6-30-2016

Novel Methods for Analyzing Longitudinal Data with Measurement Error in the Time Variable

Caroline Munindi Mulatya
University of South Carolina

Novel Methods for Analyzing Longitudinal Data with Measurement Error in the Time Variable

by

Caroline Munindi Mulatya

Bachelor of Science
University of Nairobi 2005

Master of Statistics
Universiteit Hasselt 2008

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Biostatistics

The Norman J. Arnold School of Public Health

University of South Carolina

2016

Accepted by:

Alexander C. McLain, Major Professor

Bo Cai, Committee Member

James W. Hardin, Committee Member

Paul S. Albert, External Committee Member

Lacy Ford, Senior Vice Provost and Dean of Graduate Studies

Acknowledgments

First and foremost, I am extremely grateful to God for making my career dreams come true. Thank you Lord for creating an opportunity for me to pursue a doctorate degree in Biostatistics at the University of South Carolina.

Second, I would like to express my sincere gratitude to my academic and dissertation advisor, Dr. Alexander C. McLain for his training, mentorship, patience and support over the last 3 years. You have invested a lot of time to instill the best research skills and values in me and I do not take that for granted. I also thank you for appreciating my research strengths and patiently encouraging me to improve in my weaker areas. Graduate school can be difficult especially for the international students. I will be forever grateful for the conductive work environment you created and for making me feel at home.

My sincere gratitude also goes to my dissertation committee members: Dr. Bo Cai, Dr. James W. Hardin and Dr. Paul S. Albert. I am extremely grateful for all your contributions towards my research work and the time you have invested to make my research work successful. In addition, I would like to thank the *Eunice Kennedy Shriver* NICHD for allowing us to use data from the Consortium on Safe Labor study.

Finally and not the least, I would like to thank my family for their encouragement and unwavering support throughout my career pursuit. Most of all, I would like to thank my mother Grace Kalunda Mulatya and my loving and supportive husband Dr. Daniel M. Mutisya; your prayers and support during my entire Ph.D studies are very much appreciated.

Abstract

In some longitudinal studies, the observed time points are often confounded with measurement error due to the sampling conditions, resulting into data with measurement error in the time variable. This type of data occurs mainly in observational studies when the onset of a longitudinal process is unknown or in clinical trials when individual visits do not take place as specified by the study protocol, but are often rounded to coincide with the study protocol. Methodological and inferential implications of error in time varying covariates for both linear and nonlinear models have been studied widely. In this dissertation, we shift attention to another source of measurement error in the time variable in longitudinal studies. Specifically, we develop statistical methods for analyzing longitudinal data when the onset of the process is unknown. This work has been motivated by a cervical dilation data from the Consortium on Safe Labor (CSL) study, a multi-center retrospective observational study conducted by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development. The uncertainty in onset of labor poses methodological challenges since the observed time variable is related to when women get to the hospital, not the biologic process of interest. In Chapter II, we present a Longitudinal Threshold Regression model for estimating the distribution of the time a woman's cervical dilation takes to progress from one threshold to another (in cm). In Chapter III, we present a Semi-parametric model with random shift parameters for modeling labor curves prospectively. In Chapter IV, we extend Chapter III to predict women's time to full dilation given their past measurements. We demonstrate the proposed methods using simulation studies and a data from the CSL study.

Table of Contents

Acknowledgments iii

Abstract iv

List of Tables vii

List of Figures viii

Chapter 1 Introduction 1

Chapter 2 Estimating time to event characteristics via longitudinal threshold regression models 6

2.1 Abstract 6

2.2 Introduction 6

2.3 Methodology 11

2.4 Simulation Studies 17

2.5 Application on Consortium on Safe Labor data 26

2.6 Conclusion 30

2.7 Future Research 31

Chapter 3 Penalized splines mixed effects model with random time-shift parameters 34

3.1 Abstract 34

3.2 Introduction 35

3.3 Methodology 39

3.4 Simulation 52

3.5 Application on Consortium on Safe Labor data 54

3.6 Conclusion 56

3.7 Recommendation 57

Chapter 4 Dynamic Individualized Prediction in presence of Measurement Error in the time variable 58

4.1 Abstract 58

4.2 Introduction 59

4.3 Methodology 60

4.4 Application on a simulated data 62

4.5 Conclusion 63

4.6 Recommendation 64

Chapter 5 Summary and Conclusion 65

Bibliography 69

List of Tables

Table 1.1 Example of data from the Consortium on Safe Labor study 4

Table 2.1 Parameter estimates from the LTR and IC model for the continuous outcome 19

Table 2.2 Parameter estimates from the LTR model for data rounded to the nearest 0.5 19

Table 2.3 Parameter estimates given $ni = 7$ (sparse) and $ni = 36$ (frequent) measurements for continuous outcome 21

Table 2.4 Parameter estimates given $ni = 7$ (sparse) and $ni = 36$ (frequent) measurements for data rounded to the nearest 0.5 22

Table 2.5 LTR model efficiency given $ni = 7$ (sparse) and $ni = 36$ (frequent) measurements 23

Table 2.6 Parameter estimates from unadjusted and adjusted LTR model and an unadjusted IC model with n=2000 nulliparous women . . . 27

Table 3.1 Updating the size of M at each iteration in MCEM algorithm . . . 49

Table 3.2 Summary of 200 simulated datasets with n=500 and n_i=10 53

Table 3.3 Summary of 200 simulated datasets with n=500 and n_i=5 53

Table 3.4 Summary of 200 simulated datasets with n=500, K, K_c =(2,2) knots for both ni=10 and ni=5 observations 54

Table 3.5 Summary of 200 simulated datasets with n =500, K, K_c =(5,2) knots for both ni =10 and ni=5 observations 55

Table 3.6 Parameter estimates from the semi-parametric model with n=500 nulliparous women based on K, K_c =(30,3) 55

List of Figures

Figure 1.1 Labor profiles of 10 nulliparous women from CSL study 5

Figure 2.1 Illustration of subsetting the longitudinal data for thresholds, $A_1 = 3$ and $A_2 = 4$ 13

Figure 2.2 Estimated and the true S(t) when $\beta_1 = 1, \sigma_b^2 = 0.3$ 25

Figure 2.3 Estimated and the true S(t) when $\beta_1 = 1, \sigma_b^2 = 0.15$ 25

Figure 2.4 Estimated and the true S(t) when $\beta_1 = 0.5, \sigma_b^2 = 0.15$ 26

Figure 2.5 The estimated unadjusted survival curves given $(A_1, A_2) = (3, 4)$ cm 29

Figure 2.6 The estimated unadjusted survival curves given $(A_1, A_2) = (7, 8)$ cm 29

Figure 3.1 Estimated individual profiles. $n = 500, n_i = 10, K, K_c = (30, 5)$ . . 54

Figure 3.2 Estimated mean labor curve of women from the CSL study 56

Figure 4.1 ROC curves when L=0.5 hours: Panel (a): $m = 4$ and Panel (b): $m = 7$ past measurements 63

Figure 4.2 ROC curves when L=1 hours: Panel (a): $m = 4$ and Panel (b): $m = 7$ past measurements 63

Figure 4.3 ROC curves when L=2 hours: Panel (a): $m = 4$ and Panel (b): $m = 7$ past measurements 64

Chapter 1

Introduction

Longitudinal data arises in both observational and clinical research when subjects are taken measurements on an outcome of interest or surrogate marker over time. One important outcome of interest in longitudinal studies is the distribution of the time, a subject's health status takes to progress from one clinical state to another. For example in a clinical trial, individuals at risk of Human Immunodeficiency virus (HIV) may be followed until they are diagnosed with HIV virus and subsequently the Acquired Immunodeficiency Syndrome (AIDS), with interest lying on the elapsed time between the two events (AIDS incubation period). In cancer research or cognitive impairment studies, subjects might be followed to determine transition rates across cancer stages or cognitive impairment stages. Researchers may also be interested in subject - specific mean profiles, prediction of subjects' future longitudinal profiles or survival time after a clinical diagnosis.

Presence of sophisticated statistical methods with valid inferences exists for answering the above research questions when there is no error in time variable. In addition, methodological approaches for linear and nonlinear mixed effects models with error in covariates or time varying covariates have been studied widely. For example, Wang and Davidian (1996) investigated the effect of measurement error on intra-individual covariates in a controlled variable nonlinear mixed effects model. Using simulated data, they found that though population estimates may be less biased, there are substantial bias in intra-subject variance parameters. Tosteson et al. (1998) found that in a mixed model with measurement error in time varying covariates,

the estimated variance of random effects have positive bias when the error in time varying variable is ignored. On the other end, the estimates of slope coefficients tend to be attenuated. These inferences were based on large sample estimation and were drawn from analyses of plasma levels and dietary intake of beta-carotene where the dietary intake was modeled as a time varying covariate subject to measurement error. Liang et al. (2003) analyzed the association between HIV viral load and CD4+ cell counts during HIV/(AIDS) treatment using a mixed effects varying-coefficient model with measurement error. Both HIV viral load and CD4+ cell counts are prone to measurement error. Through simulation and application on data from air population studies, Schwartz and Coull (2003) demonstrated how to account for confounding effects in multi-level models when some predictors are measured with unknown error. The proposed methods were used to investigate the association of daily deaths and gaseous air pollution while controlling for the impact of particles of different sizes , among others objectives. In multi-level models, presence of predictor variables measured with uncertainty might introduce bias on association estimates in sub-studies and in the combined estimate across studies.

Despite existence of effective methods for analyzing longitudinal data with time varying covariates and predictors prone to measurement error, there exists few statistical methods for handling longitudinal data where the time onset of the process is measured with uncertainty. When the onset of a longitudinal process is unknown or known with uncertainty, a methodological challenge arises since the observed time is related to when the subjects were enrolled in the study rather than the biological process of interest. This poses methodological challenge especially if one is interested in predicting individuals future profiles or time to a characteristic given subjects previous measurements.

In this dissertation, we present statistical methods for analyzing longitudinal data with measurement error in the time variable. The methods are motivated by a cervical

dilation data from Consortium on Safe Labor study that is explained in details in Section 1.1. We given an outline of the dissertation in Section 1.2.

1.1 Motivation data

Consortium on Safe Labor study

The Consortium on Safe Labor (CSL) study was a multi-center retrospective observational study conducted by the *Eunice Kennedy Shriver* National Institute of Child health and Human Development (NICHD). The study included women giving birth between 2002 to 2008 and its primary objective was to describe contemporary labor patterns and establish the time to cesarean delivery in women with labor protraction and arrest. The study encompassed of 19 affiliated hospitals chosen from 12 clinical centers across 9 American College of Obstetricians and Gynecologists (ACOG) US districts. In total there were 228,668 deliveries with 233,844 newborns between 2002 and 2008 included in the study. For women who had more than one delivery in the course of the study, only the first delivery was selected to avoid within woman correlation. Each woman's detailed information on delivery, labor, maternal demographic characteristics , medical history, reproductive , prenatal history, postpartum and newborn information were extracted from electronic medical records from the selected institutions, Zhang et al. (2010a,b); Laughon et al. (2012). The CSL study protocol was approved by institutional review boards for all the participating institutions.

The observations for the women in the CSL begin when they are admitted into the hospital at different stages of cervical dilation and the onset of labor may be unknown. In addition, not all women attain full dilation and obstetricians often opt for alternative birth procedures like Caesarian sections leading to censored cervical dilation curves. Another characteristic of cervical dilation measurements is that, the measurements are prone to measurement error partly due to the ad hoc methods used

in assessing the dilation process and women tend to have nonlinear labor patterns that deviate from the population curve.

The CSL data used in this dissertation comes from the Utah site and consists of the following variables, Id: the indicator variable for the subjects (woman), Dilation: the outcome variable (cervical dilation measurements in cm), BMI: the subjects BMI categories (woman Body Mass Index (kg/m^2)), Age: the subjects age (woman age in years) and Time: the time since the subjects were admitted into the hospital (observation time point in hours). An example of CSL data is shown in Table 1.1 and in Figure 1.1 we display cervical dilation curves for 10 women from the CSL study.

Table 1.1 Example of data from the Consortium on Safe Labor study

Id	Dilation	BMI	Age	Time (hours)
1	3.0	1	25	0.00
1	3.0	1	25	1.52
1	5.0	1	25	2.65
1	5.0	1	25	3.72
1	5.0	1	25	4.48
1	6.0	1	25	5.13
1	9.0	1	25	6.98
1	10.0	1	25	7.53
1	10.0	1	25	8.30
2	3.0	2	21	0.00
2	3.0	2	21	1.87
2	4.0	2	21	6.25
2	4.0	2	21	9.25
2	5.0	2	21	10.90
2	7.0	2	21	11.63
2	7.0	2	21	12.97
2	8.0	2	21	14.50
2	9.0	2	21	15.40
2	9.5	2	21	16.27
2	10.0	2	21	18.30
.	.	.	.	.
.	.	.	.	.

Figure 1.1 shows observed cervical dilation and the observed time since women were admitted in the hospital. From Figure 1.1, we infer that women's first cervical dilation

measurements since admission to the hospital tend to be greater than 0 and vary among women. We also observe that, the labor curves are nonlinear and of different trends and not all women attain full dilation thus some women have censored labor curves.

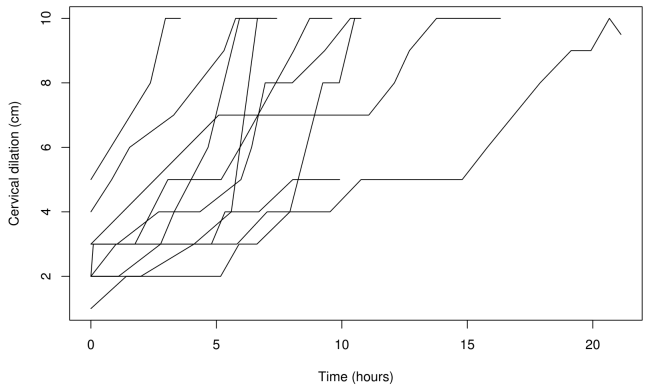


Figure 1.1 Labor profiles of 10 nulliparous women from CSL study

1.2 Outline of Dissertation

In Chapter 2 , we presence a Longitudinal Threshold Regression model for estimating the distribution of the time a woman's cervical dilation takes to progress from one threshold to another in centimeters (cm). In Chapter 3, we present a Semi-parametric model with random shift time parameters for modeling labor curves prospectively. In Chapter 4, we extend the method in Chapter 3 to predict women's time to full dilation given their past measurements. Then give a summary and overall conclusion.

Chapter 2

Estimating time to event characteristics via longitudinal threshold regression models

2.1 Abstract

In longitudinal studies, it is sometimes of interest to estimate the distribution of the time a longitudinal process takes to traverse from one threshold to another. For example, the distribution of the time it takes woman's cervical dilation to progress from 3 to 4 cm can aid the decision making of obstetricians as to whether a stalled labor should be allowed to proceed or stopped in favor of other options. Often researchers treat this type of data structure as interval censored and employ traditional survival analysis methods. However, the traditional interval censored approaches are inefficient in that they do not use all of the available data. In this chapter, we propose utilizing a longitudinal threshold regression model to estimate the distribution of the elapsed time between two thresholds of the longitudinal process from repeated measurements. A Wiener process under the first hitting time (FHT) framework is used to represent survival distribution. We demonstrate our model through simulation studies and an analysis of data from the CSL study.

2.2 Introduction

It is sometimes of interest to estimate the distribution of the time it takes a longitudinal process to progress between threshold values. For example, an important obstetrics problem is to estimate the distribution of the time it takes the cervix of

a woman in labor to dilate from, say, 3 cm to 4 cm. This distribution is useful as a benchmark during labor since it gives obstetricians an idea of what proportion of women would have reached a target dilation within a given time period. If a woman has been within the 3 cm to 4 cm interval for 4 hours, the obstetrician may consider other delivery options depending on the proportion of women expected to have progressed. As a result, a main outcome of interest is the survival distribution of the time it takes to progress from 3 to 4 cm. In most cases the exact time when the process crossed the thresholds is unknown but is known to have occurred within a given interval for both thresholds leading to a doubly interval censored data.

In Section 2.2.1, we review doubly interval censored data and a few current survival methods used for analyzing this kind of data.

2.2.1 Doubly Interval Censored Data

Doubly interval censored data arises when the time an event occurs is known to fall within an interval of two related thresholds (Sun, 2004). The survival time is then estimated as the elapsed time between the two thresholds. Doubly interval censored data arises mainly in disease progression studies where the primary interest is the time a subject takes to progress from one clinical state to another. Several authors have developed doubly interval censored models to analyze HIV data where the main focus has been the estimation of the AIDS incubation time and the factors associated with the incubation period (Sun, 2004; Bacchetti, 1990; Jewell, 1994; Jewell et al., 1994; Gómez and Lagakos, 1994; Sun, 1995, 1997; Tu, 1995). An added complexity in doubly censored data is left-truncation. For example, in an AIDS cohort follow-up, only subjects who test positive for HIV virus but do not have AIDS at the time of recruitment are included in the study (Sun, 2004). Thus the subject's AIDS incubation time is left-truncated since the time from infection and recruitment in the study is not known.

Right censored and interval censored failure time data are special cases of doubly interval censored data. Doubly interval censored data reduces to right censored time data if the upper threshold is right censored and to interval censored data if the lower threshold is exactly observed, but the upper threshold is interval censored (Sun, 2004). Previous authors have used interval censored methods to estimate the distribution of time a woman's cervical dilation takes to progress from one threshold to another. For example, Zhang et al. (2002, 2010a) used a log normal model to estimate duration of labor from one cervical dilation threshold to another in cm. Whereas Zhang et al. (2002, 2010a) studies gave better findings for contemporary labor patterns than Friedman (1954), the main critique of the interval-censored method is that it does not use all of the available data. For example, if there are 10 observations known to occur within the thresholds, the interval-censored method only uses 4 (2 to bound each threshold).

2.2.2 First Hitting Time models

First Hitting Time models are alternative methods for analyzing time to event data. In FHT models, an individual's health status is assumed to be a stochastic process deteriorating over time thus the event of interest occurs when the health status hits a predefined threshold for the first time. In this context, the first hitting time variable of the stochastic process is the survival time. The resulting model is referred to as Threshold Regression (TR) model when the effect of risk factors is evaluated on the parameters of the FHT distribution. FHT models are analogous to Cox PH model but are more flexible since the hazard is a function of time and hence they do not require the assumption of constant hazard ratio (Pennell et al., 2009).

FHT models are characterized by a parent stochastic process with an initial value, a time space, a state of the process and a boundary set contained in the state of the process (Lee and Whitmore, 2006). The underlying parent stochastic process could

be a Wiener process, Markov chain process, Bernoulli process, Poisson process or Gamma process. See Lee and Whitmore (2006) for details on parent processes of FHT models.

The FHT modeling approach has recently gained popularity as a survival analysis method for biomedical studies, (Lee et al., 2010). FHT models have been successfully applied to analyze the length of stay in the hospital, birth weight, risk of lung cancer of rail workers, and recurrent exacerbations in chronic obstructive pulmonary among others (see Horrocks and Thompson (2004); Ting Lee et al. (2004); Whitmore and Su (2007); Lee et al. (2009); Aaron et al. (2010) and the references therein). The FHT models utilize longitudinal threshold regression (LTR) of a latent "health process" to motivate the form for a survival model, (Lee et al., 2010). However, these models have not been applied to biomedical data when an actual marker of the "health process" is observed. The reliability literature on LTR has considered such a scenario. To study reliability of high reliable products, the product's degradation data is recorded over a period of time and then used to predict the product's lifetime distribution. To mention but a few, Peng and Tseng (2009) modeled laser data as a Wiener process in their paper on misspecification of degradation models, Wang (2010) proposed a Wiener process with random effects for modeling reliability of bridge beam, Ye et al. (2012) analyzed wear of hard drive disks as a Wiener process with measurement error. See Lu and Meeker (1993); Meeker et al. (1998); Gorjian et al. (2010) for more examples. Unlike actual biological health process from biomedical data, reliability experiments produce data with lower variance and less unit to unit variability.

In this chapter we propose a LTR model to estimate the elapsed time between two thresholds of a cervical dilation process and draw inferences at the population level. Our method utilizes the repeated measures of the cervical dilation process to estimate the distribution of the elapsed time between the thresholds. Difficulties that arise include creating an appropriate subset of the data, length bias induced by left

truncation, heterogeneity in the baseline measurement, and residual dependence in the curves. The proposed approach utilizes all the available data within the interval of interest.

The model is developed under the FHT framework and utilizes a Wiener process as the underlying stochastic model. A benefit of this model is that the accompanying Markov property allows one to condition on a measurement as a baseline value thus correcting for the error in time variable due to the unknown onset of labor. This is useful for our data example of cervical dilation, since it has well been documented that baseline values are arbitrary since they depend on when a woman enter the hospital (cf. (McLain and Albert, 2014)).

We applied the proposed method, and an analogous interval censored (IC) method to data from the CSL study. Other examples where similar problems arise include estimating the distribution of the time it takes women to progress from different categories of bone mineral density (i.e., normal, moderate/advanced osteopenia, osteoporosis), as measured by their bone mineral density T-score (a continuous measure) traversing between thresholds, (Gourlay et al., 2012). In general, an LTR model can be applied when there exists a surrogate marker for a disease, and well defined cut points of interest (e.g., CD4 T cell count less than 200 and AIDS). Our specific aims are to develop a framework that will allow for the estimation of an LTR model for two thresholds.

