CAR PRICE PREDICTION WITH BAYESIAN GLM

— Project Report —
Advanced Bayesian Data Analysis

Group Number : 34
Sadat Mahmud
Md Mafidul Islam

March 17, 2024

TU Dortmund University

1 Introduction

The automotive industry is a cornerstone of modern society, driving economic growth and shaping consumer preferences. With the proliferation of data analytics and machine learning techniques, the ability to predict car prices accurately has become a vital endeavour for manufacturers, dealerships, and consumers alike (PredikData, 2023). In this project, we embark on a journey into Bayesian data analysis to tackle the challenge of car price prediction using Bayesian Generalized Linear Models (GLMs).

The primary motivation behind this project is to explore the versatility of Bayesian Generalized Linear Models (GLMs) in generating multiple predictive models for car price prediction. Our focus extends beyond mere prediction; we seek to delve into applied Bayesian data analysis, where the emphasis lies on making informed decisions based on probabilistic models. Drawing from the principles learned in applied Bayesian data analysis, we will carefully consider aspects such as choosing appropriate priors, handling numerical and categorical variables, and generating and comparing multiple models. Through this project, we aim to demonstrate the practical utility of Bayesian methodologies in real-world applications, particularly in the automotive industry. By comparing and contrasting different modelling approaches within the Bayesian framework, we seek to highlight the advantages and limitations of each method.

The problem revolves around accurately estimating car prices based on various numerical and categorical predictors. The challenge lies in developing predictive models that can effectively leverage this rich information source to produce reliable car price estimates. The dataset presents a complex modelling landscape with a combination of numerical and categorical attributes. The task is to capture the relationships between these variables and car prices accurately and to navigate the challenges posed by the diverse nature of the data. Developing robust predictive models that can effectively handle numerical and categorical variables is essential to address this challenge.

Our modelling approach revolves around Bayesian Generalized Linear Models (GLMs), which offer a versatile and principled framework for regression analysis. By incorporating Bayesian principles, we can seamlessly integrate prior knowledge, handle heterogeneous data types, and quantify uncertainty in our predictions. Furthermore, the flexibility of GLMs allows us to model complex relationships between predictors and response variables, accommodating both linear and nonlinear effects.

1.1 Data

The dataset utilized in this study originates from Kaggle, a widely recognized platform and online community dedicated to data science competitions and machine learning endeavours. The dataset, "Carprice_Assignment.csv," was crafted to unravel the intricate factors influencing car pricing (Kumar, 2020). It is designed to encompass variables deemed significant in predicting the price of automobiles, thereby providing a comprehensive insight into the dynamics of pricing in the automotive industry. The dataset comprises 205 observations and 26 variables, representing a diverse mix of numerical and

categorical data Of these variables, 16 are numerical, while the remaining 10 are categorical. This dataset does not have any missing values. The inclusion of both types of variables ensures a holistic representation of the myriad factors that contribute to car pricing.
In a previous analysis utilizing this dataset, the variables affecting car prices in the American market were explored. This study aimed to identify significant predictors of car prices and assess their explanatory power.

1.1.1 Data Preprocessing

In the initial phase of data preprocessing, the "carName" column was processed to extract only the brand names, as it contained a mixture of brand names and model numbers. This step aimed to ensure clarity and consistency in the dataset. Following this, the dataset was divided into numerical and categorical columns to facilitate distinct treatment based on the data type. Subsequently, numerical columns underwent standardization to ensure uniform scaling, enabling fair comparison among different features. Categorical columns were subjected to one-hot encoding, a technique employed to convert categorical variables into a numerical format suitable for analysis. Finally, the standardized numerical columns were merged with the encoded categorical columns, resulting in a comprehensive and prepared dataset conducive to subsequent analysis. These preprocessing steps were pivotal in enhancing the dataset's suitability for comprehensive analysis of factors influencing car prices in the American market.

2 Models

In this report, we analyze the factors influencing car prices using Generalized linear model. The motivation behind this study is to understand how various features of cars, such as wheelbase, engine specifications, and brand, contribute to pricing. In this report, three distinct generalized linear regression models were generated, each incorporating a different car brand as a predictor variable.

