propose the skew-normal. This distribution accounts for the skewness present and allows for both the HRS17TOT and YOUNGTOT variables to be modeled with comparable distributions as desired. The normal distribution is a special case of the skew-normal distribution and as such provides a method of modeling the response variables as stemming from similar distributions.

The probability density function of the skew-normal distribution as derived by Sahu et al. (2003) and using multivariate elliptically symmetric distributions is:

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\psi},\boldsymbol{\Delta}) &= 2^n|\boldsymbol{\psi}+\boldsymbol{\Delta}^2|^{-1/2}\phi_n\Big\{(\boldsymbol{\psi}+\boldsymbol{\Delta}^2)^{-1/2}(\boldsymbol{y}-\boldsymbol{\mu})\Big\}P(\boldsymbol{V}>0) \\ \boldsymbol{V} &\sim N_n\{\boldsymbol{\Delta}(\boldsymbol{\psi}+\boldsymbol{\Delta}^2)^{-1}(\boldsymbol{y}-\boldsymbol{\mu}), \boldsymbol{I}-\boldsymbol{\Delta}(\boldsymbol{\psi}+\boldsymbol{\Delta}^2)^{-1}\boldsymbol{\Delta}\} \end{aligned} \tag{5.3}$$

where $\boldsymbol{\mu}$ is the location parameter vector, $\boldsymbol{\psi}$ is a positive-definite variance matrix, and $\boldsymbol{\Delta}$ is a diagonal matrix with skewness parameters $\delta_1,\ldots,\delta_n$. The distribution in equation 5.3 is denoted by $SN(\boldsymbol{\mu},\boldsymbol{\psi},\boldsymbol{\Delta})$. An alternative form of the skew-normal density function is:

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\psi},\boldsymbol{\Delta}) &= 2^n\phi_n(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\psi}+\boldsymbol{\Delta}\boldsymbol{\Delta}^T) \\ &\quad\times\Phi_n(\boldsymbol{\Delta}^T(\boldsymbol{\psi}+\boldsymbol{\Delta}\boldsymbol{\Delta}^T)^{-1}(\boldsymbol{y}-\boldsymbol{\mu})|\boldsymbol{0},(\boldsymbol{I}_n+\boldsymbol{\Delta}^T\boldsymbol{\psi}^{-1}\boldsymbol{\Delta})^{-1}) \end{aligned} \tag{5.4}$$

The mean and variance are:

$$E(\boldsymbol{Y})=\boldsymbol{\mu}+\left(\frac{2}{\pi}\right)^{1/2},\quad cov(\boldsymbol{Y})=\boldsymbol{\psi}+\left(1-\frac{2}{\pi}\right)\boldsymbol{\Delta}^2$$

The stochastic representation of the skew-normal distribution obtained from Arellano-Valle et al. (2007) (shown in Proposition 1) will be useful in simulating and adapting the OPENBUGS software (implementable in R) for fitting the specified model.

Figure 26: SN Dist with (+) Skew

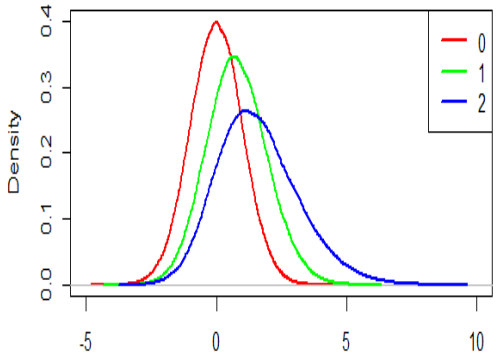
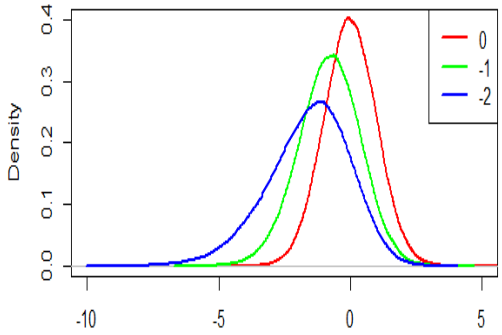


Figure 27: SN Dist With (-) Skew



Proposition 1 *Let* $\boldsymbol{Y} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{\Delta})$. *Then*

$$\boldsymbol{Y} \overset{d}{=} \boldsymbol{\Delta}|\boldsymbol{X_0}| + \boldsymbol{X_1} \tag{5.5}$$

where $\boldsymbol{X}_0 \sim N_n(\boldsymbol{0}, \boldsymbol{I}_n)$, $\boldsymbol{X}_1 \sim N_n(\boldsymbol{\mu}, \boldsymbol{\psi})$ *and are independent.*

Using the stochastic representation from Proposition 1 produces the skew-normal graphs in Figures 26 and 27 for parameters $\mu = 0$, $\Sigma = 1$, and varying skewness parameters (δ).

5.2 BAYESIAN INFERENCE WITH SKEW-NORMAL DISTRIBUTIONS

The conditional pdf on the sample $\boldsymbol{y} = (\boldsymbol{y_1^T}, \ldots, \boldsymbol{y_N^T})$ is

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\nu}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\omega}) = \prod_{i=1}^{N} & 2^{n_i} \phi_{n_i}(\boldsymbol{y_i}|\boldsymbol{X_i\beta} + \boldsymbol{1}_{n_i}\boldsymbol{\nu_i}, \boldsymbol{\psi_i} + \boldsymbol{\Delta^2}) \\ & \times \Phi_n(\boldsymbol{\Delta}(\boldsymbol{\Psi}_i + \boldsymbol{\Delta}^2)^{-1}(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{1}_n\nu_i)|\boldsymbol{0}, (\boldsymbol{I}_n + \boldsymbol{\Delta\Psi}^{-1}\boldsymbol{\Delta})^{-1}) \end{aligned} \tag{5.6}$$

Using the same methods to derive Lemma 1 from Section 4.2 I derived the following marginal distribution (see APPENDIX for proof):

Lemma 2 *Let* $\boldsymbol{Y_i} = \boldsymbol{X_i\beta} + \mathbf{1}_{\boldsymbol{n_i}}\nu_i + \boldsymbol{\epsilon_i}$*, where* $\nu_i \sim N(0, \sigma_\nu^2)$ *and* $\boldsymbol{\epsilon_i} \sim SN_{n_i}(\mathbf{0}, \boldsymbol{\psi_i}, \boldsymbol{\Delta}_i)$ *are independent. The variances are denoted by* $\sigma_\nu^2 = \sigma_\nu^2(\tau)$ *and* $\boldsymbol{\psi_i} = \boldsymbol{\psi_i}(\boldsymbol{\gamma}, \omega)$ *with* $\omega \sim N(0, \sigma_\omega^2)$ *and skewness parameter* $\boldsymbol{\Delta}_i = Diag(\delta_{i1}, \ldots, \delta_{in_i})$*. Then the marginal distribution of* $\boldsymbol{Y_i}$ *is*

$$f_{\boldsymbol{Y_i}}(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\Delta}, \sigma_\omega^2) = \int_{\mathbb{R}} 2^{n_i}\phi_{n_i}(\boldsymbol{y}|\boldsymbol{X\beta}, \boldsymbol{\Gamma}) \times \Phi_{n_i}(\boldsymbol{A}(\boldsymbol{y} - \boldsymbol{X\beta})|\mathbf{1}_{n_i}\mu + \boldsymbol{A}\mathbf{1}_{n_i}\mu, \boldsymbol{B} + \Lambda\boldsymbol{A}\boldsymbol{J}_{n_i}\boldsymbol{B}^T)\phi(\omega|0, \sigma_\omega^2)d\omega \tag{5.7}$$

where

$$\boldsymbol{\Gamma} = \boldsymbol{\Psi_i} + \boldsymbol{\Delta}_i^2 + \sigma_\nu^2\boldsymbol{J}_{n_i}, \quad \Lambda = \left(\frac{1}{\sigma_\nu^2} + \frac{1}{\sum_{j=1}^{n_i}(\sigma_{\epsilon_{ij}}^2 + \delta_j)}\right)^{-1}$$

$$\boldsymbol{A} = \boldsymbol{\Delta}_i(\boldsymbol{\Psi_i} + \boldsymbol{\Delta}_i^2)^{-1}, \quad \boldsymbol{B} = (\boldsymbol{I}_{n_i} + \boldsymbol{\Delta}_i\boldsymbol{\Psi}^{-1}\boldsymbol{\Delta}_i)^{-1}$$

$$\mu = \Delta_i\mathbf{1}'_{n_i}(\boldsymbol{\Psi}_{n_i} + \boldsymbol{\Delta}_i)^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})$$

In order to estimate the model specified in equations 5.1 and 5.2 the Bayesian MCMC procedure described in Section 4.1 will be utilized with the priors from 4.9 and added priors for the skew parameters

$$\delta_h \sim N(\delta_{h0}, \sigma_{h0}^2) \quad h = 1, \ldots, n \tag{5.8}$$

Using the marginal pdf from Lemma 2 and the specified priors gives the joint posterior distribution for the specified parameters given the observed sample $\boldsymbol{y}$

$$\begin{aligned}\pi(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \sigma_\omega^2|\boldsymbol{y}) \propto \prod_{i=1}^{N}\Bigg[&\int_{\mathbb{R}} 2^{n_i}\phi_{n_i}(\boldsymbol{y}|\boldsymbol{X\beta}, \boldsymbol{\Gamma}) \\ &\times \Phi_{n_i}(\boldsymbol{A}(\boldsymbol{y} - \boldsymbol{X\beta})|\mathbf{1}_{n_i}\mu + \boldsymbol{A}\mathbf{1}_{n_i}\mu, \boldsymbol{B} + \Lambda\boldsymbol{A}\boldsymbol{J}_{n_i}\boldsymbol{B}^T)\phi(\omega|0, \sigma_\omega^2)d\omega\Bigg] \\ &\times \prod_{k=1}^{p}\phi(\beta_k|\beta_{k0}, \sigma_{k0}^2) \times \prod_{l=1}^{r}\frac{1}{\tau_{lb0} - \tau_{la0}} \times \prod_{m=1}^{q}\frac{1}{\gamma_{mb0} - \gamma_{ma0}} \\ &\times \frac{b_0^{a_0}}{\Gamma(a_0)}\sigma_\omega^{2(-a_0-1)}\exp\left(\frac{-b_0}{\sigma_\omega^2}\right) \times \prod_{h=1}^{n}\phi(\delta_h|\delta_{h0}, \sigma_{h0}^2)\end{aligned} \tag{5.9}$$

This posterior distribution does not have a closed form solution therefore a Bayesian MCMC with Gibbs sampling will be used to determine the posterior distribution (described in Section 4.1).