This Chapter is organized as follow. In Section 2.3 we present the LTR proposed model, a brief overview of Wiener process, Robust variance estimation and a review of standard IC model. In Section 2.4 , we present simulation studies including efficiency of the proposed LTR model and model misspecification. In Section 2.5, we present the application of the proposed model and an analogous IC model on the CSL study data. We conclude the chapter in Section 2.6 and present recommendations for further study in Section 2.7.

2.3.1 Longitudinal threshold model

In this Section, we demonstrate the estimation of the distribution of the time it takes the cervical dilation processes to traverse between two thresholds when the longitudinal measurements are used for analysis. We consider the case when there are only two thresholds of interest.

We define A_1 and A_2 as the lower and upper thresholds of interest, respectively. Further, let $\boldsymbol{t}_i^* = \{t_{i1}^*, t_{i2}^*, ..., t_{in_i}^*\}$ and $\boldsymbol{Y}_i^* = \{Y_{i1}^*, Y_{i2}^*, ..., Y_{in_i}^*\}$ be the vectors of observed time points and measurements for the ith subject, respectively, where $Y_{ij}^* = Y_i^*(t_{ij}^*)$ for $i = 1, 2, ..., n,\ j = 1, 2, ..., n_i$. Let (t_0, y_0) denote the initial state, which we assume (for the moment) to be equal across subjects. The observed data consist of independent pairs of $(\boldsymbol{t}_i^*, \boldsymbol{Y}_i^*)|_{i=1,...,n}$. The survival time of interest is defined as the difference of two first hitting times, $S_{12} = S_2 - S_1$, where $S_k = \inf\{s : Y(s) \geq A_k\}$. Notice that the definition of S_k is relative to the initial state (t_0, y_0), which is commonly $(0, 0)$. Here we assume that $\Pr(S_{12} < \infty) = 1$.

Consider the general stochastic form of the longitudinal process

$$Y_{ij}^* = \mu_{ij} + \varepsilon_{ij} = \mu(t_{ij}; \boldsymbol{\beta}) + \varepsilon_{ij} \tag{2.3.1}$$

where μ_{ij} is the mean of ith subject at time t_{ij}, ε_{ij} is the error term with $\varepsilon_{ij} \sim G(\cdot|\sigma)$, and $\boldsymbol{\beta}$ is a vector of fixed parameters common to all subjects. The distribution of S_k can be defined as

$$F_{S_k}\{t|(t_0, y_0), \boldsymbol{\theta}\} = P\{S_k \leq t|(t_0, y_0), \boldsymbol{\theta}\} = P\left\{\max_{u\in[t_0,t)} Y_i^*(u) \geq A_k \Big| (t_0, y_0), \boldsymbol{\theta}\right\}. \tag{2.3.2}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$. From (2.3.2) the distribution of S_{12}, denoted by $F_{S_{12}}$, can be obtained via standard convolution formula. There are specifications of (2.3.1) where (2.3.2) and F_{S_k} simplify to known parametric distributions. For example, when ε

corresponds to a Wiener process, Gamma process, or Ornstein–Uhlenbeck process (2.3.2) corresponds to an inverse Gaussian, inverse Gamma, or Ricciardi–Sato distribution, respectively (see Lee and Whitmore Lee and Whitmore (2006) for other FHT models). Further, the Inverse Gaussian distribution is closed under convolutions and $F_{S_{12}}$ is also inverse Gaussian.

Many of the models used for (2.3.1) will have the Markov property. This is particularly convenient when the initial state (t_0, y_0) varies across subjects (as in our motivating example). When this is the case we can use an observation from $(\boldsymbol{t}_i^*, \boldsymbol{Y}_i^*)$ as the initial state. Another benefit of the Markov property is that we need not specify a model for the entire longitudinal process. Using all the longitudinal data $(\boldsymbol{t}_i^*, \boldsymbol{Y}_i^*)$ increases the efficiency of the procedure, but also increases the parametric specification required in model (2.3.1). This is unnecessary when the size of $[A_1, A_2]$ is small relative to the range of $\boldsymbol{Y}$. As a result, we limit the range of the longitudinal observations in the interest of using a more parsimonious longitudinal model.

In the interest of a parsimonious longitudinal model, we consider a subset of $(\boldsymbol{t}_i^*, \boldsymbol{Y}_i^*)$ denoted by $\mathcal{D}_i = (\boldsymbol{t}_i, \boldsymbol{Y}_i) = \{(t_{ij}, Y_{ij})_{j=1,2,\ldots,m_i}\}$ where m_i is the number of observations in $(\boldsymbol{t}_i, \boldsymbol{Y}_i)$. The subset consists of only observations that are in or immediate outside of $[A_1, A_2]$ (see Figure 2.1). To this end, let $T_{ik}^+ = \min\{t_{ij} : Y_{ij} \geq A_k\}$ be the first time an observation is equal to or above A_k, and $T_{ik}^- = \max\{t_{ij} : t_{ij} < T_{ik}^+\}$, the time that immediately preceded T_{ik}^+. For example, T_{i2}^+ is the first observation at or above A_2 and T_{i1}^- is the last observation before the first crossing of A_1. The subsetted data consists of

$$\mathcal{D}_i = (\boldsymbol{t}_i, \boldsymbol{Y}_i) = \{(t_{ij}^*, Y_{ij}^*) : T_{i1}^- < t_{ij}^* \leq T_{i2}^+\},$$

Further, let $\mathcal{D}_{i0} = (t_{i0}, Y_{i0}) = \{(t_{ij}^*, Y_{ij}^*) : t_{ij}^* = T_{i1}^-\}$ denote the initial state, and $\mathcal{D}_{ij} = \{t_{ij}, (t_{ik}, Y_{ik})_{k=0,1,\ldots,j-1}\}$ denote the observed information up to time $-t_{ij}$ ($-t_{ij} = t_{ij} - \epsilon$, where ϵ is arbitrary small). The distribution of $\mathcal{D}_i$ is modeled conditionally on the initial state $\mathcal{D}_{i0}$. In Figure 2.1, we display an example of subsetting the longitudinal

data $(\boldsymbol{t}_i^*, \boldsymbol{Y}_i^*)$ if $A_1 = 3$ and $A_2 = 4$, which includes the subset of observations $(\boldsymbol{t}_i, \boldsymbol{Y}_i)$ denoted by •, the initial state (t_{i0}, Y_{i0}) denoted by ▲, and observations excluded from the subset (X).

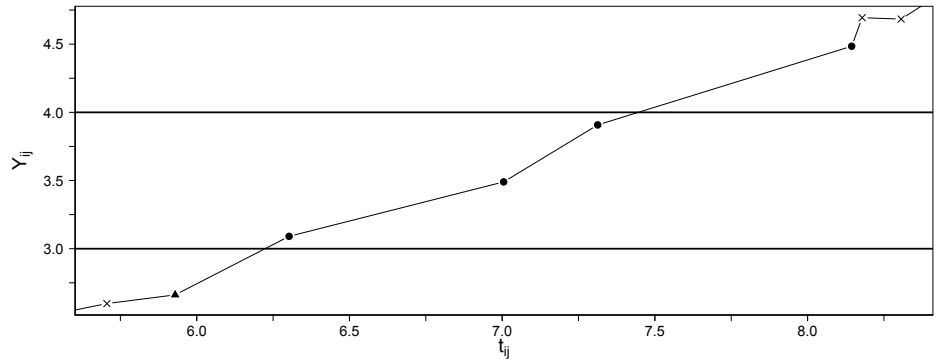


Figure 2.1 Illustration of subsetting the longitudinal data for thresholds, $A_1 = 3$ and $A_2 = 4$

To form the likelihood of the subset of the longitudinal data, we must condition on the each observation being included in the subsample. Note that all observations in $\mathcal{D}_i$ are included in the sample based on the previous observation except for (t_{i1}, Y_{i1}). That is, $\Pr\{(t_{ij}, Y_{ij}) \in \mathcal{D}_i | \mathcal{D}_{ij}\} = 1$ for $j = 2, 3, \ldots, m_i$. Conditional on $\mathcal{D}_{i0}$, (t_{i1}, Y_{i1}) is included in $\mathcal{D}_i$ if and only if $Y_{i1} > A_1$. This induces length bias on the cervical dilation measurements due to left truncation and must be corrected for in estimation. Length bias sampling occurs when the probability of inclusion of observations is proportional to their values (Cox, 1968). Thus our inclusion criteria seems to favor the cervical dilation measurements greater than the lower threshold A_1. This phenomenon is similar to what is observed in left truncated survival analysis. Similar to the survival analysis setting we correct for the length-bias by multiplying the likelihood of the (t_{i1}, Y_{i1}) observation by the inverse probability that the observation was included in the sample, $\Pr(Y_{i1} \geq A_1 | \mathcal{D}_{i1})^{-1}$ (see Wang (1989) or (Andersen et al., 1993) Chapter III.3). Notice that, as the baseline observation gets closer to the lower threshold the effect of the left truncation diminishes.

To form the likelihood we assume, without loss of generality, that $Y_{i0} < A_1$, $Y_{im_i} > A_2$ and $m_i \geq 2$. This will not always be the case in practice, but alterations are straightforward and more technical in nature. So as not to extend the parametric assumption beyond A_2, we censor the last observation at A_2. Given the observed data $\boldsymbol{\mathcal{D}} = \{(\boldsymbol{t}_i, \boldsymbol{Y}_i)_{i=1,2,\ldots,n}\}$ the likelihood is then

$$L(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}}) = \prod_{i=1}^{n} \frac{f(Y_{i1}|\mathcal{D}_{i1}, \boldsymbol{\theta})}{\bar{F}(A_1|\mathcal{D}_{i1}, \boldsymbol{\theta})} \left\{ \prod_{j=1}^{m_i-1} f(Y_{ij}|\mathcal{D}_{ij}, \boldsymbol{\theta}) \right\} \bar{F}(A_2 - Y_{im_i-1}|\mathcal{D}_{im_i}, \boldsymbol{\theta}), \quad (2.3.3)$$

where f and $\bar{F}$ are the density and survival functions, respectively, corresponding to (2.3.1). We maximize $\log\{L(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}})\}$ with respect to $\boldsymbol{\theta}$ such that $\boldsymbol{\theta} \in \Theta$. The distribution of S_{12} is then estimated using the plug-in estimator $F_{S_{12}}(t|\widehat{\boldsymbol{\theta}})$. We discuss inference procedures for $\widehat{\theta}$ and $F_{S_{12}}(t|\widehat{\boldsymbol{\theta}})$ in Section 2.3.

2.3.2 Wiener Process LTR model

Wiener process (Brownian motion) is a continuous time stochastic process with stationary and independent increments and one of useful stochastic process in applied statistics. The process has three individual level parameters: initial level (t_{i0}, Y_{i0}), the mean (drift) β and the variance σ component. Collectively these parameters define the profile of the health status of an individual. The initial level (t_{i0}, Y_{i0}) sets the starting point for the profile. The further it is from the threshold, the greater is the initial health of the individual with respect to the disease under study. The mean parameter β describes the rate per unit time at which the profile approaches the threshold while the variance σ component represents the inherent variability per unit time and gives the profile its stochastic behavior. The smaller the σ the more predictable the individual's health outcome (Lee et al., 2009).

We consider representing the underlying stochastic model as a Wiener process given for the ith subject by

$$Y_{ij} = Y_{ij-1} + \beta(t_{ij} - t_{ij-1}) + \sigma B(t_{ij} - t_{ij-1}), \quad (2.3.4)$$

which is modeled conditionally on an initial state (t_{i0}, Y_{i0}), where β is the drift, $\sigma(\sigma > 0)$ is a variance component as defined before and $B(t)$ is a standard Brownian variate corresponding to the error term where $B(t) \sim N(0, t)$. The Wiener process is a convenient choice for the underlying longitudinal process due to its mathematical applicability (Ting Lee et al., 2004). Even though cervical dilation tend to increase over time, it is possible for the observed cervical dilation measurements to exhibit non-monotone trends. For example, 7% of women in CSL data have at least one cervical dilation measurement smaller than at an earlier time point. In addition, the distribution of the FHT of a Wiener process is a well known Inverse Gaussian density that makes it possible to use analytical procedures in parameter estimation (Tweedie, 1957; Chhikara and Folks, 1989; Lee and Whitmore, 2006).

It might be of interest to examine the effect of some covariates on the longitudinal process. The covariates enter model (2.3.4) as an interaction with time, unless a model on the initial state is specified. At time t_{ij} model (2.3.4) with $p+1$ dimensional covariate vector $\boldsymbol{X}_i$ (including an intercept) would be defined as

$$Y_{ij} = Y_{ij-1} + \boldsymbol{\beta}\boldsymbol{X}_i^T(t_{ij} - t_{ij-1}) + \sigma B(t_{ij} - t_{ij-1}), \tag{2.3.5}$$

where ^{T} denotes the transpose of a vector. In matrix form (2.3.4) becomes

$$\boldsymbol{Y}_i \sim N(\boldsymbol{\beta}\boldsymbol{X}_i^T\boldsymbol{t}_i, \boldsymbol{\Sigma}_i)$$

where $\boldsymbol{X}_i^T\boldsymbol{t}_i$ is an $m_{i1} \times (p+1)$ matrix with $\boldsymbol{X}_i^T(t_{ij} - t_{ij-1})$ as row j, and $\boldsymbol{\Sigma}_i$ is a time dependent covariance structure (Hoel et al. (1972)) given as $\boldsymbol{\Sigma}_i = \sigma^2\boldsymbol{U}_i$ where

$$\boldsymbol{U}_i = \begin{pmatrix} s_{i1} & s_{i1} & \cdots & s_{i1} \\ s_{i1} & s_{i2} & \cdots & s_{i2} \\ \vdots & \vdots & \ddots & \vdots \\ s_{i1} & s_{i2} & \cdots & s_{imi} \end{pmatrix}$$

that is, $u_{ijk} = \min(s_{ij}, s_{ik})$ and $s_{ij} = t_{ij} - t_{i0}$

Given (2.3.4), $F_{S_{12}}(t)$ has an inverse Gaussian distribution with density

$$f_{S_{12}}(t) = \sqrt{\frac{\lambda}{2\pi t^3}}\ \exp\left[-\lambda\frac{(t-\mu)^2}{2\mu^2 t}\right] \tag{2.3.6}$$

where μ is the mean survival time and λ is a scale parameter. Thus, the distribution of the time it takes to progress from A_1 to A_2 can be estimated with an inverse Gaussian distribution characterized by $\mu = \frac{(A_2-A_1)}{\beta}$ and $\lambda = \frac{(A_2-A_1)^2}{\sigma^2}$. In Section 2.3.3, we discuss our inference procedures for μ, λ, and $\bar{F}_{S_{12}}$.

2.3.3 Parameter Variance Estimation

Under standard regularity conditions (e.g., see (Ferguson, 1996)) the estimated parameters from the likelihood in (2.3.3) have desirable asymptotic properties. That is, the p-dimensional vector $\widehat{\boldsymbol{\theta}}$ has an asymptotic normal distribution with variability that can be estimated via the inverse of the observed information matrix $-\widehat{J}(\widehat{\boldsymbol{\theta}})^{-1}$ with $\widehat{J}(\widehat{\boldsymbol{\theta}}) = \nabla^2_\theta l(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}})$ where ∇^2_θ is a $p \times p$ matrix of second partial derivatives with respect to $\boldsymbol{\theta}$ and $l(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}}) = \log\{L(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}})\}$

Under regularity conditions, the plug-in estimator $F_{S_{12}}(t|\widehat{\boldsymbol{\theta}})$ is asymptotically normal with variance that can be consistently estimated via

$$\widehat{\text{var}}\{F_{S_{12}}(t|\widehat{\boldsymbol{\theta}})\} = \{\nabla_\theta F_{S_{12}}(t|\widehat{\boldsymbol{\theta}})\}^T\{-\widehat{J}(\widehat{\boldsymbol{\theta}})\}^{-1}\{\nabla_\theta F_{S_{12}}(t|\widehat{\boldsymbol{\theta}})\},$$

where $\nabla_\theta F_{S_{12}}(t|\widehat{\boldsymbol{\theta}})$ is a p-dimensional vector of derivatives with respect to $\boldsymbol{\theta}$. Under the Wiener process specification the increments in the longitudinal process $\Delta Y_{ijk} = Y_{ijk} - Y_{ijk-1}$ are assumed to be independent of the previous measurements. This assumption is a reasonable approximation for a small $[A_1, A_2]$ interval, such as with the motivating example, but some residual dependence could exist. As a result, in practice a robust estimate of the variance, White (1982) is preferred. The robust estimation uses the modified sandwich estimator of the variance of $\widehat{\boldsymbol{\theta}}$, $R(\widehat{\boldsymbol{\theta}}) = \{-\widehat{J}(\widehat{\boldsymbol{\theta}})\}^{-1}\widehat{V}(\widehat{\boldsymbol{\theta}})\{-\widehat{J}(\widehat{\boldsymbol{\theta}})\}^{-1}$, where $\widehat{V}(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^n\{\nabla l_i(\boldsymbol{\theta}|\mathcal{D}_i)\}^T\nabla l_i(\boldsymbol{\theta}|\mathcal{D}_i)$ and $l(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}}) =$

$\sum_{i=1}^{n} l_i(\boldsymbol{\theta}|\mathcal{D}_i)$. To estimate the variance of $F_{S_{12}}$ we then use the robust delta-method form

$$\widehat{\text{var}}_R\{F_{S_{12}}(t|\widehat{\boldsymbol{\theta}})\} = \{\nabla_\theta F_{S_{12}}(t|\widehat{\boldsymbol{\theta}})\}^T R(\widehat{\boldsymbol{\theta}})\{\nabla_\theta F_{S_{12}}(t|\widehat{\boldsymbol{\theta}})\} \tag{2.3.7}$$

The robust variance estimation is employed in evaluating the efficiency of parameter estimates and time quantiles from the LTR model using simulation studies and for the application on the CSL data. We use a numerical approximation to all derivatives involved in computing the robust variance.

2.3.4 Standard Interval Censored model

For comparison purposes we analyzed the data using standard survival model for interval censored data. In accordance with the assumption that the underlying cervical dilation is a Wiener process, we assumed that the labor duration (i.e., survival time) has an inverse Gaussian distribution. Given A_1 and A_2, let T_i be the true unobserved time that cervical dilation would take to progress from A_1 to A_2. The true survival time is unknown and is bounded by S_{iA_1} and S_{iA_2} where $T^-_{iA_1} \leq S_{iA_1} \leq T^+_{iA_1}$ and $T^-_{iA_2} \leq S_{iA_2} \leq T^+_{iA_2}$. Therefore, the true elapsed time between the two thresholds can be estimated by $S_{iA_1A_2}$ defined as $T^-_{iA_2} - T^+_{iA_1} \leq S_{iA_1A_2} \leq T^+_{iA_2} - T^-_{iA_1}$ with a likelihood function

$$F_{S_{iA_1A_2}}(t) = \prod_{i=1}^{n} \left[F_{(S_{iA_2})} - F_{(S_{iA_1})}\right]$$

where $F_{(S_{iA_2})}$ and $F_{(S_{iA_1})}$ are inverse Gaussian distributions with parameters μ and λ.

2.4 Simulation Studies

To evaluate the performance of our model numerous simulation studies were performed. Our interest was to apply the proposed method on cervical dilation data and draw population level inference on cervical dilation progression over a given time period.

We thus simulated cervical dilation as a Wiener process where at time t^*_{ij}, the cervical dilation measurement was generated as

$$Y^*_{ij} = Y^*_{ij-1} + \beta_1(t^*_{ij} - t^*_{j-1}) + \sigma\sqrt{(t^*_{ij} - t^*_{ij-1})}\varepsilon_{ij}$$

where t^*_{ij} are uniform random variables, $t^*_{ij} \sim U(0, 25)$ is the jth time point and $\varepsilon_{ij} \sim N(0, 1)$. The initial states of the process across all subjects were defined as $Y^*_{i0} \sim N(1, 1)$. We generated 200 datasets each with sample size of 500 when (A_1, A_2) is $(2, 5)$, or $(4, 5)$ and (β_1, σ) is $(1, 1)$, or $(0.5, 1)$. Usually the cervical dilation measurements are rounded to the nearest 0.5 cm. We thus evaluated the performance of our model on the simulated continuous data, and data that were rounded to the nearest 0.5 cm.

We subset the longitudinal observations using the methods in Section 2.3.1. The average m_i for $(A_1, A_2) = (2, 5)$ was 5.8 and 4.0 for $\beta_1 = 0.5$ and 1, respectively, and for $(A_1, A_2) = (4, 5)$ it was 3.3 and 2.8 for $\beta_1 = 0.5$ and 1, respectively. All models were fitted using `R`, (R Core Team, 2015).

We compare the results of our model to estimates obtained from inverse Gaussian model for doubly interval censored data. In Table 2.1, we present parameter estimates for the proposed LTR model and inverse Gaussian IC model from a summary of 200 simulated datasets with $n = 500$ for the continuous outcome. The table displays the estimated parameters, their corresponding empirical standard errors , empirical coverage probabilities, and mean squared error given (2,5) and (4,5) thresholds. We present results for the rounded outcome to the nearest 0.5 in Table 2.2. The standard errors for parameter estimates are model based and were computed from the inverse of Hessian matrix obtained by maximizing the data log-likelihood functions under the proposed LTR model and the IC model, using Nelder-Mead optimization method in *optim* function in `R`, (R Core Team, 2015).

