THE FLORIDA STATE UNIVERSITY
COLLEGE OF ARTS AND SCIENCES

SEMIPARAMETRIC BAYESIAN SURVIVAL ANALYSIS VIA
TRANSFORM-BOTH-SIDES MODEL

By

JIANCHANG LIN

A Dissertation submitted to the
Department of Statistics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Spring Semester, 2010
The members of the committee approve the dissertation of Jianchang Lin defended on May 19, 2011.

Debajyoti Sinha  
Professor Directing Dissertation

Yi Zhou  
University Representative

Dan McGee  
Committee Member

Xu-Feng Niu  
Committee Member

Stuart Lipsitz  
Committee Member

Approved:

Dan McGee, Chair, Department of Statistics

Joseph Travis, Dean, College of Arts and Sciences

The Graduate School has verified and approved the above-named committee members.
# TABLE OF CONTENTS

List of Tables ........................................ iv
List of Figures .............................................. v
Abstract .................................................... vi

1. Introduction ................................................. 1

2. Survival Analysis via Transform-Both-Sides Model ................. 3
   2.1 Semiparametric Model .................................. 3
   2.2 Model Estimation and Inference ....................... 5
   2.3 Semiparametric Bayesian ............................. 6
   2.4 Model for Multivariate Survival Data .............. 8

3. Simulation Studies ......................................... 10

4. Application ................................................ 14
   4.1 Data Example 1 ....................................... 14
   4.2 Data Example 2 ....................................... 19

5. Discussion .................................................. 22

6. Future Work: Bayesian Regularized Median Regression ........ 24
   6.1 Introduction ......................................... 24
   6.2 Bayesian Hierarchical Model ....................... 25
   6.3 Preliminary Simulation Study ....................... 26

A. The Proofs of Theorems .................................. 30

REFERENCES ................................................ 31
### LIST OF TABLES

3.1 Results of simulation study under Exponential model: Monte Carlo approximation of the sampling mean and Mean Square Error (MSE) of different estimators of known $\beta_1 = 1$ ................................. 11

3.2 Results of simulation study under Pareto model: Monte Carlo approximation of the sampling mean and Mean Square Error (MSE) of different estimators of known $\beta_1 = 1$ ................................. 11

3.3 Quantiles of 10 copies of datasets simulated from the prior predictive models ................................. 13

4.1 Pointwise and 95% interval (within parenthesis) estimators of regression parameters for the small cell lung cancer study under different procedures ................................. 19

4.2 Pointwise (standard-error/posterior standard-deviation) estimators of regression parameters for DRS data under different methods ................................. 20
# LIST OF FIGURES

4.1 Plots of residuals versus the age at entry and the estimated median survival time using parametric TBS model for lung cancer data .......................... 16

4.2 Q-Q plots of residuals under parametric TBS model for small-cell lung cancer data ................................................................. 17

4.3 Plots of observed survival times versus Age with three estimated quartile functions for two treatment arms ......................................................... 18

6.1 Probability density function for Extreme Value variate $\epsilon: \lambda, \beta$ .................. 29
ABSTRACT

We propose a new semiparametric survival model with a log-linear median regression function as an useful alternative to the popular Cox’s (1972) models and linear transformation models (Cheng et al., 1995). Compared to existing semiparametric models, our models have many practical advantages, including the interpretation of regression parameters via median, ability to incorporate heteroscedasticity, the ease of prior elicitation and computation of Bayesian estimators.

Our Bayesian estimation method is also extended to multivariate survival model with symmetric random effects distribution. Our multivariate survival model has same covariate effects on marginal (population average) as well as conditional (given random effects) median survival time.

Our other aim is to develop a Bayesian simultaneous variable selection and estimation of median regression for skewed response variable. Our hierarchical Bayesian model can incorporates advantages of Lasso penalty albeit for skewed and heteroscedastic response variable. Some preliminary simulation studies have been conducted to compare the performance of proposed model and existing frequentist median lasso. Considering the estimation bias and Mean squared error, our proposed model performs as good as and, in some scenarios, better than competing frequentist estimators.

We illustrate our approaches and model diagnostics via reanalysis of some real life clinical studies including a small-cell lung cancer study and a retinopathy study.
CHAPTER 1

Introduction

The semiparametric models such as Cox’s (1972) [1] proportional hazards model and linear transformation models (Cheng et al., 1995, 1997; Fine et al., 1998) [2] [3] [4] and its special cases (e.g., accelerated failure time model) are very popular for modeling effects of covariates on survival response. For example, the main aim of a semiparametric model for a two-arm randomized trial for small cell lung-cancer (SCLC) patients (Ying et al., 1995) [5] is to express the effects of treatment arm and age at entry on time to death (survival time). Often, there is substantial data information available about the median. Previous semiparametric models for survival data do not focus on the effects of covariates on the median and other quantiles. Several authors including Ying et al (1995) [5] gave among others some compelling arguments put forward by in favor of focusing on the quantiles of the survival time for modeling and reporting of analysis results. The effect of treatment and age on the quantiles including median time to death is useful for describing covariate effects.

Particularly for Bayesian survival analysis, medians and other quantiles are natural choices for elicitation of experts opinions. Clinical experts on SCLC are likely to have useful prior information/opinions about survival quantiles (say, the median) and the changes in the median for varying covariate values. However, semiparametric Bayesian models for survival data, possibly with the exception of Kottas and Gelfand (2001) [6], are either based on covariate effects on the hazard ratio (see Ibrahim et al., 2001) [7] or on the mean survival time (e.g., Kuo and Mallick, 1997; Walker and Mallick, 1999; Hanson and Johnson, 2002; Hanson, 2006) [8] [9] [10] [11]. At least in two-arm cancer clinical trials, the determination of a clinically significant difference and subsequent evaluation of power of the trial, even for frequentist trial designs, are often based on the prior evaluation of the median for the control arm as well as the clinically significant effect of covariates (e.g., treatment) on median
survival time. In this paper we propose a novel semiparametric model with interpretable regression effects on the median. We show that this wide class of semiparametric models has many desirable properties including model identifiability and non-monotone hazards. Unlike previous methods for Bayesian survival analysis, our models accommodate the situation when the location/median as well the scale and shape of the survival distribution are affected by the covariate. We present Bayesian methods for analyzing univariate survival under our semiparametric model. We also show that our method can be naturally extended to multivariate survival model with associated Bayesian analysis. Unlike some of the previous methods for median regression, our model does not require all quantile functions below the median to be linear.

In chapter 2, we introduce our new semiparametric survival model using the same generalized Box-Cox transformation (Box and Cox 1964) [12] to symmetry on the log($T_i$) as well as on the log-linear median regression function. We also show the desirable properties of our large class of survival models, including expressions for other quantile functions (beside median). We present the likelihood, suitable nonparametric prior processes and MCMC (Markov Chain Monte Carlo) tools to estimate the model parameters using a semiparametric Bayesian approach. Extension of our semiparametric model to handle multivariate survival data is discussed in the chapter 2. In Chapter 3, our simulation studies reveal that estimators based on our model have better small sample performance and more robustness properties compared to competing methods for median regressions including the estimators of Portnoy (2003) [13]. In chapter 4, we consider the SCLC trial to demonstrate how our model can facilitate the determination of prior distributions based on prior opinions about two quantiles (or survival probabilities at two pre-determined time-points) of a random patient with known age and treatment arm. We also illustrate the practical utility of our Bayesian methods and model diagnostics via analyzing the SCLC study (univariate response), followed by the analysis of the bivariate survival data from the Diabetic Retinopathy Study (Huster et al., 1989) [14]. We demonstrate that our data analysis methods have the advantages of simplicity, applicability to a very large class of survival models, support of the usual justification of the Bayesian paradigm, and ease of computation and implementation. Some final remarks and future work discussion are in chapter 5.
CHAPTER 2
Survival Analysis via Transform-Both-Sides Model

2.1 Semiparametric Model

Let $T_i$ be the survival time of subject $i = 1, \ldots, n$ and let $Z_i = (1, Z_{i1}, \ldots, Z_{ip})'$ be the corresponding vector of $p$ time-constant covariates along with the intercept term. The transformation model (Cheng et al., 1995, 1997) [2] [3] assumes that

$$\eta(T_i) = \gamma'Z_i + e_i,$$

where $\eta$ is a monotone transformation, $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_p)$ is a regression parameter, and $e_i$ is unspecified error variable with common density $f_e(\cdot)$ free of covariate $Z_i$. If $E(e_i) = 0$, then (2.1) gives a mean regression model for $E[\eta(T)|Z_i]$. Usually the distribution $f_e(\cdot)$ of $e_i$ is assumed to be a member of some parametric family with location 0 and with shape and scale free of $Z_i$. Important special cases of (2.1) are the accelerated failure time model (AFT) when $\eta = \log$, the proportional odds model when $e_i$ is logistic, and Cox’s model (1972) [1] when $f_e$ is the extreme-value density.

