Ratemaking for a New Territory: Enhancing GLM Pricing Model with a Bayesian Analysis

Jing Zhang and Tatjana Miljkovic

Abstract

Motivation. This paper offers a Bayesian approach in ratemaking for a new territory where a company considers starting a new business, or for a relatively new territory where the company has very limited claims experience.
Method. A Bayesian Poisson regression model with power priors and weakly informative priors for the model parameters is proposed for modeling claims frequency. Bayesian analysis of claim severity considers a gamma regression and non-informative uniform priors for the regression coefficients.
Results. After incorporating the external information from a similar book of business in a similar territory, Bayesian analysis with power priors improved the prediction reporting a small Means Squared Prediction Error (MSPE).
Conclusions. Bayesian analysis with power priors can be used effectively in auto insurance ratemaking for pricing of a new business in a new territory, or improving pricing of a growing business in a new territory.
Availability. The original SAS code will be available for distribution pending the acceptance of this paper.

Keywords. Bayesian analysis, GLM, new territory, power priors, predictive modeling, ratemaking.

1. INTRODUCTION

1.1 Research Context

Ratemaking for a new line of business or a new territory is subject to a judgement under uncertainty. Actuaries in these situations often rely on the availability of external industry data or experience from a similar line of business, as both of these serve as heuristic benchmarks, but sometimes they lead to severe and systematic errors. If the volume of claims experience is subject to significant changes (e.g., due to catastrophic events or regulatory conditions), these estimates will be severely biased. The company may gauge some prior information about the prospective new business in a new territory by pooling this information from the existing business, assuming the new underwriting practices in a new territory will remain more or less similar to the existing underwriting practices to the territory from which this information is drawn. A new territory may also share some common demographic, geographic, or climate characteristics with one of the existing territories so that the information contained in the existing business can be utilized in the rating process of the new territory.

According to Chen and Ibrahim (2006, pp. 551), "Power priors have emerged recently as useful

informative priors for the incorporation of historical data in a Bayesian analysis" and are well-received in statistical practice. These priors can be efficiently incorporated in Bayesian analysis with generalized linear models (GLM) and help incorporate useful prior information from existing territories in the context of analyzing limited information from the new territory of interest.

Most of the insurance companies are moving away from the one-way premium calculation approach by employing GLMs with the original statistical framework discussed in the book by McCullagh and Nelder (1989). The GLM models are praised for two major advantages over ordinary linear models. First, the GLMs work with a number of discrete distributions and continuous distributions, which make them more flexible compared to the ordinary linear model that is constrained by the normal distribution only. Second, the GLMs allow for some transformation of the mean as a linear function of the covariates, with additive and multiplicative models as special cases. For more extensive theory behind non-life insurance pricing using GLMs, we refer the reader to books by Kaas et al. (2008) and Ohlsson and Johansson (2010).

A frequentist approach to predictive modeling based on GLM models has the capability to predict outcomes that best represent the company's data with insufficient regard for prior probability. The probability distributions of the parameters considered in this type of modeling rely on the sampling distributions that are based on all possible random samples of experiences that could have occurred, but they are not conditional on the actual sample that did occur. A Bayesian point of view considers inferences based on the probabilities calculated from the posterior distribution, making them conditional on the sample that actually did occur. The role of prior distributions in the Bayesian analysis is to capture "pre-data" information about the parameters, then use the prior experience that was collected to update the "pre-data" information about the parameters to "posterior" information about the parameters. Thus, the Bayesian approach considers parameters as random variables.

Recently, Bayesian methods have been actively discussed in the area of predictive modeling and ratemaking. Boucher et al. (2008) used Bayesian and frequentist models based on generalization of Poisson and negative binomial distributions to account for correlation between contracts of the same insureds. The authors showed that the models based on time dependence covariates (e.g., past experience) cannot be used in modeling of reported claims. They recommended use of random effects models in computing the next year's premium as these models show improved fit compared to other models. The same authors, Boucher et al. (2009), extended their study by considering the relationship

between number of accidents and number of claims using the generalization of the zero-inflated Poisson (ZIP) distribution. The authors proposed an approximation of the number of accidents distribution that can be used to provide insightful information about the behavior of insureds using panel count data. A Bayesian analysis was used in computation of the predictive distribution for the random effects.

Bermúdez and Karlis (2011) examined Bayesian multivariate Poisson models and their zero-inflated extensions for improving current ratemaking procedures. Brown and Buckley (2015) used a Bayesian approach to determine the number of groups in an insurance portfolio. The claim count is assumed to follow a Poisson distribution.

We consider the following scenario for pricing new business in a new territory, where there is no prior claims experience. First, we can identify a similar territory from our existing book of business for which the claim experience is established. These two territories may be neighbors that share similar climate, geography, and demographic characteristics. For pricing the new business during the first year with no data, we can borrow the information from the existing territory and set the new rates. After the first year, for pricing the business during the second year, we can borrow the experience from the similar existing territory in the analysis of the limited claim experience in the new territory. Then, we can run the proposed Bayesian model with power priors. We repeat this process for several years until we accumulate the claims experience in the new territory to be able to use the standard pricing method. The flow chart of this process is outlined in Figure 1. Our proposed Bayesian model with power priors would provide a new way of pricing the business for a new territory (framed part of Figure 1) and serves as the main contribution of this paper. The example that we provided in the subsequent sections would help the practitioners in implementation of this proposed method.

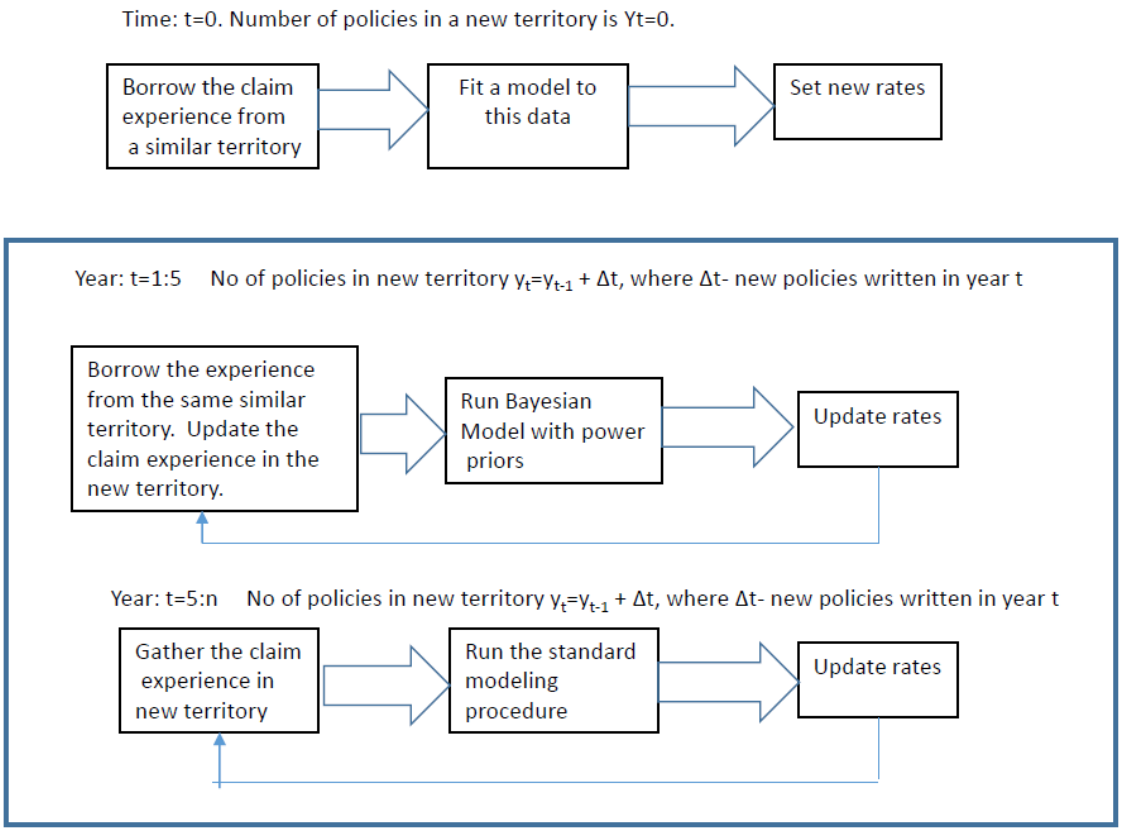


Figure 1: Flow chart of the proposed Bayesian method for pricing in a new territory.

1.2 Objective

The objective of this paper is to introduce a Bayesian approach with power priors and weakly informative priors to be used in developing frequency distribution of claims for a new territory where the company has very limited experience. The historical information can be borrowed from an adjacent territory based on geographic and demographic profiles, for purpose of the Bayesian analysis. A Bayesian analysis with non-informative priors for modeling severity of claims is also illustrated in a new territory.

To our knowledge, the Bayesian GLM claim models with a Poisson distribution have not previously been considered, either with power priors or weakly informative priors. We would like to close this gap in the actuarial literature by proposing the Bayesian frequency models that use power

priors and weak informative priors of the regression coefficients. This approach is especially appealing for determining the premium rates in a new territory that lacks claims experience.

1.3 Outline

The remainder of the paper proceeds as follows. Section 2 presents the Bayesian methodology for modeling frequency and severity of claims. Section 3 describes the analysis of real data and the results. Section 4 provides the summary of the model validation. The conclusion is provided in Section 5.

2. BACKGROUND AND METHODS

In this section, we explore the models for claims frequency and claims severity. For each model, we show frequentist and Bayesian approaches from a theoretical perspective.

2.1 Models for Claims Frequency

It is popular to assume that the number of claims follows a Poisson distribution and, hence, a generalized linear regression can be fitted to analyze the relationship between the number of claims and the relevant predictors.

$$Y_i|\theta_i \sim independent\ Poisson(E_i\,\theta_i) \tag{2.1}$$

$$\log(\theta_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}, \tag{2.2}$$

where the vertical bar "|" describes the distribution of the quantity to the left of the "|", given information to the right. Here Y_i denotes the number of claims filed by the i*th* policy holder. Here, the vector of predictors is defined as $x_i = (x_{1i}, \ldots, x_{pi})'$. The Poisson mean, $E_i\,\theta_i$, is determined by the known length of insured time (E_i, also known as the offset) and rate of claims (θ_i). Here the rate of claims is modeled as a function of the relevant predictors (θ_i), including demographic information of drivers, descriptive information of cars and residential areas. The regression coefficients, β_0, …., β_p, relate the rate of claims with these predictors.

In the frequentist approach, point estimation of model parameters can be implemented via Maximum Likelihood Estimation (MLE) or Restricted Maximum Likelihood (REML) approaches, and inferences can be made based on large sample distributions of the point estimators. Other distributional assumptions of the claims frequency can be used, such as zero-inflated Poisson or negative binomial. Initially, we considered regression models assuming these distributions as well; however, the fit of Poisson regression turns out to be the best for the data analyzed. Since the main purpose of the present study is to illustrate the incorporation of prior information from an external existing territory via Bayesian modeling of claims frequency when sample size is limited, we decided to stay with the Poisson distributional assumption.