5.3 SIMULATION RESULTS

To assess the accuracy of the model a simulation similar to the one in Section 4.3 will be used with the addition of a skew parameter that will be assumed to be constant throughout the study. The model simulated is

$$\begin{gathered}\boldsymbol{Y_i}|\boldsymbol{\nu_i}, \boldsymbol{\beta}, \boldsymbol{\psi_i}, \boldsymbol{\omega_i} \overset{ind.}{\sim} SN_{n_i}(\beta_0 + \beta_1 trt_i + \beta_2 week_j^* + \beta_3 trt_i \times week_j^* + 1_n \nu_i, \boldsymbol{\psi_i}, \boldsymbol{\Delta_i}) \\ \nu_i|\sigma_\nu^2 \overset{ind.}{\sim} N(0, \sigma_\nu^2), \quad i = 1, \ldots, N\end{gathered} \tag{5.10}$$

where the variances and skew parameters are modeled as follows

$$\begin{gathered}\sigma_\nu^2 = \exp(\tau_0 + \tau_1 \times trt_i) \\ \boldsymbol{\psi_i} = Diag(\sigma_{\epsilon_{i1}}^2, \ldots, \sigma_{\epsilon_{in}}^2) \\ \sigma_{\epsilon_{ij}}^2 = \exp(\gamma_0 + \gamma_1 trt_i + \gamma_2 week_j^* + \gamma_3 trt_i \times week_j^* + \omega_i) \\ \omega_i \sim N(0, \sigma_\omega^2) \\ \boldsymbol{\Delta_i} = Diag(\delta, \ldots, \delta)\end{gathered}$$

The week variable is transformed to $week_j^* = week_j/100$ to aid in model estimation.

Conjugate priors will be specified as in Section 4.3 with the addition of the prior distribution for the skew parameter

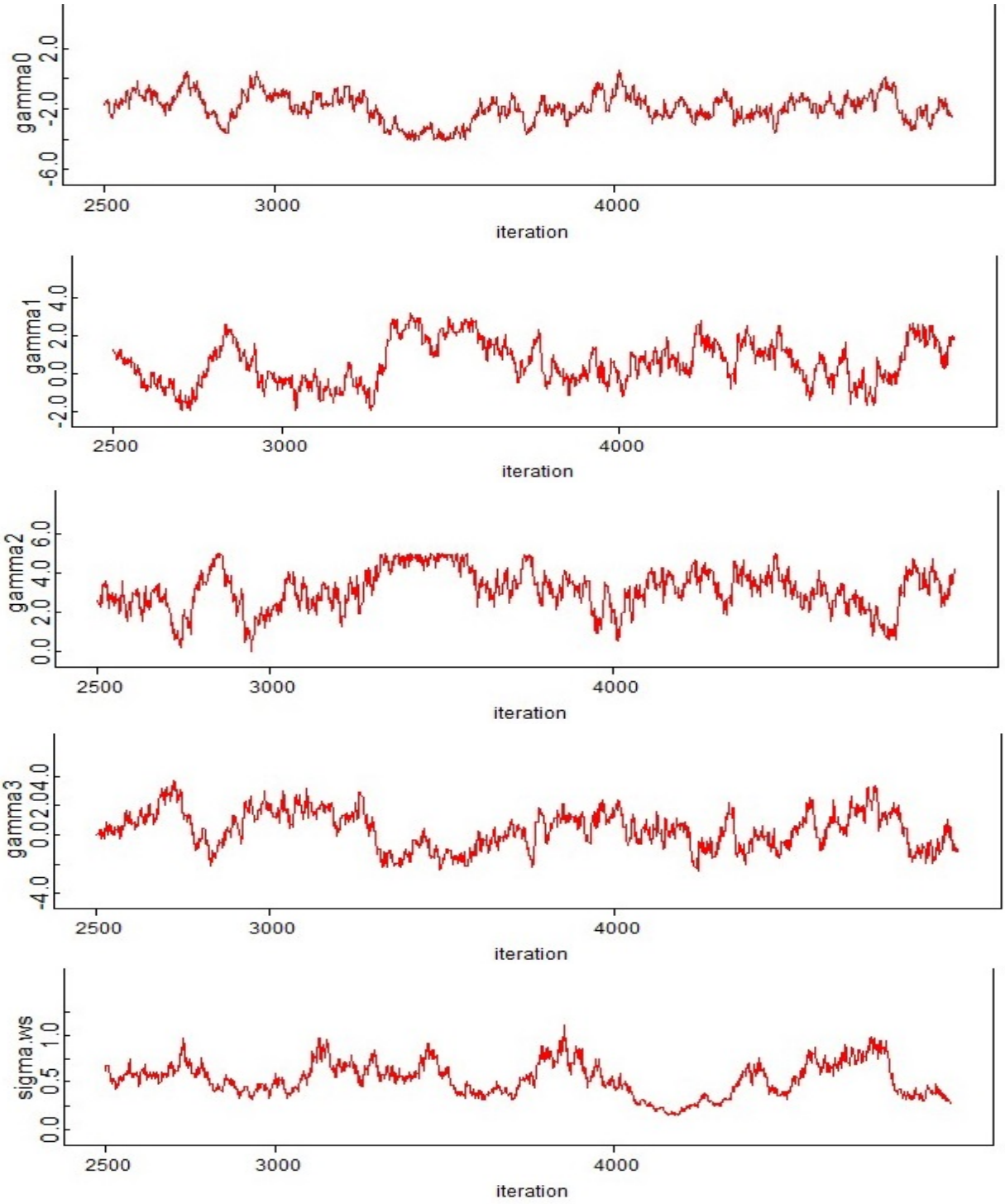
$$\begin{aligned}&\beta_k \sim N(0, 100) \quad k = 0, 1, 2, 3 \\ &\tau_l \sim U(-5, 5) \quad l = 0, 1 \\ &\gamma_m \sim U(-5, 5) \quad m = 0, 1, 2, 3 \\ &\sigma_\omega^2 \sim IG(10, 10) \\ &\delta \sim N(0, 100)\end{aligned} \tag{5.11}$$

The simulation results provided in Table 14 are based on a Bayesian draw with a 25,000 iteration burn-in period and a sample size of 2,500 obtained by selecting every 10th sample. One run takes approximately 15 minutes on a Samsung NP305E5A laptop.

Table 14: Simulation Results Based on 100 Simulations

Parameters	True Value	Estimate	SE	Bias	St. Bias	Coverage Prob (%)	Reject H_0 (%)
β_0	2	1.987	.252	-.013	-5.046	96	100
β_1	.5	.470	.364	-.030	-8.124	96	25
β_2	.2	.192	.327	-.008	-2.573	95	10
β_3	-.2	-.185	.436	.015	3.504	96	3
τ_0	.7	.703	.281	.003	1.041	95	75
τ_1	-.1	-.081	.389	.019	4.876	97	6
γ_0	-1.4	-1.731	.717	-.331	-46.129	92	77
γ_1	.2	.089	.903	-.110	-12.23	95	5
γ_2	2.5	2.702	.760	.202	26.650	97	96
γ_3	.5	.659	1.025	.159	15.522	95	7
σ_ω	.6	.629	.197	.029	14.502	95	100
δ	5	5.025	.155	.025	16.432	97	100

Figure 28: Convergence of Scale γ and σ_ω Parameters



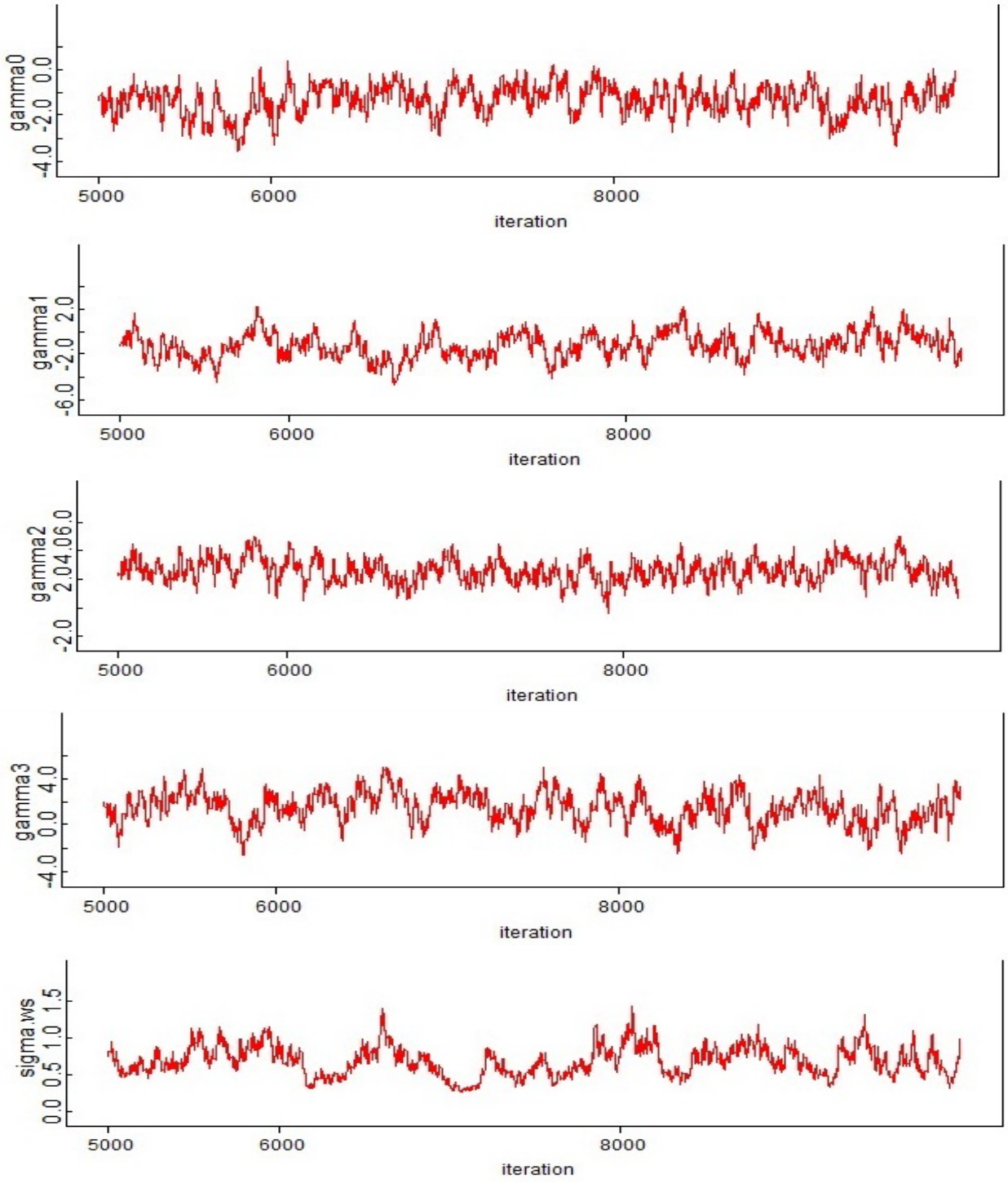
The results in Table 14 show coverage probabilities that are approximately the same as the nominal value for a 95% confidence interval. Similar to the results in Section 4.3 the results have low power which may be due to low sample size, observations per subject, effect size, or a combination of the three. However, looking at the diagnostic plots in Figure 28 for the scale parameters reveals that the inferences were not made on a stationary distribution. This problem can be alleviated by extending the number of iterations for a Bayesian MCMC draw. In addition the presence of skewness in the model structure may lead to greater difficulty in estimating, so reducing the hyperparameters in particular for σ_ω is recommended.

