Title
A Joint Marginal-Conditional For Multivariate Longitudinal Data and A Cure-Rate Model For Left-Truncated and Right-Censored Data /

Permalink
https://escholarship.org/uc/item/4206s40t

Author
Faig, Walter G.

Publication Date
2013

Peer reviewed|Thesis/dissertation
A Joint Marginal-Conditional For Multivariate Longitudinal Data
and
A Cure-Rate Model For Left-Truncated and Right-Censored Data

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Mathematics

by

Walter G. Faig

Committee in charge:

Ronghui Xu, Chair
Ian Abramson
Ery Arias-Castro
Christina Chambers
Michael Donohue
Loki Natarajan

2013
The dissertation of Walter G. Faig is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2013
# TABLE OF CONTENTS

Signature Page ......................................................... iii
Table of Contents ....................................................... iv
List of Figures ........................................................ vi
List of Tables .......................................................... vii
Acknowledgements ...................................................... viii
Vita ........................................................................... ix
Abstract of the Dissertation ............................................ x

Chapter 1 A Joint Marginal-Conditional Model For Multivariate Longitudinal Data .............................................. 1
1.1 Introduction ........................................................... 1
1.2 Model and Estimation ................................................. 3
    1.2.1 The model ...................................................... 3
    1.2.2 Estimating equations and the ES algorithm ............. 4
1.3 Asymptotic Properties ............................................... 7
1.4 Simulation Experiments ............................................ 8
1.5 Application: Nutrition Data ....................................... 11
1.6 Discussion ............................................................ 13

Chapter 2 A Cure-Rate Model For Left-Truncated and Right-Censored Data .................................................. 24
2.1 Introduction ........................................................... 24
2.2 Model and Estimation ................................................. 26
    2.2.1 The Model ...................................................... 26
    2.2.2 Estimation of $G$ ............................................ 27
    2.2.3 ES Algorithm ............................................... 30
2.3 Asymptotic Properties ............................................... 32
    2.3.1 Multiple Imputation Variance ............................... 35
2.4 Simulation Study .................................................... 35
    2.4.1 Simulating Cure-Rate Data ................................ 35
    2.4.2 Simulation Results .......................................... 37
2.5 Application: Pregnancy Outcomes Data ......................... 38
2.6 Discussion ............................................................ 41

Appendix A Longitudinal Model Notes ................................ 50
A.1 Proof of Theorem 1 .................................................. 50
LIST OF FIGURES

Figure 1.1: Boxplot of estimated standard error of $\beta_{11}$ from each of the 500 simulation runs; the dashed line represents the sample SD. .......................... 22

Figure 1.2: Plots of 30 randomly selected individuals (dashed lines) over time; left: phone counseling group, right: control group. The black solid line is the fitted mean. ...................................................... 23

Figure 2.1: Kaplan-Meier curves for the full sample data and just the non-cured individuals. ................................................................. 47

Figure 2.2: Truncation times for all individuals and just the cured individuals. 48

Figure 2.3: Kaplan-Meier estimates for truncation and residual times for the non-cured sample. ................................................................. 49
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Simulation results for two correlated outcomes, with a random intercept on each outcome. The errors were normally distributed.</td>
<td>15</td>
</tr>
<tr>
<td>1.2</td>
<td>Simulation results for two correlated outcomes, with a random intercept on each outcome. The errors were non-normally distributed.</td>
<td>16</td>
</tr>
<tr>
<td>1.3</td>
<td>Simulation results for two correlated outcomes, with a random intercept and a random slope on each outcome.</td>
<td>17</td>
</tr>
<tr>
<td>1.4</td>
<td>Correlation among the three outcomes over time</td>
<td>18</td>
</tr>
<tr>
<td>1.5</td>
<td>Comparison of two intervention groups while adjusting for baseline covariates</td>
<td>19</td>
</tr>
<tr>
<td>1.6</td>
<td>Time trajectory from baseline to 1 and 4 years (control group)</td>
<td>20</td>
</tr>
<tr>
<td>1.7</td>
<td>Time trajectory from baseline to 1 and 4 years (phone counseling group)</td>
<td>21</td>
</tr>
<tr>
<td>2.1</td>
<td>N = 200. 50% Cured. Weighted vs. Unweighted Estimates</td>
<td>43</td>
</tr>
<tr>
<td>2.2</td>
<td>75% Cured. Small vs. Large Sample Size</td>
<td>44</td>
</tr>
<tr>
<td>2.3</td>
<td>Estimates for SAB data with cure-rate model</td>
<td>45</td>
</tr>
<tr>
<td>2.4</td>
<td>Estimates for SAB data with separate logistic and proportional hazards regressions.</td>
<td>46</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would like to acknowledge Professor Ronghui Xu for her support as the chair of my committee.

Chapter 1, in full, is currently being prepared for submission for publication of the material. Faig, Walter; Xu, Ronghui; Natarajan, Loki. The dissertation author was the primary investigator and author of this material.

Chapter 2, in part, is currently being prepared for submission for publication of the material. Faig, Walter; Xu, Ronghui; Chambers, Christina. The dissertation author was the primary investigator and author of this material.
<table>
<thead>
<tr>
<th>Year</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>B.S. in Mathematics with high distinction; B.S. in Physics, University of Illinois, Champaign-Urbana</td>
</tr>
<tr>
<td>2007-2013</td>
<td>Graduate Teaching Assistant, University of California, San Diego</td>
</tr>
<tr>
<td>2009</td>
<td>M.A. in Mathematics, University of California, San Diego</td>
</tr>
<tr>
<td>2011</td>
<td>Data Analyst, Department of Pediatrics, University of California, San Diego</td>
</tr>
<tr>
<td>2013</td>
<td>Ph.D. in Mathematics, University of California, San Diego</td>
</tr>
</tbody>
</table>
ABSTRACT OF THE DISSERTATION

A Joint Marginal-Conditional For Multivariate Longitudinal Data and A Cure-Rate Model For Left-Truncated and Right-Censored Data

by

Walter G. Faig

Doctor of Philosophy in Mathematics

University of California, San Diego, 2013

Ronghui Xu, Chair

Multivariate longitudinal data frequently arise in biomedical applications, however their analysis, especially when adjusting for covariates, is often carried out one outcome at a time, or jointly using existing software in an adhoc fashion. A main challenge in the proper analysis of such data is the fact that the different outcomes are measured on different unknown scales. Methodology for handling the scale problem has been previously proposed for cross-sectional data, and here we extend it to the longitudinal setting. We consider modeling the longitudinal data using random effects, while leaving the joint distribution of the multiple outcomes unspecified.
We propose an estimating equation together with an EM-type (ES) algorithm. The consistency and the asymptotic distribution of the parameter estimates are established. The method is evaluated using simulations, and applied to a longitudinal nutrition data set from a large dietary intervention trial on breast cancer survivors, the Women’s Healthy Eating and Living (WHEL) Study.

In some medical studies, a portion of the population that is being sampled may not be susceptible to the event. Cure-rate models are used to analyze the survival regression as well as the proportion cured individuals in the population. Analysis with these models has been limited to data featuring right-censoring only, however, for some data, such as pregnancy outcomes, the time to incident should not be measured from entry into the study. In these cases, the time of entry is a truncation time, and we propose a cure-rate model that accounts for this left-truncation. The primary challenge in developing a model for such data is that left-truncation incurs biased estimates of the cure-rate. To correct for this, we incorporate inverse probability weights (IPW) based on the estimated distribution of entry times into an augmented form of the complete data likelihood. From this estimating equations are derived, and we propose parameter estimation through an EM-type (ES) algorithm. The consistency and asymptotic distribution of the parameter estimates are established. The approach is illustrated through simulated data examples as well as pregnancy data from the Organization of Teratology Information Specialists (OTIS) where we consider the outcome spontaneous abortion (SAB).
Chapter 1

A Joint Marginal-Conditional Model For Multivariate Longitudinal Data

1.1 Introduction

Multivariate outcome data are increasingly common in biomedical research; for example, in toxicology studies (Sammel and Ryan, 1996), in Alzheimer’s disease (Beckett et al., 2004), and in lifestyle intervention trials targeting multiple health behaviors. As a specific example, the Women’s Healthy Eating and Living (WHEL) Study was a randomized dietary intervention trial conducted in early stage breast cancer survivors between 1996 and 2006 (Pierce et al., 2002, 2007). Rather than target a specific nutrient, the study aimed to improve overall dietary pattern by teaching intervention participants to consume a high fiber, low fat diet emphasizing vegetables and fruits rich in carotenoids and other potentially protective phytochemicals. Detailed information on dietary intake of fruits, vegetables, fiber, fat, and other food
components were collected at study entry, and 1 and 4-years post-randomization. A major goal of the study was to examine concommitant changes over time in the multivariate outcome comprising vegetable, fiber, and fat intake, which were key targets of the behavioral intervention.

Like the WHEL study, it is increasingly common to collect multivariate data repeatedly over time. While statistical methods for analyzing longitudinal data have been well studied and broadly applied (Diggle et al., 2002; Fitzmaurice et al., 2004), the analysis of multiple outcomes longitudinal data is usually carried out by analyzing one outcome at a time. However, the scientific question of interest typically concerns the multiple outcomes as a whole, as studied in Sammel et al. (1999). While a primary comparison in a randomized study may compare the multiple outcomes at a final time point using tests like Hotelling’s $T$, it is often also of interest to model the longitudinal data simultaneously for all the outcomes. In addition, in observational studies such as birth defects, it is imperative to adjust for confounders as covariates. In this paper, we propose a general regression framework for the multivariate longitudinal data that have useful interpretations in practice.

To be more specific, the correlation among the different outcomes is typically considered a nuisance, or of secondary interest. In addition, parameter estimates can be sensitive to misspecification of the correlation structure if we choose to explicitly model that (Sammel and Ryan, 2002). Because of these considerations we propose to model the different outcomes marginally, and use an approach that is similar to the generalized estimating equations (Liang and Zeger, 1986). On the other hand, practitioners are often used to random effects modeling of the longitudinal data, which gives the interpretation of random intercepts and slopes, etc. Therefore we propose to combine the marginal and the conditional (random effects) approaches: marginal for the multivariate aspect, and conditional for the longitudinal aspect.
Inference under such a combined modeling approach carries the same combined characteristics. Since the data generating mechanism is not completely specified, likelihood based approaches are not possible. Instead estimating equations are well studied under the marginal model, albeit here with the addition of the unobserved random effects; more details are given in the following sections. This turns out to be the case of the estimating equations with missing data problem considered in Elashoff and Ryan (2004), for which they developed an expectation-substitution (ES) algorithm. The ES algorithm generalizes the expectation-maximization (EM) algorithm which is applied to the log-likelihood function. We show that under our proposed model, the estimating equations together with the ES algorithm give consistent and asymptotically normally distributed estimates of the parameters.

In the next section we describe the proposed model and inference procedure. We derive the asymptotic properties of the estimator. Finite sample properties are studied through simulations in Section 3, including its performance under misspecified correlation structures, as well as non-normally distributed errors. We apply the approach to the WHEL nutrition data, and conclude with some discussion for future work.

1.2 Model and Estimation

1.2.1 The model

Let \( y_{ijt} \) represent the value of outcome \( j \) measured on subject \( i \) at time \( t \) \((j = 1, 2, \ldots, J, t = 1, \ldots, n_i, i = 1, \ldots, N)\). We assume that the \( y_{ijt} \)'s are correlated across outcomes and over time. Let \( x_{ijt} \) be a \( p \times 1 \) vector of covariates for subject \( i \) at time \( t \) and outcome \( j \), including possibly an ‘1’ for intercept. As in many applications, we would like to model the longitudinal data using random effects
models with possibly random intercepts and slopes, etc., but would like to leave the correlation among the different outcomes unspecified. So let

\[ y_{ijt} = x_{ijt}^\top \beta_j + z_{ijt}^\top b_{ij} + e_{ijt}, \quad (1.1) \]

where \( \beta_j \) is a \( p \times 1 \) vector of fixed effects, \( b_{ij} \) is a \( q \times 1 \) vector of random effects for individual \( i \) and outcome \( j \), and the corresponding covariates with random effects are denoted by \( z_{ijt} \). We let \( \text{Var}(e_{ijt}) = \sigma_j^2 \) for outcome \( j \), \( E(b_{ij}) = 0 \) and \( \text{Var}(b_{ij}) = D_j \).

For a given \( j \), model (1.1) is a Laird and Ware (1982) type model for longitudinal data, although we do not assume normality of the random effects or errors. For the multiple outcomes, a challenge in fitting model (1.1) is the unknown different variances \( \sigma_j^2 \), sometimes referred to as scales (Roy et al., 2003). A naive approach in some applications has been to scale each outcome by the sample variance of the \( y \)'s for that outcome, and then fit a general linear mixed-effects model with common error variance and certain correlation structure (Cnaan et al., 1997), which can be done using existing software such as the \textit{lme4} package in R. It is clear, however, that the quantities which ought to be used to scale the outcomes are the \( \sigma_j \)'s, not the variances of \( y \).

### 1.2.2 Estimating equations and the ES algorithm

Under model (1.1) denote \( \theta \) the vector of parameters consisting of \( \beta = (\beta_1^\top, \beta_2^\top, ..., \beta_J^\top)^\top \), the unique entries of the \( D_j \)'s, and the \( \sigma_j \)'s. Also denote \( y_{it}^* = (y_{i1t}/\sigma_1, ..., y_{iJt}/\sigma_J)^\top \), \( b_i = (b_{i1}^\top, ..., b_{iJ}^\top)^\top \), \( X_{it} = (x_{i1t}, ..., x_{iJt})^\top \), and \( Z_{it} = (z_{i1t}, ..., z_{iJt})^\top \).