2.0.1 Model 1

Model 1 was formulated to investigate the pricing determinants of automobiles, specifically focusing on the brand "BMW". Utilizing the Generalized Linear Model (GLM) framework with a Gaussian family distribution, our analysis revealed several significant predictors of car prices. Notably, brand indicators such as 'BMW' and the use of 'gas' as the fuel type demonstrated substantial effects on pricing. Additionally, factors such as engine size, car width, and curb weight exhibited notable impacts on car prices.

Here is the model 1 using the value from Table 1 equation:

$$\begin{aligned}\text{Price} = 0.26 &- 0.03 \times \text{wheelbase} + 0.98 \times \text{carnamebmw} - 0.33 \times \text{fueltypegas}\\ &-0.14 \times \text{carlength} + 0.24 \times \text{carwidth} + 0.06 \times \text{carheight}\\ &+0.21 \times \text{curbweight} + 0.50 \times \text{enginesize} + 0.16 \times \text{horsepower}\\ &+0.14 \times \text{peakrpm} + \epsilon\end{aligned}$$

Table 1: Population-Level Effects for Model 1

Variable	**Estimate**	**95% CI Lower**	**95% CI Upper**
Intercept	0.26	0.05	0.47
wheelbase	-0.03	-0.17	0.11
carnamebmw	0.98	0.69	1.27
fueltypegas	-0.33	-0.55	-0.11
carlength	-0.14	-0.29	0.01
carwidth	0.24	0.11	0.36
carheight	0.07	-0.01	0.14
curbweight	0.20	0.01	0.40
enginesize	0.50	0.37	0.62
horsepower	0.16	0.03	0.29
peakrpm	0.14	0.07	0.22

Table 2: Family Specific Parameter for Model 1

Parameter	**Estimate**	**95% CI Lower**	**95% CI Upper**
σ	0.37	0.34	0.41

2.0.2 Model 2

Model 2 has been constructed to analyze the pricing determinants of automobiles, focusing specifically on the Audi brand. Utilizing the Generalized Linear Model (GLM) framework with a Gaussian family distribution, the coefficient estimate for engine size is 0.55, signifying a statistically significant positive correlation with price. This implies that vehicles with larger engine sizes tend to command higher prices.
Here is the model 2 equation Using Table 3 below:

$$\begin{aligned} \text{Price} = 0.30 &+ 0.04 \times \text{wheelbase} + 0.18 \times \text{carnameaudi} - 0.34 \times \text{fueltypegas} \\ &- 0.10 \times \text{carlength} + 0.11 \times \text{carwidth} + 0.07 \times \text{carheight} \\ &+ 0.15 \times \text{curbweight} + 0.55 \times \text{enginesize} + 0.22 \times \text{horsepower} \\ &+ 0.13 \times \text{peakrpm} + \epsilon \end{aligned}$$

Table 3: Family Specific Parameters for Model 2

Parameter	**Estimate**	**Est. Error**	**95% CI**
Sigma	0.42	0.02	(0.38, 0.46)

Table 4: Population-Level Effects for Model 2

Parameter	**Estimate**	**Est. Error**	**95% CI**
Intercept	0.30	0.12	(0.07, 0.53)
Wheelbase	0.04	0.08	(-0.11, 0.19)
Car Name (Audi)	0.18	0.18	(-0.17, 0.53)
Fuel Type (Gas)	-0.34	0.12	(-0.57, -0.09)
Car Length	-0.10	0.08	(-0.26, 0.07)
Car Width	0.11	0.07	(-0.03, 0.25)
Car Height	0.07	0.04	(-0.02, 0.15)
Curb Weight	0.15	0.10	(-0.06, 0.35)
Engine Size	0.55	0.07	(0.41, 0.69)
Horsepower	0.22	0.07	(0.08, 0.37)
Peak RPM	0.13	0.04	(0.06, 0.21)