The Bayesian analysis treats the parameters as unknown random variables. To implement the analysis, we need to propose a "prior distribution" for the model parameters. Combining the data likelihood and prior distribution of parameters using Bayes theorem, we are able to update the knowledge about the distribution of model parameters, and the updated knowledge is called "posterior distribution." The posterior distributions are then used for Bayesian inference. Here we begin the Bayesian analysis assuming independent normal prior distributions for the regression coefficients, i.e.,

$$\pi\left(\beta_j|\beta_j^0,\sigma_j^2\right) = N(\beta_j^0,\sigma_j^2),\ j = 0,1,\dots,p. \tag{2.3}$$

Higher level priors are then assumed for prior mean β_j^0 and prior variance σ_j^2 as follows:

$$\pi(\beta_0^0) = N(0,10) \tag{2.4}$$

$$\pi\left(\beta_j^0\right) = N(0,4),\ j = 1,\dots,p. \tag{2.5}$$

$$\pi(\sigma_0) = Uniform(0,5), \pi\left(\sigma_j\right) = Uniform(0,1), j = 1,\dots,p. \tag{2.6}$$

The hyper-parameters are chosen to incorporate weak informative prior distributions on the parameters. Besides the weakly informative priors, we also illustrate the incorporation of prior information from external data of similar region via power priors. The power priors have been

proposed in Ibrahim and Chen (2000), with applications in hierarchical modeling discussed in Chen and Ibrahim (2006) and well-received in statistical practice.

The power prior of model parameters is constructed by raising the likelihood based on the external data to a suitable power and then multiplied by an initial prior (usually non-informative or weakly informative); therefore, power prior uses the external data with a discount relative to the data of interest, which allows a discrepancy between insurance policy holders in this similar region and the current region of interest. The power prior is a useful tool to borrow strength from external data in Bayesian analysis. In the present study, we considered a second Bayesian analysis that incorporates the external data using power prior with power of 0.5, which implies a 50% discount of external information in the log-likelihood function of the joint posterior density function of model parameters; the priors used in the first Bayesian analysis (i.e. Equations (2.4)-(2.6)) are used as initial priors in this analysis.

2.2 Models for Severity

Besides the modeling of frequency of claims, it is also of interest to study whether and how the amount of each claim (severity) is related to the relevant factors (e.g., driver's age, gas type, etc.). Claim amounts are continuous measurements and can be analyzed with ordinary linear regression or generalized linear regression (e.g., log-normal regression or gamma regression). Note that the distributional assumptions that allow heavier right tails are usually a better fit to the loss data due to right-skewness of such data. When claim amounts are assumed to follow gamma distributions,

$$Z_i|\mu_i,\nu_i \sim Gamma(\mu_i,\nu_i) \tag{2.7}$$

Or equivalently,

$$f(Z_i|\mu_i,\nu_i) = \frac{1}{\Gamma(\nu_i)}\left(\frac{\nu_i}{\mu_i}\right)^{\nu_i}(z_i)^{\nu_i-1}\exp\left(-\frac{\nu_i z_i}{\mu_i}\right) \tag{2.8}$$

where ν_i is the shape parameter of the gamma distribution, and μ_i is the mean of the gamma variable and relates the covariates with the severity response. Using a log-link function, we have.

$$\log(\mu_i) = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \cdots + \gamma_p x_{pi} \tag{2.9}$$

The regression coefficients, γ_0, …., γ_p, relate the severity of claims with the set of predictors defined as $x_i = (x_{1i}, \ldots, x_{pi})$. Note that we assumed the same set of predictors are considered in the analysis of frequency and severity of claims in Equations (2) and (9), which can be modified in practice according to availability of data and prior beliefs. The two sets of covariates used in these two models are not necessarily the same.

Frequentist approaches can be used to fit these generalized linear models described above to the severity data of the new territory, and likelihood-based inference would help us determine the relationship between severity and covariates. When expert knowledge or existing analysis results concerning this relationship from a similar territory are available, the Bayesian approach would help us incorporate the information through prior elicitation. However, we believe that one should be cautious of using power priors in the analysis of severity since the potential outliers or heavy right tail in the severity observations of the "external" data might introduce misleading information in the analysis and bias the conclusions. In the present study, we used non-informative uniform priors for the regression coefficients:

$$\pi(\gamma_i) \propto 1, i = 0, 1, \ldots, p. \tag{2.10}$$

The prior distribution of shape parameter is specified through the following parameterization. Let $\kappa_i = \frac{\nu_i}{\mu_i}$ be the rate parameter, then we assume an inverse-gamma prior distribution for the rate parameter as follows,

$$k_i \sim InverseGamma(0.001, 0.001) \tag{2.11}$$

The specified prior distributions would then provide vague prior input for the analysis.

3. RESULTS AND DISCUSSION

In this section, we illustrate the proposed methodology using the data from a French insurance company, related to 677,991 motor third-party liability policies. The data set includes exposure information as well as the loss information and can be found as part of the "CASdatasets" library in the R software (CASdatasets). The discussion about the datasets used in the book by Charpentier (2014) and the book itself, can be found in the book review by Miljkovic (2017). Charpentier (2014) discussed the modeling of claims frequency and severity of this data based on a frequentist approach, using various GLM models. The rating factors include: region (R11, R23, R24, R25, R31, R53, R54, R72, R74), car age (0-100), density (2-27000), engine power (12 levels), brand (7 types), driver age (17-99), gas type (2 levels), and exposures in years (0.003-1.990).

In order to illustrate our methodology, we randomly sampled 1000 policies from the region R24 with density between 200-4500. This is the largest region in France that accounts for 39% of the business written. Miljkovic and Fernández (2018) used the policies from the same region (R24) to illustrate how the unobserved heterogeneity can be modeled in an insurance portfolio using two different mixture-based clustering approaches. The histogram of the number of claims in this region as well as the severity of the claims are shown in Figure 2. The frequency of claims in this regions is: 96.3% of zero claims, 3.5% of single claims, and 0.2% of two claims. Figure 2 also shows the density of the severity of claims in R24. Minimum claim amount in this region is 2 while maximum amount is 2,036, 833 Euros. Skewness coefficient of the claim severity data is 75.12.

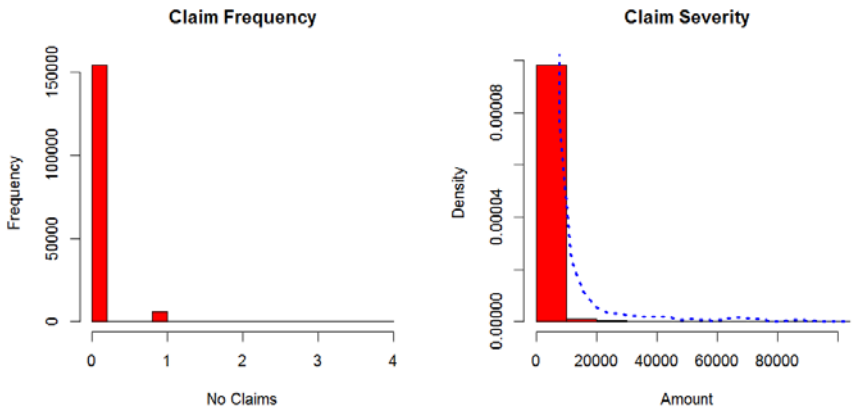


Figure 2: Frequency of claims (left) and severity of claims (right) in R24.

Our random stratified sample of 1000 policies maintains the same characteristics of R24 based on

the number of policies, gas type, density, and driver's age. Gas type has two levels, diesel and regular, with regular treated as a base level. Driver age is grouped at five levels: (17-20] (base level), (20-26], (26-42], (42-74], and 74+. Density is treated as a continuous predictor. The same variables have been used by Joan-Philippe and Arthur Charpentier (Charpentier, 2014) when modeling the same data set using Poisson and Negative Binomial regression. In the analysis of frequency or severity, we standardized the density variable since it is fairly big in numerical values and results in a numerical problem in model fitting if we use the raw measurement. The new standardized density measurements are the raw density measurements subtracted by the mean density and then divided by the standard deviation of density measurements. We used "PROC STANDARD" in SAS to assist the standardization of this variable.

Both the frequentist and Bayesian analysis are implemented with SAS with the SAS code included in Appendix B. The frequentist Poisson regression model fit was obtained via "PROC GENMOD," while the Bayesian model fit of claims frequencies was obtained via "PROC MCMC." In the analysis with the Bayesian Poisson regression model assuming weakly informative priors or power priors, 20,000 samples of parameters are simulated from the posterior distributions using Markov Chain Monte Carlo (MCMC) algorithm, which are obtained from 650,000 MCMC iterations with the first 150,000 cycles as burn-in iterations and a thinning rate of 10 (i.e., every 10th draw from the MCMC simulation is used to compute credible sets and medians of the posterior distribution).

The frequentist and Bayesian gamma regression fit was obtained via "PROC GENMOD" while the Bayesian analysis utilized the "Bayes" statement provided in "PROC GENMOD." In the Bayesian analysis of claim severity, 10,000 posterior samples are obtained from 12,000 MCMC iterations with the first 2,000 cycles as burn-in iterations and a thinning rate of 1 (i.e., no thinning was used here). Convergence of the posterior simulation was evaluated using history plots and autocorrelation (ACF) plots of the posterior samples. Figures of the posterior sample of regression coefficients are shown in the Appendix A (Figures 4-6). All of the history plots show that the posterior simulation achieved convergence, while the ACF plots show that the (thinned) posterior samples do not have strong autocorrelation.

Table-1 in the Appendix A shows the comparison of the results of the Poisson GLM regression model that has been run using a frequentist approach and a Bayesian hierarchical modeling approach with weakly informative priors and power priors. For each of these three methods we show the

coefficient estimates with their standard errors and the 95% confidence intervals. In the Poisson model, all of the coefficient for driver age are negative relative to the base group (17, 20] with the largest coefficient reported for age group (26, 42]. Thus, this age group reports on average the lowest frequency of claims relative to age group (17, 20]. These results are in line with other studies showing that young drivers (17, 20] are most likely to get into car accidents. The coefficient for regular gas type is negative relative to diesel gas type. The coefficient for population density in R24 is positive, indicating that an increase in population density results in additional claims reported.

From Gamma regression model, we observe that the coefficients for age group (20, 26] and (26, 42] are negative relative to the age group (17, 20] indicating that severity of claims for these groups is lower compared to group (17, 20]. The coefficients for age groups (42, 74] and (74+) are positive relative to age group (17, 20]. Also coefficient for density variable is positive indicating that the severity of the claims will increase on average as the population density increases.

Since the power prior is expressed as a product of the weighted likelihood of parameters, conditional on the historical information and a prior distribution of the parameters before the data are observed, a scale or discounting parameter from 0 to 1 is used to control the weight assigned to historical data. This parameter is usually controlled by user. Our Bayesian model with power priors assumes that 50% of external information is incorporated in the posterior distribution in the form of a prior input consisting of 50% of the log-likelihood of these external territory observations; thus, the scale parameter is 0.5.

