Bayesian Analysis of Heterogeneity in Modeling of Pavement Fatigue Cracking

Lu Gao[1]; José Pablo Aguiar-Moya[2]; and Zhanmin Zhang, Ph.D., A.M.ASCE[3]

Abstract: Fatigue cracking models are important for the design, analysis, and management of pavement structures. There are two types of cracking models available in the literature. The first type of model is structured to predict time to failure (given a specific failure criterion, e.g., 25% of the total area shows cracking). The other type of model is intended to directly predict the cracking area as a function of a set of covariates. For the first type of model, previous researchers have suggested a probabilistic approach such as survival modeling (or duration modeling) to take various uncertainties into consideration. However, to the authors' knowledge, no current study considers the effect of heterogeneity when applying survival models to pavement fatigue cracking. In the context of deterioration modeling, heterogeneity can be defined as the performance difference across different individuals. Ignoring unobserved heterogeneity may lead to biased and inconsistent parameter estimates. The current paper proposes applying the frailty model to address this problem by adding a random term to the hazard function. The proposed model is demonstrated using the long-term pavement performance (LTPP) database. **DOI: 10.1061/(ASCE)CP.1943-5487.0000114.** *© 2012 American Society of Civil Engineers.*

CE Database subject headings: Flexible pavements; Fatigue; Cracking; Statistics; Heterogeneity.

Author keywords: Flexible pavement; Fatigue cracking; Bayesian statistics; Survival model; Frailty model; Heterogeneity; LTPP.

Introduction

Fatigue cracking is one of the flexible pavement distress types primarily induced from repeated traffic loading and indicates structural failure. It allows moisture infiltration and may further deteriorate to a pothole. The development of fatigue cracking will accelerate the deterioration of the flexible pavement.

Extensive research has been conducted to predict the occurrence of this type of distress using various approaches. In the literature, researchers adopt two types of models to describe the development of cracking. The first type of model attempts to predict the time (or cumulative traffic load) until the area of cracking reaches a certain level or threshold. The second type of model is intended to predict the progression of the cracking area as a function of a set of covariates that are expected to affect this area.

For example, Queiroz (1981) developed separate regression equations to predict cracking initiation, and cracking progression. The first type of model was used for the cracking initiation with the number of equivalent single axles to initiation as the dependent variable and the modified structural number as the explanatory variable. The cracking progression uses the second type of model with structural and age parameters as the explanatory variables. Paterson (1987) also used a similar approach to predict the initiation and progression phase of cracking. However, to take into account the effect of aging and the inherent variability of material behavior, a survival model was adopted as an extension of the first type of model. The application of a survival model to predict cracking initiation can also be found in Shin and Madanat (2003) and Nakat and Madanat (2008).

Recently, researchers have tended to use the long-term pavement performance program (LTPP) data for the development of fatigue cracking models. The LTPP is the largest pavement performance research program ever undertaken. Starting from 1987, the LTPP program has been monitoring more than 2,400 asphalt and Portland cement concrete pavement test sections across the United States and Canada. Fatigue cracking of flexible pavements has been continuously monitored on many LTPP test sections, along with pavement structures, loading, materials, environment, and other pieces of information.

When using LTPP data, most researchers use the second type of model for fatigue cracking. For example, Ker et al. (2008) developed a generalized linear model for fatigue cracking by using the critical tensile strain under the asphalt concrete surface layer as one of the explanatory variables. Rahim et al. (2009) developed a separate fatigue cracking regression model for different climatic regions by using the percentage of the total area affected by fatigue cracking as the response variable.

In contrast, Wang et al. (2005) pointed out that the first type of model is better than the second type of model if LTPP fatigue cracking data are used. He argued that the LTPP fatigue cracking data show a large amount of variability and lacked any easily discernable trends. Observations of the LTPP fatigue cracking data revealed that most cracking in the wheel path followed the pattern that no significant cracks appear for several years and then cracks

[1]Assistant Professor, Univ. of Houston, Dept. of Construction Management, Houston, TX 77004; formerly, Ph.D. Candidate, Univ. of Texas at Austin, Dept. of Civil, Architecture and Environmental Engineering, 1 University Station C1761, Austin, TX 78712-0278 (corresponding author). E-mail: lgao5@central.uh.edu

[2]Research Unit Coordinator, National Laboratory of Materials and Structural Models, San Jose, Costa Rica; formerly, Ph.D. Candidate, Univ. of Texas at Austin, Dept. of Civil, Architecture and Environmental Engineering, 1 University Station C1761, Austin, TX 78712-0278. E-mail: jpaguiar@mail.utexas.edu

[3]Associate Professor, Univ. of Texas at Austin, Dept. of Civil, Architecture and Environmental Engineering, 1 University Station C1761, Austin, TX 78712-0278. E-mail: z.zhang@mail.utexas.edu

Note. This manuscript was submitted on May 5, 2010; approved on March 7, 2011; published online on March 9, 2011. Discussion period open until June 1, 2012; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Computing in Civil Engineering***, Vol. 26, No. 1, January 1, 2012. ©ASCE, ISSN 0887-3801/2012/1-37–43/$25.00.

appear and soon propagate to a significant level. Therefore, using the cracking area directly as a response variable is difficult. As a result, he suggested adopting a survival model that uses failure time as the response variable. Aguiar et al. (2009) also employed the first type of model for fatigue cracking using LTPP data under a multiple risk approach.