2.0.3 Model 3

Model 3 was devised to investigate the pricing determinants within the context of the Toyota car brand. Employing the GLM framework with a Gaussian family distribution, the model yielded a coefficient estimate of 0.53 for engine size, indicating a significant positive correlation with price. This suggests that vehicles featuring larger engines generally incur higher prices.
Here is the model 3 equation Using Table 5 below:

$$\begin{aligned}\text{Price} = 0.31 &+ 0.04 \times \text{wheelbase} - 0.14 \times \text{carnametoyota} - 0.31 \times \text{fueltypegas}\\ &-0.09 \times \text{carlength} + 0.12 \times \text{carwidth} + 0.06 \times \text{carheight}\\ &+0.14 \times \text{curbweight} + 0.53 \times \text{enginesize} + 0.23 \times \text{horsepower}\\ &+0.12 \times \text{peakrpm} + \epsilon\end{aligned}$$

Table 5: Family Specific Parameters for Model 3

Parameter	**Estimate**	**Est. Error**	**Lower 95% CI**	**Upper 95% CI**
Sigma	0.41	0.02	0.38	0.45

Table 6: Population-Level Effects for Model 3

Parameter	**Estimate**	**Est. Error**	**Lower 95% CI**	**Upper 95% CI**
Intercept	0.31	0.12	0.08	0.54
Wheelbase	0.04	0.08	-0.11	0.19
Carname Toyota	-0.14	0.09	-0.31	0.03
Fuel Type Gas	-0.31	0.12	-0.57	-0.07
Car Length	-0.09	0.08	-0.24	0.08
Car Width	0.12	0.07	-0.02	0.25
Car Height	0.06	0.04	-0.02	0.15
Curb Weight	0.14	0.10	-0.06	0.34
Engine Size	0.53	0.07	0.39	0.67
Horsepower	0.23	0.08	0.08	0.38
Peak RPM	0.12	0.04	0.04	0.20

2.0.4 Formula Interpretation

The formula predicts the price of an automobile based on several factors including wheelbase, car dimensions (length, width, height), curb weight, engine size, horsepower, peak RPM, and certain categorical variables such as the car being a Toyota and the fuel type being gas. Each coefficient in the formula represents the estimated effect of the corresponding predictor variable on the price of the car. For instance, a positive coefficient indicates that as the value of the predictor variable increases, the price of the car tends to increase, while a negative coefficient suggests the opposite relationship. The error term (ϵ) accounts for unexplained variability in car prices that is not captured by the predictor variables included in the model.

2.1 Prior

For all models, priors were explicitly specified to provide a clear framework for parameter estimation. The choice of prior distributions reflects a balance between incorporating existing knowledge and allowing the data to influence the parameter estimates. A normal distribution with a mean of 0 and a standard deviation of 3 was chosen as the prior distribution for all regression coefficients (class = ”b”). This choice reflects a neutral stance, assuming no strong prior beliefs about the magnitude of the coefficients. The selection of the normal distribution prior was motivated by its flexibility and simplicity. While centered at 0, allowing for unbiased estimation, the standard deviation of 3 provides sufficient flexibility to accommodate a wide range of plausible parameter values without overly constraining the model. It’s important to acknowledge that priors are inherently subjective and may introduce bias into the analysis. However, by explicitly specifying the priors, we aimed to provide transparency and consistency across all models, enabling a more interpretable and reproducible analysis.

2.2 Code

The statistical analysis for this project was conducted within the R programming language environment, version 4.2.2 (R Core Team, 2022). R served as the primary platform for data processing, statistical modeling, and visualization tasks, owing to its extensive range of packages and libraries tailored for various analytical needs. Key libraries employed in this project include brms (Bürkner, 2017) for fitting Bayesian multilevel models using Stan, loo (Vehtari et al., 2023) for efficient leave-one-out cross-validation and WAIC computation for Bayesian models, ggplot2 (Wickham, 2016) for creating elegant and customizable visualizations, rstan (Stan Development Team, 2024) for interfacing with Stan, a probabilistic programming language for Bayesian inference, and dplyr (Wickham et al., 2023) for data manipulation and transformation tasks. These libraries collectively provided the necessary tools and functionalities for conducting rigorous statistical analysis, model fitting, visualization, and data manipulation, thereby contributing to the successful execution of the project objectives.

2.2.1 Brief Explanation

We present Bayesian regression models aimed at predicting car prices for different brands, namely BMW, Audi, and Toyota. Each model considers a set of independent variables including wheelbase, car dimensions (length, width, height), curb weight, engine size, horsepower, peak RPM, and fuel type (gas). The dependent variable in all models is the price of the car. The first model focuses on BMW cars, incorporating a binary indicator for BMW, while the second and third models similarly include indicators for Audi and Toyota, respectively. These models employ Bayesian regression techniques to estimate car prices, accounting for uncertainty and leveraging prior knowledge about the relationships between the dependent and independent variables.