In this paper, we use an extension of the Box-Cox power family (Box and Cox, 1964) [12], namely the monotone power transformation $g_\lambda(y)$ (Bickel and Doksum, 1981) [15]:

$$g_\lambda(y) = \text{Sgn}(y)|y|^\lambda \text{ for } \lambda > 0,$$

where $\text{Sgn}(y) = -1$ for $y < 0$ and $\text{Sgn}(y) = +1$ otherwise. We assume that under an optimal $\lambda$, the transformed survival time $g_\lambda(\log(T_i))$ is symmetric unimodal with median $g_\lambda(\beta'Z_i) = g_\lambda(M_i)$, that is,

$$g_\lambda(\log(T_i)) = g_\lambda(M_i) + \varepsilon_i,$$

where $\varepsilon_i$ are iid with density $f_\varepsilon(\cdot)$ unimodal and symmetric around zero, $M_i = \beta'Z_i$, and $\beta$ is the vector of regression parameters. Carroll and Ruppert (1984) [16], Fitzmaurice
et al. (2007) [17], among others proposed parametric version of the transform-both-sides (TBS) regression model for an uncensored continuous response with the original Box-Cox transformation (Box and Cox, 1964) [12] and $N(0, \sigma^2)$ density for error $f_\epsilon(\cdot)$.

For a monotone transformation $g_\lambda(y)$ in (2.2) is with derivative $g'_\lambda(y) = \lambda|y|^{\lambda-1}$, $M_i = \beta'Z_i$ is the median of log($T_i$) because $P[\log(T_i) > M_i] = P[g_\lambda(\log(T_i)) > g_\lambda(M_i)] = 1/2$. As a consequence, the survival time $T_i$ has a log-linear median regression function $Q_2(Z_i) = \exp(M_i) = \exp(\beta'Z_i)$. For the SCLC study, this expression for the median implies that the ratio of medians of two patients (of the same age $z_2$) from two treatment arms is $Q_2(z_1 = 1, z_2)/Q_2(z_1 = 0, z_2) = \exp(\beta_1)$. We also get the similar straightforward interpretation of $\exp(\beta_2)$ as the relative change in the median for a unit increase in age.

The following theorem shows that the parameter $\lambda$ and the density $f_\epsilon$ of (6.5) are also identifiable, in the sense that for any survival time following (6.5), there is an unique $\lambda$ for which $g_\lambda(\log(T_i))$ has a symmetric unimodal distribution.

Theorem 1: For the model in (6.5) if there is another triplet $(\lambda^*, \beta^*, f_\epsilon^*)$ for which $g_{\lambda^*}(\log(T)) = g_{\lambda^*}(\beta^*) + e^*$, then $\lambda = \lambda^*$, $\beta = \beta^*$ and $f_\epsilon = f_\epsilon^*$.

The proof of Theorem 1 is in the Appendix. Similar to the transformation model of (2.1), we can rewrite the TBS model of (6.5) as

$$\log(T_i) = M_i + e_i,$$

(2.4)

where the asymmetric density function of the regression error $e_i$ of (2.4) is given by $f_\epsilon(u|Z_i) = f_\epsilon(g_\lambda(M_i + u) - g_\lambda(M_i)) g'_\lambda(M_i + u)$, where $g'_\lambda(y) = \lambda|y|^{\lambda-1}$. It is clear that unlike the usual assumption of the transformation model of (2.1), the median as well as the shape and scale of the error density $f_\epsilon(\cdot|Z_i)$ in (2.4) depend on the covariate $Z_i$. This allows our model to be useful for dealing with heteroscedasticity. The Bayesian median regression models of Kottas and Gelfand (2001)[6] have the linear representation of (2.4) with nonparametric asymmetric $f_\epsilon(u)$ with median 0 and free of covariate $Z$, and $f_\epsilon(|u|) = f_\epsilon(-\gamma|u|)$ for same scale $\gamma > 0$. Unlike their model, the covariate $Z$ does affect the scale and shape of the $f_\epsilon$ in our TBS models, and $f_\epsilon(u)$ for $u < 0$ is not a rescaling of $f_\epsilon(|u|)$.

A parametric log-normal model with location $M(Z) = \beta'Z$ for log($T$) is a special case of (6.5) with $\lambda = 1$ and $F_\epsilon$ being $N(0, \sigma^2)$. The hazard function $h(t|Z) = -\frac{d}{dt} \log\{P[T > t|Z]\}$ of (6.5) can be non-monotone; for example, a log-normal model has non-monotone hazard.
Although the model in (6.5) apparently focuses on modeling the median, we can easily obtain other quantiles of $\log(T)$ from (6.5) because $P[g_\lambda(\log(T)) > g_\lambda(M) + \epsilon^*_\alpha | Z] = \alpha$ for $\alpha \in (0, 1)$, where $\epsilon^*_\alpha$ is the $\alpha$-quantile for $f_\epsilon(\cdot)$ with $P[\epsilon > \epsilon^*_\alpha] = \alpha$. The regression function $Q_\alpha(Z)$ for the $\alpha$-quantile of $T$ is given by

$$Q_\alpha(Z) = \exp(M^*_\alpha(Z)) = \exp[g_{\lambda}^{-1}\{g_\lambda(\beta'Z) + \epsilon^*_\alpha\}] .$$

(2.5)

For $\alpha = 0.5$, we have $\epsilon^*_{0.5} = 0$ and get the linear median function $\beta'Z$ for $\log(T_i)$ in (6.5). This makes our model very convenient for simultaneously estimating all important quantiles of $T_i$ using the estimates of $(\beta, \lambda, \epsilon^*_\alpha)$. However, unlike the existing methods including those of Portnoy (2003) [13] and Peng and Huang (2008) [18], $Q_\alpha(Z)$ of our model are not linear in covariate $Z$ unless $\alpha = 0.5$ (median). The Bayesian models of Kottas and Gelfand (2001) [6] also have linear quantile functions $M_\alpha(Z) = \beta'_\alpha Z$ for all $1 > \alpha > 0$, and they are parallel to each other (with only the intercept of $\beta_\alpha$ different for different $\alpha \in (0, 1)$).

The expression of (2.5) for our model also implies that $Q_\alpha(Z_i) \leq Q_\alpha(Z_j) \iff Q_{\alpha'}(Z_i) \leq Q_{\alpha'}(Z_j)$ for all $\alpha, \alpha' \in (0, 1)$. This means that under this model of (6.5), ordering between two patients’ median survival times implies uniform ordering between their corresponding survival functions over the entire time-axis. This property is similar to Cox’s model where ordering between two hazards (as well as survival functions) remain the same over the entire time-axis.

### 2.2 Model Estimation and Inference

We observe $(t_{i0}, \delta_i)$ for $i = 1, \ldots, n$, where $t_{i0} = T_i \wedge C_i$ is the observed survival time and the censoring indicator $\delta_i = 1$ for $T_i = t_{i0}$ and 0 otherwise. It is assumed that $T_i$ and the random censoring time $C_i$ are conditionally independent given covariate $Z_i$. Given the observed data vector $(t_0, \delta^*)$ with $t_0 = (t_{i0}, \ldots, t_{n0})$ and $\delta^* = (\delta_1, \ldots, \delta_n)$, the likelihood function under our TBS model of (6.5) is as follows:

$$L(\beta, \lambda, f_\epsilon | t_0, \delta^*) = \prod_{i=1}^{n} \left[ \lambda | y_i |^{\lambda-1} f_\epsilon(\omega_i) \right]^{\delta_i} \left[ 1 - F_\epsilon(\omega_i) \right]^{1-\delta_i} ,$$

(2.6)

where $\omega_i = g_\lambda(y_i) - g_\lambda(\beta'Z_i)$ with $y_i = \log(t_{i0})$, $F_\epsilon(\omega) = \int_{-\infty}^{\omega} f_\epsilon(u)du$ is the cdf of the unimodal symmetric density function $f_\epsilon$.

In general, for the parametric case, any unimodal symmetric distribution, such as the logistic, can be used for $F_\epsilon$. For a $N(0, \sigma^2)$ density for $\epsilon$, the likelihood $L(\beta, \lambda, \sigma | t_0, \delta^*)$ is the
same as (2.6) with \( f_\epsilon(w) \) and \( F_\epsilon(w) \) respectively replaced by the density \( \phi_\sigma(w) \) and \( \Phi_\sigma(w) \) respectively. The corresponding posterior is

\[
p(\beta, \lambda, \sigma | t_0, \delta^*) \propto L(\beta, \lambda, \sigma | t_0, \delta^*) \pi(\beta, \lambda, \sigma),
\]
where \( \pi(\beta, \lambda, \sigma) \) is the joint prior density based on the available prior information. Markov Chain Monte Carlo (MCMC) samples from this joint posterior can be used to implement a parametric Bayesian analysis. The log-likelihood function of the (Gaussian) parametric TBS model is

\[
\ell(\beta, \lambda, \sigma | t_0, \delta^*) = r \log(\lambda) + \sum_{i=1}^{n} \left[ \delta_i \log \phi_\sigma(\omega_i) + \delta_i (\lambda - 1) \log(|y_i|) + (1 - \delta_i) \log \Phi_\sigma(\omega_i) \right],
\]

where \( r = \sum_i^n \delta_i \) and \( \Phi_\sigma(\omega) = 1 - \Phi_\sigma(\omega) \). The maximum likelihood estimator of \((\beta, \lambda, \sigma)\) is obtained via maximizing (2.7) using Newton-Raphson (NR) iterations. Under mild regularity conditions, the maximum likelihood estimators (MLE) of \( \beta \) in (6.5) (as well as the parametric Bayes estimator) is consistent and asymptotically efficient based on the regular large sample theory for MLE. Note that, when the symmetric distribution of \( g_\lambda(Y_i) \) is not \( N(g_\lambda(\beta'Z_i), \sigma^2) \), the parametric MLE \( \hat{\beta} \) based on Gaussian \( \epsilon \) yields a quasi-likelihood estimator of \( \beta \) and we can estimate the variance of \( \hat{\beta} \) using the so-called ”sandwich” variance estimator (White 1982) [19]. The loss of efficiency of this estimator under non-Gaussian symmetric model is unknown and is beyond the scope of this paper.

