A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 285, 2013

*Wioletta Grzenda**

THE SIGNIFICANCE OF PRIOR INFORMATION IN BAYESIAN PARAMETRIC SURVIVAL MODELS

Abstract. The Bayesian approach gives the possibility of using in the research additional information that is external to the sample. The primary objective of this paper is to analyse the impact of the prior information on the posterior distribution in Bayesian parametric survival models. In this work the exponential models and Weibull models with different prior distributions have been estimated and compared. The aim of this research is to investigate the determinants of unemployment duration. The models have been estimated using Markov chain Monte Carlo method with Gibbs sampling.

Key words: survival parametric models; Bayesian inference; prior distribution; MCMC method; unemployment.

I. INTRODUCTION

The Bayesian approach with the prior distribution gives the possibility of incorporating in the research additional information that is external to the sample. A prior distribution is the probability distribution that expresses the whole knowledge of a statistician on the estimated parameters before the data has been examined, it describes a degree of belief in different values of parameters (Silvey, 1978). We often have some additional prior information, for example from previous statistical analysis, which is worth using in the current research. Combining sample data and knowledge external to the sample allows us to obtain more efficient estimators. The estimate precision and credibility can be improved even with general prior information, which may be expressed as a prior distribution with large dispersion (Szreder, 1994).

The choice of the prior distribution may be determined by such factors as the experiences gained in the previous studies, the researcher's intuition or so-called expert knowledge. If prior information comes from the previous research, Bayesian estimation should lead to more precise results than classical methods. However, if prior information is subjective, obtained results might not be very credible.

* Ph.D., Institute of Statistics and Demography, Warsaw School of Economics.

The prior distributions which contain exact information and have an impact on the posterior distribution are called informative priors. In Bayesian approach, if we intend to obtain objectively correct results, we use prior distributions which have a minimal impact on the posterior distribution. Such distributions are called objective or non-informative prior distributions (Gelman et al., 2000).

In this paper the impact of prior information on the posterior distribution for the parametric survival models is scrutinized (Ibrahim *et al.*, 2001; Kim and Ibrahim, 2000). For this purpose exponential models and Weibull models have been estimated with standard non-informative prior distributions and informative prior distributions from previous studies. The aim of this research is to investigate the determinants of unemployment duration.

II. BAYESIAN PARAMETRIC SURVIVAL MODELS

While investigating prior information in the parametric survival models we will concentrate on two most popular models: exponential and Weibull models (Blossfeld and Rohwer, 1995). Further information on the Bayesian approach to parametric survival models can be found in the work of Ibrahim *et al.* (2001).

Let $\mathbf{y} = (y_1, \ldots, y_n)'$ be survival times, where y_i, $i = 1,...,n$ are independent and have an identical exponential distribution with parameter λ. The censoring indicators we denote by $\mathbf{v} = (v_1, \ldots, v_n)'$, where $v_i = 0$ if y_i is right censoring and $v_i = 1$ if y_i is failure time, $i = 1,...,n$. The density function for y_i is $f(y_i \mid \lambda) = \lambda \exp(-\lambda y_i)$, the survival function $S(y_i \mid \lambda) = \exp(-\lambda y_i)$. Let $\mathbf{X}$, ($n \times k$) be a matrix of independent variables, for which $\mathbf{x}_i'$ denotes *i*th row. Then observed data is represented as follows $D = (n, \mathbf{y}, \mathbf{X}, \mathbf{v})$.

Let us assume that $\lambda_i = \varphi(\mathbf{x}_i' \boldsymbol{\beta})$, where $\mathbf{x}_i$, ($k \times 1$) is a vector of covariates, $\boldsymbol{\beta}$, ($k \times 1$) is a vector of regression coefficients and φ is a known function. For $\varphi(\mathbf{x}_i' \boldsymbol{\beta}) = \exp(\mathbf{x}_i' \boldsymbol{\beta})$, we have the following likelihood function:

$$L(\boldsymbol{\beta} \mid D) = \prod_{i=1}^{n} f(y_i \mid \lambda_i)^{v_i} S(y_i \mid \lambda_i)^{(1-v_i)} =$$

$$= \prod_{i=1}^{n} [\exp(\mathbf{x}_i'\boldsymbol{\beta})\exp(-y_i \exp(\mathbf{x}_i'\boldsymbol{\beta}))]^{v_i} [\exp(-y_i \exp(\mathbf{x}_i'\boldsymbol{\beta}))]^{(1-v_i)} = \tag{1}$$

$$= \exp\left\{\sum_{i=1}^{n} v_i \mathbf{x}_i'\boldsymbol{\beta}\right\} \exp\left\{-\sum_{i=1}^{n} y_i \exp(\mathbf{x}_i'\boldsymbol{\beta})\right\}.$$

