Bayesian survival analysis via transform-both-sides model

Jianchang Lin* Debajyoti Sinha† Stuart Lipsitz‡ Adriano Polpo§

Abstract

We present a novel semiparametric survival model with a log-linear median regression function. As a useful alternative to existing semiparametric models, our large model class has many important practical advantages, including interpretation of the regression parameters via the median and the ability to address heteroscedasticity. We demonstrate that our modeling technique facilitates the ease of prior elicitation and computation for both parametric and semiparametric Bayesian analysis of survival data. We illustrate the advantages of our modeling, as well as model diagnostics, via a reanalysis of a small-cell lung cancer study. Results of our simulation study provide further support for our model in practice.

Key Words: Log-linear median regression; Bayesian Survival analysis; Transform-both-sides; Quantile regression

1. Introduction

Semiparametric models such as Cox's (1972) proportional hazards model and linear transformation models (Cheng et al., 1995; Fine et al., 1998) and their special cases (e.g., accelerated failure time model) are very popular for modeling effects of covariates on a 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) is to express the effects of treatment arm and age at entry on time from randomization to death (survival time). Often, there is substantial information available in the data to make inferences about the median. However, 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) gave compelling arguments in favor of focusing on the quantiles of the survival time for modeling and reporting of data analysis results. The effect of treatment and age on the quantiles including median time to death is useful for describing covariate effects. Clinical trials based on survival outcomes are often designed to detect differences in median survival between treatment arms. Models based on the median are often useful in dealing with heteroscedasticity.

Semiparametric Bayesian models for survival data, possibly with the exception of Kottas and Gelfand (2001), and Hanson & Johnson (2002), are either based on covariate effects on the hazard ratio (see Ibrahim et al., 2001) or on the mean survival time (e.g., Walker and Mallick, 1999). However, particularly for Bayesian survival analysis, medians and other quantiles are natural choices for elicitation of experts' opinions. Clinical experts on the disease under study are likely to have useful prior information/opinions about survival quantiles (say, the median). 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 treatment on median survival time (Piantadosi, 2005). In Section 2 of this paper, we propose a novel semiparametric model for the median

*Millennium: The Takeda Oncology Company, 35 Landsdowne St., Cambridge, MA 02139
†Department of Statistics, Florida State University, Tallahassee, Florida 32306
‡Division of General Medicine, Brigham and Womens Hospital, Boston, Massachusetts 02115
§Department of Statistics, Federal University of São Carlos, São Paulo, 13565-905, Brazil

survival time with interpretable covariate effects via a log-linear median regression function. This wide class of semiparametric models has many desirable properties including model identifiability, closed form expressions for all quantile functions, and non-monotone hazards. Unlike previous methods for Bayesian survival analysis (e.g., Hanson & Johnson, 2002), our model accommodates the situation when the location/median as well the scale and shape of the survival distribution are affected by the covariate. Unlike some of the previous frequentist methods for median regression, we do not require the restrictive assumption that all quantile functions below the median to be linear.

In Section 3, we present the likelihood, suitable nonparametric prior processes and MCMC (Markov Chain Monte Carlo) tools to estimate the model parameters. In section 4, we consider the SCLC trial to demonstrate how our models can facilitate the determination of prior distributions. For the SCLC study, we also compare the results of our approach to existing approaches. In Section 5, a simulation study investigates small sample performance and robustness properties compared to competing methods for median regression. Some final remarks are in Section 6.

2. Semiparametric Models

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) assumes that

$$h(T_i) = \gamma' Z_i + e_i \,, \tag{1}$$

where h is a monotone transformation, $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_p)$ is a regression parameter, and e_i is an unspecified error variable with common density $f_e(\cdot)$ free of covariate Z_i. Usually the density $f_e(\cdot)$ 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 (1) are the accelerated failure time model (AFT) when $h = \log$, the proportional odds model when e_i comes from a logistic distribution, and Cox's model (1972) when f_e is the extreme-value density.

The monotone power transformation $g_\lambda(y)$ (Bickel and Doksum, 1981),

$$g_\lambda(y) = \frac{\text{Sgn}(y)\ |y|^\lambda}{\lambda} \text{ for } \lambda > 0 \,, \tag{2}$$

where $\text{Sgn}(y) = -1$ for $y < 0$ and $\text{Sgn}(y) = +1$ otherwise, is an extension of the Box-Cox power family (Box and Cox, 1964), a popular transformation to obtain symmetric and unimodal density for the transformed random variable. We assume that for unknown λ, the transformed survival time $g_\lambda\{\log(T_i)\}$ is symmetric and unimodal with median $g_\lambda(\beta' Z_i) = g_\lambda(M_i)$, that is,

$$g_\lambda\{\log(T_i)\} = g_\lambda(M_i) + \epsilon_i \tag{3}$$

where ϵ_i are iid from a unimodal and symmetric density $f_\epsilon(\cdot)$ centered at 0, $M_i = \beta' Z_i$, and β is the vector of regression parameters. Carroll and Ruppert (1984), Fitzmaurice et al. (2007), among others proposed parametric versions of the transform-both-sides (TBS) regression model for an uncensored continuous response with the original Box-Cox transformation and $N(0, \sigma^2)$ density for error $f_\epsilon(\cdot)$.

The transformation $g_\lambda(y)$ in (2) is monotone with derivative (with respect to λ) equal to $g'_\lambda(y) = |y|^{\lambda - 1}$. The median of $\log(T_i)$ is $M_i = \beta' Z_i$ because $P[\log(T_i) > M_i]$ $= P\left[g_\lambda\{\log(T_i)\} > g_\lambda(M_i)\right] = F_\epsilon(0) = 1/2$, where F_ϵ is the cdf of ϵ. As a consequence, the survival time T_i has a log-linear median regression function $Q_{0.5}(Z_i) = \exp(M_i) = \exp(\beta' Z_i)$ and survival function $S(t|z) = 1 - F_\epsilon(g_\lambda(\log t) - g_\lambda(M))$. For the SCLC

study with $M_i = \beta_0 + \beta_1 z_1 + \beta_2 z_2$, where z_1 is a treatment indicator and z_2 denotes age, this implies that the ratio of medians from two patients of the same age but different treatment arms is $Q_{0.5}(z_1 = 1, z_2)/Q_{0.5}(z_1 = 0, z_2) = \exp(\beta_1)$. We also get a similar straightforward interpretation of $\exp(\beta_2)$ as the ratio of the medians for unit increase in age. The following theorem shows that the parameter λ and the density f_ϵ of (3) are also identifiable, in the sense that for any survival time following (3), there is a unique (λ, f_ϵ) for which $g_\lambda\{\log(T_i)\}$ has a symmetric unimodal distribution.