The scenario that led to consistence convergence used a 50,000 iteration burn-in period and a sample size of 5,000 obtained by sampling every 10th value. In addition the hyperparameters for σ_ω were lowered from 10 to 5. Each run took approximately 30 minutes on a Samsung NP305E5A laptop.

Table 15: Simulation Results Based on 50 Simulations – new priors

Parameters	True Value	Estimate	SE	Bias	St. Bias	Coverage Prob (%)	Reject H_0 (%)
β_0	2	1.985	.244	-.015	-6.102	96	100
β_1	.5	.441	.418	-.059	-14.104	92	20
β_2	.2	.175	.310	-.025	-7.989	100	14
β_3	-.2	-.178	.473	.022	4.574	92	6
τ_0	.7	.733	.254	.033	13.142	94	80
τ_1	-.1	-.166	.388	-.066	-17.077	96	4
γ_0	-1.4	-1.790	.808	-.390	-48.288	90	72
γ_1	.2	.121	1.018	-.079	-7.752	94	4
γ_2	2.5	2.646	.812	.146	18.022	94	80
γ_3	.5	.661	1.041	.161	15.483	98	6
σ_ω	.6	.728	.165	.128	77.476	96	100
δ	5	5.078	.159	.078	48.922	94	100

Figure 29: Convergence of Scale γ and σ_ω Parameters



The results in Table 15 show coverage probabilities that are comparable to the 95% nominal confidence intervals and low power in rejecting the null hypothesis for the time-by-treatment interaction for both the location and scale structures. This matches with the results from the previous scenario; however, the diagnostics plots in Figure 29 show random dispersion throughout, which improves upon the results from Figure 28. There is some concern with the results for σ_ω; however, based on the nominal estimates this should not impact the inferences of the other parameters.

Using the skew-normal distribution provides a method of dealing with the skewness present in the data without having to transform the data, but does so at a computational cost. In comparing the computation time required to fit a skew-normal versus a normal model using Bayesian MCMC the skew-normal takes twice as long. In the case of the σ_ω parameter some concern over convergence remains; whereas, the normal specification did not have this concern. To assess the impact of the different specifications and the robustness of the Bayesian MCMC estimation I simulated data arising from a skew-normal distribution (as shown previously) and modeled the data using a normal distribution (as shown in Chapter 4). The prior distributions used are the ones used to obtain the results in Table 15 without the prior for the skewness parameter. The simulations results based on 50 simulations are provided in Table 16.

Table 16: Simulation Results from Normal Model Specification

Parameters	True Value	Estimate	SE	Bias	St. Bias	Coverage Prob (%)	Reject H_0 (%)
β_0	2	5.795	.304	2.795	919.41	0	100
β_1	.5	.551	.409	.051	12.47	94	16
β_2	.2	.273	.363	.073	20.11	92	12
β_3	-.2	-.198	.583	.002	.34	92	10
τ_0	.7	.595	.368	-.105	-28.53	92	52
τ_1	-.1	-.012	.467	.088	18.84	96	4
γ_0	-1.4	2.094	.123	3.494	2,840.65	0	100
γ_1	.2	-.061	.166	-.261	-157.23	70	8
γ_2	2.5	.318	.172	-2.182	-1,268.60	0	38
γ_3	.5	.324	.243	-.176	-72.43	92	18
σ_ω	.6	.410	.082	-.190	-231.71	34	100

The results in Table 16 reveal considerable bias in most of the estimates. The coverage prob-

abilities are 0% for three parameters (β_0,γ_0, and γ_2) which can lead to erroneous clinical implications. In a MELS model the intercept term for the within-subject structure (γ_0) determines not only the statistical significance of the other WS parameters, but also their practical (or clinical) significance. Assuming no effect by the random scale effect (i.e. ω_i = 0) the following scenarios will be used to determine WS variance estimates at various time points:

1. $\gamma_0 = -1.4$, $\gamma_1 = .2$, $\gamma_2 = 2.5$, $\gamma_3 = .5$ (true values only)
2. $\gamma_0 = 2.094$, $\gamma_1 = .2$, $\gamma_2 = 2.5$, $\gamma_3 = .5$ (different γ_0)

Table 17: WS Variances at Various Configurations

	Week #	Intervention	WS Contribution	WS Variance	WS St. Dev.
Scenario 1	0	PCMM	-1.4	0.246	.496
	0	IRRI	-1.2	0.301	.549
	52	PCMM	-.1	0.905	.951
	52	IRRI	.36	1.433	1.198
Scenario 2	0	PCMM	2.094	8.117	2.849
	0	IRRI	2.294	9.914	3.149
	52	PCMM	3.394	29.785	5.457
	52	IRRI	3.854	47.181	6.869

The difference in the two scenarios is the WS intercept term (baseline for the PCMM group). The first scenario has a negative value of -1.4; whereas, the second scenario uses the estimated value from the normal model specification of 2.094. Due to the nature of the exponential function the output values for negative input values are closer together than those with positive input values. This can be seen in Table 17 where a difference of one year in the first scenario predicts an increase in the WS standard deviation for the PCMM by .7 units which contrasts with the second scenario where under the same condition the standard deviation increases by 2.6 units. In both the normal and skew-normal model specifications these increases may be found to be statistically significant; however, due to the model misspecification in the normal model the practical significance may be overstated. When fitting MELS models the Bayesian MCMC estimation approach is not robust to model misspecifications and careful consideration must be placed on the estimates to ensure accurate identification of potential clinical signficance.

5.4 APPLICATION TO BIPOLAR DATA

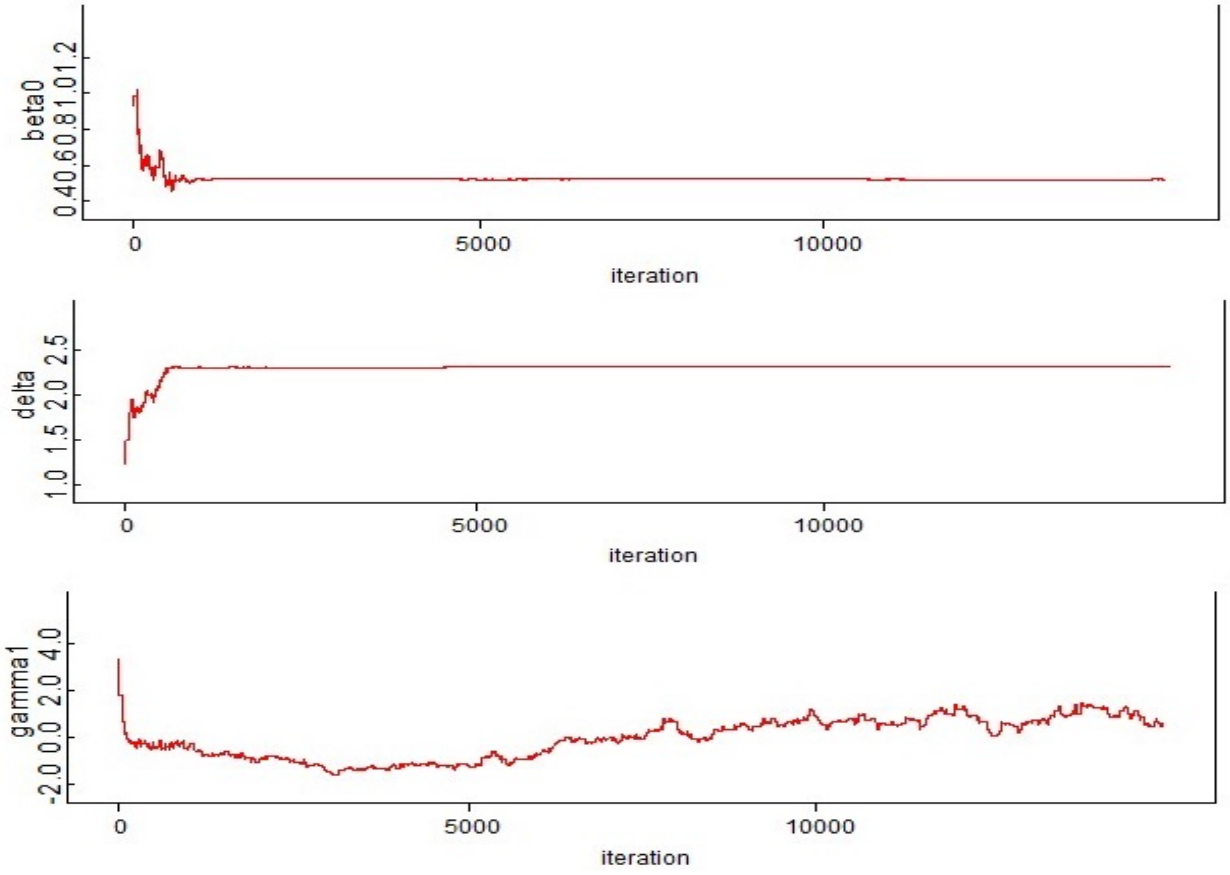
5.4.1 Skew-Normal Fit

Based on the results in the previous section the following prior distributions provide estimates with stationary distributions and coverage probabilities that match with their nominal confidence intervals:

$$\begin{aligned}
&\beta_k \sim N(0, 10) \quad k = 0, 1, 2, 3 \\
&\tau_l \sim U(-5, 5) \quad l = 0, 1 \\
&\gamma_m \sim U(-5, 5) \quad m = 0, 1, 2, 3 \\
&\sigma_\omega^2 \sim IG(5, 5) \\
&\delta \sim N(0, 100)
\end{aligned} \tag{5.12}$$