### 2.3 Semiparametric Bayesian

Any parametric assumption about \( f_\epsilon \) is deemed as a restrictive parametric assumption for some data examples in practice. In the semiparametric model of (6.5), the unimodal symmetric distribution \( F_\epsilon \) of error \( \epsilon_i \) is assumed unknown. The semiparametric likelihood of this model is given as

\[
L(\beta, \lambda, F_\epsilon | t_0, \delta^*) = \prod_{i=1}^{n} \left[ \lambda |y_i|^{\lambda-1} dF_\epsilon(\omega_i) \right]^{\delta_i} \left[ F_\epsilon(\omega_i) \right]^{1-\delta_i},
\]

where \( F_\epsilon(u) = 1 - F_\epsilon(u) \). For semiparametric maximum likelihood estimation, the likelihood of (2.8) is maximized with respect to the restriction that \( F_\epsilon \) is the cdf of a unimodal distribution symmetric around 0. The regularity conditions and asymptotic issues for semiparametric maximum likelihood estimation of (2.8) are nontrivial and beyond the scope of this paper. For semiparametric Bayesian analysis, we need the posterior

\[
p(\beta, \lambda, F_\epsilon | t_0, \delta^*) \propto L(\beta, \lambda, F_\epsilon | t_0, \delta^*) \pi_1(\beta) \pi_2(\lambda) \pi_3(F_\epsilon),
\]
where $\pi_1$, $\pi_2$ and $\pi_3$ are independent priors of $\beta$, $\lambda$ and $F_\epsilon$. This uses the simplifying, however reasonable, assumption that the prior opinions about these three quantities can be either obtained or specified independently.

We now introduce a nonparametric prior $\pi_3$ for $F_\epsilon$ defined over the space of symmetric unimodal distribution functions. For this purpose, we use the result that any symmetric unimodal distribution $F_\epsilon$ can be expressed as a scale-mixture of uniform random variables

$$F_\epsilon(u) = \int_0^\infty \zeta(u|\theta) \ dG(\theta)$$

(2.10)

for some mixing distribution $G(\theta)$ (Feller, 1971, p.158) [20], where $\zeta(u|\theta)$ for $\theta > 0$ is the uniform distribution with mean zero and support $(-\theta, +\theta)$. We use the Dirichlet process (Ferguson 1973) $[21]$ $G \sim DP(G_0, \alpha)$ prior for the unknown scale-mixing distribution $G(\theta)$ of (2.10) to define a nonparametric prior for the random (unknown) unimodal symmetric distribution $F_\epsilon$. The Dirichlet process (DP) is characterized by the known "prior guess" $G_0$ (the prior expectation of $G$), and a positive scalar parameter $\alpha$, precision around prior mean/guess $G_0$. The known mean $G_0$ of the random mixing density $G$ can be chosen appropriately to assure a desired prior mean/guess $F_*$ for unknown $F_\epsilon$. Using a result by Khintchine (1938) [22], when the density $f_*(\cdot)$ and its derivative $f'_*(\cdot)$ exist, the density $G'_0(\theta)$ of $G(\theta)$ is given as

$$G'_0(\theta) = -2\theta f'_*(\theta) \text{ for } \theta > 0 .$$

(2.11)

For example, to obtain an approximate double exponential ($Dexpo(\gamma)$) prior mean density $f_*(\epsilon) = \frac{1}{2}\gamma \exp(-\gamma|\epsilon|)$ for the regression error density $f_\epsilon$, using (2.11), we need to choose $G_0(\theta|\gamma)$ as $Gamma(2, \gamma)$ with density $G'_0(\theta|\gamma) = \gamma^{2}\theta \exp(-\gamma\theta)$. The precision parameter $\alpha$ also determines the degree of belief about how close $F_\epsilon$ should be to its prior guess $F_*$. When $\alpha$ is large enough, the unknown nonparametric $F_\epsilon$ is very close to its prior mean/guess $F_*(\cdot|\gamma)$ (which is pre-specified and often parametric). When $\alpha$ is small, the corresponding Bayes estimators of the regression parameters are expected to be very close to the semiparametric likelihood estimator of TBS model with unknown symmetric $F_\epsilon$. The details of specifications of the hyperparameters of the priors $\pi_1$, $\pi_2$ and $\pi_3$ in (2.9) are provided later in the data analysis section.
2.4 Model for Multivariate Survival Data

Multivariate survival data arise often from survival studies when survival time \( T_{ij} \) for subject \( j = 1, \ldots, m \) within cluster \( i = 1, \ldots, n \) (for example, siblings within family and patients within centers) are expected to have within cluster associations due to heterogeneity among clusters. The semiparametric model of (6.5) can be extended to address multivariate survival data using cluster-specific random effects \( W_i \). The multivariate model for survival time \( T_{ij} \) given the covariate \( Z_{ij} \) and unobservable cluster-effect \( W_i \) is

\[
g_\lambda(\log(T_{ij})) = g_\lambda(\beta'Z_{ij}) + W_i + \varepsilon_{ij}, \tag{2.12}
\]

where \( \varepsilon_{ij} \) are iid symmetric unimodal \( F_\varepsilon(\cdot) \) and \( W_i \) are iid with known common density \( f_W(\cdot|\sigma_1) \) with unknown random-effects parameter \( \sigma_1 \). To assure identifiability of the model parameters and conditional median (given \( W_i \)) of \( \log(T_{ij}) \) to be \( \beta'Z_{ij} + W_i \), we need to assume that the median of \( f_W(\cdot|\sigma_1) \) to be zero. We further need to assume that the distribution \( f_W(\cdot|\sigma_1) \) is symmetric around 0 to ensure that the marginal median of \( \log(T_{ij}) \), after integrating out the random cluster-effect \( W_i \), is equal to \( \beta'Z_{ij} \), that is

\[
P[\log(T_{ij}) > \beta'Z_{ij} | \beta, \lambda, F_\varepsilon, f_W] = \frac{1}{2}. \tag{2.13}
\]

The proof of this (omitted here) is based on the facts that convolution of two independent random variables symmetric around 0 is also symmetric around zero, and we need \( W_i + \varepsilon_{ij} \) to have symmetric density with median 0 to ensure (2.13). A possible choice for the random effects distribution of \( W_i \) is \( N(0, \sigma_1^2) \), where \( \sigma_1 \) measures the degree of the heterogeneity of the cluster-effects. However, there are other parametric symmetric choices for \( f_W(\cdot|\sigma_1) \) including double-exponential, Cauchy and even multi-modal symmetric densities. The expression for the marginal median of \( T_{ij} \) in (2.13) shows that, unlike the attenuation of the marginal covariate-effects under the frailty-random effects model (Oakes, 1989) \[23\] for multivariate survival data, there is no attenuation of the marginal effect of covariate after integrating out the cluster-effects \( W_i \) from the random-effects model of (2.12). This desirable property also facilitates the convenient elicitation of priors for \( \beta \) and \( \lambda \) based on the experts’ opinions about the change in median and the quantiles of the marginal survival times.

