University of Nevada, Reno

Developing a General Method for Analyzing Psychomotor Vigilance Task (PVT) Data: Modeling Sleep Inertia in Children

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in
Mathematics

by

Joanna M. Armstrong

Deena R. Schmidt, Ph.D. \Thesis Advisor

August 2018

We recommend that the thesis
prepared under our supervision by

JOANNA M. ARMSTRONG

entitled

**Developing a General Method for Analyzing Psychomotor Vigilance
Task (PVT) Data: Modeling Sleep Inertia in Children**

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

Deena R. Schmidt, Ph.D. Advisor

Paul J. Hurtado, Ph.D., Committee Member

Dennis Mathew, Ph.D., Graduate School Representative

David Zeh, Ph.D., Dean, Graduate School

August 2018

Abstract

Some data analysis applications may violate the assumptions of standard (e.g., linear model) frameworks. In such cases, a common solution is to develop a more suitable system-specific model, and from it derive the statistical tools necessary to conduct the desired data analysis. In this thesis, we propose a method to model Psychomotor Vigilance Task (PVT) data which is a test to measure alertness and is commonly used in psychology. In particular, we focus on applying these methods to data from studies of measuring sleep inertia in children. We will start by looking at relevant methodological and application specific background. We will then look at basic statistical analysis such as mean, median, and standard error and plot some examples to get a sense of the data. Then we will introduce a flexible model that better represents sleep inertia. We describe a maximum likelihood based estimation procedure and analyze parameter identifiability. Lastly, we use the likelihood ratio test to compare nested models.

Dedication

To my husband, Tim. Thank you for all your help and encouragement
throughout this process. To my children, Hailey and Dylan, thank you
for all your patience and letting mommy do her work.

Acknowledgements

I want to first thank my advisor, Deena Schmidt. Thank you for all your wisdom, patience, time, support, and encouragement. Words really cannot express how thankful to you I am. I also want to thank my committee member, Paul Hurtado. Thank you for taking the time to answer my questions and all your help even when you had better things to do. A special thank you to my committee member, Dennis Mathew, for all your great feedback after my defense. And lastly, thank you to Dr. Mark Splaingard for collecting and sharing the data that motivated this work.

Contents

List of Figures

Chapter 1

Introduction

In this thesis, we propose a method to model Psychomotor Vigilance Task (PVT) data. Specifically, we propose a new model to study sleep inertia in children. We analyze data from a study conducted by Dr. Mark Splaingard at Nationwide Children's Hospital (NCH) in Columbus, OH. Below we describe what sleep inertia is, how the data has been collected, show examples, and discuss the motivation for this study. Because of the highly variable reaction times in PVT data, a better mechanistic model should improve the statistical analyses of PVT data, and help open up new avenues of research.

1.1 Background on Sleep Inertia

Sleep inertia is the initial period after awakening in which delayed performance and disorientation occurs. Decreased performance caused by sleep inertia can be found in tasks such as reaction time, steadiness and coordination, visual perceptual tasks, memory tasks, logical reasoning, time estimates, complex-behavior simulation tasks, and cognitive tasks including mental arithmetic [18]. The specific sleep stage at the time of awakening can affect the amount of sleep inertia experienced [16]. A person

awakening from a slow wave sleep, a deep sleep (Non REM), is more affected by sleep inertia than a person being woken up from a stage 1-2 or REM sleep. Stage 1 consists of a person drifting in and out of sleep and can be easily awoken, while stage 2 sleep is when eye movement stops and brain waves become slower. Stages 3 and 4 are the depest sleep stages. The body temperature starts to drop and the heart rate slows. In REM sleep brain waves mimic activity during the waking state with eyes remaining closed but rapidly moving. Sleep deprivation has been found to increase the effects of sleep inertia, making it harder for sleep inertia to dissolve. Sleep inertia has also been found to be more extreme when the core temperature cycle is at its lowest point, which typically happens around 2AM [5]. Caffeine has been found to minimize the effects of sleep inertia [9]. Newman et al [9] concluded in their study that caffeine gum is a way to reverse the effects of sleep inertia.

The effects of sleep inertia can last anywhere from 1 minute to 4 hrs, which is quite a wide range of time [18]. One study found that impairment caused by sleep inertia could last for more than 2 hours after awakening [5]. The study found that the subjects alertness and cognitive throughput scores increasingly improved during the first hour and were constant by the second hour. These results supported a previous hypothesis by Folkard, S. and Akerstedt, T. [4] that sleep inertia dissipates asymptotically. The study measured alertness by using a visual analogue scale while the cognitive throughput was measured by doing two minutes of addition. Subjects were woken up after Stage 1, Stage 2, and REM sleep. The study only had a sample size of one from a slow wave sleep cycle so they could only hypothesize that sleep inertia would be worse after awakening from the slow wave sleep cycle.

Sleep deprivation occurs when a person does not get enough sleep which leads to decreased performance and alertness. The longer a person is awake the worse their performance and alertness becomes. Figure 1.1 shows the effects of sleep deprivation

as measured by the PVT at 3 time intervals: 12, 36, and 60 hours awake. When subjects experiencing sleep deprivation are given the PVT, their reaction times become slower as time goes on, as shown in Figure 1.1 [19]. Results show that sleep deprivation can be characterized by stochastic performance variability, with higher variability towards the end of the PVT.
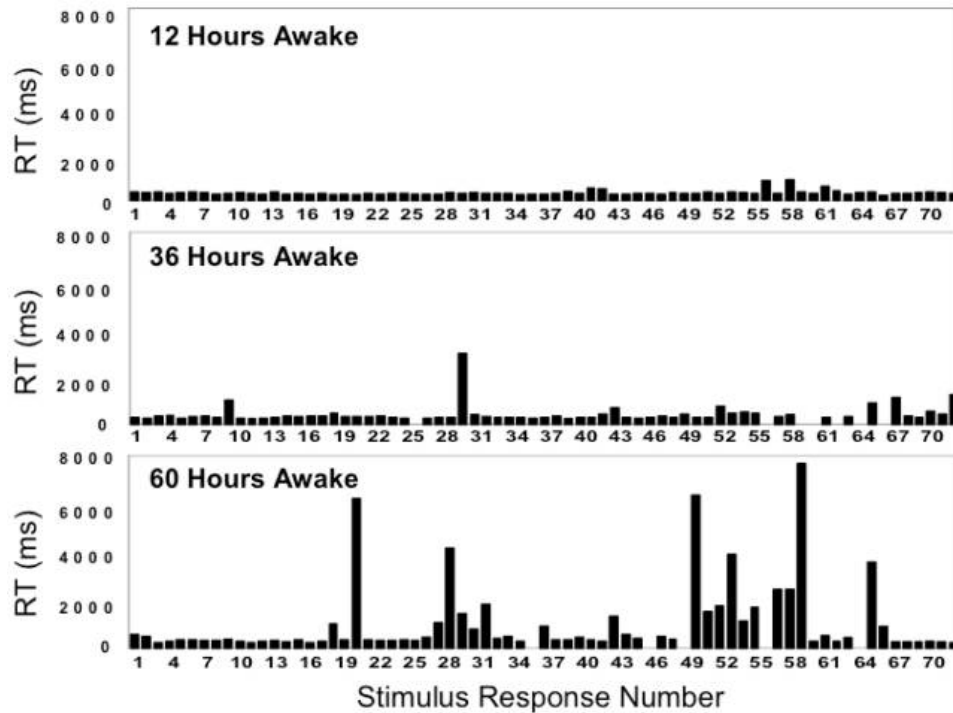
Figure 1.1: The effects of sleep deprivation as measured by the PVT from a study conducted by Van Dongen [19].

Those experiencing sleep inertia have slower reaction times during the beginning of the PVT with reaction times getting faster towards the end. Sleep inertia can be thought of as the reverse process (on a different timescale) to sleep deprivation. We will show examples of sleep inertia measured by PVT in the following section.

1.2 Psychomotor Vigilance Task (PVT)

An effective way to measure sleep inertia is by use of the psychomotor vigilance task (PVT) [2, 10]. PVT is a reaction-timed task that measures how fast an individual can respond to a visual stimulus. This "alertness test" is used to measure psychomotor impairment. It has been widely used to measure effects of sleep deprivation, and recently it has also been used to measure the effects of sleep inertia. The PVT hand-held device consists of an LED display window that provides a light to the user along with a button that the user presses when they see the light. The light flashes randomly every several seconds for 10 minutes. In particular, the inter-event times follow a random uniform distribution on the range [2sec,10sec][17]. Once the light flashes, the PVT starts a counter in milliseconds and stops the counter once the subject presses the button, recording each response time (RT) for the duration of the 10-minute test. If a subject presses the button when the light does not flash, presses a wrong button, holds down on the button too long, or does not respond after sixty-five seconds, then an error message appears instead of the reaction time to the user. Most adults have response times by 500 milliseconds, so if a response is longer than 500 milliseconds it is considered a lapse, but young children have much more variable reaction times.

1.2.1 Example Data and Motivation

Figures 1.2 and 1.3 show example PVT data. The x-axis represents the time (in seconds) when the visual stimulus appeared and the y-axis represents the time it took the child to respond to the visual stimulus (in ms). We wanted to showcase the variability in the data. Figure 1.2 shows two different PVTs from the same child, given at different times. Panel (a) is a PVT given while the child was awake (baseline), and panel (b) is a PVT given after the child is awoken from deep sleep

(cycle2). In the baseline PVT the reaction times are all close together and not highly variable. In the cycle2 PVT, the reaction times start off higher and become stationary, with some high PVT reaction times near the end of the test. These higher response times may be outliers or something else is happening toward the end of the PVT. The beginning of this PVT highlights the sleep inertia effect and the end of the PVT possibly showcases some sort of distraction and/or fatigue. Figure 1.3 shows more examples of the data. This figure shows high reaction times in the beginning of the PVT along with fatigue at the end, examples of sleep inertia, and an example of just fatigue at the end of the 10 minute PVT. There are multiple problems with the data that one could study. For example, PVT times have a bigger spread at the beginning that decreases then may increase again after a few minutes. Questions regarding our motivation for a better model include: How significant is the initial transient improvement in response times? What does the fatigue at the end (present in children but is not as apparent in adults) look like? Is the distribution of reaction times heavy tailed or bimodal? Since the data is so variable and there are many different cases, there is a need for a more flexible model.
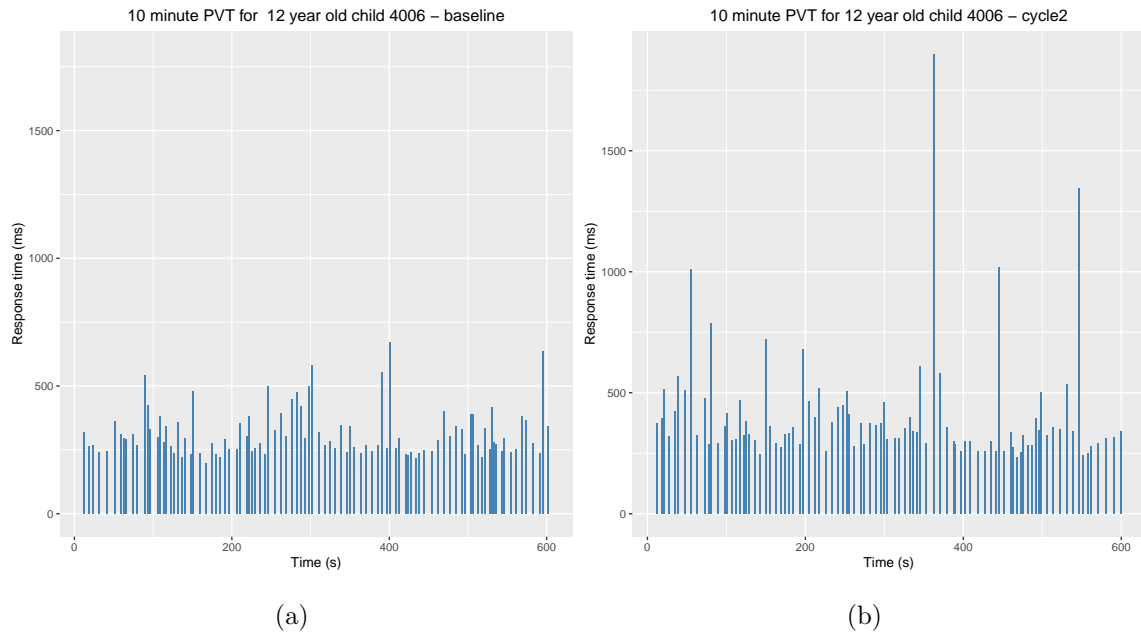
Figure 1.2: Panel (a) shows a baseline PVT of a 12 year old child with no signs of sleep inertia or fatigue. Panel (b) shows the same child after being awoken up following their second cycle of deep sleep, showing effects of sleep inertia along with some fatigue at the end.
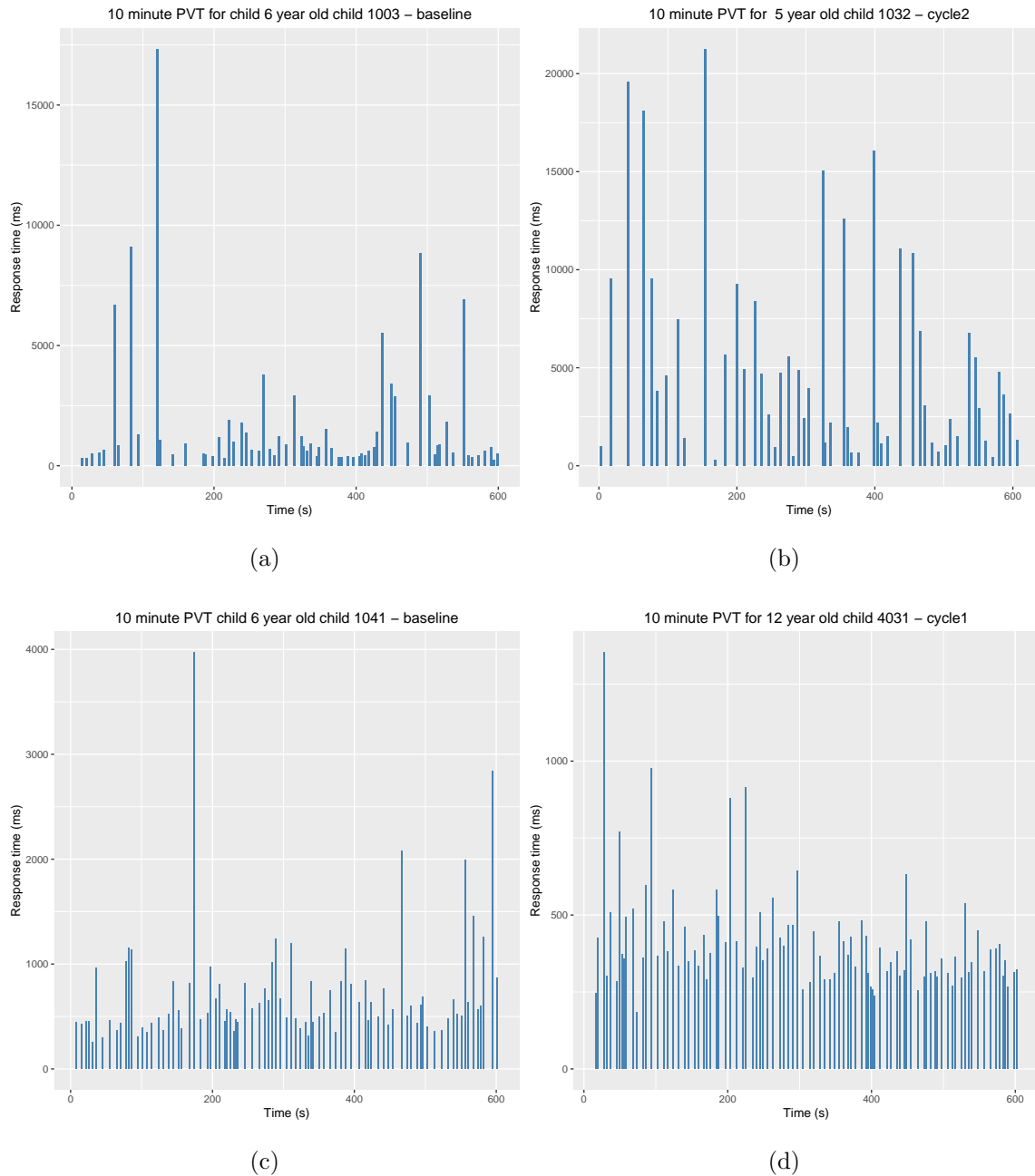
Figure 1.3: Panel (a) shows a baseline PVT from a 6 year old child with high reaction times early on and fatigue at the end. Panel (b) shows a cycle2 PVT from a 5 year old child with effects of sleep inertia. Panel (c) shows fatigue at the end of the 10 minute PVT. Panel (d) possibly shows a cycle1 PVT from a 12 year old child with effects of sleep inertia.