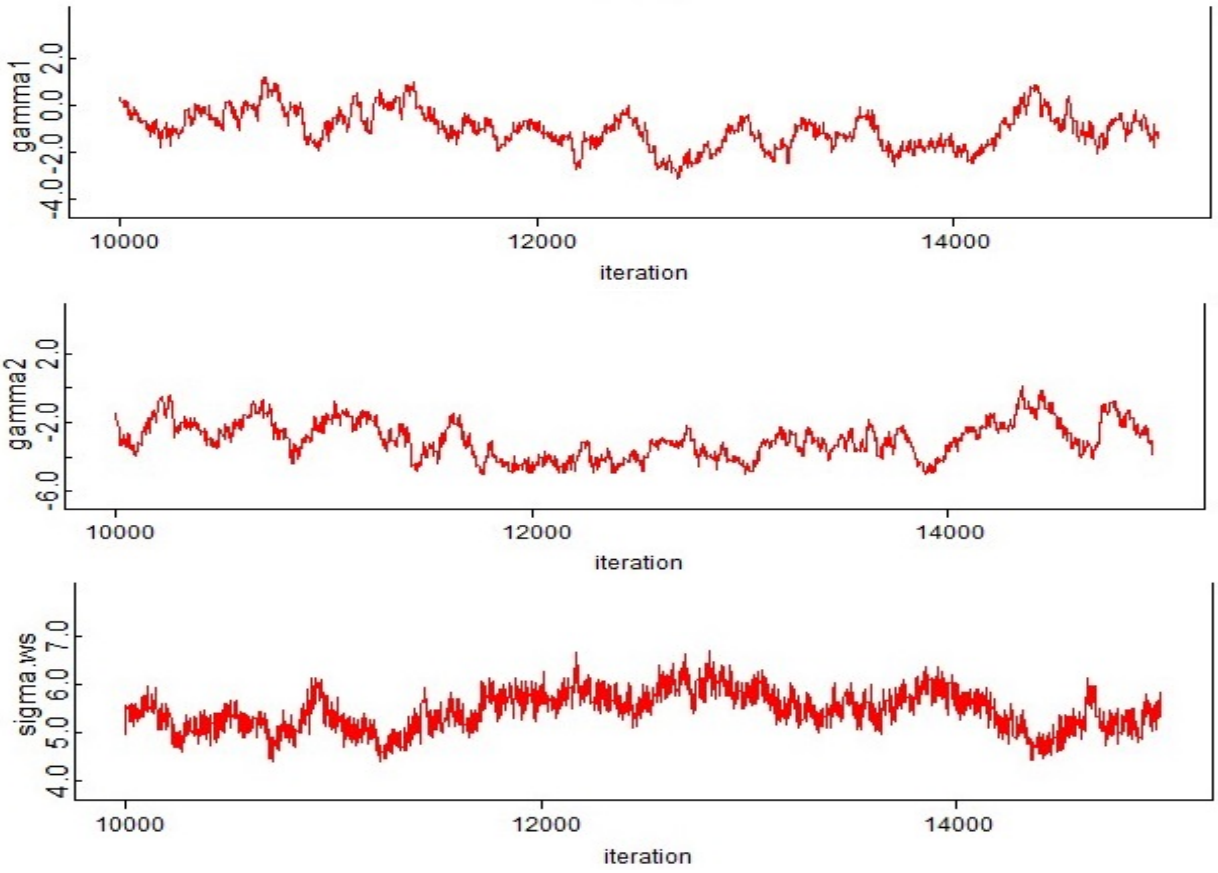
The simulated results were applied to a balanced data set;whereas, the the bipolar data set is unbalanced. This may result in a larger amount of iterations before convergence is obtained. After fitting the model in OPENBUGS to the YOUNGTOT variable the distributions of the mean parameters (β_k) and the skewness parameter (δ) converge relatively quickly, but those of the within-subject variance (γ_m) do not (see Figure 30).

Figure 30: Convergence Results After 15,000 Iterations



To obtain stationary distributions a burn-in period of 100,000 iterations is used and inference is based on 5,000 samples obtained by selecting every 10th iteration. The diagnostic plots in Figure 31 reveal stationary distributions for parameters that did not converge after 15,000 iterations.

Figure 31: Convergence Results for γ_1,γ_2, and σ_ω



Based on the results shown in Table 18 the new intervention (IRRI) has a statistically significant effect on both the mean and variability of the Young Mania Rating Scale. The location interaction term (β_3) indicates that in comparison to those on the PCMM the subjects receiving the new treatment will have a higher mania score on average. The variance interaction term (γ_3) suggests greater variability in scores for the IRRI. Based on these results it would appear that the IRRI results in higher and more variable mania scores on average. The estimate for the random scale effect (σ_ω) is on a higher magnitude than the parameter estimates associated with the covariates which indicates a greater amount of variability that goes unexplained by the model.

Looking at the parameter estimates reveals a statistically significant effect, but does not reveal

Table 18: Estimation Results for YOUNGTOT

Parameters	Mean	SD	2.5p	MED	97.5p
β_0	.528	.002	.513	.527	.544
β_1	-.494	.002	-.505	-.493	-.476
β_2	-.040	.001	-.056	-.037	-.024
β_3	.080	.004	.060	.076	.095
τ_0	-.113	.203	-.491	-.118	.313
τ_1	-.070	.283	-.618	-.069	.493
γ_0	-4.783	.216	-4.992	-4.859	-4.310
γ_1	-1.022	.775	-2.18	-1.043	.475
γ_2	-2.998	1.066	-4.502	-3.045	-1.013
γ_3	3.074	1.153	.495	3.187	4.852
σ_ω	5.447	.391	4.741	5.412	5.997
δ	2.299	.001	2.283	2.299	2.314

a clinical impact. The mean estimate for the intercept term for the within-subject variance (γ_0) is -4.783 indicating that at baseline on average a patient would have a WS variance of .008 (or .09 WS standard deviation). After 100 weeks the variance is expected to drop to .00044 (or .02 WS standard deviation). These values which are comparable to those for the IRRI do not indicate any noticeable treatment and time effects. As such based purely on the results from the application of a MELS model to the YOUNGTOT variable, there does not appear to be any noticeable difference between the two treatments. In Table 18 the standard deviations for the estimates of the location parameters ($\beta_0, \beta_1, \beta_2$, and β_3) are 0 due to the difference in convergence between the location and scale parameters. The location parameters converge in only 15,000 iterations; whereas, it takes almost 100,000 iterations for the scale parameters to converge. Due to the difference in the number of iterations it takes to converge caution is urged when basing any statistical inference on the location parameters cannot be made outside of their point estimates.

5.4.2 Normal Fit

Section 5.3 showed with simulated data that when skew-normal data is fitted with a normal model the results are biased and may result in erroneous results. The YOUNGTOT data while skewed

may not necessarily follow from a skew-normal distribution, so a normal model will be fit to assess their differences.

The model fit will resemble the one used for the HRS17TOT variable in section 4.4 with the prior distributions used in section 5.4.1. Using a 100,000 iteration burn-in period and a 5,000 sample size obtained by sampling every 10th iteration produces the results in Table 19.

Table 19: Estimation Results for YOUNGTOT Based on Normal Model

Parameters	Mean	SD	2.5p	MED	97.5p
β_0	.957	.000	.943	.957	.972
β_1	.276	.000	.261	.276	.291
β_2	.002	.000	-.012	.002	.017
β_3	-.002	.000	-.016	-.002	.013
τ_0	.629	.192	.277	.626	1.017
τ_1	.178	.280	-.365	.178	.721
γ_0	-4.756	.222	-4.992	-4.816	-4.238
γ_1	-1.727	.732	-3.117	-1.715	-.718
γ_2	.496	.729	-1.048	.534	1.712
γ_3	-.637	1.511	-2.428	-.868	2.053
σ_ω	10.43	.213	10.06	10.44	10.81

The results between the skew-normal and normal fit reveal significant differences between the two model fits. The most noticeable is the difference in estimates for the random scale effect σ_ω, 5.45 for the skew-normal and 10.43 for the normal fit. As previously discussed large values of the variance for the random scale effect result in the model failing to detect significant effects among the within-subject variance parameters. For the skew-normal fit the time parameters γ_2 and γ_3 are both significant, but they are not in the normal fit. Looking at the mean parameters shows a different conclusion for the treatment parameter β_1. The normal fit expects a higher YOUNGTOT offset for the IRRI in comparison to the PCMM group; whereas, the skew-normal fit expects the opposite. The random location effects are significant for the normal, but not for the skew-normal.

In selecting the appropriate model in Bayesian applications one of the most widely used criterion is the Deviance Information Criterion (DIC) (Gelman et al., 2014b). The DIC proposed by Spiegelhalter et al. (2002) serves as a semiformal method of identifying models that best explain the observed data. This is done by minimizing the uncertainty around the observations. When

choosing which model based on DIC the one with the lowest value will be the "best" for the data at hand. The DIC for the skew-normal fit is -28,030 and -7,046 for the normal fit. This supports the use of the skew-normal model over the normal model that was hypothesized.

5.5 DISCUSSION

In applying the MELS model it is important to preserve the original structure of the data as inference is desired both on the mean and variance components. When dealing with skewed data, data transformation fails to preserve the variability of the original response variable. I proposed a skew-normal model to account for this. The added complexity of this model results in higher number of iterations in order to obtain a stationary distribution in comparison to the normal model formulation from Chapter 4. Failure to specify a skew-normal model and using a simpler model such as the normal one may result in faster convergence, but results in highly biased estimates.

The application of the skew-normal model to the Young Mania Rating Scale revealed statistically significant effects by both intervention and time, but no practical results were noted. The results indicated that any WS variability was largely unexplained as evident by the value of the σ_ω estimate. The estimate of the skewness parameter δ reveals moderate positive skewness which based on previous works suggests a more accurate performance by the skew-normal over the normal model. This would also suggest along with the results for the HRS17TOT variable in section 4.4 that the participants regardless of intervention spend more time in the depressive state than the manic state.

6.0 EFFECT OF SAMPLE SIZE AND PARAMETER CONFIGURATIONS ON ESTIMATION OF TREATMENT EFFECTS

6.1 ASSESSING THE LIMITATIONS OF THE CURRENT MODEL

Section 4.3 assessed the accuracy of the proposed model (equation 6.1), through simulations on an data set that resembles the bipolar data discussed in Chapter 3. Here 100 subjects are assumed to be randomly assigned to two interventions: treatment and control with 50 subjects in each intervention. Measurements would be taken over a span of 108 weeks (approximately 2 years) spaced 9 weeks apart (approximately 2 months). The model was fitted using Bayesian MCMC with the priors specified in equation 6.2. The results from a simulation with 100 runs are given in table 20.

$$\begin{aligned} \boldsymbol{Y_i}|\boldsymbol{\nu_i}, \boldsymbol{\beta}, \boldsymbol{\psi_i}, \boldsymbol{\omega_i} &\overset{ind.}{\sim} N_n(\beta_0 + \beta_1 trt_i + \beta_2 week_j^* + \beta_3 trt_i \times week_j^* + 1_n \nu_i, \boldsymbol{\psi_i}) \\ \nu_i|\sigma_\nu^2 &\overset{ind.}{\sim} N(0, \sigma_\nu^2), \quad i = 1, \ldots, N \end{aligned} \tag{6.1}$$

where the variances are modeled as follows

$$\begin{aligned} \sigma_\nu^2 &= \exp(\tau_0 + \tau_1 \times trt_i) \\ \boldsymbol{\psi_i} &= Diag(\sigma_{\epsilon_{i1}}^2, \ldots, \sigma_{\epsilon_{in}}^2) \\ \sigma_{\epsilon_{ij}}^2 &= \exp(\gamma_0 + \gamma_1 trt_i + \gamma_2 week_j^* + \gamma_3 trt_i \times week_j^* + \omega_i) \\ \omega_i &\sim N(0, \sigma_\omega^2) \end{aligned}$$

The week variable is transformed to $week_j^* = week_j/100$ to aid in model estimation.

$$\beta_k \sim N(0, 100) \quad k = 0, 1, 2, 3$$
$$\tau_l \sim U(-5, 5) \quad l = 0, 1$$
$$\gamma_m \sim U(-5, 5) \quad m = 0, 1, 2, 3$$
$$\sigma^2_\omega \sim IG(10, 10) \tag{6.2}$$