Theorem 1: For the model in (3) if there is another triplet $(\lambda^*, \beta^*, f_{\epsilon^*})$ for which $g_{\lambda^*}\{\log(T)\} = g_{\lambda^*}(\beta^* x) + \epsilon^*$, 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 (1), we can rewrite the TBS model of (3) as

$$\log(T_i) = M_i + e_i, \tag{4}$$

where the error e_i in (4) has asymmetric density function $f_e(u|Z_i) = f_\epsilon\{g_\lambda(M_i + u) - g_\lambda(M_i)\}\, g'_\lambda(M_i + u)$, where $g'_\lambda(y) = |y|^{\lambda-1}$. The shape and scale of the cdf $F_\epsilon\{g_\lambda(M_i + u) - g_\lambda(M_i)\}$ of e_i depends on the covariates Z_i. The approximate variance of $\log T$ is $\sigma^2_\epsilon |M|^{2(1-\lambda)}$, where f_ϵ has finite variance σ^2_ϵ. It is clear that unlike the usual assumption of the transformation model of (1) and Bayesian models of, say, Hanson & Johnson (2002), the median as well as the shape and scale of the error density $f_e(\cdot|Z_i)$ in (4) depend on the covariate Z_i. This allows our model to be useful for dealing with heteroscedasticity of $\log T$. Thus, unlike the existing Bayes models, the covariate Z does affect the scale and shape of the f_e in our TBS models. A parametric log-normal model with location $M(Z) = \beta' Z$ for $\log(T)$ is a special case of (3) with $\lambda = 1$ and F_ϵ being $N(0, \sigma^2)$. The hazard function $h(t|Z) = -\frac{d}{dt}\log\{P(T > t|Z)\}$ of (3) can be non-monotone; for example, a log-normal model has non-monotone hazard.

Although the model in (3) apparently focuses on modeling the median, we can easily obtain other quantiles of $\log(T)$. For the TBS model of (3), the α-quantile $Q_\alpha(Z)$ of T is

$$Q_\alpha(Z) = \exp\{M^*_\alpha(Z)\} = \exp\left[g^{-1}_\lambda\{g_\lambda(\beta' Z) + \epsilon^*_\alpha\}\right], \tag{5}$$

because $P[g_\lambda\{\log(T)\} < g_\lambda(M) + \epsilon^*_\alpha \,|Z] = \alpha$ for $\alpha \in (0, 1)$, where ϵ^*_α is the α-quantile of $f_\epsilon(\cdot)$ with $P(\epsilon < \epsilon^*_\alpha) = \alpha$. For $\alpha = 0.5$, we have $\epsilon^*_{0.5} = 0$ and get the log-linear median function $\exp(\beta' Z)$ for T in (3). The expression in (5) shows that this model is very convenient for simultaneously estimating all important quantiles of T_i using the estimates of $(\lambda, \beta, \epsilon^*_\alpha)$. However, unlike the existing methods including those of Portnoy (2003) and Peng and Huang (2008), $Q_\alpha(Z)$ of the TBS model in (5) is not linear in covariate Z unless $\alpha = 0.5$ (median). The Bayesian models of Kottas and Gelfand (2001) and Hanson & Johnson (2002) also have linear quantile functions $M_\alpha(Z) = \beta'_\alpha Z$ of $\log T$ for all $1 > \alpha > 0$, and they are parallel to each other (with only the intercept of β_α different for different $\alpha \in (0, 1)$).

The expression in (5) for the TBS model also implies that $Q_\alpha(Z_i) \le Q_\alpha(Z_j) \Leftrightarrow Q_{\alpha'}(Z_i) \le Q_{\alpha'}(Z_j)$ for all $\alpha, \alpha' \in (0, 1)$. This means that under the model in (3), 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.

3. Likelihood, Prior Process and Inference

Let T_i and C_i be the survival and censoring times, respectively, for $i = 1, \cdots, n$. We observe (t_{i0}, δ_i), where $t_{i0} = T_i \wedge C_i$ is the observed follow-up time and δ_i is the censoring

indicator, with $\delta_i = 1$ 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 $\mathbf{y}_0 = (\mathbf{t_0}, \delta^*)$ with $\mathbf{t_0} = (t_{10}, \ldots, t_{n0})$ and $\delta^* = (\delta_1, \ldots, \delta_n)$, the likelihood function under our TBS model of (3) is as follows:

$$L(\beta, \lambda, F_\epsilon|\mathbf{y}_0) \propto \prod_{i=1}^{n} \left\{ |y_i|^{\lambda-1} \, dF_\epsilon \left(\omega_i\right) \right\}^{\delta_i} \left\{1 - F_\epsilon \left(\omega_i\right)\right\}^{1-\delta_i} , \tag{6}$$

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

In general, for the parametric versions of TBS model, any unimodal symmetric distribution, such as the Gaussian and logistic, can be used for F_ϵ. For example, $f_\epsilon(w)$ and $F_\epsilon(w)$ will be respectively replaced by the density $\phi_\sigma(w)$ and cdf $\Phi_\sigma(w)$ of $N(0, \sigma^2)$ for the Gaussian TBS model likelihood in (6). The corresponding posterior is $p(\tau, \sigma|\mathbf{y}_0) \propto L(\tau, \sigma|\mathbf{y}_0)\pi(\tau, \sigma)$, where $\pi(\tau, \sigma)$ is the joint prior density based on the available prior information, with $\tau = (\beta, \lambda)$. Markov Chain Monte Carlo (MCMC) samples from this joint posterior can be used to implement a parametric Bayesian analysis. Under this parametric model, the maximum likelihood estimator (MLE) of the regression parameters β can be obtained via maximizing the log-likelihood $L(\tau, \sigma|\mathbf{y}_0)$. For example, the log-likelihood function of the (Gaussian) parametric TBS model is

$$\begin{aligned} \ell(\beta, \lambda, \sigma|\mathbf{y}_0) = \sum_{i=1}^{n} \{ & \delta_i \log \phi_\sigma(\omega_i) + \delta_i(\lambda - 1) \log(|y_i|) \\ & + (1 - \delta_i) \log \overline{\Phi}_\sigma(\omega_i) \} , \end{aligned} \tag{7}$$

where $\overline{\Phi}_\sigma(\omega) = 1 - \Phi_\sigma(\omega)$ is the survival function of $N(0, \sigma^2)$. The maximum likelihood estimator (MLE) of the parameters under parametric TBS model is obtained via maximizing the corresponding log-likelihood function $\ell(\beta, \tau|\mathbf{y}_0)$ using Newton-Raphson (NR) iterations. Under mild regularity conditions, the MLE of β (as well as the parametric Bayes estimator) is consistent and asymptotically efficient based on regular large sample theory for the MLE when the modeling assumption is correct.

