Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666
Vol. 5, Issue 1 (2012): 17-32

BAYESIAN ANALYSIS OF UNEMPLOYMENT DURATION DATA IN THE PRESENCE OF RIGHT AND INTERVAL CENSORING

M. Ganjali[1] **, T. Baghfalaki**[2]
Department of Statistics, Shahid Beheshti University, G. C., Tehran, Iran.
E Mail: [1]m-ganjali@sbu.ac.ir ; [2]t.baghfalaki@yahoo.com
(Received April 07, 2012)

Abstract

In this paper, Bayesian inference for unemployment duration data in the presence of right and interval censoring, where the proportionality assumption does not hold, is discussed. In order to model these kinds of duration data with some explanatory variables, Bayesian log-logistic, log-normal and Weibull accelerated failure time (AFT) models are used. In these models, sampling from the joint posterior distribution of the unknown quantities of interest are obtained through the use of Markov chain Monte Carlo (MCMC) methods using the available WinBUGS software. These models are also applied for unemployment duration data of Iran in 2009. The models are compared using deviance information criterion (DIC). Two new sensitivity analyses are also performed to detect: (1) the modification of the parameter estimates with respect to the alteration of generalized variance of the multivariate prior distribution of regression coefficients, and (2) the change of the posterior estimates with respect to the deletion of individuals with high censoring values using Kullback-Leibler divergence measure.

Keywords: Bayesian Analysis; Interval Censoring; Kaplan-Meier Method; Kullback-Leibler Measure; MCMC; Sensitivity Analysis; Unemployment Duration; WinBUGS.

1. Introduction

The provenance of using survival analysis and duration models is in medical research. Survival analysis is a collection of methods that processes the variable of time to the occurrence of an event. In this context death or failure is considered as an event of interest. Although at the beginning the survival analysis was used to study death as an event specific to medical studies (Armitage, 1959; Lee, 1980; Elandt-Johnson and Johnson, 1980; Kalbfleisch and Prentice, 1980), nowadays these methods have evolved to special applications in several other fields. Analysis of time intervals between successive child births in demography, studies of recidivism and duration of marriage in sociology and analysis of spells of unemployment duration in labor economics are some applications of survival analyses in various fields of sciences. In this paper, special consideration is given to the study of duration data in economics.

Survival analysis, adapted in conventional econometric modeling data, received the title of duration models (Kiefer, 1988; Moffitt, 1999). In this context unemployment duration refers to the amount of time that an individual remains unemployed. Transition from unemployment into work is an important econometric problem which evaluates policies and recommends ways to facilitate these transitions, for examples see Card and Sullivan (1988), Meyer (1990), Ackum (1991), Torp (1994), Winkelmann (1997) and Nilsen et al. (2000).

Non-Bayesian parametric and semiparametric approaches for analyzing duration data are well discussed in the literature. For examples, Moffitt (1999), Kupets (2006), Berg et al. (2008), Gonzalo and Saarela (2000), Borsic and Kavkler (2009) are some references.

References that have discussed Bayesian method for survival analysis are: Arjas and Bhattacharjee (2004), Arjas and Gasbarra (1994), Campodñico and Singpurwalla (1994), Kalbfleisch (1978), Padgett and Wei (1981) and Ibrahim et al. (2005). In economic applications specially when proportionality assumption is valid in a data set, Ruggiero (1994) proposes a fully Bayesian estimator for the regression parameters, Kalbfleisch and Prentice (1980) and Campolieti (2001) describe a Semiparametric Bayesian analyses, also Paserman (2004) used semiparametric Bayesian method for duration data with unobserved and unknown heterogeneity.

The accelerated failure time (AFT) model is widely accepted as an alternative approach when the proportional hazard (PH) assumption does not hold. However, there are few studies using Bayesian analysis in the AFT model considering influential individuals. Hanson and Johnson (2004) used a Bayesian semiparametric AFT model for interval-censored data. Kudus et al. (2006) presented a Weibull model with an especial structure for modeling interval censored survival times of Acacia Mangium plantation in a spacing trial. Zhang and Lawson (2011) used Bayesian parametric AFT spatial model with a medical application. The use of full Bayesian methodology for studying unemployment duration data, particularly in parametric models with right and interval censoring, considering sensitivity of results to deletion of inflectional individual or change of prior parameters is rare. This paper, therefore, seeks to examine Bayesian method and some sensitivity analyses in this direction.

We analyze unemployment duration in Iran. The data for our empirical investigation were obtained from Statistical Center of Iran in 2009 and consist of right-censoring and interval-censoring individuals. We model, using the Bayesian approach, some explanatory variables, such as gender, age, residence place, current marital status, education status and the number of household members. These covariates may produce variations on unemployment duration. Also we assume some different distribution assumptions for logarithm of duration time and compare the results of using them by deviance information criterion (DIC) (Spiegelhalter et al. (2002)). The results of Bayesian implementation in this paper are obtained using the available software WinBUGS (Spiegelhalter et al., 2003). By an iterative approach, we shall present the sensitivity of results with respect to the change of generalized variance matrix of the multivariate prior. We shall also investigate the effect of the deletion of some individuals on inferential results.