We observe that standard errors of the posterior estimates are smaller compared to those generated with the ordinary GLM. As a result, the 95% confidence intervals are narrower than those produced with ordinary GLM or the Bayesian GLM with non-informative priors. Poisson regression results arrive at the same conclusion in terms of the risk associated with all age groups compared to age group 17-20. However, the smaller confidence intervals indicate the improvement in the estimation of the likelihood by using past information. Power priors allow for a different percent of external information to be used, which allows an actuary to judgmentally incorporate this aspect of modeling into the analysis. Another sample of 1000 losses was selected out of 16,181

policies that reported positive claim amounts. Table 2 shows the comparison of the results of Bayesian gamma regression with non-informative priors to those produced using the frequentist approach. We can also observe that standard errors and the 95% confidence intervals related to the regression coefficients are smaller compared to those produced using the frequentist approach.

4. MODEL VALIDATION

Model validation is an important part of model building. When two competing models are evaluated, common techniques such as Receiver Operating Characteristic (ROC) Curves or Double Lift Charts can be used. These techniques are appropriate, e.g logistic regression models, and require that a database of historical observations is augmented with the predictions from each of the competing models (Goldburd at al., 2016). Considering the nature of our application, the historical database is not available in a new territory where the company starts writing new business for the first time, or to an existing territory where the new business was recently introduced, so the claims experience is very limited. In absence of the historical database, we borrowed the information from the "imaginary" adjacent territory that we assumed to be R24.

Our validation is based on the "splitting data" approach and it is shown in the flowchart in Figure 3. This approach assumes drawing three samples from R24:

1) Training Set - used to perform the model building,
2) Holdout Set (Test Set) - used to perform data validation, and
3) The Bayesian "External Prior" Set - used to provide prior input information.

Both the Training Set and the "External Prior" Set consist of 1000 observations, while the Test Set consists of 100 observations. The comparison was done to evaluate the impact of incorporating the information from existing external territories on the Bayesian analysis of the Training Set. Table 3 in Appendix A summarizes the results of this validation. The Bayesian analysis with weakly information priors was applied to fit the Training Set and the predicted numbers of claims for the Test Set observations were obtained based on the corresponding posterior predictive distributions. Then we also fit the Bayesian analysis with power prior information from the External Prior Set to the Training Set and obtained the predicted number of claims for the Test Set using the new posterior prediction distributions. The two sets of predicted number of claims are both compared

with the original observed frequencies for the Test Set and MSPEs were computed: 0.51 for the Bayesian analysis with weakly information priors and 0.49 for the Bayesian analysis with power priors. The MSPE calculation includes three values based on frequency of claims shown in Figure 2.

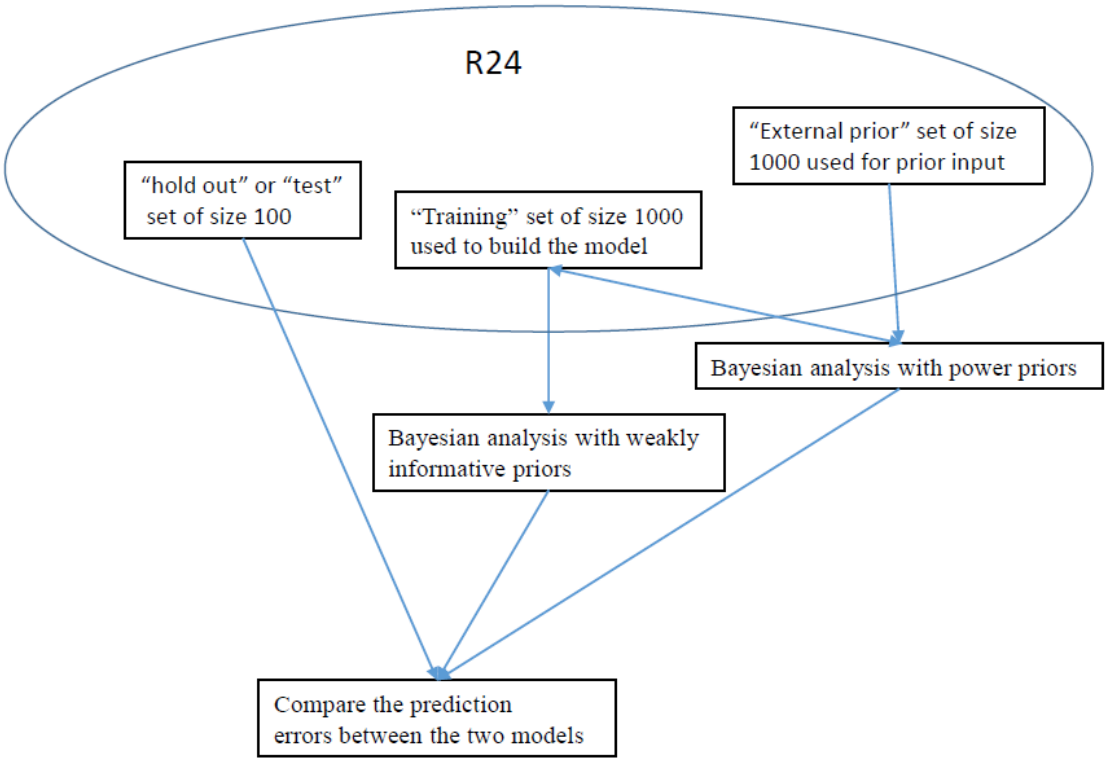


Figure 3: Flow chart of the validation process.

We also fit frequentist Poisson regression on all three samples respectively to check the similarity of the training data, validation data, and external prior information. This validation analysis indicates better prediction performance when power priors are used. However, the strength of improvement when the power priors are used in the Bayesian, is subjected to two critical factors:

(1) Sample size of the "current" data (i.e. Training Set in this validation analysis). When sample size is fairly high relative to the complexity of the model fit, the information borrowing through power priors would play a minor role in the model prediction performance.

(2) Similarity between the External Prior Set and "current" data. When the information borrowed through power priors are "misleading", it is not favorable to incorporate those in the analysis.

5. CONCLUSION

Historical claims information is available to the actuary for the purpose of ratemaking in those territories where the company has been writing business for some period of time, even when no information is available in the case of a new territory where a company is entering business for the first time. The understanding of the "new territory" can almost always be augmented by existing information. Such borrowing of strength from historical data has long been encouraged in many scientific fields. These issues motivated us to investigate the feasibility of a Bayesian power prior approach in borrowing of strength for modeling auto rates in a new territory. The goal of such an approach is to determine a practical amount of strength to borrow from the historical claims that strikes a balance between increased cost-efficiency and long-run statistical integrity. The methods for incorporating historical data should be robust to prior knowledge and consistent with the accumulating historical information. We aim to utilize historical information given strong evidence that this information would apply well in a territory that shares some common characteristics. A more attractive feature of such "information borrowing" is that the practitioners can pick multiple values of the scale or discounting parameter to compare the analysis outcomes reflecting different prior beliefs of the "similarity" between the historical data and current data.

In this paper, we showed how the Bayesian analysis with power priors and non-informative priors can be used in modeling auto claims frequency for a new territory. By borrowing prior information from an existing territory that shares some similar characteristics such as climate, population demographics, or geography, we can develop a claim frequency model. Modeling claim severity with Bayesian GLM is also shown. We illustrated our approach on a small data set drawn from the motor third-party liability data set provided by a French insurance company. An immediate future work we would like to pursue is the joint Bayesian analysis of frequency and severity of claims. The joint analysis would allow us to borrow information between the two numerical features of the "new territory" and, hence, improve the analysis with a limited amount of information. The validation of our approach was also provided. We believe that our attempt to introduce Bayesian analysis with power priors will

benefit many insurance companies as they enhance their current GLM pricing model and apply it in ratemaking of a new territory or an existing territory where the claims' experience is limited.

Acknowledgment

The authors are grateful for the comments and suggestions received by two anonymous Reviewers. We also appreciate the discussion with Professor John Zicarellie from Arizona State University whose suggestions further improved the clarity of this paper.

Supplementary Material

The SAS Code is available in Appendix B.

REFERENCES

[1] CASdatasets. URL: http://127.0.0.1:17326/library/CASdatasets/html/overview.html
[2] Charpentier, A. (2014). Computational Actuarial Science with R. CRC Press.
[3] Chen, M.H. and Ibrahim, J.G. (2006). "The relationship between the power prior and hierarchical models," *Bayesian Analysis* ,Vol 1, 551-574.
[4] Bermúdez, L. and Karlis, D. (2011). "Bayesian multivariate Poisson models for insurance ratemaking," *Insurance: Mathematics and Economics*, Vol 48, No. 2, 226-236.
[5] Boucher, J.P., Denuit, M. and Guillen, M. (2008). "Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions," *Variance*, Vol 2, No 1, 135-162.
[6] Boucher, J.P., Denuit, M. and Guillen, M. (2009). "Number of Accidents or Number of Claims? An Approach with Zero-Inflated Poisson Models for Panel Data," *Journal of Risk and Insurance*, Vol *76, No* 4, 821-846.
[7] Brown, G.O. and Buckley, W.S. (2015). "Experience rating with Poisson mixtures," *Annals of Actuarial Science*, Vol 9, No 2, 304-321.
[8] Goldburd, M., Khare, A., and Tevet, D. (2016). "Generalized Linear Models for Insurance Rating", Casualty Actuarial Society, CAS Monographs Series, No 5.
[9] Ibrahim, J.G., and Chen, M.-H. (2000). "Power prior distributions for regression models," *Statistical Science*, Vol 15, 46-60.
[10] Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2008). *Modern actuarial risk theory: using R.* Springer Science & Business Media.
[11] McCullagh, P., and J. A. Nelder, *Generalized Linear Models* (2nd edition), London: Chapman and Hall, 1989.
[12] Miljkovic, T. (2017). *Computational Actuarial Science with R.* Journal of Risk and Insurance 84(1): 267.
[13] Miljkovic T. and Fernández, D. (2018). *On Two Mixture-Based Clustering Approaches Used in Modeling an Insurance Portfolio.* Risks 6 (2). DOI: 10.3390/risks6020057.

Abbreviations and notations

GLM, generalized linear models

ROC, Receiver Operating Characteristic

REML, Restricted Maximum Likelihood

Biographies of the Authors

Dr. Jing Zhang is an Associate Professor in the Department of Statistics at Miami University. Her research focus is in Bayesian statistics, spatial analysis, statistical modeling for environmental and biological studies. Zhang holds an MBA and PhD in Statistics from University of Missouri.

Dr. Tatjana Miljkovic is an Assistant Professor and the Actuarial Science Adviser in the Department of Statistics at Miami University. Her research focus is in actuarial science and applied statistics areas. Prior to earning her PhD in statistics, she was employed by Unum Life Insurance Company (Corporate Actuarial), American National Property and Casualty Company (Corporate Actuarial), and Risk Management Solutions (Model Development). Miljkovic holds an MS Degree in Actuarial Science from University of Illinois, an MBA and PhD in Statistics from North Dakota State University.