1.2.2 Summary of Data Collection

The data we will analyze in this thesis comes from a study conducted by Dr. Mark Splaingard at Nationwide Children's Hospital in Columbus, Ohio looking at sleep inertia effects in children ages 5-12 [6]. The response times for each child were measured by using PVT for a ten minute period. Children performed the PVT about 90 minutes before going to bed, denoted baseline, to give a control for the study. The children were then woken up twice within the night, denoted cycle 1 and cycle 2, after a stage 4 sleep cycle (deep sleep) and then given the PVT. Children in the study were woken up by an alarm that went off for one minute intervals with ten seconds between each alarm. This alarm continued for up to five minutes. If the child did not wake up, then a research assistant would wake up the child and take them to the room where the PVT was administered. The room where the PVT was given was well lit and the research assistant was present for the entire duration of the PVT. One week after the first night, the children were brought back for another night of testing under the same conditions.

The data set includes reaction times for each time point within the 10-minute PVT for a set of 168 children. For example, at 10.1 seconds in the PVT test a child had a reaction time of 600.2 milliseconds. This data set also includes the child's identification number, their age, and the cycle at which the test occurred. The 168 subjects each have up to six 10-minute PVTs recorded. These six 10-minute PVT's include a baseline, cycle 1, and cycle 2 test for two separate nights. Looking at the data, there is high variability. The variability seems to increase as the age of the children decreases. There is also high variability towards the end of the PVT in the data. We hypothesize that this is due to children getting distracted and/or fatigued with the test, since ten minutes is a long time for children to stay continually focused on a mundane task.

1.2.3 Data Analysis

Before data analysis could be conducted, it was necessary to clean up the data. The raw data was stored in Matlab in an .m file. Each .m file corresponded to the child and within each .m file were six matrices representing the data collected from each trial/cycle. Some of the children did not show up for some of the testing and so those trials/cycles were stored as empty matrices. There were 168 .m files. The main goal was to take the Matlab files and store them in a SQL database called Postgres. Postgres is a free SQL program that has a graphical user, and can interface with the statistical software R [11]. Storing the data in a database makes it easier to deal with and many programming languages can run SQL queries within them. A database was created in Postgres. One table was created to store the data from the .m files and another related table was created to store the childs ID, age, and gender.

Typically, experimentalists only take into account the mean reaction time over the entire 10 minute PVT. Below, in Figure 1.4, shows the mean reaction times from each age group, for each cycle, along with the standard error.
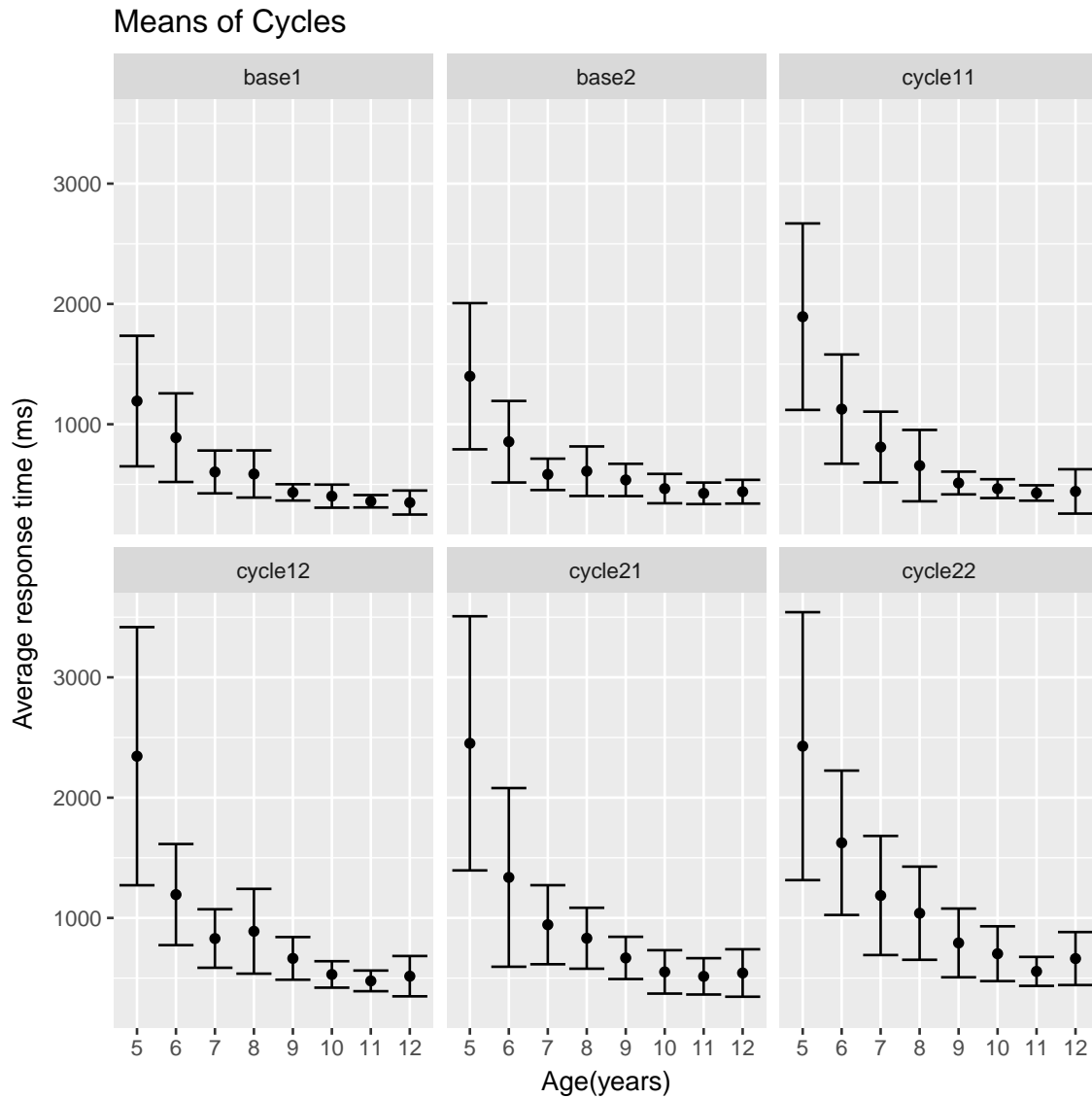
Figure 1.4: Mean reaction times with standard error of the 10 minute PVT for each age group for each cycle.

Since five year olds have a large standard error, we plotted histograms of individual five year olds. Figures 1.5 and 1.6 show histograms of six 5 year old children for the 10 minute PVT. The histograms show that the data does not follow a normal distribution and that the distribution is skewed.
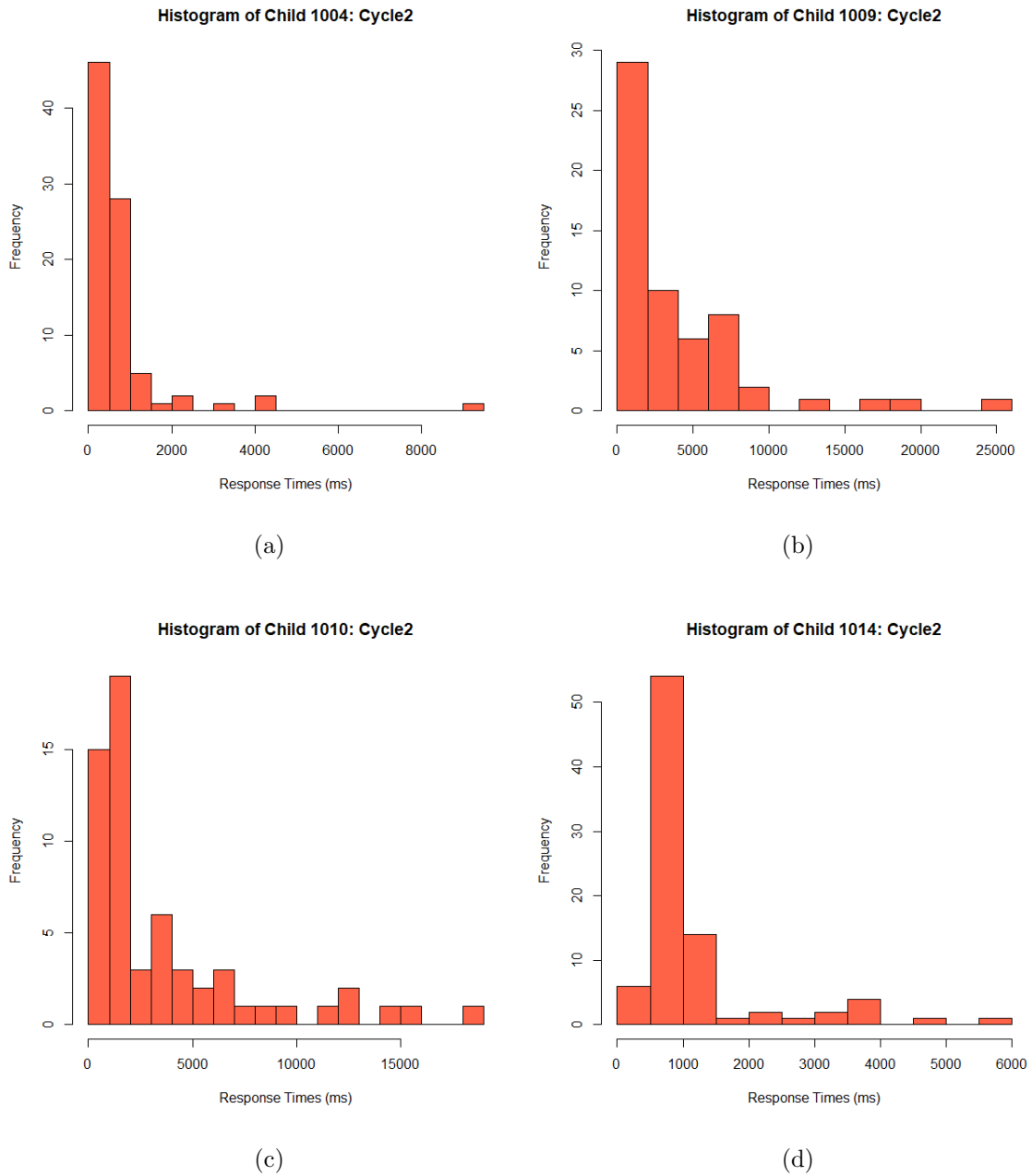
Figure 1.5: Histograms of PVT response times for 5 year old children.

Histogram of Child 1017: Cycle2

Histogram of Child 1020: Cycle2

(a)

(b)
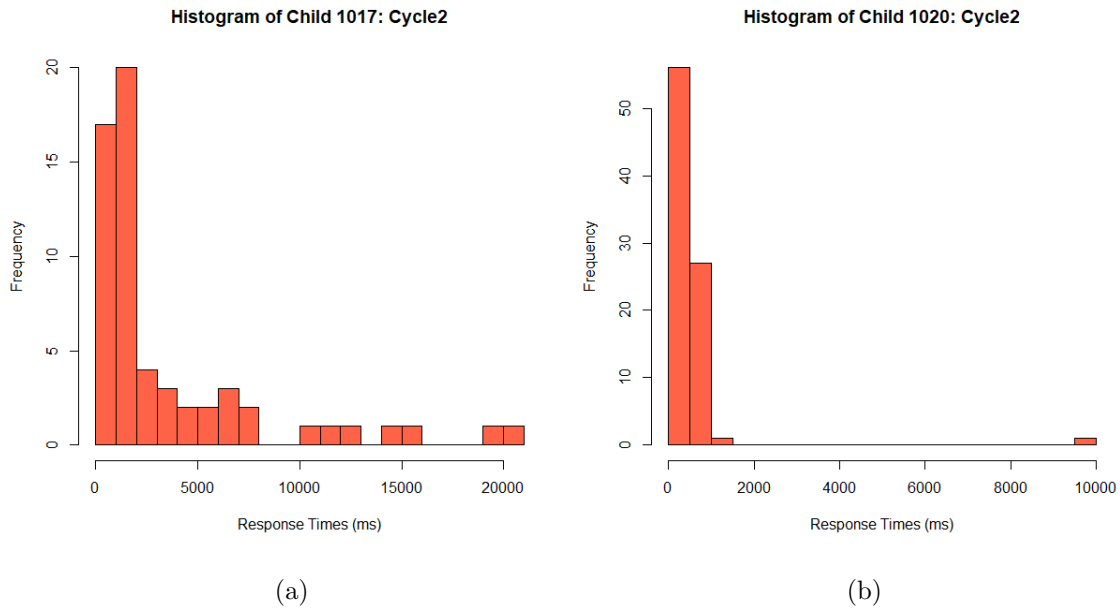
Figure 1.6: Two more histograms of 5 year old children.

