# USING TRAJECTORIES FROM A BIVARIATE GROWTH CURVE OF COVARIATES IN A COX MODEL ANALYSIS

by

**Qianyu Dang**

B.S., Fudan University, 1995

M.S., Kansas State University, 2000

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2004

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Qianyu Dang

It was defended on

July 20th 2004

and approved by

Sati Mazumdar, Ph.D., Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Stewart Anderson, Ph.D., Associate Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Howard E. Rockette, Ph.D.,Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Lisa A. Weissfeld, Ph.D., Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Charles F. Reynolds, M.D., Professor, Department of Psychiatry, University of Pittsburgh

Dissertation Director: Sati Mazumdar, Ph.D., Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

**USING TRAJECTORIES FROM A BIVARIATE GROWTH CURVE OF COVARIATES IN A COX MODEL ANALYSIS**

Qianyu Dang, PhD

University of Pittsburgh, 2004

In many maintenance treatment trials, patients are first enrolled into an open treatment before they are randomized into treatment groups. During this period, patients are followed over time with their responses measured longitudinally. This design is very common in today's public health studies of the prevention of many diseases. Using mixed model theory, one can characterize these data using a wide array of across subject models. A state-space representation of the mixed model and use of the Kalman filter allow more flexibility in choosing the within error correlation structure even in the presence of missing and unequally spaced observations. Furthermore, using the state-space approach, one can avoid inverting large matrices resulting in efficient computations. Estimated trajectories from these models can be used as predictors in a survival analysis in judging the efficacy of the maintenance treatments. The statistical problem lies in accounting for the estimation error in these predictors. We considered a bivariate growth curve where the longitudinal responses were unequally spaced and assumed that the within subject errors followed a continuous first order autoregressive (CAR (1)) structure. A simulation study was conducted to validate the model. We developed a method where estimated random effects for each subject from a bivariate growth curve were used as predictors in the Cox proportional hazards model, using the full likelihood based on the conditional expectation of covariates to adjust for the estimation errors in the predictor variables. Simulation studies indicated that error corrected estimators for model parameters are mostly less biased when compared with the nave regression without accounting for estimation errors. These results hold true in Cox

models with one or two predictors. An illustrative example is provided with data from a maintenance treatment trial for major depression in an elderly population. A Visual Fortran 90 and a SAS IML program are developed.

## PREFACE

This dissertation is organized in the following way. Chapter 1 contains the statement of the dissertation problem, the description of the study [Maintenance Therapies in Latelife Depression (MTLD)] and the motivation for using multivariate growth curves for a trajectory analysis. In Chapter 2, we present an approach for obtaining a bivariate growth curve with unequally spaced data. The corresponding paper submitted to Communication in Statistics with the title "Modeling Unequally Spaced Bivariate Growth Curve with Kalman Filter Approach" is attached as an appendix. In Chapter 3, we review the Cox model with covariate measurement error. In Chapter 5, we formulate the details of the dissertation problem and give an example. The simulation study results are shown in Chapter 5. The conclusion and directions for future research are given in Chapter 6.

For the production of this dissertation, I owe the following people my sincere gratitude. First I wish to thank my advisor, Dr. Sati Mazumdar, for her encouragement and constant support in this research. She introduced the concept of covariates error to me and provided important guidance during the course of my dissertation research. I am very grateful to Dr. Mazumdar and Dr. Anderson's patient revision of my proposal and this dissertation. I also like to thank Dr. Anderson and Dr. Tan for their previous work on the Kalman filter which provided the foundation for this dissertation. Many valuable discussions with Dr. Anderson are indispensable for me to accomplish the algorithm of Kalman filter.

I would like to express my appreciation to the rest of my committee: Dr. Charles Reynolds, Dr. Howard Rockette and Dr. Lisa Weissfeld, for giving me valuable suggestions to improve this dissertation. Special thanks to Dr. Saul Shiffman for providing research assistantship which allowed me to attend school. Finally, I dedicate this dissertation to my wife Ying and my parents for their love and support.

## TABLE OF CONTENTS

## LIST OF TABLES

# 1.0 INTRODUCTION

In this dissertation, we modify and have reprogrammed the Kalman filter algorithm developed by Jones (1993) and Tan (1993) to fit an unequally spaced, bivariate growth model with a CAR(1) error structure. We then used the estimates of each subject's trajectories obtained from the fitted model as predictors in a Cox proportional hazards model. Essentially, we have extended the existing methods of joint modeling of survival and longitudinal data with measurement error to model survival data with the predictors' estimation errors inherited from the first step of the model. We also attempt to find the best approximation to simplify the computation. Several methods are suggested for estimation of covariates coefficients and variances in the survival analysis.

## 1.1 STATEMENT OF THE PROBLEM

In many clinical trials, when participants are recruited, they are first enrolled into an open treatment period. During that time, the patients are given a standard treatment. Only those who achieve remission or improvement from the illness are subsequently eligible to be randomized for the comparison of different maintenance treatments in a clinical trial.

Many researchers notice that patients who have quick, stable responses to standard treatment during the open treatment period are more likely to have better results later in their maintenance treatment. Consequently it is important to assess the effect of each subject's trajectory of response in the first phase on their performance later.

The basic idea is to use each patient's response profile during the acute phase of the treatment to predict his later outcomes. The initial responses are usually longitudinal mea-

surements. In addition, more than one response profile may be of importance and these profiles are likely to be correlated. In the case where more than one response is observed at each time point, a multivariate growth curve is appropriate to model these data to obtain required trajectories. Since in practice there are likely many missing observations, so that ideally, the model should be able to handle unequally spaced data. Each patient's unique multi-dimensional profile can be summarized by the estimated parameters from the growth curve. These estimated parameters can then be used as predictors for the next phase of the study. Thus, we can evaluate the effect of the first phase profile on the final outcomes. The statistical problem lies in accounting for the estimation errors in these predictors.

## 1.2 MTLD STUDY AS THE MOTIVATING EXAMPLE

The maintenance therapies in late life depression (MTLD; Reynolds et al. (1999)) study was a randomized clinical trial comparing different treatments combination for older patients aged 60 to 90 with depression. The study consisted of three phases. The first was the acute treatment phase when 187 patients with recurrent depression were enrolled and openly treated with a full-dose of nortriptyline (NT) and weekly interpersonal psychotherapy (IPT) to achieve a remission of depressive symptoms. The maximum length for this phase was 26 weeks. The patients' responses were measured by the Hamilton Depression Rating Scale (17-item). Among them, only those who achieved "full remission" (HDRS score $\leq$ 10 for 3 consecutive weeks) were eligible for the next phases which were called the continuation treatments. The continuation phase consisted of 16 weeks of therapy and drug treatment to ensure stability of remission and full recovery. Patients who remained in stable remission at the end of the second phase were randomized into the maintenance treatment phase.

We selected two measurements to represent the recovery pattern during the acute and continuation treatment. The first is the HDRS-17 score, which is the direct indicator for the treatment response. The second one is the blood test results of NT concentration, which is considered to represent the body metabolism and compliance with treatment. These two measurements were obtained at the time the participants came for their weekly treatment.

They are mostly unequally spaced since patients missed their appointments due to various reasons. In addition, these two variables are likely to be correlated.

In maintenance treatment phase, the patients were assigned to four different treatment groups: combination of NT treatment and IPT, placebo treatment with IPT, NT with medication clinic and placebo treatment with medication clinic. The treatment started 6 weeks after randomization to allow for gradual double-blind tapering of NT and IPT to the placebo and medication clinic. Patients remained in maintenance therapy for three years or until recurrence of any major depression episodes. In the end, the recurrence rates ranged from 20% to 90%. The NT combined with IPT treatment group had the lowest recurrence rate among the four treatment groups.

The maintenance phase is a survival study with an event defined as a recurrence of depression. A censored observation is dropout or the three year remission. The first two phases provide data to determine trajectories for patients' responses to standard treatment. The unique patterns from these trajectories for each patient can be used as predictors for the event of recurrence in the third phase.