However, none of the previous researchers considered unobserved heterogeneity in the modeling process. Heterogeneity is defined as the performance difference across different pavement sections. The source of unobserved heterogeneity comes from factors that were captured by the available explanatory variables and that cannot be easily addressed because of the unavailability of good data sources and the lack of an appropriate model estimation approach. The undesirable results of missing or unsuccessfully addressing unobserved heterogeneity may lead to biased and inconsistent parameter estimates. From a deterioration modeling point of view, this contributes negatively to model accuracy (Hong 2007). In the cases of survival models, ignoring the existence of such risk factors has nontrivial consequences. It leads to a transformation of the hazard function and the coefficients for the measured covariates and requires carefulness in interpreting the hazard function and the effects of measured covariates without considering the possibility of unobserved covariates. To capture the deterioration process more accurately by taking the unobserved heterogeneity into consideration, this paper proposes using the frailty model to estimate the pavement fatigue cracking mechanism. The proposed model is demonstrated using the LTPP data in the case study.

Methodology

Survival Analysis

The primary goal in using survival models to analyze fatigue cracking data is to assess the dependence of time to failure attributable to fatigue cracking on external variables. One way to explore the relationship of covariates to time to failure is by using a regression model in which failure time has a probability distribution that depends on the covariates.

The specific feature that distinguishes survival analysis from classical statistical analysis is data censoring. Usually, failure time is unknown for some of the pavement sections. The only information available is that the section has survived up to a certain time. Therefore, the section is no longer followed up. This type of censoring is called right censoring. For right-censored data, the actual information of the ith pavement section, where $i = 1, \ldots, n$, is contained in the pair (t_i, d_i), where t_i = failure time and d_i = censoring indicator, taking the value 1 if the event has been observed (failed); otherwise, d_i takes the value 0 (censored). Then, the censoring indicator can be expressed in Eq. (1):

$$d_i = \begin{cases} 1 & \text{if } t_i \leq c_i \\ 0 & \text{if } t_i > c_i \end{cases} \tag{1}$$

where c_i = censoring time.

For a random time to failure, T, the probability density function of T is defined as $f(t)$ and the cumulative distribution function as $F(t) = P(T \leq t)$. Two other useful functions in this context are the survival function $S(t) = P(T > t) = 1 - F(t)$ and the hazard function $h(t) = f(t)/S(t)$, which can be interpreted as the instantaneous rate of failure given survival up until time t.

Proportional Hazards (PH) Model

The PH model is the most popular model for survival data analysis (Kleinbaum and Klein, 2005). In general. in the presence of covariates the PH model can be written as in Eq. (2):

$$h(t_i) = h_0(t_i)\exp(\mathbf{x}_i'\beta) \tag{2}$$

where $h_0(t)$ = baseline hazard function representing the deterioration rate of the pavement sections; β = parameter vector; and $\mathbf{x}_i$ = covariates vector of the ith observation. Pavement deterioration is caused by the combined effects of traffic loading and environmental factors on the structure and materials. Although addressing all of those factors in a deterioration model is impossible, the structure, material, and environment variables are used as $\mathbf{x}_i$ in the case study.

The baseline hazard function can be assumed to have a particular parametric form or can be left unspecified. A popular form for the baseline hazard function is the Weibull function, with which the PH model becomes

$$h(t_i) = pt_i^{p-1}\exp(\mathbf{x}_i'\beta), \qquad i = 1, \ldots, n \tag{3}$$

where p = shape parameter. In the PH model, the effect of the covariate has a multiplicative effect on the hazard rate.

PH Model with Frailty

When modeling fatigue cracking occurs in pavement sections, typically only a few covariates, such as pavement age, layer thickness, and environmental factors, are known. However, many other variables also influence the time to failure, for example, maintenance history and construction quality. Such factors are typically unknown, and thus cannot be explicitly included in the analysis.

If some unobserved covariates exist and they are ignored in the model, then the model estimates will be inconsistent. Vaupel et al. (1979) suggested a mathematical model for this problem by adding a random effect to Eq. (3) and assuming that the hazard function is a product of a frailty term, w_i, therefore describing the effect of the neglected covariates. Frailty models incorporate heterogeneity into the estimator by treating individual effects as random draws from a specific parametric distribution. Hence, the hazard function for a given pavement section with frailty is obtained by adding an unobservable multiplicative effect w_i on the hazard, so that conditional on the frailty is

$$h(t_i|w_i) = h_0(t_i)w_i\exp(\mathbf{x}_i'\beta), \qquad i = 1, \ldots, n \tag{4}$$

where w_i = random variable assumed to have a one-dimensional distribution.