Any parametric assumption about F_ϵ in (3) is deemed as a restrictive parametric assumption for some data examples in practice. In the semiparametric version of (3), the unimodal symmetric density of ϵ is assumed unknown. For semiparametric maximum likelihood estimation (SPMLE) under this model, the likelihood of (6) is maximized with respect to the restriction that F_ϵ is the cdf of a unimodal distribution symmetric around 0. The regularity conditions and asymptotic issues for the SPMLE under (6) are nontrivial and beyond the scope of this paper. For semiparametric Bayesian analysis, we need the posterior

$$p(\tau, F_\epsilon|\mathbf{y}_0) \propto L(\tau, F_\epsilon|\mathbf{y}_0)\pi_{12}(\tau)\pi_3(F_\epsilon) , \tag{8}$$

where π_{12} and π_3 are independent priors of $\tau = (\beta, \lambda)$ and F_ϵ. This uses the simplifying, however reasonable, assumption that the prior opinions about parametric vector τ and nonparametric function F_ϵ can be specified independently. We will discuss the practical justification of this assumption later.

Using the following result of Feller (1971, p.158), we introduce a class of nonparametric priors π_3 defined over the space of symmetric unimodal distribution functions F_ϵ in (3). Any symmetric unimodal distribution F_ϵ can be expressed as a scale-mixture of uniform random variables

$$F_\epsilon(u) = \int_0^\infty \zeta(u|\theta) \, dG(\theta) \tag{9}$$

for some mixing distribution $G(\theta)$, where $\zeta(u|\theta)$ for $\theta > 0$ is the uniform distribution with support $(-\theta, +\theta)$. We use the Dirichlet process (DP) of Ferguson (1973), $G \sim DP(G_0, \nu)$, as a nonparametric prior for the unknown scale-mixing distribution $G(\theta)$ of (9). The $DP(G_0, \nu)$ is characterized by the known "prior guess" G_0 (the prior expectation of G), and a positive scalar parameter ν, the precision parameter around the prior mean/guess G_0. The prior mean G_0 of the random mixing density G can be chosen appropriately to assure a desired prior mean/guess F_* for unknown F_ϵ. Using a result by Khintchine (1938), when the density $f_*(\cdot)$ and its derivative $f'_*(\cdot)$ exist, the density $G'_0(\theta)$ of $G(\theta)$ is given as

$$G'_0(\theta) = -2\theta\, f'_*(\theta) \;\text{ for } \theta > 0\,. \tag{10}$$

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_ϵ, using (10), 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 ν also determines the degree of belief about how close F_ϵ should be to its prior guess F_*. When ν is large enough, the unknown nonparametric F_ϵ is very close to its pre-specified (often parametric) prior mean/guess $F_*(\cdot|\gamma)$. A small ν implies very little confidence in unknown F_ϵ being close to $F_*(\cdot|\gamma)$, and the corresponding Bayes estimator of β is expected to be very close to the semiparametric likelihood estimator. The details of the specifications of the hyperparameters of the priors π_{12} and π_3 in (8) are provided in the next section.

4. Data Analysis

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); 62 cancer patients ($z_1 = 1$) were randomized to arm A (C followed by E) and 59 patients ($z_1 = 0$) to arm B (E followed by C). Apart from the treatment indicator z_1, another covariate is the patient's age at entry (z_2) centered at age 50. Each survival time (given in months) was either observed ($\delta_i = 1$) or administratively censored ($\delta_i = 0$). To evaluate the age-adjusted treatment difference, we consider the linear regression function $M_i = \beta_0 + \beta_1 z_{1i} + \beta_2 z_{2i}$. The maximum likelihood estimates of the regression parameters β under the parametric TBS model (3) with Gaussian F_ϵ are given by $\hat{\beta}_0 = 3.349$, $\hat{\beta}_1 = 0.433$, $\hat{\beta}_2 = -0.019$ with $\hat{\lambda} = 0.082$.

Now we present a parametric Bayesian analysis using the TBS model of (3) with parametric $N(0, \sigma^2)$ density for F_ϵ. One major advantage of the TBS model for Bayesian analysis is that the priors for the parameters $(\beta_0, \beta_1, \beta_2, \lambda, \sigma)$ 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^*). Without loss of generality, we assume that the priors are based on the following: (1) Prior guess and prior range of a quantile, say, the median, of the prior-predictive survival time T^* of a patient at age 50 ($z_2^* = 0$) from treatment arm B ($z_1^* = 0$); (2) Change in the median of T^* for a unit change in each age (z_2) and treatment (z_1). We point out that for most Phase 2 and 3 trials, these quantities are routinely elicited and used to design the trial and determine the power for detecting differences (e.g., Pintadosi, 1997). We first demonstrate the specification of these priors for the parametric TBS models.

We use the simplifying assumption that the joint prior is $\pi(\beta, \lambda, \sigma) = \pi_1(\beta)\pi_2(\lambda)\pi_3(\sigma|\beta, \lambda)$. This assumption can be justified in practice because the prior $\pi_1(\beta)$ is based on the median (location) of T^*, whereas the prior $\pi_2(\lambda)$ is based on the shape (skewness) of $\log(T^*)$. The specification of the prior for β_0 uses the fact that T^* with $z_1^* = z_2^* = 0$ has a prior median $\exp(\beta_0)$. For the lung cancer trial conducted before 1993, the current expert opinions about SCLC are not very appropriate. Based on the published literature about the treatment

of SCLC before this study (e.g. Jett et al., 1990; Evans et al., 1987; Comis, 1986), the median survival time for treatment arm B was thought to be between 12 to 17 months for limited-stage and 9 to 10 months for extensive-stage SCLC patients. For our SCLC study with nearly equal proportions of these two types of patients, we use a mean prior guess of 13 months and a range of (8, 18) months for T^*. These give us the prior $\beta_0 \sim N(A_1, B_1^2)$ with $A_1 = \log(13)$ and $B_1 = \{\log(18) - \log(8)\}/3$ to ensure that the prior range of β_0 has approximate length $3B_1$. Our prior opinion about β_1 is based on the prior belief about 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, but, from different treatment arms. So, the prior $\beta_1 \sim N(0, 10)$ corresponds to a 95% prior probability that the ratio of medians e^{β_1} has range $(e^{-2\sqrt{10}}, e^{2\sqrt{10}})$ and is centered at $e^0 = 1$ (indifferent opinion regarding superiority of either treatment arm). Similarly, the prior $\beta_2 \sim N(0, 10)$ corresponds to prior opinion that two patients from treatment B and with 1 year difference in age, have a ratio of medians between $(e^{-\sqrt{10}}, e^{\sqrt{10}})$ with 68% probability. We have chosen such a non-informative prior opinion about β_1 and β_2 to allow for a meaningful comparison of our analysis results with results from frequentist and previous Bayes methods based on either no prior or a non-informative prior. We would like to point out that our point-wise Bayes estimates do not change substantially ($< 4\%$ change) when we reduce the prior variances of β_1 and β_2 to 1 (instead of 10). The interval estimate of β_1 (as an example) is around 12% narrower when we use these more skeptical $N(0, 1)$ priors instead of $N(0, 10)$ priors for β_1 and β_2.