Appendix A

Method	Frequentist Approach to Poisson Regression				Bayesian Poisson Regression with Weakly Informative Priors				Bayesian Poisson Regression with Power Priors			
Parameter	Estimate	SE	95% CI		Estimate	SE	95% CI		Estimate	SE	95% CI	
Intercept	-0.4754	0.2468	-0.9590	0.0083	0.8330	0.2513	0.3292	1.3164	0.7670	0.2027	0.3531	1.1498
GasRegular	-0.2582	0.1038	-0.4617	-0.0547	-0.2588	0.1047	-0.4676	-0.0553	-0.2329	0.0839	-0.3954	-0.0661
DriverAge(20, 26]	-0.3149	0.2502	-0.8052	0.1754	-0.4907	0.2519	-0.9811	0.0053	-0.4339	0.2044	-0.8302	-0.0287
DriverAge(26, 42]	-1.0424	0.2070	-1.4481	-0.6368	-1.1942	0.2091	-1.5883	-0.7662	-1.1039	0.1711	-1.4212	-0.7554
DriverAge(42, 74]	-0.8483	0.1946	-1.2297	-0.4670	-1.0565	0.1964	-1.4231	-0.6553	-1.0441	0.1613	-1.3456	0.7111
DriverAge(74, Inf]	-0.6587	0.2609	-1.1701	-0.1473	-0.9487	0.2635	-1.4637	-0.4322	-0.9194	0.2125	-1.3389	-0.5038
Density	0.1896	0.0393	0.1126	0.2667	0.1975	0.0392	0.1179	0.2718	0.1958	0.0328	0.1301	0.2579

Table-1: Comparisons of the Regression Results for Poisson Model

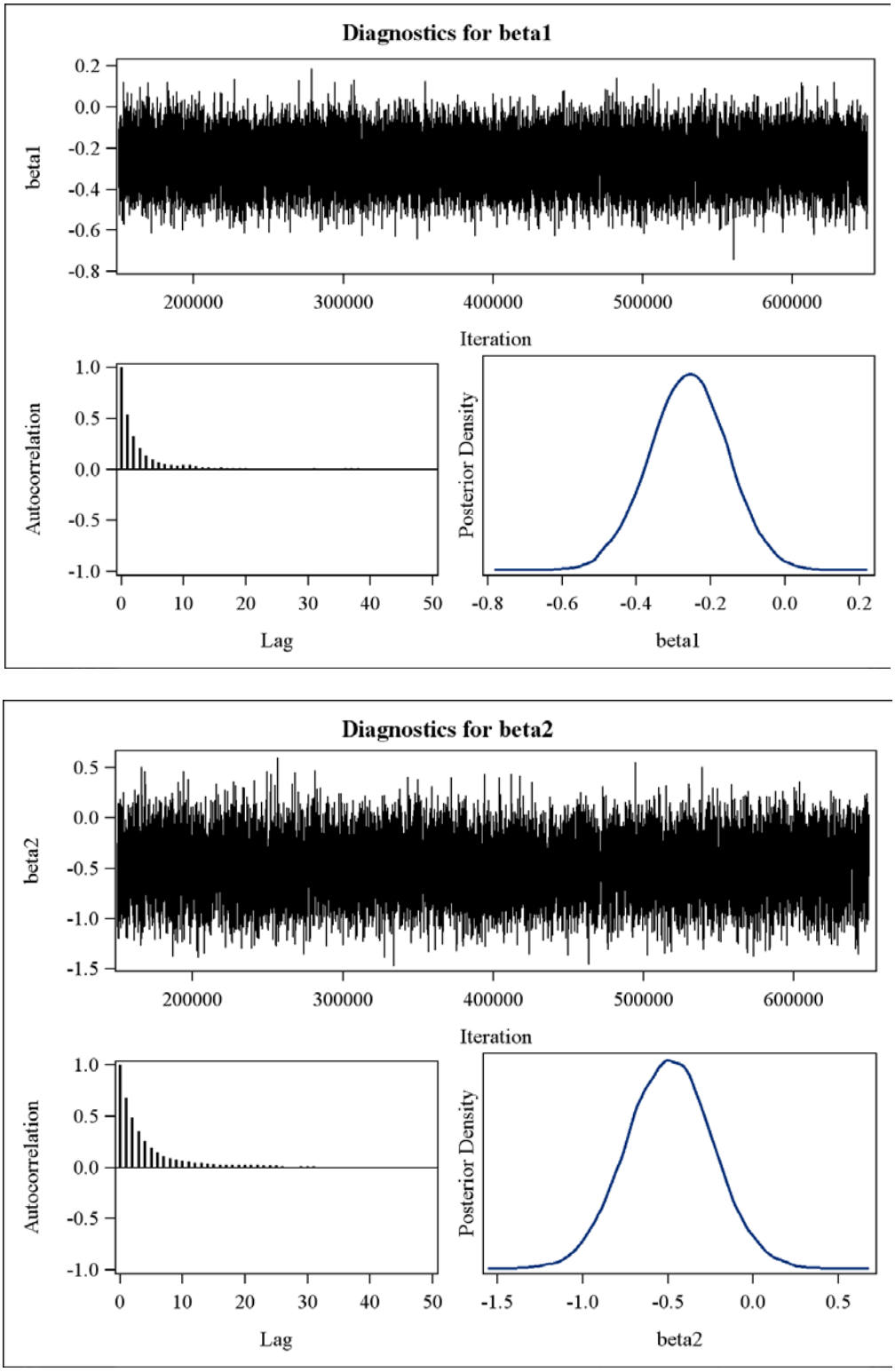
Table-2: Comparisons of the Regression Results for Gamma Model

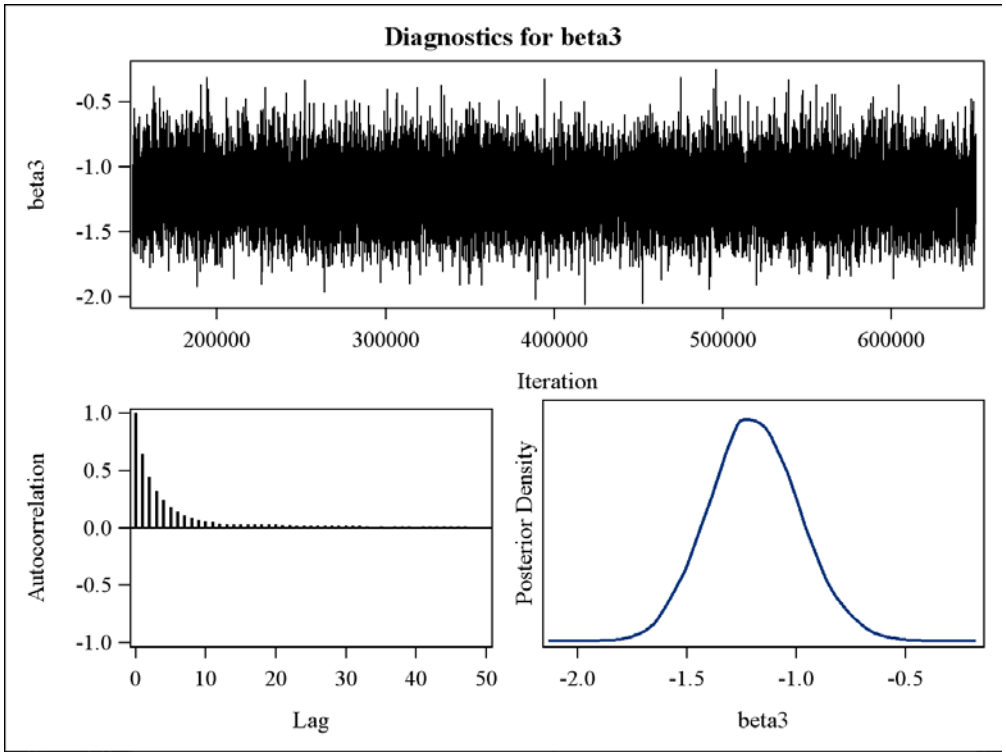
Method	Frequentist Approach to Gamma Regression				Bayesian Gamma Regression with Non-informative Priors			
Parameter	Estimate	SE	95% CI		Estimate	SE	95% CI	
Intercept	7.7746	0.1728	7.4359	8.1133	7.7869	0.1690	7.4650	8.1261
GasRegular	0.1125	0.0819	-0.0481	0.2730	0.1109	0.0818	-0.0513	0.2651
DriverAge(20,26]	-0.6454	0.2328	-1.1017	-0.1892	-0.6423	0.2320	-1.1074	-0.2037
DriverAge(26,42]	-0.4135	0.1773	-0.7611	-0.0659	-0.4218	0.1751	-0.7695	-0.0900
DriverAge(42,74]	0.2735	0.1727	-0.0650	0.6120	0.2652	0.1700	-0.0882	0.5741
DriverAge(74,Inf]	0.2555	0.2302	-0.1957	0.7067	0.2587	0.2308	-0.1743	0.7226
Density	0.0453	0.0426	-0.0382	0.1287	0.0474	0.0430	-0.0365	0.1327