For notational purposes only, consider the scaled version of (1.1):

\[ \frac{y_{ijt}}{\sigma_j} = x_{ijt}^\top \tilde{\beta}_j + z_{ijt}^\top \tilde{b}_{ij} + \tilde{e}_{ijt}, \quad (1.2) \]

where \( \tilde{\beta}_j \), \( \tilde{b}_{ij} \), and \( \tilde{e}_{ijt} \) here are the scaled counterparts of \( \beta_j \), \( b_{ij} \), and \( e_{ijt} \) in (1.1).
To estimate \( \theta \), we first consider conditioning on the random effects. Based on the general estimating equations (GEE) approach, we have

\[
\sum_{i=1}^{N} U_i^{(1)}(\theta) = \sum_{i=1}^{N} \sum_{t=1}^{n_i} X_{it}^T R^{-1} \left( y_{it} - X_{it} \tilde{\beta} - Z_{it} \tilde{b}_i \right) = 0, \tag{1.3}
\]

\[
\sum_{i=1}^{N} U_i^{(2)}(\theta) = \sum_{i=1}^{N} \left[ \sum_{t=1}^{n_i} \left\{ \frac{y_{ijt}}{\sigma_j} \left( \frac{y_{ijt}}{\sigma_j} - x_{ijt}^T \tilde{\beta}_j - z_{ijt}^T \tilde{b}_{ij} \right) - 1 \right\} \right]_{j=1,...,J} = 0, \tag{1.4}
\]

\[
\sum_{i=1}^{N} U_i^{(3)}(\theta) = \sum_{i=1}^{N} \left[ b_{ij} b_{ij}^T - D_j \right]_{j=1,...,J} = 0, \tag{1.5}
\]

where \([a_j]_{j=1,...,J} = (a_1, ..., a_J)^T\). In the above \( R \) is the working correlation matrix of the errors, just like in the GEE. The first set of estimating equations (1.3) are unbiased regardless of whether \( R \) is misspecified; note that \( E\{\sum_{i=1}^{N} U_i^{(1)}(\theta)|b_1, ..., b_N\} = 0 \) implies \( E\{\sum_{i=1}^{N} U_i^{(1)}(\theta)\} = 0 \). In addition, (1.3) is known to be efficient when \( R \) is correctly specified and there are no random effects. The second set of equations (1.4) are method-of-moments equations for \( \sigma_j^2 \). Similar equations as (1.3) and (1.4) were also used in Roy et al. (2003). The third set (1.5) are method-of-moments equations for the \( D_j \)'s, noting that the \( b_{ij} \)'s are assumed to have mean zero.

Elashoff and Ryan (2004) developed an ES algorithm for general estimating equations with missing data, in a way similar to the EM algorithm for maximum likelihood estimation. Typically the estimating equations are derived for complete data or, in other words, conditional on the missing data. The ES approach takes the conditional expectation of the complete data estimating equation given the observed data, and iterates between an E-step and a S-step, where the expected values of the missing data sufficient statistics under current parameter values are substituted in to update the unknown parameters. The approach is very similar to the EM algorithm, except that it operates on the estimating equations instead of the log-likelihoods. In the following we adapt the ES approach to our equations above.

Denote \( \mathcal{U}_i(\theta) = (U_i^{(1)}(\theta)^T, U_i^{(2)}(\theta)^T, U_i^{(3)}(\theta)^T)^T \), then we have \( \sum_{i=1}^{N} \mathcal{U}_i(\theta) = 0 \).
To take the conditional expectations of $U_i(\theta)$ given $y_i$, where $y_i$ is the vector of outcome $y_{ijt}$’s for individual $i$, we need $E(b_i|y_i)$ and $E(b_i b_i^T|y_i)$. The ES approach allows separate modeling of the missing data mechanism which, in our case, is the modeling of these two conditional expectations. Motivated by the conditional moments under the multivariate normal distribution, we consider

$$E(b_i|y_i) = (DZ_{it}^T)(Z_{it}DZ_{it}^T + R)^{-1}(y_{it}^* - X_{it}\beta),$$
$$\text{Var}(b_i|y_i) = D - (DZ_{it}^T)(Z_{it}DZ_{it}^T + R)^{-1}(DZ_{it}^T)^T,$$

where $D = \text{Cov}(b_i) = \text{diag}(D_1, ..., D_J)$ is a $Jq \times Jq$ matrix. We emphasize that these are ‘working’ assumptions; we only assume the moments condition and do not assume normality. In the simulations we will examine the above inference procedure for non-normal data. It is clear that $E(b_i b_i^T|y_i) = \text{Var}(b_i|y_i) + E(b_i|y_i)E(b_i|y_i)^T$.

To carry out the ES algorithm, one provides the initial values for all the parameters. We found it advantageous to use the estimate of $\sigma_j$ by fitting a linear mixed model as in (1.1) to outcome $j$ alone as its initial value. In our algorithm, values of $\theta$ are updated by substituting $b_{ij}$ and $b_{ij}^T b_{ij}$ in $\sum_{i=1}^N U_i(\theta)$ with their conditional expectations according to (1.6), where the expectations are evaluated under the current parameter values. In the S-step the first set of equations (1.3) can be solved using existing GEE software, while the latter two sets of equations can be solved algebraically. The convergence of the ES algorithm was discussed in Elashoff and Ryan (2004), with references to general theory on iterative solutions to nonlinear equations by Ortega and Rheinboldt (1970) and Ortega (1972), and is not pursued further here.
1.3 Asymptotic Properties

To study the asymptotic properties of \( \hat{\theta} \) obtained using the ES approach, it is important to recognize that at convergence of the algorithm \( \hat{\theta} \) solves the unconditional estimating equations \( \sum_{i=1}^{N} U_i(\theta) = 0 \), where \( U_i = E_{\theta}\{U_i(\theta)|y_i\} \). Given the closed form expressions of the conditional expectations in (1.6), one can explicitly write down the expressions for \( U_i \), and these are given in the Appendix.

In the following we show that under general regularity conditions the set of estimating equations \( \sum_{i=1}^{N} U_i(\theta) = 0 \) has a solution \( \hat{\theta} \) that is a consistent estimator for the true \( \theta_0 \) that generates the data under model (1.1).

Denote \( X_i \) and \( Z_i \) the stacked matrices of \( X_{it} \)'s and \( Z_{it} \)'s that correspond to \( y_i \). From the expression in the Appendix it is clear that \( U_i(\theta) \) is continuously differentiable with respect to \( \theta \); denote the first order Jacobian matrix \( U_i'(\theta) = \partial U_i(\theta)/\partial \theta \). We make the following assumptions:

\begin{enumerate}
  \item The triples \( (y_i, X_i, Z_i) \), \( i = 1, ..., N \), are independent and identically distributed with finite first and second moments, and satisfy model (1.1);
  \item There is a neighborhood of \( \theta_0 \) on which \( \sum_{i=1}^{N} U_i'(\theta)/N \) converges uniformly to \( E\{U_i'(\theta)\} \) in probability as \( N \to \infty \);
  \item \( E\{U_i'(\theta_0)\} \) is non-singular.
\end{enumerate}

We have the following.

**Theorem 1.** Under conditions i. – iii. above, \( \sum_{i=1}^{N} U_i(\theta) = 0 \) has a solution \( \hat{\theta} \) that is consistent for the true parameter vector \( \theta_0 \).

See the Appendix for proof.
Theorem 2. Under conditions i. – iii. above, $\sqrt{N}(\hat{\theta} - \theta_0)$ is asymptotically normally distributed with mean zero and variance that can be consistently estimated by:

$$\left\{ \frac{1}{N} \sum_{i=1}^{N} U'_i(\hat{\theta}) \right\}^{-1} \left\{ \frac{1}{N} \sum_{i=1}^{N} U_i(\hat{\theta})U_i(\hat{\theta})^\top \right\} \left\{ \frac{1}{N} \sum_{i=1}^{N} U'_i(\hat{\theta}) \right\}^{-1}.$$  

The proof of Theorem 2 is standard for estimating equations, using Taylor expansion and the central limit theorem, and can be found in for example, Yuan and Jennrich (1998).

Although the closed form of $U'_i(\theta)$ can be derived algebraically, in practice the computation is cumbersome and highly model dependent. To provide a general algorithm for fitting such models, we consider a numerical approximation using the last two iterations of $U'_i(\theta)$ at convergence. If $k$ and $k + 1$ are the last two iterations of the algorithm then the $h_{j}^{th}$ entry of $U'_i(\hat{\theta})$ is computed as

$$\frac{U_i(\hat{\theta}^{k+1})_{h} - U_i(\hat{\theta}^{k})_{j}}{\hat{\theta}_{h}^{k+1} - \hat{\theta}_{j}^{k}},$$

(1.7)

where $\hat{\theta}_{h}^{k+1}$ is the $h^{th}$ entry from the $k+1^{th}$ iteration value for $\theta$, and similarly $U_i(\hat{\theta}^{k+1})_{h}$ denotes the $h^{th}$ entry of the $k + 1^{th}$ iteration value of $U_i(\theta)$.

1.4 Simulation Experiments

We carried out simulation experiments to evaluate the inference procedure described above, including the variance estimator. We also investigated the effect of misspecified error correlation structure, as well as non-normally distributed data. Although we do not make the assumption of normally distributed data, the formulas in (1.6) for the E-step was motivated by the multivariate normal distribution. In the following we describe some representative results.

For each scenario we simulated one hundred individuals, each with two outcomes at five time points. Data were generated using fixed effect values as given in
Tables 1.1 - 1.3, where $\beta_{10}, \beta_{11}$ ... denote the fixed intercept and covariate effects for outcome 1, and similarly $\beta_{20}, \beta_{21}$ ... for outcome 2. The covariates in $x_{ijt} = x_i$ were distributed as $N(0, 2)$ unless otherwise specified. The error standard deviations are $\sigma_1$ and $\sigma_2$ for outcomes 1 and 2, and $R$ is their correlation. We considered either a random intercept or a random slope or both in the model, both random effects following a normal distribution with mean zero. In the first two tables $D_1$ and $D_2$ denote the standard deviations of the single random effect for outcomes 1 and 2, respectively, while in Table 1.3 $D_{10}, D_{11}$ and $D_{12}$ denote the standard deviation of the random intercept, the random slope, and their covariance for outcome 1, and similarly $D_{20}, D_{21}$ and $D_{22}$ for outcome 2. When there was a random slope (time), the last covariate was also time $t = 1, 2, 3, 4$ and 5. We ran five hundred trials for each simulated scenario.

Our inference makes two working assumptions: the working correlation $R$ of the errors, and the conditional expectations in (1.6) that are motivated by multivariate normal distribution. Here we examine the robustness of these two types of assumptions.

(a) Correctly and incorrectly specified error correlations:

In all the tables shown data were generated using a correlated error structure. In the top half of each table model (1.1) was fitted using an independent working correlation structure, while in the bottom half it was fitted using an exchangeable working correlation structure. Data were also generated using an independent error correlation structure, and the results were similar, with the estimated $R$ close to zero when an exchangeable working correlation structure was fitted (not shown).

(b) Normally and non-normally distributed errors:

In Table 1.1 and 1.3 the errors were normally distributed, and in Table 1.2 the
errors had a 50:50 mixture of two normal distributions with means at -3 and 3, so that the error distribution was bi-modal with mean zero. Both Table 1.1 and 1.2 had a random intercept; data were also generated using a random slope, and the results were similar (not shown). Table 1.3 included both a random intercept and a random slope.

In the tables we provide the average of the parameter estimates (‘Estimate’) over the 500 simulated trials, as well as the average of the standard errors (‘SE’) using numerical differentiation as described in the last section expression (1.7). We also give the sample standard deviation (‘Sample SD’) of the parameter estimates over the 500 runs, and the empirical coverage probabilities (‘Coverage’) of the nominal 95% confidence intervals.

From the tables we see that the parameter estimates were quite accurate, both under the misspecified error correlations and under non-normality. The ‘SE’ using numerical differentiation as described in the last section appeared to over-estimate the true standard deviation of the parameter estimates, as approximated by the ‘Sample SD’. However, a closer inspection of the individual run’s SE as summarized in Figure 1.1 shows about three dozen outliers all substantially over-estimated the true SD, leading to their average as given in the ‘SE’ column to be larger than the sample SD. On the other hand, the coverage probabilities tended to be lower than the nominal 95%, reflecting the fact that slightly over half of the ‘SE’s under-estimated the SD. The distribution of the parameter estimates appeared to be reasonably well approximated by a normal distribution (data not shown).

Since $R$ represents the parameter of a working correlation structure, like in GEE software we do not estimate its variance.
1.5 Application: Nutrition Data

Here we apply our method to data from the Women’s Healthy Eating and Living (WHEL) Study. The WHEL Study was a multisite dietary intervention trial which assessed the impact of a low-fat, high vegetable/fruit/fiber diet on preventing breast cancer recurrence (Pierce et al., 2002). Between 1995 and 2000, a total of 3088 women who had completed treatment for primary breast cancer were randomly assigned to one of two diet groups, stratified by stage of disease, age at diagnosis, and clinic site. The control group was advised to consume a diet consistent with National Cancer Institute dietary recommendations for cancer prevention, while an intensive telephone counseling approach was used to teach the intervention group to increase consumption of fruits and vegetables rich in micronutrients. Participants in the WHEL study ranged in age from 26 to 71 years at diagnosis, averaging 51 years. Approximately 40% of the WHEL sample was diagnosed with Stage I, 55% with Stage II, and 5% with Stage IIIA breast cancer. The ethnic distribution of WHEL women was 85% Caucasian-American, 5% Hispanic, 4% African-American, 4% Asian/Pacific Islander, with the remaining 2% comprised of other ethnicities. The study was completed in June 2006 (Pierce et al., 2007).

The WHEL study has a rich database of nutritional data collected longitudinally at study entry, and 1- and 4-years post randomization. The primary method for collecting information on participant dietary intake was the 24-hour recall. Repeat 24-hour recalls of dietary intake were collected during telephone interviews by trained dietary assessors blinded to the intervention or control group assignment of the participants. During these recalls participants were queried about their food intake during the previous 24-hours. Assessments at a particular time-point (study entry, year 1, or year 4) consisted of a set of four 24-hour dietary recalls collected over a three-week period, which included two weekdays and two weekend days. These four recalls were
averaged, and this average constituted the 24-hour recall intake measure for that time-point. Nutrient calculations were performed using the Nutrition Data System for Research (NDS-R) software Food and Nutrient Database 31, version 4.03, released November 2000 by Nutrition Coordinating Center at University of Minnesota.