We computed the mean, median, and standard error (SE) of response times of each age group for the three different PVT type tests. Below are three tables of summary statistics (mean, median, SE) from the data by age for baseline, cycle1, and cycle2; see Tables 1.1, 1.2, and 1.3.

Table 1.1: Baseline (Night 1) Summary Statistics - This table shows the mean, median, and standard error of PVT response times for each age group. This illustrates the difference between the mean and median; indicating a skewed distribution. Tables for Cycle1 and Cycle2 were constructed in the same manner.

Age	n	Mean	Median	SE
5	11	1192.776	667	542.574
6	8	888.3624	578	368.2887
7	23	603.4847	421	177.9456
8	16	586.3886	405	195.9833
9	26	433.4736	362	67.87197
10	26	402.1444	318	95.86562
11	22	359.8007	313	51.33567
12	18	349.1324	294	99.81827

Table 1.2: Cycle1 (Night 1) Summary Statistics

Age	n	Mean	Median	SE
5	13	1894.195	920.5	774.9956
6	8	1125.743	665	453.9564
7	23	810.3583	477	293.9945
8	16	656.0738	444	296.6032
9	25	511.4172	420	94.42216
10	26	464.5826	373	78.1233
11	21	428.5572	362	63.87834
12	18	441.4946	330	184.3794

Table 1.3: Cycle2 (Night 1) Summary Statistics

Age	n	Mean	Median	SE
5	12	2451.479	986	1056.254
6	9	1337.08	700	742.7786
7	22	943.6937	501	328.5721
8	15	831.3173	508	253.3328
9	25	667.7455	455.5	175.4443
10	24	551.6224	392	180.307
11	22	514.6285	384	150.9108
12	18	542.4616	361.5	197.1318

The summary statistics illustrate that for most age groups, the mean and median are far apart which was evident in Figure 1.6. This suggest that the data is skewed and the mean may not be the best summary statistic for these data. Looking at just the mean or median may not be the best practice for performing statistical analysis, especially for children, because the data is highly variable and those summary statistics don't capture the whole picture of what is going on in the data.

1.3 Objectives

The main objective of this thesis is to answer the question, "How do we build a better model to study sleep inertia?" Current methods assume a stationary distribution and just report the mean and/or the median. We want to build a more flexible model that allows for specific transient trends at the beginning, (e.g. learning/sleep inertia), and non-stationary effects at the end (e.g. fatigue/distraction) of the distribution so that we can quantify the stationary phase of the data (and analyze that) while

simultaneously correcting for, and thus quantifying those non-stationary effects.

We are proposing a general model for PVT data to model reaction times (RT) in children under the effects of sleep inertia. We also consider a simple model is a special case of the full model, which doesn't account for a change in mean RT over time, i.e. doesn't explicitly model sleep inertia effects. The following methods will be used. Maximum Likelihood Estimation will be used to find the best fit parameters for our models. We will also be testing to see if we have any identifiability issues with our best fit parameters in our model. To test this, we will be using likelihood profiling. To test which of the two models fit better, we use a likelihood ratio test along with the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) values.

1.4 Model Framework

Our main goal is to develop a nonlinear model that changes over time to better model PVT reaction times in children, under the effects of sleep inertia. This will also lay the groundwork for further generalizations that apply more broadly to other PVT data.

Reaction times during sleep inertia start off slow and exponentially get faster. For this reason our model for reaction times starts has an exponential decreasing mean. The model for sleep inertia can be represented by the following:

$$Y_i = y_0 + X_i \tag{1.1}$$

where Y_i is the ith response time, y_0 is the minimum possible response time, and X_i is a gamma distributed random variable with mean $\mu_i = \tau_* + \tau_o e^{-rt_i}$ and variance $c\mu_i$. To model the above non-stationary effects on the mean and variance of X_i, τ_o is

the scaling factor of the exponential decay function which is used to explicitly model sleep inertia, r is the exponent in the same exponential decay function, t_i is the ith time point (where the time vector has a length of ten minutes as in the PVT), c is the constant that is being multiplied by μ_i to make our variance proportional to the mean, and τ_* is the constant mean value for the second part of the model.

The gamma distribution was chosen as the distribution for each time index, t_i. Reaction times will never go below zero, which makes the gamma distribution preferable since its range is from zero to infinity. The gamma distribution is also a quite flexible, making it ideal for our model. For more details on the gamma distribution, see Appendix (5.1).

As a special case, where there is no initial transient improvement in performance, we will consider the special case of our model where $\tau_o = 0, r = 0$:

$$Y_i = y_0 + X_i \tag{1.2}$$

where X_i is gamma distributed with $\mu = \tau_*$ and variance $c\mu$. Simulated data was created to test our model. Figure 1.7 below shows simulated data illustrating the difference between the two different models.

Simulated Data Under Simple Model
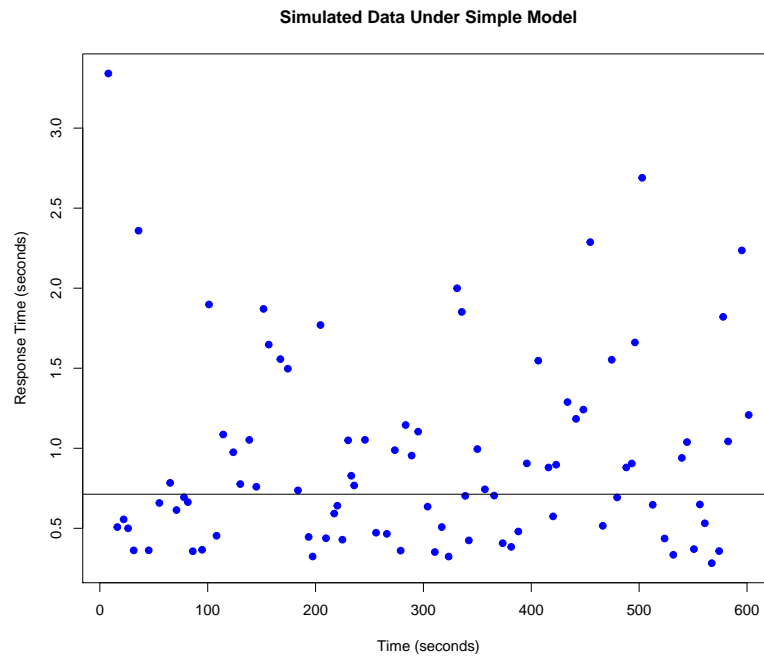
(a) Simulated data without the effects of sleep inertia with parameters $y_0 = .256, c = 0.5$, and $\tau_* = .713$. The mean is $\mu = 0.936$ and the variance is $var = 0.366$.

Simulated Data Under Full Model
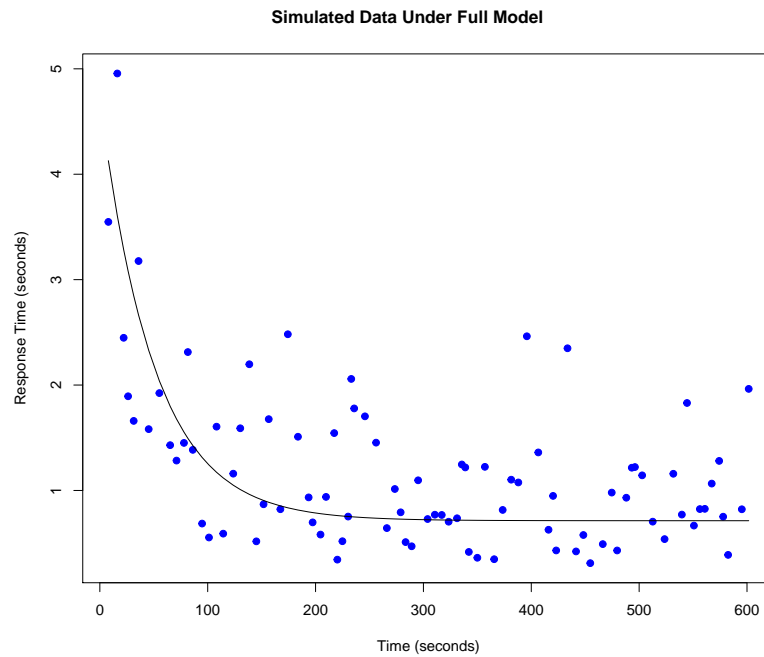
(b) Simulated data created with the effects of sleep inertia in place with parameters $\tau_o = 4$, $r = 0.02$, $y_0 = .256, c = 0.5$, and $\tau_* = .713$. The mean is $\mu = 1.183$ and variance, $var = 0.594$.

Figure 1.7: Simulated data from the simple model (a) and the full model (b).

Chapter 2

Methods

In this section we will discuss the methods to be used to implement our model, methods to develop estimators for the parameters in the model, methods for identifiability analysis, and methods to find the best fit model for specific data.

2.1 Maximum Likelihood Estimation

Maximum Likelihood Estimation is a method for estimating parameter values of a statistical model given the observations [14]. These estimates are often called MLEs. In order to find the MLE, the likelihood function is needed: $L(\theta; x)$. The likelihood for an independent and identically distributed (i.i.d) sample x of size n from a distribution with probability density function f is defined by the joint density function at the sample (data):

$$L(\theta; x) = \prod_{i=1}^{n} f(x_i; \theta) \tag{2.1}$$

where θ is the parameter vector. The maximum likelihood estimate (MLE) is the model parameter(s) $\hat{\theta}$ that maximize the likelihood function defined as

$$\hat{\theta} = \{\arg \max {}_{\theta} L(\theta; x)\} \tag{2.2}$$

To find $\hat{\theta}$ in practice it is often easier to use the negative natural log of the likelihood function: $nLL(\theta; x) = ln(L(\theta; x))$. By properties of logs, $\hat{\theta}$ also minimizes the $nLL(\theta; x)$ and we can write

$$nLL(\theta; x_i) = -\sum_{i=1}^{n} ln f(x_i | \theta) \tag{2.3}$$

One of the properties of MLEs and likelihood functions is that they can be used to test hypothesis about models and their parameters. In the model that is being generated, the MLE method will be used to find the parameters that best fit our data and are then used to perform hypothesis testing on the model.

2.1.1 Finding MLEs

The MLE $\hat{\theta}$ can be found analytically or computationally. To find MLEs analytically we use the negative likelihood function and find the partial derivatives of the function for each parameter. We then set each partial derivative to zero and solve the system of equations for θ. Specifically, if $\theta = [\theta_1, ..., \theta_r]$, we solve

$$\frac{\partial}{\partial \theta_1} nLL(\theta, x) = 0 \tag{2.4}$$

$$\vdots \tag{2.5}$$

$$\frac{\partial}{\partial \theta_r} nLL(\theta, x) = 0 \tag{2.6}$$

If this cannot be done analytically then we need to solve the system of equations

computationally.

2.1.2   Computational Optimization to Find MLE

We want an estimation method that is robust to the model having identifiability issues, and we want to retain information to use in subsequent identifiability analyses. We can use the optimization function `optimx` in R using methods Nelder-Mead and BFGS [8] for this optimization procedure, where we want to find the parameters θ that minimize the negative log likelihood function.

Each optimization run must be initialized with a parameter set close to the true MLE as possible. To define a plausible range of inital parameter values, we first looked at crude estimates as follows. For τ_*, we used the mean of the data set. The scaling factor of the variance, c, was set to 0.5. We used these crude estimations of the parameters c and τ_* along with n being set to the number of time points from our data set to solve for y_0 by using the function `uniroot`.

For the general model, we know that the data shows an exponential decay followed by a constant value. We can use this to help decide on the crude estimation of the parameters. Just as in the simple model τ_* was found by using the mean of the data set. τ_o was found by scaling τ_* by 0.2. The parameter r was found by taking $ln(20)/120$ since the data shows that the exponential decay levels out around 120 seconds. y_0 and c are the same as the simple model.

`Optimx` is a generalized wrapper function for optimization. It allows for a number of methods to all be used at once including BFGS, spg from the BB package, ucminf, nlm, and nlminb[8]. Optim uses Nelder-Mead, quasi-newton, and conjugate gradient algorithms as well as box-constraint optimization methods. An initial vector of parameters to be optimized and the function to be maximized/minimized are a requirement to run optimx. For our analysis, methods Nelder-Mead and BFGS were

used, and the maximum number of iterations was increased to 50000. The best fit parameter estimation and the convergence code was used to determine which parameter values from which method should be used. The method Nelder-Mead and BFGS produced the best fit parameters.

To ensure the best fit parameters were being estimated, and to collect information to be used in subsequent identifiability analysis, we created a function called parallel.optimx which runs N optimization routines via `optimx` at the same time, each starting with a different initial parameter set. To run N optimizations, a plausible range for each parameter was defined. We then randomly sampled N initial parameter sets from this range. These random sampled estimates were then run through `optimx` and the best fit was selected by choosing the minimum negative log likelihood value.

2.2 Parameter Estimation and Identifiability

A model is identifiable if the best fit parameter estimates of a model are unique. A model that has identifiability issues is one where at least one parameter estimate in the model can be more than one value and still correspond to the same negative log likelihood value [3, 12]. Two methods that can be used to assess if there are identifiability issues in parameter estimation are by looking at pairwise scatter plots of MLEs of the parameters via parallel.optimx and performing likelihood profiling on each parameter. Practical identifiability occurs when multiple parameter sets are very close to the optimum, which throws off computational optimization routines. We assumed model identifiability in two ways (see Sections 2.2.2 and 2.2.3 below). The likelihood profiling method relies on likelihood ratio tests, which we will first review in the next section.

2.2.1 Likelihood Ratio Test

The log likelihood ratio test is used to compare two statistical models: a null model against an alternative model [7]. The null model is a special case of the alternative model or the "simple" model, while the alternative model is the "full" model.

$$H_0 : \theta = \theta_0 \tag{2.7}$$

$$H_1 : \theta = \theta_1 \tag{2.8}$$

A critical value and a p-value are computed to determine whether or not the simple model is rejected. The probability distribution for the likelihood ratio test can be approximated using Wilk's theorem. The likelihood ratio is also justified by the Neyman-Pearson lemma which shows that the likelihood ratio test is the most powerful (the best) test at significance level α.

The likelihood ratio test is constructed as:

$$\Lambda(x) = \frac{L(\theta_0|x)}{L(\theta_1|x)} \text{ or alternatively,} \tag{2.9}$$

$$\Lambda(x) = \frac{L(\theta_0|x)}{sup\{L(\theta|x) : \theta \in \{\theta_0, \theta_1\}\}} \tag{2.10}$$

where θ_0 corresponds to the simple model and θ_1 corresponds to the full model, while the likelihood function is the suprema of the supremum function. The likelihood ratio is small if the full model fits better than the simple model. The decision rule is as follows:

$$\text{If } \Lambda > c, \text{ do not reject } H_0 \tag{2.11}$$

$$\text{If } \Lambda \leq c, \text{ reject } H_0 \tag{2.12}$$

Where c is the critical number chosen for a specified significance level, α. The

numerator in the likelihood ratio test corresponds to an observed outcome from the simple model while the denominator corresponds to the maximum likelihood of an observed outcome varying the parameters over the whole parameter space. The lower the number the less likely it is to occur under the simple model and vice versa. In order to run the likelihood ratio test on models, the models must be nested. That is, the full model can be transformed into the simple model under a certain constraint.

