mathematics

MDPI

Article

Bayesian Approach to Simultaneous Variable Selection and Estimation in a Linear Regression Model with Applications in Driver Telematics

Himchan Jeong [1] and Minwoo Kim [2,*]

[1] Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada; himchan_jeong@sfu.ca
[2] Department of Statistics, Pusan National University, Busan 46241, Republic of Korea
* Correspondence: mwkim@pusan.ac.kr

Abstract

This article proposes a novel application of the Bayesian variable selection framework for driver telematics data. Unlike the traditional LASSO, the Bayesian variable selection framework allows us to incorporate the importance of certain features in advance in the variable selection procedure so that the traditional features more likely remain in the ratemaking models. The applicability of the proposed framework in the ratemaking practices is also validated via synthetic telematics data.

Keywords: driver telematics; bayesian variable selection; LASSO; usage-based insurance

MSC: 62P05

1. Introduction

Driver telematics is a field that combines telecommunications and informatics to monitor and analyze the behavior and performance of drivers. It involves the use of devices/apps and technologies to collect data on various aspects of driving, which can be used for improving safety, efficiency, and overall driving experience. More specifically, the device or app collects various types of data including vehicle speed and acceleration, braking patterns, steering behavior, and idle time to capture the inherent risk associated with the behaviors of the driver. Driver telematics has been actively used by insurers in the format of pay-as-you-drive (PAYD) or pay-how-you-drive (PHYD), so there have been many research outputs on the effective usage of driver telematics data for auto insurance pricing, such as [1,2].

However, while driver telematics data contain rich information for more precise risk classification, most of the time the data are high-dimensional and require appropriate dimension reduction and/or regularization in order to be compatible with GLMs, which is one of the mainly used benchmarks for auto insurance pricing. For example, ref. [3] utilized a speed–acceleration heatmap to effectively use high-dimensional telematics features in auto insurance pricing. Ref. [4] discussed possible dimension reduction techniques for driver telematics data, such as territorial embedding and PCA. Recently, ref. [5] developed a methodology to summarize telematics characteristics into a one-dimensional safety score, which can be used as input for a regulatory-compliant GLM rating. Note that all the aforementioned dimension reduction approaches preserve the traditional features, whereas only the telematics features are pre-processed due to the non-ignorable importance of

check for updates

Academic Editor: Mark Kelbert

Received: 19 August 2025
Revised: 13 October 2025
Accepted: 14 October 2025
Published: 20 October 2025

Citation: Jeong, H.; Kim, M. Bayesian Approach to Simultaneous Variable Selection and Estimation in a Linear Regression Model with Applications in Driver Telematics. *Mathematics* **2025**, *13*, 3341. https://doi.org/10.3390/math13203341

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

traditional features (such as vehicle capacity or primary use of the vehicle) in auto insurance ratemaking practices [6–10] (although UBI contracts have been available to policyholders for more than twenty years, their market share remains relatively small at approximately 5% [11,12]). However, the literature on variable selection with actuarial application is still scarce, and the importance of traditional features has not yet been incorporated into the variable selection procedure [13–16]. In this regard, we propose a novel variable selection approach in a Bayesian framework for driver telematics data that enables us to incorporate traditional features as well as effectively select significant telematics features, which are often high-dimensional.

1.1. Driver Telematics Data

This section elaborates on the synthetic telematics dataset employed in our empirical analysis. The dataset originated from the work of [17], who obtained the original telematics data from a Canadian insurance company, covering auto insurance policies and claims in Ontario from 2013 to 2016. To protect sensitive information and adhere to data-sharing policies, ref. [17] utilized an algorithm known as "Extended SMOTE" to generate a synthetic version of the dataset. Note that we also utilized a more pre-processed version of the dataset in which the territorial code (with 55 categories) was converted to a real valued variable `TerritoryEmb` as in [4] for this paper.

The dataset comprises 3864 entries with positive claim amounts and includes 10 traditional variables along with 39 telematics-related variables. The outcome variable analyzed in this paper is the size of the insurance claim on a natural logarithmic scale, and descriptions of all variables are provided in Table 1.

Table 1. Description of the traditional and telematics features in the dataset.

Type	Variable	Description
Traditional	`Insured.age`	Age of insured driver, in years
	`Insured.sex`	Sex of insured driver (Male/Female)
	`Car.age`	Age of vehicle, in years
	`Marital`	Marital status (single/married)
	`Car.use`	Use of vehicle: private, commute, farmer, commercial
	`Credit.score`	Credit score of insured driver
	`Region`	Type of region where driver lives: rural, urban
	`Annual.miles.drive`	Annual miles expected to be driven declared by driver
	`Years.noclaims`	Number of years without any claims
	`TerritoryEmb`	Embedded value from the territorial location of vehicle
Telematics	`Annual.pct.driven`	Annualized percentage of time on the road
	`Total.miles.driven`	Total distance driven in miles
	`Pct.drive.xxx`	Percent of driving day xxx of the week: mon/tue/. . . /sun
	`Pct.drive.xhrs`	Percent vehicle driven within x h: 2 h/3 h/4 h
	`Pct.drive.wkday`	Percent vehicle driven during weekdays
	`Pct.drive.wkend`	Percent vehicle driven during weekends
	`Pct.drive.rush.am`	Percent of driving during a.m. rush hours
	`Pct.drive.rush.pm`	Percent of driving during p.m. rush hours
	`Avgdays.week`	Mean number of days used per week
	`Accel.xxmiles`	Number of sudden acceleration 6/8/9/11/12/14 mph/s per 1000 m
	`Brake.xxmiles`	Number of sudden brakes 6/8/9/11/12/14 mph/s per 1000 m
	`Left.turn.intensityxx`	Number of left turn per 1000 m with intensity 08/09/10/11/12
	`Right.turn.intensityxx`	Number of right turn per 1000 m with intensity 08/09/10/11/12
Response	`log(50+AMT_Claim)`	Natural logarithm of the total amount of observed claims

Here `Annual.miles.drive` is categorized as a traditional variable in Table 1; it is based on self-reported data rather than telematics. In contrast, `Total.miles.driven` reflects actual mileage recorded by telematics devices or mobile applications.

1.2. Variable Selection in a Linear Regression Model

Here we use a linear regression model for analyzing insurance claims on a natural log scale, which is given as follows:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{1}$$

Let the observable data consisting of $\boldsymbol{y} = (y_1, \ldots, y_n)^\top \in \mathbb{R}^n$ and $\boldsymbol{X} \in \mathbb{R}^{n \times p}$, where y_i represents the aggregated claim amount of the i-th policyholder on a log scale, and $\boldsymbol{X}$ is a given deterministic design matrix, which contains variables for driver telematics as well as traditional variables such as age and region. The vector $\boldsymbol{\epsilon} \sim \mathcal{N}_n(\boldsymbol{0}, \sigma^2 I_n)$ denotes additive Gaussian errors, and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of regression coefficients to be estimated. Note that the assumption of Gaussian errors in insurance claim modeling is widely adopted in the actuarial literature [18–21]. As illustrated by the histogram of the response variable in Table 1, its empirical distribution exhibits a strong resemblance to a normal distribution (Figure 1).