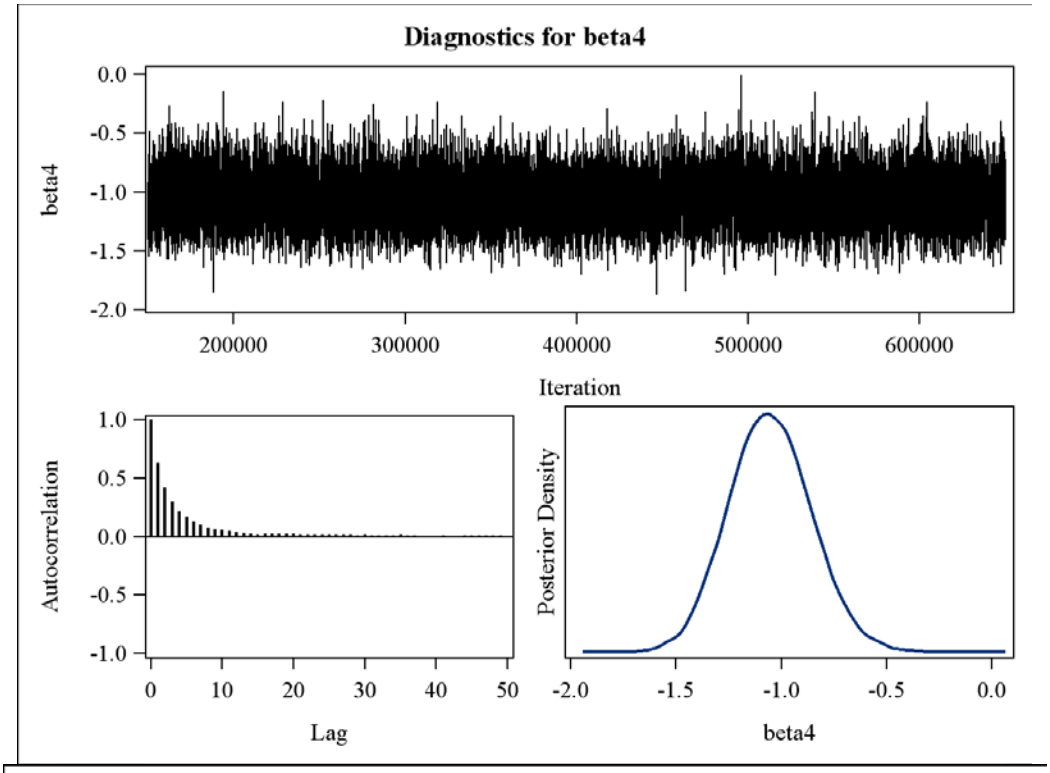
Table-3: Validation analysis

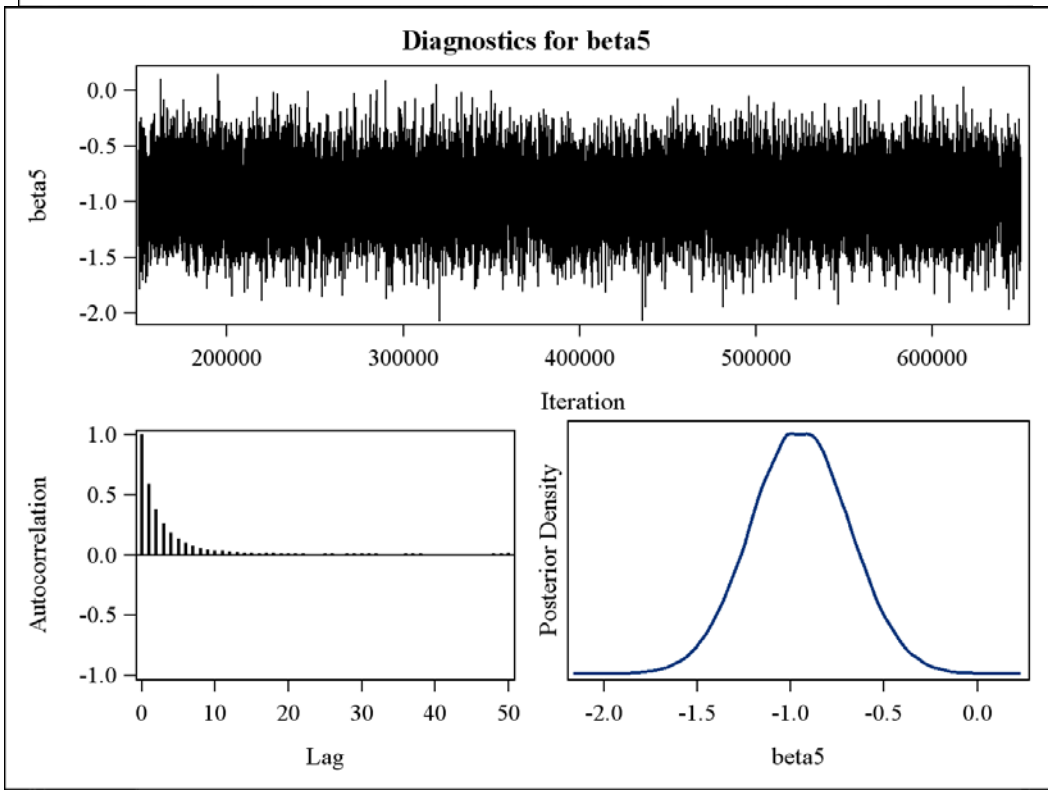
Sample Type	Training Set				Holdout Set				External Prior Set			
Parameter	Estimate	SE	95% CI		Estimate	SE	95% CI		Estimate	SE	95% CI	
Intercept	-0.4754	0.2468	-0.9590	0.0083	0.1793	0.8906	-1.5662	1.9248	-0.6566	0.2423	-1.1315	-0.1817
GasRegular	-0.2582	0.1038	-0.4617	-0.0547	-0.6379	0.3898	-1.4020	0.1261	-0.1784	0.1025	-0.3793	0.0225
DriverAge(20, 26]	-0.3149	0.2502	-0.8052	0.1754	-0.6266	0.8770	-2.3454	1.0923	-0.2215	0.2427	-0.6972	0.2542
DriverAge(26, 42]	-1.0424	0.2070	-1.4481	-0.6368	-1.2490	0.7849	-2.7875	0.2894	-0.8161	0.2024	-1.2129	-0.4193
DriverAge(42, 74]	-0.8483	0.1946	-1.2297	-0.4670	-0.8827	0.7406	-2.3343	0.5689	-0.8489	0.1955	-1.2321	-0.4658
DriverAge(74, Inf]	-0.6587	0.2609	-1.1701	-0.1473	-1.8648	1.2715	-4.3569	0.6272	-0.6206	0.2609	-1.1318	-0.1093
Density	0.1896	0.0393	0.1126	0.2667	0.2220	0.1620	-0.0955	0.5395	0.1744	0.0422	0.0916	0.2571

Figure 4. History plots, ACF plots and density curves of the model parameters in the Bayesian Poisson regression model with weakly informative priors.




Diagnostics for beta3
beta3
-0.5
-1.0
-1.5
-2.0
200000
300000
400000
500000
600000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
-2.0
-1.5
-1.0
-0.5
beta3


Diagnostics for beta4
beta4
0.0
-0.5
-1.0
-1.5
-2.0
200000
300000
400000
500000
600000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
-2.0
-1.5
-1.0
-0.5
0.0
beta4


Diagnostics for beta5
beta5
0.0
-0.5
-1.0
-1.5
-2.0
200000
300000
400000
500000
600000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
-2.0
-1.5
-1.0
-0.5
0.0
beta5

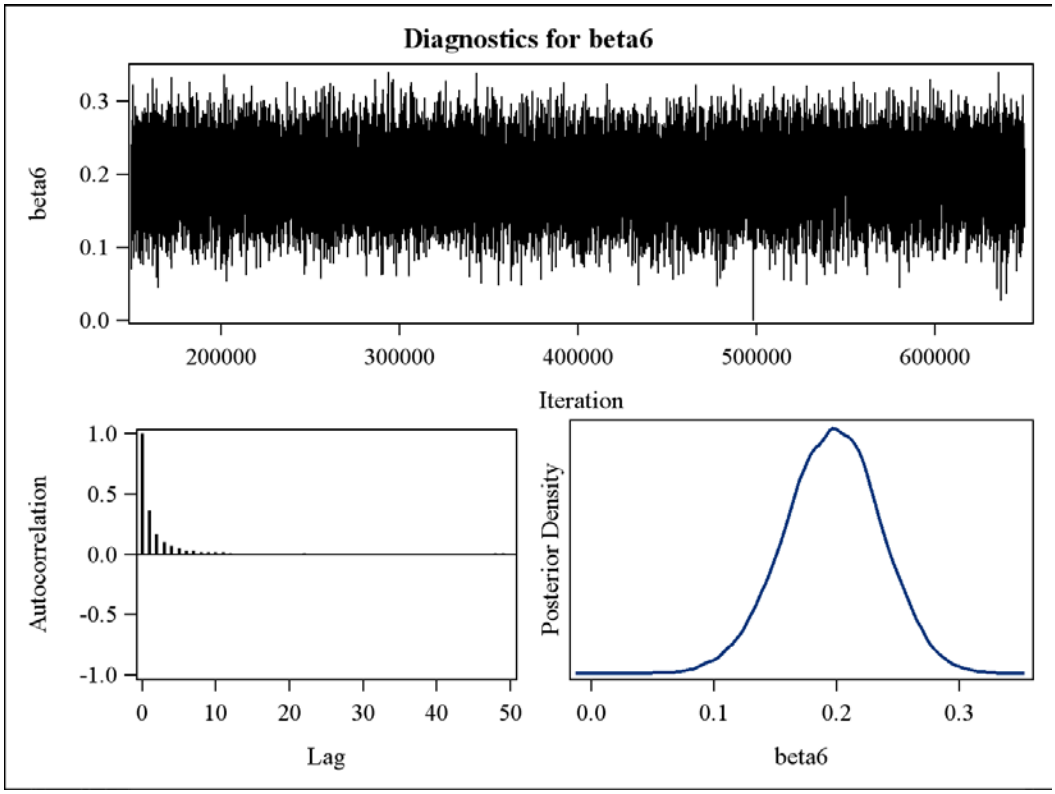
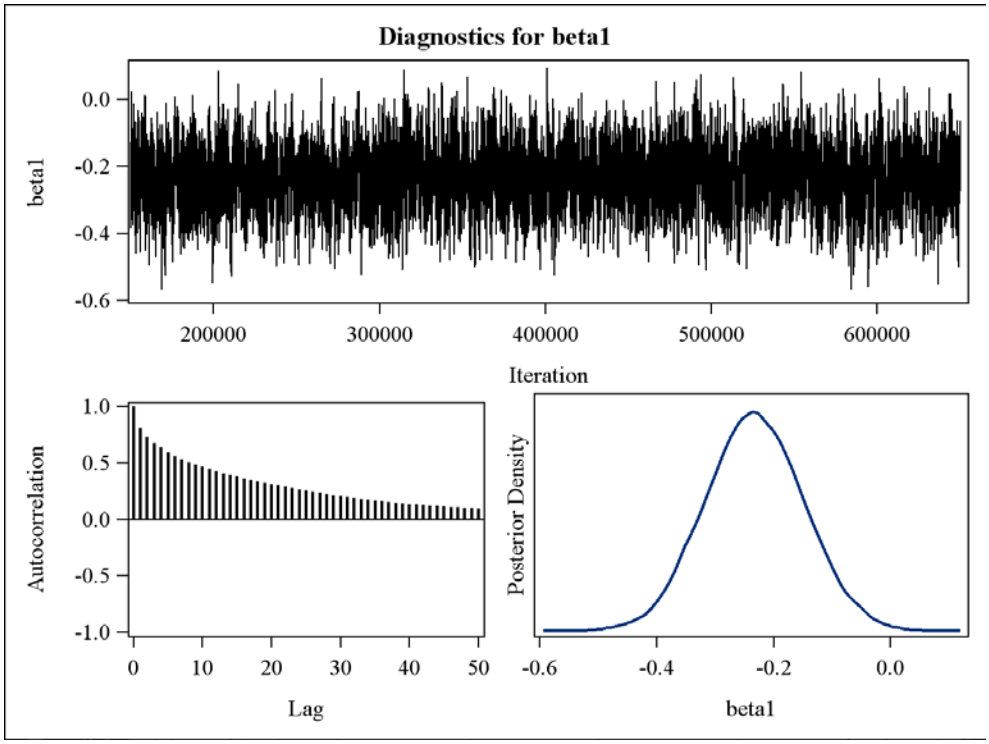
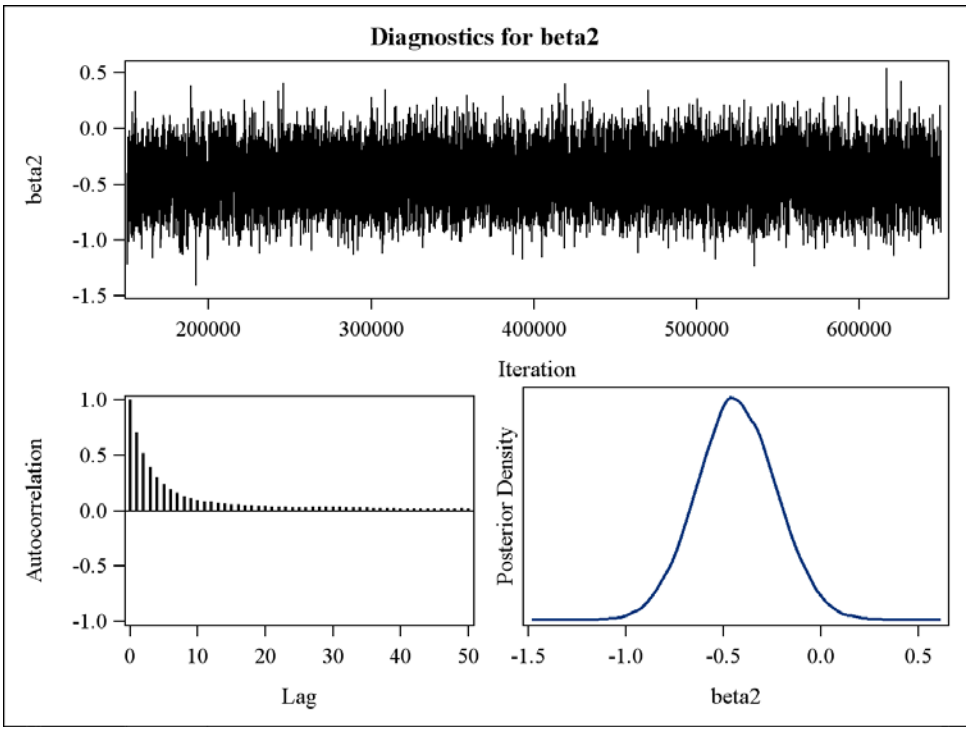
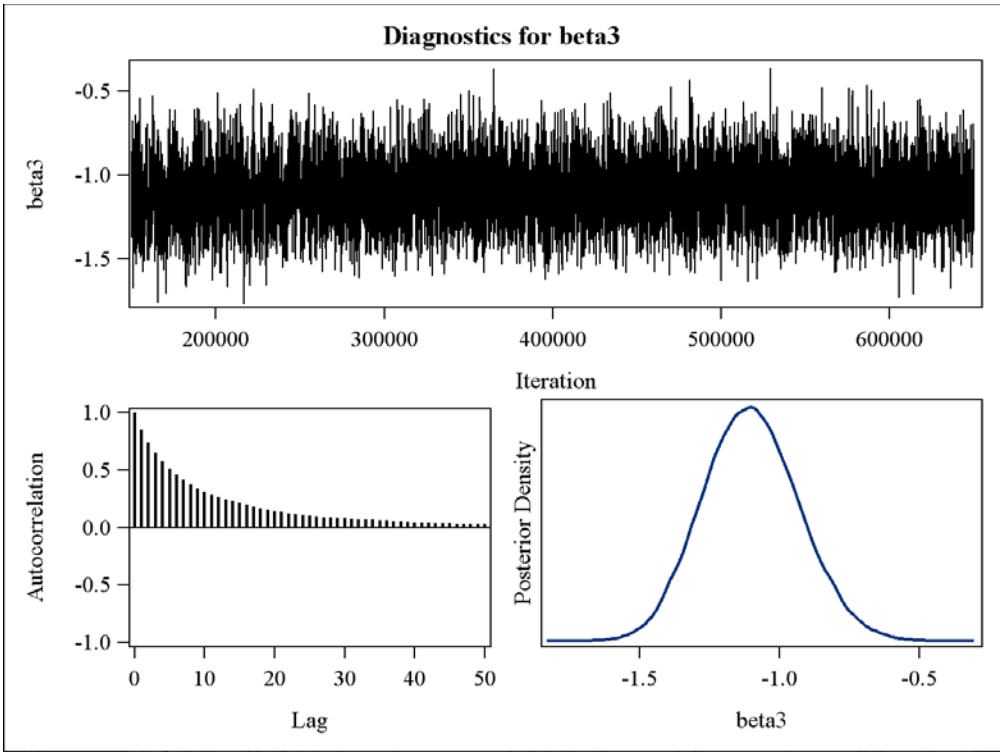
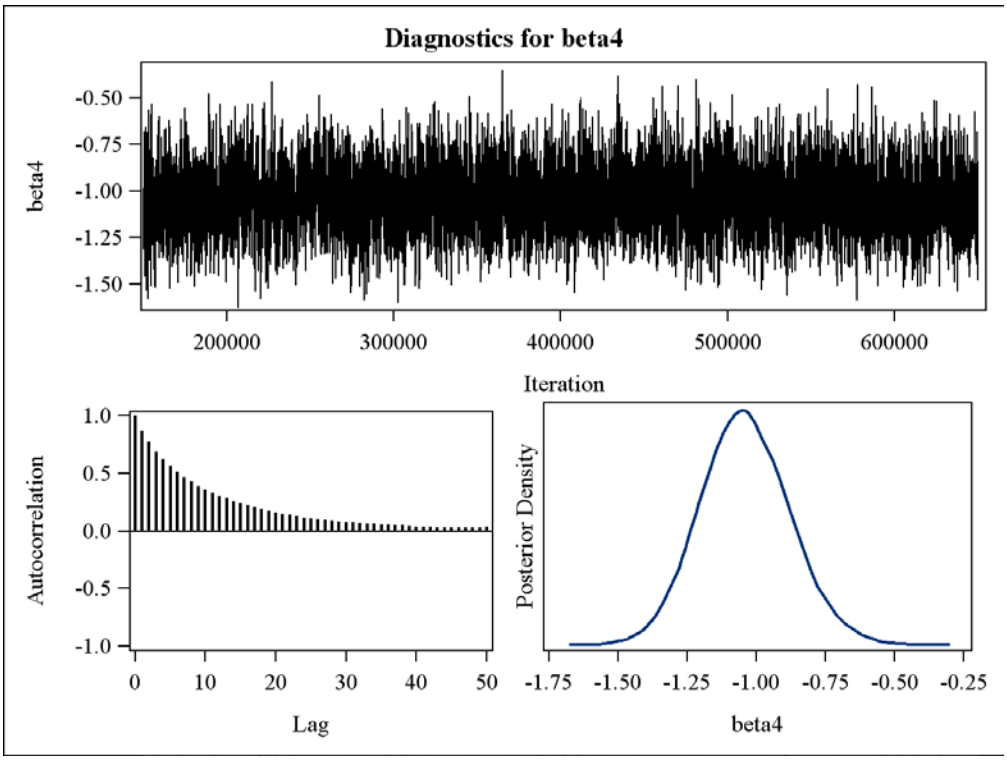