Listing 1: Bayesian Regression Model for BMW Cars

```
carprice_regr = brm(price ~ wheelbase + carnamebmw + fueltypegas +
   carlength + carwidth + carheight + curbweight + enginesize +
   horsepower + peakrpm, data = final_data, family = gaussian(), save_
   pars = save_pars(all = TRUE), set_prior("normal(0,3)", class = "b"))
```

Listing 2: Bayesian Regression Model for Audi Cars

```
carprice_regr2 = brm(price ~ wheelbase + carnameaudi + fueltypegas +
   carlength + carwidth + carheight + curbweight + enginesize +
   horsepower + peakrpm, data = final_data, family = gaussian(), save_
   pars = save_pars(all = TRUE), set_prior("normal(0,3)", class = "b"))
```

Listing 3: Bayesian Regression Model for Toyota Cars

```
carprice_regr3 = brm(price ~ wheelbase + carnametoyota + fueltypegas +
   carlength + carwidth + carheight + curbweight + enginesize +
   horsepower + peakrpm, data = final_data, family = gaussian(), save_
   pars = save_pars(all = TRUE), set_prior("normal(0,3)", class = "b"))
```

2.2.2 Explicit Parameter Choices

All three models employ a Gaussian likelihood function (`family = gaussian()`) and specify a normal prior distribution with a mean of 0 and a standard deviation of 3 for all coefficients (`set_prior("normal(0,3)", class = "b")`). The settings for the models, including the likelihood function and prior distribution, remain consistent with those utilized in Model 1. We executed 2000 iterations per chain with a warm-up period of 1000 iterations, ensuring convergence and effective parameter estimation. This moderate number of iterations permits the model to sufficiently learn from the data without succumbing to overfitting. This methodological approach is designed to capture the underlying patterns present in the data while mitigating excessive computational costs.

Listing 4: Small extract of the output of interation and warmup

```
Chain 1: Adjust your expectations accordingly!
Chain 1:
Chain 1:
Chain 1: Iteration:    1 / 2000 [  0%]  (Warmup)
Chain 1: Iteration:  200 / 2000 [ 10%]  (Warmup)
Chain 1: Iteration:  400 / 2000 [ 20%]  (Warmup)
Chain 1: Iteration: 1001 / 2000 [ 50%]  (Sampling)
Chain 1: Iteration: 1200 / 2000 [ 60%]  (Sampling)
Chain 1: Iteration: 1400 / 2000 [ 70%]  (Sampling)
Chain 1: Iteration: 1600 / 2000 [ 80%]  (Sampling)
```

2.3 Convergence Diagnostics

We conducted convergence diagnostics using the Bayesian framework implemented in the `brms` package for all the models. The convergence diagnostics were performed on the Bayesian regression model designed for predicting car prices. The model specification comprised a Gaussian likelihood function and a normal prior distribution with a mean of 0 and a standard deviation of 3 for all coefficients. The model was fitted to the data with 2000 iterations per chain and a warm-up period of 1000 iterations using 4 CPU cores. The convergence diagnostics provided valuable insights into the performance of the Bayesian regression model. Through visual examination of Figure 1, Figure 2, and Figure 3 density plots and trace plots, we assessed the Markov chains' convergence, confirming their stationarity attainment. Additionally, effective sample size measures (Bulk_ESS and Tail_ESS) were employed to evaluate the efficiency of parameter estimation. The potential scale reduction factor ($\hat{R}$) approached 1 for all parameters, indicating convergence. Overall, the convergence diagnostics validated that the model achieved reliable parameter estimation and offered valid inference for predicting car prices based on the specified predictors. A summary of the convergence diagnostics is presented in tables 7, 8, 9, 10, 11, and 12 in the appendix section.