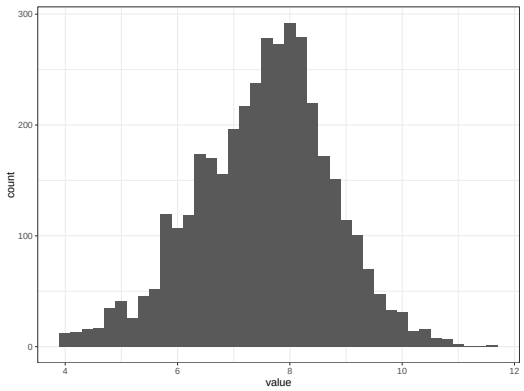


Figure 1. Histogram of the response variable.

In the standard linear model, the regression coefficients $\boldsymbol{\beta}$ are estimated by the ordinary least-squares estimator (LSE), which minimizes the residual sum of squares:

$$\hat{\boldsymbol{\beta}}^{\mathrm{LSE}} = \arg\min_{\boldsymbol{\beta}} \left\{ \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 \right\} = \left(\boldsymbol{X}^\top \boldsymbol{X} \right)^{-1} \boldsymbol{X}^\top \boldsymbol{y},$$

where $\|z\|^2 = z^\top z$ for a vector z. In `R`, we use the function `lm()` to compute the LSE.

Alternatives to the standard linear model can be employed with a least absolute shrinkage and selection operator (LASSO) and an elastic net (EN) model for simultaneous variable selection and parameter estimation. Using a popular `R` package called `glmnet`, which provides a unified framework for fitting penalized regression models, we compute the LASSO and EN estimators. `glmnet` solves the following problem, which contains two key tuning parameters α and λ:

$$\min_{\boldsymbol{\beta}} \left\{ \frac{1}{n} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda \left((1-\alpha) \|\boldsymbol{\beta}\|^2/2 + \alpha \|\boldsymbol{\beta}\|_1 \right) \right\}.$$

Setting $\alpha = 1$, we can obtain the LASSO estimator. When $\alpha \in (0, 1)$, the resulting estimator becomes an EN estimator. The tuning parameter $\lambda > 0$ controls the overall strength of the regularization. We select the optimal value of λ via grid search based on cross-validation, as implemented in the `cv.glmnet()` function.

To highlight the importance of variable selection, we compare the root mean squared errors (RMSEs) of the LASSO, EN, and linear models. We determine the optimal value of λ by performing cross-validation over a grid ranging from 0.01 to 2.00 in increments of 0.01, and we use $\alpha = 0.5$ for EN models. The original dataset described in Section 1.1 is treated as the population from which a sample is drawn. We generate 50 samples, each of

size S, and compute the RMSEs for each sample using 5-fold cross-validation. The use of cross-validation provides a more reliable assessment of predictive performance by reducing the variability associated with a single data split [22]. The sample sizes considered are $S \in \{600, 1000, 1400\}$. The results of the 5-fold cross-validation across the 50 samples are summarized in Table 2 and Figure 2.

Table 2. Summary statistics for the RMSEs of LASSO, EN, and linear models across 50 samples.

Method	S	Min	1st Quartile	2nd Quartile	Mean	3rd Quartile	Max
LASSO	600	0.923	0.942	0.953	1.122	0.965	6.343
	1000	0.927	0.940	0.948	1.021	0.958	3.771
	1400	0.925	0.936	0.942	0.945	0.948	1.079
EN	600	0.918	0.943	0.956	1.156	0.968	5.838
	1000	0.928	0.941	0.947	0.999	0.957	2.704
	1400	0.926	0.938	0.942	0.959	0.949	1.745
Linear model	600	1.023	4.359	7.981	10.915	14.320	45.505
	1000	0.971	2.128	4.268	7.253	10.418	41.810
	1400	0.977	1.452	2.326	2.897	3.250	9.854

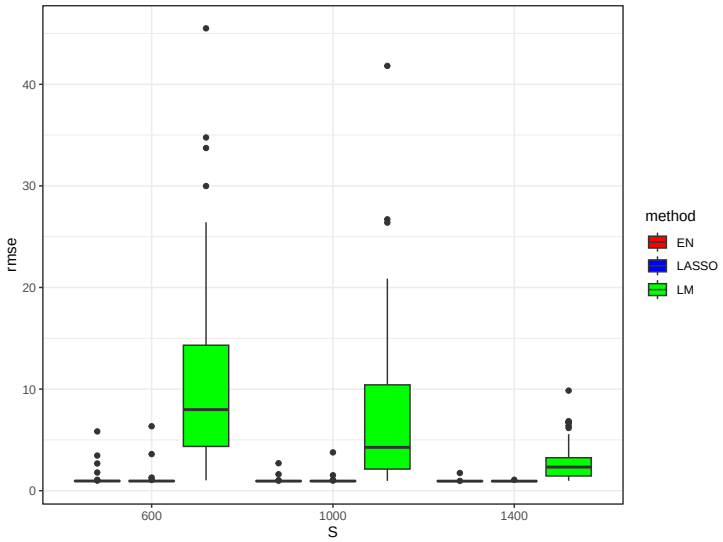


Figure 2. Boxplots of RMSEs for LASSO, EN, and linear models (LMs) across different sample sizes. The horizontal axis indicates sample sizes $S \in \{600, 1000, 1400\}$, and the vertical axis shows the RMSEs.

Notably, the LASSO and EN models significantly reduce the RMSEs for all values of S. However, they do not differentiate between traditional variables and telematics variables during selection. Given that insurance companies may prioritize traditional variables, we analyze the proportions of traditional variables included by the LASSO and EN models. Table 3 presents the inclusion rates for the five traditional variables, `Insured.age, Car.age, Car.use, Credit.score, and Region`, that are considered important by insurance companies. Indeed, we can see that the inclusion rates for the four traditional variables, except for the credit score, are very low, especially as the sample size decreases. To overcome such a limitation, we propose a Bayesian method that allows us to control the probabilities that specific variables are included in the model. While the naïve use of LASSO and EN regularizations cannot handle the distinction between the traditional variables and telematics variables, the proposed Bayesian approach allows us to incorporate the importance of the features in advance by adjusting the values of the hyperparameters.

The remainder of this article is organized as follows: Section 2 explains the theoretical foundation of the proposed Bayesian variable selection method. Section 3 presents the

applicability of the proposed method for driver telematics data. We conclude the article in Section 4 with a brief discussion of possible future research directions.

Table 3. Inclusion rates of the five traditional variables—`Insured.age`, `Car.age`, `Car.use`, `Credit.score`, and `Region`—using the LASSO and EN models.