Most frequently for regression coefficients $\boldsymbol{\beta}$ we choose uniform improper prior or normal prior distribution. In our model we take a k–dimensional normal prior $N_k(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ for $\boldsymbol{\beta}$, where $\boldsymbol{\mu}_0$ denotes the prior mean vector, and $\boldsymbol{\Sigma}_0$ denotes the prior covariance matrix. Then the posteriori distribution for $\boldsymbol{\beta}$ is given by

$$p(\boldsymbol{\beta} \mid D) \propto L(\boldsymbol{\beta} \mid D) p(\boldsymbol{\beta} \mid \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0). \tag{2}$$

where $p(\boldsymbol{\beta} \mid \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ denotes multivariate normal density.

The second model investigated in this paper is Weibull model. Let now $\mathbf{y} = (y_1, \ldots, y_n)'$ denote survival times, where y_i, $i = 1, \ldots, n$ are independent and have identical Weibull distribution with parameters: α and γ. The censoring indicators are denoted as previously. Then the density function for y_i is $f(y_i \mid \alpha, \gamma) = \alpha \gamma y_i^{\alpha-1} \exp(-\gamma y_i^{\alpha})$. For $\lambda = \ln(\gamma)$ we have the following likelihood function:

$$f(y_i \mid \alpha, \lambda) = \alpha y_i^{\alpha-1} \exp(\lambda - \exp(\lambda) y_i^{\alpha}), \tag{3}$$

whereas the survival function is given by the formula: $S(y_i \mid \alpha, \lambda) = \exp(-\exp(\lambda) y_i^{\alpha})$.

The unknown parameters α and λ are random variables, so that we can assume their independence in further considerations.

For the observed data $D = (n, \mathbf{y}, \mathbf{X}, \mathbf{v})$ and $\lambda_i = \mathbf{x}_i' \boldsymbol{\beta}$, we have the following likelihood function:

$$\begin{aligned} L(\boldsymbol{\beta}, \alpha \mid D) &= \prod_{i=1}^{n} f(y_i \mid \alpha, \lambda_i)^{v_i} S(y_i \mid \alpha, \lambda_i)^{(1-v_i)} = \\ &= \prod_{i=1}^{n} \left[\alpha y_i^{\alpha-1} \exp(\mathbf{x}_i'\boldsymbol{\beta} - \exp(\mathbf{x}_i'\boldsymbol{\beta}) y_i^{\alpha})\right]^{v_i} \left[\exp(-\exp(\mathbf{x}_i'\boldsymbol{\beta}) y_i^{\alpha})\right]^{(1-v_i)} \\ &= \alpha^{\sum_{i=1}^{n} v_i} \exp\left\{\sum_{i=1}^{n} \left[v_i (\alpha - 1) \ln(y_i) + v_i \mathbf{x}_i'\boldsymbol{\beta} - \exp(\mathbf{x}_i'\boldsymbol{\beta}) y_i^{\alpha}\right]\right\}. \end{aligned} \tag{4}$$

For regression coefficients $\boldsymbol{\beta}$ we choose *k*-dimensional normal prior distribution $N_k(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, whereas for α we choose gamma prior distribution $G(\alpha_0, k_0)$. Then the posteriori distribution is given by:

$$p(\boldsymbol{\beta}, \alpha \mid D) \propto L(\boldsymbol{\beta}, \alpha \mid D) p(\boldsymbol{\beta} \mid \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) p(\alpha \mid \alpha_0, k_0), \tag{5}$$

where $p(\boldsymbol{\beta} \mid \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ denotes multivariate normal density and $p(\alpha \mid \alpha_0, k_0)$ gamma distribution.

III. EMPIRICAL EXAMPLES

The empirical examples presented in this study refer to the event of unemployment. The analysis of the unemployment determinants with the use of classical event history models can be found in Drobnič and Frątczak's (2001) work.

The data set used in this study comes from the survey of Central Statistical Office – "Household budgets in 2009". Depending on the place of living the impact of individual considered factors on unemployment duration is different. For the purpose of this research, we take into consideration people living in the cities of more than 200 000 inhabitants, unemployed, looking for a job and ready to take it up (Eurostat). In this way 502 individuals were selected, 27 of them have already found a job and waited for it to start – for these subjects the event holds, while the others are censored individuals.