Wilk's theorem states that as the sample size n approaches infinity, the test statistic, $-2\log(n)$, for a nested model will asymptotically be chi-squared, χ^2, distributed with degrees of freedom equal to the difference in dimensionality of $\theta \in \theta_0$ when H_0 is true. This theorem is used in the likelihood ratio, Λ, to compare $-2\log(\Lambda)$ to the χ^2 value corresponding to a given statistical significance value. This statistical test uses the test statistic D:

$$D \;=\; -2ln\left(\frac{L(\theta_0|x)}{L(\theta_1|x)}\right) \tag{2.13}$$

$$=\; 2ln\left(\frac{L(\theta_1|x)}{L(\theta_0|x)}\right) \tag{2.14}$$

$$=\; 2\left[ln(L(\theta_1|x)) - ln(L(\theta_0|x))\right] \tag{2.15}$$

where the degrees of freedom are $df_1 - df_0$ i.e., the number of parameters of the simple model subtracted from the numbers of parameters from the full model.

To conduct the likelihood ratio test for two nested models we used `lr.test` in R. Inputs used to run this test are the two models to be compared, objects from a class or the negative log likelihood value for each model can be used, the significance test, and the degrees of freedom. The function returns the test statistic value, the χ_2 critical value, the p-value, and the decision. The decision "greater" indicates that we reject the null hypothesis if the p-value is less than α, otherwise we do not reject the null hypothesis.

2.2.2   Scatter Plots

To see if there are any identifiability issues with the parameter estimates, we can use pairwise scatter plots of the candidate MLEs estimated from different initial conditions by using parallel.optimx (see section 2.1.2 for details). If there are identifiability issues with the parameter estimates, the scatter plots will have an apparent pattern to them. If no pattern is present, then there is not an identifiability issue. Ideally, you want the points to be scattered randomly with best estimates clustered tightly together. For example, if an identifiable combination (i.e., a quantity that can be estimated) in a given model is $m = \dfrac{\theta_1}{\theta_2}$, where θ_1 and θ_2 are parameters best estimate, the $\theta_1 - \theta_2$ scatter plots will show a linear relationship, $\theta_1 \approx m\theta_2$. Likelihood profiling is another tool that can be used to reveal identifiability issues with parameters.

2.2.3   Likelihood Profiling

Likelihood profiling is a useful tool for finding approximate confidence intervals for best fit parameters $\hat{\theta}$ [13]. It is based on the asymptotic χ^2 distribution of the likelihood ratio test. If we have a model with parameters, θ, the profile likelihood function for the jth parameter, θ_j defined as

$$PL_j(\delta) = \operatorname*{argmin}_{\theta_i, i \neq j} nLL(\theta|x, \theta_j = \delta). \tag{2.16}$$

For every value of θ_j, $PL_j(\delta)$ is the negative log likelihood value for the best fit model with constraint $\theta_j = \delta$, where δ values range over an interval containing $\hat{\theta}_j$.

We can use likelihood profiling to construct confidence intervals as follows. Using the likelihood ratio test for the best fit (full) model and a model with fixed $\theta_j = \delta$, we can use the appropriate χ^2 distribution (See section 2.2.1) to find the negative log likelihood value threshold for the constrained model, below which there is no

significant difference between the goodness of fit of the full and constrained models to the given data. Thus the $100\%(1-\alpha)$ confidence interval for θ_j is the set of values of δ values satisfying

$$2\left[-nLL(\hat{\theta}) + PL_j(\delta)\right] < \chi^2_{1-\alpha}(1) \tag{2.17}$$

where $\chi^2_{1-\alpha}(1)$ is the $1-\alpha$ quantile of a χ^2 distribution with one degree of freedom. This yields the negative log likelihood threshold

$$nLL(\hat{\theta}) + \frac{1}{2}\chi^2_{1-\alpha}(1). \tag{2.18}$$

2.3   Model Selection

The full model that we have created can be thought of as a general model. The simple model is a special case of the full model where τ_o and r have been set to zero. Since these two models are nested models we can also use Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to assess which model fits the data better. The likelihood ratio test is an asymptotic approach that compares two models by using a test statistic that is asymptotically χ^2 distributed. Information theoretical approaches like the use of AIC or BIC for model selection provides a complimentary way of comparing which models best fit the data[15]. $AIC = 2k - 2ln(\hat{L})$ where k is the number of parameters and $\hat{L}$ is the maximum value of the likelihood function for the model. The preferred model is the model with the lowest AIC value. $BIC = ln(n)k - 2ln(\hat{L})$ where n is the sample size, k is the number of parameters, and $\hat{L}$ is the maximum value of the likelihood function is used in a similar way, but is based on the likelihood function and differs from AIC in that the penalty term for the number of parameters is larger in BIC than AIC. These three goodness of fit tests will be

used to asses which models best fit a given data set.

Chapter 3

Results

Our goal is to propose a procedure for finding the best fit parameter estimates for the full and simple model and whether they are identifiable. We will start by doing this for the simple model. Then we consider the full model and discuss approaches to deal with any identifiability issues that may arise.

3.1 Simple Model

Recall, our simple model is:

$$Y_i = y_0 + X_i \tag{3.1}$$

where $X_i \sim Gamma(\mu = \tau_*, var = c\mu)$. For details on what the parameters mean, see the Model Framework section.

3.1.1 Parameter Estimation

To use our model we first need to find the maximum likelihood estimator of our gamma distribution. The mean and variance for the gamma distribution are given

below. Let X be a gamma distributed random variable with shape α and rate β. Then

$$E[X] = \frac{\alpha}{\beta} \qquad Var(X) = \frac{\alpha}{\beta^2}. \tag{3.2}$$

Let $\mu = E[X]$ and $v = Var(X)$. Then

$$\mu = \frac{\alpha}{\beta} \tag{3.3}$$

$$v = \frac{\alpha}{\beta^2} = \frac{1}{\beta} \cdot \frac{\alpha}{\beta} \tag{3.4}$$

$$= \frac{\mu}{\beta} \implies \beta = \frac{\mu}{v}. \tag{3.5}$$

Note that,

$$\alpha = \beta\mu \tag{3.6}$$

since

$$\beta = \frac{\mu}{v} \implies \alpha = \frac{\mu^2}{v}. \tag{3.7}$$

Let $x_i = Y_i - y_0$, which represents our gamma distributed random variable and let $var = c\mu$. We can substitute the following in as such:

$$\alpha_i = \frac{\mu_i^2}{c\mu_i} = \frac{\mu_i}{c} \tag{3.8}$$

$$\beta_i = \frac{\mu_i}{c\mu_i} = \frac{1}{c} \tag{3.9}$$

$$\theta = \tau_o, r, \tau_*, y_0, c \quad \text{(full model)} \tag{3.10}$$

$$\theta = \tau_*, y_0, c \quad \text{(simple model)} \tag{3.11}$$

Will will be finding the negative log likelihood nLL function since it is easier to use in application.

$$nLL\left(\theta|\vec{x}\right) = -\ln\left(\prod_{i=1}^{n} f\left(\vec{x}|\theta\right)\right) = -\ln\left(\prod_{i=1}^{n}\left[\frac{\frac{1}{c}^{\frac{\mu_i}{c}}}{\Gamma\left(\frac{\mu_i}{c}\right)}x_i^{\frac{\mu_i}{c}-1}e^{-\frac{1}{c}x_i}\right]\right)$$

$$= -\ln\left(\frac{\prod_{i=1}^{n}\frac{1}{c}^{\frac{\mu_i}{c}}}{\prod_{i=1}^{n}\Gamma\left(\frac{\mu_i}{c}\right)}\prod_{i=1}^{n}x_i^{\frac{\mu_i}{c}-1}\prod_{i=1}^{n}e^{-\frac{1}{c}x_i}\right)$$

$$= -\ln\left(\frac{\frac{1}{c}^{\sum_{i=1}^{n}\frac{\mu_i}{c}}}{\prod_{i=1}^{n}\left(\Gamma\left(\frac{\mu_i}{c}\right)\right)}\prod_{i=1}^{n}x_i^{\frac{\mu_i}{c}-1}e^{\sum_{i=1}^{n}-\frac{1}{c}x_i}\right)$$

$$-\ln\left(\prod_{i=1}^{n} f\left(\vec{x}|\theta\right)\right) = -\ln\left(\frac{\frac{1}{c}^{\sum_{i=1}^{n}\frac{\mu_i}{c}}}{\prod_{i=1}^{n}\left(\Gamma\left(\frac{\mu_i}{c}\right)\right)}\right) - ln\left(\prod_{i=1}^{n}x_i^{\frac{\mu_i}{c}-1}\right)$$

$$- ln\left(e^{\sum_{i=1}^{n}-\frac{1}{c}x_i}\right)$$

$$= \frac{-\ln\left(\frac{1}{c}\right)}{c}\sum_{i=1}^{n}\mu_i + \sum_{i=1}^{n}\ln\left(\Gamma\left(\frac{\mu_i}{c}\right)\right)$$

$$- \sum_{i=1}^{n}\left(\frac{\mu_i}{c}-1\right)\ln\left(x_i\right) + \frac{1}{c}\sum_{i=1}^{n}x_i$$

$$(3.12)$$

Substituting equations 3.8, 3.9, and 3.11, for μ_i and x_i (where $\vec{Y}$ is the vector of data points Y_i) we have the negative log likelihood function for the simple model:

$$nLL(y_0, c, \tau_* | \vec{Y}) = \frac{\ln c}{c} \sum_{i=1}^{n} (\tau_*) + \sum_{i=1}^{n} \ln\left(\Gamma\left(\frac{\tau_*}{c}\right)\right)$$
$$- \sum_{i=1}^{n} \left(\frac{\tau_*}{c} - 1\right) \ln(Y_i - y_0) + \frac{1}{c} \sum_{i=1}^{n} (Y_i - y_0)$$

(3.13)

We can find the partial derivative of each parameter and set it equal to zero so that we have a system of equations to solve for:

$$\frac{\partial}{\partial c} nLL = \frac{-\ln c}{c^2} (n\tau_*) + \sum_{i=1}^{n} \frac{\tau_* \Gamma'\left(\frac{\tau_*}{c}\right)}{c^2 \Gamma\left(\frac{\tau_*}{c}\right)}$$
$$+ \sum_{i=1}^{n} \frac{\tau_* \ln(Y_i - y_0)}{c^2} + \frac{ny_0 - \sum_{i=1}^{n} Y_i}{c^2} = 0$$
$$\frac{\partial}{\partial y_0} nLL = \sum_{i=1}^{n} \left(\frac{\tau_* - c}{c(Y_i - y_0)}\right) - \frac{n}{c} = 0$$
$$\frac{\partial}{\partial \tau_*} nLL = \frac{n \ln c}{c} - \sum_{i=1}^{n} \frac{\Gamma'\left(\frac{\tau_*}{c}\right)}{c\Gamma\left(\frac{\tau_*}{c}\right)} - \sum_{i=1}^{n} \frac{\ln(Y_i - y_0)}{c} = 0$$

(3.14)

Re-writing the partial derivative equation for y_0, we get:

$$\frac{n}{\tau_o} - \sum_{i=1}^{n} \frac{1}{Y_i - y_0} = 0$$

(3.15)

We are able to solve for y_0 computationally for a given c and τ_*, which reduces the number of parameters we need to estimate computationally to two (c and τ_*). The function `uniroot` in R was used to find the root of y_0 over $[0, min(Y_i)]$ [1]. We are unable to analytically solve this system of equations for all three parameters. We used our function parallel.optimx in R to find the parameter estimates. Using the previously described optimiation method to find the MLE form this system of

equations, we can then use pairwise scatter plots of the MLEs and likelihood profiling to check for identifiability issues with our parameter estimation procedure.

3.1.2   Identifiability: Pairwise Scatter Plots

Figure 3.1 shows the pairwise scatter plots for the MLE's for each parameter from data simulated under the simple model with no sleep inertia effect. Looking for patterns in the scatter plots will help determine if there are identifiability issues. In 3.1, the red x's are parameter estimates for the best fit parameters with a negative log likelihood value within 5% of the best fit parameters out of 1000 runs (each starting from different initial parameter values) of our estimation procedure (see Section 2.1.2). The black dots are the top 10 best fit estimates. This plot does not show any signs of non-uniqueness with the best fit parameters since the black dots cluster tightly for all the pairwise scatter plots.
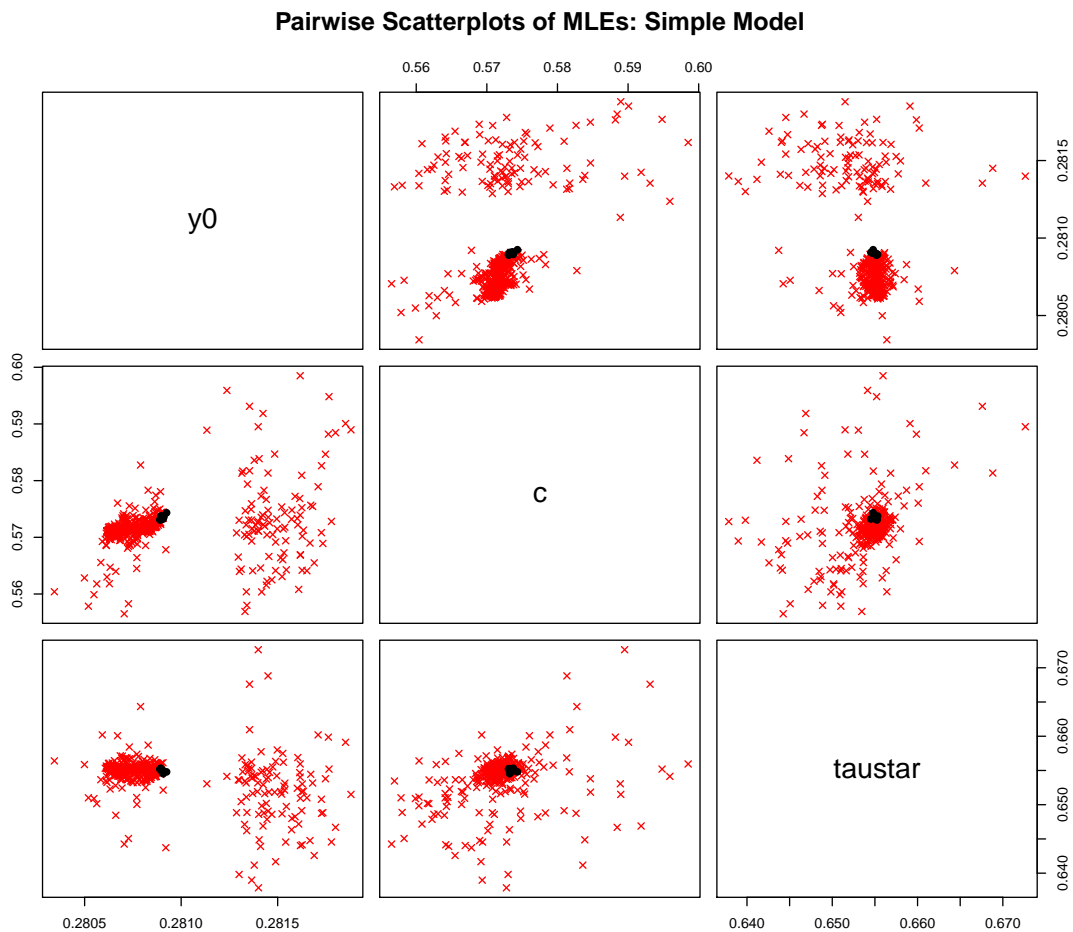
Figure 3.1: A look at the pairwise scatterplots for the best parameter estimates, for simulated data generated from and fit with the simple model. The red x's are a negative log likelihood value within 5% of the best fit parameters set out of 1000 runs starting at different initial parameter values. The black dots are the top 10 best fit estimates.