In applying our method, we consider the effect of the WHEL intervention over time on the following three outcomes: fat intake (g/day), fiber intake (g/day), and vegetable intake (servings/day) based on the 24-hour diet recalls. We include in the analysis 2279 women with complete data on the three outcomes at all three time points. In Figure 1.2 the dashed lines represent the three outcomes of 30 randomly selected individuals from each of the randomized groups over time, and Table 1.4 displays their Pearson correlation matrices at the different times. It is clear that the three outcomes are correlated, as one might expect.

In carrying out the analysis to compare the phone counseling intervention with the active control group, we use the three outcomes at 1 and 4 years as dependent variables, with the baseline fat, fiber, and vegetable intakes, together with baseline weight (kg), as covariates. We assumed an exchangeable working correlation structure for $R$ and a random slope over time. In Table 1.5 the intervention by phone counseling has helped to significantly ($p$-value < 0.01) reduce the fat intake, and increase the fiber and vegetable intakes, i.e. improve healthy eating habits overall. The estimates also confirm the time trend as we observe in Figure 1.2: from 1 to 4 years there has been an increase in fat intake, and decreases in fiber and vegetable intakes; in other words, the intervention effects seem to deteriorate over time in both groups, a feature that is common in many behavioral intervention studies whereby early successful changes are followed by some attenuation. The baseline intakes are significantly associated with all three outcomes at the later time points in most cases.

In Tables 1.6 and 1.7 we modeled the trajectory of the three outcomes from
baseline to 1 year then 4 years within each study arm. In this case, time was modeled as a categorical variable with baseline as the intercept (i.e. reference group). The estimated mean trajectories are super-imposed in Figure 1.2 using the black solid line. In the phone counseling group for all three outcomes the larger intervention effects as compared to baseline were seen at year 1, and they became attenuated at year 4 although remained significant when compared to the baseline intakes. For the active control group there was some intervention effect for fat at year 1, but by year 4 all three outcomes were worse when compared to the baseline.

1.6 Discussion

In this paper we have proposed a general regression framework for multivariate longitudinal data which combines the commonly used marginal and conditional modeling, marginal across outcomes and conditional longitudinally. An inference procedure is developed using estimating equations together with the ES algorithm, which, under standard regularity constraints, provides a unique, consistent, and asymptotically normally distributed estimate of the parameters. Via simulation studies we have found the estimator to be robust against misspecified or over-specified error covariance structure in finite samples. The simulations have also shown that the E-step formulas that were motivated by multivariate normal distribution are robust for non-normally distributed errors.

While the marginal approach and the conditional approach are both well-known in the correlated data literature, there have been few applications of the combination of the two. One of the closest works we are aware of is Rosen et al. (2000), who extended finite mixture models for i.i.d. data to correlated setting, using a GEE type approach. Interestingly Rosen et al. (2000) also proposed the ES algorithm, where the unobserved component indicators in the mixture model are the missing
data. We became aware of this work only very recently.

Our approach to study the asymptotic properties of the estimator is new to the best of our knowledge. We have explicitly expressed the conditional expectation of the complete data estimating equations at convergence of the ES algorithm, which is the actual estimating equation that defines the estimator. Although the well-known EM algorithm can be seen as a special case of the ES algorithm when applied to the likelihood score equation as an estimating equation, the asymptotic properties of the maximum likelihood estimator (MLE) are usually derived independent of the EM algorithm. The approach in this paper is broadly applicable, and can be used both for estimating equations and for likelihood methods.

It is known that for the generalized estimating equations, accurately modeling the error correlation structure can improve the efficiency in estimating the regression parameters (Liang and Zeger, 1986). It is likely that this is also the case in the presence of the unobserved random effects. Our work has focused on obtaining a consistent and asymptotically normal estimator; improving efficiency is an area open for future work.

In this paper we have focused on continuous outcomes only. Since the GEE approach also applies to other types of outcomes such as binary or Poisson, we expect that the combined modeling and inference procedure can also be extended. The E-step, however, may be more complex computationally, and may not have closed form expressions.

**Acknowledgement**

Chapter 1, in full, is currently being prepared for submission for publication of the material. Faig, Walter; Xu, Ronghui; Natarajan, Loki. The dissertation author was the primary investigator and author of this material.
Table 1.1: Simulation results for two correlated outcomes, with a random intercept on each outcome. The errors were normally distributed.

<table>
<thead>
<tr>
<th>Working Corr</th>
<th>True Value</th>
<th>Estimate</th>
<th>SE</th>
<th>Sample SD</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>$\beta_{10}$</td>
<td>1</td>
<td>0.94</td>
<td>0.05</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$\beta_{11}$</td>
<td>1</td>
<td>1.04</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>$\beta_{12}$</td>
<td>1</td>
<td>1.02</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>$\beta_{13}$</td>
<td>3</td>
<td>2.94</td>
<td>0.12</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>$\beta_{20}$</td>
<td>1.5</td>
<td>1.46</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>$\beta_{21}$</td>
<td>2</td>
<td>1.95</td>
<td>0.08</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>$\beta_{22}$</td>
<td>0.5</td>
<td>0.48</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>$\beta_{23}$</td>
<td>1.5</td>
<td>1.53</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>$\sigma_1$</td>
<td>2</td>
<td>2.02</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>$\sigma_2$</td>
<td>2</td>
<td>2.04</td>
<td>0.12</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$D_1$</td>
<td>1</td>
<td>0.97</td>
<td>0.10</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>$D_2$</td>
<td>1</td>
<td>0.98</td>
<td>0.12</td>
<td>0.07</td>
</tr>
<tr>
<td>Exchangeable</td>
<td>$\beta_{10}$</td>
<td>1</td>
<td>0.96</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>$\beta_{11}$</td>
<td>1</td>
<td>0.97</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>$\beta_{12}$</td>
<td>1</td>
<td>1.02</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>$\beta_{13}$</td>
<td>3</td>
<td>2.99</td>
<td>0.05</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$\beta_{20}$</td>
<td>1.5</td>
<td>1.51</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$\beta_{21}$</td>
<td>2</td>
<td>2.01</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$\beta_{22}$</td>
<td>0.5</td>
<td>0.51</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>$\beta_{23}$</td>
<td>1.5</td>
<td>1.54</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>$\sigma_1$</td>
<td>2</td>
<td>1.97</td>
<td>0.12</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$\sigma_2$</td>
<td>2</td>
<td>1.96</td>
<td>0.14</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>$D_1$</td>
<td>1</td>
<td>0.91</td>
<td>0.12</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>$D_2$</td>
<td>1</td>
<td>0.92</td>
<td>0.12</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$R$</td>
<td>0.3</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.2: Simulation results for two correlated outcomes, with a random intercept on each outcome. The errors were non-normally distributed.

<table>
<thead>
<tr>
<th>Working Corr</th>
<th>True Value</th>
<th>Estimate</th>
<th>SE</th>
<th>Sample SD</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{10}$</td>
<td>1</td>
<td>1.00</td>
<td>0.07</td>
<td>0.07</td>
<td>93.0</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>1</td>
<td>1.01</td>
<td>0.07</td>
<td>0.06</td>
<td>93.4</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>1</td>
<td>1.00</td>
<td>0.08</td>
<td>0.08</td>
<td>93.8</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>3</td>
<td>3.00</td>
<td>0.07</td>
<td>0.07</td>
<td>92.6</td>
</tr>
<tr>
<td>$\beta_{20}$</td>
<td>1.5</td>
<td>1.51</td>
<td>0.07</td>
<td>0.07</td>
<td>92.6</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>2</td>
<td>2.01</td>
<td>0.08</td>
<td>0.06</td>
<td>91.2</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>0.5</td>
<td>0.50</td>
<td>0.08</td>
<td>0.07</td>
<td>91.4</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>1.5</td>
<td>1.51</td>
<td>0.07</td>
<td>0.07</td>
<td>92.8</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>2</td>
<td>1.99</td>
<td>0.14</td>
<td>0.09</td>
<td>92.0</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>2</td>
<td>1.98</td>
<td>0.13</td>
<td>0.10</td>
<td>92.8</td>
</tr>
<tr>
<td>$D_1$</td>
<td>1</td>
<td>0.95</td>
<td>0.15</td>
<td>0.08</td>
<td>88.2</td>
</tr>
<tr>
<td>$D_2$</td>
<td>1</td>
<td>0.95</td>
<td>0.13</td>
<td>0.09</td>
<td>87.8</td>
</tr>
<tr>
<td>$R$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exchangeable</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{10}$</td>
<td>1</td>
<td>1.01</td>
<td>0.07</td>
<td>0.07</td>
<td>92.4</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>1</td>
<td>0.99</td>
<td>0.08</td>
<td>0.07</td>
<td>92.6</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>1</td>
<td>0.99</td>
<td>0.08</td>
<td>0.08</td>
<td>92.0</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>3</td>
<td>3.00</td>
<td>0.08</td>
<td>0.08</td>
<td>91.8</td>
</tr>
<tr>
<td>$\beta_{20}$</td>
<td>1.5</td>
<td>1.49</td>
<td>0.08</td>
<td>0.08</td>
<td>93.6</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>2</td>
<td>1.99</td>
<td>0.09</td>
<td>0.08</td>
<td>93.8</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>0.5</td>
<td>0.50</td>
<td>0.08</td>
<td>0.07</td>
<td>93.4</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>1.5</td>
<td>1.51</td>
<td>0.07</td>
<td>0.07</td>
<td>93.4</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>2</td>
<td>1.98</td>
<td>0.13</td>
<td>0.09</td>
<td>91.8</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>2</td>
<td>1.98</td>
<td>0.13</td>
<td>0.09</td>
<td>92.0</td>
</tr>
<tr>
<td>$D_1$</td>
<td>1</td>
<td>0.95</td>
<td>0.11</td>
<td>0.08</td>
<td>88.0</td>
</tr>
<tr>
<td>$D_2$</td>
<td>1</td>
<td>0.94</td>
<td>0.13</td>
<td>0.08</td>
<td>89.4</td>
</tr>
<tr>
<td>$R$</td>
<td></td>
<td>0.30</td>
<td>0.27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.3: Simulation results for two correlated outcomes, with a random intercept and a random slope on each outcome.

<table>
<thead>
<tr>
<th>Working Corr</th>
<th>True Value</th>
<th>Estimate</th>
<th>SE</th>
<th>Sample SD</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>$\beta_{10}$</td>
<td>1</td>
<td>1.01</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$\beta_{11}$</td>
<td>1</td>
<td>1.00</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>$\beta_{12}$</td>
<td>1</td>
<td>1.00</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$\beta_{13}$</td>
<td>3</td>
<td>3.00</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$\beta_{20}$</td>
<td>1.5</td>
<td>1.49</td>
<td>0.09</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>$\beta_{21}$</td>
<td>2</td>
<td>2.01</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>$\beta_{22}$</td>
<td>0.5</td>
<td>0.51</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>$\beta_{23}$</td>
<td>1.5</td>
<td>1.50</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$\sigma_1$</td>
<td>2</td>
<td>2.00</td>
<td>0.11</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>$\sigma_2$</td>
<td>2</td>
<td>2.00</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>$D_{10}$</td>
<td>1</td>
<td>0.95</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>$D_{11}$</td>
<td>1</td>
<td>0.94</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>$D_{12}$</td>
<td>0</td>
<td>0.07</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>$D_{20}$</td>
<td>1</td>
<td>0.96</td>
<td>0.11</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$D_{21}$</td>
<td>1</td>
<td>0.95</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>$D_{22}$</td>
<td>0</td>
<td>0.06</td>
<td>0.08</td>
<td>0.06</td>
</tr>
<tr>
<td>Exchangeable</td>
<td>$\beta_{10}$</td>
<td>1</td>
<td>1.00</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>$\beta_{11}$</td>
<td>1</td>
<td>1.00</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$\beta_{12}$</td>
<td>1</td>
<td>1.00</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$\beta_{13}$</td>
<td>3</td>
<td>3.01</td>
<td>0.08</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>$\beta_{20}$</td>
<td>1.5</td>
<td>1.50</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$\beta_{21}$</td>
<td>2</td>
<td>1.99</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$\beta_{22}$</td>
<td>0.5</td>
<td>0.50</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$\beta_{23}$</td>
<td>1.5</td>
<td>1.50</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>$\sigma_1$</td>
<td>2</td>
<td>2.00</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>$\sigma_2$</td>
<td>2</td>
<td>2.01</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>$D_{10}$</td>
<td>1</td>
<td>0.95</td>
<td>0.10</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$D_{11}$</td>
<td>1</td>
<td>0.95</td>
<td>0.10</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$D_{12}$</td>
<td>0</td>
<td>0.07</td>
<td>0.08</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>$D_{20}$</td>
<td>1</td>
<td>0.95</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>$D_{21}$</td>
<td>1</td>
<td>0.96</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>$D_{22}$</td>
<td>0</td>
<td>0.06</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>$R$</td>
<td>0.30</td>
<td>0.21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.4: Correlation among the three outcomes over time

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th></th>
<th>Year 1</th>
<th></th>
<th>Year 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fat</td>
<td>Fiber</td>
<td>Veg.</td>
<td>Fat</td>
<td>Fiber</td>
</tr>
<tr>
<td>Fat</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Fiber</td>
<td>0.723 1</td>
<td></td>
<td></td>
<td>0.662 1</td>
<td></td>
</tr>
<tr>
<td>Veg</td>
<td>0.740 0.710 1</td>
<td></td>
<td>0.661 0.659 1</td>
<td></td>
<td>0.715 0.695 1</td>
</tr>
</tbody>
</table>
Table 1.5: Comparison of two intervention groups while adjusting for baseline covariates