It is common to assume that the frailty term w_i follows a gamma distribution of $\Gamma(\frac{1}{\theta}, \theta)$ with shape parameter of $1/\theta$ and scale parameter of θ such that the mean and the variance of w_i is 1 and θ, respectively (Ibrahim et al. 2001). In the above equation, the frailty component w_i represents the total effect on failure of the covariates not measured when collecting information on individuals. In cases in which the frailty is greater than 1, the pavement sections experienced an increased hazard of failure and are said to be more frail than others. The Weibull hazard function conditional on frailty can be expressed as

$$h(t_i|w_i) = pt^{p-1}w_i\exp(\mathbf{x}_i'\beta), \qquad i = 1, \ldots, n \tag{5}$$

Given the relationship between the hazard and survival functions, the individual survival function conditional on the frailty can be shown as

$$S(t_i|w_i) = \exp\left[-\int_0^t h(u|\mathbf{x}_i, w_i)du\right] = S(t_i)^{w_i}, \qquad i = 1, \ldots, n \tag{6}$$

where $S(t_i) = \exp[-t_i^p\exp(\mathbf{x}_i'\beta)]$ is the survival function of a PH model without considering the frailty. Therefore, the PH model with frailty is the generalization of the usual PH model by adding a frailty term. When the variance of the frailty, θ, equals 0, the frailty model becomes the PH model.

Bayesian Approach

The traditional maximum likelihood approach to estimation is commonly used in survival analysis estimation, but it can encounter difficulties with frailty models. Standard maximum likelihood-based inference methods may not be suitable for small sample sizes or situations in which there is heavy censoring (Ibrahim et al. 2001). However, Markov chain Monte Carlo (MCMC) sampling of Bayesian analysis makes it possible to make exact inferences without resorting to asymptotic calculations. Therefore, this paper employs the Bayesian approach to estimate the parameters β, θ, and p.

Bayesian analysis combines the information from observed data with prior knowledge about the parameters to obtain their posterior distributions. Let $\pi(\beta)$, $\pi(\theta)$, $\pi(p)$, and $P(\mathbf{w}|\theta)$ to be the probability density functions of the priors of parameters p, θ, β, and $\mathbf{w}$, and let $t = \{t_i\}$, $\mathbf{w} = \{w_i\}$. The joint posterior distribution of the parameters is proportional to the likelihood times the prior, as expressed in Eq. (7):

$$P(p,\theta,\beta,\mathbf{w}|t) \propto P(\mathbf{t}|p,\theta,\beta,\mathbf{w})\pi(p)\pi(\beta)P(\mathbf{w}|\theta)\pi(\theta) \tag{7}$$

The first term on the right-hand side of Eq. (7) is the likelihood of surviving up to time $\mathbf{t}$ given the parameters p, θ, β, and $\mathbf{w}$, which can be further expressed in Eq. (8):

$$\begin{aligned} P(\mathbf{t}|p,\theta,\beta,\mathbf{w}) &= \prod_{i=1}^{n} P(t_i|p,\theta,\beta,w_i) = \prod_{i=1}^{n} S(t_i|w_i)h(t_i|w_i)^{d_i} \\ &= \prod_{i=1}^{n} \exp[-t_i^p \exp(\mathbf{x}_i'\beta)][pt^{p-1}w_i \exp(\mathbf{x}_i'\beta)]^{d_i} \end{aligned} \tag{8}$$

The fourth term on the right-hand side of Eq. (7) is the joint conditional probability function of $\mathbf{w}$ given θ. Given that w_i is gamma distributed with parameter $\Gamma(1/\theta,\theta)$, it can be expressed as

$$P(\mathbf{w}|\theta) = \prod_{i=1}^{n} P(w_i|\theta) = \prod_{i=1}^{n} \left(\frac{1}{\theta}\right)^{\frac{1}{\theta}} w_i^{\frac{1}{\theta}-1} \frac{1}{\Gamma(\frac{1}{\theta})} e^{-\frac{w_i}{\theta}} \tag{9}$$

The conditional posterior probability of each of the variables (p, θ, β, $\mathbf{w}$) can be obtained by neglecting the other variables in Eq. (7). For example, the conditional posterior probability of $\mathbf{w}$ can be obtained by treating Eq. (7) as a function of $\mathbf{w}$ for a fixed value of (p, θ, β). Therefore, by discarding the constant terms $\pi(\beta)$, $\pi(p)$, and $\pi(\theta)$ in Eq. (7), the conditional posterior probability of $\mathbf{w}$ can be obtained as in Eq. (10):

$$p(\mathbf{w}|\mathbf{t},p,\theta,\beta) \propto P(\mathbf{t}|p,\theta,\beta,\mathbf{w})P(\mathbf{w}|\theta) \tag{10}$$