Method	S	`Insured.age`	`Car.age`	`Car.use`	`Credit.score`	`Region`
LASSO	600	0.46	0.54	0.30	0.92	0.20
	1000	0.68	0.72	0.40	1.00	0.26
	1400	0.74	0.90	0.64	1.00	0.14
EN	600	0.68	0.50	0.30	1.00	0.24
	1000	0.74	0.80	0.48	1.00	0.24
	1400	0.84	0.92	0.68	1.00	0.26

2. Variable Selection in a Bayesian Framework

2.1. Prior Distributions and Likelihood

For simultaneous variable selection and estimation, we utilize a Bayesian method, aiming to identify significant predictors (or covariates) that should be included in the model while estimating unknown parameters. First, to impose sparsity on $\boldsymbol{\beta}$, we employ a latent binary vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)^\top \in \{0,1\}^p$ and use the popular spike-and-slab prior with a Gaussian slab as follows [23]:

$$\beta_j \mid \gamma_j, \sigma^2 \overset{indep}{\sim} (1-\gamma_j)\mathbb{1}(\beta_j = 0) + \gamma_j \mathcal{N}(0, \sigma^2 v^2),\ j = 1, \ldots, p, \tag{2}$$

where $\mathbb{1}(E)$ denotes an indicator function of an event E such that $\mathbb{1}(E) = 1$ if E is true; otherwise, $\mathbb{1}(E) = 0$. v^2 is a hyperparameter. Given the latent vector $\boldsymbol{\gamma}$, the spike-and-slab prior in (2) indicates that the regression coefficients are independent, with each β_j following one of two distributions. Specifically, β_j has a point mass at 0 if $\gamma_j = 0$ and follows a normal distribution with mean zero and variance $\sigma^2 v^2$ if $\gamma_j = 1$. Let $p(\boldsymbol{\beta} \mid \boldsymbol{\gamma}, \sigma^2)$ be the prior density function of $\boldsymbol{\beta}$ given $\boldsymbol{\gamma}$ and σ^2. Then,

$$p(\boldsymbol{\beta} \mid \boldsymbol{\gamma}, \sigma^2) = \prod_{j=1}^{p} p(\beta_j \mid \gamma_j, \sigma^2),$$

where $p(\beta_j \mid \gamma_j, \sigma^2)$ is the density function of β_j given γ_j and σ^2 corresponding to (2).

For the latent binary vector $\boldsymbol{\gamma}$, ref. [24] employed independent Bernoulli prior distributions for each γ_j, $j = 1, \ldots, p$ and proved the selection consistency of the posterior distribution of $\boldsymbol{\gamma}$. In contrast, ref. [25] considered a two-parameter Ising prior to capture the structural dependence among covariates, assuming that the latent variables $\gamma_1, \ldots, \gamma_p$ lie in an undirected graph. Let $\boldsymbol{J} \in \mathbb{R}^{p \times p}$ represent a matrix that determines the connectivity among $\boldsymbol{\gamma}$ called a *coupling matrix*. To construct $\boldsymbol{J}$, ref. [25] used an adjacency matrix for the underlying graph such that the diagonal entries of $\boldsymbol{J}$ are all zeros and the (j,k)-th entry of $\boldsymbol{J}$ is

$$\boldsymbol{J}(j,k) = \begin{cases} 1, & \text{if } \gamma_j \text{ and } \gamma_k \text{ are connected,} \\ 0, & \text{otherwise,} \end{cases}$$

where $j, k \in \{1, 2, \ldots, p\}$ and $j \neq k$. Thus, the prior probability mass function (pmf) of $\boldsymbol{\gamma}$ in [25] is as follows:

$$p(\boldsymbol{\gamma}) = \frac{1}{Z(a,b)} \exp\left(a \sum_{j=1}^{p} \gamma_j + b \cdot \boldsymbol{\gamma}^\top \boldsymbol{J} \boldsymbol{\gamma} \right), \tag{3}$$

where $Z(a,b) = \sum_{\boldsymbol{\gamma}} \exp\left(a\sum_{j=1}^{p}\gamma_j + b\cdot\boldsymbol{\gamma}^\top \boldsymbol{J}\boldsymbol{\gamma}\right)$ is a normalizing constant. The Ising prior described in (3) effectively accounts for the dependence among covariates, except for the special case where $b = 0$, which corresponds to independent Bernoulli priors. In order to use the Ising prior in (3), $\boldsymbol{J}$ must be pre-specified. In their simulation studies, ref. [25] assumed a linear chain dependence such that each element of $\boldsymbol{\gamma}$ was connected to two neighboring elements—one step ahead and one step behind, i.e., γ_j was connected to γ_{j-1} and γ_{j+1} for $j = 1,\ldots,p$, with the boundary conditions $\gamma_0 = \gamma_p$ and $\gamma_{p+1} = \gamma_1$.

We construct the coupling matrix $\boldsymbol{J}$ in a more data-driven manner by taking into account that (1) the inclusion probabilities of traditional variables should be higher than those of telematics variables, and (2) the correlations among telematics variables should be appropriately modeled. Let $\mathcal{I}_{\text{trad}}$ and $\mathcal{I}_{\text{tele}}$ denote index sets for traditional and telematics variables, respectively, such that $\mathcal{I}_{\text{trad}} \cup \mathcal{I}_{\text{tele}} = \{1,2,\ldots,p\}$ and $\mathcal{I}_{\text{trad}} \cap \mathcal{I}_{\text{tele}} = \emptyset$. For (1), we assign a large value to an off-diagonal entry if the corresponding indices are in $\mathcal{I}_{\text{trad}}$. Next, for (2), we define the connectivity between γ_j and γ_k for $j,k \in \mathcal{I}_{\text{tele}}$ using the sample correlation between the j-th and k-th covariates denoted by r_{jk}. The off-diagonal entries of $\boldsymbol{J}$ are then defined as $\boldsymbol{J}(j,k) = 1$ if $|r_{jk}| > \tau$, and $\boldsymbol{J}(j,k) = 0$ otherwise. Here, a threshold $\tau \in (0,1)$ controls the sparsity of $\boldsymbol{J}$. The diagonal entries of $\boldsymbol{J}$ are all zeros, and the off-diagonal entries are summarized in (4):

$$\boldsymbol{J}(j,k) = \begin{cases} J^*, \text{ if } j,k \in \mathcal{I}_{\text{trad}}\,, \\ 1, \text{ if } j,k \in \mathcal{I}_{\text{tele}} \text{ and } |r_{jk}| > \tau, \\ 0, \text{ if } j,k \in \mathcal{I}_{\text{tele}} \text{ and } |r_{jk}| \le \tau, \end{cases} \tag{4}$$

where $J^* > 1$. We emphasize that the prior pmf of $\boldsymbol{\gamma}$ remains the same as in (3), with the only difference being in $\boldsymbol{J}$. Furthermore, as the value of J^* increases, the probabilities that traditional variables are included in the model also increase. Finally, for the variance in the error terms, we set the prior on σ^2 in the form $p(\sigma^2) \propto 1/\sigma^2$, as in [25]. All the necessary random quantities $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and σ^2 are given; the likelihood function is as follows:

$$p(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2\right).$$

It should be noted that $\boldsymbol{J}$ is specified from the given dataset at the beginning and is not updated throughout the Gibbs sampling procedure. Using the prior distributions and likelihood specified in this section, we describe the procedures for variable selection and parameter estimation in Sections 2.2 and 2.3.