<table>
<thead>
<tr>
<th></th>
<th>Fat</th>
<th></th>
<th></th>
<th>Fiber</th>
<th></th>
<th></th>
<th>Veg.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
<td>p-value</td>
<td>Estimate</td>
<td>SE</td>
<td>p-value</td>
<td>Estimate</td>
<td>SE</td>
</tr>
<tr>
<td>Intercept</td>
<td>27.56</td>
<td>2.67</td>
<td>&lt;0.01</td>
<td>11.66</td>
<td>0.81</td>
<td>&lt;0.01</td>
<td>1.57</td>
<td>0.72</td>
</tr>
<tr>
<td>Weight0</td>
<td>0.01</td>
<td>0.00</td>
<td>0.09</td>
<td>0.01</td>
<td>0.00</td>
<td>0.22</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Fat0</td>
<td>0.39</td>
<td>0.00</td>
<td>&lt;0.01</td>
<td>-0.04</td>
<td>0.01</td>
<td>&lt;0.01</td>
<td>-0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Fib0</td>
<td>0.09</td>
<td>0.00</td>
<td>&lt;0.01</td>
<td>0.49</td>
<td>0.00</td>
<td>&lt;0.01</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>Veg0</td>
<td>-0.17</td>
<td>0.04</td>
<td>&lt;0.01</td>
<td>0.53</td>
<td>0.14</td>
<td>&lt;0.01</td>
<td>0.39</td>
<td>0.02</td>
</tr>
<tr>
<td>Time</td>
<td>5.24</td>
<td>2.02</td>
<td>0.01</td>
<td>-2.86</td>
<td>1.23</td>
<td>0.02</td>
<td>-0.25</td>
<td>0.13</td>
</tr>
<tr>
<td>Intervention</td>
<td>-9.85</td>
<td>2.73</td>
<td>&lt;0.01</td>
<td>7.00</td>
<td>0.15</td>
<td>&lt;0.01</td>
<td>1.49</td>
<td>0.17</td>
</tr>
<tr>
<td>σ</td>
<td>17.67</td>
<td>0.15</td>
<td>&lt;0.01</td>
<td>7.40</td>
<td>0.10</td>
<td>&lt;0.01</td>
<td>1.65</td>
<td>0.34</td>
</tr>
<tr>
<td>D</td>
<td>1.14</td>
<td>4.97</td>
<td>0.82</td>
<td>0.48</td>
<td>0.21</td>
<td>0.02</td>
<td>0.11</td>
<td>0.00</td>
</tr>
<tr>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.41</td>
</tr>
</tbody>
</table>
Table 1.6: Time trajectory from baseline to 1 and 4 years (control group)

<table>
<thead>
<tr>
<th></th>
<th>Fat Estimate</th>
<th>Fat SE</th>
<th>Fat p-value</th>
<th>Fiber Estimate</th>
<th>Fiber SE</th>
<th>Fiber p-value</th>
<th>Veg. Estimate</th>
<th>Veg. SE</th>
<th>Veg. p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>56.52</td>
<td>3.77</td>
<td>&lt;0.01</td>
<td>21.41</td>
<td>0.02</td>
<td>&lt;0.01</td>
<td>3.00</td>
<td>0.41</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Year 1</td>
<td>-4.17</td>
<td>0.65</td>
<td>&lt;0.01</td>
<td>-0.21</td>
<td>0.50</td>
<td>0.68</td>
<td>0.00</td>
<td>0.00</td>
<td>0.17</td>
</tr>
<tr>
<td>Year 4</td>
<td>0.13</td>
<td>0.04</td>
<td>&lt;0.01</td>
<td>-2.11</td>
<td>0.42</td>
<td>&lt;0.01</td>
<td>-0.06</td>
<td>0.18</td>
<td>0.71</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>21.22</td>
<td>0.00</td>
<td>&lt;0.01</td>
<td>7.82</td>
<td>0.07</td>
<td>&lt;0.01</td>
<td>1.81</td>
<td>0.07</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>D</td>
<td>2.98</td>
<td>0.05</td>
<td>&lt;0.01</td>
<td>1.07</td>
<td>0.33</td>
<td>&lt;0.01</td>
<td>0.23</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.48</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.7: Time trajectory from baseline to 1 and 4 years (phone counseling group)

<table>
<thead>
<tr>
<th></th>
<th>Fat</th>
<th></th>
<th>Fiber</th>
<th></th>
<th>Veg.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
<td>p-value</td>
<td>Estimate</td>
<td>SE</td>
<td>p-value</td>
</tr>
<tr>
<td>Intercept</td>
<td>55.67</td>
<td>4.10</td>
<td>&lt;0.01</td>
<td>21.58</td>
<td>0.35</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Year 1</td>
<td>-14.45</td>
<td>0.97</td>
<td>&lt;0.01</td>
<td>7.76</td>
<td>1.79</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Year 4</td>
<td>-8.25</td>
<td>1.78</td>
<td>&lt;0.01</td>
<td>3.88</td>
<td>0.34</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>19.04</td>
<td>0.05</td>
<td>&lt;0.01</td>
<td>9.30</td>
<td>0.11</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>D</td>
<td>2.68</td>
<td>0.15</td>
<td>&lt;0.01</td>
<td>1.30</td>
<td>0.06</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Figure 1.1:** Boxplot of estimated standard error of $\beta_{11}$ from each of the 500 simulation runs; the dashed line represents the sample SD.
Figure 1.2: Plots of 30 randomly selected individuals (dashed lines) over time; left: phone counseling group, right: control group. The black solid line is the fitted mean.
Chapter 2

A Cure-Rate Model For Left-Truncated and Right-Censored Data

2.1 Introduction

In medical studies, data will often include large numbers of long-term survivors. In some cases this may be due to the nature of the outcome, such as spontaneous abortion (SAB) in pregnancy outcomes data which is defined to occur within the first 20 weeks, but is often based on empirical observation that some participants will never experience an event. This assumption that a portion of the population is not susceptible leads to the introduction of cure-rate models to effectively analyze the survival rate of those who are susceptible along with the probability of an individual being “cured” (Farewell (1982, 1986)).

Cure-rate models have been developed in two varieties. There are mixture models, as seen in Sy and Taylor (2000) for example, or non-mixture models (Chen et al. (1999)). There has been much analysis of mixture models for various assumed dependencies of the survival function on the covariates. The most common of which, as in Kuk and Chen (1992), is Cox proportional hazards distribution (they also use the
common logistic regression to model the cured probability). Other survival models include transformation models (Lu and Ying (2004)) or the use of rich parametric models if the shape of the hazard function is of interest (Hanson et al. (2003)).

Thus far these mixture models have been explored for right-censored or interval-censored data (Kim and Jhun (2008)). The aim in this paper is to develop the mixture model using Cox proportional hazards and logistic regression that accommodates left-truncated and right-censored survival data. As many medical studies, such as those examining pregnancy outcomes, do not have participants enter precisely when they commence being at risk our model will allow for the application of cure-rate models to a broader spectrum of medical survival data.

The main challenge of data with left-truncation is that it induces selection bias as some people who would have entered a study will have an event before they can do so. This bias can be corrected using a weighted partial likelihood approach with inverse probability weights (Breslow and Wellner (2007), Sasieni (1993)). In our model we estimate the weights via the empirical distribution of the truncation times for the cured portion of the sample. The asymptotics of weighted likelihood methods for semiparametric models are presented in Li et al. (2008) and Zeng and Lin (2007).

We show that our model produces a unique consistent estimator which is asymptotically normally distributed, and illustrate its effectiveness through various simulation experiments. Particularly, we show that it corrects the selection bias from left-truncation. We conclude with example using pregnancy outcomes data from the Organization of Teratology Information Specialists (OTIS) in which we consider the outcome of spontaneous abortion (SAB).
2.2 Model and Estimation

2.2.1 The Model

Let $Y_i$ be the event indicator (possibly unobserved) for subject $i$ ($i = 1, ..., N$). Let $\tau$ be a time (possibly infinite) after which an individual is no longer considered susceptible, and let $T_i$ denote the event times. Let $Q_i$ be the left-truncation times and $C_i$ be right-censoring times. Then we define $X_i = \min(T_i, C_i, \tau)$. We assume noninformative truncation and censoring; that is $(Q_i, C_i)$ with $Q_i < C_i$ are both independent of $T_i$. This is the independence assumption found in Tsai et al. (1987) and Wang (1991). Let $\delta_i = I(X_i = T_i)$ and $Z_i$ be a vector of covariates. We allow for the first covariate to be 1 in which case denote $\tilde{Z}_i$ as the vector of nonconstant covariates. So the triples $(X_i, Z_i, \delta_i)$ are observed iid data. For $p_i = P(Y_i = 1)$ we consider the marginal cure-rate mixture model,

\[ \bar{S}_i(t) = P(T_i > t) = (1 - p_i) + p_i S_i(t), \quad (2.1) \]

where $S_i(t) = P(T_i > t | Y_i = 1), t < \tau$.

To model $p_i$ and $S_i(t)$ we use logistic and Cox proportional hazards regression respectively. These are common regression choices; Sy and Taylor (2000) uses them for the cure-rate model with right-censoring. So for $p_i$, let $\alpha$ be a parameter vector the length of $Z_i$. Then,

\[ p_i = \frac{\exp(\alpha' \tilde{Z}_i)}{1 + \exp(\alpha' \tilde{Z}_i)}. \quad (2.2) \]

For $\Lambda_0(t)$ the cumulative baseline hazard functions and parameter vector $\beta$ the length of $\tilde{Z}_i$,

\[ \Lambda_i(t) = \Lambda_0(t) \exp(\beta' \tilde{Z}_i), \quad (2.3) \]

with $S_i(t) = \exp(-\Lambda_i(t))$. Following Struthers and Farewell (1989), the observed data
The likelihood is
\[
L(\theta) = \prod_{i=1}^{N} \left\{ \frac{\bar{f}_i(X_i)}{\bar{S}_i(Q_i)} \right\}^{\delta_i} \left\{ \frac{\bar{S}_i(X_i)}{\bar{S}_i(Q_i)} \right\}^{1-\delta_i},
\]
where \( \theta = (\alpha, \beta, \Lambda_0) \), \( \bar{f}_i(t) = -d\bar{S}_i(t)/dt = p_i f_i(t) \), and \( f_i(t) = -dS_i(t)/dt \). We propose the following augmented complete data likelihood,
\[
L_c(\theta; X, \delta, Y|Q, Z) = \prod_{i=1}^{N} L_1(\alpha; Y_i|Z_i) \{ L_2(\beta, \Lambda_0; X_i, \delta_i|Y_i = 1, Q_i, Z_i) \}^{Y_i}
= \prod_{i=1}^{N} p_i^{Y_i} (1 - p_i)^{1-Y_i} \left\{ \frac{\bar{f}_i(X_i)}{\bar{S}_i(Q_i)} \right\}^{\delta_i Y_i} \left\{ \frac{\bar{S}_i(X_i)}{\bar{S}_i(Q_i)} \right\}^{(1-\delta_i)Y_i}.
\]
Note that (2.5) does not reduce to (2.4) when integrating over observed data, but it is advantageous in that the logistic and Cox regression parameters can be estimated separately. However, due to left-truncation parameter estimates are biased and we must adjust \( L_c \) to account for those individuals who are truncated out.

Our approach is to instead consider a weighted complete data likelihood,
\[
L_w(\theta) = \prod_{i=1}^{N} \left[ p_i^{Y_i} (1 - p_i)^{1-Y_i} \left\{ \frac{\bar{f}_i(X_i)}{\bar{S}_i(Q_i)} \right\}^{\delta_i Y_i} \left\{ \frac{\bar{S}_i(X_i)}{\bar{S}_i(Q_i)} \right\}^{(1-\delta_i)Y_i} \right]^{\nu_i}.
\]
Following Breslow and Wellner (2007) and Gross and Lai (1996), an appropriate choice for the \( \nu_i \) to correct the bias is to use inverse probability weights (IPW) based on the distribution of the truncation times \( G \). Estimation will be based on the IPW versions of score equations derived from the complete data likelihood. We condition on the event times \( T_i \) rather than the entry times as it performs much better in our simulation results. Conditioning on the \( Q_i \) proved to produce very large estimates for the weights. Explicitly, the weights are defined as \( \nu_i = 1/P(Q_i < T_i|T_i) \). For censored individuals (\( \delta_i = 0 \), \( T_i \) are not observed so to find estimates for \( \nu_i \) we impute a time between \( X_i \) and \( \tau \).

### 2.2.2 Estimation of \( G \)

Here we expound upon several approaches to estimating \( G(t) = P(Q < t) \) along with qualifications to their uses. Without covariates the NPMLE of \( G(t) \) is
studied in Wang (1991), which gives the following estimate,  
\[
\hat{G}(t) = \frac{\sum_{i=1}^{N} I(Q_i \leq t) / \hat{S}(Q_i)}{\sum_{i=1}^{N} 1 / \hat{S}(Q_i)},
\]  
where \( \hat{S} \) is the product-limit estimate for \( S(t) = P(t < T) \). However, for our model, we would need to consider \( \hat{S}(t) \) conditional on the covariates, and this approach does not lend itself well to closed form estimation of the variance. In the absence of censoring (2.7) reduces to a product-limit estimator for \( G \).

A similar estimate for \( G \) can be found using iterative methods where \( \nu_i \)'s are updated are each step based on the current weighted empirical distribution function for the \( Q_i \)'s,  
\[
\hat{G}^{k+1}(t) = \frac{\sum_{i=1}^{N} \hat{\nu}_i^k I(Q_i \leq t)}{\sum_{i=1}^{N} \hat{\nu}_i^k}.
\]  
Here \( \hat{\nu}_i^k = \hat{G}^k(X_i)^{-1} \). The weak convergence of weighted empirical processes is discussed in Shao and Yu (1996) and Donald and Hsu (2012) looks at the case of empirical distribution functions with IPW. At convergence we claim,  
\[
\hat{G}^{(1)}(t) = \frac{\sum_{i=1}^{N} \hat{\nu}_i I(Q_i \leq t)}{\sum_{i=1}^{N} \hat{\nu}_i}.
\]  

As, the cured portion of the sample is not subject to truncation, a third approach is to estimate \( G \) using the empirical distribution function from the cured portion of the sample,  
\[
\hat{G}^{(2)}(t) = \frac{\sum_{i=1}^{N} I(Q_i \leq t, Y_i = 0)}{\sum_{i=1}^{N} I(Y_i = 0)}.
\]  
For ease of notation throughout the remainder of this paper this is estimate we consider. Note, that without any smoothing kernel, the use of this estimator requires a large portion of the sample be cured and its smallest entry time be less than the smallest event time. In simulation studies when these conditions are met we can see that (2.9) performs identically to (2.10). So for general use we believe it is reasonable to use (2.9) although in this paper we only detail the asymptotic properties and closed form variance estimate for the model using (2.10).
Length-Biased Data and Testing for Uniformity of $G$

Though we assume non-informative truncation and censoring times for our model, here we briefly discuss the methods developed for length-biased data and methods for testing whether the truncation distribution is uniform. For nonparametric truncation distributions, two main approaches are a conditional approach developed in Wang et al. (1986) in which NPMLE estimates are derived from the log-likelihood conditional on the truncation distribution and the unconditional approach developed in Vardi (1989). When the truncation distribution is uniform Asgharian et al. (2002) show that this unconditional approach is more efficient. In Luo and Tsai (2009) they propose a pseudo-likelihood approach to estimation which gives closed-form estimators and performs well when compared with the unconditional approach.