# 2.0 MODELING UNEQUALLY SPACED BIVARIATE GROWTH CURVE WITH KALMAN FILTER APPROACH

## 2.1 THE BIVARIATE GROWTH CURVE MODEL

In order to estimate each patient's trajectories during the phase I and II treatment period, the bivariate longitudinal growth curve was modeled using a mixed effect modeling approach. A continuous time first order autocorrelation (CAR(1)) structure is assumed for within subject errors. For any individual i, the bivariate mixed effects model is given by

$$Y_i = X_i\boldsymbol{\alpha} + Z_i\boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i, \tag{2.1}$$

where $Y_i$ is an $n_i \times 2$ matrix of HDRS-17 score and NT levels for subject i; $\boldsymbol{\alpha}$ is the $6 \times 1$ fixed effects coefficients and $\boldsymbol{\theta}_i$ is the $4 \times 1$ vector of random effects coefficients consisting of intercepts and slopes for subject i. We assume $\boldsymbol{\theta}_i \sim N(0, B)$ where B is a $4 \times 4$ matrix and $\boldsymbol{\varepsilon}_i \sim i.i.d.\ N(0, W_i)$, where $W_i$ is the within subject covariance matrix for individual i. The design matrices for fixed and random effects are

$$\mathbf{X_i} = \begin{bmatrix} 1 & t_{i1} & t_{i1}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & t_{i1} & t_{i1}^2 \\ 1 & t_{i1} & t_{i1}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & t_{i1} & t_{i1}^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 1 & t_{in_i} & t_{in_i}^2 \end{bmatrix} \quad \mathbf{Z_i} = \begin{bmatrix} 1 & t_{i1} & 0 & 0 \\ 0 & 0 & 1 & t_{i1} \\ 1 & t_{i2} & 0 & 0 \\ 0 & 0 & 1 & t_{i2} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 1 & t_{in_i} \end{bmatrix}. \tag{2.2}$$

A multivariate Kalman recursive procedure was applied to find the exact likelihood of the model. Non-linear parameters are estimated through a nonlinear optimization program

to minimize the -2log likelihood obtained through Kalman filter procedure. However, the computations become complicated if more than one correlated response is measured at each time point. In these cases, the modeling strategy and numerical techniques are crucial in the estimation process.

## 2.2 SUMMARY OF THE PAPER SUBMITTED

In the attached paper in the appendix of this dissertation, we demonstrated the Kalman filter approach for a bivariate growth curve model of the MTLD data. In the paper, instead of the HDRS-17 score, we used the closely related general life function score (Mazumdar, Reynolds, et al. (1996)) and the NT level to form a bivariate mixed effects model, but the model structure was assumed to be the same. We considered a bivariate situation where the longitudinal responses were unequally spaced and assume that the within subject errors followed a continuous first order autoregressive (CAR(1)) structure. Compared to the traditional ML and REML solutions, the Kalman filter can handle more complicated correlated error structures within subjects, and the inference of association between two variables can be easily obtained. Furthermore, using the state-space approach, one can avoid inverting large matrices resulting in efficient computation. The bivariate setting requires large numbers of unknown parameters to be estimated simultaneously. This requires the estimation process to be carefully controlled to achieve convergence. In order to improve numerical stability for the multi-step procedure, the exploratory model fitting strategy was essential. We developed both a Visual Fortran and a SAS program for modeling such data. A simulation study was conducted to validate the model estimates.

# 3.0 LITERATURE REVIEW: COX PROPORTIONAL HAZARDS MODEL WITH COVARIATE MEASUREMENT ERROR

In this chapter, we review the recent progress in addressing covariate measurement errors in Cox proportional hazards models. In recent years, researchers have adjusted the estimates of the regression vector subject to measurement errors. Prentice (1982) introduced a model based on the expected hazard function to reduce the bias caused by measurement errors. Hu et al. (1998), De Gruttola et al. (1994) and Tsiatis et al. (2001) developed both parametric and non-parametric method to find the optimal estimates. Liu et al. (in press) extended Tsiatis's methods to covariates with correlated replicates.

Cox (1972) developed a proportional hazards mode for survival data. The basic concept of Cox model is to model the association between covariates and the "hazard"function $\lambda(t)$, which is closely related to the probability of failure at particular time t. Assuming $X_i(t)$ to be the observed covariate for subject i, the hazard function at time t is defined as:

$\lambda(T|X_i, T_i) = \lambda_0(T)exp(\beta' X_i)$,

where $\lambda_0(t)$ is defined as the baseline hazard at time t for all subjects. So let $R_i$ denote the set of individuals whose event or censored times exceed any time $t_i$ or whose censored times are equal to $t_i$, and $Q_i$ denote the set of all subsets of $d_i$ individuals from $R_i$. For each $q \in Q_i$ , $q$ is a $d_i$-tuple $(q_1, q_2, \ldots, q_{d_i})$ of individuals who might have failed at $t_i$. The log likelihood becomes

$$L(\beta) = \sum_{i=1}^{k} \frac{\lambda_0(t_i)exp(\beta'\mathbf{X_i})}{\sum_{q\in Q_i} \lambda_0(t_i)exp(\beta'\mathbf{X_q})} = \sum_{i=1}^{k} \frac{exp(\beta'\mathbf{X_i})}{\sum_{q\in Q_i} exp(\beta'\mathbf{X_q})}. \tag{3.1}$$

It is called the partial likelihood since all the baseline hazards $\lambda_0(.)$ are canceled from the likelihood function.

Other approaches such as pseudolikelihood (Prentice, 1986) and penalized likelihood (Li and Luan, 2003) have also been proposed to solve the Cox proportional hazards model.

To handle the measurement errors, Prentice(1982) assumed that the "observed" covariate $X(t)$ had no predictive value given the "true" but unobservable covariate value Z(t). This assumption can be written as:

$\lambda\{(t; Z(t), X(t)\} = \lambda\{t; Z(t)\}$.

It is also suggested that given the failure time $T \geq t$ and $X(t)$, the hazard function based on the observed covariate should be replaced by the conditional expectation of $\lambda(t, Z(t))$ as

$\lambda\{(t, X(t)\} = E_{\{T \geq t, X(t)\}} \lambda\{(t), Z(t)\} = \lambda_0(t) E_{\{T \geq t, X(t)\}} exp(\beta' Z(t))$.

Since the failure time T is a function of baseline hazard, the presence of $\{T \geq t\}$ in the conditioning event implies some dependence of the relative risk function $\lambda\{t, X(t)\}$on the baseline hazard function $\lambda_0(.)$. Hence the full hazard function needed to be used instead of the Cox partial likelihood function in the expression. We call this method the "full likelihood" approach, comparing to the partial likelihood for estimating the Cox regression coefficients.

Hu and Tsiatis (1998) applied this theory to the Cox regression model with covariates measurement error. Defining $V_i$ as the time to failure or censoring and $\Delta_i$ as the failure indicator, and assuming that the censoring time is independent of the failure time and the true covariates $Z_i$, the full likelihood of the Cox model is a function of

$$L = \prod_{i=1}^{n} [\int \{\lambda_0(V_i) e^{\beta z}\}^{\Delta_i} exp\{-\int_0^{V_i} \lambda_0(u) e^{\beta z} du\} f_{Z,X}(Z, X) dZ]; \tag{3.2}$$

where $f_{Z,X}(Z, X)$ is the joint density function of the random variable $Z$ and $X$. The joint density can be written as the product of conditional density $f_{X,Z}(X|Z)$and the marginal density $h(z)$ of $Z$. In an additive model, if both X and Z are normal, $X_i = Z_i + U_i$, and $U_i$ is i.i.d. and distributed as $N(0, \sigma_u^2)$, then $f_{X,Z}(X|Z_i) \sim N(Z_i, \sigma_u^2)$ while $\sigma_u^2$ is assumed to be known or estimated from the data.