2.2. Posterior Inference for $\boldsymbol{\gamma}$

We introduce the Gibbs sampling method suggested by [25] for variable selectionsunder the Ising prior in (3) with a coupling matrix in (4). Significant predictor variables are identified based on the posterior inclusion probabilities (PIPs) $Pr(\gamma_j = 1 \mid \boldsymbol{y})$, $j = 1,\ldots,p$. However, the PIPs cannot be computed in a closed form. To approximate the PIPs, ref. [25] defined $(p-1)$-dimensional vectors $\boldsymbol{\gamma}_{-j} \in \{0,1\}^{p-1}$, which are constructed by removing γ_j from $\boldsymbol{\gamma}$ and index sets $\mathcal{I}_{-j} = \{k : \gamma_k = 1, k \neq j\}$ for $j = 1,\ldots,p$. Then, the prior

conditional probabilities $Pr(\gamma_j = 0 \mid \boldsymbol{\gamma}_{-j})$ and $Pr(\gamma_j = 1 \mid \boldsymbol{\gamma}_{-j})$ are straightforward to compute using the fact that γ_j is binary:

$$Pr(\gamma_j = 0 \mid \boldsymbol{\gamma}_{-j}) = \frac{1}{1 + \exp\left(a + b \sum_{k \in \mathcal{I}_{-j}} \gamma_k\right)},$$
$$Pr(\gamma_j = 1 \mid \boldsymbol{\gamma}_{-j}) = \frac{\exp\left(a + b \sum_{k \in \mathcal{I}_{-j}} \gamma_k\right)}{1 + \exp\left(a + b \sum_{k \in \mathcal{I}_{-j}} \gamma_k\right)}.$$

It is worth noting that the conditional probabilities above no longer involve intractable normalizing constants. By the Bayes rule, the posterior conditional probability that $\gamma_j = 1$ given $\boldsymbol{\gamma}_{-j}$ and $\boldsymbol{y}$ is

$$Pr(\gamma_j = 1 \mid \boldsymbol{\gamma}_{-j}, \boldsymbol{y}) = \frac{Pr(\gamma_j = 1 \mid \boldsymbol{\gamma}_{-j})}{Pr(\gamma_j = 1 \mid \boldsymbol{\gamma}_{-j}) + BF_j^{-1} Pr(\gamma_j = 0 \mid \boldsymbol{\gamma}_{-j})}, \tag{5}$$

where $BF_j = p(\boldsymbol{y} \mid \gamma_j = 1, \boldsymbol{\gamma}_{-j}) / p(\boldsymbol{y} \mid \gamma_j = 0, \boldsymbol{\gamma}_{-j})$ is the Bayes factor. We can explicitly compute the Bayes factors by integrating out $\boldsymbol{\beta}$ and σ^2 under the prior distributions specified in Section 2.1. The detailed derivation of BF_j is described in [25]. We draw Gibbs samples of $\boldsymbol{\gamma}$ iteratively using (5) with a initial vector $\boldsymbol{\gamma}^{(0)}$ and compute the approximated PIPs denoted by $\hat{Pr}(\gamma_j = 1 \mid \boldsymbol{y})$ for $j = 1, \ldots, p$ based on the empirical distribution of the Gibbs samples. Specifically, we divide the number of Gibbs samples where $\gamma_j = 1$ by the total number of the samples excluding the burn-ins. Ref. [25] demonstrated that the Ising prior for $\boldsymbol{\gamma}$ is effective in capturing the dependence among covariates and outperforms the independent Bernoulli prior in terms of variable selection. However, they did not address the estimation of $\boldsymbol{\beta}$ or σ^2. To tackle this, we propose the Gibbs sampling method for simultaneous variable selection and parameter estimation in the following subsection.

2.3. Posterior Inference for $\boldsymbol{\beta}$ and σ^2

At the current iteration t for the Gibbs sampling introduced in the previous subsection, the current binary vector is denoted as $\boldsymbol{\gamma}^{(t)} = (\gamma_1^{(t)}, \ldots, \gamma_p^{(t)})^\top$. Given the current vector $\boldsymbol{\gamma}^{(t)}$ with $\gamma_j^{(t)} = 1$, we can derive the full posterior of β_j as follows:

$$\begin{aligned}
&p(\beta_j \mid \boldsymbol{\beta}_{-j}, \boldsymbol{\gamma}^{(t)}, \sigma^2, \boldsymbol{y}) \\
&\propto \exp\left(-\frac{1}{2} \cdot \underbrace{\frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_{ij}^2 + \frac{1}{\nu^2}\right)}_{:=(\tau_j^2)^{-1}} \left(\beta_j^2 - 2 \underbrace{\left(\sum_{i=1}^{n} x_{ij}^2 + \frac{1}{\nu^2}\right)^{-1} \left(\sum_{i=1}^{n} x_{ij}\left(y_i - \sum_{k \neq j}^{p} x_{ik}\beta_k\right)\right)}_{:=\mu_j} \beta_j \right) \right) \\
&\propto \exp\left(-\frac{1}{2\tau_j^2} (\beta_j - \mu_j)^2 \right).
\end{aligned} \tag{6}$$

Observing the last term in (6), which takes the form of a Gaussian density function, we can infer $\beta_j \mid \boldsymbol{\beta}_{-j}, \boldsymbol{\gamma}^{(t)}, \sigma^2, \boldsymbol{y} \sim \mathcal{N}(\mu_j, \tau_j^2)$ from which Gibbs samples for $\boldsymbol{\beta}$ are generated.

For the derivation of the full posterior of σ^2, we define a sparse vector $\tilde{\boldsymbol{\beta}} = \boldsymbol{\gamma} \odot \boldsymbol{\beta}$, where the symbol $\odot$ denotes a Hadamard product and an activation set $\mathcal{A}^{(t)} = \{j : \gamma_j^{(t)} = 1\}$. Then, the full posterior of σ^2 is

$$\begin{aligned} &p(\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{y}) \\ &\quad = (\sigma^2)^{-\frac{p+\left|\mathcal{A}^{(t)}\right|}{2}-1} \exp\left(-\frac{1}{\sigma^2}\left(\frac{1}{2\nu^2}\tilde{\boldsymbol{\beta}}^\top \tilde{\boldsymbol{\beta}} + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})^\top (\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})\right)\right), \end{aligned} \tag{7}$$

where $\left|\mathcal{A}^{(t)}\right|$ denotes the cardinality of $\mathcal{A}^{(t)}$. The last form in (7) indicates that $\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{y}$ follows an inverse gamma distribution with a shape parameter $a_{\sigma^2} := \left(p + \left|\mathcal{A}^{(t)}\right|\right)/2$ and a scale parameter:

$$b_{\sigma^2} := \frac{1}{2\nu^2}\tilde{\boldsymbol{\beta}}^\top \tilde{\boldsymbol{\beta}} + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})^\top (\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}).$$

Using the full posteriors $p(\beta_j \mid \boldsymbol{\beta}_{-j}, \boldsymbol{\gamma}^{(t)}, \sigma^2, \boldsymbol{y})$ and $p(\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{y})$, we can iteratively generate Gibbs samples, and the sample means serve as the Bayesian estimators for $\boldsymbol{\beta}$ and σ^2.