For left-truncated and right-censored survival data, Asgharian et al. (2006) suggest a visual test to check whether the truncation distribution is uniform. If truncation times are uniform across the study duration of the study then the distributions for the truncation times and residual life times should show no significant difference. Mandel and Betensky (2007) give the corresponding paired log-rank test to this visual test, as well as observe that for non-censored data the Kolmogorov-Smirnov test can be used to compare $U[0, 1]$ with the distribution of the $Q_i/T_i$. For the cure-rate model both tests should be utilized with the visual test and Kolmogorov-Smirnov test being applied to the non-cured and cured portions of the sample respectively.
2.2.3 ES Algorithm

From our complete data likelihood (2.6), the complete data log-likelihood is:

\[
\ell_{\nu}(\theta) = \sum_{i=1}^{N} \nu_i [Y_i \alpha' Z_i - \log\{1 + \exp(\alpha' Z_i)\} \\
+ Y_i \{\Lambda_0(Q_i) - \Lambda_0(X_i)\} \exp(\beta' \tilde{Z}_i) \\
+ \delta_i Y_i \{\beta' \tilde{Z}_i + \log \lambda_0(X_i)\}].
\]  
(2.11)

This is not the true log-likelihood, however, we can still utilize an EM-type algorithm for parameter estimation. In Elashoff and Ryan (2004), they developed an ES algorithm for generalized estimating equations in which the update to the parameters is found by substituting in the expected values of missing data based on current parameter estimates. Our score equations found through differentiation of (2.11) and estimation of \(G\) in (2.10). Here let \(\ell_{\nu_i}\) denote the weighted log-likelihood contribution from the \(i^{th}\) individual. \(0 < t_1 < ... < t_K < \tau\) are the distinct ordered event times and \(\lambda_k\) is the point mass of \(\Lambda_0\) at \(t_k\). \(0 < q_1 < ... < q_J < \tau\) are the distinct ordered truncation times of the cured individuals and \(G_j = G(q_j)\). The complete data score equations are as follows:

\[
U^{(\alpha)} = \sum_{i=1}^{N} U^{(\alpha)}_i = \sum_{i=1}^{N} \frac{\partial \ell_{\nu_i}}{\partial \alpha} = \sum_{i=1}^{N} \nu_i Z_i \left\{ Y_i - \frac{\exp(\alpha' Z_i)}{1 + \exp(\alpha' Z_i)} \right\} = 0,
\]  
(2.12)

\[
U^{(\beta)} = \sum_{i=1}^{N} U^{(\beta)}_i = \sum_{i=1}^{N} \frac{\partial \ell_{\nu_i}}{\partial \beta} = \sum_{i=1}^{N} \nu_i Y_i \tilde{Z}_i \left[ \delta_i + \{\Lambda_0(Q_i) - \Lambda_0(X_i)\} \exp(\beta' \tilde{Z}_i) \right] = 0,
\]  
(2.13)

\[
U^{(\lambda_k)} = \sum_{i=1}^{N} \frac{\partial \ell_{\nu_i}}{\partial \lambda_k} = \sum_{i=1}^{N} \nu_i Y_i \left\{ \frac{\delta_i I(X_i = t_k)}{\lambda_k} - \exp(\beta' \tilde{Z}_i) I(Q_i < t_k \leq X_i) \right\} = 0
\]  
(2.14)

\(k = 1, ..., K\).
\[ U^{(G_j)} = \sum_{i=1}^{N} U_{i}^{(G_j)} = \sum_{i=1}^{N} \left\{ \frac{G_j}{N} - \frac{I(Q_i \leq q_j, Y_i = 0)}{\sum_{i=1}^{N} I(Y_i = 0)} \right\} = 0 \quad (2.15) \]

\[ j = 1, \ldots, J - 1. \]

Note that as \( G_J = 1 \) we are only estimating the value of \( G \) at prior truncation times.

The ES algorithm proceeds iteratively; first computing the expected value of missing data based on current parameter estimates. We denote \( \omega_i = E(Y_i|\text{observed data}) \) and note when \( Y_i \) is observed \( \omega_i = Y_i \). For \( Y_i \) which are not observed we have,

\[ \omega_i = E(Y_i|T_i > X_i, T_i > Q_i) = P(T_i > X_i|T_i > Q_i, Y_i = 1)P(Y_i = 1) + P(T_i > X_i, T_i > Q_i|Y_i = 0)P(Y_i = 0) \]

\[ = \frac{\exp(\alpha'Z_i)\exp\left(\{\Lambda_0(Q_i) - \Lambda_0(X_i)\}e^{\beta'\tilde{Z}_i}\right)}{\exp(\alpha'Z_i)\exp\left(\{\Lambda_0(Q_i) - \Lambda_0(X_i)\}e^{\beta'\tilde{Z}_i}\right) + \nu_i}. \quad (2.16) \]

The S-step in the algorithm then obtains update to parameters by taking conditional expectation of the complete data score equations over the observed data and substituting the expected value of missing data found in the E-step into equations (2.12-2.15). Let \( U_i = (U^{(\alpha)}_i, U^{(\beta)}_i, U^{(\lambda)}_i, U^{(G)}_i)^T \) then denote \( U_i = E(U_i|\text{observed data}) \) and \( U = \sum_{i=1}^{N} U_i \). Parameter estimates then solve \( U(\phi) = 0 \).

Estimates for \( \alpha \) can be obtained separately from estimates for \( \beta \) and \( \Lambda_0 \). From \( U^\alpha \), we use Newton-Raphson methods to compute the estimate for \( \alpha \). For fixed \( \hat{\alpha} \) from the most recent E-step iteration we have:

\[ \frac{\partial U^{(\alpha)}(\alpha; \hat{\alpha}, \hat{\nu})}{\partial \alpha}(\hat{\alpha}; \hat{\omega}, \hat{\nu}) = \sum_{i=1}^{N} \frac{-\nu_i Z_i^\prime Z_i \exp(\alpha'Z_i)}{1 + \exp(\alpha'Z_i)} \quad (2.17) \]

The Newton-Raphson update to \( \hat{\alpha} \) is then given by:

\[ \hat{\alpha}^k = \hat{\alpha}^k - \frac{\partial U^{(\alpha)}(\hat{\alpha}^k; \hat{\omega}, \hat{\nu})}{\partial \alpha}(\hat{\alpha}^{k-1}; \hat{\omega}, \hat{\nu})^{-1} U^{(\alpha)}(\hat{\alpha}^{k-1}; \hat{\omega}, \hat{\nu}). \quad (2.18) \]
Estimates for $\beta$ and $\Lambda_0$ follow standard approach first suggested in Breslow (1972) following the initial proposal of the proportional hazards model in Cox (1972). This partial likelihood approach gives the estimate for $\beta$ that solves the following:

$$
\sum_{i:Y_i=1} \hat{\nu}_i [\beta' \tilde{Z}_i - \log \left( \sum_{Q_i < t_j \leq X_i} \exp(\beta' \tilde{Z}_j) \right)] = 0 \quad (2.19)
$$

Having obtained $\hat{\beta}$, the $\hat{\lambda}_k$ can be derived in the following manner. For given $\hat{\omega}$ the survival portion of the conditional weighted log-likelihood can be expressed as:

$$
\sum_{i=1}^{N} \hat{\nu}_i \hat{\omega}_i \{ \hat{\beta}' \tilde{Z}_i + \log \lambda_0(X_i) \} - \int_{Q_i}^{X_i} \hat{\nu}_i \hat{\omega}_i \lambda_0(s) \exp(\hat{\beta}' \tilde{Z}_i) ds. \quad (2.20)
$$

As we estimate $\Lambda_0$ at the event times, the integral reduces to

$$
\sum_{k:Q_i < t_k \leq X_i} \hat{\nu}_i \hat{\omega}_i \lambda_0(t_k) \exp(\hat{\beta}' \tilde{Z}_i). \quad (2.21)
$$

Replacing, the integral in (2.20) with this summation and taking derivatives with respect to the $\lambda_k$ we arrive at:

$$
\frac{d_k}{\hat{\lambda}_k} - \sum_{i:Q_i < t_k \leq X_i} \hat{\nu}_i \hat{\omega}_i \exp(\hat{\beta}' \tilde{Z}_i) = 0 \quad k = 1, \ldots, K. \quad (2.22)
$$

Here $d_k$ is the weighted number of events at time $t_k$. It then follows that the estimates for $\lambda_k$ are:

$$
\hat{\lambda}_k = \frac{d_k}{\sum_{i:Q_i < t_k \leq X_i} \hat{\nu}_i \hat{\omega}_i \exp(\hat{\beta}' \tilde{Z}_i)} \quad k = 1, \ldots, K. \quad (2.23)
$$

Estimation of the survival parameters is done using existing proportional hazards software for weighted left-truncated and right-censored survival data using weights $\hat{\nu}_i \hat{\omega}_i$.

### 2.3 Asymptotic Properties

To study the asymptotic properties of $\hat{\phi}$, we first observe that $\hat{G}$ is the empirical distribution function of the truncation times and not only coversges pointwise to the true distribution $G_0$, but by the Glivenko-Cantelli theorem $\hat{G} \to G_0$ uniformly (Van Der Vaart and Wellner, 1996, p. 81-82). So we primarily concern ourselves with the
asymptotic properties of \( \hat{\theta} \) and our arguments for the estimates from the weighted complete data likelihood follow closely to those developed for NPMLE.

To show the identifiability of the cure-rate model we introduce the following notation. Recall the marginal survival function:

\[
\overline{S}(t; Z) = (1 - p(Z)) + p(Z)S(t; Z), \quad t < \tau \tag{2.23}
\]

Also let \( r(Z) = \exp(\beta' \tilde{Z}) \) so that,

\[
S(t; Z) = \exp(-\Lambda_0(t))^{\exp(\beta' \tilde{Z})} = S_0(t)^{r(Z)} \tag{2.24}
\]

Define \( \mathcal{Z} \) as the covariate space and \( \mathcal{P} = \{p(Z) : p(Z) \neq 0 \text{ or } 1 \text{ for all } Z \in \mathcal{Z} \} \). For \( F(t; Z) \), the conditional distribution function corresponding to \( S(t; Z) \), the marginal distribution function is \( \overline{F}(t, Z) = p(Z)F(t; Z) \), for \( t < \tau \). \( \mathcal{F} \) is the class of conditional distribution functions defined over parameter space \( \Theta \). Our class of cure-rate models can be denoted as, \( \mathcal{H} = \{\overline{F}(t, Z) : \overline{F}(t, Z) = p(Z)F(t; Z), t < \tau, Z \in \mathcal{Z}, p \in \mathcal{P}, F \in \mathcal{F}\} \). We then have the following theorem:

**Theorem 2.3.1.** The model defined by \( \overline{S} \) is identifiable if \( Z \in \mathcal{Z} \). i.e. For any \( \overline{F}(t, Z) \) and \( \overline{F}^*(t, Z) \) in \( \mathcal{H} \), \( \overline{F}(t, Z) = \overline{F}^*(t, Z) \) if and only if \( p(Z) = p^*(Z) \) for all \( Z \in \mathcal{Z} \) and \( F(t) = F^*(t) \) for \( t < \tau \)

The proof is given in the Appendix.

To show the consistency and convergence of \( \hat{\phi} \) we assume the following regularity conditions:

**C1** The parameter space of the finite dimensional parameters \((\alpha, \beta)\) is compact, and the true values \((\alpha_0, \beta_0)\) lie in the interior of the parameter space.

**C2** The covariates in \( Z \) are bounded.

**C3** We focus on the time interval \((\tau_1, \tau_2)\) where \( 0 < \tau_1 < \tau_2 < \tau \), such that

\[ P(Q < \tau_1, C > \tau_2) > 0. \]
C4 $\lambda_0(t)$ is continuous on $(0, \tau)$ and $\Lambda_0(\tau_2) < \infty$.

C5 $E_{\phi_0}(\partial\frac{\theta}{\partial\phi})$ is nonsingular.

C6 $E_{\phi_0}(\nu|Z, X, \delta)$ is bounded above by $D < \infty$.

Note the closed form of $\frac{\partial U_i}{\partial \phi}$ is given in the Appendix. Also, we examine only the case without censoring here as $\hat{\nu}_{i}^{-1} = \hat{G}(q_j)I(q_j \leq T_i < q_{j+1})$ and the $T_i$ are only all observed when there is no censoring.

We then have the following theorem:

**Theorem 2.3.2.** Under regularity conditions C1 – C6, $\hat{\phi} = (\hat{\alpha}, \hat{\beta}, \hat{\Lambda}_0, \hat{G})$, the estimator obtained through the ES algorithm, is a consistent estimator for $\phi_0 = (\alpha_0, \beta_0, \Lambda_0, G_0)$. i.e. For each parameter in $\hat{\phi}$ there is convergence in probability to its true value in $\phi_0$. For the infinite dimensional parameters $(\Lambda_0, G)$ this convergence is pointwise.

The proof is given in the Appendix.