We use the $Unif(0, 3)$ prior for $\pi_2(\lambda)$ because it is difficult to interpret the after-transform linear model of (3) when $\lambda > 3$. In their original paper, Box and Cox (1964) recommended restricting the $\lambda \leq 2$. For a parametric Gaussian TBS model, $\log T^*$, when $z_1^* = z_2^* = 0$, can be expressed approximately as $\log T^* \simeq \beta_0 + \sigma|\beta_0|^{1-\lambda} e$ (Kettl, 1991), where β_0 is the median of $\log T^*$ and $e \sim N(0, 1)$. This allows us to obtain prior $\pi_3(\sigma|\beta_0, \lambda)$ based on prior opinion of $M_{\alpha^*}^*$ because $|\beta_0 - M_{\alpha^*}^*| \simeq \sigma|\beta_0|^{1-\lambda}|e_{\alpha^*}^*| \Rightarrow \sigma \simeq \frac{|\beta_0 - M_{\alpha^*}^*|}{|\beta_0|^{1-\lambda}|e_{\alpha^*}^*|}$, where $M_{\alpha^*}^*$ is another quantile of $\log T^*$ for $\alpha^* \neq 1/2$, and $e_{\alpha^*}^*$ is the α^*-percentile of standard normal. For example, when we take $\alpha^* = 0.75$, we have $\sigma \simeq |\beta_0 - M_{0.75}^*|\ |\beta_0|^{\lambda-1}/0.6745$. Based on the SCLC literature prior to this trial, we use the prior opinion that the third-quartile $\exp(M_{0.75}^*)$ of a patient in treatment arm with 50 years entry-age is between 10 months to 5 years with a center of 33 months. For given (β_0, λ), we use a Gamma density at the prior $\pi_3(\sigma|\beta_0, \lambda)$ with mean equal to $|\beta_0 - \log(33)||\beta_0|^{\lambda-1}/0.6745$ and approximate range between 0 and to $(\log(60) - \log(10))|\beta_0|^{\lambda-1}/0.6745$. These prior densities give us approximately the same means and ranges of $M_{0.5}^* = \beta_0$ and $|\beta_0 - M_{0.75}^*|$ that we expect from our prior opinion about these two quantiles of $\log(T^*)$. However, to simplify this further, we use an unconditional Gamma prior $\pi_3(\sigma)$ whose mean equals to $\frac{|\log(13) - \log(33)|}{0.6745}$ and variance equals to $\frac{|\log(13) - \log(60)|}{0.6745}$ (based on prior mean $\log(13)$ for β_0 and prior guess 1 for λ). We found no noticeable difference in posterior estimates using this unconditional prior for σ instead of a conditional prior $\pi_3(\sigma|\beta_0, \lambda)$. 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 the median survival time of small cell lung cancer can be very different from what we used, and that may 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 $\log(T_i)$ (subject to censoring) and $y_i^* = E[\log(T_i)|z_{1i}, z_{2i}; \mathbf{y}_0]$ is the posterior predictive expectation of $\log(T_i)$ under the model, does not show any trend of residuals under the parametric Bayes TBS model. Our plot (right-hand panel of Figure 1) of these residuals versus the estimated median survival times also does not reveal any serious inadequacy of the parametric TBS model. However, the Q-Q plot (Figure 2) of these

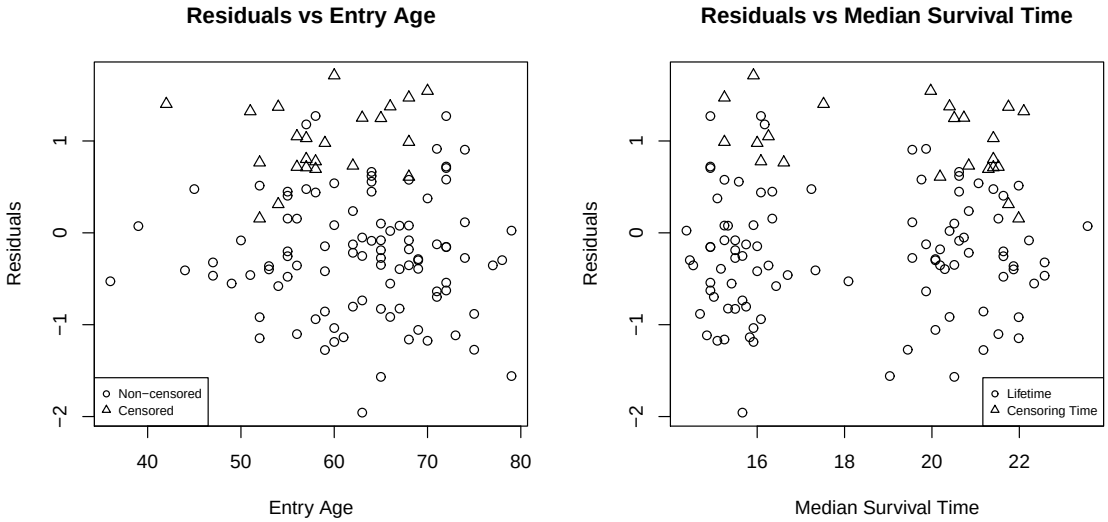


Figure 1: Plots of residuals versus the age at entry (in years) and versus the estimated median survival time (in months) using parametric TBS model for the lung cancer data.

residuals suggests that the assumption of Gaussian distribution for F_ϵ in (3) is questionable due to the plot being non-linear at the right tail. Later, we use a semiparametric Bayesian analyses to avoid the Gaussian assumption of ε_i. 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 estimated quantiles of treatment B ($z_1 = 0$) at any age z_2.