By plugging Eq. (8) and (9) into Eq. (10) and ignoring the terms that do not involve w_i, the conditional posterior of w_i is as follows:

$$P(w_i|\mathbf{t},p,\beta,\theta) \propto w_i^{d_i+\frac{1}{\theta}-1} \exp\left\{-w_i\left[\frac{1}{\theta} + t_i^p \exp(\mathbf{x}_i'\beta)\right]\right\} \tag{11}$$

This can be recognized as the core of a Gamma density function. As a result, we can conclude that the conditional posterior of w_i follows a Gamma distribution with shape and scale parameters of $d_i + \frac{1}{\theta}$ and $[\frac{1}{\theta} + t_i^p \exp(\mathbf{x}_i'\beta)]^{-1}$, respectively. Following the same procedure, we can also derive the conditional posterior of p, β, and θ. The results are as follows:

$$P(\theta|\mathbf{t},\mathbf{w},p,\beta) \propto \left(\prod_{i=1}^{n} w_i^{\frac{1}{\theta}-1}\right)\left(\frac{1}{\theta}\right)^{\frac{n}{\theta}} \Gamma\left(\frac{1}{\theta}\right)^{-n} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} w_i\right)\pi(\theta) \tag{12}$$

$$P(\beta|\mathbf{t},\mathbf{w},p,\theta) \propto \exp\left[\sum_{i=1}^{n} d_i\mathbf{x}_i'\beta - \sum_{i=1}^{n} w_i t_i^p \exp(\mathbf{x}_i'\beta)\right]\pi(\beta) \tag{13}$$

$$P(p|\mathbf{t},\mathbf{w},\beta,\theta) \propto \left(\prod_{i=1}^{n} t_i^{d_i}\right)^{p-1} p^D \exp\left[-\sum_{i=1}^{n} t_i^p \exp(\mathbf{x}_i'\beta)\right]\pi(p) \tag{14}$$

where $D = \sum_{i=1}^{n} d_i$.

Unlike Eq. (11), which can be identified as a Gamma distribution density function, the conditional posterior distributions on Eqs. (12)–(14) are intractable. They do not belong to any known probability distribution. Therefore, it is impossible to sample directly from Eqs. (12)–(14). Instead, we use the Metropolis-within-Gibbs algorithm to draw random sampling from the posterior distributions. The Metropolis-within-Gibbs algorithm can generate a Markov chain that has the posterior as its long-run distribution. Sampling from this Markov chain after an adequate burn-in period will enable us to approximate a sample from the posterior distribution (Albert, 2009). The general idea of the Metropolis-within-Gibbs algorithm is described as follows.

Metropolis-within-Gibbs Algorithm

Suppose that the parameter vector is $\theta = (\theta_1,\ldots,\theta_p)$. The joint posterior distribution of θ, which is denoted as $P(\theta|\text{data})$, may be of high dimension and difficult to summarize. Suppose we define the set of individual conditional posterior distributions as

$$\begin{aligned} &P(\theta_1|\theta_2,\ldots,\theta_p,\text{data}), \\ &P(\theta_2|\theta_1,\theta_3,\ldots,\theta_p,\text{data})\ldots P(\theta_p|\theta_1,\ldots,\theta_{p-1},\text{data}) \end{aligned} \tag{15}$$

where $[X|Y,Z]$ represents the distribution of X conditional on values of the random variables Y and Z. The idea behind Gibbs sampling is that random samples can be obtained from the joint posterior distribution by successfully simulating individual parameters from the set of p conditional distributions in Eq. (15). Draws from this simulation algorithm will converge to the target posterior distribution.

In situations in which sampling directly from the conditional distributions is inconvenient, a Metropolis algorithm with the random walk type can be used to simulate from each distribution. Let θ_i^t represent the current value of θ_i in the simulation and let $g(\theta_i)$ represent the individual conditional distribution. A candidate value of θ_i is given by

$$\theta_i^* = \theta_i^t + c_i Z \tag{16}$$

where Z = a standard normal variate and c_i = a fixed scale parameter. The next simulated value of θ_i, θ_i^{t+1}, will be equal to the candidate value with probability $P = \min\{1, g(\theta_i^*)/g(\theta_i^t)\}$ and equal to θ_i^t with probability $1 - P$. In summary, Eqs. (11)–(14) are used for the draw from posterior of w_i, β, θ, and p using the Metropolis-within-Gibbs sampling algorithm.

Case Study

In the following section, this study applies the proposed frailty model to the LTPP database and the estimated results are presented and discussed.

LTPP Dataset

For the following analyses, the failure criterion was selected according to the Mechanistic Empirical Pavement Design Guide (MEPDG) (ARA, 2004). For fatigue cracking, the failure time is defined when 25% of the total length of both inner and outer wheel paths display fatigue cracks. The LTPP database Standard Data Release 23.0 was used in this case study. The data are mainly collected from specific pavement sections (SPS) and general pavement studies (GPS). Two-hundred and ninety-four flexible pavement sections were identified containing the required information for a survival analysis for this case study. The sections are selected under the requirement that no maintenance activity was performed during the monitoring period.