The semiparametric likelihood for the multivariate model of (2.12) is

\[
L(\beta, \lambda, F_\varepsilon, W, \sigma_1 | y_0, \delta^*) = \prod_{i=1}^{n} f_W(W_i|\sigma_1) \prod_{j=1}^{m} \{\lambda|y_{ij}|^{\lambda-1} \, dF_\varepsilon(e_{ij})\}^{\delta_{ij}} \{F_\varepsilon(e_{ij})\}^{\delta_{ij}}, \tag{2.14}
\]
where $y_0$ is the set of observed log-survival times $y_{ij} = \log(\text{Min}(T_{ij}, C_{ij}))$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$; $C_{ij}$ are the censoring times (independent of $T_{ij}$); $\delta^*$ is the set of censoring indicators $\delta_{ij}$; $\bar{\delta}_{ij} = 1 - \delta_{ij}$, and $e_{ij} = g_{\lambda}(y_{ij}) - g_{\lambda}(\beta'Z_{ij})$. For semiparametric Bayesian analysis, we use MCMC samples from the posterior

$$p(\beta, \lambda, F_0, \sigma_1|t_0, \delta^*) \propto L(\beta, \lambda, F_0, W, \sigma_1|t_0, \delta^*)\pi_1(\beta)\pi_2(\lambda)\pi_3(F_0)\pi_4(\sigma_1) , \quad (2.15)$$

where $\pi_1$, $\pi_2$ and $\pi_4$ are independent priors of $\beta$, $\lambda$ and $\sigma_1$, and $\pi_3(\cdot|G_0, \alpha)$ is the Dirichlet mixture (of uniform) prior process for symmetric and unimodal $F_0$. This uses the simplifying, however reasonable, assumption that the prior opinions about these four quantities can be either obtained or specified independently.
CHAPTER 3

Simulation Studies

Through several simulation studies, we evaluated the robustness of the maximum likelihood estimators (MLE) and of the semiparametric Bayesian with DP mixture process for $F_\epsilon$ based on the TBS model of (6.5) with Gaussian $f_\epsilon$, and compare the results to the estimators based on the competing method of Portnoy (2003) [13]. We performed the simulation study with different specifications of the conditional distribution of $\log(T_i)$ given $Z_i$, namely Exponential and Pareto distributions. In both simulation models, the modeling assumptions of (6.5) are not valid.

We set the median of $Y = \log(T)$ given $Z$ to be $M(Z) = \beta_0 + \beta_1 Z = 6.5 + 1.0 Z$ with $\beta_0 = 6.5$ and $\beta_1 = 1.0$; we consider the sample sizes $n = 80, 160$, and $320$, where $Z$ can take four possible values $0, 0.5, 1.0$, and $1.5$, in equal proportions for each simulated data set. We generate at least 5000 replicates of the simulated data-set at each sample size for each distribution of $T$ considered in the study. Number of replicates of datasets for different sample sizes may vary to assure that the Monte Carlo variability of the approximate bias and MSE of the regression estimates will be smaller than 0.01.

When the conditional distribution of $T$ given $Z$ is Exponential, the skewness of $\log(T) = Y$ is relatively modest. The independent censoring variable $C$ was generated from an Exponential density, where the rate parameter $k$ was chosen to obtain various proportions of censoring. For example, the choice of $k = \log(2)/30$ results in about 20% of the subjects being censored when $n = 320$. Our estimator based on (6.5) has a disadvantage due to fact that no transformation $g_\lambda(Y)$ can have an exact symmetric distribution. From the simulation results of Table 3, we can see that MLE of $\beta_1$ based on (6.5) show little bias for $n = 80, 160$ and $320$. Further, the approximate MSE of Portnoy’s estimators $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ are about 70% larger than MSE of the MLE and Bayers estimators from TBS model. The semiparametric
Bayesian estimator of $\beta_1$ has a much smaller MSE compared to the MLE, while its bias is comparable to those of its competitors. Due to small biases, the approximate sampling variances of the estimates of $\beta_1$ are very close to the corresponding MSE values for all sample sizes.

Table 3.1: Results of simulation study under Exponential model: Monte Carlo approximation of the sampling mean and Mean Square Error (MSE) of different estimators of known $\beta_1 = 1$

<table>
<thead>
<tr>
<th>Sample-size</th>
<th>TBS MLE</th>
<th>SP Bayes</th>
<th>Portnoy’s estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>MSE</td>
<td>Mean</td>
</tr>
<tr>
<td>80</td>
<td>0.91</td>
<td>2.66</td>
<td>0.92</td>
</tr>
<tr>
<td>160</td>
<td>0.97</td>
<td>1.35</td>
<td>1.08</td>
</tr>
<tr>
<td>320</td>
<td>0.94</td>
<td>0.69</td>
<td>0.93</td>
</tr>
</tbody>
</table>

In Table 4, we present the results of the simulation study when $T$ has a Pareto distribution with scale parameter equal to 1 and median $\exp(6.5 + 1 \times 1.5)$. For a Pareto distribution, $\log(T) = Y$ has an extremely skewed and heavy-tailed density and $g_\lambda(Y)$ is skewed for all values of $\lambda$. As we expect, the MLE $\hat{\beta}_1$ based on Gaussian TBS model has slightly higher bias than that of Portnoy’s estimator because the parametric modeling assumption of (6.5) is clearly not valid for the Pareto simulation model. The semiparametric Bayesian estimator performs as good as Portnoy’s estimator regarding bias. At the same time, it has a much smaller MSE than other competing estimates.

Table 3.2: Results of simulation study under Pareto model: Monte Carlo approximation of the sampling mean and Mean Square Error (MSE) of different estimators of known $\beta_1 = 1$

<table>
<thead>
<tr>
<th>Sample-size</th>
<th>TBS’s MLE</th>
<th>SP Bayes</th>
<th>Portnoy’s estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>MSE</td>
<td>Mean</td>
</tr>
<tr>
<td>80</td>
<td>1.00</td>
<td>12.69</td>
<td>1.03</td>
</tr>
<tr>
<td>160</td>
<td>0.99</td>
<td>6.00</td>
<td>1.01</td>
</tr>
<tr>
<td>320</td>
<td>0.95</td>
<td>2.75</td>
<td>1.02</td>
</tr>
</tbody>
</table>

To investigate the effects of different choices of the priors, we conduct the following simulation study of our prior predictive model. The true median survival time here is set to be $\exp(6.5 + 1 \times 1.5) = 2980.96$ for $z = 1.5$. The priors used here are: $\beta_0 \sim N(6, 10)$,
\( \beta_1 \sim N(0, 1) \). This prior for \( \beta_1 \) implies that there is almost 5% prior probability that median can change by a factor larger than 7.4 for unit change in \( z \). This prior is very non-informative about the effect of covariate because it allows high degree of covariate, however, the prior is also somewhat skeptical because the prior of covariate is somewhat concentrated around the prior guess of no-covariate effect. If a Bayes method can demonstrate good performance for detecting covariate-effect with this prior, that will signify that even a skeptical prior may not hinder the Bayes method’s ability to detect the covariate effect. The priors of a Bayes model involving multiple parameters is much better described via a summary of the prior predictions of the observables/responses. For this purpose of demonstrating the non-informative nature of our prior predictive model, we sample/simulate 500 copies of the parameters from the priors model, then we evaluate the values of various summary statistics of 500 survival times simulated from each simulated set of parameters. The quantiles of 10 (randomly selected) of these 500 datasets from prior predictive models are reported in the Table 5. The high variability of the quantiles of these 500 datasets indicate that our prior predictive model is very non-informative and can cover a wide range of survival patterns. Given that the true median of the underlying simulation model is \( \approx 2981 \) for \( z = 1.5 \), we can see that our prior predictive model is actually far from being accurate in predicting the true median. This also shows that the ranges of our prior predictions are also much higher than the true 5-percentile point of 12884 for \( z = 1.5 \) under the simulation model. In practice, we expect to use a much useful and somewhat informative prior predictive model using often available information about a range of responses (even under a prior model with skeptical view about the covariate effect).

In summary, when the distribution of \( \log(T_i) \) after an optimal transformation has a moderate degree of asymmetry, the estimator of \( \beta_1 \) based on parameteric version of (6.5) has finite sample bias very similar to that of Portnoy (2003)’s [13] estimator. More importantly, the MLE and Bayes estimators have much smaller MSE compared to that of Portnoy (2003)’s [13] estimator. Thus, the precision of our estimators based on TBS are better even when the underlying assumption of (6.5) is not entirely valid. The semiparametric Bayes estimators have the best precisions and smallest biases among all of its competitors. In this situation, the relative bias is almost negligible compared to the reduction in the MSE when using TBS model. However, the MLE’s performance depends on the degree of symmetry of the distribution of \( g_\lambda(Y_i) \) under optimal \( \lambda \). For the sake of brevity, we do not present
Table 3.3: Quantiles of 10 copies of datasets simulated from the prior predictive models

<table>
<thead>
<tr>
<th></th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.69</td>
<td>152.42</td>
<td>2160.05</td>
<td>76215.88</td>
<td>2.930955 × 10^8</td>
</tr>
<tr>
<td>2</td>
<td>4.63</td>
<td>157.08</td>
<td>3127.49</td>
<td>53758.61</td>
<td>2.572787 × 10^7</td>
</tr>
<tr>
<td>3</td>
<td>5.21</td>
<td>114.82</td>
<td>2740.81</td>
<td>60917.41</td>
<td>5.524163 × 10^7</td>
</tr>
<tr>
<td>4</td>
<td>4.17</td>
<td>133.84</td>
<td>2570.37</td>
<td>61349.55</td>
<td>1.041445 × 10^8</td>
</tr>
<tr>
<td>5</td>
<td>7.03</td>
<td>140.28</td>
<td>1638.64</td>
<td>55417.74</td>
<td>5.335909 × 10^7</td>
</tr>
<tr>
<td>6</td>
<td>5.68</td>
<td>210.62</td>
<td>4051.06</td>
<td>89269.98</td>
<td>1.652006 × 10^8</td>
</tr>
<tr>
<td>7</td>
<td>5.15</td>
<td>91.37</td>
<td>2010.46</td>
<td>68037.51</td>
<td>1.809062 × 10^7</td>
</tr>
<tr>
<td>8</td>
<td>3.94</td>
<td>137.22</td>
<td>2916.45</td>
<td>92941.55</td>
<td>8.267014 × 10^6</td>
</tr>
<tr>
<td>9</td>
<td>3.60</td>
<td>79.57</td>
<td>2034.13</td>
<td>56994.82</td>
<td>2.308840 × 10^7</td>
</tr>
<tr>
<td>10</td>
<td>4.65</td>
<td>106.06</td>
<td>2540.70</td>
<td>39685.28</td>
<td>9.764055 × 10^6</td>
</tr>
</tbody>
</table>