For the semiparametric Bayesian analysis with a symmetric unimodal f_ϵ in (3), we need to specify the prior guess/mean F^* of F_ϵ and a prior precision parameter ν. We take the precision parameter $\nu = 1$ to imply a very low confidence around our parametric prior guess F_* of the nonparametric error distribution F_ϵ. We take the prior mean f_* of f_ϵ to be $N\{0, (\sigma_0)^2\}$ where $\sigma_0 = \frac{|\log(60) - \log(10)|}{0.6745}$. This makes f_* equal to the prior mean of f_ϵ used for the parametric Bayes analysis of the TBS model. Using (10), this $N\{0, (\sigma_0)^2\}$ density for f_* corresponds to a $Gamma(3/2,\ 1/\{2(\sigma_0)^2\})$ for G_0 in (10). The constructive definition of the DP mixture prior process for F_ϵ is $F_\epsilon(u) = \sum_{k=1}^{\infty} p_k \zeta(u|\theta_k)$ (Sethuraman, 1994), where $\theta_k \overset{i.i.d.}{\sim} G_0$, $p_k = V_k \prod_{j=1}^{k-1}(1 - V_j)$ with $V_j \overset{i.i.d.}{\sim} Beta(1, \nu)$. The actual implementation of the MCMC tool to sample from (8) is based on a finite approximation $F_\epsilon(u) \simeq \sum_{k=1}^{K} p_k \phi(u|\theta_k)$ of Sethuraman's construction with, say, $K = 1,000$ and $V_K = 1$. The MCMC computational tool can be implemented, even via a standard package such as Winbugs. The rest of the conditional posteriors are the same as those used for the parametric Bayes.

We get the semiparametric Bayes point estimates $\hat{\beta}_0 = 3.086$, $\hat{\beta}_1 = 0.304$ and $\hat{\beta}_2 = -0.006$ for $(\beta_0, \beta_1, \beta_2)$ along with 95% credible intervals $(2.836, 3.315)$, $(0.003, 0.577)$ and $(-0.022, 0.011)$ respectively, with $\hat{\lambda} = 0.629$. The results of the Bayes estimators of regression parameters (β_1 and β_2) under parametric and semiparametric TBS models along with the ML estimator based on a 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) using the model of (4) with $f_e(u) = (1/2)\eta^{sgn(u)} f_0(\eta^{sgn(u)}|u|)$ for a nonparametric density $f_0(u)$ defined on $u > 0$.

The point estimates of the regression parameters of the median functional under different methods are not strikingly different to the corresponding point estimator obtained

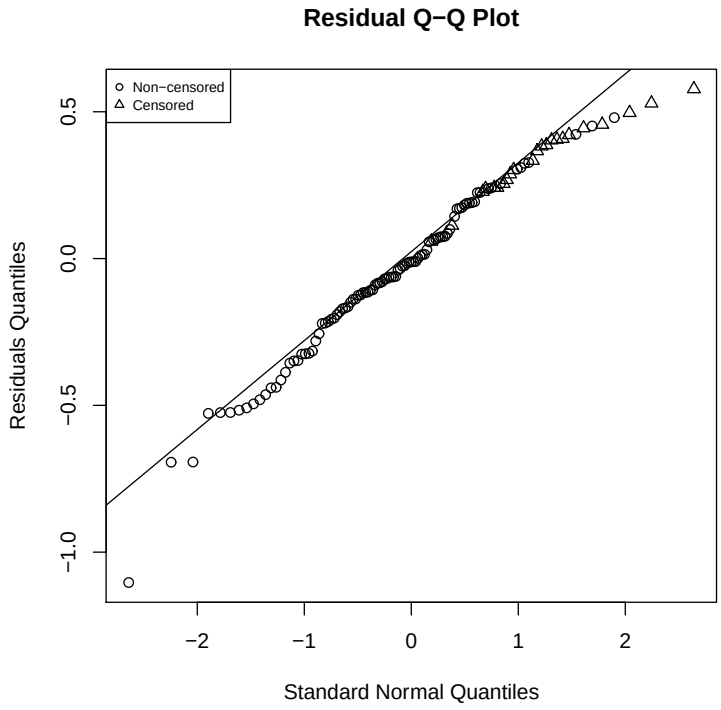
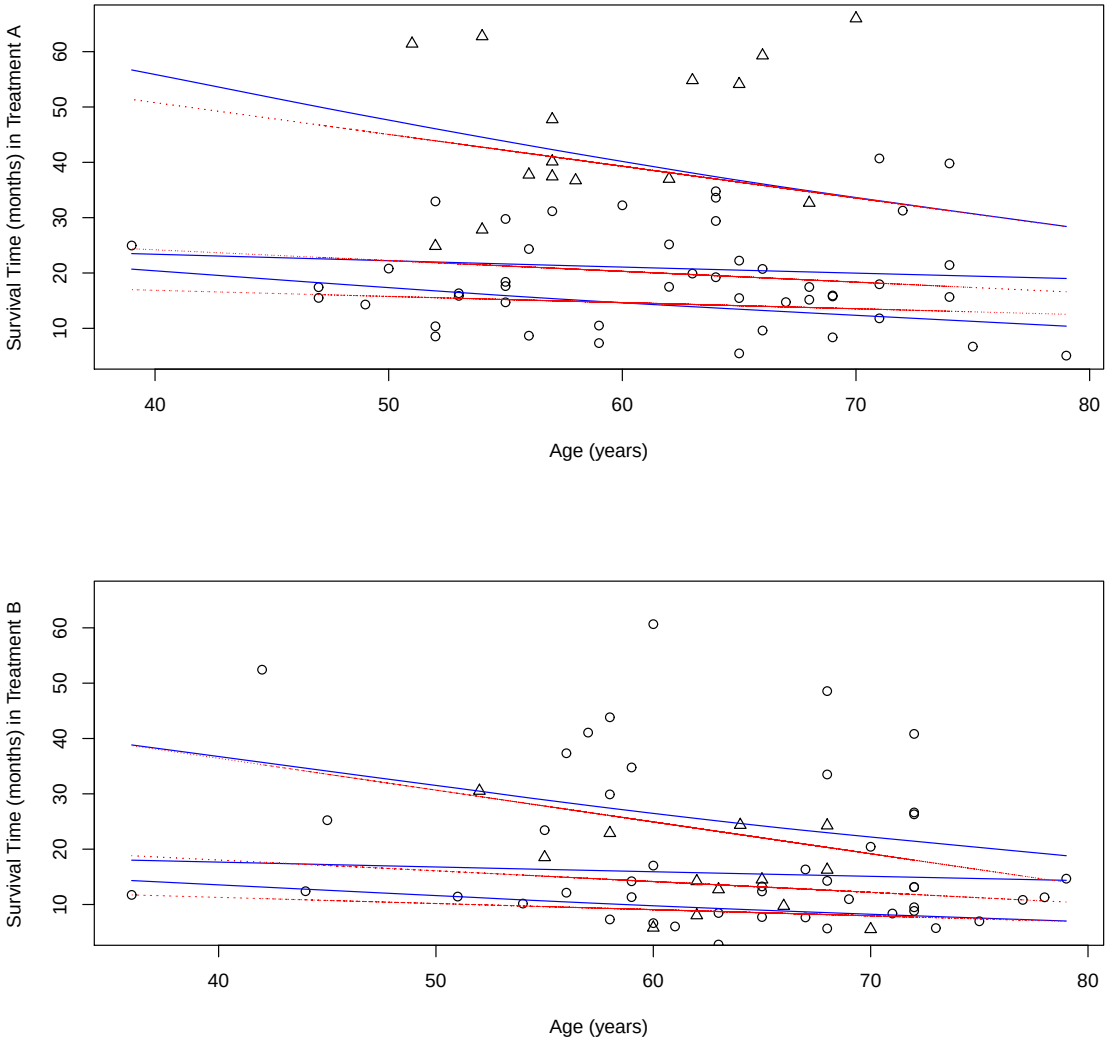