Overall, results in Table 2.1 show that both models estimate the true mean survival time and the scale parameter closely. However, the point estimates from the LTR

Table 2.1 Parameter estimates from the LTR and IC model for the continuous outcome

		Continuous Outcome					
		LTR model			IC model		
	TRUE	EST(SE)	ECP	MSE	EST(SE)	ECP	MSE
$(A_1, A_2) = (2, 5)$							
μ	3	3.01 (0.08)	0.95	0.01	2.91 (0.11)	0.88	0.02
λ	9	9.05 (0.53)	0.93	0.28	9.83 (1.59)	0.98	3.22
μ	6	6.14 (0.22)	0.93	0.07	6.16 (0.28)	0.94	0.10
λ	9	8.96 (0.38)	0.95	0.15	8.79 (0.96)	0.97	1.05
$(A_1, A_2) = (4, 5)$							
μ	1	1.00 (0.03)	0.95	0.001	0.95 (0.09)	0.91	0.01
λ	1	1.00 (0.08)	0.92	0.01	1.29 (0.47)	0.97	0.31
μ	2	2.04 (0.12)	0.96	0.01	1.91 (0.16)	0.89	0.03
λ	1	1.00 (0.06)	0.92	0.004	0.99 (0.21)	0.95	0.04

Table 2.2 Parameter estimates from the LTR model for data rounded to the nearest 0.5

	EST(SE)	ECP	MSE
$(A_1, A_2) = (2, 5)$			
μ	3.01 (0.07)	0.96	0.01
λ	9.05 (0.46)	0.97	0.22
μ	6.16 (0.23)	0.94	0.08
λ	8.89 (0.32)	0.97	0.11
$(A_1, A_2) = (4, 5)$			
μ	0.999 (0.03)	0.94	0.001
λ	1.01 (0.06)	0.97	0.004
μ	2.06 (0.12)	0.94	0.02
λ	0.997 (0.06)	0.93	0.004

model have smallest bias and mean squared error (MSE) in all occasions. Both models capture the true mean survival time well and have empirical coverage probabilities close to the nominal value, except the IC model for interval (2,5) with $\beta = 1$ and (4,5) with $\beta = 0.5$ which show considerable type I error. Similar to the mean estimate, the lambda estimate from LTR model has the smallest bias and MSE than that from

the IC model. The empirical coverage probabilities from both models cover the true lambda value well. Table 2.2 presents results from the LTR model where the data have been rounded to units of 0.5. For this model there is some additional variation induced by the rounding mechanism. Thus, we expect some positive bias in σ^2 and negative bias in λ. Accounting for this, the results are similar to the results from LTR model with the continuous outcome for all settings.

2.4.1 Efficiency of LTR model

We checked the impact that number of measurements per individual have on parameter estimates, 50%, 75% and 95% time quantiles in terms of their standard errors, empirical coverage probabilities and MSE. We mimicked sparse and frequently measured longitudinal data by setting n_i to 7 (sparse) or 36 (frequent) and subsetted the longitudinal observations using the methods in Section 2.3.1. For the sparse setting this resulted in an average m_i of 2.3 and 2.6 for $\beta_1 = 1$ and 0.5, respectively, and for the frequent setting an average m_i of 3.5 and 4.9 for $\beta_1 = 1$ and 0.5, respectively. The robust variances as discussed in section 2.3.3 were used in these simulation studies.

Table 2.3 displays the average estimated parameters, corresponding standard errors, MSE, and the empirical coverage probabilities using the robust delta method presented in Section 2.3.3 for the μ and λ parameters, along with estimates of the quantile function denoted by $F^{-1}_{S12}(q)$ for the qth quantile of S_{12} from a summary of 1000 simulated datasets with $n = 500$. For the sparsely measured data the interval censored (IC) model produces erratic estimates of the λ parameter. Further, the coverage probabilities for all parameters are well below the nominal level. The LTR model produces estimates that are relatively unbiased, with coverage probabilities close to the nominal level. Moreover, the MSE of the LTR model is markedly lower than the MSE for the IC model. For the frequently measured setting, both models estimate the true mean survival time and the scale parameter closely. For the $\mu = 1$

Table 2.3 Parameter estimates given $ni = 7$ (sparse) and $ni = 36$ (frequent) measurements for continuous outcome

		LTR model			IC model		
	TRUE	EST (SE)	ECP	MSE	EST (SE)	ECP	MSE
				Sparse measurements			
μ	1	1.00 (0.04)	0.944	0.001	1.04 (0.68)	0.293	0.462
λ	1	1.01 (0.07)	0.955	0.006	10.87 (15.64)	0.200	342.103
$F_{S12}^{-1}(0.5)$	0.68	0.68 (0.02)	0.960	<0.001	0.81 (0.79)	0.266	0.647
$F_{S12}^{-1}(0.75)$	1.24	1.24 (0.04)	0.947	0.002	1.15 (0.82)	0.314	0.678
$F_{S12}^{-1}(0.95)$	2.92	2.92 (0.18)	0.939	0.032	2.39 (0.52)	0.793	0.551
μ	2	2.02 (0.12)	0.955	0.015	1.75 (0.18)	0.702	0.095
λ	1	1.00 (0.06)	0.945	0.004	0.64 (0.22)	0.526	0.178
$F_{S12}^{-1}(0.5)$	1.03	1.03 (0.04)	0.961	0.001	0.75 (0.16)	0.582	0.103
$F_{S12}^{-1}(0.75)$	2.28	2.30 (0.11)	0.952	0.011	1.83 (0.26)	0.616	0.28
$F_{S12}^{-1}(0.95)$	7.11	7.17 (0.59)	0.955	0.358	6.74 (0.69)	0.877	0.616
				Frequent measurements			
μ	1	1.00 (0.05)	0.947	0.002	0.99 (0.05)	0.928	0.003
λ	1	1.00 (0.05)	0.946	0.002	0.96 (0.13)	0.917	0.020
$F_{S12}^{-1}(0.5)$	0.68	0.68 (0.02)	0.952	<0.001	0.66 (0.04)	0.920	0.002
$F_{S12}^{-1}(0.75)$	1.24	1.24 (0.05)	0.947	0.003	1.23 (0.06)	0.927	0.004
$F_{S12}^{-1}(0.95)$	2.92	2.92 (0.18)	0.939	0.032	2.93 (0.20)	0.935	0.040
μ	2	2.09 (0.14)	0.932	0.029	1.99 (0.13)	0.937	0.016
λ	1	1.00 (0.04)	0.945	0.001	0.97 (0.10)	0.924	0.012
$F_{S12}^{-1}(0.5)$	1.03	1.05 (0.04)	0.921	0.002	1.01 (0.06)	0.930	0.005
$F_{S12}^{-1}(0.75)$	2.28	2.37 (0.13)	0.922	0.023	2.26 (0.13)	0.928	0.017
$F_{S12}^{-1}(0.95)$	7.11	7.27 (0.64)	0.934	0.432	7.14 (0.59)	0.950	0.347

setting the LTR model has preferable bias, MSE, and empirical coverage probabilities. For the $\mu = 2$ setting the IC model appears to have more precise estimates of the μ parameter, while the LTR model produces more accurate estimates of the λ parameter. The overall results of our simulation agree with the findings of Lu et al. (1996), who found that models based on repeated measurements of a longitudinal process give more precise estimates of quantiles of event time distribution than do traditional time to event models.

Table 2.4 Parameter estimates given $ni = 7$ (sparse) and $ni = 36$ (frequent) measurements for data rounded to the nearest 0.5

	TRUE	EST(SE)	ECP	TRUE	EST(SE)	ECP
μ	1	1.00 (0.06)	0.943	2	2.03 (0.15)	0.949
λ	1	0.98 (0.06)	0.935	1	0.98 (0.05)	0.936
$F^{-1}_{S12}(0.5)$	0.68	0.70 (0.03)	0.951	1.03	1.04 (0.04)	0.965
$F^{-1}_{S12}(0.75)$	1.24	1.26 (0.06)	0.948	2.28	2.38 (0.11)	0.942
$F^{-1}_{S12}(0.95)$	2.92	2.92 (0.19)	0.934	7.11	7.17 (0.59)	0.934

We also evaluated the effect that number of observations per an individual have on parameter estimates when the cervical dilation measurements are rounded to the nearest 0.5 cm. The analysis were based on a summary of 1000 simulated datasets with $n = 500$. Table 2.4 presents results from the LTR model for this setting with sparsely measured data. The results for the frequently measured data were similar to the pattern observed in Table 2.3. As indicated in Table 2.1, rounding up the cervical dilation measurements has induced additional variation. Nevertheless, the results are similar to results observed in Table 2.3.

To quantify the efficiency of LTR model parameter estimates, we computed the ratio of the MSE of LTR and IC model parameters given sparsely observed and frequently observed data, $Eff = 100(\frac{MSE_{LTR}}{MSE_{IC}})$. The analyses were based on a summary of 1000 simulated datasets with $n = 500$. Overall, the parameter estimates from the proposed model LTR are more efficient than estimates from the IC model.

2.4.2 Model Misspecification

As stated in section 2.3.3, Wiener process assumes independent increments of the longitudinal process. While this is reasonable for our motivating example, we expect in general circumstances measurements from the same individual to be correlated. In this section, we present additional simulation studies ran to investigate any biases that might arise when residual dependence is present. Specifically, we tested the ability

Table 2.5 LTR model efficiency given $ni = 7$ (sparse) and $ni = 36$ (frequent) measurements

		LTR model		IC model		
	TRUE	EST (SE)	MSE	EST (SE)	MSE	EFF
			Sparse measurements			
μ	1	1.00 (0.04)	0.001	1.04 (0.68)	0.462	0.002
λ	1	1.01 (0.07)	0.006	10.87 (15.64)	342.103	0.000
$F^{-1}_{S12}(0.5)$	0.68	0.68 (0.02)	<0.001	0.81 (0.79)	0.647	0.002
$F^{-1}_{S12}(0.75)$	1.24	1.24 (0.04)	0.002	1.15 (0.82)	0.678	0.003
$F^{-1}_{S12}(0.95)$	2.92	2.92 (0.18)	0.032	2.39 (0.52)	0.551	0.058
μ	2	2.02 (0.12)	0.015	1.75 (0.18)	0.095	0.158
λ	1	1.00 (0.06)	0.004	0.64 (0.22)	0.178	0.022
$F^{-1}_{S12}(0.5)$	1.03	1.03 (0.04)	0.001	0.75 (0.16)	0.103	0.010
$F^{-1}_{S12}(0.75)$	2.28	2.30 (0.11)	0.011	1.83 (0.26)	0.28	0.039
$F^{-1}_{S12}(0.95)$	7.11	7.17 (0.59)	0.358	6.74 (0.69)	0.616	0.581
			Frequent measurements			
μ	1	1.00 (0.05)	0.002	0.99 (0.05)	0.003	0.667
λ	1	1.00 (0.05)	0.002	0.96 (0.13)	0.020	0.100
$F^{-1}_{S12}(0.5)$	0.68	0.68 (0.02)	<0.001	0.66 (0.04)	0.002	0.500
$F^{-1}_{S12}(0.75)$	1.24	1.24 (0.05)	0.003	1.23 (0.06)	0.004	0.750
$F^{-1}_{S12}(0.95)$	2.92	2.92 (0.18)	0.032	2.93 (0.20)	0.040	0.800
μ	2	2.09 (0.14)	0.029	1.99 (0.13)	0.016	1.812
λ	1	1.00 (0.04)	0.001	0.97 (0.10)	0.012	0.083
$F^{-1}_{S12}(0.5)$	1.03	1.05 (0.04)	0.002	1.01 (0.06)	0.005	0.400
$F^{-1}_{S12}(0.75)$	2.28	2.37 (0.13)	0.023	2.26 (0.13)	0.017	1.353
$F^{-1}_{S12}(0.95)$	7.11	7.27 (0.64)	0.432	7.14 (0.59)	0.347	1.245

of the proposed method to estimate the marginal survival function from a Weiner process model with a random slope term. To test the properties of our method under the model misspecification, simulation studies were performed. We simulated the longitudinal outcome as a Wiener process with both fixed and random slope. At time t^*_{ij}, the cervical dilation measurement was generated as

$$Y^*_{ij} = Y^*_{ij-1} + (\beta_1 + b_i)(t^*_{ij} - t^*_{j-1}) + \sigma\sqrt{(t^*_{ij} - t^*_{ij-1})}\varepsilon_{ij}$$

where $\varepsilon_{ij} \sim N(0,1)$ and $b_i \sim N(0, \sigma^2_b)$ is the random slope for ith woman. We did simulations assuming the variance of the random slopes were $\sigma^2_b = (0.15^2, 0.3^2)$. The

initial states of the process across all subjects were defined as $Y_{i0}^* \sim N(1,1)$. We generated 1000 datasets each with sample size of 500 when (A_1, A_2) is $(4,5)$ and (β_1, σ) is $(1,1)$, or $(0.5,1)$. For the $\beta_1 = 0.5$ specification only the $\sigma_b^2 = 0.15^2$ was used to avoid negative $(\beta_1 + b_i)$. Note that when negative $(\beta_1 + b_i)$ are present the survival function of interest is improper. We subset the longitudinal observations to $(\boldsymbol{t}_i, \boldsymbol{Y}_i) = \{(t_{ij}, Y_{ij})_{j=1,2,\ldots,m_i}\}$ which are relevant in estimating S_{12} (see Section 2.3.1. The average m_i for $(\beta_1 = 1, \sigma_b^2 = 0.3^2)$, $(\beta_1 = 1, \sigma_b^2 = 0.15^2)$, and $(\beta_1 = 0.5, \sigma_b^2 = 0.15^2)$ were 2.6, 2.3 and 2.5, respectively.

We compare the average estimated survival function to the marginal survival function of the random slopes model

$$\bar{F}_{S_{12}}^M(t|\mu, \lambda, \sigma_b^2) = \int \bar{F}_{S_{12}}(t|\mu + b, \lambda) dF_b(b|\sigma_b^2) \tag{2.4.1}$$

where $\bar{F}_{S_{12}}(\cdot|\mu, \lambda)$ denotes the survival function of an inverse Gaussian with parameters, mean μ and scale parameter λ, and $F_b(\cdot|\sigma_b^2)$ denotes a $N(0, \sigma_b^2)$ distribution function. Note that $\bar{F}_{S_{12}}^M$ is not an inverse Gaussian distribution. As a result, the proposed model is misspecified in this setting. For example, the random slopes model will appear to have more residual variability when fitted with the proposed method.

In Figure 2.2 to 2.4 we compare the true and estimated curves for $(\beta_1 = 1, \sigma_b^2 = 0.3^2)$, $(\beta_1 = 1, \sigma_b^2 = 0.15^2)$, and $(\beta_1 = 0.5, \sigma_b^2 = 0.15^2)$ settings. For the proposed model we present the average estimated survival curve

$$\hat{\bar{F}}_{S_{12}}(t|\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\lambda}}) = \frac{1}{1000} \sum_{k=1}^{1000} \bar{F}_{S_{12}}(t|\hat{\mu}_k, \hat{\lambda}_k)$$

where $\{\hat{\mu}_k, \hat{\lambda}_k\}$ are the estimated parameters at the kth iteration, along with the 2.5th and the 97.5th percentiles of the set $\{\bar{F}_{S_{12}}(t|\hat{\mu}_k, \hat{\lambda}_k)|_{k=1,2,\ldots,1000}\}$. Figures 2.2 to 2.4 also contain the true marginal survival curve $\bar{F}_{S_{12}}^M$ given in (2.4.1). For the $(\beta_1 = 1, \sigma_b^2 = 0.3^2)$ setting there is a some negative biases present early in the curve, and positive bias later in the curve. The average estimated curve for the $(\beta_1 = 1, \sigma_b^2 = 0.15^2)$ and $(\beta_1 = 0.5, \sigma_b^2 = 0.15^2)$ settings are virtually unbiased, with the true marginal

survival curves falling within the 95% point-wise confidence intervals throughout. As a result, it appears that the proposed method accurately estimate the marginal survival function $\bar{F}^M_{S_{12}}$, especially when there is small-to-moderate variability in the random slopes.

Figures 2.2 to 2.4 display the average estimated survival curve (solid black line) with 95% point-wise empirical confidence intervals (black dotted line) versus the true marginal survival curve $\bar{F}^M_{S_{12}}$ (dashed gray line).

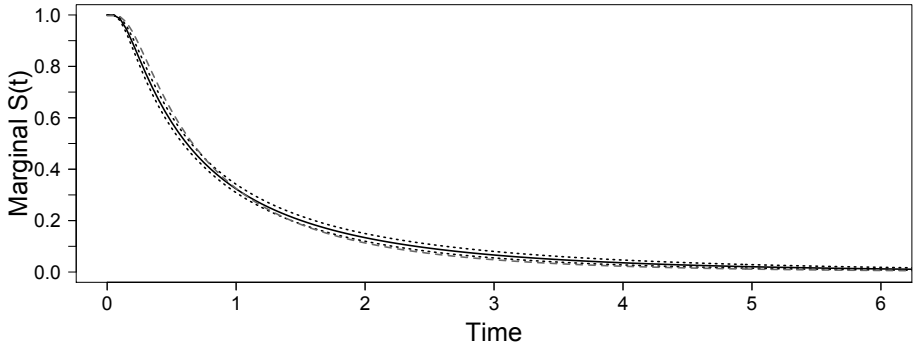


Figure 2.2 Estimated and the true S(t) when $\beta_1 = 1, \sigma_b^2 = 0.3$

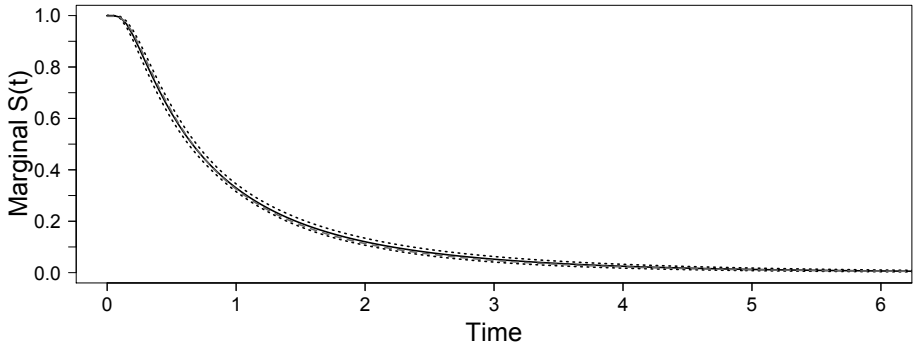


Figure 2.3 Estimated and the true S(t) when $\beta_1 = 1, \sigma_b^2 = 0.15$

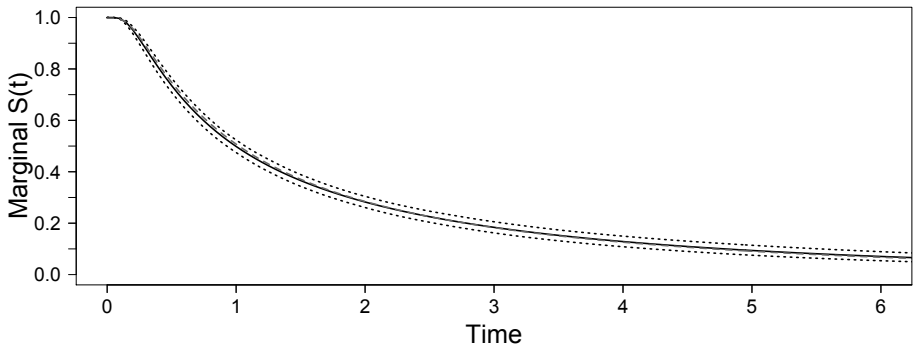


Figure 2.4 Estimated and the true S(t) when $\beta_1 = 0.5, \sigma_b^2 = 0.15$

2.5 Application on Consortium on Safe Labor data

We applied the proposed model and the analogous doubly interval censored model on the CSL data. We randomly selected a sample of 2,000 out of 11,521 nulliparous women from Utah site of the CSL study. We carried out two analyses. First, we compared the estimates from unadjusted analyses using the LTR and IC models. In the second analysis, we fitted the LTR model adjusted for maternal BMI levels (adjusted model). Maternal BMI has been established to be a key factor in labor progression and management, Vahratian et al. (2004); Kominiarek et al. (2011), however this has never been investigated for specific portions of the labor curve. The maternal BMI levels were categorized as normal (18.5 to 24.9 kg/m^2), overweight (25 to 29.9 kg/m^2) and obese (above 30 kg/m^2). We performed the analyses for two sets of $[A_1, A_2]$; $(3, 4)$ and $(7, 8)$ which respectively corresponds to early and late stages of labor. The average m_i was 2.6 and 2.3 for the $(3, 4)$ and $(7, 8)$ thresholds, respectively.