3. Application for Telematics Feature Selection

Now we assess the applicability of the proposed method for variable selection with driver telematics data. In Section 3.1, we examine how the inclusion probabilities of traditional variables vary with the values of J^*, demonstrating that our method effectively adjusts these probabilities in a flexible way.

3.1. Inclusion Probabilities of Traditional Variables

J^* is a user-specified value that can be chosen according to the purpose of the analysis. A larger value of J^* increases the probability that the traditional variables are included in the model. Using the five values of $J^* \in \{2, 3, 5, 7, 10\}$, we follow the steps below to compute the average of the inclusion probabilities with its standard deviations:

Step 0. We choose $J^* \in \{2, 5, 10\}$ and $S \in \{600, 1000, 1400\}$.

Step 1. We generate a sample of size S from the population.

Step 2. Given a sample, we implement the Gibbs procedure described in Section 2.2 to draw $\boldsymbol{\gamma}^{(1)}, \ldots, \boldsymbol{\gamma}^{(2000)}$. The first 1000 iterations are considered burn-ins, and the last 1000 are used.

Step 3. We compute each marginal PIP as follows:

$$\hat{Pr}(\gamma_j = 1 \mid \boldsymbol{y}) = \frac{1}{1000}\sum_{t=1001}^{2000} \gamma_j^{(t)}, \; j = 1, \ldots, p.$$

Step 4. We perform Steps 1, 2, and 3 fifty times and compute the average and standard deviation of PIPs across the fifty repetitions.

Table 4 shows the PIPs for the five traditional variables, `Insured.age`, `Car.age`, `Car.use`, `Credit.score`, and `Region`. The observed increase in PIPs with increasing values of J^* implies that the inclusion rates of traditional variables can be effectively controlled through appropriate choices of J^*. Additional results on the inclusion rates of the remaining traditional and telematics variables are summarized in Appendix A.

3.2. RMSE Comparison

In the previous subsection, we demonstrated that the proposed Bayesian approach allows flexible control over the inclusion probabilities of traditional variables in the model. However, such flexibility would be meaningless if it comes at the cost of poor accuracy. In this subsection, we compare the RMSEs of the proposed method with those of the LASSO model. As in Section 1.2, we compute the RMSEs using the 5-fold cross-validation with the 50 samples generated, each of size $S \in \{600, 1000, 1400\}$. As shown in Table 4, setting $J^* = 10$ ensures the inclusion of all five traditional variables. Therefore, we fix J^* at 10. As summarized in Table 5, the proposed Bayesian method outperforms LASSO, showing that variable inclusion can be flexibly controlled without compromising—and even enhancing—accuracy.

Table 4. Inclusion rates of the five traditional variables—`Insured.age`, `Car.age`, `Car.use`, `Credit.score`, and `Region`—using the proposed Bayesian model.

J^*	S	`Insured.age`	`Car.age`	`Car.use`	`Credit.score`	`Region`
2	600	0.56	0.69	0.50	0.99	0.40
	1000	0.49	0.76	0.49	1.00	0.33
	1400	0.43	0.86	0.54	1.00	0.29
3	600	0.73	0.81	0.69	1.00	0.62
	1000	0.66	0.86	0.67	1.00	0.53
	1400	0.61	0.92	0.69	1.00	0.48
5	600	0.96	0.98	0.96	1.00	0.94
	1000	0.95	0.98	0.95	1.00	0.92
	1400	0.93	0.99	0.95	1.00	0.90
7	600	0.99	1.00	0.99	1.00	1.00
	1000	0.99	1.00	0.99	1.00	0.99
	1400	0.99	1.00	0.99	1.00	0.99
10	600	1.00	1.00	1.00	1.00	1.00
	1000	1.00	1.00	1.00	1.00	1.00
	1400	1.00	1.00	1.00	1.00	1.00

Table 5. Summary statistics for the RMSEs of LASSO and the proposed Bayesian method with $J^* = 10$ across the 50 samples.

Method	S	Min	1st Quartile	2nd Quartile	Mean	3rd Quartile	Max
LASSO	600	0.923	0.942	0.953	1.122	0.965	6.343
	1000	0.927	0.940	0.948	1.021	0.958	3.771
	1400	0.925	0.936	0.942	0.945	0.948	1.079
EN	600	0.918	0.943	0.956	1.156	0.968	5.838
	1000	0.928	0.941	0.947	0.999	0.957	2.704
	1400	0.926	0.938	0.942	0.959	0.949	1.745
Bayesian method	600	0.919	0.948	0.959	0.960	0.969	1.021
	1000	0.923	0.941	0.951	0.949	0.958	0.975
	1400	0.926	0.934	0.941	0.941	0.948	0.969

4. Conclusions

In this article, it was shown that the proposed Bayesian variable selection can effectively address the predetermined importance of available covariates, which maintains the traditional variables in the ratemaking scheme without losing too much of the prediction accuracy. While such results are promising, we acknowledge that the aforementioned Bayesian variable selection framework is constrained by the assumption of Gaussian white noise. In this regard, it is expected that this research can be extended to incorporate the

proposed Bayesian variable selection scheme for response variables with different distributional assumptions, such as the Poisson distribution. It should also be noted that due to the limited availability of publicly available telematics data, all the analyses are based on synthetic telematics data (which originated from actual data) provided by So et al. [17]. That being said, we recommend that practitioners test the external validity of the proposed framework, provided that an actual telematics dataset is available.

Author Contributions: Conceptualization, M.K. and H.J.; methodology, M.K.; software, M.K.; validation, M.K.; formal analysis, M.K.; investigation, M.K.; data curation, H.J.; writing—original draft preparation, M.K. and H.J.; writing—review and editing, M.K. and H.J.; visualization, M.K.; supervision, H.J.; project administration, M.K. and H.J.; funding acquisition, M.K. and H.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by PNU-RENovation (2024–2025).

Data Availability Statement: The synthetic dataset and `R` codes used in this article are available at https://github.com/ssauljin/telematics_Bayes (accessed on 17 October 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PAYD	Pay-as-you-drive
PHYD	Pay-how-you-drive
GLM	Generalized Linear Model
PCA	Principal Component Analysis
SMOTE	Synthetic Minority Oversampling Technique
RMSE	Root Mean Squared Error
LASSO	Least Absolute Shrinkage and Selection Operator
PIP	Posterior Inclusion Probability

Appendix A

Table A1 presents the inclusion probabilities of the remaining traditional variables, excluding the five variables already summarized in Tables 3 and 4. Tables A2–A9 present the inclusion probabilities of the 39 telematics variables.

Table A1. Inclusion rates of the remaining five traditional variables—`Insured.sex`, `Marital`, `Annual.miles.drive`, `Years.noclaims`, and `TerritoryEmb`—using the LASSO, EN, and proposed Bayesian models.