Table 20: Simulation Results from Scenario 2 in Section 4.3 with 100 runs

Parameters	True Value	Estimate	Bias	St. Bias	Cov. Prob. (%)	Reject H_o (%)
β_0	2	1.989	-.011	-5.272	93	100
β_1	.5	.523	.023	7.57	89	45
β_2	.2	.199	-.001	-.639	95	34
β_3	-.2	-.202	-.002	-1.325	98	22
τ_0	.7	.698	-.002	-1.106	94	91
τ_1	-.1	-.131	-.031	-10.483	95	6
γ_0	-1.4	-1.419	-.019	-11.920	90	100
γ_1	.2	.220	.020	8.900	89	24
γ_2	2.5	2.495	-.005	-2.097	93	100
γ_3	-.5	-.494	.006	2.066	93	41

Under the specified conditions the Bayesian MCMC estimated approach provides unbiased estimates of all the parameters; however, the number of instances where the approach would reject H_o (i.e. that the parameter is not significant) varies from 5% to 100%. In detecting whether there is a treatment-by-time effect in both the mean and variance β_3 and γ_3 respectively, the percentage of times where the effect is detected is only 22% and 41%. In designing an experiment with the aim of detecting the treatment-by-time effects, these values would not provide investigators with enough confidence to move forward with a similar experiment. Due to the time commitment involved for both the patients and researchers, appropriate sample sizes are required in both the number of subjects and number of observations per subject. This chapter will determine the appropriate sample sizes required to detect this difference with sufficient confidence using the following 8 scenarios:

1. 50 subjects, 7 observations/subject (Week Max = 54)

2. 50 subjects, 13 observations/subject (Week Max = 108)
3. 50 subjects, 19 observations/subject (Week Max = 162)
4. 50 subjects, 25 observations/subject (Week Max = 216)
5. 100 subjects, 7 observations/subject (Week Max = 54)
6. 100 subjects, 13 observations/subject (Week Max = 108) (**base case**)
7. 100 subjects, 19 observations/subject (Week Max = 162)
8. 100 subjects, 25 observations/subject (Week Max = 216)

6.2 COMPARING OPENBUGS AND PYMC3

In fitting the model using Bayesian MCMC I previously used the OPENBUGS software. This software is one of many that can be used and in this chapter, along with exploring the effects of varying observations on the percent of runs that reject H_o for the treatment-by-time interaction effect, the use of another software to fit the model will be investigated. The software studied is the PyMC3 Python module implemented in the Python programming language (Salvatier et al., 2016).

A major benefit of using PyMC3 over OPENBUGS is that the source code is readily available which allows anyone to modify it as desired. This provides the flexibility to adapt the procedures for different prior distributions and optimization techniques that are not supported by OPENBUGS. A disadvantage in using PyMC3 is the lack of "ready-to-use" features in specifying the number of iterations in fitting a Bayesian MCMC. The developers of OPENBUGS have already accounted for burn-in iterations as well as thinning methods. Both of these are not built in to the current version of PyMC3. In OPENBUGS I applied a thinning method by selecting one out of every 50 observations along the Markov chain random samples in order to deal with the dependency Markov chains have among successive draws. In order to implement this feature in PyMC3 at the present state this would require recording the entire sequence of random draws for each parameter and then manually selecting every 50th iteration. Depending on the number of iterations required, this could lead to large data sets that could result in memory issues. The rest of the analysis in this chapter will not use a thinning approach in selecting the samples to use. I would suggest future researchers to adapt the PyMC3 module to handle this limitation.

The results in table 21 are from 100 runs in Python and use the same model structure and prior distributions specified in section 6.1 in PyMC3 (the Python code is provided in the Appendix). The only difference is the number of iterations used for a burn-in period (5,000) and the iterations used for inference (1,000). These iterations also did not use a thinning procedure (discussed previously). The simulation results using OPENBUGS is provided in table 20 and the results from PyMC3 are in table 21.

Table 21: Simulation Results from Scenario 2 in Section 4.3 fitted with PyMC3

Parameters	True Value	Estimate	Bias	St. Bias	Cov. Prob. (%)	Reject H_0 (%)
β_0	2	1.981	-.019	-9.578	89	100
β_1	.5	.523	.023	8.019	91	55
β_2	.2	.196	-.004	3.362	93	32
β_3	-.2	-.179	.021	13.177	97	16
τ_0	.7	.688	-.012	-5.695	96	92
τ_1	-.1	-.063	.037	12.137	95	4
γ_0	-1.4	-1.233	.166	99.359	60	100
γ_1	.2	.189	-.010	4.383	82	39
γ_2	2.5	2.471	-.028	-13.512	86	100
γ_3	-.5	-.451	.049	16.702	89	48

When comparing the metrics in the two tables, especially the standardized bias column, those from OPENBUGS outperform those from PyMC3 with respect to providing unbiased estimates. There is not much difference between the coverage and reject H_o %. The results from PyMC3 may be due to the the lack of thinning and the amount of iterations used. It also may be due to a potential misspecification of a model. It is worth noting in the literature that relatively few examples exist for fitting a mixed-effects model with this Python module; whereas, there are numerous cases in OPENBUGS. Despite these differences especially with the parameter γ_0, the results for the treatment-by-time interaction parameters (β_3 and γ_3) of interest appear to be unbiased estimates of the true parameter value. The coverage probability and reject H_o columns are practically identical between the two tables. Simulation studies performed on the other 7 scenarios discussed in section 6.1 (not shown) revealed identical results with respect to the nature of the estimates. This provides confidence in using the PyMC3 module to determine effective sample sizes for detecting the treatment-by-time effects.

6.3 SAMPLE SIZE SPECIFICATIONS

Figures 6.3, 33, and 6.3 provide the % of simulations the reject H_o (i.e. there is no parameter effect) for the 10 parameters (excluding the random scale component) specified in the model. These graphs show how this percentage varies when the number of subjects in the study is either 50 or 100, as well as observations varying from 7 to 25 per subject. The parameters of interest are β_3 and γ_3. These parameters measure the treatment-by-time interaction effect for the mean and variance components, respectively. As can be seen in the graph for β_3 the reject % increases from 30% for 7 observations per subject to 40% for 25 observations per subject when a study has 100 subjects. For γ_3 the increase is more dramatic. For studies where there are either 50 or 100 subjects, the number of simulations that would detect an effect increases from 20% to 100% with 100 subjects reaching 100% at a quicker rate than the one with 50.

While both parameter effects may not be detected at the same rate, a mixed-effects location-scale model would be able to detect the variance treatment-by-time interaction effect for small effects. For the base case where there were only 13 observations per subject the number of successful treatment-by-time interactions in the variance component was low (39 out of 100 simulations). However, by increasing the number of observations per subject from 13 to 19 for 100 subjects, this rate increases to over 90%. These simulations suggest that within the confines of a typical clinical trial, it is possible to simultaneously measure the impacts of both the mean and variance without having large amounts of repeated measurements per observation. Hedeker's motivating studies relied on over 30 measures per subject over a span of one week. However, my results show that such effects can be detected with about half as many measures over a multi-year clinical trial.

In looking at the figures for β_1, β_2, τ_0, and τ_1 there are some concerns since the graphs are not monotonically increasing. We would expect that as the per subject sample size increases the treatment and time effects would be more detectable not less. These results may be due to the random variation resulting from using 100 runs for each simulation. I would expect that by increasing the number of runs the current behavior would be removed. Another potential remedy may be to apply a monotone smoothing condition. I would suggest future researchers to look into and attempt to remedy this behavior before proceeding.