3.1.3   Identifiability: Likelihood Profiling

Another way to check for identifiability is by using likelihood profiling. By plotting the likelihood profiling for each parameter we can check for any non-uniqueness in the estimate for an isolated parameter. A look at the likelihood profiling with confidence intervals are in Figure 3.2. The red horizontal line represents the 95% confidence

interval which was found by

$$nLL(\hat{\theta}) + \frac{1}{2}\chi^2_{0.05}(1) \tag{3.16}$$

Note that the vertical lines are best fit parameter estimates (i.e. the $\hat{\tau}_*$ with the minimum negative log likelihood value). Figure 3.2 (a) shows y_0 fixed while c and τ_* are allowed to vary. The likelihood profiling for y_0 has a minimum point and does not have indications of identifiability issues. Figure 3.2 (b) shows c fixed while y_0 and τ_* are allowed to vary. The likelihood profiling for c also has a minimum point without identifiability issues. Figure 3.2 (c) shows τ_* fixed while y_0 and c are allowed to vary. The likelihood profiling for τ_* also has a minimum point and again, without apparent identifiability issues.
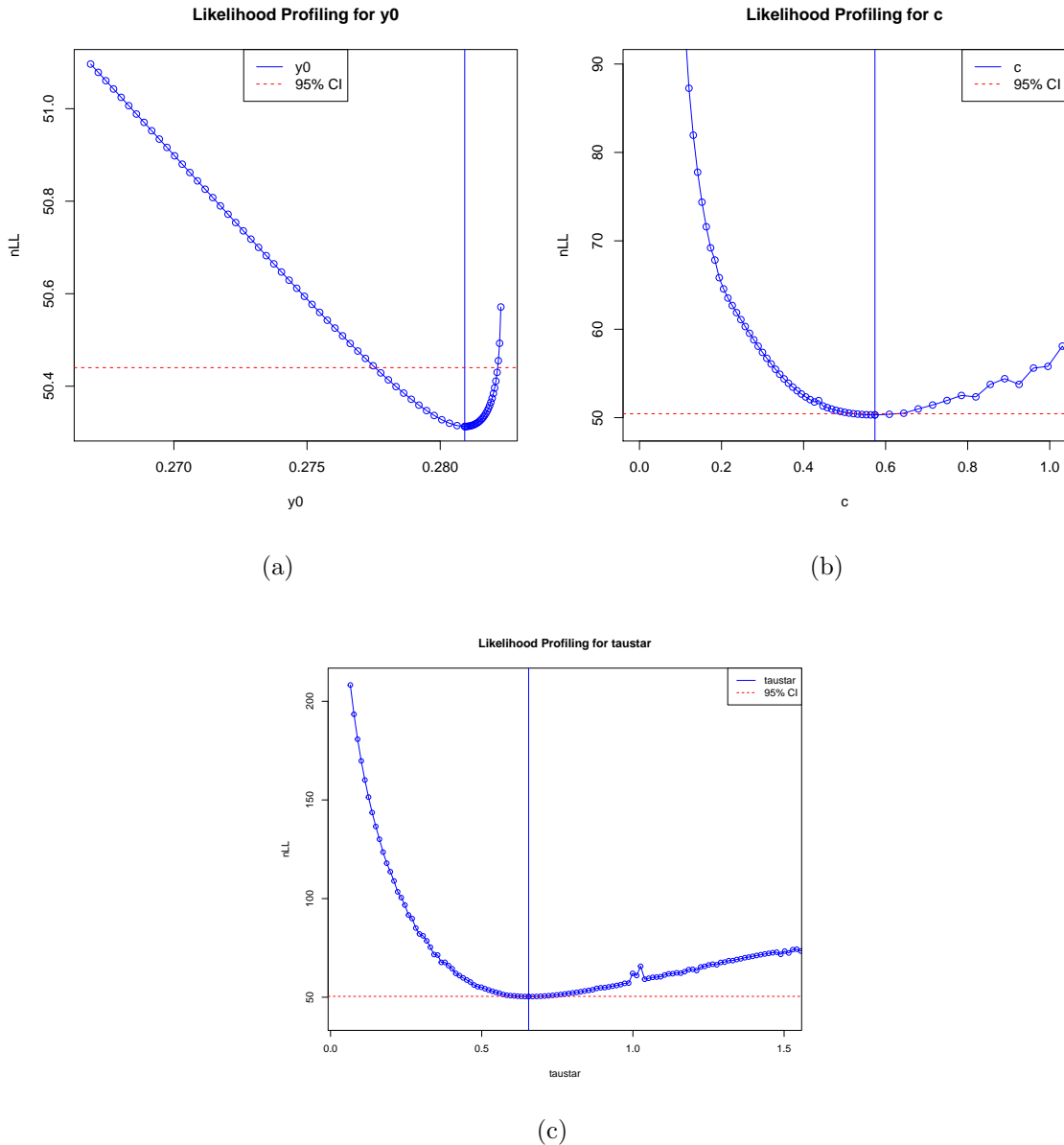
Figure 3.2: Likelihood profiling fitting the simple model to simulated data without an initial transient improvement in performance for (a) y_0, (b) c, and (c) τ_*. Simulated data came from the simple model with $y_0 = .256$, $c = 0.5$, and $\tau_* = .713$.

These results show that, for the simulated data with no sleep inertial (generated under the simple model), no identifiability issues are present and we have a best estimate for the parameters, y_0, c, and τ_*. Thus, we see that in the simple model, the MLEs are unique. This shows that this model has an identifiable (or unique)

parameter set.

3.2 Full Model

Recall, from Section 1.4 that the full model is:

$$Y_i = y_0 + X_i \tag{3.17}$$

where $X_i \sim \ Gamma(\mu = \tau_o e^{-rt_i} + \tau_*, v = c\mu)$.

3.2.1 Parameter Estimation

The negative log likelihood function for the full model is the following:

$$nLL = \frac{\ln c}{c} \sum_{i=1}^{n} \left(\tau_o e^{-rt_i} + \tau_* \right) + \sum_{i=1}^{n} \ln \left(\Gamma \left(\frac{\tau_o e^{-rt_i} + \tau_*}{c} \right) \right)$$
$$- \sum_{i=1}^{n} \left(\frac{\tau_o e^{-rt_i} + \tau_*}{c} - 1 \right) \ln \left(Y_i - y_0 \right) + \frac{1}{c} \sum_{i=1}^{n} \left(Y_i - y_0 \right) \tag{3.18}$$

The MLE for the full model is a root of the following system of equations:

$$\frac{\partial}{\partial c}nLL = \frac{-\ln c}{c^2}\left(n\tau_* + \sum_{i=1}^{n}\tau_o e^{-rt_i}\right) + \sum_{i=1}^{n}\frac{\tau_o e^{-rt_i} + \tau_*\Gamma'\left(\frac{\tau_o e^{-rt_i} + \tau_*}{c}\right)}{c^2\Gamma\left(\frac{\tau_o e^{-rt_i} + \tau_*}{c}\right)}$$

$$+ \sum_{i=1}^{n}\frac{\tau_o e^{-rt_i} + \tau_*\ln(Y_i - y_0)}{c^2} + \frac{-ny_0 + \sum_{i=1}^{n}Y_i}{c^2}$$

$$\frac{\partial}{\partial y_0}nLL = \sum_{i=1}^{n}\left(\frac{\tau_o e^{-rt_i} + \tau_* - c}{c(Y_i - y_0)}\right) + \frac{n}{c}$$

$$\frac{\partial}{\partial r}nLL = \frac{-\ln c}{c}\sum_{i=1}^{n}\tau_o t_i e^{-rt_i} + \sum_{i=1}^{n}\frac{\tau_o t_i e^{-rt_i}\Gamma'\left(\frac{\tau_o e^{-rt_i} + \tau_*}{c}\right)}{c\Gamma\left(\frac{\tau_o e^{-rt_i} + \tau_*}{c}\right)} \quad (3.19)$$

$$+ \sum_{i=1}^{n}\frac{\tau_o t_i e^{-rt_i}\ln(Y_i - y_0)}{c}$$

$$\frac{\partial}{\partial \tau_o}nLL = \frac{\ln c}{c}\sum_{i=1}^{n}e^{-rt_i} - \sum_{i=1}^{n}\frac{e^{-rt_i}\Gamma'\left(\frac{\tau_o e^{-rt_i} + \tau_*}{c}\right)}{c\Gamma\left(\frac{\tau_o e^{-rt_i} + \tau_*}{c}\right)} - \sum_{i=1}^{n}\frac{e^{-rt_i}\ln(Y_i - y_0)}{c}$$

$$\frac{\partial}{\partial \tau_*}nLL = \frac{n\ln c}{c} - \sum_{i=1}^{n}\frac{\Gamma'\left(\frac{\tau_o e^{-rt_i} + \tau_*}{c}\right)}{c\Gamma\left(\frac{\tau_o e^{-rt_i} + \tau_*}{c}\right)} - \sum_{i=1}^{n}\frac{\ln(Y_i - y_0)}{c}$$

As in the system of equations with three equations of the simple model, we are not able to analytically solve for the system of equations for the full model. Our parallel.optimx function in R was also used to find the parameter estimates for the five parameters.

3.2.2 Identifiability Analysis

Mathematically, we have identifiability issues with the full model while our simple model does not. When we look at $\mu = \tau_o e^{-rt_i} + \tau_*$ and take the taylor expantion of

e^{-rt_i}, we get

$$e^{-rt_i} = 1 - rt_i + \frac{(rt_i)^2}{2!} - \frac{(rt_i)^3}{3!} + ... \approx 1 - rt_i + O(rt_i)^2 \qquad (3.20)$$

Then,

$$\tau_o e^{-rt_i} + \tau_* \approx \tau_o + \tau_* - \tau_o rt_i \qquad (3.21)$$

There are three cases when we expect to have identifiability issues. The first case is when r is small (i.e., roughly when $0 \leq r < 0.01$), the mean of the gamma random variable in our model is

$$\tau_o e^{-rt_i} + \tau_* \approx \tau_o + \tau_* \qquad (3.22)$$

This presents an identifiability problem because we are trying to estimate τ_o and τ_* separately, but only the combination $\tau_o + \tau_*$ is identifiable. This leads to trade-offs between the MLEs for τ_o and τ_* when they yield the same sum.

Secondly, when r is really big, then the sleep inertia effect only lasts for a very short time since the curve e^{-rt_i} drops down quickly as a function of t_i. In that case, $\tau_o e^{-rt_i} + \tau_* \approx \tau_*$ for mot or all t_i and hence the likelihood calculation is basically independent of τ_o.

Lastly, if $\tau_o \approx 0$, then r has no input on the likelihood values. Thus if τ_o is small, the likelihood also be essentially independent of r.

3.2.3 Identifiability: Pairwise Scatter Plots

Figure 3.3 shows the pairwise scatter plots for the MLE's for each parameter from data simulated under the full model with sleep inertia effect. The red x's are param-

eter estimates for the best fit parameters with a negative log likelihood value within 5% of the best fit parameters out of 1000 runs (each starting from different initial parameter values) our estimate procedure (see Section 2.1.2). The black dots are the top 10 best fit estimates. The top 10 best fit estimates all cluster closely together, showing no signs of non-uniqueness with the parameter estimates, which implies no identifiability issues.

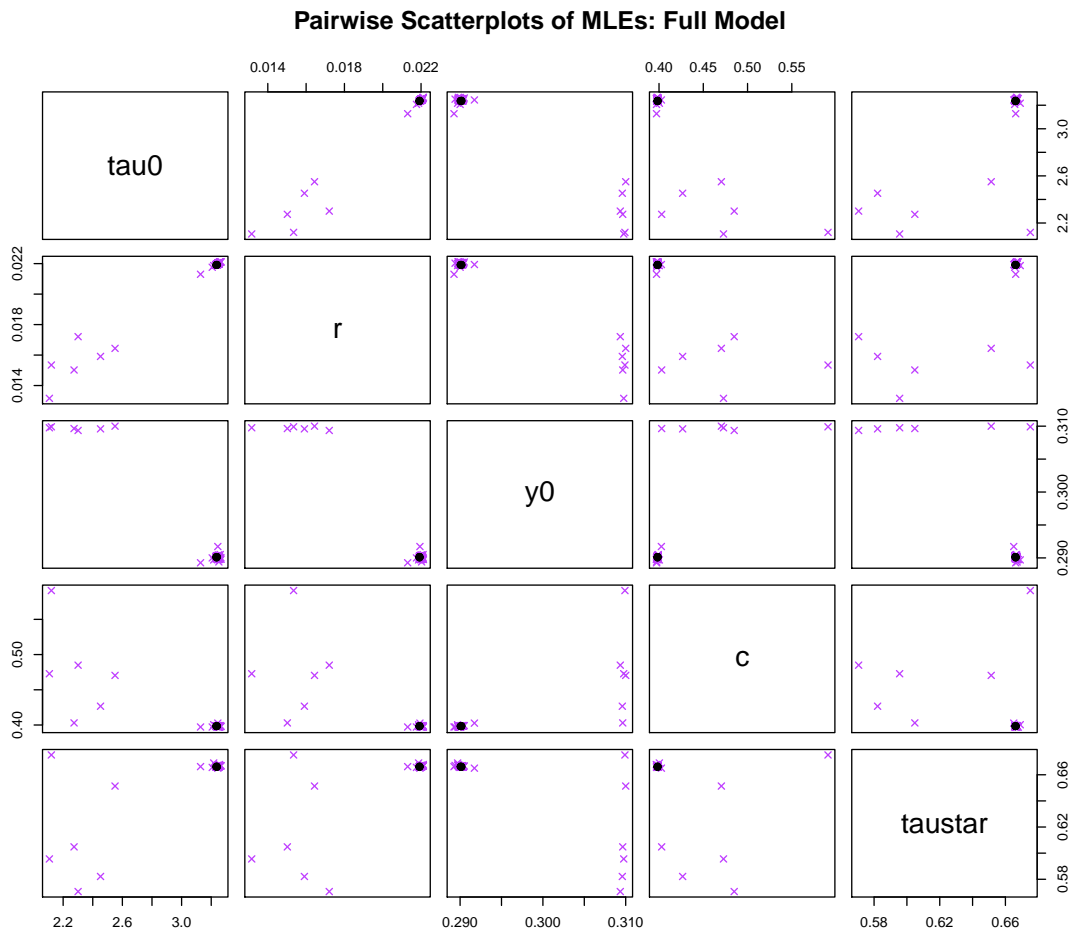Figure 3.3: Pairwise scatter plots for the best parameter estimates, for simulated data generated from and fit with the full model. The purple x's are a negative log likelihood value within 5% of the best fit parameter set out of 1000 runs starting at different initial parameter values. The black dots are the top 10 best fit estimates, which cluster tightly and indicate unique parameter estimates.