Diagnostics for beta6
beta6
0.3
0.2
0.1
0.0
200000
300000
400000
500000
600000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0.0
0.1
0.2
0.3
beta6

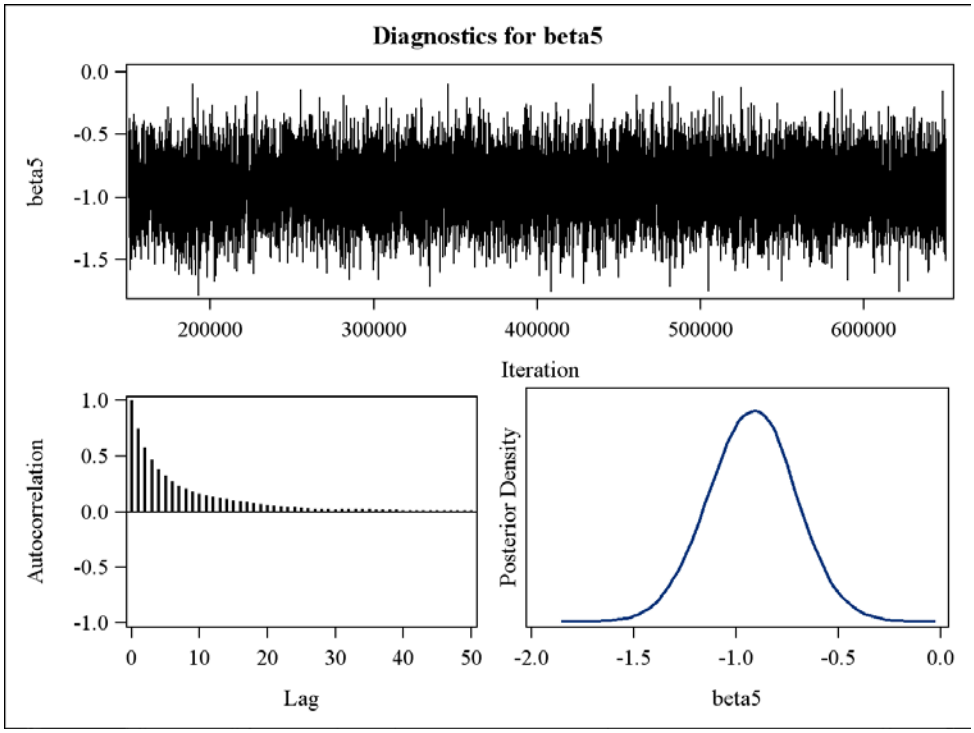
Figure 5. History plots, ACF plots and density curves of the model parameters in the Bayesian Poisson regression model with power priors.






Diagnostics for beta3
beta3
-0.5
-1.0
-1.5
200000
300000
400000
500000
600000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
-1.5
-1.0
-0.5
beta3


Diagnostics for beta4
beta4
-0.50
-0.75
-1.00
-1.25
-1.50
200000
300000
400000
500000
600000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
-1.75
-1.50
-1.25
-1.00
-0.75
-0.50
-0.25
beta4


Diagnostics for beta5
beta5
0.0
-0.5
-1.0
-1.5
200000
300000
400000
500000
600000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
-2.0
-1.5
-1.0
-0.5
0.0
beta5