The next step is defining the probability function of Z. Hu and Tsiatis (1998) suggested three methods: The first one was the fully parametric algorithm, which assumed that the distribution of X and Z was specified in terms of a finite number of parameters. The second was the fully nonparametric approach, which originated from Laird (1978) and Mallet(1986)'s

nonparametric MLE of a mixing distribution. This approach restricted the estimator to the class of discrete distributions with at most $n$ support points. For $j \leq n$, letting $\{s_i \ldots\ldots s_j\}$ to be the locations of these points with probability mass $m_i = Pr(Z = s_i)$. The probability function of $Z$ was

$$h(z, \nu) = \prod_{i=1}^{j} Pr(X = s_i)^{I(X=s_i)} = \prod_{i=1}^{j} m_i^{I(X=s_i)}, \tag{3.3}$$

where $\nu$ was a summarization of unknown parameters$\{j, s_i \ldots\ldots s_j, m_i \ldots\ldots m_j\}$.

A third way of defining $h(z)$ is the semi-parametric method which gives the best results in their examples. It defines

$$h(z, \nu) = \frac{1}{c}(a_0 + a_1 x + \ldots + a_k x_k)^2 n(x; \mu_x, \sigma_x^2) \tag{3.4}$$

where $c$ is a integration constant. If $k$ is 0, it reduces to the fully parametric approach as above. Alternatively, it approximates the nonparametric method if $k$ is very large. By defining $h(z)$, the likelihood function can be maximized to find MLE's by various types of numerical optimizations. Hu (1998) and De Gruttola (1994) applied the EM algorithm to estimate all the unknown parameters.

Besides the likelihood based approach, Carroll (1995) and Gleser (1990) suggested the regression calibration approach on this problem by replacing the covariates of interest $Z$ with the regression of true value $Z$ on the observed value $X$ in the standard analysis to obtain parameter estimates. For subject $i$, it is assumed that $X_i = Z_i + U_i$ where $U_i$ is the Gaussian error. Let $\sigma_z^2$ and $\sigma_u^2$ to be the unknown variance of $X_i$ and $U_i$, the expected value of $Z_i$ conditioned on $X_i$ is

$$E(Z_i|X_i) = (\frac{\sigma_z^2}{\sigma_z^2 + \sigma_u^2})X_i. \tag{3.5}$$

So the estimated calibration function

$$\hat{E}(Z_i|X_i) = (\frac{\hat{\sigma}_z^2}{\hat{\sigma}_z^2 + \hat{\sigma}_u^2})X_i. \tag{3.6}$$

By replacing $X_i$ with $\hat{E}(Z_i|X_i)$ in the partial likelihood function, the regression calibration estimators for $\beta$ can be found with maximum likelihood algorithm.

Sometimes the additive model is not appropriate to represent the association between the observed value and true value of covariates. Wulfsohn and Tsiatis (1997) developed a joint model of longitudinal covariates measured concurrently with the survival process. They used a random effects model to represent the longitudinal measurements of $X_i$ at time $t_{ij}$, that is, $X_{ij} = \theta_{0i} + \theta_{1i}t_{ij} + e_{ij}$, where $\boldsymbol{\theta}_i \sim N(\boldsymbol{\theta}, B)$. They modeled the covariates and the failure time process through the proportional hazards model where the hazard depended on the covariates through its current value. The hazard function can be represented as

$\lambda(t|\boldsymbol{\theta}_i, X_i, t_i) = \lambda(t|\boldsymbol{\theta}_i) = \lambda_0(t)exp\{\beta(\theta_{0i} + \theta_{1i}t)\}$.

Using the same notation as above, they followed Prentice's concept of conditional expected hazard function. The observed data likelihood becomes

$$L = \prod_{i=1}^{n}[\int_{-\infty}^{+\infty}\{\prod_{j=1}^{m_i} f(x_{ij}|\boldsymbol{\theta}_i, \sigma_e^2)\}f(\boldsymbol{\theta}_i|\boldsymbol{\theta}, B)f(V_i, \Delta_i|\boldsymbol{\theta}_i, \lambda_0, \beta)d\boldsymbol{\theta}_i], \tag{3.7}$$

where

$f(x_{ij}|\boldsymbol{\theta}_i, \sigma_e^2) = (2\pi\sigma_e^2)^{-\frac{1}{2}}exp\{-(x_{ij} - \theta_{0i} - \theta_{1i}t_{ij})^2/2\sigma_e^2\}$,

$f(\boldsymbol{\theta}_i|\boldsymbol{\theta}, B) = (2\pi|B|)^{-\frac{1}{2}}exp\{-(\boldsymbol{\theta}_i - \boldsymbol{\theta})'B^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\theta})/2\}$,

and

$$f(V_i, \Delta_i|\boldsymbol{\theta}_i, \lambda_0, \beta) = [\lambda_0(V_i)e^{[\beta(\theta_{0i}+\theta_{1i}V_i)]}]^{\Delta_i}exp[-\int_0^{V_i}\lambda_0(u)e^{[\beta(\theta_{0i}+\theta_{1i}u)]}du]. \tag{3.8}$$

The density function of the survival data assumes that the current value of covariates is the proper component among all covariate history to be used in the model. The authors showed that the parameter estimates could also be found by EM algorithm in the joint model.

Tsiatis and Davidian (2001) made another improvement later. They reported that the normal assumption for the random effects $\boldsymbol{\theta}$ was not even necessary when the conditional score approach of Stefanski and Carroll (1987) was applied. This is another semi-parametric approach in the sense that the random effects distribution can be left unspecified.

Recently Liu et al. (2001) extended the likelihood procedure by assuming multivariate normal distribution $mvn(\mathbf{z}; \boldsymbol{\mu}_\mathbf{z}, \boldsymbol{\Sigma}_\mathbf{z})$ for replicated measures. Two variants of likelihood-based approach were used to account for measurement errors. To make the optimization procedure to be simpler, the baseline hazards were kept to be fixed while searching for the unknown parameters. Simulation studies done by Liu et al. (in press) with one continuous covariate

measured with error and one categorical covariate measured without error indicate that the likelihood-based approach has improved bias reduction over the regression calibration approach for all covariates which show any bias. These results hold true in both main effects and interaction models used in Cox regression. Among the likelihood approaches, the semi-parametric method turned out to be more robust than the fully parametric approach.

# 4.0 USING PREDICTIONS FROM A GROWTH CURVE AS COVARIATES IN A COX MODEL ANALYSIS

In this chapter, we present the formulation of the problem of using the predictions from the growth curve described above as covariates in the Cox survival model and its application. A naive approach with Cox partial likelihood is to use the predicted values as covariates and ignore any estimation errors. To adjust for the errors of covariates inherited form the first step model estimation, we incorporated the probability function of the estimates and used the full likelihood for Cox proportional hazards model. These two methods were examined and applied to the MTLD data to analyze the relationship between recurrence of depression and the first stage recovery pattern.

The full likelihood method can be considered as an extension of the work of Hu and Tsiatis (1998), Liu, K. (2001, in press) reviewed in the previous chapter. First we will derive the formula for single predictor Cox model, and then we will extend the method to a dual-predictor mode allowing the two predictors to be correlated.

## 4.1 COX SURVIVAL MODEL WITH TRAJECTORIES FROM THE LONGITUDINAL MEASUREMENTS AS PREDICTORS

In Chapter 2, we applied the Kalman filter approach on bivariate growth curve model. The unique pattern of each individual's growth was summarized by the predicted random effects including intercepts and slopes for both variables. We also obtained the variance-covariance matrix for the random predictions and the estimated joint distribution of the random effects.