Method	S	`Insured.sex`	`Marital`	`Annual.miles.drive`	`Years.noclaims`	`TerritoryEmb`
LASSO	600	0.12	0.20	0.14	0.88	0.18
	1000	0.20	0.24	0.28	1.00	0.28
	1400	0.14	0.28	0.34	1.00	0.24
EN	600	0.12	0.14	0.12	0.98	0.20
	1000	0.18	0.28	0.32	1.00	0.28
	1400	0.20	0.34	0.38	1.00	0.32
Bayesian method ($J^* = 2$)	600	0.39	0.41	0.42	0.84	0.40
	1000	0.31	0.35	0.38	0.92	0.36
	1400	0.27	0.32	0.36	0.97	0.31
Bayesian method ($J^* = 3$)	600	0.61	0.62	0.63	0.91	0.61
	1000	0.51	0.55	0.57	0.95	0.55
	1400	0.46	0.51	0.53	0.98	0.49
Bayesian method ($J^* = 5$)	600	0.94	0.94	0.94	0.99	0.94
	1000	0.91	0.92	0.93	0.99	0.92
	1400	0.90	0.91	0.91	1.00	0.90
Bayesian method ($J^* = 7$)	600	0.99	0.99	0.99	1.00	0.99
	1000	0.99	0.99	0.99	1.00	0.99
	1400	0.98	0.99	0.99	1.00	0.99
Bayesian method ($J^* = 10$)	600	1.00	1.00	1.00	1.00	1.00
	1000	1.00	1.00	1.00	1.00	1.00
	1400	1.00	1.00	1.00	1.00	1.00

Table A2. Inclusion rates of five telematics variables—`Annual.pct.driven`, `Total.miles.driven`, `Pct.drive.mon`, `Pct.drive.tue`, and `Pct.drive.wed`—using the LASSO, EN, and proposed Bayesian models.

Method	S	`Annual.pct.driven`	`Total.miles.driven`	`Pct.drive.mon`	`Pct.drive.tue`	`Pct.drive.wed`
LASSO	600	0.20	0.30	0.20	0.26	0.34
	1000	0.12	0.58	0.32	0.38	0.62
	1400	0.26	0.80	0.36	0.48	0.78
EN	600	0.20	0.36	0.20	0.30	0.42
	1000	0.18	0.62	0.38	0.42	0.64
	1400	0.22	0.84	0.42	0.54	0.78
Bayesian method ($J^* = 2$)	600	0.24	0.57	0.22	0.27	0.27
	1000	0.20	0.64	0.19	0.24	0.29
	1400	0.20	0.70	0.17	0.24	0.31
Bayesian method ($J^* = 3$)	600	0.25	0.58	0.22	0.27	0.27
	1000	0.20	0.65	0.19	0.24	0.29
	1400	0.21	0.71	0.16	0.24	0.30
Bayesian method ($J^* = 5$)	600	0.25	0.60	0.22	0.27	0.27
	1000	0.20	0.67	0.19	0.24	0.29
	1400	0.21	0.74	0.16	0.24	0.30
Bayesian method ($J^* = 7$)	600	0.25	0.60	0.22	0.27	0.27
	1000	0.20	0.67	0.19	0.24	0.29
	1400	0.21	0.73	0.17	0.24	0.30
Bayesian method ($J^* = 10$)	600	0.25	0.60	0.22	0.27	0.27
	1000	0.20	0.67	0.19	0.24	0.29
	1400	0.21	0.73	0.17	0.24	0.30

Table A3. Inclusion rates of five telematics variables—`Pct.drive.thr`, `Pct.drive.fri`, `Pct.drive.sat`, `Pct.drive.sun`, and `Pct.drive.2hrs`—using the LASSO, EN, and proposed Bayesian models.

Method	S	`Pct.drive.thr`	`Pct.drive.fri`	`Pct.drive.sat`	`Pct.drive.sun`	`Pct.drive.2hrs`
LASSO	600	0.18	0.26	0.08	0.22	0.40
	1000	0.12	0.32	0.16	0.42	0.64
	1400	0.10	0.44	0.16	0.52	0.72
EN	600	0.18	0.30	0.08	0.28	0.42
	1000	0.14	0.34	0.20	0.48	0.66
	1400	0.10	0.52	0.24	0.62	0.82
Bayesian method ($J^* = 2$)	600	0.23	0.23	0.32	0.34	0.38
	1000	0.18	0.20	0.30	0.34	0.37
	1400	0.16	0.21	0.29	0.36	0.37
Bayesian method ($J^* = 3$)	600	0.23	0.23	0.32	0.34	0.38
	1000	0.18	0.21	0.29	0.34	0.37
	1400	0.15	0.21	0.29	0.35	0.37
Bayesian method ($J^* = 5$)	600	0.23	0.23	0.32	0.34	0.38
	1000	0.18	0.21	0.29	0.34	0.37
	1400	0.16	0.21	0.29	0.36	0.36
Bayesian method ($J^* = 7$)	600	0.23	0.23	0.32	0.34	0.38
	1000	0.17	0.21	0.29	0.34	0.37
	1400	0.15	0.21	0.29	0.36	0.37
Bayesian method ($J^* = 10$)	600	0.23	0.23	0.32	0.34	0.38
	1000	0.17	0.21	0.29	0.34	0.37
	1400	0.15	0.21	0.29	0.36	0.36

Table A4. Inclusion rates of five telematics variables—`Pct.drive.3hrs`, `Pct.drive.4hrs`, `Pct.drive.wkday`, `Pct.drive.wkend`, and `Pct.drive.rush.am`—using the LASSO, EN, and proposed Bayesian models.

Method	S	`Pct.drive.3hrs`	`Pct.drive.4hrs`	`Pct.drive.wkday`	`Pct.drive.wkend`	`Pct.drive.rush.am`
LASSO	600	0.32	0.22	0.02	0.02	0.22
	1000	0.56	0.32	0.0	0.0	0.36
	1400	0.66	0.38	0.02	0.02	0.52
EN	600	0.40	0.24	0.02	0.02	0.16
	1000	0.58	0.34	0.0	0.0	0.44
	1400	0.70	0.38	0.04	0.04	0.54
Bayesian method ($J^* = 2$)	600	0.45	0.35	0.37	0.37	0.26
	1000	0.43	0.31	0.34	0.34	0.26
	1400	0.45	0.35	0.33	0.34	0.24
Bayesian method ($J^* = 3$)	600	0.45	0.35	0.37	0.37	0.26
	1000	0.43	0.31	0.34	0.34	0.27
	1400	0.46	0.35	0.33	0.34	0.25
Bayesian method ($J^* = 5$)	600	0.45	0.35	0.37	0.37	0.27
	1000	0.43	0.32	0.34	0.34	0.27
	1400	0.46	0.35	0.33	0.34	0.26
Bayesian method ($J^* = 7$)	600	0.45	0.35	0.37	0.37	0.27
	1000	0.43	0.32	0.34	0.34	0.28
	1400	0.46	0.35	0.33	0.34	0.27
Bayesian method ($J^* = 10$)	600	0.45	0.35	0.37	0.37	0.27
	1000	0.43	0.32	0.34	0.34	0.28
	1400	0.46	0.35	0.33	0.34	0.27

Table A5. Inclusion rates of five telematics variables—`Pct.drive.rush.pm`, `Avgdays.week`, `Accel.06miles`, `Accel.08miles`, and `Accel.09miles`—using the LASSO, EN, and proposed Bayesian models.