Table 2.6 shows parameter estimates from unadjusted and adjusted LTR model and an unadjusted IC model with n=2000 nulliparous women. The displayed are the point estimates, robust standard errors and corresponding 95% confidence intervals. For the adjusted model the mean survival times for the three BMI groups were estimated as: $\widehat{\mu}_1 = \frac{(A_2 - A_1)}{(\widehat{\beta} + \widehat{\beta}_{1_{BMI}})}$, $\widehat{\mu}_2 = \frac{(A_2 - A_1)}{(\widehat{\beta} + \widehat{\beta}_{2_{BMI}})}$ and $\widehat{\mu}_3 = \frac{(A_2 - A_1)}{\widehat{\beta}}$ for women with

.

Table 2.6 Parameter estimates from unadjusted and adjusted LTR model and an unadjusted IC model with n=2000 nulliparous women

	EST (SE)	95% CI	EST (SE)	95% CI
Unadjusted LTR model versus IC model				
	LTR model		IC model	
$(A_1, A_2) = (3, 4)$				
μ	2.50 (0.06)	(2.38, 2.62)	2.58 (0.12)	(2.35, 2.81)
$F_{S12}^{-1}(0.5)$	1.21 (0.11)	(0.99, 1.43)	1.79 (0.11)	(1.57, 2.01)
$F_{S12}^{-1}(0.95)$	9.18 (0.51)	(8.17, 10.19)	7.34 (0.51)	(6.33, 8.34)
λ	1.10 (0.20)	(0.71, 1.49)	2.82 (0.47)	(1.91, 3.73)
$(A_1, A_2) = (7, 8)$				
μ	1.29 (0.05)	(1.20, 1.38)	0.71 (0.05)	(0.61, 0.82)
$F_{S12}^{-1}(0.5)$	0.40 (0.01)	(0.38, 0.43)	0.37 (0.04)	(0.29, 0.46)
$F_{S12}^{-1}(0.95)$	5.45 (0.26)	(4.95, 5.96)	2.52 (0.38)	(1.76, 3.27)
λ	0.27 (0.02)	(0.24, 0.30)	0.37 (0.11)	(0.16, 0.58)
Adjusted LTR model				
	$(A_1, A_2) = (3, 4)$		$(A_1, A_2) = (7, 8)$	
β	0.32 (0.04)	(0.24, 0.39)	0.57 (0.03)	(0.50, 0.63)
$\beta_{1_{BMI}}$	0.11 (0.03)	(0.04, 0.18)	0.29 (0.01)	(0.27, 0.31)
$\beta_{2_{BMI}}$	0.07 (0.001)	(0.069, 0.07)	0.12 (0.06)	(-0.00, 0.24)
λ	1.07 (0.09)	(0.90, 1.23)	0.26 (0.02)	(0.23, 0.29)
Women with normal BMI				
μ_1	2.33 (0.08)	(2.17, 2.50)	1.17 (0.05)	(1.06, 1.27)
$F_{S12}^{-1}(0.5)$	1.15 (0.002)	(1.14, 1.15)	0.38 (0.0001)	(0.38, 0.38)
$F_{S12}^{-1}(0.75)$	2.62 (0.11)	(2.41, 2.82)	1.05 (0.19)	(0.68, 1.42)
$F_{S12}^{-1}(0.95)$	8.48 (1.62)	(5.31, 11.66)	4.92 (0.18)	(4.57, 5.26)
Overweight women				
μ_2	2.58 (0.24)	(2.11, 3.06)	1.46 (0.14)	(1.18, 1.75)
$F_{S12}^{-1}(0.5)$	1.21 (0.30)	(0.62, 1.79)	0.40 (0.001)	(0.40, 0.41)
$F_{S12}^{-1}(0.75)$	2.83 (0.40)	(2.04, 3.63)	1.20 (0.03)	(1.14, 1.26)
$F_{S12}^{-1}(0.95)$	9.63 (1.55)	(6.61, 12.66)	6.35 (2.61)	(1.23, 11.46)
Obese women				
μ_3	3.14 (0.37)	(2.41, 3.87)	1.77 (0.10)	(1.58, 1.96)
$F_{S12}^{-1}(0.5)$	1.32 (0.04)	(1.24, 1.40)	0.43 (0.03)	(0.37, 0.48)
$F_{S12}^{-1}(0.75)$	3.27 (0.07)	(3.13, 3.42)	1.32 (0.03)	(1.27, 1.38)
$F_{S12}^{-1}(0.95)$	12.24 (2.03)	(8.26, 16.23)	7.79 (0.50)	(6.81, 8.78)

normal BMI, overweight and obese (reference) women, respectively. The regression coefficients, $\widehat{\beta}_{1_{BMI}}$ and $\widehat{\beta}_{2_{BMI}}$ estimate the difference in the rate of change of cervical dilation per hour (cm/hr) between the obese women and the normal and overweight

women, respectively. For the adjusted model we tested allowing the variance parameter λ to vary by BMI categories, but a likelihood ratio test preferred a model with one λ parameter. Equations for estimated μ and λ for the unadjusted LTR model and the IC model are as explained in Section 2.3.2. The parameter and percentile estimates, denoted by $F^{-1}_{S12}(q)$ for the qth percentile of S_{12}, can be viewed in Table 2.6. Wald-type confidence intervals were constructed using the robust delta method presented in Section 2.3.3.

In Figure 2.5 and Figure 2.6 we display the estimated unadjusted survival curves for both threshold definitions for the LTR and IC models. From the unadjusted analysis, both LTR model and the IC model estimate the mean and the median survival time closely. For example, to progress from 3 to 4 cm we estimate the mean survival time as 2.50 hours and median survival time as 1.21 hours from the LTR model. Similarly, using the IC model we estimate mean survival time of 2.58 hours and median survival time of 1.79 hours. However, it will take 9.18 hours and 7.34 hours for 95% of women's cervical dilation to progress from 3 to 4 cm based on the LTR model and the IC model, respectively. To progress from 7 to 8 cm, based on the LTR model we estimate the mean and median survival time as 1.29 hours and 0.40 hours, and $95th$ time percentile of 5.45 hours. On the other hand from IC model we estimate mean and median survival time from 7 to 8 cm as 0.71 hours and 0.37 hours with 95% of women crossing the two thresholds within 2.52 hours. The estimates from LTR model have smaller standard errors in most occasions.

The results from adjusted LTR model show that, the estimated rate of change in cervical dilation per hour is significantly faster for women with normal BMI compared to obese women for early ($\beta_{1_{BMI}} = 0.11, 95\%$ CI : $[0.04, 0.18]$) and late ($\beta_{1_{BMI}} = 0.29, 95\%$ CI : $[0.27, 0.31]$) stages of labor. However the rate for the overweight women is significant only for the early stage ($\beta_{2_{BMI}} = 0.07, 95\%$ CI : $[0.069, 0.07]$). Overall, the rate of change appears to be slower at the early stage of labor and

gradually accelerates in the later stage. For example, for the obese women the rate is 0.32 cm/hr for the early stage and 0.57 cm/hr at the later stage. Overall, the mean and median values show small differences between the BMI groups. For the means these differences are statistically significant, but are unlikely to be significant in practice. The biggest difference between BMI groups is in the latter percentiles of the distributions. For example, in the later stage of labor the $95th$ percentile is estimated to be 4.92 and 7.79 hours for normal and obese women, respectively, a difference of 2.87 hours.

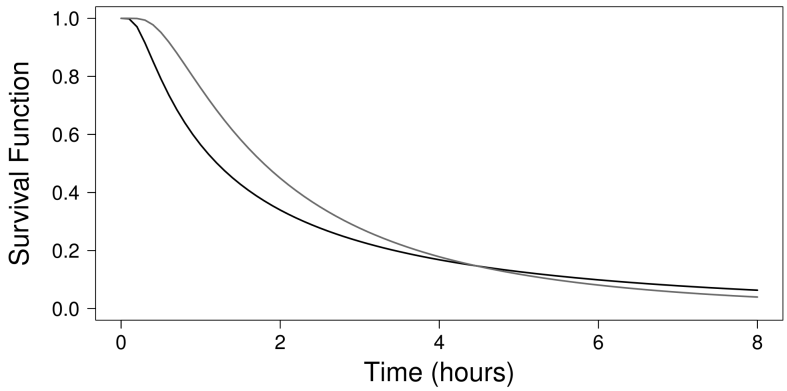


Figure 2.5 The estimated unadjusted survival curves given $(A_1, A_2) = (3, 4)$ cm

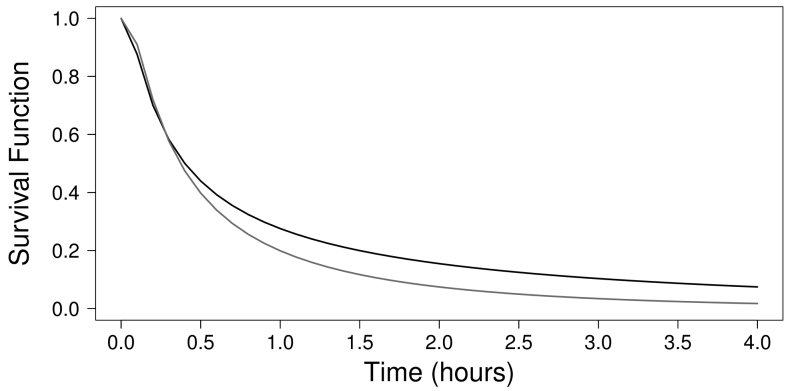


Figure 2.6 The estimated unadjusted survival curves given $(A_1, A_2) = (7, 8)$ cm

2.6 Conclusion

In this chapter, we have demonstrated that the distribution of the elapsed time between two pre-specified thresholds can be estimated via a LTR model. By relying on the repeated measurements we incorporate the impact of the intermittent changes of the underlying process. Our proposed method enables the utilization of all the available data within a bounded interval, and appears to have more favorable properties than the competing interval censored approach especially when the longitudinal measurements are sparse.

Our analysis of the CSL study data lead to some interesting conclusions regarding the results from the LTR and IC models. While these models estimated similar means and medians, marked differences in the later percentiles were evident. For example, for women going from 7 to 8 cm dilation the LTR model estimated the $95th$ percentile to be 5.45 hours (95% , CI $[4.95, 5.96]$) while the IC model estimated the same quantity to be 2.52 hours (95% , CI $[1.76, 3.27]$). This is a stark difference when one considers that an obstetrician might use the $95th$ percentile as a cutoff to advocate for stopping natural labor in favor of other options (e.g., Cesarean section). When using the results from the LTR model the cutoff is almost 3 hours later. Consequently, if the IC results were used it could result in a higher rate of unnecessary Cesarean sections. The longitudinal measurements were sparse (average m_i was 2.6 and 2.3 for $(3, 4)$ and $(7, 8)$, respectively), thus the simulation studies suggest that the IC results should be used with caution. The adjusted LTR model found larger differences in cervical dilation progression among the BMI groups at earlier stages of labor versus the later stages. Women with normal BMI levels have significantly higher rate of change in cervical dilation in both early and late stages of labor compared to obese women. Further, at early stages of labor the cervical dilation of overweight women progressed at a significantly faster rate than for obese women.

The proposed LTR model can be generalized to accommodate multiple thresholds and allow for non-linearity in the longitudinal process. As a result, the model is a viable method for representing the entire longitudinal process. This is an interesting endeavor and my collaborator and dissertation advisor, Dr. Alexander McLain has outlined the theoretical aspects of generalizing the current LTR model to multiple thresholds. Below we present his proposed extensions.

2.7.1 LTR for Multiple Thresholds

We now extend the LTR model in Section 2.3.1 to multiple threshold values. For simplicity we demonstrate for two consecutive threshold pairs then extend to multiple pairs below. Let A_1, A_2 and A_3 denote the thresholds of interest with the corresponding survival outcomes of interest $S_{12} = S_2 - S_1$ and $S_{23} = S_3 - S_2$. For example, we may be interested in the distribution of the time it takes a woman's cervical dilation to progress from 3cm to 4cm, and 4cm to 5cm. As a result, we extend the methods from Section 2.3.1 to this situation. To this end, we expand the notation from Section 2.3.1 and include an extra subscript to differentiate data for $[A_1, A_2)$ and $[A_2, A_3)$. That is, let $\mathcal{D}_{ik} = (\boldsymbol{t}_{ik}, \boldsymbol{Y}_{ik}) = \{(t_{ijk}, Y_{ijk})_{j=1,2,\ldots,m_{ik}}\}$, $\mathcal{D}_{i0k} = (t_{i0k}, Y_{i0k})$, and $\mathcal{D}_{ijk}$ denote the subset data, initial state, and history over $[A_k, A_{k+1})$, respectively, for subject i. Recall that $\mathcal{D}_{ijk}$ is the longitudinal history over subset $[A_k, A_{k+1})$ up to time $-t_{ijk}$ (where t_{ijk} is predictable).

Without loss of generality we assume that the distributions of the longitudinal process over $[A_1, A_2)$ and $[A_2, A_3)$ only differ by their parameter values denoted by $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively. Having multiple thresholds allows for non-linearity in the longitudinal process. Note that $\mathcal{D}_{i1}$ and $\mathcal{D}_{i2}$ will overlap. Specifically, last observation in $\mathcal{D}_{i1}$ and the first observation in $\mathcal{D}_{i2}$ are the same, i.e., $(t_{im_{i1}1}, Y_{im_{i1}1}) = (t_{i12}, Y_{i12})$.

Further, the distribution of Y changes at A_2 assuming $\theta_1 \neq \theta_2$ and the likelihood will need to account for this change. Using the law of total probability, we condition on the latent observation $S_{i2} = \min(t : Y_i(t) = A_2)$, and integrate out the conditional density of Y_{i12} given S_{i2} over all possible S_{i2}. Conditional on S_{i2}, the density of Y_{i12} is akin to adding (S_{i2}, A_2) to the history $\mathcal{D}_{i12}$.

Putting these factors together, the density for Y_{i12} is given by

$$f(Y_{i12}|\mathcal{D}_{i12}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \int_{t_{i0}}^{t_{i1}} f\{Y_{i12}|(\mathcal{D}_{i12}, s, A_2), \boldsymbol{\theta}_2\} f_{S_2}(s|\mathcal{D}_{i02}, \boldsymbol{\theta}_1) ds. \tag{2.7.1}$$

Here, Y_{i12} is broken into two components, the first occurs over the interval $(Y_{i02}, A_2]$ and is represented by f_{S_2} in the integrand, which is a function of $\boldsymbol{\theta}_1$ only. The second component $(A_2, Y_{i12}]$ is modeled via the density of Y_{i12} given the latent (S_2, A_2) observation. This technique could be used with only two thresholds (A_1, A_2) to allow for a different distribution for $Y < A_1$, where we would then include (2.7.1) in the likelihood given in (2.3.3).

Given the observed data $\boldsymbol{\mathcal{D}} = \{(\boldsymbol{t}_{ik}, \boldsymbol{Y}_{ik})_{k=1,2;i=1,2,\ldots,n}\}$ the likelihood is then

$$\begin{aligned} L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\boldsymbol{\mathcal{D}}) &= \prod_{i=1}^{n} \frac{f(Y_{i11}|\mathcal{D}_{i11}, \boldsymbol{\theta}_1)}{\bar{F}(A_1|\mathcal{D}_{i11}, \boldsymbol{\theta})} \left\{ \prod_{j=2}^{m_{i1}-1} f(Y_{ij1}|\mathcal{D}_{ij1}, \boldsymbol{\theta}_1) \right\} \frac{f(Y_{i12}|\mathcal{D}_{i12}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\bar{F}(A_2|\mathcal{D}_{i12}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)} \\ &\quad \left\{ \prod_{j=2}^{m_{i2}-1} f(Y_{ij2}|\mathcal{D}_{ij2}, \boldsymbol{\theta}) \right\} \bar{F}\{A_2|\mathcal{D}_{im_{i2}}, \boldsymbol{\theta}_2\}, \end{aligned} \tag{2.7.2}$$

where

$$\begin{aligned} &\bar{F}(A_2|\mathcal{D}_{i12}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \Pr(Y_{i12} \geq A_2|\mathcal{D}_{i12}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\ &= \int_{t_{i0}}^{t_{i1}} \bar{F}\{0|(\mathcal{D}_{i12}, s, A_2), \boldsymbol{\theta}_2\} f_{S_2}(s|\mathcal{D}_{i02}, \boldsymbol{\theta}_1) ds. \end{aligned} \tag{2.7.3}$$

In general, using similar notation to above, the likelihood in (2.7.2) can be expanded to a disjoint partition $(A_k, A_{k+1}]$ for $k = 1, 2, \ldots, K$ via

$$L(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}}) = \prod_{i=1}^{n}\prod_{i=1}^{n}\frac{f(Y_{i11}|\mathcal{D}_{i11},\boldsymbol{\theta}_1)}{\bar{F}(A_1|\mathcal{D}_{i11},\boldsymbol{\theta})}$$
$$\prod_{k=1}^{K-1}\left[\left\{\prod_{j=2}^{m_{ik}-1} f(Y_{ijk}|\mathcal{D}_{ijk},\boldsymbol{\theta}_k)\right\}\frac{f(Y_{i1k+1}|\mathcal{D}_{i1k+1},\boldsymbol{\theta}_k,\boldsymbol{\theta}_{k+1})}{\bar{F}(A_{k+1}|\mathcal{D}_{i1k+1},\boldsymbol{\theta}_k,\boldsymbol{\theta}_{k+1})}\right]$$
$$\left\{\prod_{j=2}^{m_{iK}-1} f(Y_{ijK}|\mathcal{D}_{ijK},\boldsymbol{\theta}_K)\right\}\bar{F}\{A_K - Y_{im_{ik}K}|\mathcal{D}_{im_{iK}},\boldsymbol{\theta}_K\}, \qquad (2.7.4)$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K)$, $\boldsymbol{\mathcal{D}} = \{(\boldsymbol{t}_{ik}, \boldsymbol{Y}_{ik})_{k=1,2,\ldots,K;i=1,2,\ldots,n}\}$ and $f(Y_{i1k+1}|\mathcal{D}_{i1k+1},\boldsymbol{\theta}_k,\boldsymbol{\theta}_{k+1})$ is equivalent to (2.7.1), and $\bar{F}(A_{k+1}|\mathcal{D}_{i1k},\boldsymbol{\theta}_k,\boldsymbol{\theta}_{k+1})$ for $k = 2,\ldots,K-1$ is given in (2.7.3).

Chapter 3

Penalized splines mixed effects model with random time-shift parameters

3.1 Abstract

Monitoring cervical dilation patterns is important in labor management for it may help obstetricians make pragmatic decisions whether a stalled labor should be allowed to continue or stopped for other options. Nevertheless, modeling labor curves remains a challenge since women are usually admitted into the hospital at different stages of the cervical dilation and the onset (i.e., time=0) of labor is often unknown. This constitutes error in the time variable that needs to be adjusted for in the analyses. Several nonlinear methods have been proposed to model labor curves where the time to full dilation (women are fully dilated at 10 cm of dilation) is used as the benchmark point and the time run backwards. However, these methods have drawbacks as they do not include women whose cervical dilation do not get to 10 cm, and cannot be used for prospective prediction. Recently, a change point model was proposed to characterize labor curves. The time since admission to the hospital for each individual woman was scaled with an additive factor, to adjust for the unknown onset of labor. This model adjusts for the unknown onset of labor but we feel that a change point model is not a good approximation of labor curves which tend to be nonlinear. Motivated by cervical dilation data from the CSL study, we propose to model labor curves with a semi-parametric model with random time shift parameters. We view each woman as having unknown time shift which when adjusted for appropriately

aligns her curve. Our proposed model uses flexible truncated basis functions to capture the non-linear relationship between cervical dilation and time. We use penalized smoothing splines to balance the complexity of the model with the goodness of fit. The model is formulated in mixed effects framework where the basis coefficients enter the model as random effects. The proposed model offers many advantages including easy implementation in standard statistical software, less computational cost and adjusts for the unknown onset of cervical dilation process. To implement the random time shift parameters, a Monte Carlo Expectation Maximization (MCEM) procedure was used. The random shift parameters were generated using a rejection sampling algorithm in E-step. In M-step, we propose an augmented data approach which allows convenient updating of the parameters using Restricted Maximum Likelihood Estimation (REML) procedure in *nlme* package in R. A non-parametric bootstrap approach was used to compute the parameter standard errors and point-wide confidence intervals. We demonstrate the proposed method through simulation studies and real data from Consortium of Safe Labor study.

3.2 Introduction

There has been increased interest in developing methods for characterizing labor patterns for contemporary obstetric practice. Defining labor curves however remains a challenging task due to the fact that women are admitted into the hospital at different stages of cervical dilation and the onset of labor may be unknown. One approach to this problem is to set the time to full dilation as the benchmark time and recode the time for prior dilation measurements as the number of hours before 10 cm. In this context earlier authors have modeled cervical dilation process with a polynomial repeated measures model, Zhang et al. (2002, 2010a) and with nonlinear mixed effects models, (Conell-Price et al., 2008; Arunajadai, 2010; Elmi et al., 2011, 2014). In our view, the aforementioned methods are not efficient enough for handling

labor curves given different subpopulations. That is, methods that use time at 10 cm dilation or time at delivery as the benchmark offer valid inferences only when there are no women with censored cervical dilations. With this in mind, McLain and Albert (2014) proposed a change point model with mixed effects, and rescaled each individual time with an additive factor to adjust for the unknown onset of labor. Whereas this method tends to alleviate the time zero problem thus correcting the error in time variable, cervical dilations exhibit nonlinear trends that may not be captured will by a random change point model. Alternatively, we propose to model cervical dilation process semi-parametrically with a penalized splines (P-splines) model as outlined by (Eilers and Marx, 1996).