Figure 2: Q-Q plots of the residuals under parametric TBS model for the lung cancer data

Table 1: Pointwise and 95% interval estimates (within parenthesis) of regression parameters (β_1 for treatment z_1 and β_2 for age z_2) for the lung cancer study under different procedures

Estimator	Treatment	Age
MLE (TBS model)	0.433 (0.141, 0.727)	-0.019 (-0.037, -0.002)
Parametric Bayes (TBS)	0.318 (0.036, 0.604)	-0.008 (-0.023, 0.008)
Semiparametric Bayes (TBS)	0.304 (0.083, 0.577)	-0.009 (-0.021, -0.002)
Portnoy	0.369 (0.149, 0.591)	-0.009 (-0.031, 0.012)
KG Bayes	0.389 (0.037, 0.845)	-0.018 (-0.028, -0.007)



Portnoy's method; △: censored observation)

via Portnoy's method (2003). This is also evident from Figure 3, where 3 estimated quantiles for Portnoy's method (dotted straight lines) and for semiparametric Bayes TBS model (solid curved lines) are plotted (separately for 2 treatment arms).

Figure 3: Plots of observed survival times versus Age (z_2) with three estimated quartile functions for two treatment arms. (Solid lines: estimated via TBS model; Dotted straight lines: estimated via
Portnoy's method; △: censored observation)

We find that the proportion of observations in each quantile-interval is closer to the expected proportions for Bayes estimates of quantile functions compared to Portnoy's. However 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 estimates based on TBS models have smaller estimated standard errors for the treatment effect compared to competing procedures. The widths of the interval estimates from parametric and semiparametric Bayes are substantially smaller than widths of the corresponding estimates based on Portnoy's method (at least for the age-effect). This is

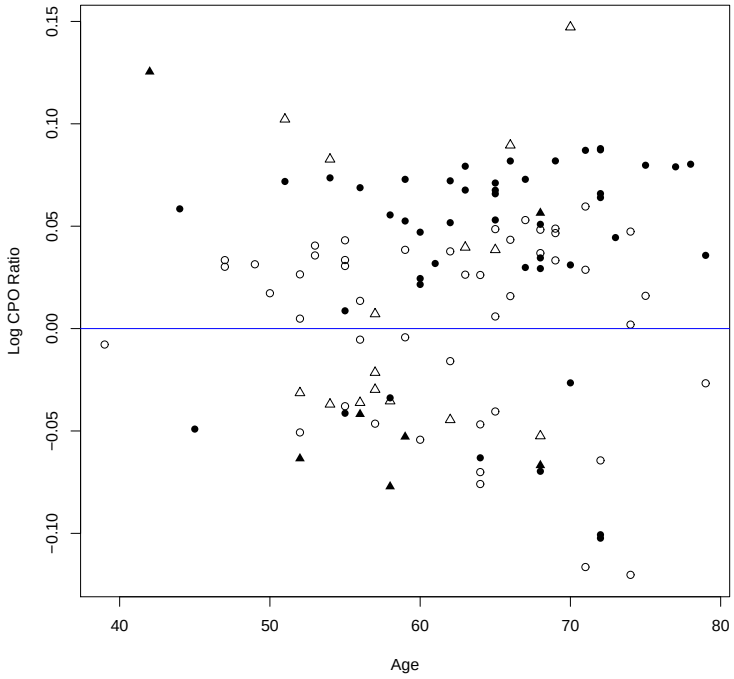


Figure 4: Plot of the log-ratio of two CPOs obtained from semiparametric TBS and Gaussian TBS model (y-axis), versus Age (x-axis): ○ uncensored from treatment A; △ censored from treatment A; ● uncensored from treatment B; ▲ censored from treatment B)

not surprising because Portnoy's median regression methods have a far larger number of regression parameters than the finite dimensional regression parameter β in (3). The posterior standard deviations of the TBS estimators are also smaller than those from Kottas and Gelfand (2001). Figure 4 plots the logarithm of the ratio of the CPO (Conditional Predictive Ordinate) of the semiparametric TBS model and CPO of the Gaussian TBS model against the observation numbers. A value greater than 0 for this supports a semiparametric model over a Gaussian model. In this example, approximately 67% of observations favor the semiparametric TBS model over the Gaussian TBS model, i.e., a substantially higher proportion of observations supporting the semiparametric model over parametric model. The final conclusion is that semiparametric model fits the data better than other competing parametric models for entire range of age and for both treatments.

5. Simulation Study

For our simulation models, we set the median of $Y = \log(T)$ given Z to be $M(Z) = \beta_0 + \beta_1 Z = 6.5 + Z$, i.e., $\beta_0 = 6.5$ and $\beta_1 = 1.0$, where Z can take four possible values $0, 0.5, 1.0,$ and 1.5, in equal proportions for each simulated data set. For each simulation distribution of T considered in the study, we simulate at least 5000 datasets with sample sizes $n = 80, 160,$ and 320. The number of simulated datasets for different sample sizes may vary to assure that the Monte Carlo variability of the approximate bias and MSE of the regression estimates are smaller than 0.01.