Any simulation study of the comparison of the MLE and Bayes’ estimators with Portnoy’s estimator when the data is simulated from TBS model in (6.5), because the MLE (and as a consequence Bayes’ estimator) is known to be the most efficient estimator when the modeling assumption is correct.
CHAPTER 4

Application

4.1 Data Example 1

Here we analyze the data set from the randomized cross-over trial of Etoposide (E) and Cisplatin (C) for small cell lung cancer patients (Ying et al., 1995) [5]; 62 cancer patients ($Z_1 = 0$) are randomized to arm A (C followed by E) and 59 patients ($Z_1 = 1$) to arm B (E followed by C). Apart from treatment indicator $Z_1$, another covariate is the patient’s age at entry ($Z_2$). Each survival time was either observed ($\delta_i = 1$) or administratively censored ($\delta_i = 0$). To evaluate the age-adjusted treatment difference, we consider the regression function $M_i = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i}$. The parametric maximum likelihood estimates of the regression parameters under TBS model (6.5) with Gaussian $F_\epsilon$ are given by $\hat{\beta}_0 = 7.455$, $\hat{\beta}_1 = -0.402$, $\hat{\beta}_2 = -0.014$.

Next we perform a parametric Bayesian analysis using the TBS model of (6.5) with parametric $N(0, \sigma^2)$ density for $F_\epsilon$. One major advantage of the TBS model for Bayesian analysis is that the priors for the parameters ($\beta_1, \beta_2, \lambda, \sigma$) can be determined based on prior opinions about some key quantities related to the prior-predictive survival time $T^*$ of a patient with known covariate values, say, $(z_1^*, z_2^*)$. These prior opinions are related to: (1) Prior guess and prior range of the median (or any other quantile) of prior-predictive survival time $T^*$; (2) Change in median survival for change in each covariate value. For the lung cancer trial, we do not have any access to the expert opinions, however, we present here the prior opinions about the survival responses that corresponds to the priors we use. This should demonstrate how these priors can be determined after eliciting the expert/prior opinions about the above key quantities related to survival responses.

For the convenience of presentation, we work with age $Z_2$ centered at the median age of the population. Without loss of generality, we take $z_1^* = 0$ and $z_2^* = 0$ (i.e., $T^*$ is the survival
time of a random patient with median age and receiving treatment A). The specification of the prior for $\beta_0$ uses the fact that $T^*$ has a prior median $\exp(\beta_0)$. Thus, a mean prior guess of median of $T^*$ as 18 months and range of median in (6, 30) months give us the prior $\beta_0 \sim N(A_1, B_1^2)$, where $A_1 = \log(18)$ and $B_1 = (\log(30) - \log(6))/4$ (prior range of $\beta_0$ should have length of $4B_1$). A prior opinion about $\beta_1$ is based on the observation that the ratio of medians $[Q_{0.5}(z_1 = 1, z_2^*)/Q_{0.5}(z_1 = 0, z_2^*)] = \exp(\beta_1)$ of two patients with identical age (standardized median equal to 0), but, from different treatment arms. So, a 95% prior probability of this ratio of medians $e^{\beta_1}$ being in $(e^{-2}, e^2)$ and centered at $e^0 = 1$ (indifference in opinion regarding superiority of any treatment arm), corresponds to the prior $\beta_1 \sim N(0, 1)$. Similarly, the prior $\beta_2 \sim N(0, 1)$ corresponds to prior opinion that 1 year change in age of $T^*$ can change the median survival by a factor between $(e^{-1}, e)$ with 68% probability.

We use the $Unif(0,3)$ prior $\pi_2(\lambda)$ of $\lambda$ because it is difficult to interpret the linear model of (6.5) after transformation to symmetry when the transformation parameter $\lambda$ is more than 3. In their original paper, Box and Cox (1964) recommended restricting the $\lambda \leq 2$. Centering $\pi_2$ at 1 represents the mean prior opinion that $\log(T^*)$ itself has symmetric and unimodal density. The prior of $\sigma$ is based on prior opinion about possible values of another quantile $Q^* = Q_{\alpha^*}(z_1^*, z_2^*)$ (different from the median) of $T^*$, where $P[T^* > Q^*|\pi(\cdot)] = \alpha^*$ with $\pi$ representing the joint prior. The prior predictive probability $P[T^* > Q^*|\lambda = \lambda^*, \beta = \beta^*, \sigma] = 1 - \Phi(g_{\lambda^*}(\log(Q^*)) - g_{\lambda^*}(\beta_0^*))$ should be approximately equal to the $\alpha^*$, where $(\lambda^*, \beta^*)$ is the known prior mean of $(\lambda, \beta)$ and $\Phi(u)$ is the cdf of $N(0, \sigma)$. This gives us a prior opinion about the range $(R_1, R_2)$ and prior mean $\sigma_0$ of $\sigma$ from the prior range $(A_3, B_3)$ and prior mean $A_0$ of $Q^*$ because predictive density of $[g_{\lambda^*}(\log(Q^*)) - g_{\lambda^*}(\beta_0^*)]/z_{\alpha^*}$ is approximately the prior density of $\sigma$, when $z_{\alpha^*}$ is the $\alpha^*$-quantile of standard normal. For this data analysis, we use the prior that 1-standard deviation away from prior predictive mean of $T^*$ lies in the interval (18, 60) months. Arguably, one can also use a conditional prior with $\pi_3(\sigma|\beta_0, \lambda)$ representing that center and support for the prior of $\sigma$ should depend on the parameters $(\beta_0, \lambda)$. We remind the reader that the priors used in our analysis are solely for demonstrating the method of development of one set of priors for the Bayesian analysis of the lung-cancer study. An expert’s prior opinions on median survival time of small cell lung cancer can be very different from what we used, and that will lead to different prior specification of the parameters.

Our plot (left-hand panel of Figure 1) of residuals $y_i - y_i^*$ versus the patient’s age at
entry, where \( y_i \) is the observed log-survival time (subject to censoring) and \( y_i^* \) is the posterior predictive expectation of \( \log(T) \) under the model, does not show any trend of residuals under parametric Bayes model. Our plot (right-hand panel of Figure 1) of these residuals versus the estimated median survival times also does not reveal any inadequacy of the parametric model. However, the Q-Q plot (Figure 2) of these residuals suggests that the assumption of Gaussian distribution for \( F_\epsilon \) in (6.5) is questionable due to the plot being non-linear at the right tail. Later, we use the semiparametric Bayesian analysis to avoid the Gaussian assumption of \( \epsilon \). Our posterior means (Bayes estimates) of three quartiles \( Q_\alpha(z_1,z_2) \) for \( \alpha = 0.25, 0.50, 0.75 \) of treatment A (\( z_1 = 1 \)) are higher than the corresponding quantiles of treatment B (\( z_1 = 0 \)) at any age (\( z_2 > 0 \)) (See Figure 3).

This prior process assumes a rich class of unimodal distributions symmetric around 0 for the support of \( F_\epsilon \). This constructive formulation for the prior process allows us to implement the semiparametric Bayesian estimation via MCMC computational tools to sample from the posterior of \( (\beta, \lambda, F_\epsilon) \) in (2.9). The actual implementation of the MCMC tool to sample from (2.9) is based on a finite approximation \( F_\epsilon(u) \simeq \sum_{k=1}^{N} p_k \phi(u|\theta_k) \) of the construction.
Figure 4.2: Q-Q plots of residuals under parametric TBS model for small-cell lung cancer data described in (4.1).