Let $\hat{\theta}_i$ denote the estimated random effects for subject $i$, which could be a scalar or

an m-dimensional vector. Since the true value of $\theta_i$ is unknown, we will use $\hat{\theta}_i$ instead as predictors in the subsequent survival analysis, but the associated estimation error has to be accounted for. Assuming $\theta_i \sim N(0, B)$, the mixed effects model in the first step will provide the maximum likelihood estimator of $\hat{\theta}_i$. Harville(1976) showed that the is also the unbiased estimator of $\theta_i$ given the observation and fixed effects. Hence $\hat{\theta}_i$ is normally distributed and centered at $\theta_i$. This method is very similar to the covariate measurement error models described in the previous chapter where the observed covariates are also functions of their true values and observation errors. Assuming for subject $i$, the true predictor $\theta_i$ is a function of the estimated $\hat{\theta}_i$ and error, we can model the covariates $\theta_i$ and failure process jointly in one model combining the proportional hazards and expression of $\hat{\theta}_i$ through $\theta_i$.

Letting $\Delta_i$ be the censoring indicator and $T_i$ be the event time, for $n$ individuals, the full likelihood is given by

$$L = \prod_{i=1}^{n} [\int_{-\infty}^{+\infty} f(T_i, \Delta_i|\theta_i, \lambda_0, \beta) f(\hat{\theta}_i|\theta_i) f(\theta_i) d\theta_i], \tag{4.1}$$

where

$$f(T_i, \Delta_i|\theta_i, \lambda_0, \beta) = [\lambda_0(T_i) e^{(\beta^T \theta_i)}]^{\Delta_i} exp[-\int_0^{T_i} \lambda_0(u) e^{(\beta^T \theta_i)} du] \tag{4.2}$$

and $\beta$ is the unknown coefficients for $\theta_i$ . If we also let $Var(\hat{\theta}_i) = \Sigma_{\hat{\theta}_i}$, then it follows that

$$f(\hat{\theta}_i|\theta_i) = ((2\pi)^m |\Sigma_{\hat{\theta}_i}|)^{-\frac{1}{2}} exp\{-\frac{1}{2}(\hat{\theta}_i - \theta_i)' \Sigma_{\hat{\theta}_i}^{-1} (\hat{\theta}_i - \theta_i)\} \tag{4.3}$$

and

$$f(\theta_i) = ((2\pi)^m |B|)^{-\frac{1}{2}} exp(-\frac{\theta_i' B^{-1} \theta_i}{2}). \tag{4.4}$$

Here $B$ is the variance matrix of $\theta_i$ and we will use the estimated variance $\hat{B}$ to approximate $B$. The accuracy of estimation can be improved by including the information of $f(\hat{B}|B)$ and $Var(\hat{B})$; If $B$ is univariate, $\hat{B}$ follows a Pearson type III distribution centered at $B$. But if $\theta_i$ is an m-dimensional vector, then $B$ becomes an $m \times m$ matrix, the computation can be very complicated and time consuming for any $m \geq 2$.

It is reasonable to assume that the underlying hazard function, $\lambda_0(t)$ has mass only at the failure times. Let $C_i$ denote the times of censoring; Also let $V_i = min(T_i, C_i)$. For m distinct failure times $t_1 \ldots\ldots t_m$, the total likelihood is

$$L = \prod_{i=1}^{n} [\int_{-\infty}^{+\infty} \prod_{j=1}^{m} \lambda_{0(t_j)}^{I(V_i=t_j)} e^{(\beta^T\theta_i)\Delta_i} e^{-e^{(\beta^T\theta_i)} \sum_{j=1}^{m} \lambda_0(t_j)^{I(V_i \geq t_j)}} f(\hat{\theta}_i|\theta_i) f(\theta_i) d\theta_i]. \tag{4.5}$$

The initial values of baseline hazards $\lambda_0(t)$ can be computed from the survival function $S_0$ estimated from the naive approach without considering estimation errors. In SAS PROC PHREG, $S_0$ is estimated through product-limit method. Any ties in the event time are handled by Breslow's approximation to the partial likelihood function. Let $C_i$ denote the set of individuals censored in the half-open interval $[t_i, t_{i+1})$, where $t_0 = 0$ and $t_{k+1} = \infty$. So if $\gamma_l$ is the censoring times in $[t_i, t_{i+1})$ where $l$ ranges over $C_i$, the likelihood function for all individuals is given by

$$L = \prod_{i=0}^{k} \{ \prod_{l \in D_i} ([S_0(t_i)]^{exp(\hat{\beta}^T\theta_l)} - [S_0(t_{i+0})]^{exp(\hat{\beta}^T\theta_l)}) \prod_{l \in C_i} [S_0(\gamma_l + 0)]^{exp(\hat{\beta}^T\theta_l)} \} \tag{4.6}$$

where $D_i$ is the set of individuals fail at time $t_i$ and $D_0$ is an empty set. The likelihood $L$ is maximized by taking $S_0(t) = S_0(t_i + 0)$ for $t_i < t < t_{i+1}$ and assuming that the probability mass to fall only on the observed event times $t_1, \ldots, t_k$. By considering a discrete model with hazard contribution $1 - \alpha_i$ at $t_i$, we take $S_0(t_i) = S_0(t_{i-1} + 0) = \prod_{j=0}^{i-1} \alpha_j$, where $\alpha_0 = 1$. Substituting all these into the likelihood, we have

$$L = \prod_{i=0}^{k} \{ \prod_{j \in D_i} (1 - \alpha_i^{exp(\hat{\beta}^T\theta_j)}) \prod_{l \in R_i - D_i} \alpha_i^{exp(\hat{\beta}^T\theta_l)} \} \tag{4.7}$$

where $R_i$ denote the risk set just before event time $t_i$ and $\hat{\beta}$ is estimated from the Cox partial likelihood, the maximum likelihood estimates of $\hat{\alpha}_i (i = 1 \ldots k)$ is the solution of

$$\sum_{j \in D_i} \frac{exp(\hat{\beta}^T\theta_j)}{1 - \hat{\alpha}_i^{exp(\hat{\beta}^T\theta_j)}} = \sum_{l \in R_i} exp(\hat{\beta}^T\theta_l). \tag{4.8}$$

If only a single event occurs at $t_i$, $\alpha_i$ can be found explicitly. Otherwise, an iterative solution is obtained by Newton algorithm. The survival function $S_0(t)$ can be estimated by

$\hat{S}_0(t) = \prod_{j=0}^{i-1} \hat{\alpha}_j$, where $t_{j-1} < t < t_j$. The estimated baseline cumulative hazard function, $\hat{\Lambda}_0$ at any event time $t$ is $\hat{\Lambda}_0 = -log(\hat{S}_0(t))$, from which $\hat{\lambda}_0(t)$ can be obtained through $\hat{\Lambda}_0(t)$.

The variance of $\hat{\beta}$ is estimated by inverting the observed information matrix. The unknown parameters are $\beta$ and the $\lambda_0(t_i)$ as its dimension could be very large. To avoid inverting a large-dimensional matrix, Hu (1998) suggested using the profile likelihood to simplify the procedure. Let $L_p(\beta)$ denote the profile likelihood, which is computed by maximizing the full likelihood given $\beta = \hat{\beta}$. The variance of $\hat{\beta}$ evaluated at $\beta = \hat{\beta}$ can be estimated by the negative inverse of the second derivative of the profile log likelihood $LL_p(\beta) = log(L_p(\beta))$ and be written as

$$Var(\hat{\beta}) = -\frac{1}{\partial^2/\partial\beta^2 LL_p(\beta)|_{\beta=\hat{\beta}}}. \tag{4.9}$$

As the analytic differentiation of $LL_p(\beta)$ could be complicated, a numerical approximation are used to simplify $\partial^2/\partial\beta^2 LL_p(\beta)$ by

$$\frac{LL_p(\hat{\beta}-\nu) - 2LL_p(\hat{\beta}) + LL_p(\hat{\beta}+\nu)}{\nu^2} + O(\nu^2), \tag{4.10}$$

where $\nu$ is some arbitrary small number. Thus the variance can be estimated by

$$Var(\hat{\beta}) = -\{\frac{LL_p(\hat{\beta}-\nu) - 2LL_p(\hat{\beta}) + LL_p(\hat{\beta}+\nu)}{\nu^2}\}^{-1}. \tag{4.11}$$

The method can be applied to both univariate and multivariate models. In the multivariate model where $\boldsymbol{\beta}$ is an m-dimensional vector, we fix $\beta_i = \hat{\beta}_i$ for $i = 1 \ldots m$ and compute $Var(\hat{\beta}_i)$ individually.