Flexible pavement is usually composed of several layers, including surface course, base course, and subbase course. The properties of these structures have a direct impact on pavement performance. Environment (rain and snow) is also an important factor in the pavement deterioration process. The effect of environmental factors on pavement performance is usually reflected through a change in material properties attributable to their environment-sensitive characteristics. In this case study, several explanatory variables were selected to enter the survival analysis (see Table 1) based on previous research (Wang et al. 2005; Aguiar et al. 2009). By using the variables in Table 1, the frailty model in Eq. (4) becomes

$$\begin{aligned} h(t_i|w_i) = h_0(t_i)w_i \exp(&\beta_0 + \beta_1 \times \text{Thickness\_AC} + \beta_2 \times \text{AV} \\ &+ \beta_3 \times \text{Total\_Precip} + \beta_4 \times \text{Total\_Snow} \\ &+ \beta_5 \times \text{Days} > 89.6°\text{F} + \beta_6 \times \text{FI}), \\ i = 1, \ldots, n \end{aligned} \tag{17}$$

This study assumed very flat priors of normal distribution $N(0, 100)$ for β and gamma distribution $\Gamma(0.1, 10)$ for θ and p. In forming the priors, the more uncertainty the decision maker feels, the more the priors should be spread out. Therefore, choosing a flat prior means allowing the data to completely determine the posterior or the posterior is determined completely by the likelihood. In other words, the precise form of these priors had no appreciable impact on the posterior inferences (Kheiri et al. 2007).

Table 1. Explanatory variables

Category	Variable	Abbreviation	Unit
Structural	Total thickness of AC layers	Thickness AC	In
Material	Air void content	AV	%
Environmental	Average total annual precipitation	Total precipitation	In
	Average total annual snowfall	Total snow	In
	Average number of days above 89.6 °F	Days > 89.6 °F	#
	Average freezing index	FI	°F-days

Results

The proposed model was implemented by using the statistical software *R*, version 2.9.2 (http://www.r-project.org/). The following table shows the parameter estimation result of the survival analysis with the frailty term.

Table 2 shows that strong posterior evidence exists of a high degree of heterogeneity in the dataset. The mean value of θ is 0.5193 and the 95% credible interval is [0.0543, 1.2043], which means that the posterior probability that θ lies in the interval from 0.0543 to 1.2043 is 0.95. This probability indicates a significant effect of unobserved heterogeneity in the survival model, and the effect of heterogeneity should be considered in the estimation process. Failure to take this effect into consideration will result in inconsistent estimates.

The estimation results from Table 2 also show that the thickness of the asphalt concrete layers has a positive impact on the fatigue life of the pavement. As was originally expected, an increase in the asphalt layer thickness will defer the time of fatigue cracking failure. The air void content was also found to be significant in influencing the timing of the fatigue cracking failure. The model indicates that an increase in the air void content will accelerate the failure process, potentially attributable to the fact that a larger internal air void structure reduces the effective area of asphalt mixture used in resisting the microcracking and crack propagation process.

Additionally, the environmental factors were also found to be significant. Total precipitation, total snow, and freezing index have a negative impact on the pavement life in terms of fatigue cracking. Interestingly, an increase in days > 89.6 °F deg increases the pavement life in terms of fatigue cracking. This effect is also expected since the asphalt mixture has a more elastic behavior at higher temperatures. The estimation results can be better illustrated by Fig. 1, which shows the posterior probability distribution of the parameters. The posterior distribution is the conditional probability distribution of the parameters after the information is taken into account.

With the estimated parameters, the failure time of a pavement section in terms of fatigue cracking can be predicted. The conditional predictions (condition on the frailty) can be calculated using Eq. (6). The population prediction can be obtained by integrating out the frailty term. When w_i is distributed as gamma with mean 1 and variance θ, the population survival function can be shown to be

$$S_\theta(t_i) = \{1 - \theta \ln[S(t_i)]\}^{-1/\theta}, \qquad i = 1, \ldots, n \tag{18}$$

The population survival function in Eq. (18) represents the expected survival probability across the population considering the heterogeneity. To demonstrate the prediction of pavement fatigue

Table 2. Estimation Results of Frailty Model

Parameter	Mean	Median	Standard deviation	2.5 percent	97.5 percent
Intercept	−3.6223	−3.6395	1.7107	−6.6569	−0.2406
Thickness AC	−0.0899	−0.0894	0.0720	−0.2281	0.0476
AV	0.0390	0.0371	0.0639	−0.0853	0.1669
Total precipitation	0.0012	0.0012	0.0007	−0.0001	0.0025
Total snow	0.0001	0.0001	0.0005	−0.0008	0.0010
Days > 89.6 °F	−0.0142	−0.0141	0.0052	−0.0243	−0.0043
FI	0.0010	0.0010	0.0008	−0.0006	0.0024
p	1.4986	1.4837	0.2618	1.0247	2.0533
θ	0.5193	0.5087	0.3108	0.0543	1.2043