There is immense literature where semi-parametric models have been used to analyze longitudinal data with nonlinear trends. The methods have evolved and more flexible model representations have emerged. Semi-parametric models have been expressed in linear mixed model (LMM) framework thus allowing the basis coefficients to enter the model as random effects. In LMM formulation, the problem of estimating an optimal smoothing parameter reduces to estimating the variance components and the smoothing parameter is estimated as a function of the basis coefficients variances (Durbán et al., 2005). Many researchers have explored the LMM formulation to tackle complex nonlinear models. For example, Zeger and Diggle (1994) used a semi-parametric mixed model to study the evolution of CD4+ cell count numbers of HIV sero-converters. The model was implemented by specifying a parametric model for the covariates adjustment, a non-parametric estimation of a smooth time trend and a serial correlation structure for measurements from the same subject. Verbyla et al. (1999) used a cubic smoothing spline mixed effects model to investigate the effect of treatment on cow growth profiles. Using an additive non-parametric model, Coull et al. (2001) investigated subjects' characteristics that affect their response to air pollution. The analyses were carried on the Utah Valley respiratory health and

air pollution study data. A first order auto-regressive correlation structure was used to capture the correlation between observations from the same subject. Wang et al. (2003) proposed the use of a semi-parametric nonlinear mixed model to determine disease effects on circadian rhythms of cortisol, an hormone that is affected by stress. The model was specified as an invariant-shape mixed model where the covariate effects were modeled using a smoothing spline ANOVA decomposition. Meyer (2005); Boligon et al. (2012) used B-splines random effects models to analyze the growth of Australian Angus and Nellore cattle, respectively. The analyses utilized weights data recorded from birth to 820 days of age and a weight data obtained from birth to adulthood, respectively. Durbán et al. (2005) has given a detailed demonstration on how to represent a semi-parametric model for longitudinal data as a linear mixed model and offers various examples from a random intercept model, a model with random intercept and random slope to a model with interactions of time and categorical variables. The formulation was demonstrated by analyzing a lymphoblastic leukaemia data to investigate the long term effect of radiation therapy on the health of children suffering from acute lymphoblastic leukaemia. See Zhang et al. (1998); Rice and Wu (2001); Guo (2002); Currie and Durban (2002); Wand (2003); Gurrin et al. (2005); Kauermann and Opsomer (2011) for more illustrations and inferences on linear mixed model representation of semi-parametric models.

To adjust for the unknown onset of labor, we introduce random shift parameters to align each individuals time points using the MCEM method as outlined in 3.3. The MCEM is a modification of Expectation Maximization (EM) algorithm whereby at the E-step the expectation of the conditional log-likelihood of the complete data is approximated by Monte Carlo sampling and the resulting expected log-likelihood is optimized in the M-step. The MCEM method has been implemented widely to approximate messy or intractable log-likelihoods in E-step. For example, through simulation studies Ganguli et al. (2005) implemented MCEM method to fit an ad-

ditive model where the predictor variables were subject to measurement error. Using a random effects model, Liu et al. (2005) incorporated the MCEM method to analyze latent genetic and environmental effects on censored outcomes in twin studies. Golan and Rosset (2011) fitted a random effects model to estimate heritability in genome wide studies where the parameter estimation was carried out using the MCEM method. Gad and El Kholy (2012) implemented MCEM method to fit a generalized linear mixed model whereby the random effects were assumed to have a log-normal distribution. Kang et al. (2013) used MCEM method to assess diagnostic test accuracy for cervical neoplasis in women with atypical glandular cells of undetermined significance (AGC). The true disease status was non-observed (latent variable) but observations on at least 3 conditionally independent diagnosis tests were available. See McCulloch (1997); Booth and Hobert (1999); Levine and Casella (2001); Kim et al. (2003); Neath et al. (2013) for more illustrations of MCEM and inferences. To implement the MCEM method, usually researchers use Markov chain Monte Carlo (MCMC) procedures such as Metropolis-Hastings algorithm McCulloch (1997); Kim et al. (2003), rejection sampling Booth and Hobert (1999) or important sampling Booth and Hobert (1999); Levine and Casella (2001) to obtain the Monte Carlo samples at the E-step. We chose to implement our MCEM procedure by sampling the random shift parameters using a rejection sampling algorithm. Our choice for a rejection sampling was motivated by guaranteed independent samples.

This project exploits the connection between penalized spline smoothing and linear mixed effects model to analyze labor curves while adjusting for the uncertainty in the onset of the process. This alternative approach has many advantages including easy implementation in standard software such as the *nlme* package in R Pinheiro et al. (2009) and less computational cost. To our knowledge, our research is the first to formulate a semi-parametric model in LMM framework, to analyze longitudinal data when the time onset of the process is prone to measurement error.

The main objective of this project is to formulate and fit the proposed model on a simulated data and CSL data. Specific objectives are (i) demonstrate the proposed model through simulation studies in terms of bias and Mean Squared Error (MSE) of the parameter estimates and estimate mean labor curves using simulated data (iii) apply the proposed model on the CSL data.

This Chapter is organized as follow. In Section 3.3, we present the methodology. We first present a review on the classical smoothing techniques, connection between P-splines model and the linear mixed model, briefly outline the EM method, rejection sampling algorithm and finally the proposed semi-parametric model. In Section 3.4, we present the simulation studies. In Section 3.5, we present the application. We present the conclusion and recommendation for further study in Section 3.6 and 3.7, respectively.

3.3 Methodology

3.3.1 Overview of classical smoothing techniques

For the moment we assume the time shift parameters are known variables. At the jth time point (j= 1,2,...,n_i), the cervical dilation of the ith woman (i=1,2,...,n) can be represented as a simple semi-parametric model

$$y_{ij} = f(t_{ij}) + \varepsilon_{ij} \tag{3.3.1}$$

where $f(t_{ij})$ function represents the mean cervical dilation at the jth time point and ε_{ij} is the measurement error. The function $f(.)$ can be approximated by linear combinations of K basis functions.

$$y_{ij} = \sum_{k=1}^{K} \beta_k q_k(t) + \varepsilon_{ij} \tag{3.3.2}$$

There are several types of basis functions like Fourier basis, wavelets basis and spline basis that are commonly used. Spline basis are a collection of piecewise polynomials

that connect smoothly at knots and the total number of basis functions depends on the number of knots (Coffey et al., 2014). Some of basis functions are B-splines, truncated line basis functions or truncated time power basis. We shall utilize truncated line basis functions due to their easy implementation in complicated models like our proposed model. Expressing the function $f(.)$ in 3.3.1 as a linear combination of truncated line basis functions the model becomes

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \sum_{k=1}^{K} \gamma_k (t_{ij} - q_k)_+ + \varepsilon_{ij} \tag{3.3.3}$$

where $(t_{ij} - q_k)_+ = \begin{cases} 0 & t_{ij} \leq q_k \\ (t_{ij} - q_k) & t_{ij} > q_k \end{cases}$

and the parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \boldsymbol{\gamma})$ are population intercept, effect of the linear slope and a vector of basis coefficients corresponding to the basis functions, respectively. The term $(t_{ij} - q_k)_+$ captures the nonlinearity of the mean function.

Expressing 3.3.3 as a linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ the regression coefficients $\boldsymbol{\beta}$ can be estimated by minimizing the residual sum of squares $RSS = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ where $\mathbf{X}$ is a design matrix given by

$$\mathbf{X} = \begin{pmatrix} 1 & t_{i1} & (t_{i1} - q_1)_+ & (t_{i1} - q_2)_+ & \cdots & (t_{i1} - q_k)_+ \\ 1 & t_{i2} & (t_{i2} - q_1)_+ & (t_{i2} - q_2)_+ & \cdots & (t_{i2} - q_k)_+ \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & t_{ini} & (t_{ini} - q_1)_+ & (t_{ini} - q_2)_+ & \cdots & (t_{ini} - q_k)_+ \end{pmatrix}$$

and one can evaluate the optimal number of basis functions by Cross Validation (CV), Generalized Cross Validation (GCV) or Akaike Information Criteria (AIC). Alternatively, to obtain the parameter estimates $\boldsymbol{\beta}$ one can employ smoothing splines regression model and minimize the penalized RSS function

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \alpha \int f''(t)^2 dt \tag{3.3.4}$$

Smoothing splines regression model requires placing the knots at each unique time point and imposes a penalty on the minimization function to control over-fitting of the

mean curve by shrinking the basis coefficients towards a linear fit. An added complex arises in solving the integral in the penalty term since it requires numerical integration to solve and becomes more complex as the sample size increases. A more flexible approach would be to use a Penalized splines smoothing (P-splines) model that is less computational for it imposes a discrete penalty on the basis coefficients and is relatively less sensitive to the number of basis functions Ruppert (2002). Ruppert (2002) recommends the number of knots K to be $K = \min(\frac{1}{4}\text{number of unique } t'_{ij}s,\ 40)$. In a P-splines model the penalized RSS function becomes

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \alpha\boldsymbol{\beta}^T\mathbf{D}\boldsymbol{\beta} \tag{3.3.5}$$

where $\mathbf{D}$ is a penalty matrix such that $\boldsymbol{\beta}^T\mathbf{D}\boldsymbol{\beta} = \sum_{k=1}^{K}\gamma_k^2$. Both P-splines and smoothing splines regression uses a smoothing parameter α to control the smoothness of the curve and an optimal smoothing parameter can be obtained using CV, GCV or AIC.

3.3.2 P-splines model as a LMM model

As demonstrated in Section 3.3.1, the classical smoothing techniques are more cumbersome to implement and less flexible. In addition, estimating $\boldsymbol{\beta}$ by minimizing the least squares tends to give rough curves due to fluctuations in the basis coefficients estimates γ_k. A better approximation to the mean function and less cumbersome modeling approach would be to shrink the γ_k estimates towards zero by letting $\gamma_k \sim N(0, \sigma_\gamma^2)$ which is equivalent to penalization and works by replacing γ_k with BLUPs (Best Linear Unbiased Prediction) in mixed effects model Wand (2003); Gurrin et al. (2005).

Expressing the P-splines model as a LMM model replaces the function $f(.)$ with a parametric form

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{3.3.6}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)$ and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_k)$

$$\mathbf{X}_i = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{ini} \end{bmatrix}, \quad \mathbf{Z}_i = \begin{bmatrix} (t_{i1} - q_1)_+ & (t_{i1} - q_2)_+ & \cdots & (t_{i1} - q_k)_+ \\ (t_{i2} - q_1)_+ & (t_{i2} - q_2)_+ & \cdots & (t_{i2} - q_k)_+ \\ \vdots & \vdots & \cdots & \vdots \\ (t_{ini} - q_1)_+ & (t_{ini} - q_2)_+ & \cdots & (t_{ini} - q_k)_+ \end{bmatrix}$$

To achieve smoothness, a penalty term is imposed on the basis coefficients resulting into the penalized likelihood function

$$l(\boldsymbol{\beta}; \boldsymbol{\gamma}; \alpha) = -\frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\gamma})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\gamma}) - \frac{1}{2}\boldsymbol{\gamma}^T\mathbf{D}\frac{\boldsymbol{\gamma}}{\alpha} \tag{3.3.7}$$

where $\mathbf{D}$ is the penalty matrix. The penalized likelihood function 3.3.7 corresponds to the penalized minimization criterion PRSS $= ||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\gamma}||^2 + \alpha||\boldsymbol{\gamma}||^2$ and the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are obtained as the BLUPs of the mixed model.

3.3.3 Expectation Maximization

Expectation Maximization algorithm is a method for parameter estimation using likelihood functions in presence of missing or incomplete data (Dempster et al., 1977). The method partitions data into $\mathbf{y} = (\mathbf{y}_{obs}, \mathbf{y}_{mis})$ where $\mathbf{y}_{obs}$ is the observed "incomplete" data, and $\mathbf{y}_{mis}$ is the unobserved "missing data". The interest is to maximize the log-likelihood of the complete data $l(\theta; y) = log[f(\mathbf{y}; \theta)]$. However, the log likelihood function of the complete data can not be maximized since $\mathbf{y}$ is not wholly observed. Instead, the EM method replaces the $log[f(\mathbf{y}; \theta)]$ with its conditional expectation given the observed data $\mathbf{y}_{obs}$ at current values of $\theta \in \Theta$. The EM method follows the following steps to estimate parameter $\boldsymbol{\theta}$.

Step 0: Choose the initial values of parameters $\boldsymbol{\theta}^0$. Step 1: At the E-step, at rth iteration, the conditional expectation of the complete data log-likelihood is computed given the observed data $\mathbf{y}_{obs}$ at the current parameter $\boldsymbol{\theta}^r$ as

$$\mathbf{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^r) = E_{\boldsymbol{\theta}}[l(\boldsymbol{\theta}; y)|y_{obs}, \boldsymbol{\theta}^r]$$

Step 2: At the M-step, the new parameter $\boldsymbol{\theta}^{r+1}$ is found by maximizing $\mathbf{Q}(\boldsymbol{\theta};\boldsymbol{\theta}^{r})$ such that

$$\mathbf{Q}(\boldsymbol{\theta}^{r+1};\boldsymbol{\theta}^{r}) \geq \mathbf{Q}(\boldsymbol{\theta};\boldsymbol{\theta}^{r})$$

The algorithm then iterates between the Expectation (E-step) and the Maximization (M-step) steps to get maximum likelihood estimates (MLE) of the parameter $\boldsymbol{\theta}$. The EM method is applicable when the E-step is tractable which does not always happen. The MCEM method is commonly applied when the equation in E-step is intractable or messy. The MCEM is a modification of EM algorithm whereby at the E-step the expectation of the conditional log likelihood of the complete data is approximated by Monte Carlo sampling and the resulting expected log-likelihood is optimized in the M-step.

3.3.4 Rejection Sampling

Rejection sampling is a sampling technique used for drawing random samples from a given distribution to obtain numerical results. Rejection sampling does not require knowledge of the conditional posterior distribution of the parameters. However, the method requires a candidate distribution $q(\theta)$ which is easy to sample from and resembles the posterior density function $p(\theta|y)$ of interest in terms of location and support. The method requires the "envelope property" to hold. That is, for all unknown parameters

$$p(\theta|y) \leq c(q(\theta)) \tag{3.3.8}$$

where the posterior density $p(\theta|y)$ is the target distribution, $q(\theta)$ is the candidate distribution and c is a normalizing constant (Ghosh et al., 2007). The posterior density function $p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \propto p(y|\theta)p(\theta)$ since $p(y)$ is a constant in reference to θ. The $p(\theta)$ is the prior distribution of θ. Rejection sampling is implemented by drawing θ from the candidate distribution and $u \sim U(0,1)$ independently. At the rth

draw, accept θ^r to be from the target distribution if

$$u \leq \frac{p(\theta|y)}{c(q(\theta))} \tag{3.3.9}$$

otherwise reject and again draw θ from the candidate distribution and compare to a new u via 3.3.9. This process is repeated until M samples are drawn.

3.3.5 Proposed Semi-parametric model

We utilize the connection between linear mixed effects model and penalized smoothing splines to fit the proposed semi-parametric model and incorporate random time shift parameters.

Modeling framework

Let $\mathbf{t}_i = (t_{i1}, t_{i2}, ..., t_{in_i})$ and $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{in_i})$ be the vector of the observed time points and repeated measurements of the outcome for the ith subject, $(i = 1, 2, ..., n, j = 1, 2, ..., n_i)$. The $\mathbf{t}_i$ is the recorded time points since the woman was admitted in the hospital. Define $(\mathbf{t}_i - \Delta_i)$ as the shifted time for the ith subject where $\Delta_i \sim N(0, \sigma^2_\Delta)$ is a random shift parameter.

At time point t_{ij}, we model the outcome of the ith subject as

$$y_{ij} = f(t_{ij} - \Delta_i) + g_i(t_{ij} - \Delta_i) + \varepsilon_{ij} \tag{3.3.10}$$

where $f(.)$ is the population mean, $g_i(.)$ captures the deviation from the overall mean and $\varepsilon \sim N(0, \sigma^2_\varepsilon)$ is the measurement error for the jth measurement. We estimate the functions $f(.)$ and $g_i(.)$ in (3.3.10) by K and K_c truncated line basis functions respectively. The model imposes a penalty on the basis coefficients in $f(.)$ function to control for over fitting of the mean curve. We evaluate the basis functions at fixed knots $q_1 < q_2, ..., < q_k$ and $q_1 < q_2, ..., < qk_c$ where the knots are defined as quantiles

of the observed time. In this regard, (3.3.10) becomes

$$Y_{ij} = \beta_0 + \beta_1(t_{ij} - \Delta_i) + \sum_{k=1}^{K} \gamma_k (t_{ij} - \Delta_i - q_k)_+ + b_i(t_{ij} - \Delta_i) + \sum_{k=1}^{K_c} \lambda_{ik}(t_{ij} - \Delta_i - q_k)_+ + \varepsilon_{ij} \tag{3.3.11}$$

For simplicity of notations, we shall assume $K = K_c$ unless stated otherwise.

In model 3.3.11, the parameters γ_k are the basis coefficients common across subjects and λ_{ik} are subject-specific basis coefficients. The term $(t_{ij} - \Delta_i - q_k)_+$ captures the nonlinear trends of the labor curves and is such that $(t_{ij} - \Delta_i - q_k)\mathbf{1}(t_{ij} - \Delta_i > q_k)$. As demonstrated in Section 3.3.2, we implement the proposed semi-parametric model as a linear mixed model where the basis coefficients $\gamma_k \sim N(0, \sigma_\gamma^2)$ and $\lambda_{ik} \sim N(0, \sigma_\lambda^2)$ enter the model as random effects.

In the linear mixed model formulation, the matrix form of (3.3.11) is

$$\mathbf{Y} = \mathbf{X}(\Delta)\boldsymbol{\beta} + \mathbf{Z}(\Delta)\mathbf{u} + \boldsymbol{\varepsilon} \tag{3.3.12}$$

where $\mathbf{Y}^T = [\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n]$, $\boldsymbol{\beta}^T = [\beta_0, \beta_1]$

$$\mathbf{X}(\Delta) = \begin{bmatrix} \mathbf{1} & \mathbf{X}_1 \\ \mathbf{1} & \mathbf{X}_2 \\ \vdots & \vdots \\ \mathbf{1} & \mathbf{X}_n \end{bmatrix}, \quad \mathbf{Z}(\Delta) = \begin{bmatrix} \mathbf{Z}_{1p} & \mathbf{X}_1 & 0 & \cdots & 0 & \mathbf{Z}_{1s} & 0 & \cdots & 0 \\ \mathbf{Z}_{2p} & 0 & \mathbf{X}_2 & \cdots & 0 & 0 & \mathbf{Z}_{2s} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_{np} & 0 & 0 & \cdots & \mathbf{X}_n & 0 & 0 & \cdots & \mathbf{Z}_{ns} \end{bmatrix}$$

For the ith subject, $\mathbf{X}_i$ is a vector of shifted time points and $\mathbf{Z}_{ip}$ is a matrix of the basis functions corresponding to the basis coefficients common across subjects

$$\mathbf{X}_i = \begin{bmatrix} t_{i1} - \Delta_i \\ t_{i2} - \Delta_i \\ \vdots \\ t_{in_i} - \Delta_i \end{bmatrix}, \quad \mathbf{Z}_{ip} = \begin{bmatrix} (t_{i1} - \Delta_i - q_1)_+ & ... & (t_{i1} - \Delta_i - q_k)_+ \\ \vdots & \ddots & \vdots \\ (t_{ini} - \Delta_i - q_1)_+ & ... & (t_{ini} - \Delta_i - q_k)_+ \end{bmatrix}$$

and $\mathbf{Z}_{is}$ is a matrix of the basis functions corresponding to the basis coefficients specific to the ith subject. The random vectors $\mathbf{u} = [\boldsymbol{\gamma}, b_1, b_2, ..., b_n, \boldsymbol{\lambda}]$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_k)$,

and $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_n)$ where $\boldsymbol{\lambda}_i = (\lambda_{i1}, \lambda_{i2}, ..., \lambda_{ik})$. The vectors $\boldsymbol{\varepsilon}$ and $\mathbf{u}$ are assumed to be independent. That is,

$$\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \sigma_\varepsilon^2 \mathbf{I} \end{bmatrix} \right) \text{ where } \boldsymbol{\Sigma} = \text{cov}(\mathbf{u}) = \begin{bmatrix} \sigma_\gamma^2 \mathbf{I} & 0 & 0 \\ 0 & \sigma_b^2 \mathbf{I} & 0 \\ 0 & 0 & \sigma_\lambda^2 \mathbf{I} \end{bmatrix}$$

As explained in Section 3.3.2, for a given σ_ε^2 and $\boldsymbol{\Sigma}$ the estimates of $\boldsymbol{\beta}$ and $\mathbf{u}$ of a linear mixed model are equivalent to solutions to the penalized least squares minimization problem

$$\begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{u}} \end{bmatrix} = \text{argmin}_{\beta, \mathbf{u}} \left(||\mathbf{Y} - \mathbf{X}(\Delta)\boldsymbol{\beta} - \mathbf{Z}(\Delta)\mathbf{u}||^2 + \alpha ||\boldsymbol{\gamma}||^2 \right)$$

where α is a smoothing parameter expressed as $\alpha = \sigma_\varepsilon^2 / \sigma_\gamma^2$

Estimation & Inference

We estimate the unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_\Delta^2, \sigma_\varepsilon^2, \boldsymbol{\Sigma})$ of model (3.3.12). If the shift parameters $\boldsymbol{\Delta}$ were known variables model (3.3.12) would be fit as a linear mixed effects model. Typically a REML procedure (Patterson and Thompson, 1971) is used as the estimation criteria whereby the restricted log likelihood

$$l_R(\sigma_\gamma^2, \sigma_\lambda^2, \sigma_b^2, \sigma_\varepsilon^2) = -\frac{1}{2}\log|\mathbf{V}| \tag{3.3.13}$$
$$-\frac{1}{2}\log|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{y}^T(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1})\mathbf{y}$$

is optimized with respect to the variance components $(\sigma_\varepsilon^2, \sigma_b^2, \sigma_\gamma^2, \sigma_\lambda^2)$, $\mathbf{V} = \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}^T + \mathbf{R}$ and $\mathbf{R} = \boldsymbol{\sigma}_\varepsilon^2 I_n$. The fixed effects $\boldsymbol{\beta}$ and random coefficients $\mathbf{u}$ can then be determined as the solutions of the mixed model equations (Henderson, 1973; Robinson, 1991).