Now let $h = (h_1, h_2, h_3, h_4)$ where $h_1$ and $h_2$ are vectors of the same dimension as $\alpha$ and $\beta$ respectively. $h_3$ and $h_4$ are functions of bounded variation on $[0, \tau]$. We denote $H_N$ as the vector with elements $h_1, h_2, h_3(X_i)$ at those $X_i$ where $\delta_i = 1$, and $h_4(Q_i)$ from those cured individuals. $H = \{h = (h_1, h_2, h_3, h_4)\}$ and $H_P = \{h \in H : ||h_1||_v, ..., ||h_4||_v \leq p\}$ for some $0 < p < \infty$.

**Theorem 2.3.3.** Under regularity conditions C1 – C6, the process $\sqrt{N}(\hat{\alpha} - \alpha_0, \hat{\beta} - \beta_0, \hat{\Lambda}_0 - \Lambda_0, \hat{G} - G_0)$ converges in distribution to a mean zero Gaussian process in the functional space $\ell^\infty(H_P)$ on $H_P$.

An outline of the proof is given in the Appendix.

For the following sandwich estimator for the variance,

$$\hat{V}_N = \left\{ \sum_{i=1}^{N} \frac{\partial U_i(\hat{\phi})}{\partial \hat{\phi}} \right\}^{-1} \left[ \frac{\sum_{i=1}^{N} U_i(\hat{\phi})U_i(\hat{\phi})^\top}{\sum_{i=1}^{N} \frac{\partial U_i(\hat{\phi})}{\partial \hat{\phi}} \left( \frac{\partial U_i(\hat{\phi})}{\partial \hat{\phi}} \right)^\top} \right] \left\{ \sum_{i=1}^{N} \frac{\partial U_i(\hat{\phi})}{\partial \hat{\phi}} \right\}^{-1},$$

(2.25)

we have the following theorem:
Theorem 2.3.4. Under regularity conditions $C1 - C6$, $N h'_N V_N h_N$ converges uniformly in probability to the asymptotic variance of

$$\sqrt{N} \{h_1'(\hat{\alpha} - \alpha_0) + h_2'(\hat{\beta} - \beta_0) + \int_0^\tau h_3(u) d(\hat{\Lambda}_0 - \Lambda_0)(u) + \int_0^\tau h_4(u) d(\hat{G} - G_0)(u)\}.$$ 

Proof. The proof follows the Theorem 3’s in the respective papers of Parner (1998) and Zeng and Lin (2007).

2.3.1 Multiple Imputation Variance

In order to allow for possibly right-censored data when computing the variance we use a multiple imputation approach. For censored individuals, we impute an event time at each iteration in our algorithm using the most recent update for $\hat{\alpha}$ by $(1 - \hat{p}_i)\tau + \hat{p}_i U[X_i, \tau]$. This imputed event time is then used to update $\hat{\nu}_i$. The resulting variance estimate is the sum of the average variance estimates using the imputed values and an additional component due to the imputation (Chen and Sun (2010), Pan (2000)). Let $\hat{\phi}^{(m)}$ be the parameter estimate based on the $m^{th}$ set of imputed data for $m = 1, ..., M$ and let $\bar{\phi}$ be the average of these $M$ parameter estimates. Then our variance estimate with imputation is:

$$\hat{\Sigma} = \frac{1}{M} \sum_{m=1}^M V_N(\hat{\phi}^{(m)}) + (1 + \frac{1}{M}) \sum_{m=1}^M (\hat{\phi}^{(m)} - \bar{\phi})(\hat{\phi}^{(m)} - \bar{\phi})^T \frac{M - 1}{M - 1}.$$ 

(2.26)

In our simulation studies and data example we use $M = 10$ and the last 10 sets of imputed data to compute this variance estimate.

2.4 Simulation Study

2.4.1 Simulating Cure-Rate Data

Simulating cure-rate model data presents its own challenges, and we do not give a comprehensive examination of all types through these simulation studies. Here
we detail our data simulation procedure for all of simulation results. We only consider cases that have finite $\tau$, which we set to be 20 (weeks) across all simulations to match that from the pregnancy outcomes data we examine in the section. We begin by generating covariates in $Z$ for a larger sample than we desire to account for those who will be left out due to truncation. Values for $\alpha$ are chosen to procure the desired percentage of cured individuals on average in the population, and when we refer to the percentage of cured individuals in a study we are referring to this population percentage. The values are determined through experimentation and adjustment until we reach the desired cure percentage. An individual is then either designated as cured or susceptible with probability determined from the logistic model.

We choose values for $\beta$ and then for those susceptible individuals we generate an event time. In Bender et al. (2005) they propose generating event times from $\Lambda_0^{-1}(−\log U/\exp \beta' \tilde{Z}_i)$ with $U$ a random variable with distribution $U[0,1]$. Our baseline hazard function is selected to guarantee that all event times occur before $\tau$. Specifically, we use $\Lambda_0^{-1}(t) = 20\{1 - \exp(−t)\}$. The values of $\beta$ are also selected via experimentation and adjustment so that the conditions for use of (2.10) to estimate $G$ will almost always be met.

Truncation times are generated from a uniform distribution $U[0,a]$ for some $a < 15$ chosen so that on average the desired percentage of susceptible individuals are truncated out. When we refer to percentage truncation we mean this percentage of the susceptible population who will on average be truncated out. Once the truncation times are generated, all individuals with event times less than their truncation time are removed. For the remaining individuals if we have censoring the censoring times are generated from $U[15,b]$ for some $b > \tau$ so that on average the desired percentage of the remaining individuals (including those who are cured) will have a censoring time less than $\min(T_i, \tau)$. When we refer to censoring percentage it means this percentage
of the sample who are censored. As can be seen, to facilitate the generation of our cure-rate data we chose truncation and censoring distributions which never overlap. Finally, we reduce the generated data to the desired sample size by taking the first \( N \) individuals from those who remain.

2.4.2 Simulation Results

The results of simulation experiments are provided here to show that our inference procedure does indeed account for the bias due to left-truncation. Our investigations also show its efficacy over varying sample size and percentage of cured individuals in the population. In addition, we examine the effects over varying truncation and censoring percentages.

All simulations are done with 500 trials. We include two covariates: a normal, \( N(4, 1) \), with corresponding parameters \((\alpha_1, \beta_1)\) and a Bernoulli \((p = 0.3)\) with corresponding parameters \((\alpha_2, \beta_2)\). The logistic part also includes an intercept with parameter \(\alpha_0\). For each simulation we provide the average parameter estimate over the 500 trials (“Estimate”), the sample standard deviation (“Sample SD”), the average over 500 trials of the standard errors based on our variance estimation (“SE”), and empirical coverage percentage (“Coverage”) of the nominal 95% confidence intervals. We ran simulations over two truncation percentages (10%, 20%) and two censoring percentages (0%, 20%).

In Table 2.1, we give the results of eight simulation experiments with 50% of the population cured. The weighted results are using our inference method while the unweighted results are when the \(\nu_i\)’s are forced to be 1 throughout the estimation. The weighted simulations show efficacy of our model with good parameter estimates and coverage. Standard errors can be quite larger than the sample standard deviation in some instances which is likely an effect from the use of a sandwich estimator along
with the variance due to imputation. However, it did not seem to effect the coverage much except perhaps for the Cox regression parameters. When weights are forced to 1, the inference severely underestimates the intercept in the logistic model. As truncation percentage increases, the bias gets worse as is expected.

In Table 2.2, we gives the results of eight simulation experiments with 75% of the population cured. Here we compare small ($N = 200$) versus large ($N = 1000$) sample sizes over our truncation and censoring amounts. Aside from improved variance estimation, the results show no significant improvement with the larger sample size. Furthermore, we note that even when there is a larger cured population, we still obtain good parameter estimates in both the logistic and Cox regression.

### 2.5 Application: Pregnancy Outcomes Data

Here we apply our method to data from the Organization of Teratology Information Specialists (OTIS). OTIS is a North American network of university or hospital based teratology servies that counsel between 70,000 and 100,000 pregnant women every year. From counsultations some women are asked are invited to enroll in studies in which the mothers and their babies are followed over time. Phone interviews are conducted through the length of the pregnancy along with pregnancy diaries recorded by the mother. An outcome phone interview is conducted shortly after the pregnancy ends and if it results in a live birth a dysmorphology exam is done within 6 months with follow-ups at one year and possibly later dates (Xu and Chambers (2011)). One collection of studies examines risks of autoimmune disease drugs on pregnancy outcomes including spontaneous abortion (SAB). Our data is from a combined autoimmune disease registry, and our sample includes pregnant women who entered a study between 2005 and 2012.

By definition SAB occurs within the first 20 weeks of pregnancy so our sample
consists of \( N = 964 \) women who entered the study before week 20 of their pregnancies taken from the larger sample data. There were 74 SAB events and 21 loss to follow up (including elective abortions). While for clinical pregnancies the rate of SAB is estimated to be about 12% (Wilcox et al. (1988)), in our sample it is less than 10%. This is due to the selection bias from left-truncation. This, along with knowing that any pregnancy lasting more than 20 weeks is no longer susceptible to SAB, provides support for use our cure-rate model. In Figure 2.1 we present the Kaplan-Meier curves for the full sample and just the non-cured portion.

As noted, it can be useful to determine whether the truncation distribution of the data is uniform and length-biased methods can be used. In Figure 2.2, we provide histograms of entry times for the entire sample and just the cured portion where it is evident that the distribution of times in either case is not uniform. Furthermore in 2.2.2 we discuss the two tests for uniformity of the truncations distribution. Kolmogorov-Smirnov test comparing \( Q/20 \) for the cured individuals with \( U[0, 1] \) yields \( p \)-value < 0.01 and the visual inspection of the Kaplan-Meier curves for truncation times and residual times for the non-cured portion agrees as there is a wide disparity between the two (shown in Figure 2.3).

Chambers et al. (2013), identified prior elective abortion (TAB) as a significant factor in vaccine trials. Our own analysis identified maternal age at pregnancy start to be a significant factor in logistic regression for the cure-rate. We also adjust for smoking and exposure to autoimmune disease drugs. There were 106 (11.0%) individuals with a prior TAB, maternal age ranged from 18-47 years with mean 32.5 years and standard deviation 4.9, there were 103 (10.7%) identified as smokers, 499 (51.8%) individuals in the combined exposure group for autoimmune disease drugs (Enbrel, Humira, etc.), 272 (28.2%) in the disease-matched control group, and 193 in the healthy control group (20.0%). For these covariates we provide the results
of our model regression in Table 2.3. As in the simulation results, we used ten imputed values for the censored survival times to compute the imputation variance. In addition to right-censoring there were 10 individuals with a SAB event who were interval-censored for which we took the midpoint of the interval as the event time. There was also one individual with an interval-censored TAB for which we took the midpoint of the interval as the censoring time.

The results appear reasonable with maternal age and smoking seen to increase the probability of SAB. Furthermore the comparison of the exposure group to the disease control and healthy control groups shows increased risk of SAB from exposure to autoimmune disease drugs. In addition to the intercept (which in absence of other covariates estimates SAB rate in population) our model showed maternal age and the comparison to the disease control group to be significant factors in determining the probability of SAB. The Cox regression identified prior TAB as a significant factor in the survival model. Furthermore, we see that prior TAB and smoking decrease the survival time. Interestingly, the direction of the disease control group parameter suggests that autoimmune disease drugs may delay the time of an SAB event. There is of course some concern that the standard errors might be larger than desired due to the use of the sandwich estimator and multiple imputation in computing the variance. In the discussion we outline future work to reduce the variability in the estimate.

We also include an analysis of the data using separate logistic and Cox regression models for the data in Table 2.4. We removed those individuals who were lost to follow up in this analysis. In the logistic model we get the same direction of the effects from the covariates and it does identify the same covariates as significant. The comparison does show that our model corrects for the sampling bias the intercept coefficient is lower in this unweighted model. The direction of effects of the covariates in the Cox regression is the same as in our model. However, this analysis
fails recognize prior TAB as significant as in our cure-rate model.

2.6 Discussion

In this paper we have proposed a regression model mixture cure-rate models that accommodates left-truncation in addition to right-censoring. Our inference method utilizes an augmented complete data likelihood to produce estimating equations and an EM-type algorithm to determine parameter estimates. The method gives a unique, consistent, and asymptotically normally distributed set of estimates. Through simulation we have shown the estimator to perform well over different covariate structures and truncation and censoring distributions. And importantly corrects for the selection bias due to left truncation.

It would be desirable to use an unweighted form of the complete data likelihood, as then we would have MLEs for our parameters. It should be possible to write the complete data likelihood without weights using the cure probability conditional on an individual entering the study. However, the implementation of the estimation procedure would need to be adjusted as the estimation of the logistic and Cox regressions could no longer be decoupled. Variance estimation would be more straightforward, and the Fisher information could be found directly using methods in Louis (1982). This corrects the overestimation from use of the sandwich estimator (Kauermann and Carroll (2001)). Inverse probability weights from the estimated truncation distribution may still need to be utilized to ensure a good initial guess and convergence of the algorithm but would be computationally superfluous.

A feature of our model which we did not examine is that it also works for possibly interval-censored data. Closed form estimates for proportional hazards models with purely interval-censored data can be found in Sun (2006). As we employ a multiple imputation approach, however, estimates can be obtained by imputing an event
time for interval-censored individuals along with the right-censored ones.