Method	S	`Pct.drive.rush.pm`	`Avgdays.week`	`Accel.06miles`	`Accel.08miles`	`Accel.09miles`
LASSO	600	0.16	0.34	0.46	0.12	0.02
	1000	0.22	0.58	0.78	0.06	0.02
	1400	0.24	0.80	0.80	0.16	0.0
EN	600	0.16	0.32	0.50	0.14	0.04
	1000	0.22	0.64	0.80	0.12	0.02
	1400	0.26	0.82	0.84	0.18	0.0
Bayesian method ($J^* = 2$)	600	0.22	0.39	0.35	0.41	0.44
	1000	0.18	0.50	0.34	0.35	0.39
	1400	0.15	0.59	0.29	0.31	0.35
Bayesian method ($J^* = 3$)	600	0.22	0.39	0.35	0.41	0.44
	1000	0.18	0.51	0.35	0.36	0.39
	1400	0.15	0.59	0.29	0.32	0.35
Bayesian method ($J^* = 5$)	600	0.23	0.40	0.35	0.41	0.44
	1000	0.18	0.51	0.35	0.36	0.39
	1400	0.15	0.60	0.29	0.32	0.35
Bayesian method ($J^* = 7$)	600	0.23	0.40	0.35	0.41	0.44
	1000	0.18	0.51	0.35	0.36	0.39
	1400	0.15	0.60	0.29	0.31	0.35
Bayesian method ($J^* = 10$)	600	0.23	0.40	0.35	0.41	0.44
	1000	0.18	0.51	0.35	0.36	0.39
	1400	0.15	0.60	0.30	0.32	0.35

Table A6. Inclusion rates of five telematics variables—`Accel.11miles`, `Accel.12miles`, `Accel.14miles`, `Brake.06miles`, and `Brake.08miles`—using the LASSO, EN, and proposed Bayesian models.

Method	S	`Accel.11miles`	`Accel.12miles`	`Accel.14miles`	`Brake.06miles`	`Brake.08miles`
LASSO	600	0.02	0.04	0.02	0.42	0.58
	1000	0.04	0.04	0.06	0.74	0.64
	1400	0.0	0.04	0.06	0.78	0.74
EN	600	0.04	0.04	0.08	0.62	0.72
	1000	0.04	0.08	0.08	0.80	0.74
	1400	0.04	0.04	0.08	0.86	0.84
Bayesian method ($J^* = 2$)	600	0.46	0.47	0.48	0.42	0.47
	1000	0.41	0.43	0.41	0.46	0.46
	1400	0.38	0.41	0.37	0.56	0.48
Bayesian method ($J^* = 3$)	600	0.46	0.47	0.48	0.42	0.47
	1000	0.41	0.43	0.41	0.46	0.46
	1400	0.38	0.41	0.38	0.56	0.48
Bayesian method ($J^* = 5$)	600	0.46	0.48	0.48	0.41	0.47
	1000	0.41	0.43	0.41	0.45	0.46
	1400	0.38	0.41	0.38	0.55	0.48
Bayesian method ($J^* = 7$)	600	0.46	0.47	0.48	0.41	0.47
	1000	0.41	0.43	0.41	0.45	0.47
	1400	0.38	0.40	0.37	0.55	0.48
Bayesian method ($J^* = 10$)	600	0.46	0.47	0.48	0.41	0.47
	1000	0.41	0.43	0.41	0.45	0.46
	1400	0.38	0.40	0.37	0.55	0.48

Table A7. Inclusion rates of five telematics variables—`Brake.09miles`, `Brake.11miles`, `Brake.12miles`, `Brake.14miles`, and `Left.turn.intensity08`—using the LASSO, EN, and proposed Bayesian models.

Method	S	`Brake.09miles`	`Brake.11miles`	`Brake.12miles`	`Brake.14miles`	`Left.turn.intensity08`
LASSO	600	0.28	0.50	0.04	0.08	0.12
	1000	0.30	0.66	0.12	0.10	0.20
	1400	0.16	0.90	0.08	0.02	0.30
EN	600	0.42	0.56	0.14	0.04	0.12
	1000	0.38	0.82	0.16	0.08	0.28
	1400	0.22	0.92	0.14	0.06	0.30
Bayesian method ($J^* = 2$)	600	0.51	0.63	0.49	0.45	0.40
	1000	0.51	0.68	0.45	0.42	0.35
	1400	0.56	0.72	0.39	0.35	0.33
Bayesian method ($J^* = 3$)	600	0.51	0.62	0.49	0.45	0.40
	1000	0.52	0.68	0.44	0.42	0.35
	1400	0.56	0.72	0.40	0.35	0.34
Bayesian method ($J^* = 5$)	600	0.51	0.63	0.49	0.45	0.40
	1000	0.52	0.68	0.45	0.42	0.35
	1400	0.56	0.71	0.40	0.35	0.34
Bayesian method ($J^* = 7$)	600	0.52	0.63	0.49	0.46	0.40
	1000	0.52	0.68	0.45	0.42	0.35
	1400	0.56	0.72	0.39	0.35	0.33
Bayesian method ($J^* = 10$)	600	0.52	0.63	0.49	0.46	0.40
	1000	0.52	0.68	0.45	0.42	0.35
	1400	0.56	0.71	0.39	0.35	0.33

Table A8. Inclusion rates of five telematics variables—`Left.turn.intensity09`, `Left.turn.intensity10`, `Left.turn.intensity11`, `Left.turn.intensity12`, and `Right.turn.intensity08`—using the LASSO, EN, and proposed Bayesian models.

Method	S	`Left.turn.intensity09`	`Left.turn.intensity10`	`Left.turn.intensity11`	`Left.turn.intensity12`	`Right.turn.intensity08`
LASSO	600	0.02	0.04	0.02	0.10	0.20
	1000	0.06	0.0	0.04	0.10	0.36
	1400	0.04	0.0	0.04	0.12	0.40
EN	600	0.06	0.02	0.10	0.0	0.20
	1000	0.12	0.02	0.04	0.14	0.30
	1400	0.04	0.0	0.08	0.18	0.48
Bayesian method ($J^* = 2$)	600	0.42	0.44	0.44	0.44	0.44
	1000	0.38	0.38	0.38	0.37	0.42
	1400	0.36	0.36	0.35	0.34	0.40
Bayesian method ($J^* = 3$)	600	0.42	0.44	0.44	0.44	0.44
	1000	0.38	0.38	0.38	0.37	0.42
	1400	0.36	0.36	0.35	0.34	0.40
Bayesian method ($J^* = 5$)	600	0.42	0.44	0.44	0.44	0.44
	1000	0.38	0.38	0.38	0.37	0.42
	1400	0.36	0.36	0.35	0.34	0.40
Bayesian method ($J^* = 7$)	600	0.42	0.44	0.44	0.44	0.44
	1000	0.38	0.38	0.38	0.37	0.42
	1400	0.36	0.36	0.35	0.34	0.40
Bayesian method ($J^* = 10$)	600	0.42	0.44	0.44	0.44	0.44
	1000	0.38	0.38	0.38	0.37	0.42
	1400	0.36	0.36	0.35	0.34	0.40

Table A9. Inclusion rates of five telematics variables—`Right.turn.intensity09`, `Right.turn.intensity10`, `Right.turn.intensity11`, and `Right.turn.intensity12`—using the LASSO, EN, and proposed Bayesian models.