2. The Data Set

The data set that will be used in this paper is extracted from a follow up study conducted by Statistical Center of Iran. In these data the labor force status of people in two seasons of spring and summer in 2009 are recorded. We have selected the individuals who are observed on both seasons and are unemployed in spring (in spring unemployed individuals answer a question about their duration of unemployment). The data contain detail individual information for a random sample of age 14 and older

population. The vector of covariates includes personal characteristics such as gender, age, the place of residence, current marital status, education status and the number of household members. Details of categories of covariates and their percentages are described in Table 1. Table 2 gives frequencies and percentage of different categories of unemployment status in summer 2009 of unemployed individuals in spring 2009. This table shows that from 1337 individuals in the study, 743 individuals stayed in unemployment status in summer and $473+121$ individuals have moved from unemployment to employment status. Unfortunately, only for 121 individuals the exact duration of unemployment is recorded. For other 473 remaining individuals we only know that their movement to employment happened during a 3-month period. Employment duration of these individuals can be considered as interval censoring. Figure 1 presents the survival curves for unemployment duration of the above described data set. Points on this curve estimate the proportion of individuals who will be in unemployment status at least at a given period of time.

Explanatory variable	**Categories**	Percentage
Current marital status	married	0.294
	widow or divorced	0.007
	single	0.699
Gender	female	0.225
	Male	0.775
Age	< 20	0.126
	$21-25$	0.378
	$26-30$	0.249
	$30>$	0.247
Education status	under diploma	0.464
	diploma	0.307
	associate of arts (or science)	0.081
	MA and upper	0.148
Number of household members	one or two	0.082
	three	0.883
	four and more	0.034
Residence	rural	0.277
	urban	0.723

Table 1: Different levels of the chosen explanatory variables along with their percentages.

	Frequency	Percent
Right-censored data [1]	743	55.572
Completed observation [2]	121	9.050
Interval--censored data [3]	473	35.378

Table 2: Employment status of unemployment individuals of the spring of 2009 in the summer of 2009. [1]: Still unemployed in summer, [2]: Duration is recorded, [3]: Duration is recorded in an interval.

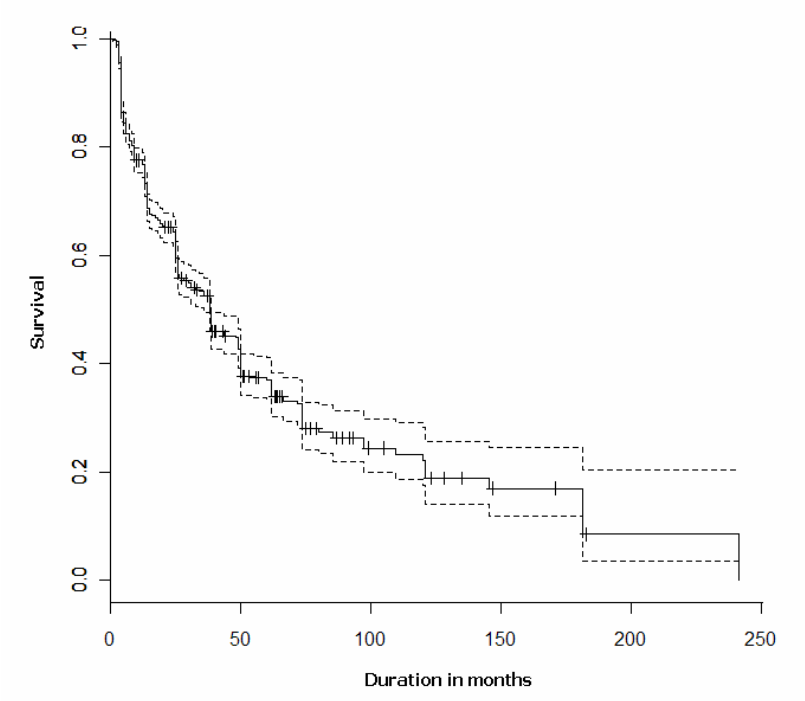

Figure 1: Survival curve of unemployment duration along with its 95% confidence bands.

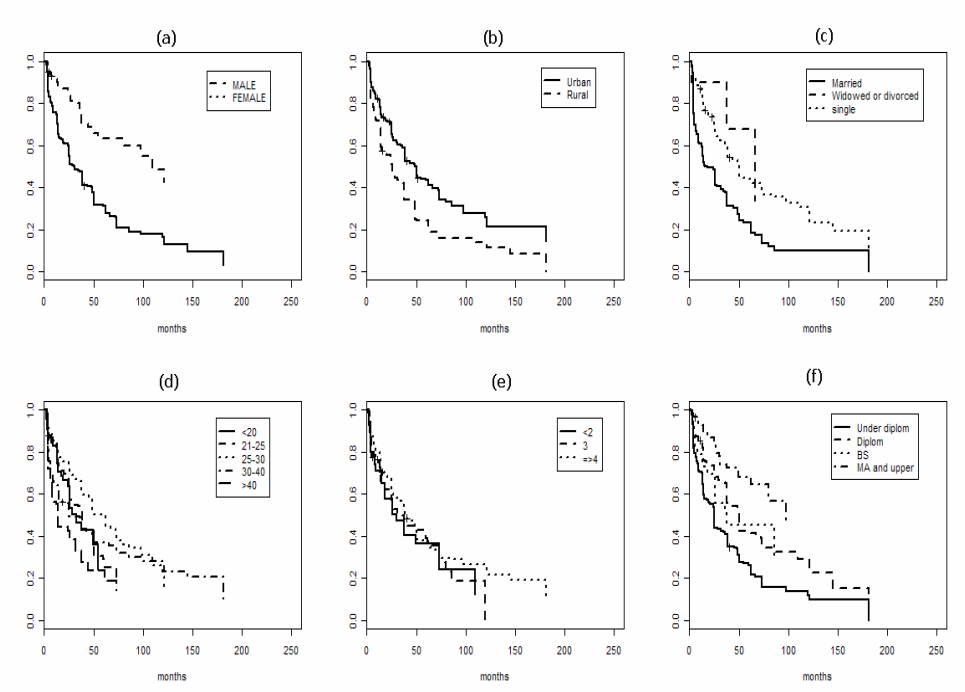

Figure 2: Kaplan-Meier estimators of the survival curves of unemployment duration on covariates groups, (a): Gender, (b): Place of residence, (c): Current marital status, (d): Age group, (e): Number of household members, (f): Educational level.