3.2.4 Identifiability: Likelihood Profiling

Figure 3.4 shows the likelihood profiling with the confidence intervals for the five parameters, τ_o, r, y_0, c, and τ_*. Just as in the simple model, the 95% confidence interval is the red horizontal line and the blue vertical line is the best fit parameter. The likelihood profiling plots for each parameter have a minimum point, showing no identifiability issues.
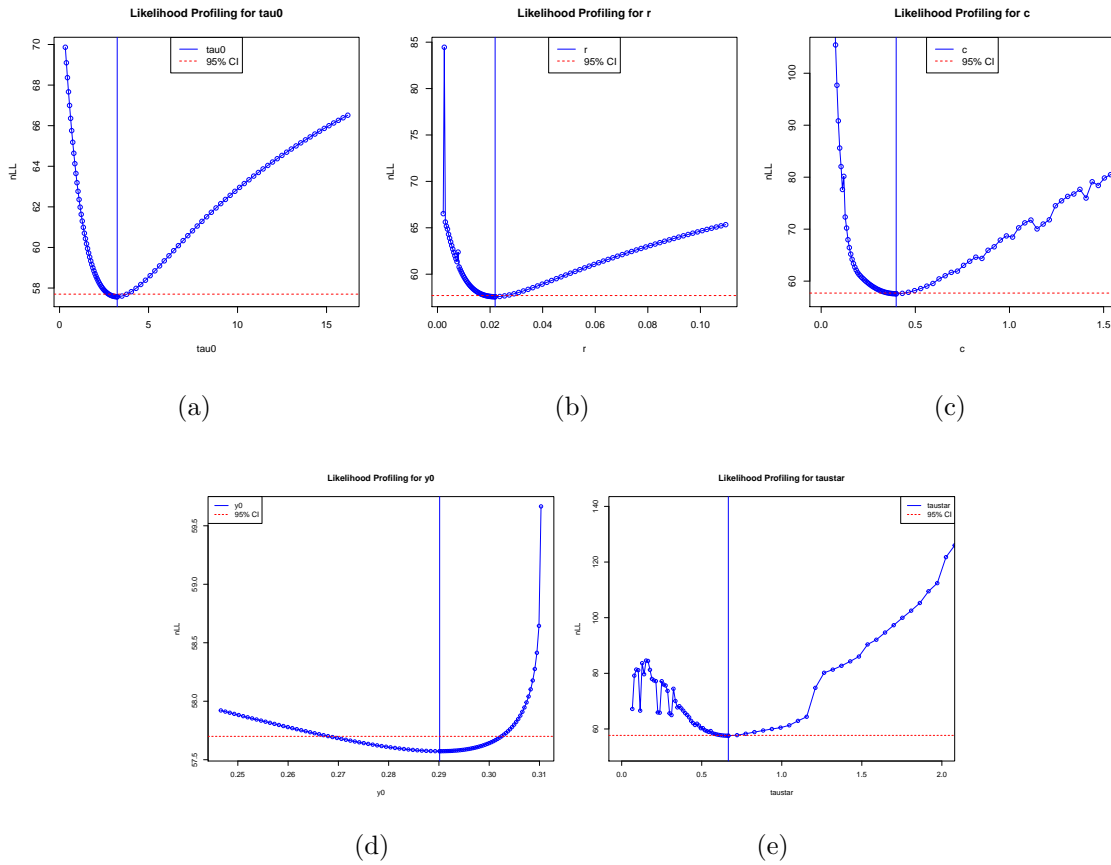

Figure 3.4: Likelihood profiling fitting the full model to simulated data with an initial transient improvement in performance for (a)τ_o, (b) r, (c) c, (d) y_0, and (e) τ_*. Simulated data came from the full model with $\tau_o = 4$, $r = 0.02$, $y_0 = .256$, $c = 0.5$, and $\tau_* = .713$.

3.3 Model Comparison

We now want to compare the nested models to test for a transient sleep inertia effect. Recall that we have already fit simulated data with no sleep inertia effect to the simple model, and simulated data with a sleep inertia effect to the full model. There we illustrate how to fit both models to a given data set, and then use the likelihood ratio test and AIC and BIC values to asses whether or not there is a significant initial transient improvement in performances.

3.3.1 Simulated Data with no Sleep Inertia Effect

Recall, we are using the simulated data that was created under the simple model, so it has no sleep inertia effect. Here we fit the full model to that data. Figure 3.5 shows scatter plots of pairwise MLEs fit to the full model. By fitting the the full model to the simple data, we are able to check for identifiability issues within the model. Here we can see a trend line between τ_o and r which shows trade-offs between these two parameters. There are also trade-offs between τ_o and τ_*. There is a straight line when plotting these two parameters against each other.
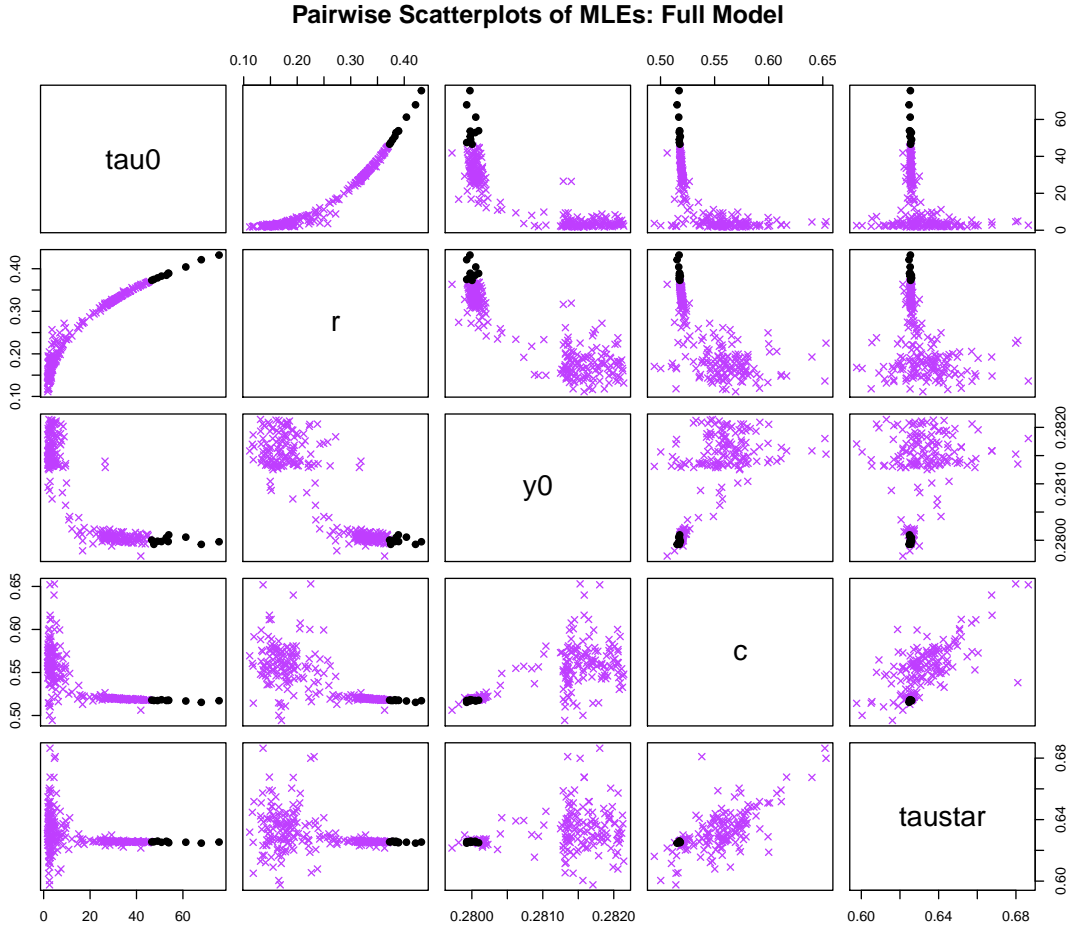
Pairwise Scatterplots of MLEs: Full Model

Figure 3.5: Pairwise scatter plots for the best parameter estimates, for simulated data generated from the simple model (no sleep inertia effect) and fit with the full model. The purple x's are a negative log likelihood value within 5% of the best fit parameter set out of 1000 runs starting at different initial parameter values. The black dots are the top 10 best fit estimates. Note the apparent patterns between τ_o, r, and τ_* highlighting identifiability issues.

Figure 3.6 shows the likelihood profiling results for all five parameters of the full model fit to the simple data. The likelihood profiling for τ_o shows the fixed parameter values leveling out by 200. This shows that multiple values of τ_o correspond to the same negative log likelihood value, which highlights identifiability issues. The same can be shown for r. There are multiple values for r that correspond to the same negative log likelihood value. The other three parameters show unique best fit values,
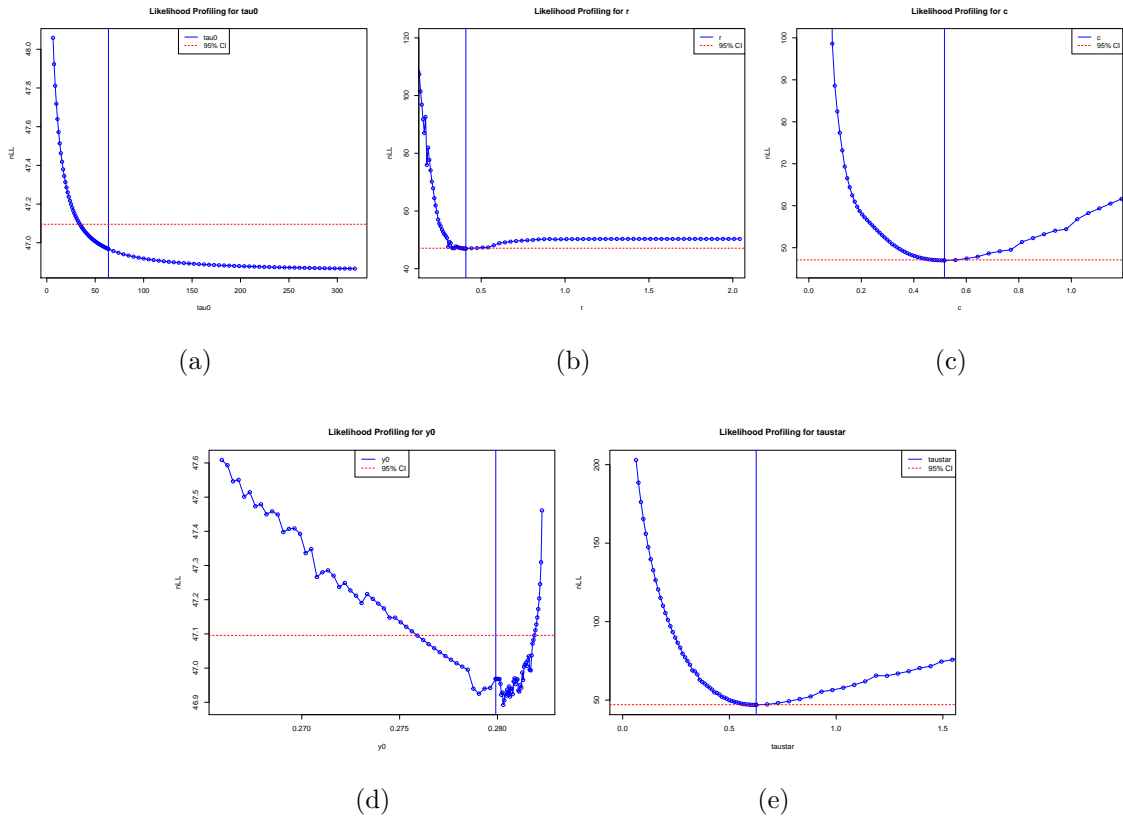
so they are identifiable.

Figure 3.6: Likelihood profiling fitting the full model to simulated data with no initial transient improvement in performance for (a)τ_o, (b) r, (c) c, (d) y_0, and (e) τ_*. Simulated data came from the simple model with $y_0 = .256$, $c = 0.5$, and $\tau_* = .713$.

We used the likelihood ratio test to see whether there was a significant difference in model fits. Results from the likelihood ratio test were the as follows: the likelihood ratio was 6.7302, the χ_2 critical value was 5.9915, degrees of freedom was 2, and the p-value was 0.03456. Table 3.1 shows the AIC and BIC results. AIC and BIC values with a difference of more than 3 or 4 are often interpreted as significantly different model fits.

Table 3.1: Data with no sleep inertia effect

Model	$nLL(\hat{\theta}; x)$	n	AIC	BIC	Δ AIC	Δ
Simple	50.31289	88	106.6258	114.0578	2.7302	0
Full	46.94778	88	103.8956	116.2822	0	2.22642

Based on the likelihood ratio test and the AIC and BIC, the decision would be that the simple model fits the data without sleep inertia better.