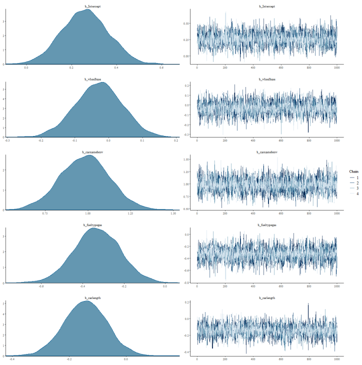

Figure 1: Density plot of Convergence Diagnostics (1)

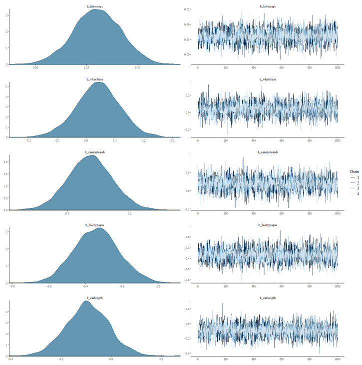

Figure 2: Density plot of Convergence Diagnostics (2)

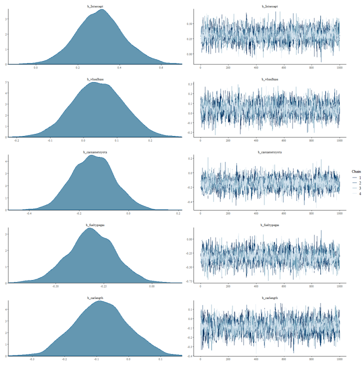

Figure 3: Density plot of Convergence Diagnostics (3)

2.4 Posterior Predictive Checks and model re-iteration

In the context of our analysis, posterior prediction was applied to our car price regression model to generate simulated outcomes for the target variable (car prices) based on the estimated model parameters and uncertainty. The posterior prediction process was executed using the `posterior_predict()` function in R, which yields a matrix of simulated outcomes for each observation in the dataset. In this matrix, each row represents a different observation, while each column represents a simulated outcome. Specifically, the `pp` variable contains the posterior predictive values generated by the regression model. It is structured as a matrix with dimensions [4, 200], indicating 4 rows (possibly corresponding to distinct observations or samples) and 200 columns (likely representing various predictions or iterations). Each entry in the matrix denotes a simulated value for a particular observation or prediction. These simulated values are drawn from the posterior predictive distribution, encompassing both parameter uncertainty and data variability.

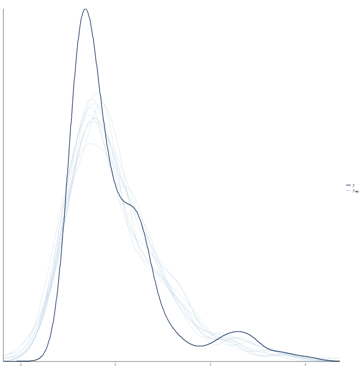

Figure 4: Posterior predictive checks for Model 2

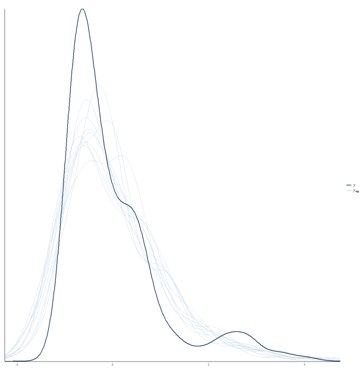

Figure 5: Posterior predictive checks for Model 1

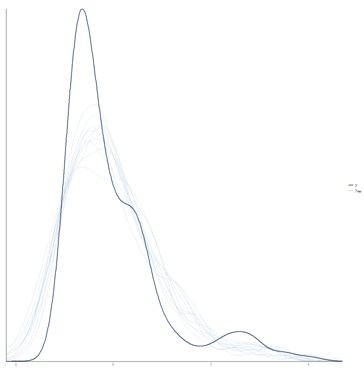

Figure 6: Posterior predictive checks for Model 3

2.5 Model comparison

In this section, we compare the performance of three different regression models developed for predicting car prices. The models under consideration are denoted as `carprice_regr`, `carprice_regr2`, and `carprice_regr3`.