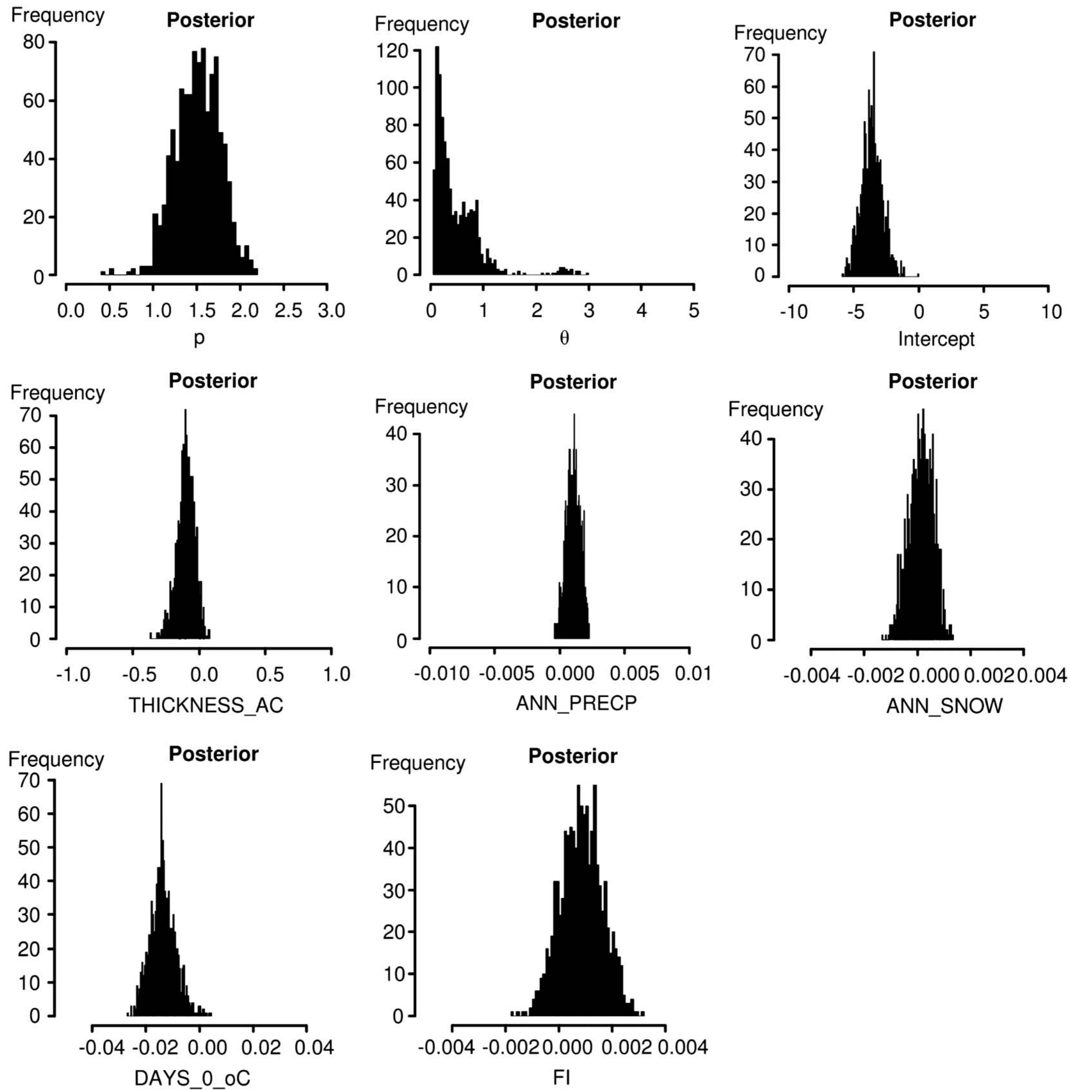

Fig. 1. Posterior distribution of parameters

life using the proposed model, we pick up a test pavement section with explanatory variables values around the average of the dataset (thickness AC $= 6.0$, AV $= 5.3$, total precip. $= 850$, total snow $= 769$, days $> 89.6\,°\mathrm{F} = 100$, FI $= 350$). Applying Eq. (18) results in the survival curve in Fig. 2. The predicted median value was 12 years, whereas the actual observed failure time was 13.44 years. This estimation result shows that the survival model can provide pavement agencies with general information on the most likely fatigue failure time.

Fig. 2 shows that the probability of failure attributable to fatigue cracking follows a three-stage process. During the initial stage, the probability of failure is low and slightly increasing up to approximately three years into the service life of the pavement structure. In the second stage, failures attributable to fatigue cracking occur at a more constant rate, which is expected since fatigue cracking can normally be observed after several years (five or more) of service life of the pavement structure. Finally, there is a third stage in which failure attributable to fatigue cracking appears to be reduced, possibly because pavement structures with design lives of more than 20 years tend to have more conservative designs and, therefore, their probabilities of failure are lower.