3.3.2 Simulated Data with a Sleep Inertia Effect

Here, we will be using the simulated data that was created by the full model. Figure 3.7 shows the pairwise scatter plots of the simple model fit to the simulated data with a sleep inertia effect. This figure does not show identifiability issues since the top 10 best fit parameters cluster tightly. Figure 3.8 shows the likelihood profiling of the simple model fit to the sleep inertia data. The likelihood profiling shows a minimum point which suggest that there are no identifiability issues with these parameters.
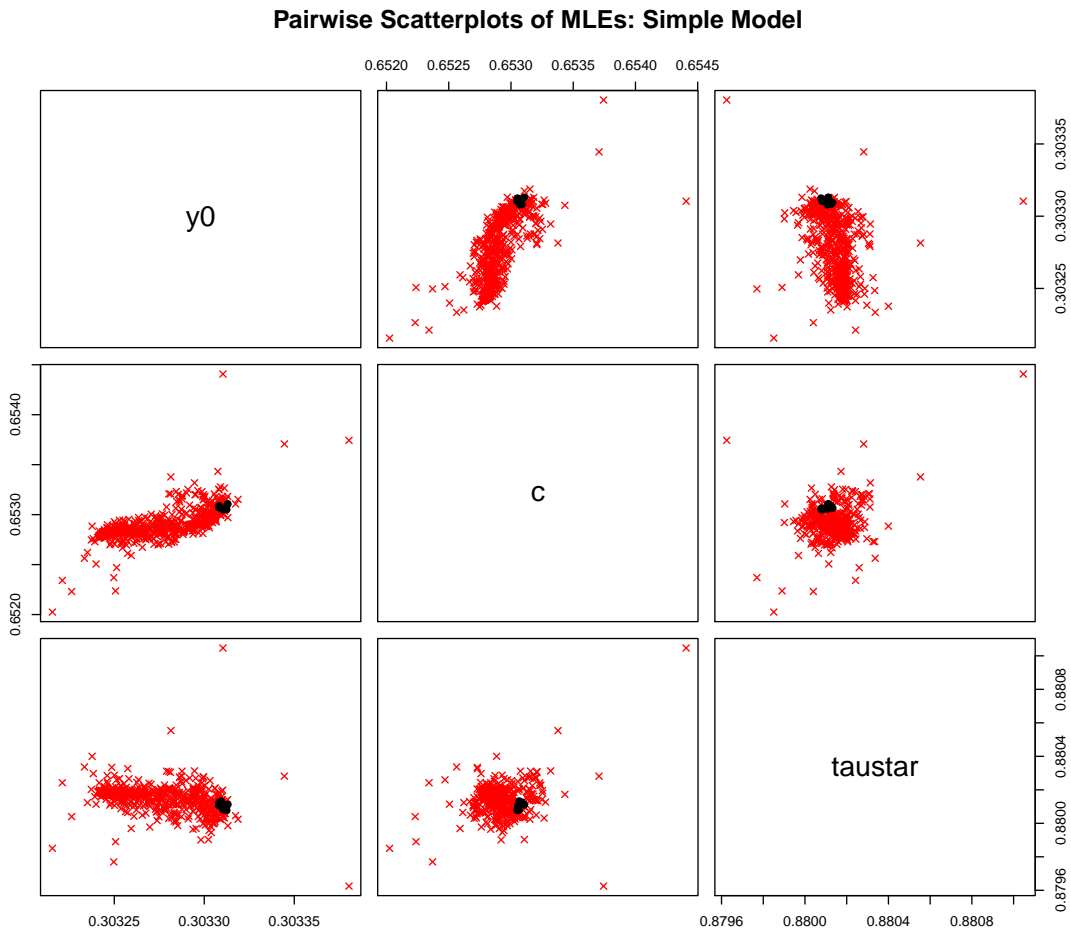
Figure 3.7: Simple model fit to the simulated data without a sleep inertia effect. As before, the purple x's are a negative log likelihood value within 5% of the best fit parameter set out of 1000 runs starting at different initial parameter values. The black dots are the top 10 best fit estimates, which cluster tightly and indicate no identifiability issues.
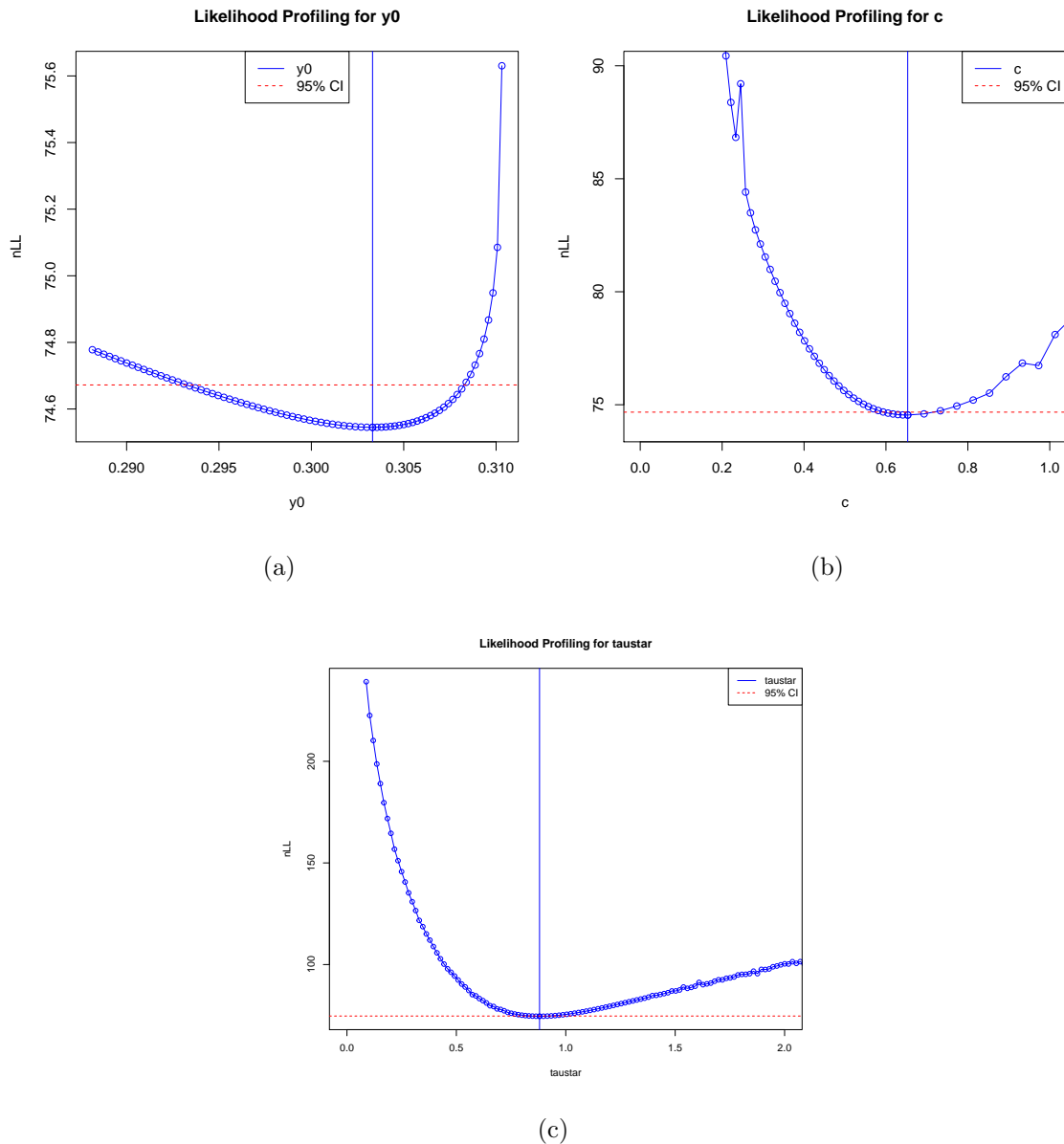
Figure 3.8: Likelihood profiling fitting the simple model to simulated data with a sleep inertia effect for (a) y_0, (b) c, and (c) τ_*. Likelihood profiling plots for each parameter have a minimum point which indicates a unique parameter estimate.

Results from the likelihood ratio test were as follows: the likelihood ratio was 33.9447, the χ_2 critical value was 5.9915, degrees of freedom was 2, and the p-value was 4.256e-08. In this case it doesn't make a difference if $\alpha = 0.05$ or $\alpha = 0.01$, in either case the decision would be to reject the null hypothesis and conclude that the

full model fits the data better. Table 3.2 shows the AIC and BIC results.

Table 3.2: Data with sleep inertia effect

Model	$nLL(\hat{\theta}; x)$	n	AIC	BIC	Δ AIC	Δ BIC
Simple	74.54499	88	155.09	162.522	29.9448	24.9901
Full	57.57262	88	125.1452	137.5319	0	0

Both the AIC and BIC are lower for the full model confirming that the full model fits the data with sleep inertia better.

3.3.3 Real Data with a Sleep Inertia Effect

We used 12 year child (ID 4031) as an example from the real data set. There looks to be a weak trend of sleep inertia and so we would like to do a model comparison with this data. Figure 3.9 shows pairwise scatter plots of the parameters of the simple model fit to this data. These plots show tight clustering of the best fit parameter estimates, indicating that they are identifiable (or unique).

Figure 3.9: Pairwise scatter plots for the best parameter estimates, for the simple model fit to real data (12 year child ID 4031). The red x's are a negative log likelihood value within 5% of the best fit parameters set out of 1000 runs starting at different initial parameter values. The black dots are the top 10 best fit estimates.

Figure 3.10 shows likelihood profiling of the simple model fit to this data. No signs of identifiability issues are present in any of the likelihood profiling for the parameters; each plot has a minimum value which corresponds to a unique best estimate.

Likelihood Profile for y0

(a)

Likelihood Profile for c

(b)

Likelihood Profile for taustar

(c)

Figure 3.10: Likelihood profiling fitting the simple model to real data (child ID 4031) for (a) y_0, (b) c, and (c) τ_*. Each likelihood profiling plot appears to have a unique best estimate.

Figure 3.11 shows the pairwise scatter plots of the MLEs for each parameter of the full model to this data. Here we can see identifiability issues between the parameters.

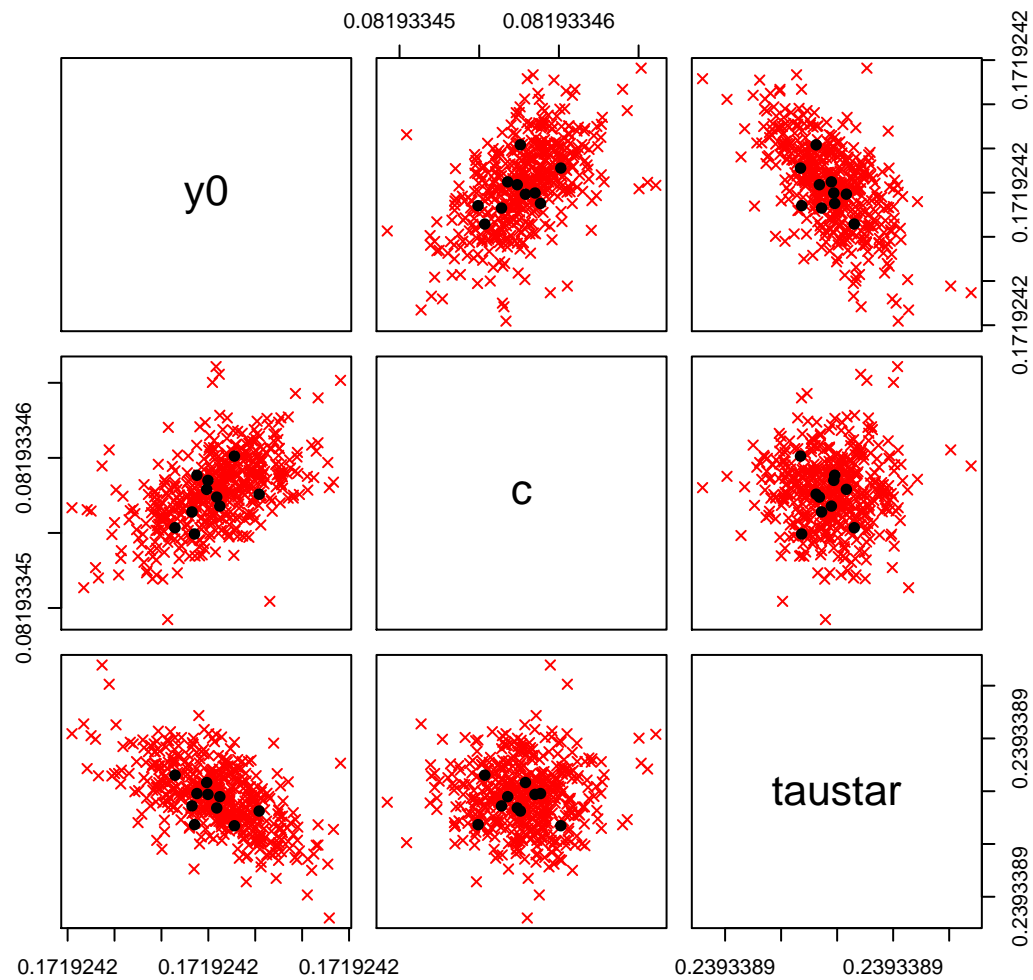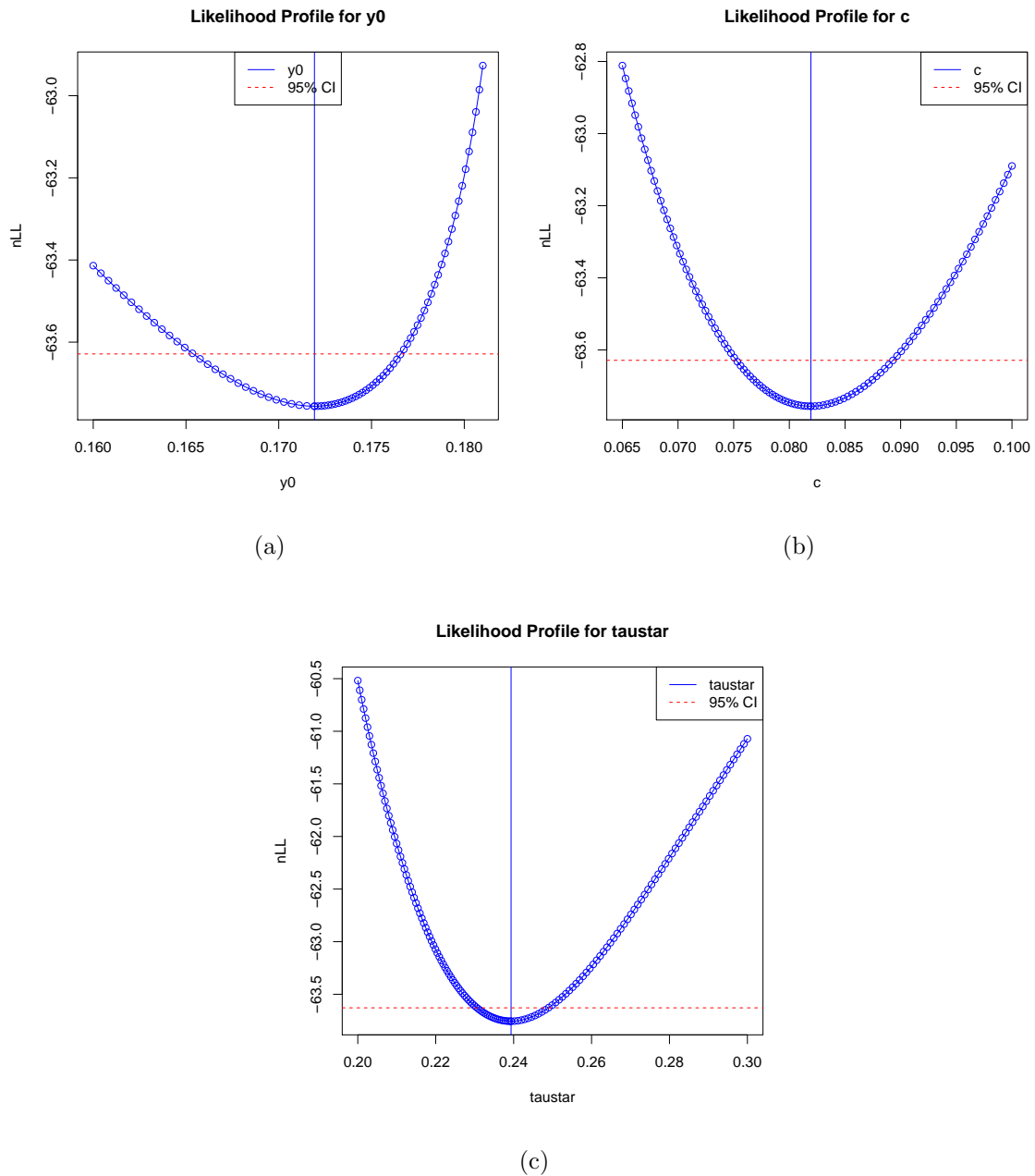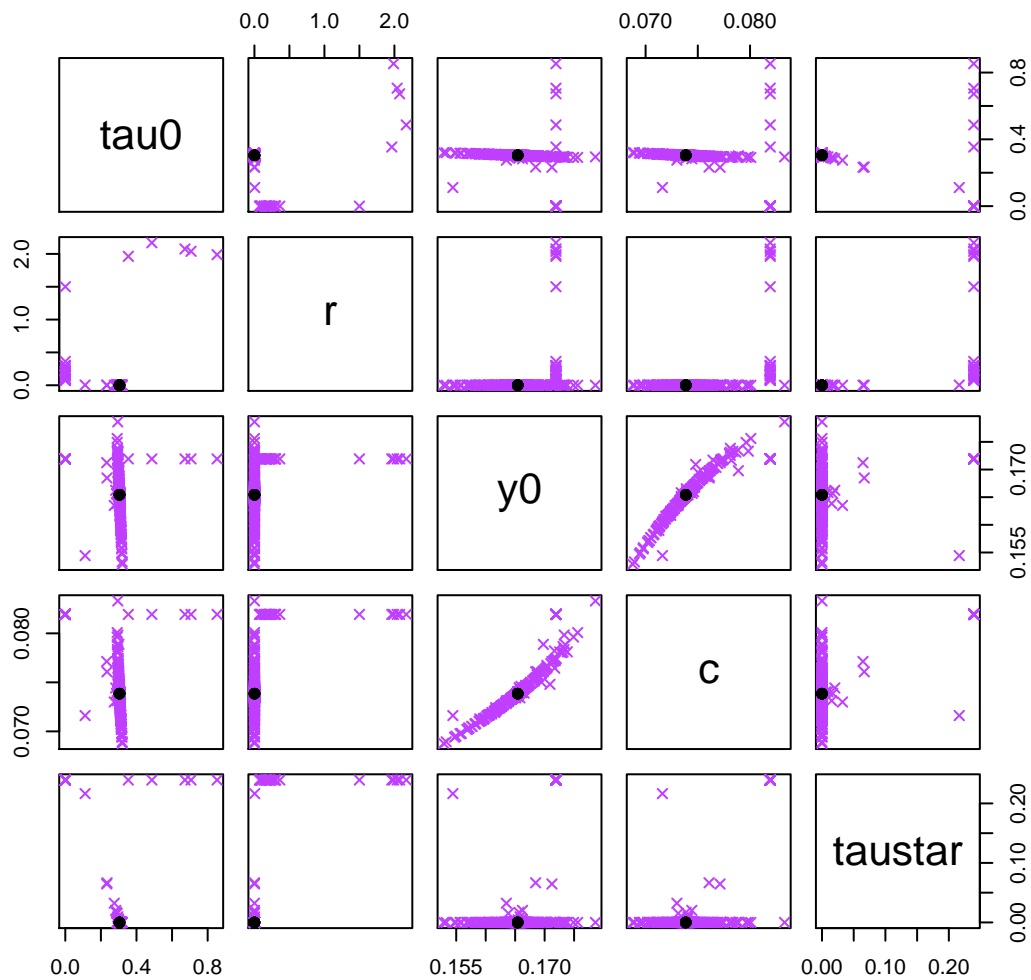**Pairwise Scatterplots of Estimates: Full Model**