For a primary description of the explanatory variables in the data set, Figure 2 shows Kaplan-Meier estimators of the survival curves of unemployment duration for different covariates groups. Description of this figure is simple, for example according to figure 2(a) females have longer unemployment duration than that of males.

3. Models for Duration Data

In the unemployment duration literature, analyzing the relationship between unemployment duration and one or more explanatory variables is of most interest. Cox proportional hazards model (Cox, 1972) is a broadly applicable and the most widely used method in duration modelling with explanatory variables. The Cox proportional hazards model is given by:

$$h_i(t) = h_0(t)\exp(x_i'\beta)$$

where β denotes a $p\times 1$ vector of unknown regression parameters, x_i is a $p\times 1$ vector of explanatory variables and $h_0(t)$ is the baseline hazard function. This model is semiparametric because while the baseline hazard can take any form, the explanatory variables enter the model linearly in the exponential scale. The Cox model has had tremendous success in applied work, because of the availability of software to perform

estimation and inference in the model. It is clear that one can not use this method any time and the first stage in using a statistical method is checking the validity of it. There are some famous tests which check the validity of the proportional hazards assumption (vide for example, Deshpande and Purohit, 2005). If the proportionality assumption is not valid, the Cox proportional hazrd models cannot be used in modelling, the main purpose of this paper is the use of Bayesian modelling approach.

In a parametric model, the distribution of outcomes (duration of unemployments) is specified in term of a finite number of unknown parameters. One of the famous parametric models is accelerated failure time (AFT) model in which duration is assumed to be a function of explanatory variables. Let T_i, $i:1,2,...,n$ be a duration time and x_i is a $p\times 1$ vector of explanatory variables. The AFT model assumes that the relationship of logarithm of T and x is linear and can be written as

$$\log(T_i) = x_{i'}\beta + \varepsilon_i,\ i:1,2,...,n \qquad (1)$$

where β is a $p\times 1$ vector of regression parameters and ε_i is a residual term with a specified distribution, let $\varepsilon \sim F_\varepsilon(.|\sigma)$, such that $F_\varepsilon(.|\sigma)$ is the known cdf associated with density $f_\varepsilon(.|\sigma)$ with scale parameter σ. The survival and hazard function of ε are $S_\varepsilon(.|\sigma) = 1 - F_\varepsilon(.|\sigma)$ and $h_\varepsilon = f_\varepsilon / S_\varepsilon$, respectively. The commonly used distributions for ε are the extreme value, logistic and normal distributions. These three distributions are, respectively, log-transformation of Weibull, the log-logistic and the log-normal distributions. These distributions are appropriate distributions for analyzing unemployment duration (Greene, 2003).

Suppose we observe n independent vectors of (T_i, δ_i, x_i), where T_i is the time to the event and δ_i is the indicator telling us whether T_i is un-censored or censored. The values $\delta_i = -1$, $\delta_i = 0$ and $\delta_i = 1$ indicate an interval censored observation, a completely known observation and a right censored observation, respectively. The x_i is a $p\times 1$ vector of explanatory variables. Let θ denote the set of unknown parameters in the model (1). The above elements have used to build model (1) and will be denoted by $T_i \overset{iid}{\sim} AFT(F_\varepsilon, \theta | x_i)$. In the first stage we need the survival function for an individual, let the precision parameter be denoted by $\tau = 1/\sqrt{\sigma}$, then (Christensen et al., 2011; page 325),

$$S(t_i|x_i,\theta) = 1 - F_\varepsilon[(\log(t_i) - x_{i'}\beta)\sqrt{\tau}],$$

$$f(t_i|x_i,\theta) = \frac{\sqrt{\tau}}{t} f_\varepsilon[(\log(t_i) - x_{i'}\beta)\sqrt{\tau}],$$

$$h(t_i|x_i,\theta) = \frac{\sqrt{\tau}}{t} h_\varepsilon[(\log(t_i) - x_{i'}\beta)\sqrt{\tau}].$$

The log-likelihood function of the set of unknown parameters, θ, in the presence of right and interval censoring (for our data set) can be written as:

$$\ell(\theta|t,x) = \sum_{i=1}^{n}\Big(\log(f(t_i|\theta,x_i))\times I_{\{\delta_i=0\}} + \log(S(t_i|\theta,x_i))\times I_{\{\delta_i=1\}} + \log(S(t_i|\theta,x_i) - S(t_i+3|\theta,x_i))\times I_{\{\delta_i=-1\}}\Big),$$

where $f(t_i|\theta,x_i)$ and $S(t|\theta,x_i)$ are the density and survival distributions, respectively. The Bayesian AFT model (Ntzoufras, 2009; Ibrahim et al., 2005; Christensen et al., 2011) for log-logistic and log-normal models can be obtained by assumption $\mu_i = x_{i'}\beta$. In these models, when both of β and σ are unknown, no joint conjugate prior is available. A typical joint prior specification can be expressed as a product of a multivariate normal (for parameter $\beta|\sigma^2$) and an inverse gamma prior (for σ^2), that is $\beta|\sigma^2 \sim N_p(\mu_0, \sigma^2 V_0), \quad \sigma^2 \sim IG(a,b).$

A direct way to state the Weibull AFT model is to let

$$T_i \sim \text{Weibull}(\lambda_i,\gamma) \quad and \quad \log(\lambda_i) = -x_i'\beta,$$

and a joint prior specification is to take $\beta \sim N_p(\mu_0, V_0)$ and $\gamma \sim IG(a,b)$.