Figure 32: Reject Hypothesis % for β parameters

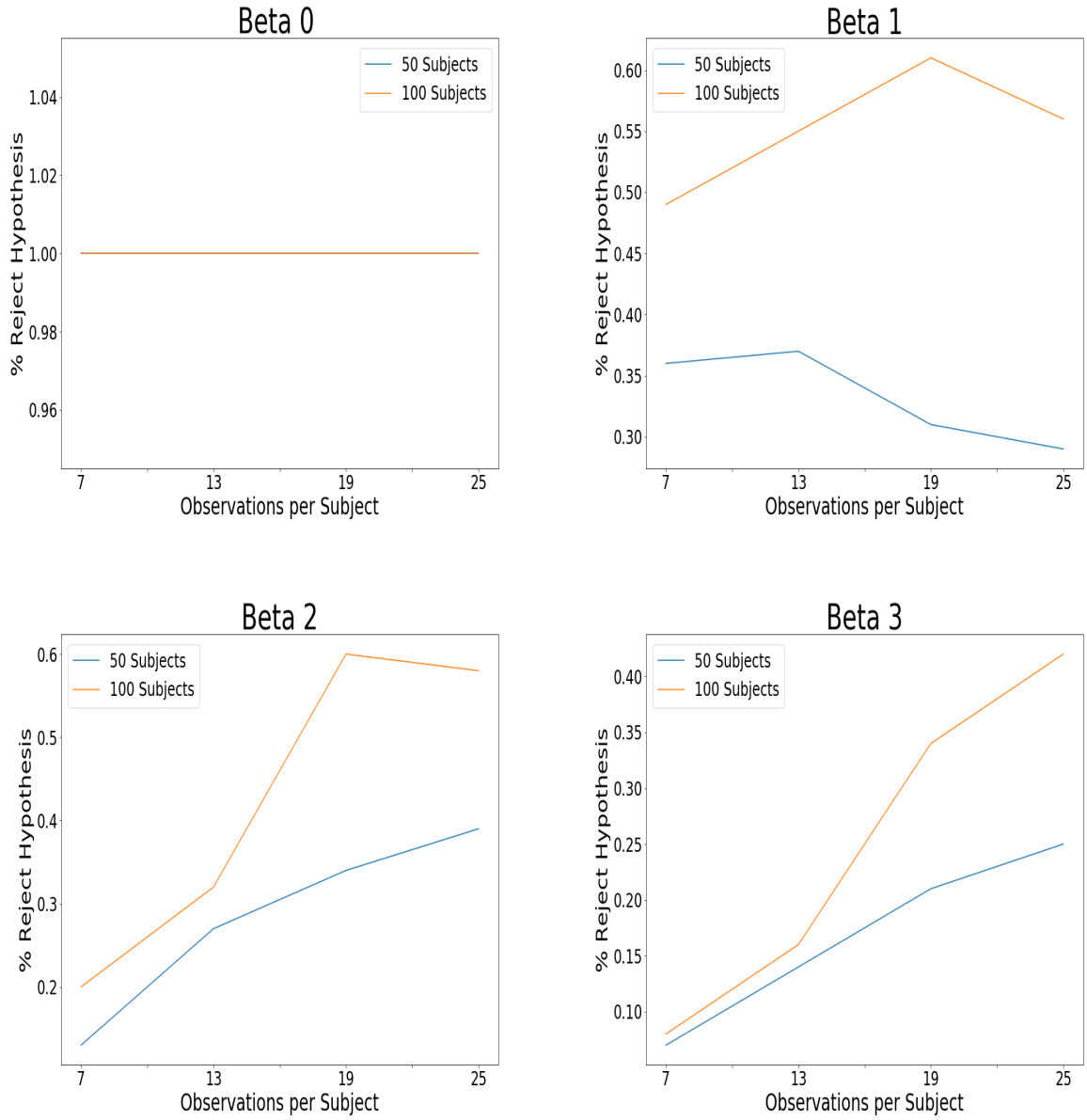


Figure 33: Reject Hypothesis % for γ parameters

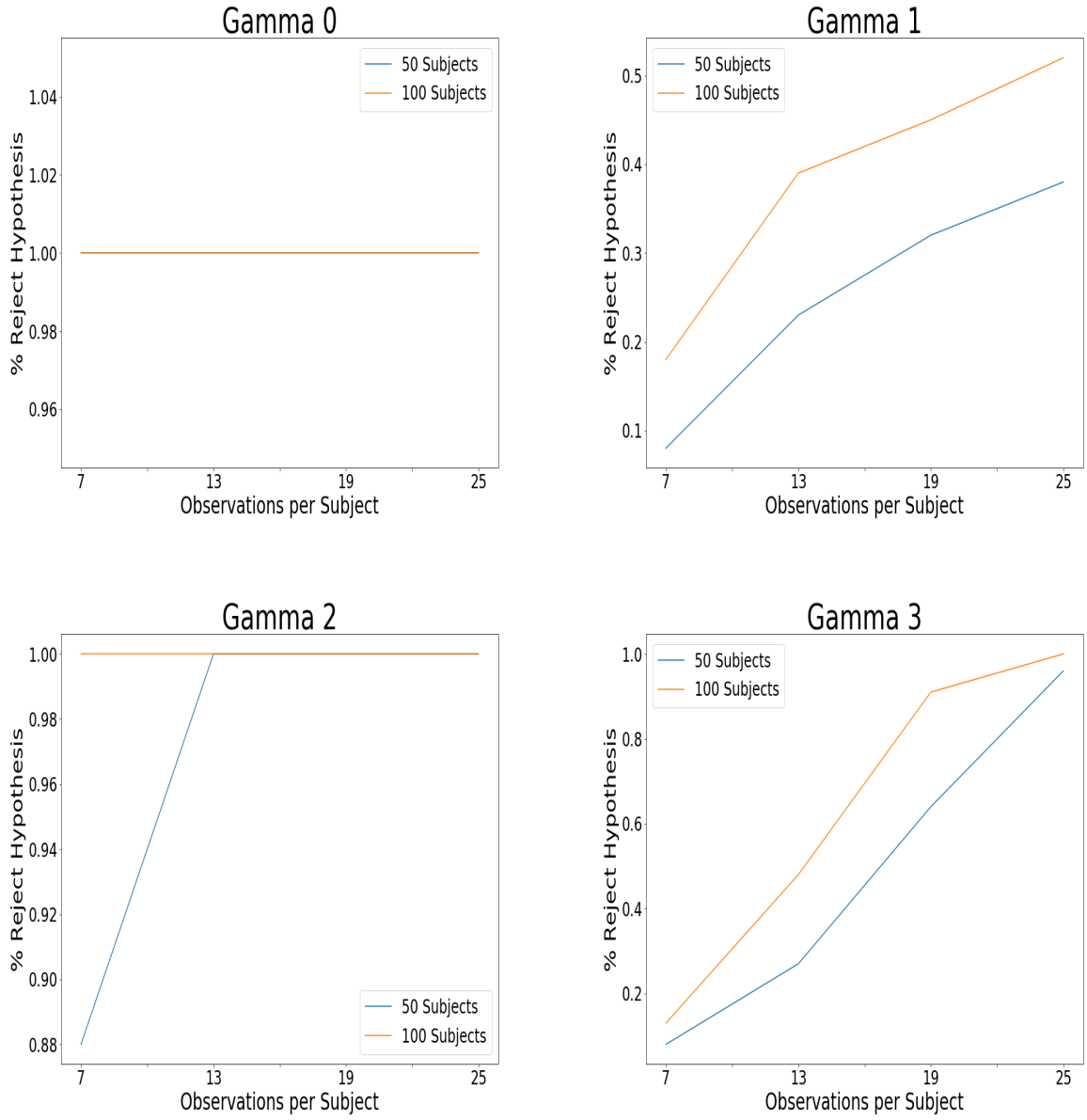
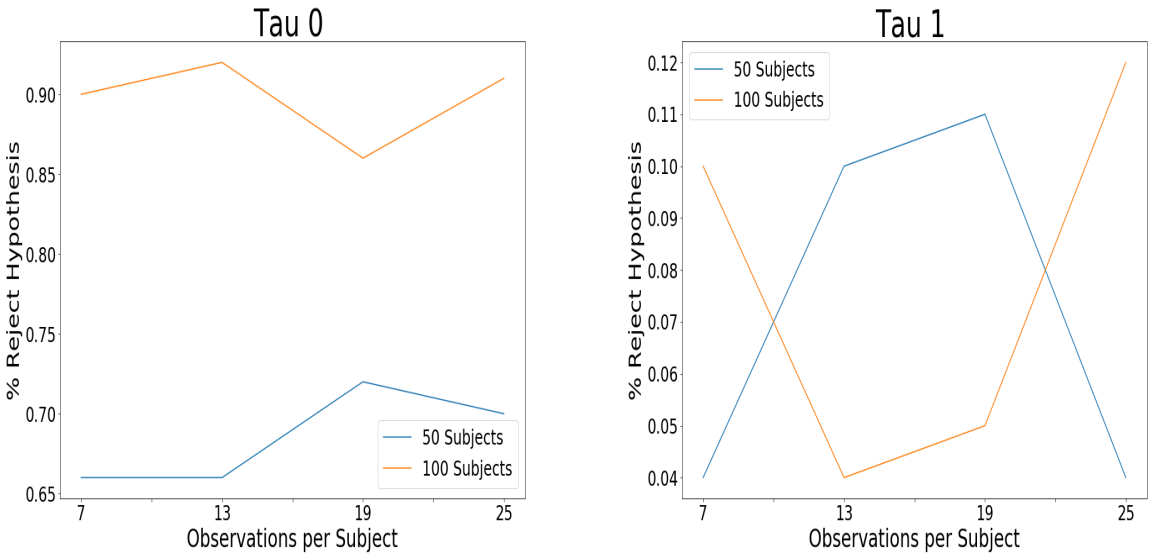


Figure 34: Reject Hypothesis % for τ parameters



7.0 FUTURE WORK

The MELS model is a relatively unused model with recent literature coming from Hedeker and his students. I speculate that this is due to the difficulty involved in fitting this model which is currently done using MML. As discussed in this dissertation the MML technique provides unbiased estimates for data similar to Hedeker's; however, the amount of similar data sets that provide results when a MELS model is estimated with MML is limited. My Bayesian MCMC approach is able to expand the scope of data sets that can be used with this model. One extension is to a clinical trial where patients with bipolar symptoms are followed for almost 2 years.

The benefit of the Bayesian MCMC over the MML approach is its ability to provide estimates in cases where we previously were not. A major disadvantage is the amount of time it takes to fit the model. Depending on the computing system the MML can fit the model in under a minute compared to almost 20 minutes for the Bayesian approach. This issue can be mitigated by obtaining more computing power. Another solution is investigating the different softwares available for fitting these models. This dissertation looked at two different programming languages/software in OPENBUGS and Python; however, there are countless others available that may be optimal in fitting nonlinear models of this nature.

Another area that we can extend the MELS model into is bivariate modeling. Similar to Pugach et al. (2014), this model can be extended to modeling two response variables simultaneously. The HRS17TOT and YOUNGTOT variables (see chapter 3) are modeled separately in this dissertation; however, they are outcomes from the same data and should be modeled simultaneously. Future work should look into this bivariate aproach while allowing for the variables to follow two different error distributions such as the normal and skew-normal. Additional research also needs to be conducted in improving the convergence rates for the Bayesian MCMC approach. This can be done by specifying different prior distributions as well as transforming the parameters. Section

5.4.2 showed an instance where the location parameters achieved a stationary distribution after only 10,000 iterations; however, it took another 90,000 iterations in order for the scale parameters to achieve a stationary distribution. This discrepancy led to zero standard deviations for the location parameters. This prevents the use of hypothesis testing and confidence intervals to perform statistical tests.

Chapter 6 showed the required sample sizes to detect treatment-by-time interaction for specified scale parameters at various observation configurations (looking at different subjects and different number of observations per subject). This was only done for one combination of parameters. If another set of parameters were used the current process is too time-consuming. A better approach for calculating sample sizes in order to detect the treatment-by-time parameter will be investigated in future work.

APPENDIX A

SAS CODE FOR 2.2.2.2

```
PROC NLMIXED DATA= SIM_INPUT GCONV=1E-12 METHOD=GAUSS
QTOL = .01 TECH=NEWRAP;
    PARMS B0 = &&INTERCEPT_&_JJ_. B1 = &&TREATMENT_&_JJ_. B2 = &&WEEK_&_JJ_.
    TAU0 = &TAU_EST GAMMA0 = &GAMMA_EST TAU1 = 0 GAMMA1 = 0 COV= 0 VARU2 =
    0;

    Z = B0 + B1*TREATMENT + B2*WEEK + U1;
    VARU1 = EXP(TAU0 + TAU1*TREATMENT);
    VARE = EXP(GAMMA0 + GAMMA1*TREATMENT + U2);
    LOGLIK = -0.5*LOG(2*3.1415926*VARE)-0.5*(RESPONSE - Z)**2/VARE;

    MODEL RESPONSE  GENERAL(LOGLIK);
    RANDOM U1 U2  NORMAL([0,0],[VARU1,COV,VARU2]) SUBJECT=SUBJECT;

  ODS OUTPUT PARAMETERESTIMATES = SIM.PAREST_&_JJ_ CONVERGENCESTATUS = CON-
VERG; RUN;
```

APPENDIX B

USEFUL RESULTS FROM ARELLANO ET AL. (2007)

LEMMA A1 *Let* $\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x} \sim N_p(\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{x}, \boldsymbol{\Sigma})$ *and* $\boldsymbol{X} \sim N_q(\boldsymbol{\eta}, \boldsymbol{\Omega})$. *Then*,

$$\begin{aligned}\phi_p(\boldsymbol{y}|\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{x}, \boldsymbol{\Sigma})\phi_q(\boldsymbol{x}|\boldsymbol{\eta}, \boldsymbol{\Omega}) = \phi_p(\boldsymbol{y}|\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{\eta}, \boldsymbol{\Sigma} + \boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{A}^T) \\ \times \phi_q(\boldsymbol{x}|\boldsymbol{\eta} + \boldsymbol{\Lambda}\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{m}\boldsymbol{u} - \boldsymbol{A}\boldsymbol{\eta}), \boldsymbol{\Lambda}),\end{aligned}$$

where $\boldsymbol{\Lambda} = (\boldsymbol{\Omega}^{-1} + \boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}$.