Fig. 3 shows the expected failure time of the pavement section used in Fig. 2 as a function of the frailty w. The horizontal line in Fig. 3 reflects the level of estimated failure time expectancy without considering frailty. The curved line shows the individual expected failure time as a function of the frailty. The difference between those two results indicates that if the frailty in the estimation is not considered, the failure time of sections with a frailty of $w = 1$ tends to be overestimated. This is even more serious for sections with a high frailty (e.g., $w > 2$). In summary, Fig. 3 shows that the expected failure time of average individual sections with a frailty of $w = 1$ is below the expected failure time of the population (using the PH model), and if the estimation does not consider the impact of heterogeneity, individual failure time tends to be overestimated.

Model Comparison

Several criteria are available for frailty model comparison, for example, deviance information criteria, Bayesian information criterion, and Bayes factor (Kheiri et al. 2007). This paper selected the Bayes factor to compare whether the frailty PH model fits the data better than the PH model. The Bayes factor is the posterior odds ratio between two models. For example, the odds ratio between model i and j can be expressed in Eq. (19):

$$BF_{ij} = \frac{P(y|M_i)}{P(y|M_j)} \tag{19}$$

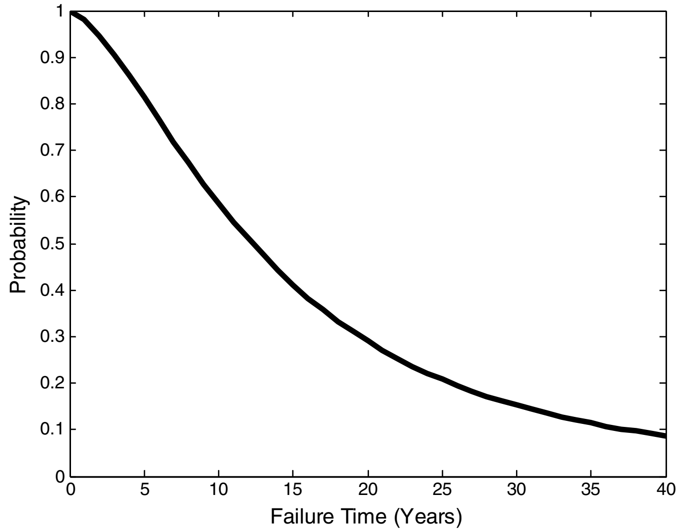

Fig. 2. Predicted failure time

where $P(y|M_i)$ = marginal likelihood of model i with M_i representing the ith model, which can be further expressed as

$$P(y|M_i) = \int P(y|\theta^i, M_i)P(\theta^i|M_i)d\theta^i \tag{20}$$

where θ^i represents the parameters of model i. However, with the likelihood function in Eq. (8), analytically working out the integration in Eq. (20)) is impossible. As an alternative, this paper proposes using the Savage-Dickey Density ratio (Koop, 2003) to calculate the Bayes factor. The general idea of the Savage-Dickey Density is described as follows.

Suppose M_2 is the unrestricted version of a model with a parameter vector $\theta(\omega, \psi)$. The likelihood and prior for this model are given by $P(y|\omega, \psi, M_2)$ and $P(\omega, \psi|M_2)$. The restricted version of the model, denoted by M_1, has $\omega = \omega_0$ where ω_0 is a vector of constants. The parameters in ψ are unrestricted in each model. The likelihood and prior for M_1 are given by $P(y|\psi, M_1)$ and $P(\psi|M_1)$. The Savage-Dickey density ratio states that the Bayes

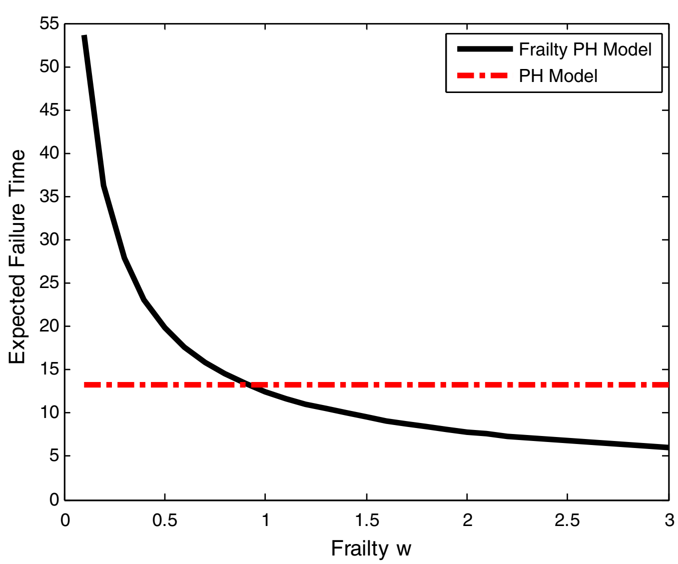

Fig. 3. Expected failure time

factor in Eq. (19) comparing M_1 with M_2 can be obtained by Eq. (21) instead of calculating Eq. (20)

$$BF_{12} = \frac{P(\omega = \omega_0|y, M_2)}{P(\omega = \omega_0|M_2)} \tag{21}$$