For the semiparametric Bayesian analysis with a symmetric unimodal $f_\epsilon$ in (6.5), we need to specify the prior guess/mean of $F^*$ of $F_\epsilon$ and a prior precision parameter $\alpha$. The constructive definition of the DP mixture prior process for $F_\epsilon$ is (Sethuraman, 1994) [24]

$$F_\epsilon(u) = \sum_{k=1}^{\infty} p_k \zeta(u|\theta_k),$$

(4.1)

where $\theta_k \overset{i.i.d.}{\sim} G_0$, $p_k = V_k \prod_{j<k} (1 - V_j)$ with $V_j \overset{i.i.d.}{\sim} Beta(1, \alpha)$. We take the precision parameter $\alpha = 1$ to imply a very low confidence around our parametric prior guess $F_*$ of the nonparametric error distribution $F_\epsilon$. We take $f_*$ to be $N(0, (0.51)^2)$, the same as the prior mean of $f_\epsilon$ for the parametric Bayes analysis. This prior mean $f_*$ of $f_\epsilon$ give us the prior mean of $f_\epsilon$ used in the parametric Bayes analysis. Using (2.11), this $N(0, (0.51)^2)$ density for $f_*$ corresponds to a $Gamma(3/2, 1/\{2(0.51)^2\})$ for $G_0$ in (4.1). To obtain a $N(0, \gamma^2)$ density as prior mean/guess $f_*$, we need to use $Gamma(3/2, 1/(2\gamma^2))$ cdf for $G_0$. The MCMC computational computational tool can be implemented, even in standard package such as
Winbugs, using a finite (say, $N = 10$) number (maximum number of $\phi_k$ in (4.1)) of $V_1, \ldots, V_N$ and $p_1, \ldots, p_N$. The priors for $V_1, \ldots, V_{N-1}$ for $N = 10$ are i.i.d. $Beta(1, \alpha)$ and $V_N = 1$. After simulating $V_1, \ldots, V_N$, we obtain $p_1 = V_1$, $p_2 = V_2(1 - V_1)$, $p_3 = V_3(1 - V_2)(1 - V_1), \ldots$, $p_M = V_M(1 - V_{M-1})(1 - V_{M-2}) \ldots (1 - V_1)$. The rest of the conditional posteriors are same as those used for parametric Bayes.

We get the semiparametric Bayes point estimates $\hat{\beta}_1 = -0.409$ and $\hat{\beta}_2 = -0.023$ of the regression parameter $(\beta_1, \beta_2)$ along with 95% credible intervals $(-0.693, -0.255)$ and $(-0.029, -0.013)$ (respectively). The results of the Bayes estimators of regression parameters $(\beta_1$ and $\beta_2)$ under parametric and semiparametric TBS models along with ML estimator based on parametric Gaussian error TBS model are presented in Table 1. The last line of Table 1 is the result for the Bayesian median regression model of Kottas and Gelfand (2001) [6] using the model of (2.4) with $f_\alpha(u) = (1/2)|\eta|^{sgn(u)}f_0(|\eta|^{sgn(u)}u)$ for a nonparametric density $f_0(u)$ defined on $u > 0$.

The point estimates of each regression parameter under different methods are similar to the corresponding point estimator obtained via Portnoy’s method (2003) [13]. However
Table 4.1: Pointwise and 95% interval (within parenthesis) estimators of regression parameters for the small cell lung cancer study under different procedures

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Treatment</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>-0.402</td>
<td>(-0.576, -0.228)</td>
</tr>
<tr>
<td>Parametric Bayes</td>
<td>-0.421</td>
<td>(-0.703, -0.168)</td>
</tr>
<tr>
<td>Semiparametric Bayes</td>
<td>-0.409</td>
<td>(-0.693, -0.255)</td>
</tr>
<tr>
<td>Portnoy</td>
<td>-0.360</td>
<td>(-0.660, -0.060)</td>
</tr>
<tr>
<td>KG Bayes</td>
<td>-0.169</td>
<td>(-0.367, -0.016)</td>
</tr>
</tbody>
</table>

ML and Bayes methods yield smaller estimated standard errors and substantially narrower interval estimates than those obtained using Portnoy’s method. For this data example, the parametric MLE has the smallest estimated standard error for the treatment effect among all procedures. The corresponding posterior standard deviations from parametric and semiparametric Bayes are smaller than the standard errors from Portnoy’s method. This is not surprising because existing median regression methods have a far larger number of regression parameters than the finite dimensional regression parameter $\beta$ in (6.5).

### 4.2 Data Example 2

We apply the multivariate TBS model of (2.12) to the Diabetic Retinopathy Study (DRS) trial (Hougaard, 2000) [25] of $n = 197$ high-risk diabetes patients to assess the effectiveness of the laser treatment in delaying severe vision loss. For each patient, one eye was randomly assigned to the laser photocoagulation, and the other served as an untreated control. The event of interest for eye $j = 1, 2$ of patient $i = 1, \ldots, n = 197$ was the time $T_{ij}$ when visual acuity was less than 5/200 for two consecutive visits (time to ”blindness” in months). Two covariates of the study are: indicator $Z_1 = 0$ for eye/treatment (=0 for treated right eye and =1 for untreated left eye), and indicator $Z_2$ of patient’s diabetes onset-time (1 for juvenile onset and 2 for adult onset).

To the best of our knowledge, no existing works handle median regression in a multivariate survival data scenario. Thus, for the DRS trial, we compare the Bayesian methods based on a multivariate TBS model with some existing multivariate survival methods: a marginal Cox model (Wei et al., 1989) [26] and the log-normal frailty model (Hougaard, 1995, 2000) [27]
For the $j^{th}$ event and the $i^{th}$ cluster, the marginal hazard for $T_{ij}$ under the multivariate survival model of Wei et al. (1989) \[26\] is

$$h_{ji}(t|Z_{ij}) = h_0(t) \exp(\beta_1 z_{1ji} + \beta_2 z_{2ji} + \beta_3 z_{1ji} z_{2ji}),$$

(4.2)

where $h_0(t)$ is the unspecified marginal baseline hazard and $\beta$ is the marginal regression coefficient. The conditional hazard of $T_{ij}$ of the log-normal frailty model (Hougaard, 1989, 2000) has the form:

$$h_{ji}(t) = h_0(t) \exp(\beta_1 z_{1ji} + \beta_2 z_{2ji} + \beta_3 z_{1ji} z_{2ji}) \nu_i,$$

(4.3)

where the random and unobservable cluster effects (frailty) $\nu_i$ are assume to be i.i.d. log-normal with mean 1. Even though we are taking the liberty of using the same notation $h_0(t)$ for the baseline hazards and $\beta$ for the regression parameters in both models, the actual form and interpretation of $(h_0(t), \beta)$ in (4.2) and (4.3) are very different and there is no simple functional relationship between $(h_0(t), \beta)$ in (4.2) and $(h_0(t), \beta)$ in (4.3). The estimators of regression parameters ($\beta$) from the DRS data along with the corresponding standard errors are obtained via SAS routines using partial likelihood based methods of Wei et al. (1989) \[26\] for (4.2) and an EM-type algorithm described in Hougaard (2000) \[25\] for (4.3). The point estimators of $\beta$ (along with corresponding variability estimators in the paranthesis) under different methods (including semiparametric Bayes for multivariate TBS model) are shown in Table 2.

Table 4.2: Pointwise (standard-error/posterior standard-deviation) estimators of regression parameters for DRS data under different methods

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_1 \times \beta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBS with Bayes</td>
<td>1.03 (0.17)</td>
<td>1.19 (0.23)</td>
<td>-0.72 (0.16)</td>
</tr>
<tr>
<td>TBS with Semi-Bayes</td>
<td>0.99 (0.15)</td>
<td>1.17 (0.21)</td>
<td>-0.68 (0.14)</td>
</tr>
<tr>
<td>Marginal Cox Model</td>
<td>-1.27 (0.24)</td>
<td>-0.34 (0.20)</td>
<td>0.85 (0.30)</td>
</tr>
<tr>
<td>Log-normal Frailty</td>
<td>-1.59 (0.30)</td>
<td>-0.45 (0.27)</td>
<td>1.04 (0.37)</td>
</tr>
</tbody>
</table>

Table 2 shows that the semiparametric Bayes estimates of $\beta$ are close to the estimates obtained from frequentist estimates based on (4.2) and (4.3). Because the increase/decrease

\[^1\] Negative of the Marginal Cox and Log-normal frailty describe changes in median, and are comparable to the TBS model.
in hazard corresponds to decrease/increase in median, the signs of the estimated regression parameters are reversed in marginal Cox and frailty models compared to those from TBS models. Though the multivariate TBS model of (2.12) has a different formulation and interpretation of regression parameters $\beta$ from those under (4.2) and (4.3), we can compare them if we assume that $h_0(t)$ in (4.2) and (4.3) are approximately constant. For the Wei et al. (1989) [26] model of (4.2) with constant $h_0(t) = \gamma$, the marginal median is $t^* = \log(2)/(\gamma \exp(\beta'Z))$. This implies that the effect of the parameter $\beta$ on median is actually negative of effect of $\beta$ on median in (6.5). For the frailty model of (4.3), there is no direct interpretation of $\beta$ on marginal median unless the frailty variance is 0 and $h_0(t) = \gamma$ (constant), and in this case the interpretation of $\beta$ is same as in (4.2). Table 2 shows that the estimates we have obtained from different models/methods are very comparable, except the posterior standard deviation from a semiparametric multivariate TBS model is lower than the standard errors from other competing methods.
CHAPTER 5

Discussion

In this paper, we present a new class of semiparametric models amenable to Bayes estimation of the median regression parameter $\beta$ for censored survival data. Similar to previous semiparametric models (e.g., Cox’s model), it has one non-parametric function $f_\epsilon$ and finite dimensional parameter vector $(\beta, \lambda)$. Our Bayes methods have the advantage of ease of computation, determination of priors and a simple interpretation of the regression parameters of the models. Our model is different from Cox’s model, however, our model class is large since the after-transformation nonparametric error density $f_\epsilon$ has no functional form except the assumption of symmetry and unimodality. Our method can be applied when the covariate $Z$ affects the location as well as the scale and shape of the regression error (hetero-scedasticity).