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}, \quad \widehat{\mathbf{u}} = \widehat{\boldsymbol{\Sigma}}\mathbf{Z}^T\widehat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})$$

Here $\boldsymbol{\Delta}$ being random complicates the estimation procedure by inducing an intractable integral in the E-step. To circumvent the estimation problem, we incorpo-

rate the random time shifts parameters via an MCEM procedure (Wei and Tanner, 1990) and sample the shift parameters using a rejection sampling algorithm in E-step.

MCEM approximation and variables definitions

Let $\mathbf{y}$ be the observed data with density function $f(\mathbf{y};\boldsymbol{\theta})$ and $\boldsymbol{\Delta}$ the unobserved variable with density function $f(\boldsymbol{\Delta})$ and posterior distribution

$$h(\boldsymbol{\Delta};\mathbf{y},\boldsymbol{\theta}) \propto f(\mathbf{y}|\boldsymbol{\Delta};\boldsymbol{\theta}) f(\boldsymbol{\Delta}|\sigma^2_{\boldsymbol{\Delta}}) \tag{3.3.14}$$

Further let $(\mathbf{y},\boldsymbol{\Delta})$ be the complete data with conditional log-likelihood

$$\log\{f(\mathbf{y},\boldsymbol{\Delta};\boldsymbol{\theta})|\mathbf{y}\}$$

In linear mixed effects model, the Expectation Maximization algorithm evaluates the log-likelihood of the complete data to obtain the MLE of $\boldsymbol{\theta}$ over the marginal distribution of $\mathbf{y}$ obtained after integrating out the random effects $\mathbf{u} = (\boldsymbol{\gamma},\boldsymbol{\lambda},\mathbf{b})$. Note that given $\boldsymbol{\Delta}$

$$f(\mathbf{y};\boldsymbol{\theta}) = \int\int\int f(\mathbf{y},\boldsymbol{\gamma},\boldsymbol{\lambda},\mathbf{b}|\boldsymbol{\theta}) d\boldsymbol{\gamma} d\boldsymbol{\lambda} d\mathbf{b} = \int f(\mathbf{y},\mathbf{u}|\boldsymbol{\theta}) d\mathbf{u}$$

has a closed form unlike when $\boldsymbol{\Delta}$ is unknown and one requires solving the integral

$$\int f(\mathbf{y};\boldsymbol{\theta}) dF(\boldsymbol{\Delta}|\sigma^2_{\boldsymbol{\Delta}}) = \int f(\mathbf{y},\boldsymbol{\Delta}|\boldsymbol{\theta}) d\boldsymbol{\Delta}$$

Our approach is to use Monte Carlo simulations to approximate the expected conditional log-likelihood of the complete data in E-step which at the r iteration is given by

$$\mathbf{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) = \mathrm{E}_{\boldsymbol{\theta}^{(r)}}[\log\{f(\mathbf{y},\boldsymbol{\Delta};\boldsymbol{\theta})|\mathbf{y}\}] = \int \log\{f(\mathbf{y},\boldsymbol{\Delta};\boldsymbol{\theta})\} h(\boldsymbol{\Delta};\mathbf{y},\boldsymbol{\theta}^{(r)}) d\boldsymbol{\Delta} \tag{3.3.15}$$

Let Δ_{il}be the lth sample of the Δ value for the ith subject, l=1,2,...,M. In E-step we replace (3.3.15) with the Monte Carlo approximation

$$\mathbf{Q}_M(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r-1)}) = \frac{1}{M}\sum_{l=1}^{M}\sum_{i=1}^{n} \log\left\{f(\mathbf{y}_i,\Delta_{il}^{(r)};\boldsymbol{\theta})\right\} \tag{3.3.16}$$

where $\mathbf{y}_i$ is the observed data for the ith subject.

MCEM Algorithm steps

We carry out the estimation procedure by first getting the initial parameters $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\beta}, \sigma^2_\varepsilon, \sigma^2_\Delta, \boldsymbol{\Sigma})$ by fitting model (3.3.12) as a linear mixed effects model in *nlme* R package. We initiate Δ as $\Delta \sim N(0,1)$ and follow the following algorithm steps until convergence:

Step 1: At the rth iteration generate Δ_{il} for the ith subject using rejection sampling algorithm given $\boldsymbol{\theta}^{(r-1)}$. We sample Δ_{il} from their posterior distribution $h(\boldsymbol{\Delta}; \mathbf{y}_i, \boldsymbol{\theta})$ by (a): Drawing Δ_{il} from a candidate distribution $g(\Delta)$, a $N(0, \sigma^2_\Delta)$ and independently drawing $u \sim U(0,1)$ (b): We accept Δ_{il} as a random sample from $h(\boldsymbol{\Delta}; \mathbf{y}_i, \boldsymbol{\theta})$ if

$$u \leq \frac{f(\mathbf{y}_i | \Delta^{(r)}_{il}; \boldsymbol{\theta}^{(r-1)}, \widehat{\boldsymbol{\gamma}}^{(r-1)}) f(\Delta)}{c g(\Delta)} \tag{3.3.17}$$

where $c = \sup_\Delta \left\{ f(\mathbf{y}_i | \Delta; \boldsymbol{\theta}^{(r-1)}, \widehat{\boldsymbol{\gamma}}^{(r-1)}) \right\}$. Otherwise we reject and repeat Step (a) to (b) until M samples are drawn. The acceptance criteria equation (3.3.17) is equivalent to

$$u \leq \frac{f(\mathbf{y}_i | \Delta^{(r)}_{il}; \boldsymbol{\theta}^{(r-1)}, \widehat{\boldsymbol{\gamma}}^{(r-1)})}{c}$$

since the candidate distribution $g(\Delta)$ is chosen to be equal to the prior distribution $f(\Delta)$. It is important to note that, though the basis coefficients $\boldsymbol{\gamma}$ are modeled as random effects, the $\boldsymbol{\gamma}$'s are population level parameters. Thus, in the acceptance criteria 3.3.17 we assume $\mathbf{y}$ has a marginal multivariate normal distribution with mean $\mathbf{X}(\Delta)\boldsymbol{\beta} + \mathbf{Z}_p\boldsymbol{\gamma}$ and variance structure $\mathbf{Z}_{ts}\mathbf{G}\mathbf{Z}^T_{ts} + \mathbf{R}$, where $\mathbf{R}$ is as defined before and $\mathbf{G}$ is the covariance matrix of subject-specific random effects $\mathbf{b}_i$ and $\boldsymbol{\lambda}_i$. The design matrix $\mathbf{Z}_p$ is as defined before and $\mathbf{Z}_{ts}$ is the design matrix for subject-specific random effects

$$\mathbf{Z}_{ts} = \begin{bmatrix} t_{i1} - \Delta_i & (t_{i1} - \Delta_i - q_1)_+ & \dots & (t_{i1} - \Delta_i - q_k)_+ \\ \vdots & \vdots & \ddots & \vdots \\ t_{ini} - \Delta_i & (t_{ini} - \Delta_i - q_1)_+ & \dots & (t_{ini} - \Delta_i - q_k)_+ \end{bmatrix}$$

Step 2: In the E-step replace (3.3.15) with the Monte Carlo approximation as shown in equation 3.3.16. Step 3: In the M-step, $\boldsymbol{\theta}$ is updated by optimizing the log-likelihood of the resulting augmented data using the REML procedure in the *nlme* R package. The augmented data is the duplicate of the original data such that for the ith subject, the "new" data is

$$(\mathbf{Y}_i, \mathbf{t}_i)^T_{new} = \begin{bmatrix} \mathbf{Y}_i & \mathbf{Y}_i & ... & \mathbf{Y}_i \\ \mathbf{t}_i - \Delta_{i1} & \mathbf{t}_i - \Delta_{i2} & ... & \mathbf{t}_i - \Delta_{iM} \end{bmatrix}$$

The estimates from Step 3 are used as the values for the $(r+1)$th iteration. At rth iteration, we update σ^2_Δ with the $\boldsymbol{\Delta}$ sample variance from $(r-1)$ iteration given by $\frac{1}{nM-1}\sum_{i=1}^{n}\sum_{l=1}^{M}(\Delta_{il} - \mu_\Delta)^2$.

The MCEM procedure is an approximation to the EM so there is a possibility of premature convergence due to the Monte Carlo sampling error and no guarantee that the likelihood will improve or the ascent property will hold. We therefore repeat Steps (1) to (3) until $|l_R(\boldsymbol{\theta})^{(r)} - l_R(\boldsymbol{\theta})^{(r-1)}| < 0.005$ for three consecutive iterations and take the final estimated parameters as the MLE. Here $l_R(\boldsymbol{\theta})$ is the restricted log likelihood obtained from the REML procedure in Step 3. Usually the estimates from first iterations are further from the MLE. Therefore, we generate a small number of Δ's per subject in the early iterations and increase gradually as the algorithm progresses. Table 3.1 shows the generated number of Δ samples relative to the number of iterations.

Table 3.1 Updating the size of M at each iteration in MCEM algorithm

Iterations	1 - 9	10 - 24	25 - 39	40 - 54	55 - 69	70 and above
Δ samples	10	50	100	200	500	1000

3.3.6 Model checking and Goodness of Fit

When analyzing longitudinal data, one is required to chose the best mean function and variance-covariance structure that leads to a simple and yet a parsimonious fit that adequately explains the variability in the outcome variable. The proposed semi-parametric model (3.3.12) introduces basis functions (nonlinear terms) in both the population mean and subject-specific slopes. To test whether a simpler model is more appropriate for modeling the labor curves, one can test whether the variances of the basis coefficients are zero. Using the hierarchy of random effects, one can test whether the nonlinear term in the random slopes should be omitted using the following hypothesis $H(0) : \sigma^2_\lambda = 0 :$ versus $H(1) : \sigma^2_\lambda > 0$

In linear mixed models, model building usually involves using a likelihood ratio (LRT) test based on REML to compare models with different covariance structures and a LRT test based on ML to compare models with different mean structures. A LRT test based on the REML is of the form

$$LRT_{Res} = -2[l_R(H0) - l_R(H1)] \tag{3.3.18}$$

where $l_R(H0)$ and $l_R(H1)$ are the restricted log likelihood under the null and under the alternative hypotheses respectively. The test statistic is assumed to have a χ^2_v under H0 where v is the difference in the number of parameters between models under H1 and H0.

It is worthy to note that, in marginal models the covariance structure $\mathbf{V}$ is required to be positive definite which is a weaker assumption compared to when subject-specific inferences are of interest. For example, in a hierarchical model (like the proposed model) the covariance structure $\mathbf{\Sigma}$ of the random effects is required to be positive definite. Evaluating variance components in this case results into one-sided hypothesis in a constrained parameter space. This is as a result of all the hypothesis laying on the boundary of parameter space (Molenberghs and Verbeke, 2007). Thus

evaluating null hypothesis such as $\sigma^2_\lambda = 0$ places the covariance structure on the boundary of parameter space since $\sigma^2_\lambda \in [0, \infty]$. Therefore the above hypothesis test is of "nonstandard" form and the test statistic (3.3.18) may be inappropriate in this case. A better alternative would be to approximate the distribution of LRT_{Res} with a mixture of chi-square distributions

$$\frac{1}{2}\chi^2_q + \frac{1}{2}\chi^2_{q+1} \tag{3.3.19}$$

where q is the number of fixed effects under the null hypothesis. Self and Liang (1987); Stram and Lee (1994) have demonstrated that under the assumption that the vector $\mathbf{y}$ can be subsetted into independent sub vectors and that the number of subvectors tend to infinity, then the asymptotic distribution of LRT_{Res} is a mixture of chi-square distribution. However, this assumption does not hold under the alternative hypothesis for semi-parametric model and some authors have suggested use of simulations to determine the distribution of likelihood ratio test under the null hypothesis (Crainiceanu et al., 2002). The simulation approach entails estimating model parameters at the null hypothesis then simulating the distribution of the likelihood ratio test under the null hypothesis at the estimated parameters (Durbán et al., 2005). An R package *RLRsim* exists can be used to implement this approach. The method however is not feasible for complex models especially with large number of observations like our proposed model.

For illustration purposes, Durbán et al. (2005) used a mixture of chi-square test to examine if a simpler semi-parametric model that assumes equal covariance structures across treatment groups was more adequate than a model that specified different covariance structures per treatment group. Our proposed model is a semi-parametric model that employs the MCEM method in parameter estimation. In reference to this, we feel that none of the mentioned goodness of fit tests are appropriate. In future, we shall investigate more on model diagnosis for the proposed model.

3.4 Simulation

We used simulation studies to demonstrate the performance of our proposed model in terms of parameter bias, standard errors and mean squared error. The model was applied on 200 simulated datasets each with 500 sample size. We generated the data of the ith subject as in model (3.3.12) and compared the parameter estimates when different number of knots (K, K_c) are used in fitting the proposed model. We also checked the impact that the number of measurements n_i per subject had on the parameter estimates for a given set of (K, K_c). In this regard, we did simulations assuming n_i is 5 and 10 and the observed time is $t_{ij} \sim U(0, 5)$. For n_i =10, we set (K, K_c) to be either (2,2), (2,5), (5,2), (10,5), (30,5) or (30,3) and for n_i =5, (K, K_c) was either (2,2), (10,3), (30,3) or (5,2). The true values of parameters, $\boldsymbol{\theta}$ were specified as β_0 =1, β_1=1.5, σ^2_ε=0.5, σ^2_Δ=1, σ^2_γ=0.015, σ^2_λ=0.05, σ^2_b=1.

Tables 3.2 to 3.5 show results from simulated data. From the results we observe that, when the number of observations recorded per subject are n_i =10, the parameter estimates from all the models have similar bias and MSE except when K, K_c = (2,2) where the estimates have relatively larger bias. When n_i =5 the estimates have similar bias and MSE across all the models. When same number of knots is used, the estimates have relatively smaller MSE and higher bias when ni=10 than when ni=5. Overall, all the estimates had small bias and MSE across models regardless of the number of knots used or number of observations per subjects. We also plotted the mean individual profiles against the true profiles as displayed in Figure 3.1. The red lines represent the estimated mean profiles and the dark lines the true profiles for 20 subjects. The estimated mean profiles are close to the true profiles as observed in the Figure.

Table 3.2 Summary of 200 simulated datasets with n=500 and n_i=10

	TRUE	EST	BIAS	SE	MSE
			$K, K_c = (10, 5)$		
β_0	1	0.9762	-0.0238	0.0584	0.0040
β_1	1.5	1.5263	0.0263	0.1096	0.0127
σ^2_ε	0.5	0.5035	0.0035	0.0128	0.0002
σ^2_Δ	1	0.9751	-0.0249	0.0794	0.0069
σ^2_λ	0.05	0.0590	0.0090	0.0229	0.0006
σ^2_b	1	0.9887	-0.0113	0.0722	0.0053
			$K, K_c = (30, 5)$		
β_0	1	0.9851	-0.0149	0.0586	0.0036
β_1	1.5	1.4943	-0.0057	0.1215	0.0148
σ^2_ε	0.5	0.5073	0.0073	0.0125	0.0003
σ^2_Δ	1	0.9934	-0.0066	0.0761	0.0058
σ^2_λ	0.05	0.0748	0.0248	0.0185	0.0009
σ^2_b	1	0.9354	-0.0646	0.0797	0.0106

Table 3.3 Summary of 200 simulated datasets with n=500 and n_i=5

	TRUE	EST	BIAS	SE	MSE
			$K, K_c = (10, 3)$		
β_0	1	1.0000	0.0000	0.0704	0.0050
β_1	1.5	1.4911	-0.0089	0.1384	0.0193
σ^2_ε	0.5	0.5045	0.0045	0.0201	0.0004
σ^2_Δ	1	0.9896	-0.0104	0.0849	0.0073
σ^2_λ	0.05	0.0278	-0.0222	0.0498	0.0030
σ^2_b 1		1.0250	0.0250	0.0924	0.0091
			$K, K_c = (30, 3)$		
β_0	1	0.9957	-0.0043	0.0726	0.0053
β_1	1.5	1.5013	0.0013	0.1455	0.0212
σ^2_ε	0.5	0.5101	0.0101	0.0205	0.0005
σ^2_Δ	1	0.9934	-0.0066	0.0886	0.0079
σ^2_λ	0.05	0.0418	-0.0082	0.0566	0.0033
σ^2_b	1	1.0007	0.0007	0.0989	0.0098

Table 3.4 Summary of 200 simulated datasets with n=500, K, K_c =(2,2) knots for both ni=10 and ni=5 observations

	TRUE	EST	BIAS	SE	MSE
			n_i=10		
β_0	1	0.9554	-0.0446	0.0644	0.0061
β_1	1.5	1.5247	0.0247	0.0663	0.0050
σ^2_ε	0.5	0.5032	0.0032	0.0137	0.0002
σ^2_Δ	1	0.9552	-0.0448	0.0742	0.0075
σ^2_λ	0.05	0.0843	0.0343	0.0649	0.0054
σ^2_b	1	0.9863	-0.0137	0.0713	0.0053
			n_i=5		
β_0	1	0.9897	-0.0103	0.0637	0.0042
β_1	1.5	1.5020	0.0020	0.0786	0.0062
σ^2_ε	0.5	0.5055	0.0055	0.0218	0.0005
σ^2_Δ	1	0.9727	-0.0273	0.0866	0.0082
σ^2_λ	0.05	0.0489	-0.0011	0.0620	0.0038
σ^2_b	1	1.0056	0.0056	0.0844	0.0071

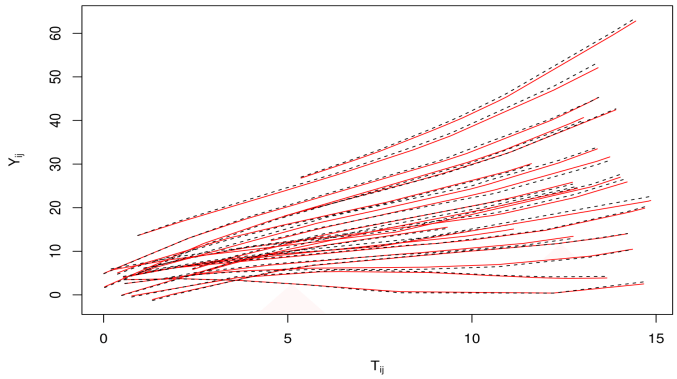


Figure 3.1 Estimated individual profiles. $n = 500, n_i = 10, K, K_c = (30, 5)$

3.5 Application on Consortium on Safe Labor data

We applied the proposed semi-parametric model on the CSL data. We sampled 500 women from 11,521 nulliparous women from the Utah site and fitted the proposed model (3.3.12) with time variable only. We used K, K_c = (30,3) knots. Table 3.6 shows the parameter estimates, their corresponding standard errors and 95% point-

Table 3.5 Summary of 200 simulated datasets with n =500, K, K_c =(5,2) knots for both ni =10 and ni=5 observations

	TRUE	EST	BIAS	SE	MSE
			n_i=10		
β_0	1	0.9896	-0.0104	0.0577	0.0034
β_1	1.5	1.5556	0.0556	0.1396	0.0226
σ^2_ε	0.5	0.5036	0.0036	0.0118	0.0001
σ^2_Δ	1	0.9617	-0.0383	0.0762	0.0073
σ^2_λ	0.05	0.0743	0.0243	0.0514	0.0032
σ^2_b	1	0.9837	-0.0163	0.0695	0.0051
			n_i=5		
β_0	1	0.9897	-0.0103	0.0637	0.0042
β_1	1.5	1.5020	0.0020	0.0786	0.0062
σ^2_ε	0.5	0.5055	0.0055	0.0218	0.0005
σ^2_Δ	1	0.9727	-0.0273	0.0866	0.0082
σ^2_λ	0.05	0.0489	-0.0011	0.0620	0.0038
σ^2_b	1	1.0056	0.0056	0.0844	0.0071

wise confidence intervals. The standard errors and point-wise confidence intervals were computed using 250 non-parametric bootstraps. We set the initial variance of Δ as 1.