Acknowledgement

Chapter 2, in part, is currently being prepared for submission for publication of the material. Faig, Walter; Xu, Ronghui; Chambers, Christina. The dissertation author was the primary investigator and author of this material.
Table 2.1: $N = 200$. 50% Cured. Weighted vs. Unweighted Estimates

<table>
<thead>
<tr>
<th></th>
<th>Weighted Estimates</th>
<th></th>
<th>Unweighted Estimates</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True Value</td>
<td>Estimate</td>
<td>Sample SD</td>
<td>SE</td>
</tr>
<tr>
<td>10% Truncation, 0% Censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>1.00</td>
<td>1.01</td>
<td>0.69</td>
<td>0.69</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.39</td>
<td>-0.40</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>2.00</td>
<td>2.03</td>
<td>0.39</td>
<td>0.39</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.20</td>
<td>-0.20</td>
<td>0.12</td>
<td>0.13</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.30</td>
<td>0.31</td>
<td>0.23</td>
<td>0.37</td>
</tr>
<tr>
<td>20% Truncation, 0% Censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>1.00</td>
<td>0.95</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.39</td>
<td>-0.38</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>2.00</td>
<td>2.04</td>
<td>0.43</td>
<td>0.44</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.20</td>
<td>-0.20</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.30</td>
<td>0.31</td>
<td>0.23</td>
<td>0.37</td>
</tr>
<tr>
<td>10% Truncation, 20% Censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>1.00</td>
<td>1.03</td>
<td>0.81</td>
<td>0.94</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.39</td>
<td>-0.41</td>
<td>0.20</td>
<td>0.23</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>2.00</td>
<td>2.07</td>
<td>0.48</td>
<td>0.50</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.20</td>
<td>-0.21</td>
<td>0.12</td>
<td>0.24</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.30</td>
<td>0.33</td>
<td>0.24</td>
<td>0.45</td>
</tr>
<tr>
<td>20% Truncation, 20% Censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>1.00</td>
<td>1.05</td>
<td>0.81</td>
<td>0.91</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.39</td>
<td>-0.42</td>
<td>0.20</td>
<td>0.21</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>2.00</td>
<td>2.09</td>
<td>0.48</td>
<td>0.55</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.20</td>
<td>-0.19</td>
<td>0.13</td>
<td>0.20</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.30</td>
<td>0.33</td>
<td>0.25</td>
<td>0.49</td>
</tr>
</tbody>
</table>
Table 2.2: 75% Cured. Small vs. Large Sample Size

<table>
<thead>
<tr>
<th></th>
<th>True Value</th>
<th>Estimate</th>
<th>Sample SD</th>
<th>SE</th>
<th>Coverage</th>
<th>Estimate</th>
<th>Sample SD</th>
<th>SE</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$N = 1000$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>1.00</td>
<td>1.00</td>
<td>0.77</td>
<td>0.77</td>
<td>95.2</td>
<td>0.97</td>
<td>0.36</td>
<td>0.36</td>
<td>94.2</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.63</td>
<td>-0.64</td>
<td>0.20</td>
<td>0.20</td>
<td>95.8</td>
<td>-0.62</td>
<td>0.09</td>
<td>0.09</td>
<td>94.2</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>1.00</td>
<td>0.99</td>
<td>0.41</td>
<td>0.41</td>
<td>94.8</td>
<td>0.99</td>
<td>0.17</td>
<td>0.17</td>
<td>95.2</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.20</td>
<td>-0.21</td>
<td>0.19</td>
<td>0.23</td>
<td>93.2</td>
<td>-0.20</td>
<td>0.07</td>
<td>0.07</td>
<td>95.6</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.30</td>
<td>0.33</td>
<td>0.35</td>
<td>0.72</td>
<td>97.6</td>
<td>0.30</td>
<td>0.15</td>
<td>0.20</td>
<td>97.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>True Value</th>
<th>Estimate</th>
<th>Sample SD</th>
<th>SE</th>
<th>Coverage</th>
<th>Estimate</th>
<th>Sample SD</th>
<th>SE</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$N = 1000$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>1.00</td>
<td>1.01</td>
<td>0.85</td>
<td>0.85</td>
<td>94.2</td>
<td>1.00</td>
<td>0.38</td>
<td>0.39</td>
<td>94.6</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.63</td>
<td>-0.65</td>
<td>0.22</td>
<td>0.23</td>
<td>95.6</td>
<td>-0.64</td>
<td>0.10</td>
<td>0.11</td>
<td>95.0</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>1.00</td>
<td>0.98</td>
<td>0.42</td>
<td>0.43</td>
<td>95.0</td>
<td>0.98</td>
<td>0.17</td>
<td>0.18</td>
<td>95.2</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.20</td>
<td>-0.22</td>
<td>0.18</td>
<td>0.25</td>
<td>93.2</td>
<td>-0.21</td>
<td>0.08</td>
<td>0.10</td>
<td>96.0</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.30</td>
<td>0.29</td>
<td>0.35</td>
<td>0.81</td>
<td>97.6</td>
<td>0.29</td>
<td>0.17</td>
<td>0.24</td>
<td>98.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>True Value</th>
<th>Estimate</th>
<th>Sample SD</th>
<th>SE</th>
<th>Coverage</th>
<th>Estimate</th>
<th>Sample SD</th>
<th>SE</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$N = 1000$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>1.00</td>
<td>1.09</td>
<td>0.85</td>
<td>0.97</td>
<td>94.0</td>
<td>1.04</td>
<td>0.42</td>
<td>0.42</td>
<td>94.2</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.63</td>
<td>-0.65</td>
<td>0.21</td>
<td>0.26</td>
<td>94.4</td>
<td>-0.64</td>
<td>0.09</td>
<td>0.10</td>
<td>94.6</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>1.00</td>
<td>1.03</td>
<td>0.42</td>
<td>0.47</td>
<td>92.8</td>
<td>1.01</td>
<td>0.17</td>
<td>0.17</td>
<td>94.8</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.20</td>
<td>-0.20</td>
<td>0.19</td>
<td>0.28</td>
<td>93.0</td>
<td>-0.19</td>
<td>0.08</td>
<td>0.11</td>
<td>96.4</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.30</td>
<td>0.30</td>
<td>0.35</td>
<td>1.08</td>
<td>97.4</td>
<td>0.30</td>
<td>0.16</td>
<td>0.47</td>
<td>97.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>True Value</th>
<th>Estimate</th>
<th>Sample SD</th>
<th>SE</th>
<th>Coverage</th>
<th>Estimate</th>
<th>Sample SD</th>
<th>SE</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$N = 1000$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>1.00</td>
<td>1.09</td>
<td>0.95</td>
<td>1.05</td>
<td>94.0</td>
<td>1.05</td>
<td>0.45</td>
<td>0.48</td>
<td>94.8</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.63</td>
<td>-0.65</td>
<td>0.24</td>
<td>0.30</td>
<td>95.8</td>
<td>-0.64</td>
<td>0.10</td>
<td>0.11</td>
<td>94.6</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>1.00</td>
<td>1.02</td>
<td>0.43</td>
<td>0.60</td>
<td>93.8</td>
<td>1.00</td>
<td>0.17</td>
<td>0.23</td>
<td>94.0</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.20</td>
<td>-0.18</td>
<td>0.20</td>
<td>0.37</td>
<td>92.8</td>
<td>-0.19</td>
<td>0.18</td>
<td>0.35</td>
<td>93.2</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.30</td>
<td>0.31</td>
<td>0.37</td>
<td>1.40</td>
<td>97.8</td>
<td>0.30</td>
<td>0.18</td>
<td>0.57</td>
<td>97.8</td>
</tr>
</tbody>
</table>
### Table 2.3: Estimates for SAB data with cure-rate model

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Logistic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.52</td>
<td>1.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Maternal Age</td>
<td>0.09</td>
<td>0.03</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Prior TAB</td>
<td>-0.53</td>
<td>0.42</td>
<td>0.21</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.20</td>
<td>0.36</td>
<td>0.58</td>
</tr>
<tr>
<td>Disease Control</td>
<td>-1.34</td>
<td>0.46</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Healthy Control</td>
<td>-0.07</td>
<td>0.32</td>
<td>0.83</td>
</tr>
<tr>
<td><strong>Cox PH</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal Age</td>
<td>0.01</td>
<td>0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>Prior TAB</td>
<td>-0.61</td>
<td>0.24</td>
<td>0.01</td>
</tr>
<tr>
<td>Smoking</td>
<td>-0.46</td>
<td>0.26</td>
<td>0.08</td>
</tr>
<tr>
<td>Disease Control</td>
<td>-0.59</td>
<td>0.38</td>
<td>0.12</td>
</tr>
<tr>
<td>Healthy Control</td>
<td>0.08</td>
<td>0.29</td>
<td>0.78</td>
</tr>
</tbody>
</table>
Table 2.4: Estimates for SAB data with separate logistic and proportional hazards regressions.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Logistic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-5.10</td>
<td>0.88</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Maternal Age</td>
<td>0.08</td>
<td>0.03</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Prior TAB</td>
<td>-0.23</td>
<td>0.40</td>
<td>0.56</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.61</td>
<td>0.35</td>
<td>0.08</td>
</tr>
<tr>
<td>Disease Control</td>
<td>-1.07</td>
<td>0.45</td>
<td>0.02</td>
</tr>
<tr>
<td>Healthy Control</td>
<td>-0.14</td>
<td>0.27</td>
<td>0.61</td>
</tr>
<tr>
<td><strong>Cox PH</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal Age</td>
<td>0.02</td>
<td>0.02</td>
<td>0.47</td>
</tr>
<tr>
<td>Prior TAB</td>
<td>-0.40</td>
<td>0.39</td>
<td>0.31</td>
</tr>
<tr>
<td>Smoking</td>
<td>-0.29</td>
<td>0.34</td>
<td>0.40</td>
</tr>
<tr>
<td>Disease Control</td>
<td>-0.49</td>
<td>0.50</td>
<td>0.33</td>
</tr>
<tr>
<td>Healthy Control</td>
<td>0.26</td>
<td>0.27</td>
<td>0.33</td>
</tr>
</tbody>
</table>
Figure 2.1: Kaplan-Meier curves for the full sample data and just the non-cured individuals.
**Figure 2.2**: Truncation times for all individuals and just the cured individuals.
Figure 2.3: Kaplan-Meier estimates for truncation and residual times for the non-cured sample.
Appendix A

Longitudinal Model Notes

At the convergence of the ES algorithm \( \hat{\theta} \) solves the estimating equations

\[
\sum_{i=1}^{N} U_i(\theta) = 0,
\]

where

\[
U_i(\theta) = \begin{cases}
\sum_{t=1}^{n_i} X_{it}^\top R^{-1} \left[ y_{it}^* - X_{it}\beta - Z_{it}\{(DZ_{it}^\top)(Z_{it}DZ_{it}^\top + R)^{-1}(y_{it}^* - X_{it}\beta)}\right] \\
\sum_{t=1}^{n_i} \frac{y_{ijt}}{\sigma_j} \left\{ \frac{y_{ijt}}{\sigma_j} - x_{ijt}\beta_j + z_{ijt}[(DZ_{it}^\top)(Z_{it}DZ_{it}^\top + R)^{-1}(y_{it}^* - X_{it}\beta)]_j \right\} - 1, \\
[D - (DZ_{it}^\top)(Z_{it}DZ_{it}^\top + R)^{-1}(DZ_{it}^\top)^\top + (DZ_{it}^\top)(Z_{it}DZ_{it}^\top + R)^{-1}(y_{it}^* - X_{it}\beta) \\
(y_{it}^* - X_{it}\beta)^\top ((Z_{it}DZ_{it}^\top + R)^{-1})^\top (DZ_{it}^\top)^\top)_j - D_j,
\end{cases}
\text{for } j=1,\ldots,J.
\]

A.1 Proof of Theorem 1

Proof. Under model (1.1) and conditional on the \( b_i \)’s, \( \sum_{i=1}^{N} U_i(\theta) = 0 \) is a set of unbiased estimating equations; that is \( E\{\sum_{i=1}^{N} U_i(\theta_0)|b_i\} = 0 \). We then have \( E\{\sum_{i=1}^{N} U_i(\theta_0)\} = E\{\sum_{i=1}^{N} U_i(\theta_0)\} = 0 \). The rest of the proof follows immediately from Yuan and Jennrich (1998): basically the non-singularity of \( E\{U'_i(\theta_0)\} \) in Assumption iii. ensures that \( \theta_0 \) is the unique zero of \( E\{\sum_{i=1}^{N} U_i(\theta)\} = 0 \) in a neighborhood of \( \theta_0 \); the uniform convergence in Assumption ii. together with the fact that \( E\{\sum_{i=1}^{N} U_i(\theta_0)\} = 0 \) establishes that with probability one \( \sum_{i=1}^{N} U_i(\theta_0) = 0 \) has
a solution $\hat{\theta}$ in any neighborhood of $\theta_0$ for $N$ sufficiently large; this in turn gives the consistency.
Appendix B

Cure-Rate Model Notes

B.1 Variance Computations

Here we provide calculations giving the closed form of $\partial U / \partial \phi$. We begin with partial derivatives of $\omega_i$:

\[
\frac{\partial \omega_i}{\partial \alpha} = \omega_i(1 - \omega_i)Z_i, \quad \text{(B.1)}
\]

\[
\frac{\partial \omega_i}{\partial \beta} = \omega_i(1 - \omega_i)\{\Lambda_0(Q_i) - \Lambda_0(X_i)\} \exp(\beta' \tilde{Z}_i) \tilde{Z}_i, \quad \text{(B.2)}
\]

\[
\frac{\partial \omega_i}{\partial \lambda_k} = \omega_i(1 - \omega_i)\{I(t_k \leq Q_i) - I(t_k \leq X_i)\} \exp(\beta' \tilde{Z}_i) \quad k = 1, ..., K, \quad \text{(B.3)}
\]

\[
\frac{\partial \omega_i}{\partial G_j} = \frac{-\omega_i(1 - \omega_i)}{\nu_i} \frac{\partial \nu_i}{\partial G_j} \quad j = 1, ..., J - 1, \quad \text{(B.4)}
\]

with

\[
\frac{\partial \nu_i}{\partial G_j} = -\nu_i^2 I(q_j \leq T_i < q_{j+1}) \quad j = 1, ..., J - 1. \quad \text{(B.5)}
\]
Now $\partial U / \partial \phi$ can be partitioned as follows:

$$
\frac{\partial U}{\partial (\theta, G)} = \begin{pmatrix}
\frac{\partial U^{(\theta)}}{\partial \theta} & \frac{\partial U^{(G)}}{\partial \theta} \\
\frac{\partial U^{(\theta)}}{\partial G} & \frac{\partial U^{(G)}}{\partial G}
\end{pmatrix}
$$

(B.6)

The upper left component is the variance formula in Louis (1982) which we compute via methods used in Sy and Taylor (2000). It is comprised of the following:

$$
\frac{\partial U^{(\alpha)}}{\partial \alpha} = \sum_{i=1}^{N} \nu_i [\{\omega_i (1 - \omega_i) - p_i (1 - p_i)\} Z_i Z_i^T],
$$

(B.7)

$$
\frac{\partial U^{(\alpha)}}{\partial \beta} = \sum_{i=1}^{N} \nu_i \{\Lambda(Q_i) - \Lambda_0(X_i)\} \exp(\beta^T \tilde{Z}_i) \omega_i (1 - \omega_i) Z_i \tilde{Z}_i^T,
$$

(B.8)

$$
\frac{\partial U^{(\alpha)}}{\partial \lambda_k} = \sum_{i=1}^{N} \nu_i \exp(\beta^T \tilde{Z}_i) \{I(t_k \leq Q_i) - I(t_k \leq X_i)\} \omega_i (1 - \omega_i) Z_i
$$

$$
\quad k = 1, ..., K,
$$

(B.9)

$$
\frac{\partial U^{(\beta)}}{\partial \beta} = \left[ \sum_{i=1}^{N} \nu_i \omega_i (1 - \omega_i) \exp(\beta^T \tilde{Z}_i)^2 \{\Lambda_0(Q_i) - \Lambda_0(X_i)\}^2 \tilde{Z}_i \tilde{Z}_i^T + \right.