2.5.1 Bayes Factor Analysis

Bayes factor analysis provides a quantitative measure to compare the relative strength of evidence between pairs of models. The estimated Bayes factors for pairwise model comparisons are as follows:

- Bayes factor in favor of `carprice_regr` over `carprice_regr2`: 634683743.94522
- Bayes factor in favor of `carprice_regr` over `carprice_regr3`: 530589955.52914
- Bayes factor in favor of `carprice_regr2` over `carprice_regr3`: 0.82389

These results suggest strong evidence in favor of `carprice_regr` over both `carprice_regr2` and `carprice_regr3`, indicating that `carprice_regr` performs better in predicting car prices compared to the other two models. Additionally, there is weak evidence in favor of `carprice_regr2` over `carprice_regr3`.

2.5.2 Widely Applicable Information Criterion (WAIC)

The WAIC provides a measure of model fit and complexity, with lower values indicating better-performing models. The WAIC estimates for each model are presented below:

- **Model** `carprice_regr`:
  - Effective Log Pointwise Predictive Density (elpd_waic): -98.2 (SE: 15.5)
  - Penalty term (p_waic): 17.2 (SE: 3.7)
  - WAIC: 196.4 (SE: 31.1)
  - Percentage of *p_waic* estimates greater than 0.4: 4.9%
- **Model** `carprice_regr2`:
  - elpd_waic: -118.7 (SE: 15.0)
  - p_waic: 15.0 (SE: 2.7)
  - WAIC: 237.3 (SE: 30.0)
  - Percentage of *p_waic* estimates greater than 0.4: 3.4%
- **Model** `carprice_regr3`:
  - elpd_waic: -117.8 (SE: 14.9)

- p_waic: 15.2 (SE: 2.7)
- WAIC: 235.5 (SE: 29.7)
- Percentage of *p_waic* estimates greater than 0.4: 3.9%

The WAIC analysis indicates that `carprice_regr` has the lowest WAIC value, suggesting superior model fit and complexity compared to `carprice_regr2` and `carprice_regr3`. Both `carprice_regr2` and `carprice_regr3` exhibit slightly higher WAIC values, indicating comparatively poorer performance in terms of model fit and complexity.

2.6 Limitation and Potential Improvements

Our project exhibits promising results, yet enhancements are feasible in several aspects. Although we diligently preprocessed the data, potential inconsistencies may persist, thereby influencing model accuracy. Acquiring a larger, high-quality dataset holds the potential to bolster the robustness of our analyses. Assumptions underpinning our analysis, including linearity, and independence are critical. Violations of these assumptions could compromise the validity of our findings. Addressing this issue entails conducting comprehensive diagnostic checks and exploring alternative modeling approaches that relax these assumptions.
The regression models employed may inadequately capture the intricacies inherent in the relationship between car features and prices. Delving into advanced modeling techniques, such as ensemble methods or deep learning architectures, presents an avenue for potentially enhancing predictive accuracy. Moreover, concerns regarding model generalization to unseen data or disparate geographical regions persist. Assessing model generalization across external datasets or diverse contexts could elucidate the models' robustness and applicability in real-world scenarios.

2.7 Conclusion

In conclusion, our study endeavors to elucidate the dynamics influencing car prices through robust regression modeling techniques. Despite certain limitations, including potential data inconsistencies and assumptions underlying our analyses, our findings offer valuable insights into the intricate relationship between car features and prices. Ultimately, this project contributes to the broader discourse on pricing dynamics in the automotive industry.

2.8 Reflection on own learnings

In reflecting on our learnings from this project within the framework of Applied Bayesian Data Analysis, several key insights emerge. The project has reinforced the importance of rigorous data preprocessing and model diagnostics in Bayesian analysis. This project has underscored the significance of Bayesian model comparison techniques in evaluating alternative hypotheses and selecting the most appropriate model for the data. By leveraging methods such as Bayes factors and information criteria like WAIC, we have learned to quantify the relative strengths of competing models and make informed decisions. Additionally, our engagement with this project has fostered a deeper appreciation for the iterative nature of Bayesian analysis, where model refinement and validation are ongoing processes. Embracing uncertainty and iteratively refining our models based on new insights and data is fundamental to the Bayesian paradigm. This project has provided valuable experience in communicating complex analytical findings effectively. Through the process of preparing reports and presentations, we have honed our skills in translating technical concepts into accessible insights for diverse audiences. Overall, the journey through this project has been instrumental in consolidating our understanding of Bayesian data analysis principles and methodologies.