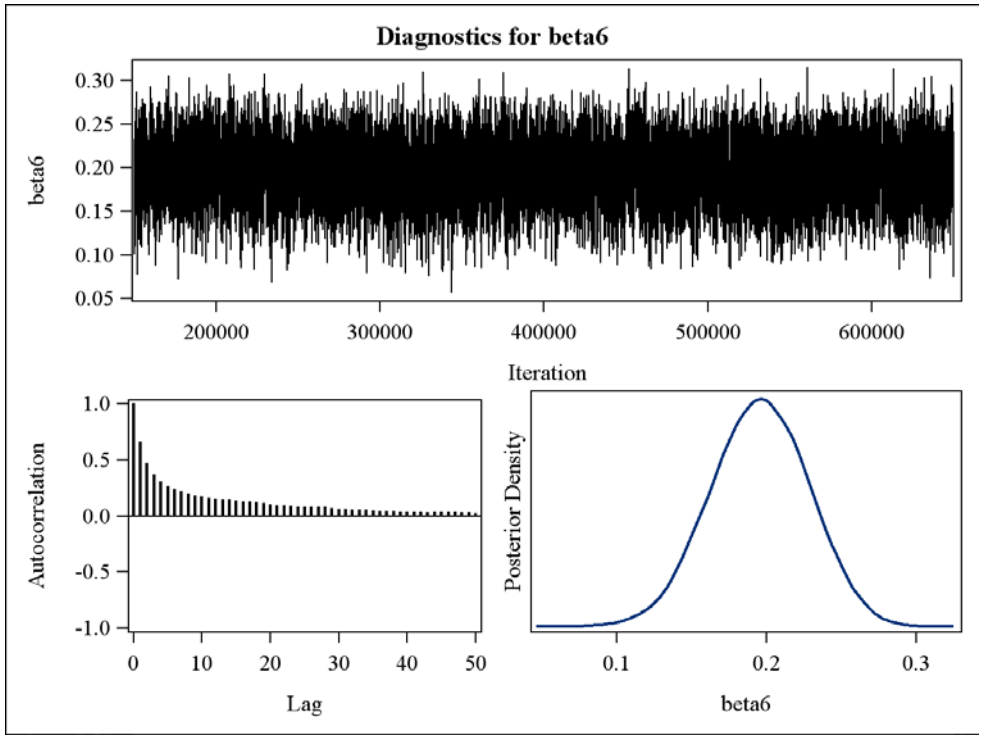
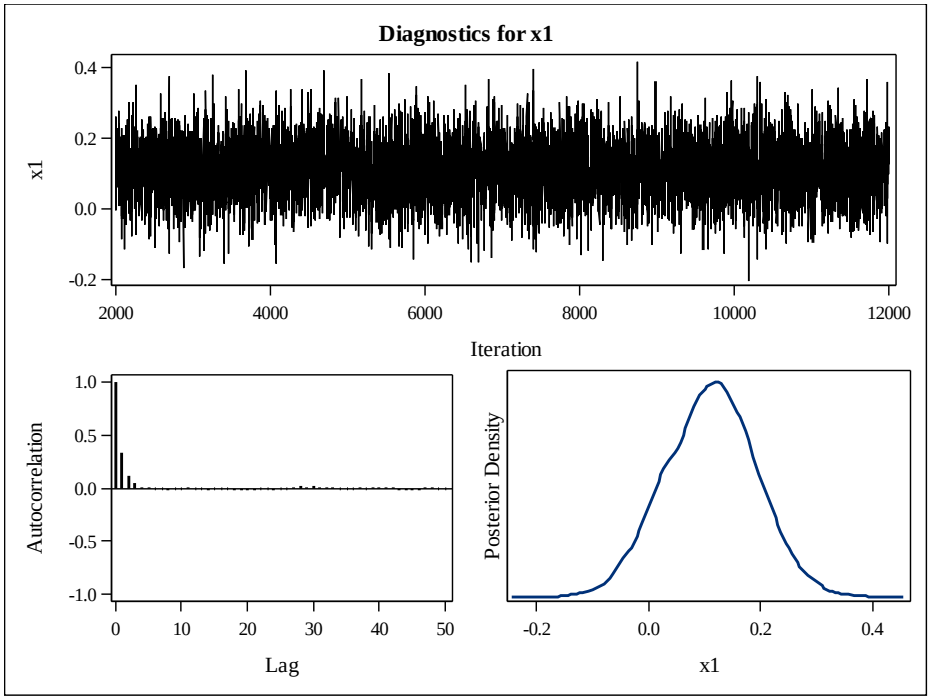
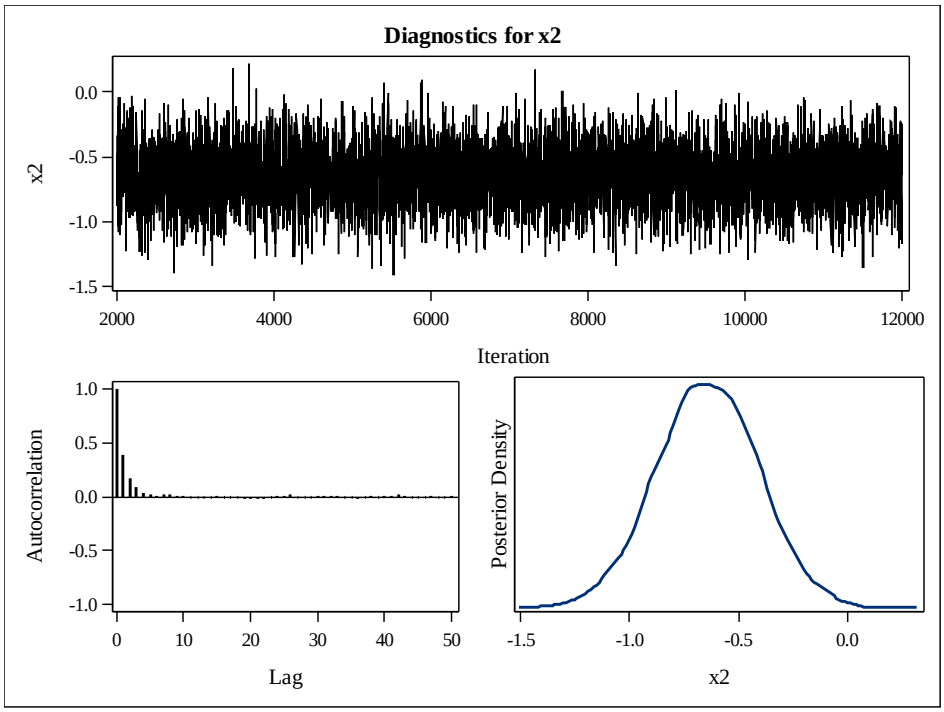
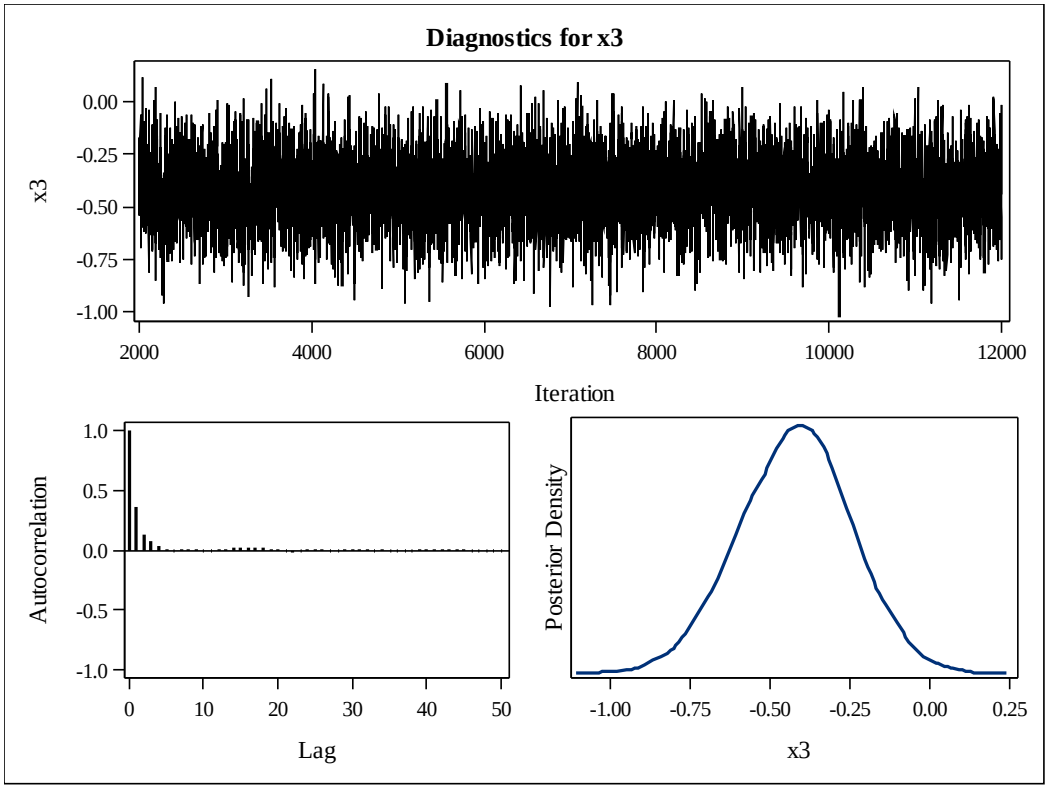
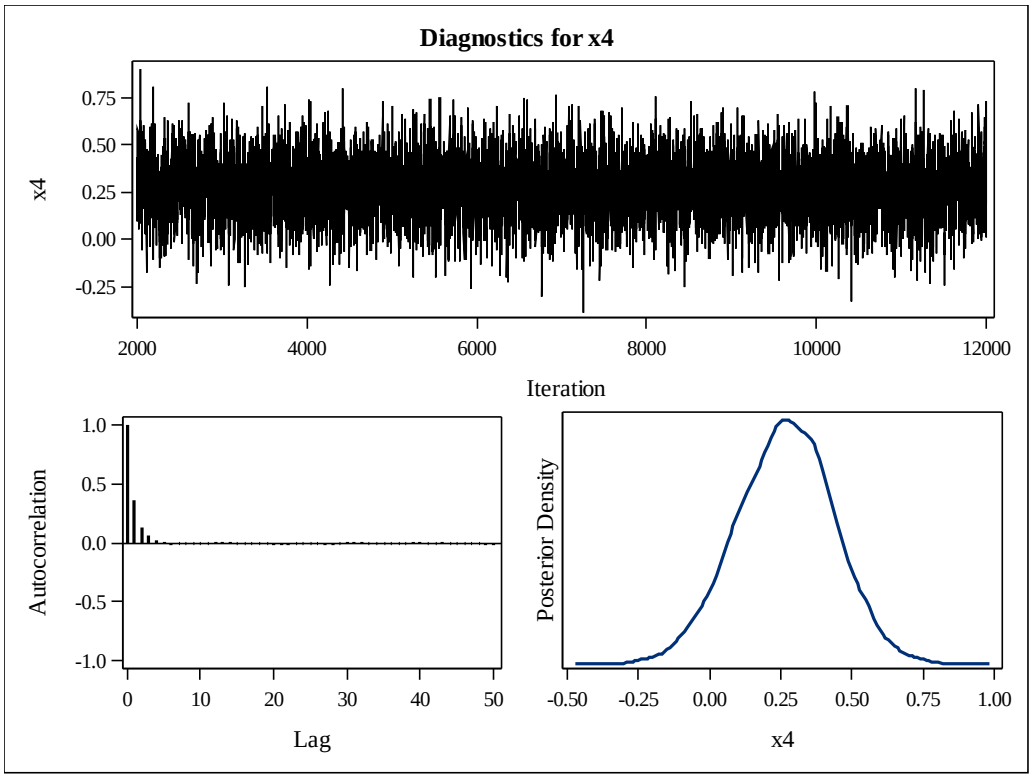

Diagnostics for beta6
beta6
0.30
0.25
0.20
0.15
0.10
0.05
200000
300000
400000
500000
600000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0.1
0.2
0.3
beta6

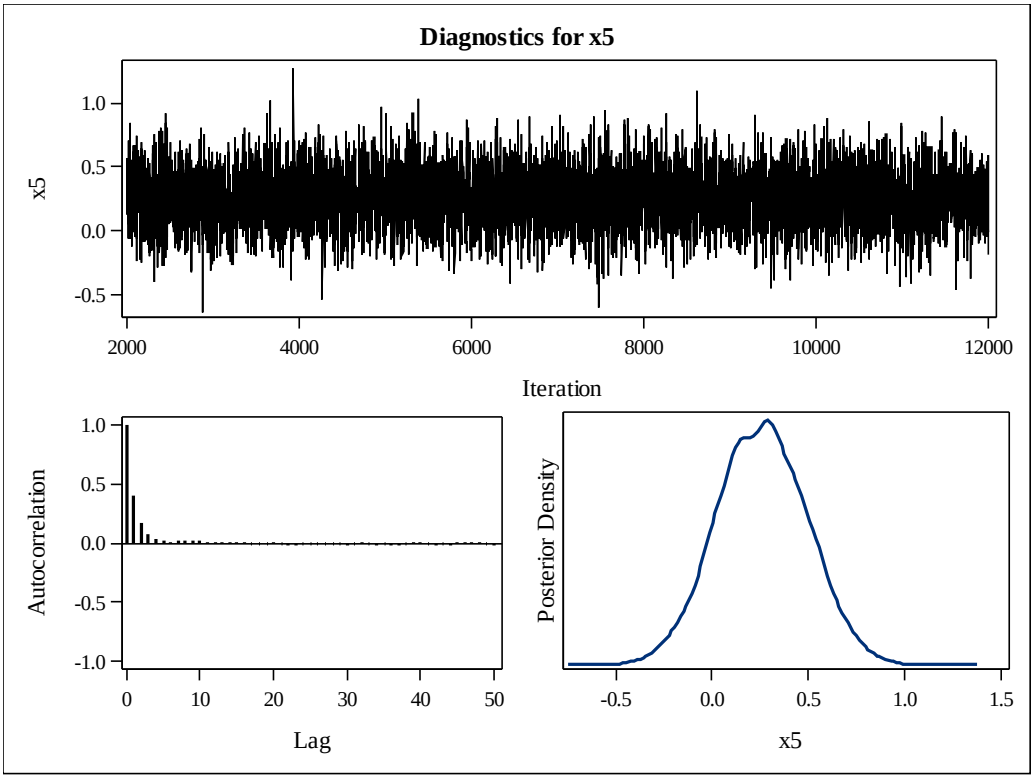
Figure 6. History plots, ACF plots and density curves of the model parameters in the Bayesian gamma regression model.