Let M_1 represent the PH model with the restriction that the frailty term $\mathbf{w} = 1$ and M_2 represent the unrestricted frailty model, in which w_i follows a one-dimensional distribution. Then, by using the Savage-Dickey density ratio, the Bayes factor, BF_{12}, comparing M_1 with M_2 has the form

$$BF_{12} = \frac{P(\mathbf{w} = 1|t, M_2)}{P(\mathbf{w} = 1|M_2)} \tag{22}$$

The numerator on the right-hand side of Eq. (22) is the joint probability that the frailties equal one under the posterior distribution. The denominator is the joint probability that the frailties equal one under the prior distribution. Since the prior and posterior distribution of $\mathbf{w}$ all follow a Gamma distribution, the joint probabilities in Eq. (22) can be obtained analytically. The study estimated the Bayes factor BF_{12} to be 0.117 for the given dataset, which is the odds ratio between the PH model and the frailty model. This value indicates that the frailty model fit the data better than the PH model and proves the necessity of taking heterogeneity into consideration when modeling pavement fatigue cracking failure using the survival analysis.

Conclusion

This paper proposes modeling the fatigue cracking of flexible pavement using the survival model. A frailty term is added to the proportional hazard function to identify and quantify the effect of unobserved heterogeneity. For the parameter estimation, this study adopted the Bayesian approach using a MCMC algorithm and then applied the model to an LTPP dataset.

The results indicate that the effect of unobserved heterogeneity is significant in the analysis and should be taken into consideration when estimating time to failure. The results also indicate that asphalt layer thickness, air void content, and environmental factors have a significant effect on the probability that a pavement performs appropriately for a given number of years. The approach presented here is believed to be valuable at a state or even at a district level because the model is able to capture the factors that cause pavement structures to fail at a faster rate. The model can also provide general information on the most likely fatigue failure time.

References

Aguiar, J. P., Prozzi, J. A., and Banerjee, A. (2009). “Survival analysis of flexible pavements based on LTPP data.” *CD Proc. of the 89th Transportation Research Board Annual Meeting*, Transportation Research Board, Washington, DC.

Albert, J. (2009). *Bayesian computation with R*, Springer, New York.

ARA, Inc. ERES Consultants Div. (2004). “Guide for mechanistic-empirical 19 design of new and rehabilitated pavement structures.” *NCHRP Research Rep. 1-20 37A*, Transportation Research Board, National Research Council, Washington, DC.

Hong, F. (2007). “Modeling heterogeneity in transportation infrastructure deterioration application to pavement.” Ph.D. thesis, Univ. of Texas at Austin, Austin, TX.

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001). *Bayesian survival analysis*, Springer, New York.

Ker, H. W., Lee, Y. H., and Wu, P. H. (2008), “Development of fatigue cracking prediction models using long-term pavement performance database.” *J. Transp. Eng.*, 134(11), 477–482.

Kheiri, S., Kimber, A., and Meshkani, M. R. (2007). “Bayesian analysis of an inverse gaussian correlated frailty model.” *Comput. Stat. Data Anal.*, 51(11), 5317–5326.

Kleinbaum, D. G., and Klein, M. (2005). *Survival analysis: A self-learning text*, Springer, New York.

Koop, G. (2003). *Bayesian econometrics*, Wiley, London.

Nakat, Z. S., and Madanat, S. M. (2008), “Stochastic duration modeling of pavement overlay crack initiation.” *J. Infrastruct. Syst.*, 14(3), 185–192.

Paterson, W. D. O. (1987). *Road deterioration and maintenance effects models for planning and management*, John Hopkins University Press, Baltimore.

Quieroz, C. A. V. (1981). “Performance prediction models for pavement management in Brazil.” Ph.D. thesis, Univ. of Texas at Austin, Austin, TX.

Rahim, A. M., Fiegel, G., and Ghuzlan, K. (2009), “Evaluation of fatigue (alligator) cracking in the LTPP SPS-6 experiment.” *Int. J. Pavement Res. Technol.*, 2(1), 26–32.

Shin, H. C., and Madanat, S. (2003), “Development of a stochastic model of pavement distress initiation.” *J. Infrastruct. Plann. Manage.*, 744, 61–67.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). “The impact of heterogeneity in individual frailty on the dynamics of mortality.” *Demography*, 16(3), 439–454.

Wang, Y. H., Mahboub, K. C., and Hancher, D. E. (2005), “Survival analysis of fatigue cracking for flexible pavements based on long-term pavement performance data.” *J. Transp. Eng.*, 131(8), 608–616.