References

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Bürkner, P.-C. (2017). brms: An R package for bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Kumar, M. (2020). *Car price prediction dataset*. https://www.kaggle.com/datasets/hellbuoy/car-price-prediction

R Core Team. (2022). *R: A language and environment for statistical computing* [R version 4.2.2 (2022-10-31)]. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

PredikData. (2023). Predictive analytics in the automotive industry [Accessed: 2024-03-07].

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., & Gelman, A. (2023). *loo: Efficient leave-one-out cross-validation and waic for bayesian models* [R package version 2.6.0]. https://mc-stan.org/loo/

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation* [R package version 1.1.4]. https://CRAN.R-project.org/package=dplyr

Stan Development Team. (2024). RStan: The R interface to Stan [R package version 2.32.5]. https://mc-stan.org/

3 Appendix

3.1 A Additional Tables

Table 7: Summary Table of Convergence Diagnostics for Model 1

Variable	**Estimate**	**Est. Error**	**l-95% CI**	**u-95% CI**	**Rhat**	**Tail ESS**
Intercept	0.26	0.10	0.06	0.47	1.00	3044
wheelbase	-0.03	0.07	-0.17	0.11	1.00	2743
carnamebmw	0.98	0.14	0.71	1.26	1.00	2902
fueltypegas	-0.33	0.11	-0.55	-0.12	1.00	3221
carlength	-0.14	0.08	-0.29	0.00	1.00	2736
carwidth	0.24	0.06	0.11	0.37	1.00	2990
carheight	0.07	0.04	-0.01	0.14	1.00	2763
curbweight	0.20	0.09	0.02	0.38	1.00	2763
enginesize	0.50	0.06	0.37	0.63	1.00	3147
horsepower	0.16	0.07	0.02	0.29	1.00	2957
peakrpm	0.14	0.04	0.07	0.21	1.00	3029

Table 8: Family Specific Parameters for table 1

Parameter	**Estimate**	**Est. Error**	**l-95% CI**	**u-95% CI**	**Rhat**	**Tail ESS**
Sigma	0.37	0.02	0.34	0.41	1.00	2688

Table 9: Summary Table of Convergence Diagnostics for Model 2

Variable	**Estimate**	**Est. Error**	**l-95% CI**	**u-95% CI**	**Rhat**	**Tail ESS**
Intercept	0.30	0.12	0.08	0.53	1.00	2924
wheelbase	0.04	0.08	-0.10	0.20	1.00	2631
carnameaudi	0.17	0.18	-0.18	0.53	1.00	2963
fueltypegas	-0.34	0.12	-0.58	-0.10	1.00	2819
carlength	-0.09	0.08	-0.26	0.07	1.00	3016
carwidth	0.11	0.07	-0.03	0.26	1.00	2976
carheight	0.07	0.04	-0.01	0.15	1.00	3115
curbweight	0.14	0.11	-0.05	0.35	1.00	2830
enginesize	0.55	0.07	0.41	0.69	1.00	2746
horsepower	0.22	0.08	0.08	0.37	1.00	3013
peakrpm	0.14	0.04	0.06	0.21	1.00	2912

Table 10: Family Specific Parameters for Model 2

Variable	**Estimate**	**Est. Error**	**l-95% CI**	**u-95% CI**	**Rhat**	**Tail ESS**
Sigma	0.42	0.02	0.38	0.46	1.00	2803

Table 11: Summary Table of Convergence Diagnostics for Model 3

Variable	**Estimate**	**Est. Error**	**l-95% CI**	**u-95% CI**	**Rhat**	**Tail ESS**
Intercept	0.30	0.11	0.08	0.53	1.00	3424
wheelbase	0.04	0.08	-0.11	0.19	1.00	2903
carnametoyota	-0.14	0.08	-0.30	0.02	1.00	3166
fueltypegas	-0.31	0.12	-0.55	-0.07	1.00	3164
carlength	-0.09	0.08	-0.25	0.07	1.00	2745
carwidth	0.12	0.07	-0.01	0.25	1.00	2916
carheight	0.06	0.04	-0.02	0.15	1.00	2990
curbweight	0.14	0.10	-0.06	0.34	1.00	2774
enginesize	0.53	0.07	0.40	0.67	1.00	3083
horsepower	0.23	0.08	0.08	0.38	1.00	2622
peakrpm	0.12	0.04	0.04	0.20	1.00	2954