Median regression offers a useful alternative to the popular regression functions of Cox’s model (1972) [1] and the transformation model of (2.1). There is a substantial literature on median regression for censored survival data. Ying et al. (1995) [5], Yang (1999) [28], McKeague et al. (2001) [29] and Bang and Tsiatis (2003) [30] proposed semiparametric procedures based on modifications of the least absolute deviation (LAD) of Powell (1984, 1986) [31] [32]. These methods involve solving non-linear discontinuous estimating equations often with multiple solutions. However, since the estimating equations involve discontinuous estimators of survival and hazard functions, the computation can be difficult to implement in practice. The recursive nature off Portnoy’s (2003) [13] self-consistency based method makes the asymptotic justifications and computation complicated. Peng and Huang’s (2008) [18] martingale based estimating equations involve minimization of an $L_1$-type discontinuous convex functions. Unlike estimation with Cox’s model (Cox, 1972) [1], martingale based methods may not be the most efficient for estimating regression parameters of the median...
survival time. For most of these methods, every quantile functional is assumed to be linear in $Z$, that is $Q_\alpha(Z) = \beta_\alpha Z$ for all $\alpha \in (0, 1)$, where $P[T > Q_\alpha(Z)] = \alpha$. This restrictive modeling assumption may not hold true for a study and very few known stochastic models can satisfy this. Unlike our semiparametric model of (6.5), these models have an infinite number of regression parameters $\beta_\alpha$ for $\alpha \in (0, 1)$. For more in depth discussion about the implementations, comparisons, and consequences of the restrictive assumptions for existing quantile regression approaches for censored survival data (particularly on the asymptotic rate of convergence of any quantile estimate), one can look at the excellent review by Koenker (2008) [33].

Although we focus on modeling the median functional, our method can be used via (2.5) to compute the point estimate and can even compute the joint confidence band of any quantile functional. However, for some diseases, such as cancers with very good prognosis, the main interest may not be on the median, and the goal may be on modeling the quantile $Q_\alpha(Z)$ as a log-linear function with $P[T > Q_\alpha(Z)] = \alpha$ for $\alpha > 1/2$. In this case, we can use a modification of the log-linear model in (6.5) with $P[\epsilon > 0] = 1 - F_\epsilon(0) = \alpha$. We can use the scale-mixture of uniform model of (2.11) with the modification that $\zeta(u|\theta)$ as the uniform density with support $(2\theta(\alpha - 1), 2\theta\alpha)$ with Dirichlet process for the unknown mixing density $G$. The rest of the methodology can be extended from our median regression method with minor changes to adopt this modified $\zeta(\theta)$. For the sake of brevity, we skip the details of the MCMC steps. This model allows only the $(1 - \alpha)$-percentile of $T$ to have log-linear function of covariate.

Our method can also predict the outcome of a future patient with known covariate values. We have also addressed the median regression for multivariate survival data. Our simulation results show that the efficiency gain of semiparametric Bayes estimators is substantial compared to existing frequentist estimators even when our assumption of (6.5) does not hold (e.g. for Pareto distribution). We do not present any separate simulation study of parametric Bayes estimators because these estimators under diffuse prior information are numerically close to parametric ML estimators. All of these advantages make our proposed method an extremely attractive alternative to other existing semiparametric method for censored data.
CHAPTER 6

Future Work: Bayesian Regularized Median Regression

6.1 Introduction

We consider the following general linear regression model of the form:

\[ Y = \mu 1_n + X\beta + \epsilon, \]

where \( Y = (Y_1, \ldots, Y_n)' \) is the \( n \times 1 \) vector of response, \( \mu \) is the overall mean, \( X = (X_1, X_2, \ldots, X_n)' \) is the \( n \times p \) matrix of covariates and \( \epsilon \) is the \( n \times 1 \) vector of iid normal errors with mean 0 and variance \( \sigma^2 \).

The Least Absolute Shrinkage and Selection Operater (LASSO) is proposed by Tibshirani (1996) [34] for simultaneous variable selection and parameter estimation. This procedure is constructed within the penalized likelihood framework and shrinks some regression coefficients towards exactly zero. The lasso minimizes the residual sum of square with a constraint which is expressed in term of the \( L_1 \)-norm of the coefficient vector of \( \beta \):

\[ \hat{\beta}_L = \arg \min_{\beta} (\tilde{y} - X\beta)'(\tilde{y} - X\beta) + \lambda \sum_{j=1}^{p} |\beta_j|, \]

where \( \tilde{y} = y - \bar{y} 1_n \) is the centered observed response vector and \( \lambda \geq 0 \) is the tuning parameter of penalty. When \( \lambda = 0 \), \( \hat{\beta}_L \) is the ordinary least-square estimate, and it would be shrinked towards zero when \( \lambda \) is sufficiently large. The lease angle regression algorithm (LARS) (Efron et al., 2004; Osborne et al.,2000) [35][36]provides an efficient implementation of LASSO computation.

Tibshirani (1996) [34] pointed out that the lasso estimator can be interpreted as the Bayes posterior mode estimates of \( \beta \) using gaussian likelihood and iid Laplace (double-exponential) priors for \( \beta_1, \ldots, \beta_p \). Bae and Mallick (2004) [37] used a two-level hierarchical Bayesian

However, if the normality assumption of (6.1) does not hold, penalized median/quantile regression provides a useful alternative to classical LASSO estimates for its superior robustness properties, richer information and better prediction accuracy. The natural extension of LASSO penalty in quantile regression context can be defined as (Koenker, R. 2005) [40]:

\[ \hat{\beta}_L(\tau) = \arg\min_{\beta} \sum_{i=1}^{n} \rho_\tau(\tilde{y}_i - X_i \beta) + \lambda \sum_{j=1}^{p} |\beta_j|, \]

(6.3)

where \( \rho_\tau(\cdot) \) is the loss function

\[ \rho_\tau(t) = \begin{cases} \tau t, & \text{if } t > 0 \\ -(1-\tau)t, & \text{if } t < 0 \end{cases} \]

The above penalized estimators (6.3) of quantile LASSO can be easily solved by standard linear programming techniques. Li, Xi and Lin (2010) [41] studied regularization in quantile regression from a Bayesian perspective. Hierarchical models are proposed to give a generic treatment of different penalties including lasso, group lasso and elastic net penalties.

In this paper, Our aim is to develop a Bayesian simultaneous variable selection and parameter estimation procedure in median regression using the Transform-both-sides model. The parameters of median regression are estimated via gaussian Bayesian hierarchical LASSO model after applying the generalized Box-Cox transformation (Bickel and Doksum, 1981) [15] to both the response and linear predictor term. The rest of paper proceeds as follows. In Section 2, we introduce the proposed Beyesian regularized median regression via Transform-both-sides Model. Hierarchical model is developed to get estimators. In section 3, we compare the proposed Bayesian median lasso to the frequentist counterparts via simulation studies. Real data examples are illustrated in the Section 4. In section 5, we have the final discussion and conclusion.

### 6.2 Bayesian Hierarchical Model

Bickel and Doksum (1981) [15] proposed a monotone power transformation, an extension of the Box-Cox power family,

\[ g_\eta(y) = \frac{y^\eta \text{sign}(y) - 1}{\eta}, \text{ for } \eta > 0, \]

(6.4)
where \( \text{sign}(y) = 1 \) if \( y \geq 0 \) and \( \text{sign}(y) = -1 \) if \( y < 0 \). We assume that under an optimal \( \eta \), the transformed response \( g_\eta(Y_i) \) has normal distribution with mean \( g_\eta(x_i^T \beta) \) and variance \( \sigma^2 \), that is

\[
g_\eta(Y_i) = g_\eta(x_i^T \beta) + \epsilon_i ,
\]

where \( \epsilon_i \sim N(0, \sigma^2) \). Suggested by Tibshirani (1996) [34], the conditional Laplace prior has the form as follows:

\[
\pi(\beta|\sigma^2) = \prod_{j=1}^{p} \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_j|/\sqrt{\sigma^2}} ,
\]

Park and Casella (2008) [39] noted that conditioning on \( \sigma^2 \) is important and meaningful because it guarantees a unimodal full posterior. Andrews and Mallows (1974) [42] also pointed out that the Laplace prior can be represented as a scale mixture of normals with an exponential distribution.

\[
\frac{\lambda}{2} e^{-\lambda |z|} = \int_0^\infty \frac{1}{\sqrt{2\pi s}} \exp \left( -\frac{z^2}{2s} \right) \frac{\lambda^2}{2} \exp(-\frac{\lambda^2}{2s}) ds ,
\]

So, our Bayesian Hierarchical model is represented as follows:

\[
g_\eta(y_i)|\beta, \sigma^2, \eta \sim N(g_\eta(x_i^T \beta), \sigma^2) \\
\beta|\tau \sim \prod_{k=1}^{p} \frac{1}{\sqrt{2\tau_k}} \exp \left( -\frac{\beta_k^2}{2\tau_k} \right) \\
\tau|\lambda \sim \prod_{k=1}^{p} \lambda^2 \exp \left( -\frac{\lambda^2}{2\tau_k} \right) \\
\pi(\lambda^2) \sim (\lambda^2)^{r-1} \exp(-\delta\lambda^2)
\]

As suggested by Park and Casella (2008) [39], the improper prior \( \pi(\sigma^2) \propto 1/\sigma^2 \) can be used as the prior of model the error variance \( \sigma^2 \). We can also use \( U(0, 3) \) for the prior of \( \eta \), since the transformation of 6.4 is hard to interpret when \( \eta \) is too large, say \( \eta > 3 \). For \( \lambda^2 \), if \( r = \delta = 0 \), \( \lambda^2 \) have a noninformative prior.