## 4.2 THE COMPUTATION AND NUMERICAL METHODS

We have developed a Visual FORTRAN 90 program for the computation. Numerical integration is performed by calling IMSL library functions attached to the software package. The IMSL Library function, DQDAG, is used in computing one-dimensional integrations. It integrates a function using a globally adaptive scheme based on Gauss-Kronrod rules. The two-dimensional integration is carried out by calling the function DQAND, which integrates

a function on a hyper-rectangle. A direct search complex algorithm of DBCPOL is used to find the minimum of -2 log likelihood. It is an algorithm designed for non-smooth functions by minimizing a function of a number of variables subject to bounds on the variables with a direct search complex algorithm. We found that it was more stable than other functions using the finite-difference gradients or Hessians in multi-dimensional searching. All the functions and variables are used in double precision.

Since there are exponential and double exponential parts in the integration as in (4.5), the computation could easily overflow or underflow if the value of $\beta$ is too large. Thus, it is ideal to keep the search range for unknown $\hat{\beta}$ as small as possible. We found that the naive approach ignoring estimation errors usually gave very good initial values for and $\lambda_0'$s, and often the search range could be safely controlled within $\pm 3\hat{\beta}_{naive}$. Even after narrowing down the search range, rescaling may still be necessary to avoid the overflow problems. Most of the times, however, the problem was caused by the double exponential part of formula (4.5). It happens more frequently in the multi-dimensional integrations.

Another purpose of data rescaling is to improve the stability of the numerical integration and non-smoothing optimization procedure. Since the optimization is designed to minimize the -2log likelihood, it is sensitive to the precision of the numerical integration. If $\beta$ is large, the result of integration could be a very small number close to the machine precision limit; thus, a small change of $\beta$ would have a large impact on the results. Under this situation, most of the numerical integration functions do not perform well, neither does the optimization process as the -2log likelihood alters dramatically for small changes of $\beta$. The detail of data rescaling and transformation will be illustrated in the examples of next section.

Unfortunately, if the dimension of the integration is more than two, the numerical procedure becomes very slow and makes impossible for the optimization procedure to find the MLE. Even in a dual predictor model, the computing time for searching for the MLE's of all the baseline hazards and $\beta$'s is about 20 to 60 hours. The computing time also depends on the searching ranges and accuracy of the initial values. Approximation methods to shorten the computing time will be demonstrated in the next chapter of the simulation studies.

## 4.3 EXAMPLE: TIME TO RECURRENCE OF DEPRESSION IN MTLD STUDY

Data from the study of maintenance therapies in late life depression (MTLD) (Reynolds et al., 1999) are used to illustrate the statistical approach. In Chapter 2, each patient's unique pattern of HDRS score and NT level measurement during the acute and continuation treatment period was summarized by the estimated random effects through a bivariate mixed effects model. The random effects include the intercept and slope for both variables, indicating each individual's starting level and recovery rate. So here $\boldsymbol{\theta}$ is a 4-dimensional vector distributed as $mvn(0, B)$. The estimate of variance-covariance matrix for the random effects is

$$\hat{B} = \begin{bmatrix} 3.056 & 3.325 & -0.004604 & 0 \\ 3.325 & 10.334 & 0 & -0.03828 \\ -0.004604 & 0 & 0.009605 & 0 \\ 0 & -0.03828 & 0 & -0.03279 \end{bmatrix}.$$

The estimates of $\hat{\boldsymbol{\theta}}_i$ were used as predictors in the Cox survival model for maintenance treatment period to predict the recurrence of depression during the trial. For simplicity, we did not include other covariates in the model.

As seen in the data, $\boldsymbol{\theta}_i$ is a $4 \times 1$ vector whose components are not independent with each other, which makes the numerical computation of model parameter estimation to be very difficult. An alternative way is to model each predictor individually first, and using only those who are significant in the final model. We followed this strategy by using intercepts and slopes of HDRS score and NT level as individual predictors in four different models, then choosing those significant to fit a multi-predictor model based on the individual p-values.

First we used the naive approach by plugging in the four predictors in the Cox survival model and ignoring all the estimation errors. The models were fitted using SAS PROC PHREG. Only the slope of HDRS was significant with p = 0.0215. The intercept of HDRS had a marginal p value of 0.1112. The two variables from NT level were both insignificant. Then we fit both the intercept and slope of HDRS in the survival model, the results showed only the slope was significant with p = 0.0916. It indicates that there could be some

correlation between the two covariates. Starting from the naive model estimates, we can conveniently obtain the baseline hazards $\lambda_0$'s at all event times from the predicted survival function.

Using $\hat{\lambda}_0(t)$ and $\hat{\beta}$ as the initial value of the unknown parameters, we first fitted random intercept and slope of HDRS score and NT level individually as single predictor in the model using likelihood function of (4.5). The standard deviations of parameter estimates were obtained from the profile likelihood.

From the results in table 1 we can see that the coefficient estimates are different form the naive approach where all predictors are used "as is" without considering the estimation error associated with them. We selected the two most significant variables and fit a Cox survival model. The bivariate algorithm was demonstrated using the random slopes of HDRS-17 and NT-level as the only two covariates. To avoid the overflow problem during two-dimensional numerical integration, the data was first transformed by multiplying the predictors by an appropriate constant. It is also necessary to transform the variance covariance matrix accordingly. For instance, if we multiply the estimates of $\hat{\boldsymbol{\beta}}$ by 10, we also need to multiply the corresponding variance covariance matrix of the random effects $\hat{B}$ and $Var(\hat{B})$ by 100. The final estimates of $\hat{\boldsymbol{\beta}}$ should be rescaled back by multiplying 10. We found it is usually safe to keep the searching range of $\hat{\boldsymbol{\beta}}$ between -0.5 and 0.5.

Now the predictor is two-dimensional where $\boldsymbol{\theta} = (\theta_1, \theta_2)$, which has a bivariate normal distribution. The computation could be simplified by writing the formula as

$$f(\theta_1, \theta_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp[-\frac{\sigma_2^2\theta_1^2 - 2\rho\sigma_1\sigma_2\theta_1\theta_2 + \sigma_1^2\theta_2^2}{2\sigma_1^2\sigma_2^2(1-\rho^2)}]. \tag{4.12}$$

The variance-covariance matrix of $(\theta_1, \theta_2)$ is $\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$ and correlation $\rho = corr(\theta_1, \theta_2) = \frac{\sigma_{12}}{\sigma_1\sigma_2}$ The formula of $f(\hat{\theta}_1, \hat{\theta}_2|\theta_1, \theta_2)$ can also be rewritten with regard to $\sigma_1, \sigma_2$ and $\rho$ as following:

$$f(\hat{\theta}_1, \hat{\theta}_2|\theta_1, \theta_2) = \\ \frac{1}{2\pi\sigma_3\sigma_4\sqrt{1-\rho_2^2}} exp[-\frac{\sigma_4^2(\hat{\theta}_1-\theta_1)^2 - 2\rho_2\sigma_3\sigma_4(\hat{\theta}_1-\theta_1)(\hat{\theta}_2-\theta_2) + \sigma_3^2(\hat{\theta}_2-\theta_2)^2}{2\sigma_3^2\sigma_4^2(1-\rho_2^2)}], \tag{4.13}$$

where variance-covariance matrix of $(\hat{\theta}_1, \hat{\theta}_2|\theta_1, \theta_2)$ is $\begin{bmatrix} \sigma_3^2 & \sigma_{34} \\ \sigma_{43} & \sigma_4^2 \end{bmatrix}$ and $\rho = corr(\hat{\theta}_1, \hat{\theta}_2|\theta_1, \theta_2) = \frac{\sigma_{34}}{\sigma_3\sigma_4}$.