Table 12: Family Specific Parameters for Model 3

Variable	**Estimate**	**Est. Error**	**l-95% CI**	**u-95% CI**	**Rhat**	**Tail ESS**
Sigma	0.41	0.02	0.38	0.46	1.00	2948

3.2 B Additional Figures

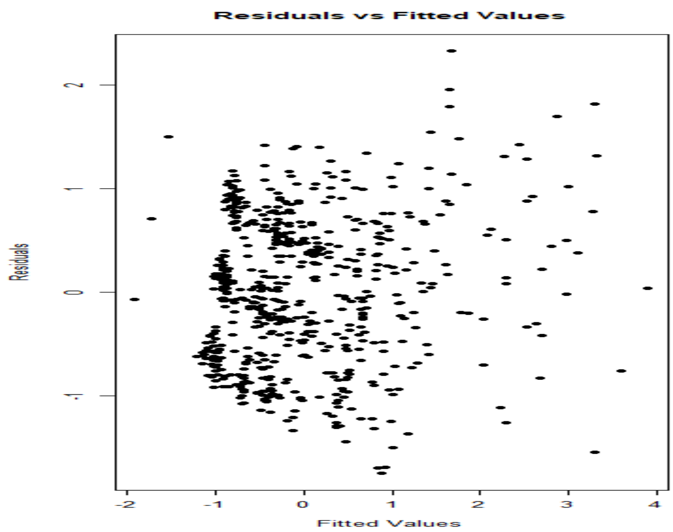

Figure 7: Residual Plot for model 1

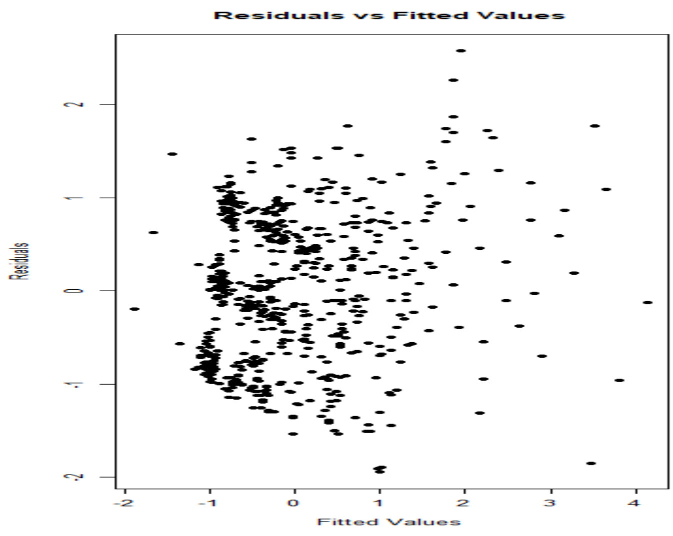

Figure 8: Residual Plot for model 2

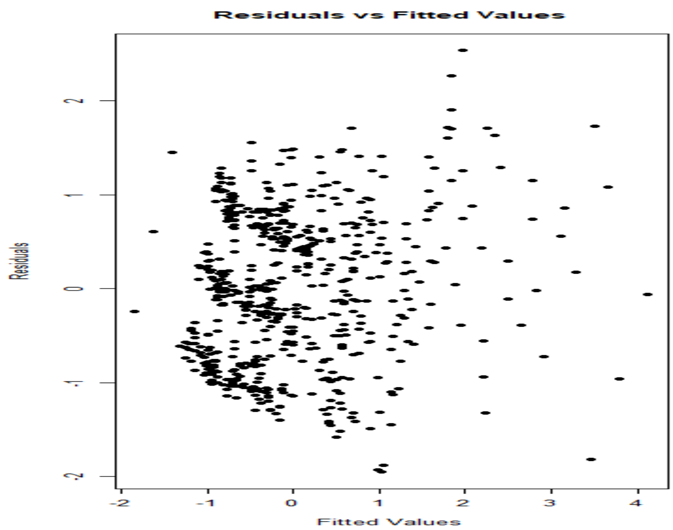

Figure 9: Residual Plot for model 3