In this model, time is a dependent variable defined as the number of months of unemployment. Since different factors can determine unemployment depending on its duration, only such individuals were investigated who had remained unemployed maximally for 24 months. Characteristics of human capital which most frequently diversify unemployment rates such as *age* (in years), *sex* (1 – man, 2 – woman), *education level* (1 – higher, 2 – post-secondary, 3 – secondary professional, 4 – secondary general, 5 – basic vocational, 6 – primary school were chosen as independent variables.

The models were estimated using Markov chain Monte Carlo method with Gibbs sampling (Casella and George, 1992). The number of burn-in samples is assumed to be 2000 and the posterior samples 10000. Using Geweke's test (Geweke, 1992) it was found that there is no indication that the Markov chain has not converged for all the parameters of investigated models, with the significance level of 0.01.

Table 1. The prior distributions and the posterior distributions

Parameter	Model 1				
	Prior distributions		Posterior distributions		
	Mean	Standard dev.	Mean	Standard dev.	HPD
Intercept	0	10^6	4.3154	0.0291	(4.2598, 4.3728)
Sex 1	0	10^6	0.0329	0.00985	(0.0144, 0.0531)
Education 1	0	10^6	–2.2786	0.0256	(–2.3331, –2.2319)
Education 2	0	10^6	–1.8051	0.0340	(–1.8715, –1.7386)
Education 3	0	10^6	–0.6908	0.0283	(–0.7464, –0.6360)
Education 4	0	10^6	–1.6778	0.0270	(–1.7339, –1.6263)
Education 5	0	10^6	–1.5169	0.0263	(–1.5700, –1.4667)
Age	0	10^6	0.0617	0.00052	(0.0607, 0.0628)

Source: own calculations.

For non-informative prior distributions results similar to maximum likelihood estimates were received, but their significance is different. In table 2 posterior distributions for different prior distributions for variables such as *sex* and *age* are presented. In all models for other parameters normal prior distributions were chosen. To obtain credible results informative prior distributions from the same research in the year 2008 were used. In models: I, II, III and IV the mean from the year 2008 for cities over 200 000 inhabitants was used. In models V and VI the mean from the year 2008 was used referring to the data for the whole country.

For data from 2009, estimated with non-informative prior distributions, parameter *sex* has a lower value than for the data from 2008, under the same assumption. For big variance we received similar estimations independent of the assumed mean. Often instead of non-informative prior distributions, least-informative prior distributions are chosen, for example normal distribution with zero means and variance 1. With this assumption the results are similar to the previous results for zero means and variance 10^6. When choosing small variance one must be careful. In model VI (Table 2) for parameter *sex* in informative prior with mean for the whole country we received another direction of interdependence. For parameter *age* in model VI (Table 2) the mean for big cities and the whole country is similar, therefore results in all models are similar.

Table 2. The prior distributions and the posterior distributions for *sex* and *age*.

Model	*Sex*				*Age*			
	Prior distributions		Posterior distributions		Prior distributions		Posterior distributions	
	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
I	0	10^6	0.0329	0.00985	0	10^6	0.0617	0.00052
II	0.1532	10^6	0.0331	0.00966	0.0178	10^6	0.0617	0.00052
III	0.1532	1	0.0331	0.00973	0.0178	1	0.0617	0.00053
IV	0.1532	0.0001	0.0506	0.00114	0.0178	0.0001	0.0616	0.00052
V	–0.3350	1	0.0330	0.01000	0.0160	1	0.0617	0.00052
VI	–0.3350	0.0001	–0.0180	0.00037	0.0160	0.0001	0.0615	0.00052

Source: own calculations.

With similar assumptions for regression coefficients two Weibull models were estimated, in which a non-informative gamma prior distribution with the shape parameter equalling 0.001 and the same value for the inverse scale parameter were additionally chosen for the shape parameter.

Table 3. The prior distributions and the posterior distributions for Weibull models.

Parameter	Weibull model 1				Weibull model 2			
	Prior distributions		Posterior distributions		Prior distributions		Posterior distributions	
	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
Intercept	0	10^6	3.9213	0.0262	0	10^6	3.9081	0.0247
Sex 1	0	10^6	0.00644	0.00833	0.1532	0.0001	0.0344	0.00119
Education 1	0	10^6	–1.9597	0.0234	0	10^6	–1.9532	0.0223
Education 2	0	10^6	–1.5210	0.0298	0	10^6	–1.5229	0.0291
Education 3	0	10^6	–0.5956	0.0249	0	10^6	–0.5931	0.0241
Education 4	0	10^6	–1.4952	0.0240	0	10^6	–1.4870	0.0229
Education 5	0	10^6	–1.3207	0.0236	0	10^6	–1.3222	0.0227
Age	0	10^6	0.0551	0.00046	0.0178	0.0001	0.0550	0.000457
Weibull shape	gamma		1.1939	0.00427	gamma		1.1932	0.00424

Source: own calculations.