For the simulation study, the Bayes estimators considered by us are based only on the semiparametric model of (3). The priors used for Bayes estimation in the simulation study

are: $\beta_0 \sim N(6, 10)$ and $\beta_1 \sim N(0, 1)$. The prior mean for the Dirichlet process is $N(0, 1)$ and the precision is $\nu = 0.01$. This prior for β_1 implies that there is almost 5% prior probability that the ratio of medians is larger than 7.4 for a unit change in z. In order to avoid undue influence of the choice of the priors on the results of our simulation study, we use these somewhat vague priors here. However, the prior can also be viewed as a skeptical prior because the prior of the regression parameters is centered at the prior guess of no-covariate effect ($\beta = 0$). If a Bayes estimator can demonstrate good performance for detecting covariate-effects with this prior, this suggests that even a skeptical and unusually "flat" prior may not hinder the Bayes method's ability to detect the covariate effect. The implications of chosen priors for multiple model parameters are best described via various summaries of the prior predictions of the observables/responses. We generate various summary statistics including the sample median, range and width of the range of 500 survival times using a single set of parameters simulated from the joint prior. We then replicate the whole process of simulating these summary statistics 1000 times. We found the range of these 1000 sample medians is between 1400-5200 for $z = 1.5$, compared to the true median of $\simeq 2981$ for the simulation model. The range of survival times from the prior predictive models may have width as large as 10^8. These summaries indicate that our prior predictive models are very non-informative and can cover a wide range of survival patterns. In practice, we expect to use a more informative prior predictive model using often available information about the range of responses (even after incorporating a skeptical prior view about the covariate effect).

First we evaluate the robustness of the maximum likelihood estimators (MLE) and of the Bayes estimates based on (3). We compare performances (bias and MSE) of these estimators to the competing frequentist estimator of Portnoy (2003). For this aim, we simulate survival data from parametric exponential and Pareto densities. Both exponential and Pareto simulation densities, being heteroscedastic and skewed for all λ, do not satisfy the assumptions of (3). The independent censoring distribution was generated from an exponential density $(\Lambda e^{\Lambda C})$ with rate parameter Λ chosen to obtain desired proportions of censoring. For example, the choice of $\Lambda = \log(2)/30$ results in approximately 20% censoring for exponential simulation model.

Table 2 presents the summary of the approximate sampling mean and mean-square-error (MSE) of various competing estimators of β_1 under different simulation models. Results in Table 2 under an exponential and Pareto simulation model show that the MLE based on (3), and the Bayes estimators based on (3) have comparable biases relative to competing estimators. Further, the MSE of Portnoy's estimators are much larger than the corresponding MSE of the MLE and Bayes estimators. The Bayes estimators under (3) have much smaller MSE compared to the MLE.

For Pareto simulation model, $g_\lambda(Y)$ has an extremely skewed and heavy-tailed density for all values of λ. In smaller samples $(n \leq 160)$, Portnoy's estimator has the most bias. For the largest sample size $(n = 320)$, the bias of the Gaussian MLE $\hat{\beta}_1$ is highest. The bias of semiparametric Bayes estimators have the smallest bias for all samples, and also have much smaller MSE than other competing estimators.

For the last part of Table 2, we investigate the performance of the semiparametric Bayes estimator using data simulated from a TBS model of (3) with $\lambda = 0.5$ and double-exponential density for ϵ. We see that the Bayes estimators have substantial improvement in MSE compared to competing estimators. The bias of the MLE under the Gaussian TBS model is similar for $n = 160$ and $n = 320$.

In summary, when the distribution of $\log(T)$ after an optimal transformation has a moderate degree of asymmetry, the MLE and Bayes estimators based on (3) have finite sample biases very similar to that of Portnoy (2003)'s estimator. More importantly, the precision

Table 2: Results of simulation study under Exponential and Pareto models: Monte Carlo approximation of the sampling mean and Mean Square Error (MSE) of different estimators of known $\beta_1 = 1$

Simulation		Gaussian TBS MLE		Portnoy		SP TBS	
Model	Sample	Mean	MSE	Mean	MSE	Mean	MSE
Exponential	80	0.91	2.66	0.93	4.27	0.92	0.90
	160	0.97	1.35	1.11	2.28	1.08	0.65
	320	0.94	0.69	0.96	1.20	0.93	0.48
Pareto	80	1.03	12.01	1.10	19.89	1.03	0.95
	160	0.93	5.41	0.91	8.60	1.01	0.85
	320	0.92	2.68	0.98	4.25	1.02	0.68
TBS (Double Exponential)	80	0.99	1.94	1.01	1.52	1.04	0.72
	160	0.96	0.97	0.98	1.69	0.97	0.48
	320	0.97	0.51	0.98	1.35	1.03	0.30

of the Bayes estimators based on TBS is better even when the underlying assumptions of (3) are not entirely valid. However, the MLE's performance depends on the degree of symmetry of the distribution of $g_\lambda(Y)$ under optimal λ. The semiparametric Bayes estimators have excellent biases and smallest MSE among all of its competitors. When the modeling assumption of (3) is correct, the Bayes estimator based on (3) shows much smaller MSE compared to any competing estimators. This implies that the semiparametric Bayes estimator based on (3) is a safer and more robust estimator to use in practice compared to its competitors.

6. Discussion

In this paper, we present a new class of semiparametric models amenable to Bayes estimation of the log-linear median regression function for censored survival data. Similar to previous semiparametric models (e.g., Cox's model), our model has a finite dimensional parameter vector and one non-parametric symmetric unimodal function f_ϵ. We argue that our assumption of unimodality of f_ϵ justifies the importance of median as the location parameter of interest. Previous research, including Box and Cox (1964), has found that the transformation in (2) is often an effective tool to obtain symmetry and accommodate heteroscedasticity. Our method can be applied when the covariate Z affects the location as well as the scale and shape of $\log(T)$.

Median regression offers a useful alternative to the popular regression functions of Cox (1972) and the transformation model of (1). There is a substantial literature on median regression for censored survival data, including Ying et al. (1995), Yang (1999), McKeague et al. (2001) and Bang and Tsiatis (2003). These methods involve non-linear discontinuous estimating equations that are difficult to solve, often with multiple solutions. The recursive nature of some of these methods (e.g. that of Portnoy (2003)) make the asymptotic justifications and computations complicated. Peng and Huang's (2008) martingale based estimating equations involve minimization of an L_1-type discontinuous convex functions. Unlike estimation with Cox's model (Cox, 1972), 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 $P\{T > Q_\alpha(Z)\} = \alpha$. Unlike our model of (3), these frequentist linear quantile models have an infinite number of regression parameters β_α for all $\alpha \in (0,1)$. As a consequence, unlike the model of (3), there is no simple expression available for survival functions for these models. For more in-depth discussion about the implementation, comparisons, asymptotic rate of convergence and consequences of the restrictive assumptions for existing quantile regression approaches, we suggest the excellent review by Koenker (2008). This restrictive assumption of linearity of all quantile functions may not hold true for any real study and very few known stochastic models can satisfy this, except when $\log T = M + M_i^* e$ with $e \sim N(0, \sigma^2)$, $M_i^* = \gamma Z$ and $M_i = \beta Z$ (Kettl, 1991). As an alternative to the semiparametric model of (3), we can also consider $g_\lambda\{\log(T_i)\} = g_\lambda(M_i) + |M_i|^\gamma \eta_i$ with a symmetric unimodal density for η_i. However, both of these models are less parsimonious than (3) due to using separate parameters to address skewness and heteroscedasticity. Our preliminary simulation studies (omitted due to brevity) also cast doubts about the practical advantages of these alternatives to (3).