The limits of the integration were determined by the value range of the parameters, usually it is not necessary to set them to be infinite. In our example after rescaling, both slopes of HDRS score and NT-level are centered at 0 and have standard deviation less than 1. Also because in the double integration, the inner part should has less impact on the results, thus the limits for the inner part was set to be [-20, +20] and [-140, +140] for outer integration. It showed little difference if larger interval range of inner integration and infinite range for outer integration were used. The variance estimates were obtained by profile likelihood described in previous section.

Table 2 showed the results for the example with two covariates; the differences between ignoring and adjusting estimation errors are bigger than the single predictor models. Notice that the coefficient for NT-level slope becomes negative after adjusting for estimation errors. This could be the indication that the adjustment makes it less biased. As we learned from chapter 2, the HDRS score and NT level showed reverse trend during the acute and continuation treatment period; the HDRS scores were decreasing while NT-levels were increasing for most patients when they were achieving remission. Thus a negative coefficient of NT-level will make more sense since it implies higher positive slope will reduce the risk of recurrence of depression.

Nevertheless, due to the increase of estimation variance, adjusting for the estimation error did not change much for the significance for the predictors. Since it is very difficult for us to verify the estimates from the real data, the model validation will be completed in the next chapter through several simulation studies.

# 5.0 SIMULATION STUDIES

In this chapter, we will use data simulation to investigate potential biases caused by covariates estimation errors in a Cox proportional hazards model, and the effectiveness of our approach to correct these biases. These simulations are implemented to mimic real applications in the set of both single and multi-predictor models. The computation program was executed on a PC with Pentium IV processor. The simulations were performed in SAS IML environment while the analysis was programmed with FORTRAN 90 language.

There are two objectives for the simulation studies:

1. to validate approximations to handle the nuisance parameters with the intention of reducing computing time;
2. to evaluate the effectiveness of our full likelihood approach on bias reduction in both single and dual predictor Cox model.

## 5.1 APPROXIMATIONS WITH FIXED BASELINE HAZARDS

In my simulation, every set of data has 100 subjects and about 50 different event times. Therefore the algorithm needed to find out the MLE for 50 unknown parameters to maximize the total likelihood. The procedure is very unstable and time consuming, and the variance of final estimates has to be obtained through profile likelihood to avoid inverting a large dimension matrix.

From the discussion in last chapter, we know that the numerical problem in our full likelihood approach is that it requires intensive numerical computation. Part of it is caused by estimating large numbers of baseline hazard $\lambda_0$ at all event times. But since $\hat{\lambda}_0$'s are

usually not of interest, we can consider them as nuisance parameters in the full likelihood function. One way to handle nuisance parameter is replacing it by an estimator thereof (Burguette,1982). The estimator of $\lambda_0$ can be easily obtained through the survival function estimates of naive approach described in chapter 4. Thus only the regression coefficient is left to be estimated. In order to validate this approximation approach, a simulation study of 500 datasets with single predictor was carried out to find out the difference of estimates with and without $\lambda_0$ to be fixed at $\hat{\lambda}_{naive}$.

To estimate the standard deviation of $\hat{\beta}$, the full likelihood approach uses profile likelihood by fixing $\beta$ at $\hat{\beta}$ as described above. But if all the $\lambda_0$'s are replaced by their estimators, the only unknown parameters are the $\beta$'s, then the variance of $\hat{\beta}$ may be estimated by inverting the observed information matrix. Let $L(\beta)$ to be the log likelihood of $\beta$, the variance can be estimated by the negative inverse of the second derivative of $L(\beta)$, which is:

$$Var(\hat{\beta}) = -\frac{1}{\partial^2/\partial\beta^2 L(\beta)|_{\beta=\hat{\beta}}}. \tag{5.1}$$

There is a simple approach to approximate the analytic differentiation of $L(\beta)$ by define an arbitrarily small number $\nu$, then

$$\partial^2/\partial\beta^2 L(\beta) \approx \frac{L(\hat{\beta}-\nu) - 2L(\hat{\beta}) + L(\hat{\beta}+\nu)}{\nu^2} + O(\nu^2). \tag{5.2}$$

Thus the variance becomes

$$Var(\hat{\beta}) \approx -\{\frac{L(\hat{\beta}-\nu) - 2L(\hat{\beta}) + L(\hat{\beta}+\nu)}{\nu^2}\}^{-1}. \tag{5.3}$$

There were 500 Monte Carlo datasets generated with each has sample size of n = 100. The first step was generating a linear growth curve for each sample in the baseline. The number of observations was uniformly distributed between 2 and 12. The time interval between observations had an exponential distribution with scale parameter a = 0.1. The random effect was standard Gaussian multiplied by the time interval between observations. For simplicity, there were no fixed effects in the model, and the errors within each subject were standard normal and had an AR(1) correlation structure with coefficient $\Phi = 0.3$. The growth curve was fitted using univariate random effects model. The estimated random coefficients and their variance were outputted to be used in the next step. Subsequently,

the survival dataset was simulated based on the true values of random effects from the growth curve. The failure time T, given the true random effect X, was generated from the exponential distribution with hazard $e^{0.1\times X}$ . The censoring time C, was created from the exponential distribution with mean of 1. The true parameter value for predictor X in the Cox survival model is $\beta_X = 0.1$. No replicate measurements were generated.

The standard deviation of $\hat{\beta}$ were estimated from observed information matrix if $\lambda_0$ was fixed, and by profile likelihood if otherwise.

The simulation results in table 3 shows that very little differences (less than 3%) in parameter estimation were observed if the baseline hazards were fixed, so that this approximation is valid under above situation. The computation time was reduced from 5 days to about 15 minutes for these 500 sets of single predictor models. In the dual-predictor simulation studies, it took about 20 hours searching for the MLE's using the full likelihood algorithm for a single dataset, while the computing time was only about 15 to 20 minutes after applying fixed baseline hazards approximations. The results were also not very close

## 5.2 SIMULATION WITH SINGLE PREDICTOR

Three sets of simulation studies were conducted to investigate the performance of the full likelihood approach on correcting biases caused by estimation errors from predictors. Each set of simulation has 500 datasets and each dataset has 100 subjects. The rest of the procedure followed the same steps illustrated in section 5.1 except for the time $\delta_t$ between consecutive observations within each subjects in the step of growth curve simulation. We found that as $\delta_t$ becoming smaller, the parameters of the growth curve estimates became less accurate, as a consequence the resulted estimates in the Cox model became more biased. We set $\delta_t$ distributed as one parameter exponential with scale parameter to be 0.1, 0.5 and 1 in three simulated datasets. In all the simulations, the true values of $\beta$ were set to be 0.1.

As in the MTLD data analysis, we used the Cox model to model the association between time to event and predicted single random effect from the growth curve. The naive approach ignored all the estimation errors and considered the predicted random effect as true value while the full likelihood approach would account for such errors. The analysis was done with

approximations of fixed baseline hazards, and the standard deviations of estimates were obtained from the approximation method showed in section 5.1.

From the results in table 4 we can see that by ignoring the estimation errors carried from the model in the first step, the estimates of $\beta$ were biased towards 0. This nature is a typical bias of attenuation to the null for additive covariates errors. Let the true value of predictor to be Z, and the error from model estimation be U, then the estimates from the first step model X = Z + U. The extent of bias depends on the accuracy of first step model estimation, in other words, the magnitude of estimation errors. If we use X instead of Z as the predictor, the attenuation effect is unavoidable. This is a very similar situation as ignoring the measurement error of covariates described by Hu, Tsiatis and Liu et al. For all three simulations above, the full likelihood approach provided less biased estimates for $\beta$. In different settings, was about 39.5%, 27% and 25% less biased than the naive approach estimates. The estimated standard deviations of $\beta$ were also higher than the naive approach. This could the result of additional variance induced by the variability of $\hat{\theta}$. Again, similar results were also shown in the measurement error simulations described by Liu and Mazumdar (in press).