Method	S	`Right.turn.intensity09`	`Right.turn.intensity10`	`Right.turn.intensity11`	`Right.turn.intensity12`
LASSO	600	0.04	0.02	0.06	0.08
	1000	0.06	0.4	0.04	0.26
	1400	0.02	0.04	0.06	0.32
EN	600	0.10	0.04	0.04	0.14
	1000	0.18	0.08	0.08	0.30
	1400	0.08	0.06	0.16	0.34
Bayesian method ($J^* = 2$)	600	0.46	0.46	0.46	0.45
	1000	0.43	0.43	0.43	0.40
	1400	0.41	0.41	0.41	0.36
Bayesian method ($J^* = 3$)	600	0.46	0.46	0.46	0.45
	1000	0.43	0.43	0.43	0.40
	1400	0.41	0.41	0.41	0.36
Bayesian method ($J^* = 5$)	600	0.46	0.46	0.46	0.45
	1000	0.43	0.43	0.43	0.40
	1400	0.41	0.42	0.41	0.36
Bayesian method ($J^* = 7$)	600	0.46	0.46	0.46	0.45
	1000	0.43	0.43	0.43	0.40
	1400	0.41	0.42	0.41	0.36
Bayesian method ($J^* = 10$)	600	0.46	0.46	0.46	0.45
	1000	0.43	0.43	0.43	0.40
	1400	0.41	0.42	0.41	0.37

References

1. Ayuso, M.; Guillén, M.; Pérez-Marín, A.M. Time and distance to first accident and driving patterns of young drivers with pay-as-you-drive insurance. *Accid. Anal. Prev.* **2014**, *73*, 125–131. [CrossRef]
2. Cheng, J.; Feng, F.Y.; Zeng, X. Pay-as-you-drive insurance: Modeling and implications. *N. Am. Actuar. J.* **2023**, *27*, 303–321. [CrossRef]
3. Gao, G.; Meng, S.; Wüthrich, M.V. Claims frequency modeling using telematics car driving data. *Scand. Actuar. J.* **2019**, *2019*, 143–162. [CrossRef]
4. Jeong, H. Dimension reduction techniques for summarized telematics data. *J. Risk Manag.* **2022**, *33*, 1–24. [CrossRef]
5. Peiris, H.; Jeong, H.; Zou, B. Development of Telematics Risk Scores in Accordance with Regulatory Compliance. Available at SSRN 5049191. 2024. Available online: https://ssrn.com/abstract=5049191 (accessed on 31 January 2025).
6. Guillen, M.; Nielsen, J.P.; Ayuso, M.; Pérez-Marín, A.M. The use of telematics devices to improve automobile insurance rates. *Risk Anal.* **2019**, *39*, 662–672. [CrossRef]
7. Jiang, Q.; Shi, T. Auto insurance pricing using telematics data: Application of a hidden markov model. *N. Am. Actuar. J.* **2024**, *28*, 822–839. [CrossRef]
8. Chan, I.W.; Tseung, S.C.; Badescu, A.L.; Lin, X.S. Data mining of telematics data: Unveiling the hidden patterns in driving behavior. *N. Am. Actuar. J.* **2025**, *29*, 275–309. [CrossRef]
9. Guillen, M.; Nielsen, J.P.; Pérez-Marín, A.M. Near-miss telematics in motor insurance. *J. Risk Insur.* **2021**, *88*, 569–589. [CrossRef]
10. So, B.; Jeong, H. Simulation Engine for Adaptive Telematics data. *Variance* **2025**, *18*. [CrossRef]
11. Holzapfel, J.; Peter, R.; Richter, A. Mitigating moral hazard with usage-based insurance. *J. Risk Insur.* **2024**, *91*, 813–839. [CrossRef]
12. Peiris, H.; Jeong, H.; Kim, J.K.; Lee, H. Integration of traditional and telematics data for efficient insurance claims prediction. *ASTIN Bull. J. IAA* **2024**, *54*, 263–279. [CrossRef]

13. Williams, B.; Hansen, G.; Baraban, A.; Santoni, A. A practical approach to variable selection—A comparison of various techniques. In Proceedings of the Casualty Actuarial Society E-Forum, Philadelphia, PA, USA, 9–11 March 2015.
14. Devriendt, S.; Antonio, K.; Reynkens, T.; Verbelen, R. Sparse regression with multi-type regularized feature modeling. *Insur. Math. Econ.* **2021**, *96*, 248–261. [CrossRef]
15. McGuire, G.; Taylor, G.; Miller, H. Self-assembling insurance claim models using regularized regression and machine learning. *Variance* **2021**, *14*, 1–22. [CrossRef]
16. Jeong, H.; Chang, H.; Valdez, E.A. A non-convex regularization approach for stable estimation of loss development factors. *Scand. Actuar. J.* **2021**, *2021*, 779–803. [CrossRef]
17. So, B.; Boucher, J.P.; Valdez, E.A. Synthetic dataset generation of driver telematics. *Risks* **2021**, *9*, 58. [CrossRef]
18. Brazauskas, V.; Jones, B.L.; Zitikis, R. Robust fitting of claim severity distributions and the method of trimmed moments. *J. Stat. Plan. Inference* **2009**, *139*, 2028–2043. [CrossRef]
19. Shi, P.; Basu, S.; Meyers, G.G. A Bayesian log-normal model for multivariate loss reserving. *N. Am. Actuar. J.* **2012**, *16*, 29–51. [CrossRef]
20. Jeong, H.; Dey, D. Application of a vine copula for multi-line insurance reserving. *Risks* **2020**, *8*, 111. [CrossRef]
21. Bae, T.; Miljkovic, T. Loss modeling with the size-biased lognormal mixture and the entropy regularized EM algorithm. *Insur. Math. Econ.* **2024**, *117*, 182–195. [CrossRef]
22. Abu-Mostafa, Y.S.; Magdon-Ismail, M.; Lin, H.T. *Learning from Data*; AMLBook: New York, NY, USA, 2012; Volume 4.
23. Ishwaran, H.; Rao, J.S. Spike and slab variable selection: Frequentist and Bayesian strategies. *Ann. Stat.* **2005**, *33*, 730–773. [CrossRef]
24. Narisetty, N.N.; He, X. Bayesian variable selection with shrinking and diffusing priors. *Ann. Stat.* **2014**, *42*, 789–817. [CrossRef]
25. Li, F.; Zhang, N.R. Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. *J. Am. Stat. Assoc.* **2010**, *105*, 1202–1214. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.