The posterior distribution for the model specification above does not have closed form solutions for the parameters. To conduct the Bayesian analysis, Markov chain Monte Carlo (MCMC) techniques can be used to sample the joint posterior distribution of these models. One special MCMC type approach, which requires only the specification of the conditional posterior distribution for each parameter, is the Gibbs sampler. In situations where those distributions are simple to sample from the approach is easily implemented. In other situations, as in our situation, the more complex Metropolis-Hastings approach needs to be considered.

Combining the likelihood function (2) with the prior distributions on (β,σ^2) in the above models, the full conditional distributions for unknown parameters in log-logistic and log-normal models are given by:

$$\pi(\beta|\sigma^2,t,x) \propto \prod_{i=1}^{n}\Big(f(t_i\,|\,x_i,\beta,\sigma^2)^{I(\delta_i=0)}\times S(t_i\,|\,x_i,\beta,\sigma^2)^{I(\delta_i=1)}\times \times (S(t_i\,|\,x_i,\beta,\sigma^2) - S(t_i+3\,|\,x_i,\beta,\sigma^2))^{I(\delta_i=-1)}\Big)\times \pi(\beta|\sigma^2)$$

$$\pi(\sigma^2|\beta,t,x) \propto \prod_{i=1}^{n}\Big(f(t_i\,|\,x_i,\beta,\sigma^2)^{I(\delta_i=0)}\times S(t_i\,|\,x_i,\beta,\sigma^2)^{I(\delta_i=1)}\times \times (S(t_i\,|\,x_i,\beta,\sigma^2) - S(t_i+3\,|\,x_i,\beta,\sigma^2))^{I(\delta_i=-1)}\Big)\times \pi(\beta|\sigma^2)\times\pi(\sigma^2),$$

where $\pi(\beta|\sigma^2)$ and $\pi(\sigma^2)$ are prior distributions for β and σ^2 in the log-logistic and log-normal model. Also, the full conditional distributions for Weibull model are given in the same notations, where σ^2 replaced by γ and $\pi(\beta)$ and

$\pi(\gamma)$ are independent prior distributions. For these models, the Gibbs sampler can be implemented using the WinBUGS software (Spiegelhalter et al., 2003).

4. Selection of Theoretical Distribution Based on the Probability Plotting

The probability plot is one of the most applicable method for checking distributional assumption. In this paper, we have used Lee and Wang (2003) method with some adjusment based on the Kaplan-Meier method.

As mentioned by Lee and Wang (2003), if the theoretical distribution is adequate for the data, a graph of $log(t)$ versus a function of the sample cumulative distribution function will be close to a straight line. In another word, a fitted linear regression for $log(t)$ and the function of cumulative distribution function is a good index for selection of a theoretical distribution.

The fitted regression lines for log-normal, log-logistic and Weibull distributions are given by:

$$\log t_i = \frac{1}{\gamma}\log\frac{1}{\lambda} + \frac{1}{\gamma}\log(\log(\frac{1}{1-F(t_i)})),$$
$$\log t_i = \mu + \sigma\Phi^{-1}(F(t_i)),$$
$$\log t_i = \mu + \sigma\log(\frac{1}{1-F(t_i)} - 1).$$

Thus, a quick goodness of fit test is a regression line of $log(t_i)$ versus a function of $\hat{F}(t_i)$, where $\hat{F}(t_i)$ is an estimate of $F(t_i)$. This method can be summarized in the following steps:

• Select a theoretical distribution for the survival time T.

• Estimat the cumulative distribution function. There are some methods for this: the most famous is Kaplan-Meier method, another method which is used by Lee and Wang (2003) is the use of $(i-0.5)/n$ for the i^{th} ordered t values, $i:1,2,...,n$. In this method right censored observations are considered only in sorting the index(i). For interval censored observation a midpoint imputation may be used.

• Fit a linear regression line for $log(T)$ and the function of cumulative distribution function.

The best selected theoretical distribution would be obtained by finding the best fitted linear regression.