\quad \left. \sum_{i=1}^{N} \nu_i \omega_i \{\Lambda_0(Q_i) - \Lambda_0(X_i)\} \exp(\beta^T \tilde{Z}_i) \tilde{Z}_i \tilde{Z}_i^T \right],
$$

(B.10)

$$
\frac{\partial U^{(\beta)}}{\partial \lambda_k} = \left[ \sum_{i=1}^{N} \nu_i \exp(\beta^T \tilde{Z}_i)^2 \{I(t_k \leq Q_i) - I(t_k \leq X_i)\} \omega_i (1 - \omega_i) \{\Lambda_0(Q_i) - \Lambda_0(X_i)\} \tilde{Z}_i \right.

\quad \left. + \sum_{i=1}^{N} \nu_i \omega_i \{I(t_k \leq Q_i) - I(t_k \leq X_i)\} \exp(\beta^T \tilde{Z}_i) \tilde{Z}_i \right]
$$

$$
\quad k = 1, ..., K,
$$

(B.11)

$$
\frac{\partial U^{(\lambda_k)}}{\partial \lambda_k} = \left[ \sum_{i=1}^{N} \nu_i \omega_i (1 - \omega_i) \exp(\beta^T \tilde{Z}_i)^2 \{I(t_k \leq Q_i) - I(t_k \leq X_i)\}^2 - \right.

\quad \left. \sum_{i=1}^{N} \nu_i \delta \omega_i \{X_i = t_k\} \left( \frac{1}{\lambda_0(X_i)^2} \right) \right]
$$

$$
\quad k = 1, ..., K,
$$

(B.12)
and for \( j \neq k \):

\[
\frac{\partial U^{(\lambda_k)}}{\partial \lambda_j} = \sum_{i=1}^{N} \nu_i \omega_i (1 - \omega_i) \exp(\beta^T \tilde{Z}_i)^2 \{ I(t_k \leq Q_i) - I(t_k \leq X_i) \} \{ I(t_j \leq Q_i) - I(t_j \leq X_i) \}.
\]

(B.13)

The values of \( \frac{\partial U^{(G)}}{\partial \theta} \) are all 0 as \( U^{(G)} \) does not involve \( \theta \). \( \frac{\partial U^{(G)}}{\partial G} \) is the \((J-1) \times (J-1) \) identity matrix. The components of \( \frac{\partial U^{(\theta)}}{\partial G} \) are as follows:

\[
\frac{\partial U^{(\alpha)}}{\partial G_j} = \sum_{i=1}^{N} \left( \nu_i \frac{\partial \omega_i}{\partial G_j} + (\omega_i - p_i) \frac{\partial \nu_i}{\partial G_j} \right) Z_i \quad j = 1, \ldots, J - 1,
\]

(B.14)

\[
\frac{\partial U^{(\beta)}}{\partial G_j} = \sum_{i=1}^{N} \left( \nu_i \frac{\partial \omega_i}{\partial G_j} + \omega_i \frac{\partial \nu_i}{\partial G_j} \right) [\delta_i + \{ \Lambda_0(Q_i) - \Lambda_0(X_i) \} \exp(\beta^T \tilde{Z}_i)] \tilde{Z}_i \quad j = 1, \ldots, J - 1,
\]

(B.15)

and for each \( k = 1, \ldots, K \),

\[
\frac{\partial U^{(\lambda_k)}}{\partial G_j} = \sum_{i=1}^{N} \left( \nu_i \frac{\partial \omega_i}{\partial G_j} + \omega_i \frac{\partial \nu_i}{\partial G_j} \right) \frac{\delta_i I(X_i = t_k)}{\lambda_k} - \exp(\beta^T \tilde{Z}_i) I(Q_i < t_k \leq X_i) \]

\[ j = 1, \ldots, J - 1. \]

(B.16)

### B.2 Proofs of Asymptotic Properties

#### Proof of Theorem 2.3.1

Proof. We assume condition C2 holds, and that without loss of generality the covariate space contains zero. Although the logistic regression could have a nonzero intercept, we use a slight abuse of notation \( Z = 0 \) to mean \( \tilde{Z} = 0 \) here. This proof is follows that of the more generalized case II in Li et al. (2001). If the model is not identifiable then there exists \((p^*, S_0^*, r^*) \neq (p, S_0, r) \) so that

\[
\frac{p(Z)}{p^*(Z)} = \frac{1 - S^*(t; Z)}{1 - S(t; Z)} = \frac{1 - S_0^*(t)}{1 - S_0(t)} \frac{r^*(Z)}{r(Z)} = c(Z), \quad Z \in \mathcal{Z}, t < \tau
\]

(B.17)

For some positive function \( c(Z) \) that does not depend on \( t \).

Solving for * terms, we get the following:
\[ S_0^*(t)r^*(Z) = 1 - c(Z) + c(Z)S_0(t)r(Z) \]  \hspace{1cm} (B.18)

\[ p^*(Z) = \frac{p(Z)}{c(Z)} \]  \hspace{1cm} (B.19)

And when \( Z = 0 \), note that \( S_0^*(t) = 1 - c(0) + c(0)S_0(t) \) since \( r(0) = 1 \) so that:

\[ r^*(Z) = \frac{\log\{1 - c(Z) + c(Z)S(t; Z)\}}{\log\{1 - c(0) + c(0)S(0)\}} \]  \hspace{1cm} (B.20)

Defined the right hand side above as the appropriate function \( h(c(Z), u, c(0), r(0)) \)

where \( u = S_0(t_u) \) for some \( t_u \) such that \( S_0(t_u) < 1 \) and also \( Z \) is chosen so that \( r(Z) \neq 1 \).

Now let \( c(Z) = c(0) + \Delta \). Then for some \( c^* \) between \( c(0) \) and \( c(0) + \Delta \), the first order Taylor series of \( h \) about \( c(0) \) gives,

\[ r^*(Z) = h(c(0), u, c(0), r(Z)) + \Delta h^{(1)}(c^*, u, c(0), r(Z)) \]  \hspace{1cm} (B.21)

And through differentiation we have,

\[ h^{(1)}(c^*, u, c(0), r(Z)) = \frac{u^{r(Z)} - 1}{(1 - c^* + c^*u^{r(Z)}) \log(1 - c(0) + c(0)u)} \]  \hspace{1cm} (B.22)

Thus, there is no way for \( r^*(Z) \) to not depend on \( u \) unless \( \Delta = 0 \) and \( c(0) = 1 \). This implies \( c(Z) = c(0) + \Delta = 1 \) for all \( Z \) which gives us our contradiction. \( \Box \)

**Proof of Theorem 2.3.2**

**Proof.** Firstly, as noted \( \hat{G} \) is the empirical distribution function of the cured sample truncation times. So by the strong law of large numbers, for every \( t \), \( \hat{G}(t) \rightarrow G_0(t) \) almost surely uniformly in \( t \) (Van Der Vaart and Wellner, 1996, p. 81-82). So \( \hat{G}(t) \) is a consistent estimator for \( G_0(t) \).

To show the existence of our estimator \( \hat{\theta} \) observe that \( \ell_\nu \) is continuous and by assumption \( \alpha \) and \( \beta \) have compact support; so it remains to show that \( \hat{\Lambda}_0 \) has bounded jumps. For some \( B > 0 \) and suppose \( \hat{\Lambda}_0(X(K)) > B \) where \( X(K) \) is the largest event
time. As $\hat{\alpha}$ and $\hat{\beta}$ are bounded and as $\hat{\nu} = G^{-1}(q_i)I(q_j \leq T_i < q_{j+1})$ the $\hat{\nu}$ are bounded above by $\hat{\nu} < 1/(1/N) = N$, the weighted completed data likelihood is bounded above by $\prod_{i=1}^{N} ab\hat{\Lambda}_0(Q(x)) - \hat{\Lambda}_0(x)$ for some $0 < a, b < \infty$. Then as $B \to \infty$, and $Q(x) < C(x) < \tau_2$, the weighted complete data likelihood goes to 0 with probability 1. This gives our contradiction so we have the existence of $\hat{\theta}$. Furthermore, as $\ell_{\nu}$ is concave the estimator is unique.

In addition, we show that $\hat{\Lambda}_0(\tau_2)$ is bounded and thusly over $[\tau_1, \tau_2]$ the parameter space of $\Lambda_0$ is compact. The estimated cumulative hazard function at $\tau_2$ is:

$$\hat{\Lambda}_0(\tau_2) = \sum_{i=1}^{N} \frac{d_i I(X_i \leq \tau_2)}{\sum_{j:Q_j < t_i \leq X_j, \nu_j \omega_j \exp{\beta_j Z_j}}}$$

(B.23)

The bound in the denominator is because the $\nu_j$’s are bounded below by 1, $Z_j$ and $\beta$ have compact support, and by condition $C3$ there is positive probability of an outcome being observed so at least some of the $\omega_j$’s are 1’s. To show (B.23) is bounded we have the following limit as $N \to \infty$ from the law of large numbers:

$$\frac{1}{N} \sum_{i=1}^{N} d_i I(X_i \leq \tau_2) \to E\{\delta I(T \leq \tau_2)\} < 1.$$  

(B.24)

Also by assumption the denominator is bounded away from 0 with probability 1; so $\hat{\Lambda}_0(\tau_2)$ is bounded.

We conclude the proof following methods of Li et al. (2008) and Su and Wang (2012). $\hat{\theta}$ has compact support so by Helly’s selection theorem there exists a convergent subsequence $\hat{\theta}_{nk}$ of any subsequence $\hat{\theta}_n$. We suppose the convergent subsequence $\hat{\theta}_{nk}$ converges to $\theta^*$ and we can conclude by showing that $\theta^* = \theta_0$.

Firstly, we define the following:

$$\tilde{\Lambda}_0(t) = \sum_{i=1}^{N} \frac{d_i I(X_i \leq t)}{\sum_{j:Q_j < t_i \leq X_j, \nu_j \omega_j \exp{\beta_0 Z_j}}}$$

(B.25)

By the Glivenko-Cantelli Theorem $\tilde{\Lambda}_0$ converges uniformly to $\Lambda_0$ on $(\tau_1, \tau_2)$. Observe that $\ell_{\nu}(\hat{\alpha}, \hat{\beta}, \hat{\Lambda}_0) - \ell_{\nu}(\alpha_0, \beta_0, \Lambda_0) \geq 0$. Let $p_\theta$ be the joint density of $(Z, X, \delta)$ for parameter $\theta$. 


Then,
\[
E_{\theta_0} \{ \ell_{\nu}(\hat{\theta}) - \ell_{\nu}(\theta) \} = E_{\theta_0} \{ \ell_{\nu}(\theta^*) - \ell_{\nu}(\theta_0) \} \\
= E_{\theta_0} \{ \nu \log \frac{p^{\theta^*}}{p^{\theta_0}} \} \\
= E_{\theta_0} \{ \log \frac{p^{\theta^*}}{p^{\theta_0}} \} - E_{\theta_0} \{ \nu |Z,X,\delta} \} \\
\leq D \log E_{\theta_0} \{ \frac{p^{\theta^*}}{p^{\theta_0}} \} \quad \text{by Jensen’s inequality} \\
= D \log 1 = 0.
\]

Thus, \( p^{\theta^*} \) is equivalent to \( p^{\theta_0} \) with probability 1, and as model is identifiable we have that \( \theta^* = \theta_0 \).

\[\Box\]

**Proof of Theorem 2.3.3**

*Proof.* Here we outline the proof for the asymptotic normality. We provide the closed form of \( \frac{\partial U}{\partial \phi} \) earlier in Appendix B.1, and, along with condition \( C5 \), we note that it is continuously invertible in a neighborhood of \( \phi_0 \). The proof is similar to that in (Van Der Vaart and Wellner, 1996, p. 310-311). We denote \( \tilde{U} = E_{\phi_0}(U) \) and consider the following:

\[
\sqrt{N} \{ \tilde{U}(\hat{\phi}) - \tilde{U}(\phi_0) \} = \sqrt{N} \{ \tilde{U}(\hat{\phi}) - U(\hat{\phi}) \} + o_p(1) \\
= -\sqrt{N}(U - \tilde{U})(\phi_0) + o_p(1 + \sqrt{N}||\hat{\phi} - \phi_0||). \quad (B.27)
\]

From the differentiability of \( \tilde{U} \), the left side of (B.27) can be replaced by \( \sqrt{N} E_{\phi_0}(\frac{\partial U}{\partial \phi})(\hat{\phi} - \phi_0) + o_p(\sqrt{N}||\hat{\phi} - \phi_0||) \). From the law of large numbers and the consistency of \( \hat{\phi} \), we have that \( \sqrt{N} E_{\phi_0}(\frac{\partial U}{\partial \phi})(\hat{\phi} - \phi_0) \) converges weakly to a mean zero Gaussian process. The invertibility of \( E_{\phi_0}(\frac{\partial U}{\partial \phi}) \) along with the inverse mapping theorem then completes the proof. \[\Box\]
Bibliography