For Weibull models (Table 3) we can observe the same impact of prior distributions on posterior distribution as in exponential models. The exponential model was estimated as a special case of Weibull model. At the significance level of 0.05 the hypothesis that shape parameter of Weibull model equals 1 has to be rejected. Moreover, a smaller value of DIC statistics shows the superiority of Weibull model.

IV. SUMMARY AND CONCLUSIONS

The abovementioned examples indicate how changes in prior distributions influence posterior distribution. The researcher's belief in having information before the research is expressed as a value of standard deviation in prior distributions. The big standard deviation indicates a lack of precise information about the problem in question. It was suggested that if a big enough variance is selected, slight changes of mean do not influence the results of estimates.

The sample data is the foundation of statistical inference. If we have much sample information, the significance of prior information decreases. Even significant changes in prior distributions do not alter greatly posterior distributions (Silvey, 1978). Therefore in this paper a small size sample limited to the inhabitants of big cities was investigated in order to show the influence of prior distributions on posterior distributions. As the classical approach leverages boundary theorems and requires a big sample, for a small sample it is essential to use Bayesian approach, even if we do not have any prior information, because we can choose non-informative prior distributions (Gelman et al. 2000).

The presented empirical examples seem to reveal that all variables are statistically significant for 0.05, except *sex* variable in Weibull model with non-informative prior distributions. For people living in big cities it might be assumed that having a higher education level than primary gives a better chance of finding a job. It was indicated that the likelihood of finding a job decreases by about 5% as the age of a respondent increases by one year. For these two characteristics the findings for big cities are similar to these observed in the whole country (Grzenda, 2011), whereas for the *sex* determinant the results suggest different trends. In fact, for big cities it was found that men are about 3% less likely to find a job than women, which is unlike in the entire country.

REFERENCES

Blossfeld H.P., Rohwer G. (1995), *Techniques of event history modeling, New approaches to causal analysis*, Hillsdale, NJ: L. Erlbaum.

Casella G., George E.I. (1992), Explaining the Gibbs sampler, *The American Statistician*, 46, 167–74.

Drobnič S., Frątczak E. (2001), Employment patterns of married women in Poland, *Careers of couples in contemporary society*, New York.

Gelman A., Carlin J.B., Stern H.S., Rubin D.B. (2000), *Bayesian data analysis*, Chapman & Hall/CRC, London.

Geweke J. (1992), Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In Bernardo J., Berger J., Dawiv A., Smith A., *Bayesian Statistics*, 4, 169–193.

Grzenda W. (2011), Bayesian exponential survival model in analysis of unemployment duration determinants, *Acta Universitatis Lodziensis*, *Folia Oeconomica*, (Przyjęte do publikacji).

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

Kim S.W., Ibrahim J.G. (2000), On Bayesian Inference for Parametric Proportional Hazards Models Using Noninformative Priors, *Lifetime Data Analysis*, 6, 331–341.
Silvey S. D. (1978), *Wnioskowanie statystyczne*, PWN, Warszawa.
Szreder M. (1994), *Informacje a priori w klasycznej i bayesowskiej estymacji modeli regresji*, Uniwersytet Gdański, Gdańsk.

Wioletta Grzenda

ZNACZENIE INFORMACJI *A PRIORI* W BAYESOWSKICH PARAMETRYCZNYCH MODELACH PRZEŻYCIA

W pracy przedstawiono parametryczne modele przeżycia w ujęciu bayesowskim. Podejście bayesowskie wymaga zadania rozkładów *a priori* dla szacowanych parametrów modelu. Rozkład *a priori* parametru jest rozkładem prawdopodobieństwa, który wyraża całą wiedzę badacza o szacowanym parametrze przed sprawdzeniem aktualnych danych. W literaturze przedmiotu często spotyka się nieinformacyjne rozkłady *a priori*, które wyrażają brak wstępnej wiedzy badacza o szacowanych parametrach modelu. W celu pokazania znaczenia informacji *a priori* oraz jej wpływu na rozkład *a posteriori* oszacowano kilka parametrycznych modeli przeżycia przy różnych rozkładach *a priori*. Przedmiot badań stanowią determinanty długości czasu pozostawania bez pracy.