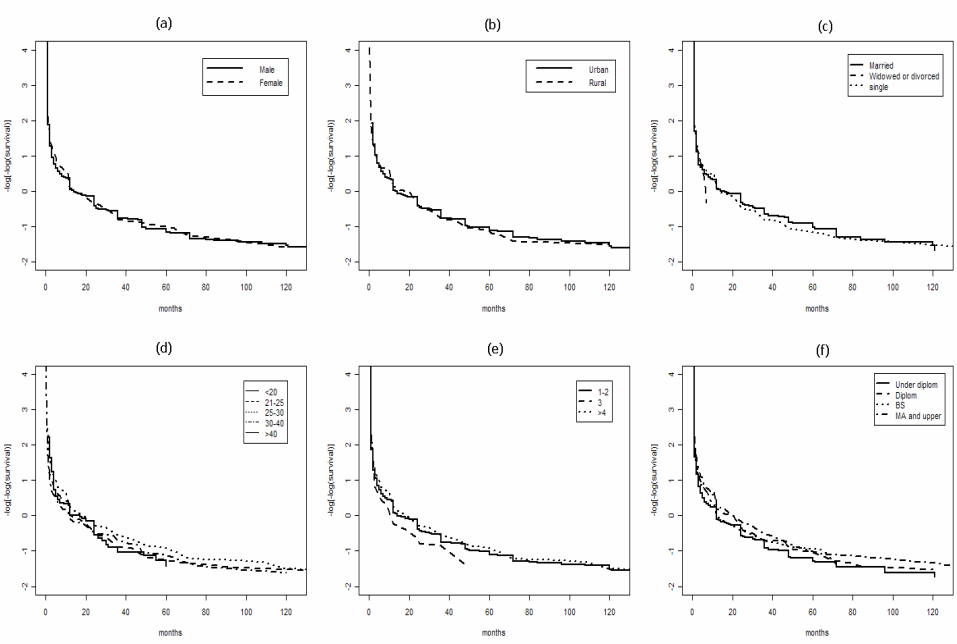

Figure 3: Graphical test for proportional hazards, Kaplan-Meier log-log plots of unemployment duration on covariates groups, (a) Gender, (b) Place of residence, (c) Current marital status, (d) Age group, (e) No. of household members, (f) Educational level.

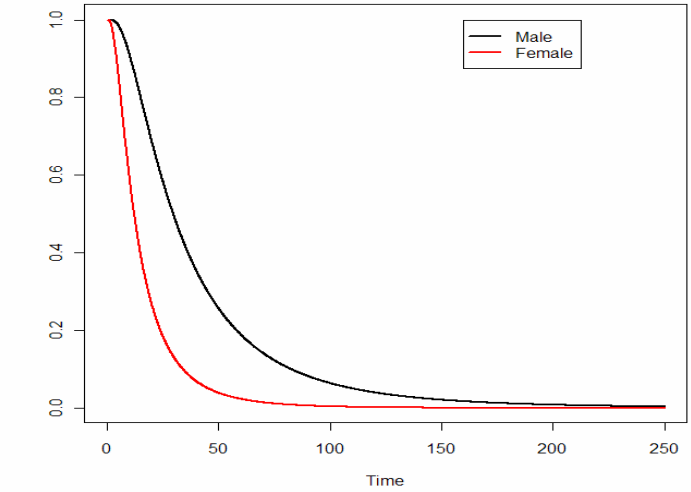

Figure 4: $S(t\,|\theta,x)$ **for an under diploma married twenty-nine old person with more than three members of family in urban.**

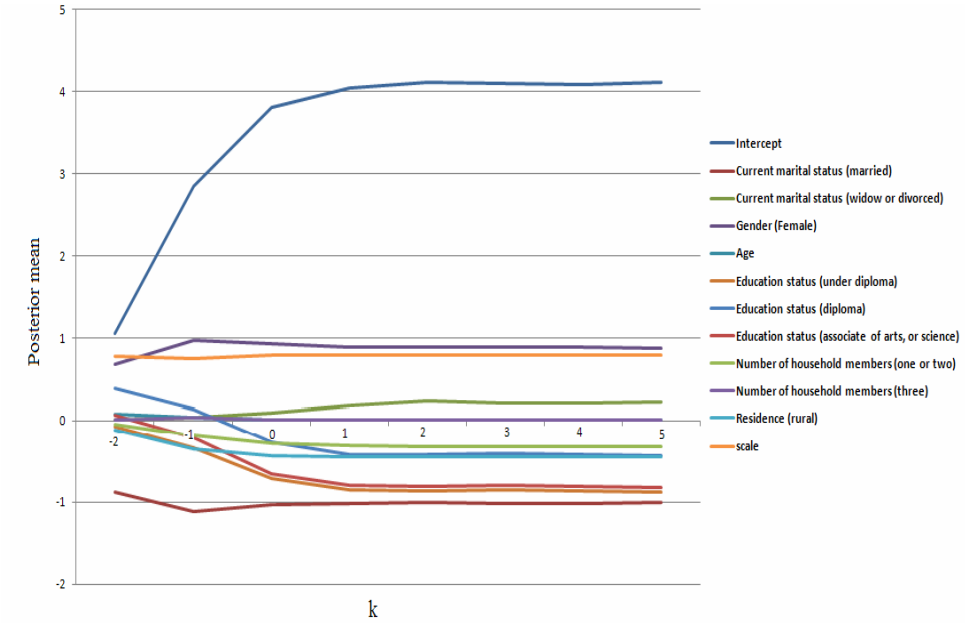

Figure 5: Sensitivity plot of posterior mean of different parameters for different values of k in log-logistic AFT model.

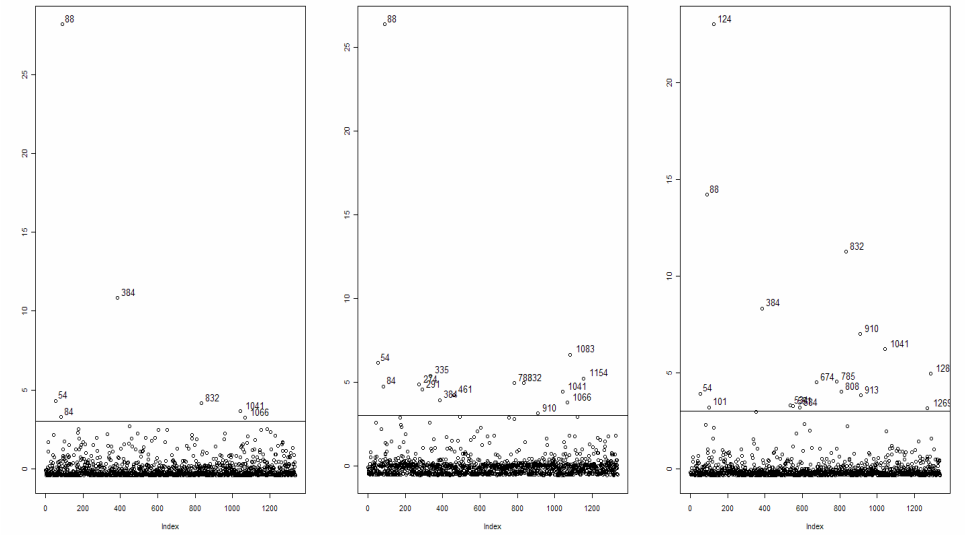

Figure 6: Standardized Kullback-Leibler divergence under log-logistic AFT model (left panel), log-normal AFT model (middle panel) and Weibull AFT model (right panel).