Figure 3.11: Pairwise scatter plots for the best parameter estimates, for the full model fit to real data (12 year child ID 4031). The purple x's are a negative log likelihood value within 5% of the best fit parameters set out of 1000 runs starting at different initial parameter values. The black dots are the top 10 best fit estimates.

Figure 3.12 shows likelihood profiling for each parameter of the full model fit to this data. Plots for τ_o and r contain near-horizontal lines which shows that every parameter of τ_o and r correspond to the same negative log likelihood value. We also see identifiability issues with τ_* getting better fits the closer it gets to 0. This is likely due to the trade-off between τ_o and τ_*.
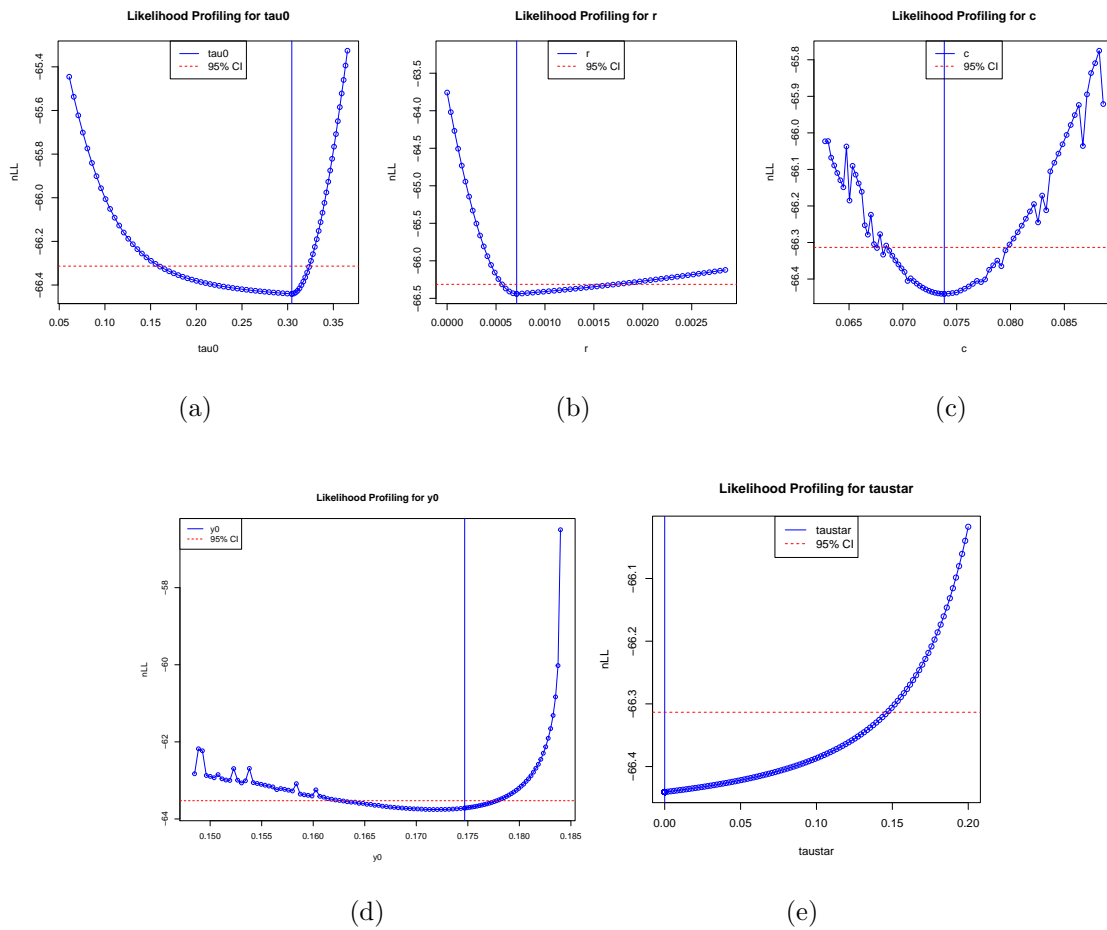
Figure 3.12: Likelihood profiling fitting the full model to real data (child ID 4031) for (a) τ_o, (b) r, (c) c, (d) y_0, and (e) τ_*.

Results from the likelihood ratio test were as follows: the likelihood ratio was 5.3692, the χ_2 critical value was 5.991, degrees of freedom was 2, and the p-value was 0.06825. These results show that the full model fits the data reasonably well, and fits better than the simple model at any significance level $\alpha > 0.06825$. Table 3.3 shows the AIC and BIC results.

Table 3.3: Model Comparison for Real Data (child ID 4031)

Model	$nLL(\hat{\theta};x)$	n	AIC	BIC	Δ AIC	Δ BIC
Simple	-63.75605	95	-121.5121	-113.8505	1.3692	0
Full	-66.44065	95	-122.8813	-110.1119	0	3.738554

The AIC value is lower for the full model whereas BIC is lower for the simple model, and Δ AIC and Δ BIC are both less than 4, which suggests that both the full model and the simple model are fitting the data well but not significantly differently. The mixed results for AIC and BIC are expected given that the p-value for likelihood ratio test was close to 0.05 but not less than it.

Chapter 4

Discussion

4.1 Conclusion

Our goal was to come up with a flexible model that better captured PVT data including the effects of sleep inertia. PVT data are skewed and non-stationary, so using just the mean to model the data is not the best method. The data has high variability, causing the mean to be inflated. This can be seen when we compare the mean and the median. Using just the mean misses the whole effect that sleep inertia has in the data. This shows the need for a more flexible model specific to this biological process. In addition kids have the special issue of getting distracted and/or fatigued towards the end of the 10 minute PVT.

We came up with a general model that can be used to model sleep inertia and that could also be transformed into a simple model to be used on PVT data with no sleep inertia present. We used techniques such as finding the likelihood function to find best fit parameters, pairwise scatter plots of MLES of the parameters and likelihood profiling to test for identifiability issues in our data, and since we had nested models we were able to use the likelihood ratio test to make decisions on which model fits our data best. The full model showed identifiability issues in some cases. Analyzing

real data showed that the full model can be a better fit than the simple model, but also outlined a need for future work.

Our results showcase that data with a strong sleep inertia effect fit the full model better while data with no clear sleep inertia effect better fit with the simple model.

4.2   Future Work

The models in this thesis do not include the fatigue that is present in some of the data. The real data showed that the identifiability issues were complicated which signifies the need for future work in adding in the fatigue. Possible ways to model the distraction and/or fatigue would be looking at time varying variance, mixture gamma distributions, etc. We also didn't look at any bias in the estimates. Future work would eventually include applying these methods to the full data set and repeating age and gender level analysis.

Chapter 5

Appendix

5.1 Gamma Distribution

The Gamma Distribution is a continuous probability distribution consisting of two parameters, the shape parameter and the scale parameter. Special cases for the Gamma distribution are the Exponential distribution, the Erlang distribution, and the Chi-squared distiribution. A random variable, X that is Gamma distributed with shape α and rate β is denoted:

$$X \sim \Gamma(\alpha, \beta)$$

The PDF for the Gamma Distribution is the following:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \text{ for x,} \alpha, \beta > 0$$

Where $\Gamma(\alpha)$ is the gamma function.

References

[1] Brent, R. P. (2013). *Algorithms for minimization without derivatives.* Courier Corporation.

[2] Burke, T. M., Scheer, F. A. J. L., Ronda, J. M., Czeisler, C. A., and Wright, K. P. (2015). Sleep inertia, sleep homeostatic and circadian influences on higher-order cognitive functions. *Journal of Sleep Research*, 24(4):364–371.

[3] Eisenberg, M. C. and Hayashi, M. A. (2014). Determining identifiable parameter combinations using subset profiling. *Mathematical Biosciences*, 256:116 – 126.

[4] Folkard, S. and Åkerstedt, T. (1992). A three-process model of the regulation of alertness-sleepiness. *Sleep, arousal and performance*, pages 11–26.

[5] Jewett, M. E., Wyatt, J. K., Ritz-DeCecco, A., Khalsa, S. B., Dijk, D.-J., and Czeisler, C. A. (1999). Time course of sleep inertia dissipation in human performance and alertness. *Journal of sleep research*, 8(1):1–8.

[6] Kinderknecht, K. (2013). Sleep inertia in children.

[7] Morris H. DeGroot, M. J. S. (2002). *Probability and Statistics.*

[8] Nash, J. C., Varadhan, R., et al. (2011). Unifying optimization algorithms to aid software system users: optimx for R. *Journal of Statistical Software*, 43(9):1–14.

[9] Newman, R. A., Kamimori, G. H., Wesensten, N. J., Picchioni, D., and Balkin, T. J. (2013). Caffeine gum minimizes sleep inertia. *Perceptual and Motor Skills*, 116(1):280–293. PMID: 23829154.

[10] Powell, J. W. (1999). *PVT-192 User's Guide and Reference Manual*. Ambulatory Monitoring, Inc.

[11] R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[12] Ran, Z.-Y. and Hu, B.-G. (2014). Determining structural identifiability of parameter learning machines. *Neurocomputing*, 127:88 – 97. Advances in Intelligent Systems.

[13] Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmuller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929.

[14] Richard J. Larsen, M. L. M. (2012). *An Introduction to Mathematical Statistics and its Applications*.

[15] Sheather, S. J. (2009). *A Modern Approach to Regression with R*.

[16] Splaingard, M., Hayes, J., and Smith, G. A. (2007). Impairment of reaction time among children awakened during stage 4 sleep. *Sleep*, 30(1):104–108.

[17] Sun, Y., Lim, J., Kwok, K., and Bezerianos, A. (2014). Functional cortical connectivity analysis of mental fatigue unmasks hemispheric asymmetry and changes in small-world networks. *Brain and Cognition*, 85:220 – 230.

[18] Tassi, P. and Muzet, A. (2000). Sleep inertia. *Sleep Medicine Reviews*, 4(4):341 – 353.

[19] Van Dongen, H. P., Belenky, G., and Krueger, J. M. (2011). A local, bottom-up perspective on sleep deprivation and neurobehavioral performance. *Current topics in medicinal chemistry*, 11(19):2414–2422.