### 5.3 SIMULATION WITH DUAL-PREDICTOR

The simulation used to assess the dual-predictor model is very similar to those of the single predictor models. 500 datasets were generated with 100 subjects in each set. In the first step of the bivariate growth curve, each subject has between 2 and 12 observations. The time intervals between observations followed an exponential distribution with mean of 0.1. The two random effects were distributed as bivariate normal with a variance-covariance matrix $\begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}$ and mean $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$. The curve was fitted by the bivariate growth curve procedure with random effects. The predictions of each subject's random effects were used as predictors in the survival model of the second step. Similar to the univariate model, the survival data was generated by randomly taking failure time T from a exponential distribution with hazard $e^{\beta_1 X_1 + \beta_2 X_2}$. The true values were set as $\beta_1 = \beta_2 = 0.1$. The censoring time C was

also distributed as a standard exponential with mean 1. About half of the subjects were censored.

From the results in Table 5, we can see that compared to the single predictor models, the bias caused by ignoring estimation errors in dual predictor model tend to be larger and not biased towards the null. The results agree with the bivariate analysis results from the MTLD data. A possible explanation is that in the naive approach for dual predictor examples, we ignored both the error from model estimation and the correlation between two covariates. Generally, the Cox model assumes that all the predictors are fixed and are measured without error. Violation of such assumptions could alter the estimates of coefficients to be either higher or lower determined by the joint effects of the two factors. Nevertheless, the full likelihood approach we developed provides better estimates of coefficients, which are closer to their true values. The variance of $\hat{\boldsymbol{\beta}}$ measures the variability of the estimates from $500\hat{\boldsymbol{\beta}}$ 's. The result indicates that by accounting for the errors and correlations, the full likelihood estimates are generally more accurate with lower variability. Thus, correction for the estimation error can improve both the accuracy and precision of model estimation.

The standard error estimates of the coefficient from the full likelihood are also lower. This is different from the univariate examples where the standard errors were slightly higher. This could be the results of including correlation information between two predictors as well as larger bias correction than the single predictor models. Liu and Mazumdar (2001, in press) also reported that correction of the measurement error can also improve the precision of parameter estimates, and this could offset the variance caused by the additional parameters in the full likelihood function. Further investigation should be done by changing the covariance structure between the two variables and see how that would affect the precision and accuracy of estimators.

Generally in the Cox proportional hazards models where the predictors are actually parameters estimated from other models, ignoring the estimation errors from the first step model can create bias in the coefficient estimation. The bias is usually larger in multi-predictor models, where the naive approach ignores the estimation errors and the possible correlation between covariates. Under both situations, our simulation studies show that the full likelihood based approach can reduce the bias effectively in coefficients estimates for

covariates of the Cox regression model by accounting for the estimation errors. In the dual predictor models, since we include both errors and the correlation between two predictors, the improvements are even larger than the single predictor models.

## 6.0 DISCUSSION AND DIRECTIONS FOR FURTHER RESEARCH

Multi-step models are used often in data analysis of clinical studies when the studies are divided into several different stages. It is a simple way to find out causal relationship between outcomes of different stage. De Gruttola and Tu (1994) pointed out that the bias of survival estimates in the second step model was mainly caused by estimation errors of the first step outcomes. Hence we developed a full likelihood methodology to account for these errors and correct the bias. The simulation studies and the application on MTLD data show that our methodology effectively reduced the bias of parameter estimation. If we can obtain the variance estimates of the random variance, $Var(\hat{B})$, and replace the distribution of B by $f(\hat{B}|B)$just like what we have done to the parameter $\theta$, we could further reduce the bias. However, the computation is difficult and need further investigation.

One disadvantage of a two-step model is its low efficiency. In recent years, there are discussions about improving the statistical efficiency and reducing bias by using a joint model instead of a two-step model. Most of the studies have been done on the joint model of longitudinal data and survival analysis. The joint model usually relies on the joint likelihood or conditional likelihood to estimate all the parameters, so both models has to be likelihood based and easy to combine. For instance, it is very challenging to combine the Cox regression and growth curve model with a Kalman filter approach due to the iterative nature of the Kalman recursion. In addition, the joint model usually has more unknown parameters to be estimated simultaneously through the optimization algorithm. Sometimes such approach is not practical in the multivariate models due to the instability in the numerical procedure. So far the EM algorithm is best way for parameter estimation, but if there is no close-form MLE for some parameters, a Newton-Raphson algorithm has to be called at each iteration step. On the other hand, the full likelihood approach actually does not require any specific modeling

in the first step as long as the solution can provide parameter estimators and their errors. Thus, the two step approach could be easily extended to handle various parametric models in the first step such as non-linear mixed models. The only requirements are parameters and variances estimators and their exact distributions. This flexibility is a major advantage a two step method over a joint modeling approach.

Alternatively, it is obvious that the joint model is more appropriate if the outcomes are collected concurrently. De Gruttola and Tu (1994) indicated that under that situation part of the bias was also caused by possible in-time underlying association between two outcomes. So the joint model would be more capable to handle such problems using conditional likelihood function. For example, if the longitudinal measurements of disease progression and survival are taken concurrently, and the survival is depended on the progression, then the two-step model is not appropriate since it could not account for the "informative early failure bias" (patients who have severe progression are likely to fail early, but they are not properly weighted in the first step model) in the growth curve estimation. Otherwise if the longitudinal measurements and survival are not taken concurrently, such bias does not exist and the two step model could be better with all the advantages listed above.

If other methods such as ML or REML are used to solve the growth curve model in the first step model, one has to make sure the model estimates are in correct form. For instance, in SAS PROC MIXED, the standard error estimates of random predictions $\hat{\theta}$ are actually in the form of $Var(\hat{\theta} - \theta)$. The expression of $Var(\hat{\theta} - \theta)$ is

$$Var(\hat{\theta} - \theta) = B - BZ_i'V_i^{-1}Z_iB + BZ_i'V_i^{-1}X_i(\sum_1^m X_i'V_i^{-1}X_i)^{-1}X_i'V_i^{-1}X_iB = B - Var(\hat{\theta}_i), \tag{6.1}$$

where $V_i$ is the variance of the outcome $y_i$ and $Z_i$ is the design matrix for the random effects. It is obvious that

$$Var(\hat{\theta}_i) = B - Var(\hat{\theta} - \theta). \tag{6.2}$$

If B is unknown then $\hat{B}$ can be used instead.

One problem in the current setting of full likelihood approach in multi-predictor Cox model is the numerical multiple integration. The FORTRAN subroutine was not very efficient on multi-dimensional numerical integration. In the simulated examples, even with

fixed baseline hazard approximation, it took about 15 minutes for the program to find out the MLE for $\hat{\boldsymbol{\beta}}$. So it is practically almost impossible to computing models with three or more correlated predictors. Also in the high-dimension integration, it is very difficult to find the appropriate transformation for all the variables and the variance-covariance matrix to avoid the numerical overflow problem. One possible solution for this problem is using Monte Carlo (MC) or Markov chain Monte Carlo (MCMC) algorithm to speed up the process with some technical details need to be completed.

There are also several directions for future research on this topic. The survival outcome is currently modeled by the popular Cox proportional hazards model, but many other survival models can be used likewise. In some occasions, the parametric survival model is more suitable than the proportional hazards model. These models are less widely used than Cox model in clinical data analysis but still conceptually attractive. One advantage is its flexibility on defining the transformed failure time as exponential, Weibull, lognormal, loglogistic, and gamma distributions. The model can also be solved using maximum likelihood estimates, so it should not be very difficult to follow a similar procedure to adjust for the estimation error carried by the predictors. In addition, since the likelihoods of these parametric models do not contain the baseline hazards, the procedure could become much easier. The formula of (4.1) is still valid but we need to replace (4.2) the proportional hazard function with the parametric survival likelihood function without the baseline hazards $\lambda_0$. A simulation study can be used to compare the performance of the full likelihood approach on different types of parametric models in the second step.