5 Application

5.1 Model for Unemployment Duration in Iran

In this section, we analyze the data set described in Section 2 based on the proposed methods. The explanatory variables in this data set were described in table 1. In this section, we consider age as a continuous variable. We use some dummy variables (if a categorical variable has g categories then $g-1$ dummy variables need to be created for the corresponding terms in a model and consequently $g-1$ regression coefficients have to be estimated) to describe the variable levels. For example, for current marital status dummy variables are as follows:

$$\text{mar1}=\begin{cases}1 & \text{married}\\ 0 & \text{o.w.}\end{cases}, \qquad \text{mar2}=\begin{cases}1 & \text{widow or divorced}\\ 0 & \text{o.w.}\end{cases},$$

and $\text{mar1}=\text{mar2}=0$ defines single status which will be taken as the baseline category.

The Cox proportional hazard model is the most general of the regression models because it is not based on any assumptions concerning the nature or shape of the underlying survival distribution, therefore one may be interested to use this methodology. However, one may use graphical test for checking the validity of the proportionality assumption (Deshpande and Purohit, 2005; page 189). Figure 3 shows $\log-\log$ plots of unemployment duration for different levels of covariates. In some panels of this figure (for example panel c) log-log of survival function are not parallel for different levels of covariates, therefore proportional hazards assumption is not valid for our data set.
We consider the following AFT model for unemployment duration data:

$$\begin{aligned}\log(T_i) = {} & \beta_0+\text{mar1}_i\beta_1+\text{mar2}_i\beta_2+\text{sex}_i\beta_3+\text{age}_i\beta_4+\text{edu1}_i\beta_5+\text{edu2}_i\beta_6\\ & +\text{edu3}_i\beta_7+\text{num1}_i\beta_8+\text{num2}_i\beta_9+\text{res}_i\beta_{10}+\varepsilon_i, \quad i:1,...,n.\end{aligned}$$

where ε_i has mean 0. For theoretical distribution of t_i, we use the methodology of section 3. R-squared of fitted lines for log-normal, log-logistic and Weibull model are presented in table 3. This table shows that these three distributions are adequate for analysing these data set.

Therefore, we assume three distributional assumption for ε_i, logistic, normal and extreme value distributions. As described in Section 3, they lead to log-logistic, log-normal and Weibull AFT model, respectively. For prior distributions in the first two models, we have

$$\beta|\sigma^2 \sim N_p(\mu_0,\sigma^2V_0), \quad \sigma^2 \sim IG(a,b).$$

where, we assume $\mu_0=0$ and $V_0=10^4 I$. Such distributions are low informative. Also, for a and b we consider small values to holding low informative prior

assumption. In Weibull AFT model, we assume $\beta \sim N_p(0,10^4 I)$ and $\gamma \sim IG(a,b)$, with small values of a and b.

In our analysis, we ran two parallel MCMC chain for 50000 iterations, then, we discarded the first 20000 iterations as burn-in and retained 30000 for the posterior analysis. We have checked the convergence of the chains using Gelman-Rubin convergence diagnosis and other available criteria in BOA package.

Table 4 shows the obtained results which are posterior mean and standard deviation from the three models introduced in section 3, one idea is that model with smaller DIC should be preferred to model with larger DIC, therefore this table shows that the log-logistic model has the best fit between these models.

This table shows that gender, current marital status, education level and living area are effective factors on unemployment duration of Iran, such that married persons have shorter unemployment duration than singles and widow or divorced people, also unemployment duration for females is longer than that for males. Also in Iran the under diploma people have the shortest unemployment duration. Unemployment duration for families with one or two members is shorter than that for families with larger members, and at last rural people have shorter unemployment duration than urbans.

Figure 4 shows predicted log-logistic AFT survival function, $S(t|\theta,x)$, for an under diploma married twenty-nine old person with more than three members of family. This plot shows that for people with these characteristics, unemployment duration for males is shorter than that for females. A residual analysis shows no important paths in the graphs.

	Multiple R-squared	
	$(i-0.5)/n$	Kaplan-Meier
log-normal model	0.977	0.914
log-logistic model	0.958	0.965
Weibull model	0.915	0.935

Table 3: R-squared for fitting described regressions in (3).