LEMMA A2 *Let* $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. *Then for any fixed k-dimensional vector* $\boldsymbol{a}$ *and* $k \times n$ *matrix* $\mathbf{B}$,

$$E[\Phi_k(\boldsymbol{a} + \boldsymbol{B}\boldsymbol{Y}|\boldsymbol{\eta}, \boldsymbol{\Omega})] = \Phi_k(\boldsymbol{a}|\boldsymbol{\eta} - \boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{\Omega} + \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^T).$$

APPENDIX C

PROOFS OF LEMMAS 1 AND 2

Proof 1 (for Lemma 1) *Begin by dropping subscript i and substituting $n_i = n$. We can write the marginal density as*

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\tau},\boldsymbol{\gamma},\sigma_\omega^2) &= \int_{\mathbb{R}^2} f(\boldsymbol{y}|\nu,\boldsymbol{\beta},\boldsymbol{\gamma},\omega) f(\nu|\boldsymbol{\tau}) f(\omega|\sigma_\omega^2) d\nu d\omega \\ &= \int_{\mathbb{R}^2} \phi_n(\boldsymbol{y}|\boldsymbol{X\beta} + \boldsymbol{1_n}\nu, \boldsymbol{\psi_i}) \phi(\nu|0,\sigma_\nu^2) \phi(\omega|0,\sigma_\omega^2) d\nu d\omega \end{aligned} \tag{C.1}$$

Using Lemma A1 from Arellano-Valle et al. (2007) the first two pdfs of equation C.1 can be rewritten as

$$\phi_n(\boldsymbol{y}|\boldsymbol{X\beta} + \boldsymbol{1_n}\nu, \boldsymbol{\psi_i}) \phi(\nu|0,\sigma_\nu^2) = \phi_n(\boldsymbol{y}|\boldsymbol{X\beta}, \boldsymbol{\psi_i} + \sigma_\nu^2 \boldsymbol{J_n}) \phi(\nu|\Lambda \boldsymbol{1_n^T \psi_i^{-1}} (\boldsymbol{y} - \boldsymbol{X\beta}), \Lambda)$$

where $\boldsymbol{J_n}$ is a n x n matrix of 1s and

$$\Lambda = \left(\frac{1}{\sigma_\nu^2} + \sum_{j=1}^{n} \frac{1}{\sigma_{\epsilon_{ij}}^2} \right)^{-1}$$

We can rewrite C.1 as:

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\tau},\boldsymbol{\gamma},\sigma_\omega^2) &= \int_{\mathbb{R}^2} \phi_n(\boldsymbol{y}|\boldsymbol{X\beta}, \boldsymbol{\psi_i} + \sigma_\nu^2 \boldsymbol{J_n}) \phi(\nu|\Lambda \boldsymbol{1_n^T \psi_i^{-1}} (\boldsymbol{y} - \boldsymbol{X\beta}), \Lambda) \phi(\omega|0,\sigma_\omega^2) d\nu d\omega \\ &= \int_{\mathbb{R}} \phi_n(\boldsymbol{y}|\boldsymbol{X\beta}, \boldsymbol{\psi_i} + \sigma_\nu^2 \boldsymbol{J_n}) \phi(\omega|0,\sigma_\omega^2) d\omega \end{aligned}$$

Proof 2 (for Lemma 2) *Begin by dropping the subscript i and substituting $n_i = n$. Using the alternative skew-normal representation given in equation 5.4 and the assumptions in the lemma the marginal distribution of $\boldsymbol{Y}$ can be written as*

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\tau},\boldsymbol{\gamma},\boldsymbol{\Delta},\sigma_\omega^2) &= \int_{\mathbb{R}^2} f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\Delta},\nu,\omega) f(\nu|\boldsymbol{\tau}) f(\omega|\sigma_\omega^2) d\nu d\omega \\ &= \int_{\mathbb{R}^2} 2^n \phi_n(\boldsymbol{y}|\boldsymbol{X\beta} + \nu_i \mathbf{1}_n, \boldsymbol{\Psi}_i + \boldsymbol{\Delta}_i^2) \\ &\quad \times \Phi_n(\boldsymbol{\Delta}_i(\boldsymbol{\Psi}_i + \boldsymbol{\Delta}_i^2)^{-1}(\boldsymbol{y} - \boldsymbol{X\beta} - \nu \mathbf{1_n})|\mathbf{0}, (\boldsymbol{I_n} + \boldsymbol{\Delta_i}\boldsymbol{\Psi}_i^{-1}\boldsymbol{\Delta}_i)^{-1}) \\ &\quad \times \phi(\nu|0,\sigma_\nu^2)\phi(\omega|0,\sigma_\omega^2) d\nu d\omega \end{aligned} \tag{C.2}$$

Define $\boldsymbol{A} = \boldsymbol{\Delta}_i(\boldsymbol{\Psi_i} + \boldsymbol{\Delta}_i^2)^{-1}$ and $\boldsymbol{B} = (\boldsymbol{I}_{n_i} + \boldsymbol{\Delta}_i \boldsymbol{\Psi}^{-1} \boldsymbol{\Delta}_i)^{-1}$ and substitute into C.2 to obtain

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\tau},\boldsymbol{\gamma},\boldsymbol{\Delta},\sigma_\omega^2) &= \int_{\mathbb{R}^2} 2^n \phi_n(\boldsymbol{y}|\boldsymbol{X\beta} + \nu_i \mathbf{1}_n, \boldsymbol{\Psi}_i + \boldsymbol{\Delta}_i^2) \Phi_n(\boldsymbol{A}(\boldsymbol{y} - \boldsymbol{X\beta} - \nu \mathbf{1}_n|\mathbf{0}, \boldsymbol{B}) \\ &\quad \times \phi(\nu|0,\sigma_\nu^2)\phi(\omega|0,\sigma_\omega^2) d\nu d\omega \end{aligned} \tag{C.3}$$

Using Lemma A1 from Arellano-Valle et al. (2007)

$$\begin{aligned} \phi_n(\boldsymbol{y}|\boldsymbol{X\beta} + \nu_i, \boldsymbol{\Psi}_i + \boldsymbol{\Delta}_i^2)\phi(\nu_i|0,\sigma_\nu^2) = \; & \phi_n(\boldsymbol{y}|\boldsymbol{X\beta}, \boldsymbol{\Psi_i} + \boldsymbol{\Delta}_i^2 + \sigma_\nu^2 \boldsymbol{J}_n) \\ & \times \phi(\nu_i|\Lambda \mathbf{1}_n'(\boldsymbol{\Psi}_i + \boldsymbol{\Delta}_i^2)^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}), \Lambda) \end{aligned} \tag{C.4}$$

where $\Lambda = \left(\frac{1}{\sigma_\nu^2} + \frac{1}{\sum_{j=1}^{n_i}(\sigma_{\epsilon_{ij}}^2 + \delta_{ij})}\right)^{-1}$

Define $\boldsymbol{\Gamma} = \boldsymbol{\Psi}_i + \boldsymbol{\Delta}_i^2 + \sigma_\nu^2 \boldsymbol{J}_n$ and $\mu = \Lambda \mathbf{1}_n'(\boldsymbol{\Psi}_i + \boldsymbol{\Delta}_i^2)^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})$ and substitute along with C.4 into C.3 to obtain

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\tau},\boldsymbol{\gamma},\boldsymbol{\Delta},\sigma_\omega^2) &= \int_{\mathbb{R}^2} 2^n \phi_n(\boldsymbol{y}|\boldsymbol{X\beta}, \boldsymbol{\Gamma}) \phi(\nu_i|\mu,\Lambda) \\ &\quad \times \Phi_n(\boldsymbol{A}(\boldsymbol{y} - \boldsymbol{X\beta} - \nu \mathbf{1}_n)|\mathbf{0}, \boldsymbol{B}) \phi(\omega|0,\sigma_\omega^2) d\nu d\omega \end{aligned} \tag{C.5}$$

Using Lemma A2 from Arellano-Valle et al. (2007) with

$$E[\Phi_n(\boldsymbol{A}(\boldsymbol{y} - \boldsymbol{X\beta} - \nu_i \mathbf{1}_n)|\mathbf{0}, \boldsymbol{B}], \quad \nu_i \mathbf{1}_n \sim N(\mu \mathbf{1}_n, \Lambda \boldsymbol{J}_n) \tag{C.6}$$

gives

$$f_{\boldsymbol{Y_i}}(\boldsymbol{y_i}|\boldsymbol{\beta},\boldsymbol{\tau},\boldsymbol{\gamma},\boldsymbol{\Delta},\sigma^2_\omega) = \int_{\mathbb{R}} 2^{n_i}\phi_{n_i}(\boldsymbol{y}|\boldsymbol{X\beta},\boldsymbol{\Gamma}) \\ \times \Phi_{n_i}(\boldsymbol{A}(\boldsymbol{y}-\boldsymbol{X\beta})|\mathbf{1}_{n_i}\mu + \boldsymbol{A}\mathbf{1}_{n_i}\mu, \boldsymbol{B}+\Lambda\boldsymbol{A}\boldsymbol{J}_{n_i}\boldsymbol{B}^T)\phi(\omega|0,\sigma^2_\omega)d\omega$$

which completes the proof.

APPENDIX D

OPENBUGS CODE FOR SECTION 4.4

```
hrs17tot_model=function(){
  for (i in 1:N) {
      y[i] ~ dnorm(mu[i],tau.ws[i])
      mu[i] <- beta0 + beta1*treatment[i] + beta2*(week[i]/100)
      + beta3*treatment[i]*(week[i]/100) + alpha[person[i]]

      log.sigma2[i] <- gamma0 + gamma1*treatment[i] + gamma2*(week[i]/100)
      + gamma3*treatment[i]*(week[i]/100) + zeta[i]
      sigma2.ws[i] <- exp(log.sigma2[i])
      tau.ws[i] <- 1/sigma2.ws[i]

      zeta[i] ~ dnorm(0,tau.ws2)
  }

  for(j in 1:114){
  alpha[j] ~ dnorm(0, tau.bs[j])
  sigma2.bs[j] <- exp(tau0 + tau1*treatment2[j])
  tau.bs[j] <- 1/sigma2.bs[j]
  }
```

```
  #Prior Distribution of Parameters

  #Location Parameters
  beta0 ~ dnorm(0,.01)
  beta1 ~ dnorm(0,.01)
  beta2 ~ dnorm(0,.01)
  beta3 ~ dnorm(0,.01)


  #BS Parameters

  tau0 ~ dunif(-5,5)
  tau1 ~ dunif(-5,5)

  #WS Parameters

  gamma0 ~ dunif(-5,5)
  gamma1 ~ dunif(-5,5)
  gamma2 ~ dunif(-5,5)
  gamma3 ~ dunif(-5,5)


  tau.ws2 ~ dgamma(.1,.1)
  sigma.ws <- pow(tau.ws2,-1/2)


}
```