Table 1: REGRESSION COEFFICIENTS AND STANDARD ERRORS FROM A SINGLE PREDICTOR COX MODEL

Regression coefficients and standard errors from single predictor Cox model for MTLD data; the predictors are random predictions of intercept and slope of HDRS score and NT level from univariate growth curve; each was modeled individually as single predictor for time to recurrence of depression.

| Method | Predicted trajectories from HDRS growth model | | Predicted trajectories from NT-level growth model | |
|---|---|---|---|---|
| | Intercept (SE) (p-value) | Slope (SE) (p-value) | Intercept (SE) (p-value) | Slope (SE) (p-value) |
| Estimated $\hat{\beta}$ from naive approach ignoring estimation errors | 0.07448 (0.04604) (0.1112) | 0.04380 (0.01904) (0.0215) | -24.7310 (16.5187) (0.4615) | -4.5154 (6.1319) (0.1346) |
| Estimated $\hat{\beta}$ from full likelihood approach | 0.08811 (0.08445) (0.1712) | 0.07004 (0.03172) (0.02083) | -39.5818 (26.3467) (0.5128) | -3.7444 (4.7875) (0.1425) |

Table 2: REGRESSION COEFFICIENTS AND STANDARD ERRORS FROM THE DUAL-PREDICTOR COX MODEL

Regression coefficients and standard errors from the dual-predictor Cox model for MTLD data: the predictors are random predictions of slope of HDRS score and NT level from bivariate growth curve; both were modeled together for time to recurrence of depression

| Method | Predicted random slope of HDRS (SE) | Predicted random slope of NT-level (SE) |
|---|---|---|
| Estimated $\hat{\beta}$ from naive approach ignoring estimation errors | 0.04729 (0.02078) | 2.6565 (6.4471) |
| Estimated $\hat{\beta}$ from full likelihood adjusted for estimation errors | 0.1647 (0.06054) | -14.0847 (11.1203) |

Table 3: SIMULATION RESULTS FOR ASSESS EFFECT OF FIXED BASELINE HAZARDS IN SINGLE PREDICTOR COX MODEL ESTIMATES

Simulation results for assess effect of fixed baseline hazards in single predictor Cox model estimates, the true value for $\beta$ is 0.1.

| Method | Mean of the estimates for $\hat{\beta}$ (SE) | Variance of 500 estimates of $\hat{\beta}$ |
|---|---|---|
| Estimates includes baseline hazards | 0.07641 (0.02805) | 0.03279 |
| Fixed baseline hazards | 0.07434 (0.02754) | 0.03384 |
| Difference | 0.002068 (0.0051) | -0.00105 |

Table 4: SIMULATION RESULTS FOR SINGLE PREDICTOR COX MODELS

Simulation results for assessing effectiveness of correcting biases caused by estimation errors from single predictor in Cox regression model, the true values for all coefficients are 0.1.

| | Scale parameter of exponential distribution of $\delta_t$ | Mean of $\hat{\beta}$ from naive approach ignoring estimation errors<br><br>(SE) | Mean of $\hat{\beta}$ from full likelihood approach adjusted for estimation errors<br>(SE) |
|---|---|---|---|
| Simulation1 | 0.1 | 0.04484<br>(0.1529) | 0.07434<br>(0.2754) |
| Simulation2 | 0.5 | 0.06384<br>(0.1904) | 0.08774<br>(0.3261) |
| Simulation3 | 1 | 0.07539<br>(0.1778) | 0.1005<br>(0.2639) |

Table 5: SIMULATION RESULTS FOR DUAL PREDICTOR COX MODELS

Simulation results for assessing effectiveness of correcting biases caused by estimation errors from dual predictor in Cox regression model, the true values for both coefficients are 0.1.

| | Mean of $\hat{\beta}$ from naive approach ignoring estimation errors (SE) | Mean of $\hat{\beta}$ from full likelihood approach adjusted for estimation errors (SE) |
|---|---|---|
| $\hat{\beta}_1$ | -0.02231 (0.6813) | 0.1352 (0.5848) |
| $\hat{\beta}_2$ | 0.2598 (1.0732) | 0.09321 (0.9073) |
| S.E. of 500 $\hat{\beta}_1$ | 0.3518 | 0.009886 |
| S.E. of 500 $\hat{\beta}_2$ | 0.5524 | 0.01731 |

# APPENDIX

## MODELING UNEQUALLY SPACED BIVARIATE GROWTH CURVE WITH KALMAN FILTER APPROACH

qianyu_dang_2004_appendix.pdf

## BIBLIOGRAPHY


Burguete, J., R. Gallant & G. Souza (1982). On unification of the asymptotic theory of nonlinear econometric models. *Econometric Reviews.* 1, 151-190.

Carroll, R. J., Ruppert, D. & Stefanski, L. A. (1995). *Measurement error in nonlinear models.* London: Chapman and Hall.

Cox, D.R. (1972).Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B.* 34, 187-220.

De Gruttola, V. & Tu, X. M. (1994). Modeling progression of cd4-lymphocyte count and its relationship to survival time. *Biometrics.* 50, 1003-1014.

Glester, L. J. (1990). Improvement of the naive approach to estimation in non-linear errors-in-variables regression models. *In Statistical Analysis of Measurement Error Models and Application, P. J. Brown and W. A. Fuller, editors.* American Mathematics Society, Providence.

Harville, D.A. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *Annals of Statistics.* 4, 384-395.

Hu., P., Tsiatis, A. A. & Davidian, M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrika.* 54, 1047-1419..

Jones, R.H. (1993). *Longitudinal Data with Serial Correlation: A State-space Approach.* First edition, Chapman and Hall, London.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association.* 73, 805-811.

Li, H. & Luan, Y. (2003). Kernel Cox Regression models for linking gene expression profiles to censored survival data. *Pacific Symposium of Biocomputing.* 8, 65-76.

Liu, Kenneth (2001). Cox regression analysis with measurement error in covariates. *Doctoral Dissertation, University of Pittsburgh.*

Liu, K., Mazumdar, S., Stone, R. A., Dew, M.A., Houck, R. H., Reynolds, C. F. (2000). Accounting for covariate measurement error in a Cox model analysis of recurrence of depression. *Journal of Psychiatric Research.* 35, 177-185.

Liu, K., Mazumdar, S., Stone, M.A., Houck, R. H., Reynolds, C. F. (In press). Covariate measurement error in a Cox model analysis: a simulation study of recurrence of depression. *Communications in Statistics: Simulation and Computation.*

Mallet, A. (1986). A maximum likelihood estimation method for random coefficient regression models. *Biometrika.* 73, 645-656.

Mazumdar, S. , Reynolds, C. , Houck, P. et al. (1996). Quality of life in elderly patients with recurrent major depression: a factor analysis of the general life functioning scale. *Psychiatry Research.* 63, 183-190.

Prentice, R.L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika.* 69, 331-342.

Prentice, R.L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika.* 73, 1-11.

Reynolds C.F., Frank E., Perel J.M., Imber S.D., Cornes C., Miller M.D., et al. (1999). Nortriptyline and interpersonal psychotherapy as maintenance therapies for recurrent major depression. *Journal of American Medical Association.* 281, 39-45.

Stefanski, L. A. & Carroll, R. J. (1987). Conditional scores and optimal scores in generalized linear measurement error models. *Biometrika.* 74, 703-716.

Tan, L., (1993). A multivariate growth curve model with random effects and car(1) errors. *Doctoral Dissertation, University of Pittsburgh.*

Tsiatis, A. A. & Davidian. M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika.* 88, 447-458.

Wulfsohn, M.S. & Tsiatis, A.A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics.* 53, 330-339.