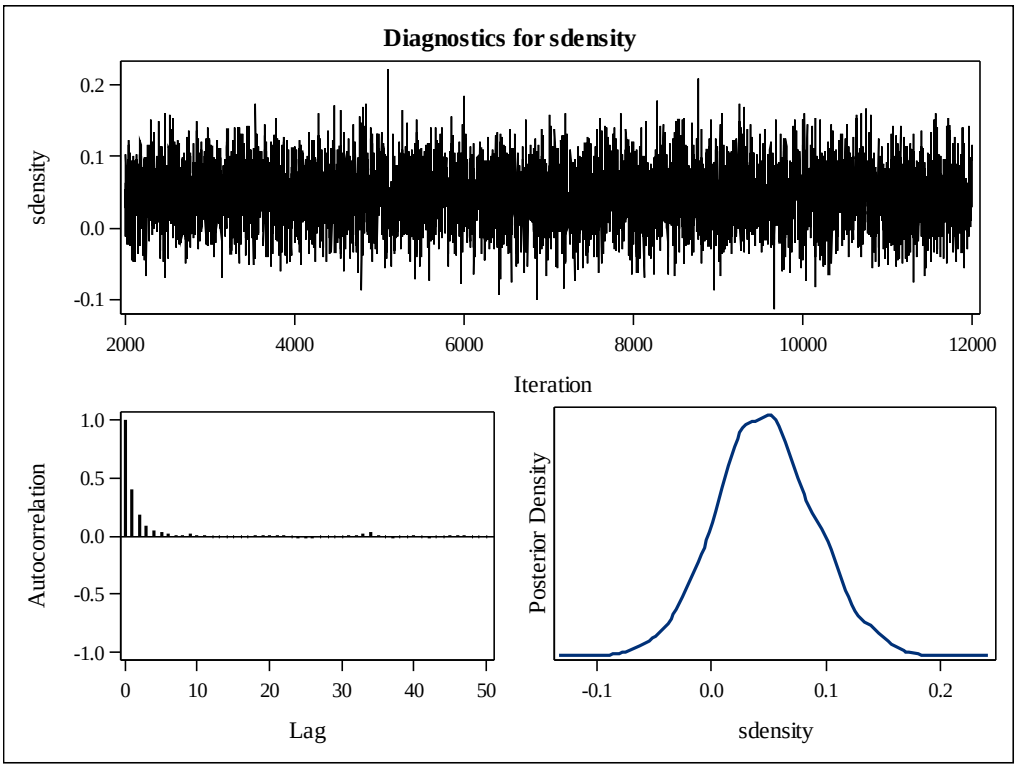





Diagnostics for x3
x3
0.00
-0.25
-0.50
-0.75
-1.00
2000
4000
6000
8000
10000
12000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
-1.00
-0.75
-0.50
-0.25
0.00
0.25
x3


Diagnostics for x4
x4
0.75
0.50
0.25
0.00
-0.25
2000
4000
6000
8000
10000
12000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
-0.50
-0.25
0.00
0.25
0.50
0.75
1.00
x4


Diagnostics for x5
x5
1.0
0.5
0.0
-0.5
2000
4000
6000
8000
10000
12000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
-0.5
0.0
0.5
1.0
1.5
x5


Diagnostics for sdensity
sdensity
0.2
0.1
0.0
-0.1
2000
4000
6000
8000
10000
12000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
-0.1
0.0
0.1
0.2
sdensity

Appendix B

```
SAS code for the analysis.
Part 1:
/*Poisson regression with weakly-informative priors*/
/*
ClaimNb: response variable, number of claims occurred during a
given time period in the region for a customer;
logoffset: logarithm of the exposure (duration of the policy),
offset variable we used in the Poisson regression;
Gas: indicator variable, Gas=1 if the car insured uses regular
gas;
x2: indicator variable, x2=1 if the insured driver is older than
22 but is 26 or younger;
x3: indicator variable, x3=1 if the insured driver is older than
26 but is 42 or younger;
x4: indicator variable, x4=1 if the insured driver is older than
42 but is 74 or younger;
x5: indicator variable, x5=1 if the insured driver is older than
74;
Density: population density of the region;
alpha: intercept;
beta1-beta6: regression coefficients associated with Gas, X2-X5
and Density;

Priors used for alpha: alpha ~ N(mua,sda^2), with higher level
priors mua~N(0, 10) and sda~Uniform(0,5).

Priors used for alpha and all the betai's: N(mubi,sdbi), i=1,...,6,
with higher level priors mubi~N(0,4) and sdbi~Uniform(0,1).
*/



proc mcmc data=datasetname seed=1181 nmc=500000 nbi=150000 thin=10
propcov=quanew monitor =(_parms_ ) outpost=out1000prior2;
ods select Parameters PostSummaries PostIntervals tadpanel;
parms alpha 0 beta1 0 beta2 0 beta3 0 beta4 0 beta5 0 beta6 0;
parms mua 0 mub1 0 mub2 0 mub3 0 mub4 0 mub5 0 mub6 0;
parms sda 0.5 sdb1 0.5 sdb2 0.5 sdb3 0.5 sdb4 0.5 sdb5 0.5 sdb6 0.5;
prior alpha ~ normal(mua, var=sda**2);
prior beta1 ~ normal(mub1, var=sdb1**2);
prior beta2 ~ normal(mub2, var=sdb2**2);
prior beta3 ~ normal(mub3, var=sdb3**2);
prior beta4 ~ normal(mub4, var=sdb4**2);
prior beta5 ~ normal(mub5, var=sdb5**2);
```

```
prior beta6 ~ normal(mub6, var=sdb6**2);
prior mua ~ normal(0, var=10);
prior mub: ~ normal(0, var=4);
prior sda ~ uniform(0,5);
prior sdb: ~ uniform(0,1);

mu      =      exp(logoffset      +      alpha      +      beta1*Gas      +
beta2*x2+beta3*x3+beta4*x4+beta5*x5+beta6*Density);
model ClaimNb ~ poisson(mu);
run;

Part 2:

/*Poisson regression with Power prior, a0 fixed*/

/*
ClaimNb: response variable, number of claims occurred during a
given time period in the region for a customer;
logoffset: logarithm of the exposure (duration of the policy),
offset variable we used in the Poisson regression;
Gas: indicator variable, Gas=1 if the car insured uses regular
gas;
x2: indicator variable, x2=1 if the insured driver is older than
22 but is 26 or younger;
x3: indicator variable, x3=1 if the insured driver is older than
26 but is 42 or younger;
x4: indicator variable, x4=1 if the insured driver is older than
42 but is 74 or younger;
x5: indicator variable, x5=1 if the insured driver is older than
74;
Density: normalized population density of the region;

alpha: intercept;
beta1-beta6: regression coefficients associated with Gas, X2-X5
and Density;

Initial Priors used for alpha: alpha ~ N(mua,sda^2), with higher
level priors mua~N(0, 10) and sda~Uniform(0,5).

Initial Priors used for alpha and all the betai's: N(mubi,sdbi),
i=1,...,6,   with   higher   level   priors   mubi~N(0,4)   and
sdbi~Uniform(0,1).

Power prior is used here with fixed power a0=0.5.
*/
proc mcmc data=datasetname seed=1181 nmc=500000 nbi=150000 thin=10
```

```
propcov=quanew monitor =(_parms_ ) outpost=out1000power50;
ods select Parameters PostSummaries PostIntervals tadpanel;
parms alpha 0 beta1 0 beta2 0 beta3 0 beta4 0 beta5 0 beta6 0;
parms mua 0 mub1 0 mub2 0 mub3 0 mub4 0 mub5 0 mub6 0;
parms sda 0.5 sdb1 0.5 sdb2 0.5 sdb3 0.5 sdb4 0.5 sdb5 0.5 sdb6 0.5;
prior alpha ~ normal(mua, var=sda**2);
prior beta1 ~ normal(mub1, var=sdb1**2);
prior beta2 ~ normal(mub2, var=sdb2**2);
prior beta3 ~ normal(mub3, var=sdb3**2);
prior beta4 ~ normal(mub4, var=sdb4**2);
prior beta5 ~ normal(mub5, var=sdb5**2);
prior beta6 ~ normal(mub6, var=sdb6**2);
prior mua ~ normal(0, var=10);
prior mub: ~ normal(0, var=4);
prior sda ~ uniform(0,5);
prior sdb: ~ uniform(0,1);
begincnst;
a0=0.5;
endcnst;
mu      =      exp(logoffset      +      alpha      +      beta1*Gas      +
beta2*x2+beta3*x3+beta4*x4+beta5*x5+beta6*Density);
llike=logpdf('poisson',ClaimNb,mu);
if (city='old') then llike=a0*llike;
model general(llike);
run;
```

Part 3:

```
/*Gamma regression with noninformative prior for severity
analysis*/

/*
AggClaimAmount: response variable, severity of claims;
ClaimNb: number of claims occurred during a given time period in
the region for a customer, used as the exponential family
dispersion parameter weight for each observation;
X1: indicator variable, Gas=1 if the car insured uses regular gas;
x2: indicator variable, x2=1 if the insured driver is older than
22 but is 26 or younger;
x3: indicator variable, x3=1 if the insured driver is older than
26 but is 42 or younger;
x4: indicator variable, x4=1 if the insured driver is older than
42 but is 74 or younger;
x5: indicator variable, x5=1 if the insured driver is older than
74;
sdensity: normalized population density of the region;
Default uniform priors are used for all regression coefficients;
Default INV-Gamma(0.001, 0.001) used for the rate parameter (see
```

```
paper).
*/


proc genmod data= datasetname;
class Gas;
Weight ClaimNb;
model AggClaimAmount = x1 x2 x3 x4 x5 sdensity/ dist=gamma
                                link=log;
bayes seed=4 outpost=postgamma diagnostics=all summary=all;
run;
```