Table 3.6 Parameter estimates from the semi-parametric model with n=500 nulliparous women based on K, K_c =(30,3)

	EST	SE	95% CI
β_0	2.530	0.139	(2.382, 2.974)
β_1	0.285	0.313	(-0.412, 0.702)
σ^2_ε	0.854	0.090	(0.788, 1.154)
σ^2_Δ	3.080	0.976	(1.654, 5.051)
σ^2_λ	0.448	0.163	(0.338 , 0.929)
σ^2_b	0.512	0.158	(0.297 , 0.867)

From Table 3.6, the parameter estimates of fixed intercept and slope are (β_0=2.530, 95%IC: [2.382, 2.974]) and (β_1=0.285, 95%IC: [-0.412, 0.702]). We estimated the variance components as: measurement error variance (0.854, 95%IC: [0.788, 1.154]), variance of the random shift parameter (3.080, 95%IC: [1.654, 5.051]), variance of

subject-specific basis coefficients (0.448, 95%IC : [0.338 , 0.929]) and the variance of random effect for the linear time slope (0.512, 95%IC: [0.297 , 0.867]).

We also plotted the mean labor curve of women from the CSL data. From the fitted model on the CSL data, we obtained the population level estimates $(\hat{\beta}_0, \hat{\beta}_1, \hat{\boldsymbol{\gamma}})$ and the variances of subject-specific random effects $\widehat{\sigma}_b^2$ and $\widehat{\sigma}_\lambda^2$. Then we randomly generated 125 samples of $\mathbf{b} \sim MVN(\mathbf{0}, \widehat{\sigma}_b^2\mathbf{I})$, and $\boldsymbol{\lambda} \sim MVN(\mathbf{0}, \widehat{\sigma}_\lambda^2\mathbf{I})$. Given the population level estimates and the sampled subject-specific random effects, we then estimated 125 mean labor curves. Figure 3.2 shows the average mean labor curve (dashed black line), 25th (gray line), 50th (blue line) and 75th (gray line) percentiles. We estimate that 50% of the mean labor curves are within the gray lines as shown in the Figure.

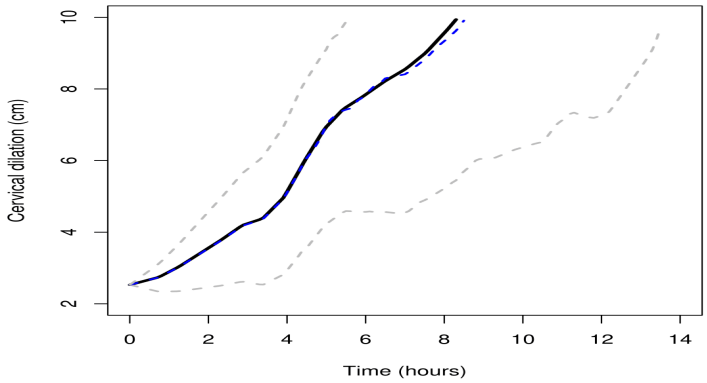


Figure 3.2 Estimated mean labor curve of women from the CSL study

3.6 Conclusion

We have demonstrated the proposed model through simulation studies and application on a real dataset from CSL study. Through simulation studies we found that, the estimated parameters have small bias and MSE regardless of the number of knots. Our findings collaborate with the Ruppert (2002) recommendations on knot selection for a P-splines model. It seems that when the number of observations per subject

is frequent, the parameter estimates have relatively higher bias but smaller MSE in many occasions than when sparse observations are taken per subject. This observation was based on 10 and 5 observations per subject thus further investigation is needed since there is no much difference between 5 and 10 observations. Overall, the parameter estimates from our proposed model have small bias and MSE given different number of knots and observations per a subject.

3.7 Recommendation

The reported analyses have focused on simulation studies and the application on CSL data without considering factors associated with cervical dilation like maternal Body Mass Index (BMI). We shall extend the proposed semi-parametric model to accommodate covariates like BMI and estimate the mean labor curves for women in different BMI categories. We shall categorize the BMI groups as: normal (18.5 to 24.9 kg/m^2), overweight (25 to 29.9 kg/m^2) and obese (above 30 kg/m^2). The estimated mean labor curves will be compared with the observed cervical dilation curves as a way to determine the appropriateness of the proposed model in approximating cervical dilation curves. During the actual labor management, some variables are measured while the women are already admitted in the hospital. These variables are time varying covariates that need to be considered in the analysis. Therefore, we plan to extend the proposed model to accommodate time varying covariates in order to depict the true scenario in labor management. The current CSL data does not have time dependent covariates but this extension can be demonstrated using simulation studies.

Chapter 4

Dynamic Individualized Prediction in presence of Measurement Error in the time variable

4.1 Abstract

Predicting a woman's time to full cervical dilation is important in labor management for it may aid obstetricians to decide whether a woman should be allowed to dilate or booked for alternative procedures like Cesarean section. Such decisions are often based on the woman's labor patterns and their past measurements. A challenge arises since the recorded time points are based on when the woman was admitted to the hospital and not when the cervical dilation started. This project aims to extend Chapter 3 to a unified dynamic prediction approach where the remaining time to full dilation of a woman is predicted based on her past measurements. The prediction was implemented via a Monte Carlo method that takes into account the uncertainty in the population level estimates and in the variance components estimates. We used Receiver operating characteristic (ROC) curves to check the ability of our proposed method to discriminate between women who will achieve a full dilation within a given time frame and those who will not. The Area Under the Curve (AUC) was used to quantify the predictive ability of the method. The prediction ability was evaluated using a simulated data.

4.2 Introduction

In 1950s Friedman (1954) characterized the first stage of labor as composed of latent and active phases. The latent phase is the onset of regular uterine contraction and the active phase starts with a noticeable change in cervical dilation that ends with full dilation at 10 centimeters (cm). Friedman depicted abnormal labor progression in the active phase of labor as change in cervical dilation less than 1.2 cm per hour for nulliparous women and 1.5 cm per hour in multiparous women (Zhang et al., 2002; Vahratian et al., 2006; Zhang et al., 2010a). While the Friedman description of the labor curve gave guidance on the relationship between labor duration and cervical dilation, women have labor patterns that deviate from mean curve and obstetricians often monitor labor by taking measurements on cervical dilation periodically then use Friedman labor definition to make decision.

The management of labor has notably changed in contemporary practice and there is increased interventions by obstetricians like labor induction, oxycotin use and epidural analgesia which have the potential to alter normal labor progression. For example, cases of labor induction have increased from 9.0% in 1989 Vahratian et al. (2006) to 23.3% in 2012 (Osterman and Martin, June 2014). This calls for better statistical methods for studying labor progression and prediction of time to full dilation in contemporary practice.

In this Chapter, we extend the proposed model in Chapter 3 to dynamic individualized prediction of women's time to full dilation based on their past measurements. We fit the proposed model in Chapter 3 on a training data and implement predictions on a test data. To carry out prediction, we sample random effects of women in the test data from their posterior distribution given the estimated parameters and a small sample of their data. We use a rejection sampling algorithm to generate the random effects samples.

This Chapter is organized as follow. In Section 4.3, we present methodology and outline the prediction procedure including the sampling algorithm and method used in computing sensitivity and specificity for plotting ROC curves. In Section 4.4, we present ROC curves and AUC from a single simulated data. We conclude in Section 4.5 and in Section 4.6 we present recommendation for further study.

4.3 Methodology

4.3.1 Modeling framework

Let $\mathbf{D}_{im} = (\mathbf{t}_{im}, \mathbf{y}_{im})$ be the training set of the ith individual from the test dataset, $\mathbf{t}_{im} = (t_{i1}, t_{i2}, ..., t_{im})$ and $\mathbf{y}_{im} = (y_{i1}, y_{i2}, ..., y_{im})$ for i= 1,2,...n. For example, if a woman has 10 observations , we may decide to use the first 4 observations in prediction. Thus $\mathbf{t}_{im} = (t_{i1}, t_{i2}, t_{i3}, t_{i4})$ and $\mathbf{y}_{im} = (y_{i1}, y_{i2}, y_{i3}, y_{i4})$. Define the subject-specific random effects as $\mathbf{\Theta} = (\Delta, b_i, \boldsymbol{\lambda})$ where as explained before, Δ is a random shift time parameter for the ith woman, b_i is a random effect for the linear time slope and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_k)$ is a vector of subject-specific basis coefficients for the ith woman.

Assume $\widehat{\boldsymbol{\beta}}_{pred} = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\boldsymbol{\gamma}})$ are the population level parameter estimates obtained by fitting model (3.3.12) on the training data where $\widehat{\beta}_0$ is the estimated fixed intercept, $\widehat{\beta}_1$ the fixed effect for the linear time slope and $\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_1, \widehat{\gamma}_2, ..., \widehat{\gamma}_k)$ are basis coefficients common across subjects. Let $\widehat{\omega}_{pred} = c(\widehat{\sigma}^2_{\varepsilon}, \widehat{\sigma}^2_{\Delta}, \widehat{\sigma}^2_{b}, \widehat{\sigma}^2_{\lambda})$ be the estimated variance components, $\widehat{\sigma}^2_{\varepsilon}$ is the estimated variance of measurement error and the estimated variance structure of $\mathbf{\Theta}$, $\widehat{\Sigma}_{\Theta}$ is given as $\begin{bmatrix} \widehat{\sigma}^2_{\Delta} & 0 & \mathbf{0} \\ 0 & \widehat{\sigma}^2_{b} & \mathbf{0} \\ 0 & 0 & \widehat{\sigma}^2_{\lambda} I \end{bmatrix}$

Further let $f(\mathbf{D}_{im}; \widehat{\boldsymbol{\beta}}_{pred}, \widehat{\omega}_{pred})$ be the density function of $\mathbf{D}_{im}$ and $f(\mathbf{\Theta}|\widehat{\mathbf{\Sigma}}_{\mathbf{\Theta}})$ the estimated prior distribution of $\mathbf{\Theta}$ with a posterior distribution $h(\mathbf{\Theta}|\mathbf{D}_{im}; \widehat{\boldsymbol{\beta}}_{pred}, \widehat{\omega}_{pred}) \propto f(\mathbf{D}_{im}; \widehat{\boldsymbol{\beta}}_{pred}, \widehat{\omega}_{pred}) f(\mathbf{\Theta}|\widehat{\mathbf{\Sigma}}_{\mathbf{\Theta}})$.

4.3.2 Sampling Algorithm

We used rejection sampling procedure to generate M samples of $\boldsymbol{\Theta}$ using the following steps: (i) Fit model (3.3.12) on the training data to obtain the population level estimates $\widehat{\boldsymbol{\beta}}_{pred}$ and the variance components $\widehat{\omega}_{pred}$ (ii) For the ith subject at the rth iteration, sample $\boldsymbol{\Theta}$ from the posterior distribution $h(\boldsymbol{\Theta}|\mathbf{D}_{im}; \widehat{\boldsymbol{\beta}}_{pred}, \widehat{\omega}_{pred})$ by drawing $\boldsymbol{\Theta}^{(r)}$ from a candidate distribution $g(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}})$ where $g(.|\widehat{\Sigma}_{\Theta})$ is normal distribution with mean $\mathbf{0}$ and variance $\widehat{\Sigma}_{\Theta}$ and independently draw u from the uniform(0,1) distribution. (iii) Accept $\boldsymbol{\Theta}^{(r)}$ as a random sample from $h(\boldsymbol{\Theta}|\mathbf{D}_{im}; \widehat{\boldsymbol{\beta}}_{pred}, \widehat{\omega}_{pred})$ if

$$u \leq \frac{f(\mathbf{D}_{im}|\boldsymbol{\Theta}^{(r)}; \widehat{\boldsymbol{\beta}}_{pred}, \widehat{\omega}_{pred})}{c} \tag{4.3.1}$$

where $c = \sup_{\boldsymbol{\Theta}} \left\{ f(\mathbf{D}_{im}|\boldsymbol{\Theta}; \widehat{\boldsymbol{\beta}}_{pred}, \widehat{\omega}_{pred}) \right\}$. Otherwise reject and repeat step (i) to (iii) until M samples are drawn.

4.3.3 Receiver operating characteristic curves

In this section, we outline the approach used to check the discriminate ability of our proposed model. Specifically, we want to investigate how our model discriminates against women who will achieve full dilation within a given time period against those who will not.

Let T_i be the true time to 10 cm for the ith woman, t_{im} the last measurement of the individual training data $\mathbf{D}_{im}$ and L the length of the time period of interest (in hours). To determine women who will take more than L hours to attain full dilation, we compute an indicator variable X_i such that $X_i = \begin{cases} 0 \text{ if } (T_i - t_{im}) \leq L \\ 1 \text{ if } (T_i - t_{im}) > L \end{cases}$

Let $\widehat{\boldsymbol{\Theta}} = (\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, ..., \boldsymbol{\Theta}_M)$, l=1,2,...,M be the final generated sample for the ith subject. For each $\boldsymbol{\Theta}$, we predict the time to 10 cm as $\widehat{T}_l = argmin_t(E\{Y_l(t)\} - 10)^2$ where $E\{Y_l(t)\}$ is the predicted mean labor curve given $\boldsymbol{\Theta}_l$ and $\widehat{\boldsymbol{\beta}}_{pred}$.

The $E\{Y_l(t)\}$ is computed as

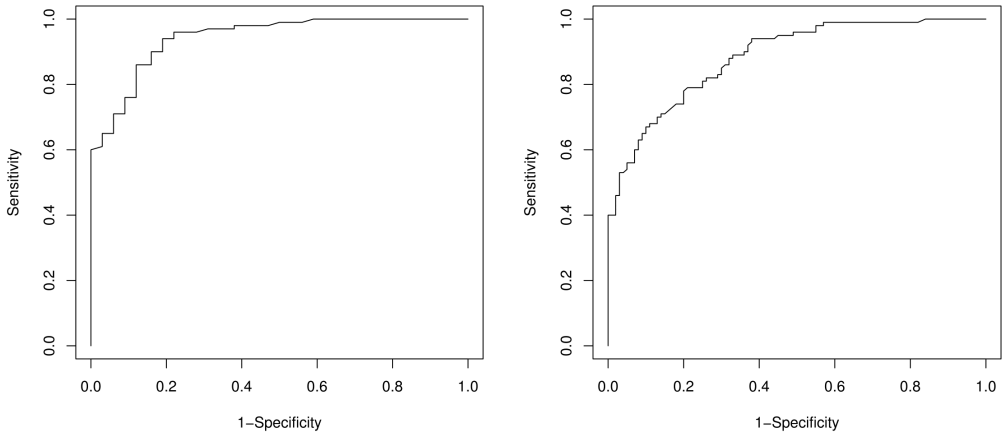
$$E\{Y_l(t)\} = \hat{\beta}_0 + \hat{\beta}_1(t-\Delta^{(l)}) + \sum_{k=1}^{K} \hat{\gamma}_k(t-\Delta^{(l)}-q_k)_+ + b^{(l)}(t-\Delta^{(l)}) + \sum_{k=1}^{K_c} \lambda_k^{(l)}(t-\Delta_{(l)}-q_k)_+ \tag{4.3.2}$$

Given the set of the predicted times to 10 cm, $\widehat{T}_i = c(\widehat{T}_1, \widehat{T}_2, ..., \widehat{T}_M)$, we identify subjects who will require more than L hours to attain full dilation based on the prediction model by defining another indicator variable $S_l = \begin{cases} 0 \text{ if } (\widehat{T}_l - t_{im}) \leq L \\ 1 \text{ if } (\widehat{T}_l - t_{im}) > L \end{cases}$ and then compute the probability $P_i = \frac{1}{M}\sum_{l=1}^{M} S_l$ that the ith woman will attain 10 cm after L hours. To create the ROC curves, we obtain the sensitivity (true positive rate) and specificity (true negative rate) as $\text{Se}_j = P(\widehat{P}_i > \widehat{P}_{(j)} | X_i = 1)$ and $\text{Sp}_j = P(\widehat{P}_i \leq \widehat{P}_{(j)} | X_i = 0)$ where $\widehat{P}_{(j)}$ is the jth ordered $\widehat{P}_i$

4.4 Application on a simulated data

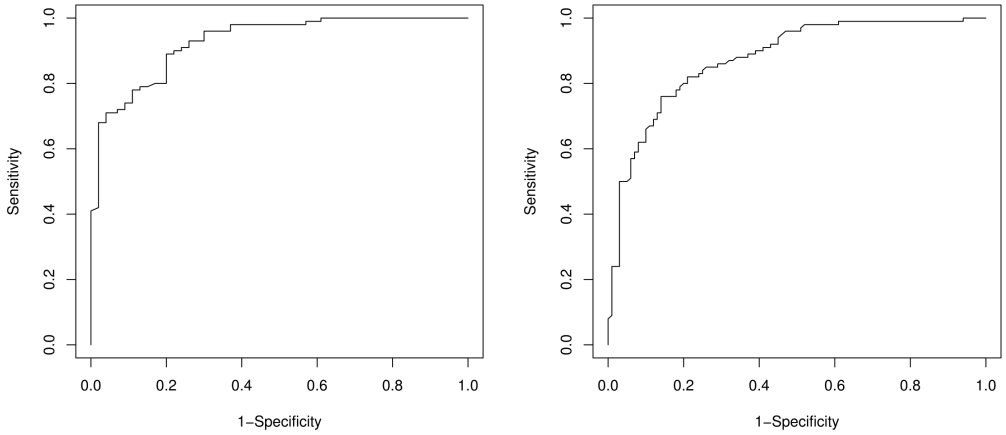
We simulated data with n=600 and used $N_1 = 300$ as training set and $N_2 = 300$ as the test set. We assumed ni=10 observations per subject, $K, K_c = (30,4)$ knots then generated M=1000 $\boldsymbol{\Theta}$ samples per subject. The simulated data was generated assuming the following true values: β_0 =1, β_1=1.5, σ^2_ε=0.5, σ^2_Δ=1, σ^2_γ=0.02, σ^2_λ=0.05 and σ^2_b=1. Using the procedure outlined in Section 4.3, we evaluated the predictive ability of our proposed model when the size of the individual training set is m =(4,7) and the length of the time of interest is $L = (0.5, 1, 2)$ hours.

From the analysis we observe that, the estimated AUC ranges from 0.84 to 0.94. When the targeted length of time L is 0.5 hours, the estimated AUC = (0.943,0.884), when L = 1 hour, AUC = (0.928,0.877) and when L = 2, AUC = (0.878,0.84) for m =4 and m =7 respectively. As the length of targeted period of time widens, so does the estimated AUC lessens. When m =4, we observe a higher AUC than when m = 7 regardless of the value of L. Figure 4.1 to 4.3 shows the estimated ROC curves for m =(4,7) given L =(0.5,1,2).



(a) AUC= 0.943 (b) AUC= 0.884

Figure 4.1 ROC curves when L=0.5 hours: Panel (a): $m = 4$ and Panel (b): $m = 7$ past measurements

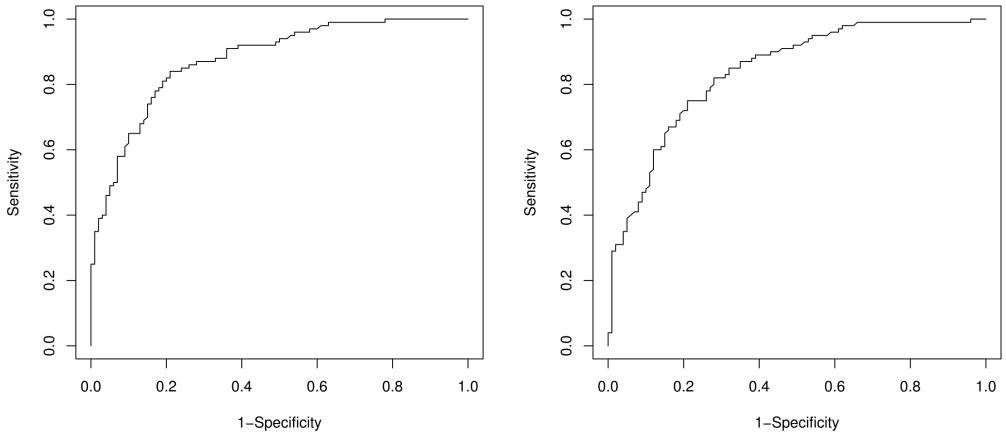


(a) AUC= 0.928 (b) AUC= 0.877

Figure 4.2 ROC curves when L=1 hours: Panel (a): $m = 4$ and Panel (b): $m = 7$ past measurements

4.5 Conclusion

The focus of this chapter was to evaluate the prediction ability of the proposed model in Chapter 3 using a simulated data. Based on AUC, it seems that the predictive ability of the proposed model diminishes as the size of individual training data in-



(a) AUC= 0.878 (b) AUC= 0.84

Figure 4.3 ROC curves when L=2 hours: Panel (a): $m = 4$ and Panel (b): $m = 7$ past measurements

creases. However, it is important to note that the subjects included in the prediction all attained 10 cm which might be a small subset. In addition, it is more likely that many subjects will have at least 4 measurements during the observation period but fewer subjects will have at least 7 measurements. Thus the sample size of the test data may decrease as the size of the individual training data increases. This might lead to poor predictive ability of the proposed model.