Existing Bayesian median regression models of Kottas and Gelfand (2001) and Hanson & Johnson (2002) have the linear representation of (4) with $f_e(u)$ free of covariate Z. As a consequence, all quantile functions of $\log T$ are linear with the same slope (regression coefficient) for each covariate. As we mentioned before, these previous Bayes models cannot accommodate heteroscedasticity of $\log(T_i)$, a very common phenomena in most popular survival models including Weibull and Cox's model (1972). We believe that our models achieve a sensible compromise between existing frequentist and Bayesian models via accommodating heteroscedasticity while not restricting to linear functional for all quantiles. Note that the parametric MLE $\hat{\beta}$ based on an assumed Gaussian ϵ yields a consistent quasi-likelihood estimator of β as long as the true ϵ is symmetric around 0 (even if it is not Gaussian); the variance of $\hat{\beta}$ can be estimated using the so-called "sandwich" variance estimator (White 1982). The loss of efficiency of this estimator under a non-Gaussian model is beyond the scope of this paper.

Although we focus on modeling the median functional, our method can be used to compute the joint confidence band of any other quantile functional via (5) involving $(\beta, \lambda, \epsilon^*_\alpha)$. For brevity, we have omitted the results of our simulation study showing an excellent accuracy of joint confidence bands of all these quantile functions $(Q_{0.25}(z), Q_{0.5}(z), Q_{0.75}(z))$ under the Bayes TBS model of (3) (even when the simulation model is Pareto). For some diseases, such as cancers with very good prognosis, the main interest may center on modeling the quantile $Q_\alpha(Z)$ as a log-linear function with $P\{T < Q_\alpha(Z)\} = \alpha$ for $\alpha > 1/2$ (different than the median). In this case, we can use a modification of (3) with assumptions $P(\epsilon < 0) = F_\epsilon(0) = \alpha$ and $\zeta(u|\theta)$ (9) being the uniform density with support $\{2\theta(\alpha - 1),\ 2\theta\alpha\}$. For the sake of brevity, we again omit the details of the rest of the methodology and related MCMC steps. Our methods can also predict the outcome of a future patient with known covariate values. 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 methods for censored data.

REFERENCES

Bang, H. and Tsiatis, A. A. (2003). Median regression with censored cost data. *Biometrics* **58,** 643-649.

Bickel, P.J. and Doksum, K.A. (1981). An analysis of transformations revisited. *J. Amer. Statist. Assoc.* **76,** 296–311.

Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B* **26,** 211–243.

Carroll, R.J. and Ruppert, D. (1984) Power-transformations when fitting theoretical models to data. *J. Amer. Statist. Assoc.* **79,** 321–328.

Cheng, S.C., Wei, L.J. and Ying, Z. (1995) Analysis of transformation models with censored data. *Biometrika* **82,** 835–845;.

Comis R.L. (1986), Clinical trials of cyclophosphamide, etoposide and vincristine in the treatment of small-cell lung cancer, *Semin Oncol* **13,** 40–44.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34,** 187–200.

Evans, W.K., Feld, R., Murray, N., et al. (1987), Superiority of alternating non-cross-resistant chemotherapy in extensive small cell lung cancer, *Ann Intern Med* **107,** 451–458.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*. Wiley.

Ferguson, T.S. (1973), Bayesian Analysis of some nonparametric problems, *The Annals of Statistics* **1,** 209–230.

Fine, J.P., Ying, Z. and Wei, L.J. (1998), On the linear transformation model with censored data, *Biometrika* **85,** 980–986.

Fitzmaurice, G.M., Lipsitz, S.R. and Parzen, M. (2007) Approximate median regression via the Box-Cox transformation. *The American Statistician*. **61,** 233–238.

Hanson, T. and Johnson, W.O., (2002), Modeling regression error with a mixture of Polya trees, *J. Amer. Statist. Assoc.* **97,** 1020–1033.

Hanson, T. (2006), Inference for mixtures of finite Polya tree models, *J. Amer. Statist. Assoc.* **101,** 1548–1565.

Ibrahim, J.G., Chen, M.-H., Sinha, D. (2001). *Bayesian Survival Analysis*. Springer-Verlag.

Jett JR, Everson L, Therneau TM (1990), Treatment of limited-stage small-cell lung cancer with cyclophosphamide, doxorubicin and vincristine with or without etoposide: A randomized trial of the North Central Cancer Treatment Group, *J Clin Oncol* **8,** 33–38.

Koenker, R. (2008), Censored quantile regression redux, *J. Statistical Software* **27**.

Khintchine, A.Y. (1938). *On Unimodal Distributions*. *Inst. Mat. Mech. Tomsk. Gos. Univ.* **2,** 1–7.

Kettl, S. (1991). Accounting for heteroscedasticity in the transform both sides regression model. *Applied statistics*, **49**, 261-268.

Kottas, A. and Gelfand, A.E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, **456**, 1458-1468.

McKeague, I.W., Subramanian, S., Sun, Y. (2001) Median regression and the missing information principle. *J. Nonparametric Statist.* **13,** 709–727.

Peng, L., Huang, Y. (2008). Survival Analysis With Quantile Regression Models. *J. Amer. Statist. Assoc.* **103,** 637–649.

Portnoy, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.* **98,** 1001–1012.

Piantadosi, S. (2005). *Clinical Trials: A Methodologic Perspective*. Wiley series in probability and statistics. Wiley-Interscience, 2nd ed..

Sethuraman, J. (1994). Constructive definition of dirichlet priors. *Statistica Sinica* **4,** 639–650.

Walker, S. and Mallick, B.K., (1999). A Bayesian semiparametric accelerated failure time model. *Biometrics* **55,** 477–483.

White, H. (1999). Maximum likelihood estimation under misspecified models. *Econometrica* **50,** 1–26.

Yang, S. (1999). Censored median regression using weighted empirical survival and hazard functions. *J. Amer. Statist. Assoc.* **94,** 137–145.

Ying, Z., Jung, S. H., Wei, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90,** 178–184.