\

	Log-Normal Model		Log-Logistic Model		Weibull Model	
Parameters	Est.	S.E.	Est.	S.E.	Est.	S.E.
Intercept	4.060	0.245	4.109	0.260	4.256	0.247
Gender						
Female	0.824	0.133	0.882	0.142	0.848	0.142
Baseline (male)	-	-	-	-	-	-
Current marital status						
Married	-0.966	0.121	-1.010	0.122	-0.801	0.100

Widow or divorced	0.211	0.589	0.614	0.194	0.179	0.645
Baseline (single)	-	-	-	-	-	-
Education level						
Under diploma	-0.827	0.171	-0.875	0.182	-0.808	0.176
Diploma	-0.391	0.174	-0.428	0.185	-0.389	0.182
BS	-0.768	0.216	-0.817	0.228	-0.733	0.220
Baseline(MA and higher)	-	-	-	-	-	-
Age	0.010	0.007	0.010	0.007	0.014	0.006
Number of household members						
One or two	-0.267	0.199	-0.319	0.207	-0.225	0.178
Three	0.032	0.129	0.007	0.130	-0.076	0.116
Baseline (four and more)	-	-	-	-	-	-
Living area						
Rural	-0.406	0.099	-0.444	0.099	-0.391	0.089
Baseline (urban)	-	-	-	-	-	-
Scale	1.838	0.113	0.795	0.026	0.961	0.031
DIC	4548.860		3999.700		4599.944	

Table 4: Bayesian parameters estimates for AFT models

5.2 Sensitivity Analysis

Sensitivity analysis is defined as a measure of the change of a given input on a given output, in other words, the purpose is to find the sensitivity of the results with respect to inputs modification. Sensitivity analysis is an important part of any applied Bayesian work. In this paper, we want to check two types of sensitivity analyses, the sensitivity of a posterior distribution with respect to changes in the prior parameters and with respect to deletion of some individuals in the data set. In previous subsection, we found that log-logistic AFT model is the best fitted model, therefore the main consideration of this section is on sensitivity analysis on log-logistic AFT model.

Firstly, the sensitivity of the posterior mean over different values of the prior parameter k, which controls the generalized prior variance $|V_0|$ (constant part of the determinant of the prior covariance matrix) is investigated. We consider $|V_0| = 10^{p \times k}$ where p is the dimension of matrix V_0 (here is 11) and values of $k = -2,-1,0,1,2,3,4,5$. Figure 5 is a graphical display of the sensitivity of the posterior mean of all parameters about k with respect to the changes in $|V_0|$. This figures shows that robustness of parameter estimates with respect to variation of k is obtained after $k = 1$.

In checking sensitivity analysis with respect to omission of individuals, instead of focusing on inference based on $\pi(\theta|t)$, posterior density for all of the individuals, we estimate parameters based on $\pi(\theta|t_{-i}), i:1,2,...,n$, the posterior density with the i^{th} observation removed. The more difference between parameter estimates based on all of the data set and parameter estimates after removing an individual, the more effective is the deleted individual.

An appropriate tool for measuring this difference is Kullback-Leibler divergence between predictive densities based on full and deleted case data. This for individual i is given by:

$$K_i = \int \pi(\theta|t) \log\left(\frac{\pi(\theta|t)}{\pi(\theta|t_{-i})}\right) d\theta$$

which can be approximated by: (Christensen et al., 2011; page 341)

$$K_i. = \log\left(\frac{1}{m}\sum_{j=1}^{m}[1/L(\theta_j|t_i,x_i)]\right) - \frac{1}{m}\sum_{j=1}^{m}\log[1/L(\theta_j|t_i,x_i)].$$

where t_i and x_i are unemployment duration and the vector of explanatory variables for individual i, θ_j is a random draw from $\pi(\theta|t_i,x_i)$, and $L(\theta|t_i,x_i)$ is the likelihood function for i^{th} individual.

Figure 6 presents standardized Kullback-Leibler divergence, $\frac{K_i - \overline{K}}{S_K}$ [where, $\overline{K} = \frac{1}{n}\sum K_i$ and $S_K^2 = \frac{1}{n-1}\sum (K_i - \overline{K})^2$] versus the index of individuals. The index of individuals with standardized Kullback-Leibler divergence more than 3 are given insides the plots. We consider these points as unusual cases. This figure shows that the number of unusual points detected in log-logistic AFT model are less than other models. After investigating the data set, we can conclud that these points are often widow or divorced persons with high values of right censored unemployment duration.
In our data set, for making sure of robustness of our results, we delete unusual individuals in log-logistic AFT model and reanalysis the data set. The new results show that except the regression coefficient of widow or divorced which has higher estimate than before, other parameters estimates are almost non-sensitive with respect to the deletion of this group of people.

6. Conclusion

In this paper, we used Bayesian method and WinBUGS software as tools for analyzing unemployment duration data when proportionality assumption is not valid.
In our analysis, we applied Bayesian accelerated failure time model under three distributional assumptions. We used a method for sensitivity analysis of results with respect to different prior parameters, and we investigated sensitivity analyses with respect to individual deletion using Kullback-Leibler divergence. Also, in another analysis, all of the unusual people are omitted. Parameter estimates in these cases show that the results are not sensitive.