APPENDIX E

OPENBUGS CODE FOR SECTION 5.4

```
youngtot_model=function(){
  for (i in 1:N) {
    x[i] ~ dnorm(0,1)%_%I(0,)
    z[i] <- mu[i] + delta*x[i]
    mu[i] <- beta0 + beta1*treatment[i] + beta2*(week[i]/100)
    + beta3*treatment[i]*(week[i]/100) + alpha[person[i]]
    y[i] ~ dnorm(z[i],tau.ws[i])

    log.sigma2[i] <- gamma0 + gamma1*treatment[i] + gamma2*(week[i]/100)
    + gamma3*treatment[i]*(week[i]/100) + zeta[i]
    sigma2.ws[i] <- exp(log.sigma2[i])
    tau.ws[i] <- 1/sigma2.ws[i]

    zeta[i] ~ dnorm(0,tau.ws2)
  }

  for(j in 1:114){
    alpha[j] ~ dnorm(0, tau.bs[j])
    sigma2.bs[j] <- exp(tau0 + tau1*treatment2[j])
```

```
  tau.bs[j] <- 1/sigma2.bs[j]
}

#Prior Distribution of Parameters

#Location Parameters

beta0 ~ dnorm(0,.1)
beta1 ~ dnorm(0,.1)
beta2 ~ dnorm(0,.1)
beta3 ~ dnorm(0,.1)

#Skew Parameter

delta ~ dnorm(0,.1)

#BS Parameters

tau0 ~ dunif(-5,5)
tau1 ~ dunif(-5,5)

#WS Parameters

gamma0 ~ dunif(-5,5)
gamma1 ~ dunif(-5,5)
gamma2 ~ dunif(-5,5)
gamma3 ~ dunif(-5,5)


tau.ws2 ~ dgamma(.5,.5)
```

```
  sigma.ws <- pow(tau.ws2,-1/2)


}
```

APPENDIX F

PYTHON CODE FOR SECTION 6.1

```
import pandas as pd
import numpy as np
import pymc3 as pm
from sim_mixed_normal import sim_mix_normal
import theano.tensor as tt
import theano
import datetime


if __name__ == "__main__":

    sim_number = 33
    file_number = 1

    week_max = 162

    week = range(0,week_max+9,9)

    bipolar_full = sim_mix_normal(sim_num = sim_number)
```

```
#Initialize Data Frame
full_data = pd.DataFrame()

for i in range(sim_number):

    data = bipolar_full[bipolar_full.Simulation == i+1]

    data['HRS17TOT'] = data['HRS17TOT'].astype(
      theano.config.floatX)
    num_trt = len(data.Treatment.unique())
    subject_names = data.Subject.unique()
    n_subjects = len(data.Subject.unique())
    subject_idx = np.repeat(range(n_subjects), len(week))

    n_subjects = len(data.Subject.unique())

    with pm.Model() as hierarchical_model:

        # Hyperpriors for group nodes

        #Intercept
        beta0 = pm.Normal('beta0',mu = 0, sd = 100**2)

        #Slopes
        beta1 = pm.Normal('beta1',mu = 0, sd = 100**2)
        beta2 = pm.Normal('beta2',mu = 0, sd = 100**2)
        beta3 = pm.Normal('beta3',mu = 0, sd = 100**2)

        #Subject Specific Parameters
```

```
#BS

tau0 = pm.Uniform('tau0',-5,5)
tau1 = pm.Uniform('tau1',-5,5)

#WS
gamma0 = pm.Uniform('gamma0', -5, 5)
gamma1 = pm.Uniform('gamma1', -5, 5)
gamma2 = pm.Uniform('gamma2', -5, 5)
gamma3 = pm.Uniform('gamma3', -5, 5)

bs_notrt = pm.Normal('bs_notrt', mu = 0,
                     sd = np.sqrt(np.exp(tau0)),
                       shape=n_subjects/2)
bs_trt = pm.Normal('bs_trt', mu = 0,
                   sd = np.sqrt(np.exp(tau0+tau1)),
           shape = n_subjects/2)

     bs_vec = tt.concatenate((bs_notrt,bs_trt))

     # WS error (random scale)

     eps = pm.InverseGamma('eps', 10,10)
ws_error = pm.Normal('ws_error', mu = 0 ,
                     sd = eps , shape = len(data))


#Location
hrs_est = beta0 + beta1*data.Treatment.values +
```

```
            beta2*data.Week.values/float(100) +
            beta3*data.Treatment.values*
        data.Week.values/float(100) +
          bs_vec[subject_idx]

        #Scale
        hrs_error = tt.sqrt(tt.exp(gamma0 +
              gamma1*data.Treatment.values +
            gamma2*data.Week.values/float(100) +
              gamma3*data.Treatment.values*
        data.Week.values/float(100) + ws_error))

        # Data likelihood
        hrs_like = pm.Normal('hrs_like', mu=hrs_est,
              sd=hrs_error, observed=data.HRS17TOT)

    with hierarchical_model:
        step = pm.Metropolis()

        #Warmup
        hierarchical_trace1 = pm.sample(5000,step,
              init = None, njobs=2)

        #Sample for Statistics
        hierarchical_trace2 = pm.sample(1000, step,
              init = None, njobs=2,
              start = hierarchical_trace1[-1])
```

```
output1 = pm.df_summary(hierarchical_trace2,
  varnames = ['beta0','beta1', 'beta2', 'beta3',
'gamma0','gamma1','gamma2','gamma3','tau0',
'tau1','eps'])

output1['Simulation'] = i + 1

full_data = full_data.append(output1)
```

BIBLIOGRAPHY

IM Anderson, PM Haddad, and J Scott. Bipolar disorder. *BMJ*, 345, 2012. doi: 10.1136/bmj.e8508.

RB Arellano-Valle, H Bolfarine, and VH Lachos. Bayesian inference for skew-normal linear mixed models. *Journal of Applied Statistics*, 34:663–682, 8 2007.

A Azzalini. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12:171 – 178, 1985.

RD Bock. *Multivariate Statistical Methods in Behavioral Research*. McGraw-Hill, New York, 1975.

RD Bock. Measurement of human variation: A two stage model. In *Multilevel Analysis of Educational Data*, pages 319 – 342. Academic Press, New York, 1989.

Ren Chen. *Bayesian Inference on Mixed-effects Models with Skewed Distributions for HIV longitudinal data*. PhD thesis, University of South Florida, 1 2012.

VM Chinchilli, JD Esinhart, and WG Miller. Partial likelihood analysis of within-unit variances in repeated measurement experiments. *Biometrics*, pages 205 – 216, 1995.

L Clark, WS Cleveland, L Denby, and C Liu. Modeling customer survey data. *Case Studies in Bayesian Statistics IV*, pages 3–57, 1999.

WS Cleveland, L. Denby, and C. Liu. Random scale effects. Technical report, 2000. URL http://cm.bell-labs.com/doc/randomscale.ps.

DR Cox and PJ Solomon. Analysis of variability of with large numbers of small samples. *Biometrika*, pages 543 – 554, 1986.

E Frank, M Wallace, M Hall, B Hasler, J Levenson, C Janney, I Soreca, M Fleming, J Buttenfield, F Ritchey, and D Kupfer. An integrated risk reduction intervention can reduce body mass index in individuals being treated for bipolar i disorder: results from a randomized trial. *Bipolar Disorders*, pages 424–437, 2015.

A Gelfand and A Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398 – 409, 1990.

A Gelman, B Carlin, H Stern, D Dunson, Aki Vehtari, and D Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, FL, 3 edition, 2014a.

A Gelman, J Huang, and A Vehtari. Understanding predictive information criteria for bayesian models. *Statistics and Computing*, 24:997 – 1016, 2014b.

S Geman and D Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721 – 741, 1984.

WK Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1 1970.

D Hedeker and R Nordgren. Mixregls: A program for mixed-effects location scale analysis. *Journal of Statistical Software*, 52(12), 2013.

D Hedeker, M Berbaum, and RJ Mermelstein. Location-scale models for multilevel ordinal data: Between- and within-subjects variance modeling. *Journal of Probability and Statistical Science*, pages 1–20, 2006.

D Hedeker, R Mermelstein, and H Demirtas. An application of a mixed effects location scale model for analysis of ecological momentary assesment data. *Biometrics*, 64(2), 2008.

D Hedeker, M Berbaum, and RJ Mermelstein. A mixed ordinal location-scale model for analysis of ecological momentary assessment data. *Stat Interface*, 2:391 – 401, 2009.

D. Hedeker, RJ Mermelstein, and H. Demirtas. Modeling between- and within-subject variance in ecological momentary assessment (ema) data using mixed-effects location scale models. *Statistics in Medicine*, (31):3328–3336, 2012.

AT James, WN Venables, IB Dry, and JT Wiskich. Random effects and variances as a synthesis of nonlinear regression analyses of mitochondrial electron transport. *Biometrika*, pages 219 – 235, 1994.

VE Johnson. An alternative to traditional gpa for evaluatiing student performance. *Statistical Science*, pages 251–269, 1997.

K Kapur, X Li, E Blood, and D Hedeker. Bayesian mixed-effects location and scale models for multivariate longitudinal outcomes: an application to ecological momentary assessment data. *Statistics in Medicine*, pages 630 – 651, 2015.

NM Laird and JH Ware. Random effects models for longitudinal data. *Biometrics*, (38):963–974, 1982.

X Li and D Hedeker. A three-level mixed-effects location scale model with an application to ecological momentary assessment data. *Statistics in Medicine*, pages 3192 – 3210, 2012.

X Lin, J Raz, and SD Harlow. Linear mixed models with heterogenous within-cluster variances. *Biometrics*, pages 910 – 923, 1997.

DV Lindley. *The Estimation of Many Parameters*, pages 435–455. Toronto, Holt, Rinehart, and Winston, 1971.

R Mermelstein, D Hedeker, B Flay, and S Shiffman. Situational versus intra-individual contributions to adolescents' subjective mood experience of smoking. Annual Meeting for the Society for Research on Nicotine and Tobacco, Savannah, GA, 2002.

N Metropolis, AW Rosenbluth, Rosenbluth MN, AH Teller, and E Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 6:1087 – 1092, 1953.

JR Nesselroade and N. Ram. Studying intraindividual variability: What we have learned that will help us understand lives in context. *Research in Human Development*, (1):9–29, 2004.

O Pugach, D Hedeker, and R Mermelstein. A bivariate mixed-effects location scale model with application to ecological momentary assessment data. *Health Services Outcome Research Method*, pages 194 – 212, 2014.

N Reisby, LF Gram, P Bech, A Nagy, GO Petersen, J Ortmann, I Ibsen, SJ Dencker, O Jacobsen, O Krautwald, I Sondergaard, and J Christiansen. Imipramine:clinical effects and pharmacokinetic variability. *Psychopharmacology*, 54:263 – 272, 1977.

S Sahu, D Dey, and M Branco. A new class of multivariate skew distributions with applications to bayesian regressian models. *The Canadian Journal of Statistics*, 31(2):129 – 150, 2003.

J Salvatier, TV Wiecki, and C Fonnesbeck. Probabilistic programming in python using pymc3. *PeerJ Computer Science*, page e55, 2016.

R Scarpa, M Thiene, and K Train. Utility in willingness to pay space: A tool to address confounding random scale effects in destination choice to the alps. *American Journal of Agricultural Economics*, pages 994 – 1010, 2008.

Lei Shu. *Stepwise Model Building for Mixed-Effects Models with Random Scale Effects*. PhD thesis, Purdue University, 8 2008.

D Spiegelhalter, N Best, B Carlin, and A van der Linde. Bayesian measures of model complexity and fit. *Journal of the Rotal Statistical Society*, 64:583 – 939, 2002.

G Verbeke and G Molenberghs. *Linear Mixed Models for Longitudinal Data*. Springer, New York, 2000.