### 6.3 Preliminary Simulation Study

In this section, we conduct the simulation study to evaluate the performance of proposed Bayesian median regression model via TBS. The Lasso Penalized Quantile Regression (Koenker, R. 2005) [40] estimator is used for comparison.
The data in the simulation studies come from the following model:

\[ Y_i = X_i^T \beta + \epsilon_i, \ i = 1, \ldots, n, \]

(6.9)

- **Case 1** Let \( \beta = (3,1.5,0,2,0,0,0,0) \) and \( \epsilon_i \) comes from Normal distribution with mean=0, variance=1, \( n = 20 \) and 200. The correlation between \( X_i \) and \( X_j \) is \( 0.5^{|i-j|} \). 100 simulations were conducted here. The hyperpriors for the proposed model are assigned as follows:

\[ \pi \left( \frac{\lambda^2}{2} \right) \sim \text{Gamma}(a,b), \]

(6.10)

where \( a \) and \( b \) are set to be 0.1 and 0.1, which corresponds to nearly noninformative prior.

<table>
<thead>
<tr>
<th>Simulation 1</th>
<th>Method</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( \beta_4 )</th>
<th>( \beta_5 )</th>
<th>( \beta_6 )</th>
<th>( \beta_7 )</th>
<th>( \beta_8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta_{\text{true}} )</td>
<td>3.000</td>
<td>1.500</td>
<td>0.000</td>
<td>2.000</td>
<td>3.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>( n = 20 )</td>
<td>( \beta_{\text{TBS.lasso}} )</td>
<td>3.000</td>
<td>1.323</td>
<td>0.088</td>
<td>1.854</td>
<td>0.021</td>
<td>0.007</td>
<td>0.004</td>
<td>0.017</td>
</tr>
<tr>
<td>(MSE)</td>
<td></td>
<td>0.073</td>
<td>0.179</td>
<td>0.036</td>
<td>0.124</td>
<td>0.032</td>
<td>0.019</td>
<td>0.027</td>
<td>0.020</td>
</tr>
<tr>
<td>( n = 20 )</td>
<td>( \beta_{\text{PQ.lasso}} )</td>
<td>3.074</td>
<td>1.388</td>
<td>0.051</td>
<td>1.905</td>
<td>0.043</td>
<td>0.004</td>
<td>-0.008</td>
<td>0.030</td>
</tr>
<tr>
<td>(MSE)</td>
<td></td>
<td>0.130</td>
<td>0.208</td>
<td>0.100</td>
<td>0.158</td>
<td>0.108</td>
<td>0.084</td>
<td>0.070</td>
<td>0.106</td>
</tr>
<tr>
<td>( n = 200 )</td>
<td>( \beta_{\text{TBS.lasso}} )</td>
<td>2.999</td>
<td>1.482</td>
<td>0.017</td>
<td>2.000</td>
<td>0.000</td>
<td>-0.001</td>
<td>0.007</td>
<td>0.000</td>
</tr>
<tr>
<td>(MSE)</td>
<td></td>
<td>0.004</td>
<td>0.008</td>
<td>0.005</td>
<td>0.008</td>
<td>0.004</td>
<td>0.004</td>
<td>0.005</td>
<td>0.006</td>
</tr>
<tr>
<td>( n = 200 )</td>
<td>( \beta_{\text{PQ.lasso}} )</td>
<td>2.997</td>
<td>1.464</td>
<td>0.012</td>
<td>2.002</td>
<td>-0.007</td>
<td>0.005</td>
<td>-0.002</td>
<td>0.009</td>
</tr>
<tr>
<td>(MSE)</td>
<td></td>
<td>0.008</td>
<td>0.013</td>
<td>0.012</td>
<td>0.016</td>
<td>0.012</td>
<td>0.014</td>
<td>0.016</td>
<td>0.012</td>
</tr>
</tbody>
</table>

The simulation results in the Table 1 shows that both estimators have very little bias for \( n = 20 \) and 200. However, the mean square errors (MSEs) of the proposed Bayesian lasso is smaller the frequentist counterpart.

- **Case 2** Let \( \beta = (3,3,0,0,3,0,0,0) \) and \( \epsilon_i \) comes from Laplace distribution with mean=0, scale=1  \( n = 20 \) and 200. The correlation between \( X_i \) and \( X_j \) is \( 0.5^{|i-j|} \). 100 simulations were conducted here.
Simulation results in Table 2 show that the proposed Bayesian median lasso has comparative performance to the frequentist regularized method regarding the bias part. Also, the proposed method has smaller MSEs in most of the parameter estimation.

· **Case 3** Let $\beta = (3, 3, 0, 0, 3, 0, 0, 0)$ and $\beta = (5, 0, 0, 0, 0, 0, 0, 0)$ and $\epsilon_i$ comes from Extreme Value (Gumbel) distribution, with median of $\epsilon_i$ is 0, $n = 20$. We performed 100 simulation replicates here. The Gumbel distribution is a left skew distribution, and the cumulative distribution function has form as follows:

$$F(\epsilon_i|\lambda, \beta) = e^{-e^{-(\epsilon_i - \lambda)/\beta}}, \quad (6.11)$$

The median is $\lambda - \beta \ln(\ln 2)$. We set $\lambda = -1, \beta = -1/\log(\log(2))$ in simulation, which result in variance = $\pi^2/6$, median = 0.
regularized estimator in the esymmetry error distribution. The bias is almost negligible considering the MSEs gain in most parameter estimations.
APPENDIX A

The Proofs of Theorems

Proof of Theorem 1: Without loss of generality, it will be sufficient to prove the following: when

\[ Y = \beta x + \epsilon^* \text{ and } Y^\lambda = (\beta^* x)^\lambda + \epsilon^{**} \]

for two symmetric \( \epsilon^* \sim F^* \) and \( \epsilon^{**} \sim F^{**} \) with \( F^*(0) = F^{**}(0) = 1/2 \) imply \( F^* = F^{**}, \beta = \beta^* \) and \( \lambda = 1 \).

We can show that \( P[Y < y|x] = F^*(y - \theta) = F^{**}(y^\lambda - (\theta^*)^\lambda) \) for all \( y \), where \( \theta = \beta x \) and \( \theta^* = \beta^* x \). Taking \( y - \theta = 0 \), we show \( F^*(0) = F^{**}(0) = F^{**}(\theta^\lambda - (\theta^*)^\lambda) = 1/2 \) implies \( \theta^\lambda = (\theta^*)^\lambda \). Using this last result, we can show that \( \theta = \theta^* \) and the rest of the proof follows from there.

Remark 1. Under the TBS model, the hazard function is expressed as follows:

\[
h(t|Z) = -\frac{d}{dt} \log S_Z(t) = -\frac{d}{dt} \log \left[ \Phi_0 \left( \frac{g_\lambda(\log t) - g_\lambda(\beta^* Z)}{\sigma} \right) \right] = \frac{\phi_0(\frac{g_\lambda(\log t) - g_\lambda(\beta^* Z)}{\sigma})}{\Phi_0(\frac{g_\lambda(\log t) - g_\lambda(\beta^* Z)}{\sigma})} \frac{1}{\sigma t} (|\log t|^\lambda - 1) = \frac{\phi_0(\omega)}{\Phi_0(\omega)} \frac{1}{\sigma t} (|\log t|^\lambda - 1)
\]

where \( \omega = \frac{g_\lambda(\log t) - g_\lambda(\beta^* Z)}{\sigma} \), \( \phi_0(\omega) \) is the standard normal \( N(0, 1) \) density function, \( \Phi_0(\omega) = \int_{-\infty}^{+\infty} \phi_0(u)du \) is the survival function corresponding to the density \( \phi_0(\omega) \).
REFERENCES