References

1. Ackum. S. (1991). Youth unemployment, Labour market programmes and subsequent earnings, Scandinavian Journal of Economics, 93, p. 531-543.
2. Arjas. E. and Bhattacharjee, M. (2004). Modeling heterogeneity in repeated failure time data: A hierarchical Bayesian approach, In Mathematical Reliability, Kluwer Academic Publishers, Norwell, MA.
3. Arjas, E. and Gasbarra. D. (1994). Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler, Statistica Sinica, 4, p. 505-524.
4. Armitage, P. B. G. (1959). Statistical Methods in Medical Research. Blackwell.
5. Berg. V., Gerard. J., Gijsbert. A., Lomwel. C., and Jan, C. Van Ours (2008). Nonparametric Estimation of a Dependent Competing Risks Model for Unemployment Durations, Empirical Economics 34(3), p. 477-491.
6. Borsic. D. and Kavkler. A. (2009). Modeling Unemployment Duration in Slovenia using Cox Regression Models, Transit Stud Rev, 16, p. 145-156.
7. Card, D. and Sullivan, D. (1988). Measuring the effect of subsidized training programs on movements in and out of employment, Econometrica, 56, p. 497-530.
8. Campolietim M. (2001). Bayesian semiparametric estimation of discrete duration models: an application of the dirichlet process prior, Journal of Applied Econometrics, 16, p. 1-22.
9. Campodnico. S. and Singpurwalla, N. D. (1994). A Bayesian analysis of the logarithmic-Poisson execution time model based on expert opinion and failure data, IEEE Transactions on Software Engineering, 20, p. 677-683.
10. Christensen. R., Johnson. W., Branscum A., Hanson. T. E., (2011). Bayesian Ideas and Data Analysis, An Introduction for Scientists and Statisticians, Chapman & Hall / CRC.
11. Cox, D. R. (1972). Regression models and life tables, Journal of Royal Statistical Society, B34, p. 187-220.
12. Deshpande. J. V and Purohit. S. G., (2005). Life Time Data: Statistical Models and Methods, World Scientific.
13. Elandt-Johnson, R. C. and Johnson, N. L. (1980). Survival Models and Data Analysis. Wiley, New York.
14. Gilks. W., Richardson. S. and Spiegelhalter. D. (1996). Markov Chain Monte Carlo in Practice, Interdisciplinary Statistics, Chapman & Hall, Suffolk, UK.
15. Gonzalo, M. T. and Saarela. J. (2000). Gender Differences in Exit Rates from Unemployment: Evidence from a Local Finnish Labour Market. Finnish Economic Papers, Autumn 2000.
16. Greene. W. H. (2003). Econometric Analysis, 5th edition, Prentice Hall.
17. Hamada, M. S., Wilson, A., Reese, C. S. and Martz, H. (2008). Bayesian Reliability, Springer.
18. Hanson, T. and Johnson, W.O. (2004). A Bayesian semiparametric AFT model for interval-censored data, J. Comput. Graph. Stat., 13, p. 341-361.
19. Ibrahim. J. G., Chen. M. and Sinha. D. (2005). Bayesian Survival Analysis, Springer.
20. Kalbfleisch, J. D. (1978). Non-parametric Bayesian analysis of survival time data, Journal of Royal Statistical Society Ser. B, 40 (2), p. 214-221.

21. Kalbfleisch. J. D. and Prentice. R. L. (1980). The Statistical Analysis of Failure Time Data, Wiley, New York.
22. Kiefer. N. M, (1988). Economic Duration Data and Hazard Functions, Journal of Economic Literature, 26, p. 646-679.
23. Kudus, K. A, Kimbera, A. C. and Lapongan, J. (2006). A parametric model for the interval censored survival times of acacia mangium plantation in a spacing trial, Journal of Applied Statistics, 33, p. 1067–1074.
24. Kupets, O. (2006). Determinants of unemployment duration in Ukraine, J Comp Econ. 34, p. 228-247.
25. Lee, E. T. (1980). Statistical Methods for Survival Data Analysis. Lifetime Learning Publications, Belmont. N. D.
26. Lee, E. T. and Wang, G. W. (2003). Statistical Methods for Survival Data Analysis, Lifetime Learning Publications, Wiley, New York.
27. Meyer, B. D. (1990). Unemployment insurance and unemployment spells, Econometrica, 58 p. 757-782.
28. Moffitt, R. A. (1999). New developments in econometric methods for labor market analysis, In: O. Ashenfelter, and D. Card (eds), Handbook of Labor Economics, Chapter 24, p. 1367-1397.
29. Nilsen. A., Risa. A. E. and Torstensen, A. (2000). Transitions from employment among young Norwegian workers, J. Popul. Econ., 13, p. 21-34
30. Ntzoufras, I. (2009). Bayesian Modeling Using WinBUGS. Wiley Series in Computational Statistics, Hoboken, USA.
31. Padgett, W. J. and Wei, L. J. (1981). A Bayesian nonparametric estimator of survival probability assuming increasing failure rate, Communications in Statistics – Theory and Methods, A10 (1), 49-63.
32. Paserman, M. D. (2004). Bayesian Inference for Duration Data with Unobserved and Unknown Heterogeneity: Monte Carlo Evidence and an Application, No 996, IZA Discussion Papers, Institute for the Study of Labor (IZA).
33. Ruggiero, M. (1994). Bayesian Semiparametric Estimation of Proportional Hazards Models, Journal of Econometrics, 62(2), p. 277-300.
34. Spiegelhalter, D.J., Best. N.G., Carlin, B.P. and Lindevan, der. A. (2002). Bayesian measures of model complexity and fit, J. R. Stat. Soc. Ser. B 64, p. 583–616.
35. Spiegelhalter, D.J., Thomas, A., Best, N. and Lunn, D. (2003). WinBUGS Examples, MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology and Public Health, Imperial College School of Medicine, UK, available at $http://www.mrc-bsu.cam.ac.uk/bugs$.
36. Torp, H. (1994). The impact of training on employment: Assessing a Norwegian labour market programme, Scandinavian Journal of Economics, 96, p. 531-550.
37. Winkelmann, R. (1997). How young workers get their training: A survey of German versus the United States, Journal of Population Economics, 10(2), p. 159-170.
38. Zhang, J. and Lawson, A. B. (2011). Bayesian parametric accelerated failure time spatial model and its application to prostate cancer, Journal of Applied Statistics, 38(3), p. 591-603.