4.6 Recommendation

The reported results are based on a single simulated data thus more simulation studies are needed for better inference. The ultimate goal of this project was to predict the time to full dilation of women from CSL data and it will be important to carry out prediction on the CSL data for women who attained 10 cm and those who did not. We intend to apply the proposed model on CSL data adjusting for maternal Body Mass Index (BMI) and predict women's time to full dilation in each of the following BMI categories: normal: (18.5 to 24.9 kg/m2), overweight: (25 to 29.9 kg/m2) and obese: (above 30 kg/m2).

Chapter 5

Summary and Conclusion

This dissertation is about developing novel statistical methods for analyzing longitudinal data with measurement error on time variable. The methods have been demonstrated on a cervical dilation data from the Consortium of Safe Labor study. The dissertation constitutes of three projects: (1) Development of an efficient method for estimating the distribution of time it takes a woman's cervical dilation to progress from one threshold to another when the origin of the process is uncertain. (2) Development of a flexible nonlinear method for modeling cervical dilation curves while simultaneously adjusting for the uncertainty in the onset of labor. (3) An extension of project (2) to dynamic individualized prediction of women's time to full dilation given their past cervical dilation measurements.

There is increased interest in developing methods for characterizing labor patterns in contemporary obstetric practice. The challenges arises since the women are admitted at different stages of cervical dilation and the onset of labor is often unknown. In assessing the labor progression, obstetrician usually measure the cervical dilation periodically. This is important for it may help the obstetrician to decide whether a stalled labor should be allowed to dilate or opt for other options like C-section.

Interval censored models have been used before to estimate the distribution of time a cervical dilation takes to progress from 1 cm to another. However, the current methods are not efficient for they do not utilize all the available data and do not adjust for left truncation. There exists nonlinear models for estimating the labor curves and prediction of time to full dilation. Many of the current methods use the

time to full dilation as the benchmark time and the time is run backwards. These methods are only applied when all the women have attained a full dilation and are not efficient for data with censored labor curves. Recently a change point model with random effects has been proposed. The method corrects for the unknown onset of labor by re-scaling the observed time for each woman by an additive factor. However, the labor curves are known to be nonlinear and a change point model may not be a good approximation.

In this dissertation, I first developed a Longitudinal Threshold Regression model for estimating the distribution of time a woman's cervical dilation takes to progress from one threshold to another in cm. The model is developed under the First Hitting Time framework where the event of interest occurs when the process hits a pre-defined threshold for the first time. The proposed model utilizes the repeated measurements of the cervical dilation to estimate the time distribution. The underlying cervical dilation process is modeled as a Wiener process whose FHT has an inverse Gaussian distribution. The Wiener process has a Markov property that helps us to adjust for the unknown onset of labor by conditioning on a baseline observation.

The main goal of the proposed model is to estimate the distribution of the time it takes women to progress between cervical dilation "stations" at the population level. We have presented unadjusted survival curves for the time it takes women to traverse between 3 and 4 cm, along with 7 and 8 cm. These survival curves are to be interpreted on the population level of all women that would be observed within the interval of interest. Another method of estimation would be to fit a random-slope model, resulting in a conditional estimate of the survival function. A survival curve from such a model would be interpreted on the population of all women with a particular value for the random slope (i.e., the variance of the random slope $= 0$). This interpretation is not optimal since a woman's random effect level is unknown. The random effect could be integrated out to obtain an estimate of the *marginal survival*

function, however then one loses the parametric form of the survival function. Further, the resulting survival function can only be displayed graphically or in a table form as it cannot be given as a function of the parameters. As a result, in the proposed research using a fixed effect model has benefits over the random effect approach.

The proposed model assumes a linear trend over time, which will not always hold. For our motivating data, we consider 1 cm differences in cervical dilation where we believe the assumption of linearity is a reasonable approximation. If the interest lies in examining larger intervals, a model relaxing this assumption might be more appropriate. Typically we expect women's cervical dilation to progress differently over time due to inherent heterogeneity. Some authors Peng and Tseng (2009); Wang (2010) have modeled degradation data as a Wiener process with random effects to examine mean failure time given a single threshold. When interest lies in subject specific inferences a LTR model with random effects is appropriate. This type of approach will result in conditional inferences on the survival functions of interest. It was the goal of this research to provide population-level inference to use as a standard for labor progression that will be observed in the hospital. The survival function estimates, as well as the regression estimates for the effect of BMI on labor progression, can all be interpreted on the population-level of women that will be observed in the hospital within the range of interest. As we have shown in our robustness simulation studies under model misspecification (see Section 2.4.2) our method accurately estimates the marginal survival function from dependent data, however, a random effects model will be beneficial when prediction is of interest. In the future, we look to develop a method that can predict a woman's survival curve based on her past measurements via the random effects approach.

In the second project, I developed a semi-parametric model for estimating labor curves prospectively. The model adjusts for the unknown onset of labor by introducing random time shift parameters. The model is formulated in linear mixed effects

framework where the basis coefficients are model as random effects. This model has advantage over the current methods for it captures the nonlinear relationship of cervical dilation and time, it is easy to implement in standard statistical software and adjusts for the uncertainty in onset of labor thus correcting for the measurement error in the time variable. The random shift parameters are incorporated using the MCEM procedure where we used rejection sampling algorithm to draw the shift parameters at the E-step and the model parameters updated by optimizing the log-likelihood of the resulting augmented data in M-step. The proposed model introduces nonlinear functions in the population mean and in the subject-specific slopes. It would be important to evaluate the adequacy of the proposed model against a simpler model. For example, one can test whether the subject-specific slopes are linear functions as opposed to the proposed model that assumes that the slopes are nonlinear. Model diagnosis has not been carried out for the proposed model and in the future we plan to investigate more on goodness of fit test in semi-parametric models when MCEM method is used in estimation.

In the third project, I have extended the proposed model in project (2) to predict women's time to full dilation given their past measurements. The prediction is carried out using a Monte Carlo method that takes into account the uncertainty in the population level estimates and in the variance components estimates. The prediction is applied on a single simulated data. The prediction was done only for subjects who attained a full dilation. In the future we plan to do prediction on the real CSL data for both women who attained full dilation and those who did not.

Finally, the methods developed in this dissertation are not limited to cervical dilation data only. The methods can be used to analyze other longitudinal data with measurement error in time variable.

Bibliography

SD Aaron, T Ramsay, K Vandemheen, and GA Whitmore. A threshold regression model for recurrent exacerbations in chronic obstructive pulmonary disease. *Journal of clinical epidemiology*, 63(12):1324–1331, 2010.

Per Kragh Andersen, Ørnulf Borgan, Richard D. Gill, and Niels Keiding. *Statistical models based on counting processes*. Springer Series in Statistics. Springer-Verlag, New York, 1993. ISBN 0-387-97872-0.

Srikesh G Arunajadai. A nonlinear model for highly unbalanced repeated time-to-event data: Application to labor progression. *Statistics in medicine*, 29(26): 2709–2722, 2010.

Peter Bacchetti. Estimating the incubation period of aids by comparing population infection and diagnosis patterns. *Journal of the American Statistical Association*, 85(412):1002–1008, 1990.

AA Boligon, MEZ Mercadante, RB Lôbo, F Baldi, and Lucia Galvão de Albuquerque. Random regression analyses using b-spline functions to model growth of nellore cattle. *animal*, 6(02):212–220, 2012.

James G Booth and James P Hobert. Maximizing generalized linear mixed model likelihoods with an automated monte carlo em algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1):265–285, 1999.

Raj S Chhikara and J Leroy Folks. *The inverse Gaussian distribution: theory, methodology, and applications*, volume 95. CRC Press, 1989.

Norma Coffey, John Hinde, and Emma Holian. Clustering longitudinal profiles using p-splines and mixed effects models applied to time-course gene expression data. *Computational Statistics & Data Analysis*, 71:14–29, 2014.

Jessamyn Conell-Price, Jennifer B Evans, Daewha Hong, Steven Shafer, and Pamela Flood. The development and validation of a dynamic model to account for the

progress of labor in the assessment of pain. *Anesthesia & Analgesia*, 106(5): 1509–1515, 2008.

Brent A Coull, J Schwartz, and MP Wand. Respiratory health and air pollution: additive mixed model analyses. *Biostatistics*, 2(3):337–349, 2001.

DR Cox. Some sampling problems in technology, 1968.

C Crainiceanu, D Ruppert, and TJ Vogelsang. Probability that the mle of a variance component is zero with applications to likelihood ratio tests. *Unpublished manuscript*, 2002.

Iain D Currie and Maria Durban. Flexible smoothing with p-splines: a unified approach. *Statistical Modelling*, 2(4):333–349, 2002.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

María Durbán, Jaroslaw Harezlak, MP Wand, and RJ Carroll. Simple fitting of subject-specific curves for longitudinal data. *Statistics in medicine*, 24(8):1153–1167, 2005.

Paul HC Eilers and Brian D Marx. Flexible smoothing with b-splines and penalties. *Statistical science*, pages 89–102, 1996.

Angelo Elmi, Sarah J Ratcliffe, Samuel Parry, and Wensheng Guo. A b-spline based semiparametric nonlinear mixed effects model. *Journal of Computational and Graphical Statistics*, 20(2):492–509, 2011.

Angelo Elmi, Sarah J Ratcliffe, and Wensheng Guo. The estimation of branching curves in the presence of subject-specific random effects. *Statistics in medicine*, 2014.

Thomas S. Ferguson. *A course in large sample theory*. Texts in Statistical Science Series. Chapman & Hall, London, 1996. ISBN 0-412-04371-8.

E Friedman. The graphic analysis of labor. *American Journal of Obstetrics and Gynecology*, 68(6):1568, 1954.

Ahmed M Gad and Rasha B El Kholy. Generalized linear mixed models for longitudinal data. *International Journal of Probability and Statistics*, 1(3):41–47, 2012.

Bhaswati Ganguli, John Staudenmayer, and MP Wand. Additive models with predictors subject to measurement error. *Australian & New Zealand Journal of Statistics*, 47(2):193–202, 2005.

Jayanta K Ghosh, Mohan Delampady, and Tapas Samanta. *An introduction to Bayesian analysis: theory and methods*. Springer Science & Business Media, 2007.

David Golan and Saharon Rosset. Accurate estimation of heritability in genome wide studies using random effects models. *Bioinformatics*, 27(13):i317–i323, 2011.

Guadalupe Gómez and Stephen W Lagakos. Estimation of the infection time and latency distribution of aids with doubly censored data. *Biometrics*, pages 204–212, 1994.

Nima Gorjian, Lin Ma, Murthy Mittinty, Prasad Yarlagadda, and Yong Sun. *A review on degradation models in reliability analysis*, pages 369–384. Springer, 2010.

Margaret L Gourlay, Jason P Fine, John S Preisser, Ryan C May, Chenxi Li, Li-Yung Lui, David F Ransohoff, Jane A Cauley, and Kristine E Ensrud. Bone-density testing interval and transition to osteoporosis in older women. *New England Journal of Medicine*, 366(3):225–233, 2012.

Wensheng Guo. Functional mixed effects models. *Biometrics*, 58(1):121–128, 2002.

Lyle C Gurrin, Katrina J Scurrah, and Martin L Hazelton. Tutorial in biostatistics: spline smoothing with linear mixed models. *Statistics in medicine*, 24(21):3361–3381, 2005.

Charles R Henderson. Sire evaluation and genetic trends. *Journal of Animal Science*, 1973(Symposium):10–41, 1973.

Paul G Hoel, Sidney C Port, and Charles J Stone. *Introduction to stochastic processes*. Waveland Press, 1972.

Julie Horrocks and Mary E Thompson. Modeling event times with multiple outcomes using the wiener process with drift. *Lifetime Data Analysis*, 10(1):29–49, 2004.

Nicholas P Jewell. Non-parametric estimation and doubly-censored data: General ideas and applications to aids. *Statistics in medicine*, 13(19-20):2081–2095, 1994.

Nicholas P Jewell, Hina M Malani, and Eric Vittinghoff. Nonparametric estimation for a form of doubly censored data, with application to two problems in aids. *Journal of the American Statistical Association*, 89(425):7–18, 1994.

Le Kang, Randy Carter, Kathleen Darcy, James Kauderer, and Shu-Yuan Liao. A fast monte carlo expectation–maximization algorithm for estimation in latent class model analysis with an application to assess diagnostic accuracy for cervical neoplasia in women with atypical glandular cells. *Journal of applied statistics*, 40 (12):2699–2719, 2013.

Göran Kauermann and Jean D Opsomer. Data-driven selection of the spline dimension in penalized spline regression. *Biometrika*, 98(1):225–230, 2011.

Inyoung Kim, Noah D Cohen, and Raymond J Carroll. Semiparametric regression splines in matched case-control studies. *Biometrics*, 59(4):1158–1169, 2003.

Michelle A Kominiarek, Jun Zhang, Paul VanVeldhuisen, James Troendle, Julie Beaver, and Judith U Hibbard. Contemporary labor patterns: the impact of maternal body mass index. *American journal of obstetrics and gynecology*, 205 (3):244–e1, 2011.

S Katherine Laughon, D Branch, Julie Beaver, and Jun Zhang. Changes in labor patterns over 50 years. *American journal of obstetrics and gynecology*, 206(5): 419–e1, 2012.

Mei-Ling Ting Lee and GA Whitmore. Threshold regression for survival analysis: modeling event times by a stochastic process reaching a boundary. *Statistical Science*, pages 501–513, 2006.

Mei-Ling Ting Lee, GA Whitmore, Francine Laden, Jaime E Hart, and Eric Garshick. A case–control study relating railroad worker mortality to diesel exhaust exposure using a threshold regression model. *Journal of statistical planning and inference*, 139(5):1633–1642, 2009.

Mei-Ling Ting Lee, G. A. Whitmore, and Bernard A. Rosner. Threshold regression for survival data with time-varying covariates. *Statistics in Medicine*, 29(7-8): 896–905, 2010. ISSN 1097-0258.

Richard A Levine and George Casella. Implementations of the monte carlo em algorithm. *Journal of Computational and Graphical Statistics*, 10(3):422–439, 2001.

Hua Liang, Hulin Wu, and Raymond J Carroll. The relationship between virologic and immunologic responses in aids clinical research using mixed-effects varying-coefficient models with measurement error. *Biostatistics*, 4(2):297–312, 2003.

I-Chao Liu, Ronghui Xu, Deborah L Blacker, Garrett Fitzmaurice, Michael J Lyons, and Ming T Tsuang. The application of a random effects model to censored twin data. *Behavior genetics*, 35(6):781–789, 2005.

C Joseph Lu and William O Meeker. Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, 35(2):161–174, 1993.

C Joseph Lu, William Q Meeker, and Luis A Escobar. A comparison of degradation and failure-time analysis methods for estimating a time-to-failure distribution. *Statistica Sinica*, 6(3):531–546, 1996.

Charles E McCulloch. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American statistical Association*, 92(437):162–170, 1997.

Alexander C McLain and Paul S Albert. Modeling longitudinal data with a random change point and no time-zero: Applications to inference and prediction of the labor curve. *Biometrics*, 70(4):1052–1060, 2014.

William Q Meeker, Luis A Escobar, and C Joseph Lu. Accelerated degradation tests: modeling and analysis. *Technometrics*, 40(2):89–99, 1998.

Karin Meyer. Random regression analyses using b-splines to model growth of australian angus cattle. *Genetics Selection Evolution*, 37(5):473–500, 2005.

Geert Molenberghs and Geert Verbeke. Likelihood ratio, score, and wald tests in a constrained parameter space. *The American Statistician*, 61(1):22–27, 2007.

Ronald C Neath et al. On convergence properties of the monte carlo em algorithm. In *Advances in Modern Statistical Theory and Applications: A Festschrift in Honor of Morris L. Eaton*, pages 43–62. Institute of Mathematical Statistics, 2013.

Michelle J.K. Osterman and Joyce A. Martin. Recent declines in induction of labor by gestational age. (155), June 2014.

H Desmond Patterson and Robin Thompson. Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3):545–554, 1971.

Chien-Yu Peng and Sheng-Tsaing Tseng. Mis-specification analysis of linear degradation models. *Reliability, IEEE Transactions on*, 58(3):444–455, 2009.

Michael L Pennell, GA Whitmore, and Mei-Ling Ting Lee. Bayesian random-effects threshold regression with application to survival data with nonproportional hazards. *Biostatistics*, page kxp041, 2009.

J Pinheiro, D Bates, S DebRoy, D Sarkar, et al. The nlme package, 2009.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL `http://www.R-project.org/`.

John A Rice and Colin O Wu. Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, 57(1):253–259, 2001.

George K Robinson. That blup is a good thing: the estimation of random effects. *Statistical science*, pages 15–32, 1991.

David Ruppert. Selecting the number of knots for penalized splines. *Journal of computational and graphical statistics*, 2002.

Joel Schwartz and Brent A Coull. Control for confounding in the presence of measurement error in hierarchical models. *Biostatistics*, 4(4):539–553, 2003.

Steven G Self and Kung-Yee Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610, 1987.

Daniel O Stram and Jae Won Lee. Variance components testing in the longitudinal mixed effects model. *Biometrics*, pages 1171–1177, 1994.

Jianguo Sun. Empirical estimation of a distribution function with truncated and doubly interval-censored data and its application to aids studies. *Biometrics*, pages 1096–1104, 1995.

Jianguo Sun. Self-consistency estimation of distributions based on truncated and doubly censored survival data with applications to aids cohort studies. *Lifetime Data Analysis*, 3(4):305–313, 1997.

Jianguo Sun. Statistical analysis of doubly interval-censored failure time data. *Handbook of Statistics–23: Advances in Survival Analysis*, pages 105–122, 2004.

Mei-Ling Ting Lee, GA Whitmore, Francine Laden, Jaime E Hart, and Eric Garshick. Assessing lung cancer risk in railroad workers using a first hitting time regression model. *Environmetrics*, 15(5):501–512, 2004.

Tor D Tosteson, John P Buonaccorsi, and Eugene Demidenko. Covariate measurement error and the estimation of random effect parameters in a mixed model for longitudinal data. *Statistics in Medicine*, 17(17):1959–1971, 1998.

Xin Ming Tu. Nonparametric estimation of survival distributions with censored initiating time, and censored and truncated terminating time: application to transfusion data for acquired immune deficiency syndrome. *Applied statistics*, pages 3–16, 1995.

Maurice CK Tweedie. Statistical properties of inverse gaussian distributions. i. *The Annals of Mathematical Statistics*, pages 362–377, 1957.

Anjel Vahratian, Jun Zhang, James F Troendle, David A Savitz, and Anna Maria Siega-Riz. Maternal prepregnancy overweight and obesity and the pattern of labor progression in term nulliparous women. *Obstetrics & Gynecology*, 104(5, Part 1):943–951, 2004.

Anjel Vahratian, James F Troendle, Anna Maria Siega-Riz, and Jun Zhang. Methodological challenges in studying labour progression in contemporary practice. *Paediatric and perinatal epidemiology*, 20(1):72–78, 2006.

Arũnas P Verbyla, Brian R Cullis, Michael G Kenward, and Sue J Welham. The analysis of designed experiments and longitudinal data by using smoothing splines. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48(3):269–311, 1999.

Matt P Wand. Smoothing and mixed models. *Computational statistics*, 18(2):223–249, 2003.

Mei-Cheng Wang. A semiparametric model for randomly truncated data. *Journal of the American Statistical Association*, 84(407):742–748, 1989. ISSN 0162-1459.

Naisyin Wang and Marie Davidian. A note on covariate measurement error in nonlinear mixed effects models. *Biometrika*, 83(4):801–812, 1996.

Xiao Wang. Wiener processes with random effects for degradation data. *Journal of Multivariate Analysis*, 101(2):340–351, 2010.

Yuedong Wang, Chunlei Ke, and Morton B Brown. Shape-invariant modeling of circadian rhythms with random effects and smoothing spline anova decompositions. *Biometrics*, 59(4):804–812, 2003.

Greg CG Wei and Martin A Tanner. A monte carlo implementation of the em algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, 85(411):699–704, 1990.

Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25, 1982.

GA Whitmore and Yi Su. Modeling low birth weights using threshold regression: results for us birth data. *Lifetime data analysis*, 13(2):161–190, 2007.

Z-S Ye, Yu Wang, K-L Tsui, and Michael Pecht. Degradation data analysis using wiener processes with measurement errors. 2012.

Scott L Zeger and Peter J Diggle. Semiparametric models for longitudinal data with application to cd4 cell numbers in hiv sero converters. *Biometrics*, pages 689–699, 1994.

Daowen Zhang, Xihong Lin, Jonathan Raz, and MaryFran Sowers. Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, 93(442):710–719, 1998.

Jun Zhang, James F Troendle, and Michael K Yancey. Reassessing the labor curve in nulliparous women. *American Journal of Obstetrics and Gynecology*, 187(4): 824–828, 2002.

Jun Zhang, Helain J Landy, D Ware Branch, Ronald Burkman, Shoshana Haberman, Kimberly D Gregory, Christos G Hatjis, Mildred M Ramirez, Jennifer L Bailit, Victor H Gonzalez-Quintero, et al. Contemporary patterns of spontaneous labor with normal neonatal outcomes. *Obstetrics and gynecology*, 116(6):1281, 2010a.

Jun Zhang, James Troendle, Uma M Reddy, S Katherine Laughon, D Branch, Ronald Burkman, Helain J Landy, Judith U Hibbard, Shoshana Haberman, Mildred M Ramirez, et al. Contemporary cesarean delivery practice in the united states. *American journal of obstetrics and gynecology*, 203(4):326–e1, 2010b.