conclusionSection

Trevor Isaacson

2022-10-12

```
set.seed(551)
```

```
##   Price Age Mfg_Month Mfg_Year    KM Fuel_Type HP Metallic Automatic   CC Doors
## 1 13500  23        10     2002 46986    Diesel 90        1         0 2000     3
## 2 13750  23        10     2002 72937    Diesel 90        1         0 2000     3
## 3 13950  24         9     2002 41711    Diesel 90        1         0 2000     3
## 4 14950  26         7     2002 48000    Diesel 90        0         0 2000     3
## 5 13750  30         3     2002 38500    Diesel 90        0         0 2000     3
## 6 12950  32         1     2002 61000    Diesel 90        0         0 2000     3
##   Gears QuartTax Weight Guarantee BOVAG Period
## 1     5      210   1165         0     1      3
## 2     5      210   1165         0     1      3
## 3     5      210   1165         1     1      3
## 4     5      210   1165         1     1      3
## 5     5      210   1170         1     1      3
## 6     5      210   1170         0     1      3
```

Scale data and log transform `Price`.

```
fit4 <- stan_glm(Price ~ Age + Mfg_Month + Mfg_Year + KM + Fuel_Type + HP +
                   Metallic + QuartTax + Weight + Guarantee + BOVAG + Period,
                 data = data_scale,
                 refresh = 0,
                 iter = 5000)
```

Conclusion

Final Model and Fitted Equation

For our final model, we decided to use model 4. This model is constructed using horse-shoe selected predictors with the scaled data and default priors. The horseshoe selected predictors include `Age`, `Mfg_Month`, `Mfg_Year`, `KM`, `Fuel_Type`, `HP`, `Metallic`, `QuartTax`, `Weight`, `Guarantee`, `BOVAG`, and `Period`. Also, using the scaled numeric predictors and a log-transformation of car price to restrict prediction prices to positive values only, this model is able to better predict selling price compared to other models.

Tables of Estimated Coefficients/Standard Errors

```
print(fit4, digits = 3)
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      Price ~ Age + Mfg_Month + Mfg_Year + KM + Fuel_Type + HP + Metallic +
##     QuartTax + Weight + Guarantee + BOVAG + Period
##  observations: 694
##  predictors:   28
## ------
##                 Median MAD_SD
## (Intercept)      9.169  0.643
## Age             -0.120  0.379
## Mfg_Month2       0.022  0.028
## Mfg_Month3       0.033  0.051
## Mfg_Month4       0.012  0.075
## Mfg_Month5       0.034  0.097
## Mfg_Month6       0.017  0.122
## Mfg_Month7      -0.003  0.146
## Mfg_Month8       0.014  0.169
## Mfg_Month9      -0.039  0.194
## Mfg_Month10     -0.027  0.218
## Mfg_Month11     -0.030  0.242
## Mfg_Month12     -0.002  0.267
## Mfg_Year2000    -0.001  0.289
## Mfg_Year2001    -0.013  0.577
## Mfg_Year2002     0.100  0.862
## Mfg_Year2003     0.131  1.153
## Mfg_Year2004     0.102  1.447
## KM              -0.078  0.007
## Fuel_TypeDiesel  0.030  0.034
## Fuel_TypePetrol  0.171  0.035
## HP               0.050  0.006
## Metallic1        0.022  0.009
## QuartTax         0.076  0.009
## Weight           0.040  0.008
## Guarantee1       0.030  0.009
## BOVAG1           0.050  0.015
## Period           0.015  0.005
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 0.102  0.003
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

Interpretation and Discussion

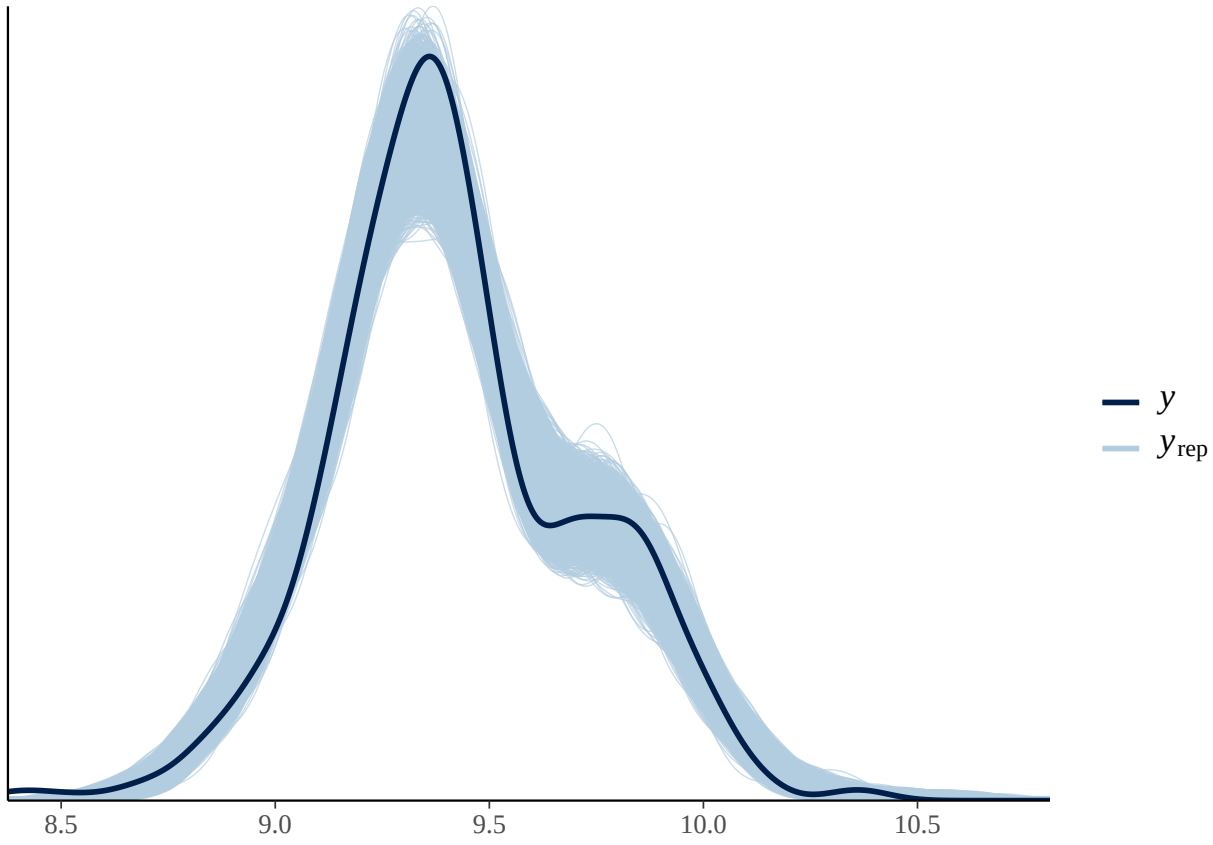
The intercept coefficient can be interpreted as the expected log(price) of a non-metallic, CNG-fueled car that was manufactured in January of 1999 with no manufacturer or BOVAG guarantee and a value of 0 for all numeric predictors included in the model. The coefficients estimates for the numeric predictors can be

interpreted as the expected percent difference in predicted car price associated with a one standard deviation increase in the value of the predictor, with all other predictors remaining constant. Overall, this model had an in-sample Bayesian R^2 value of 0.882 and a leave one out adjusted R^2 value of 0.87. This confirms our model isn't overfitting while also showing a relatively high R^2 value.

Predictive Plots

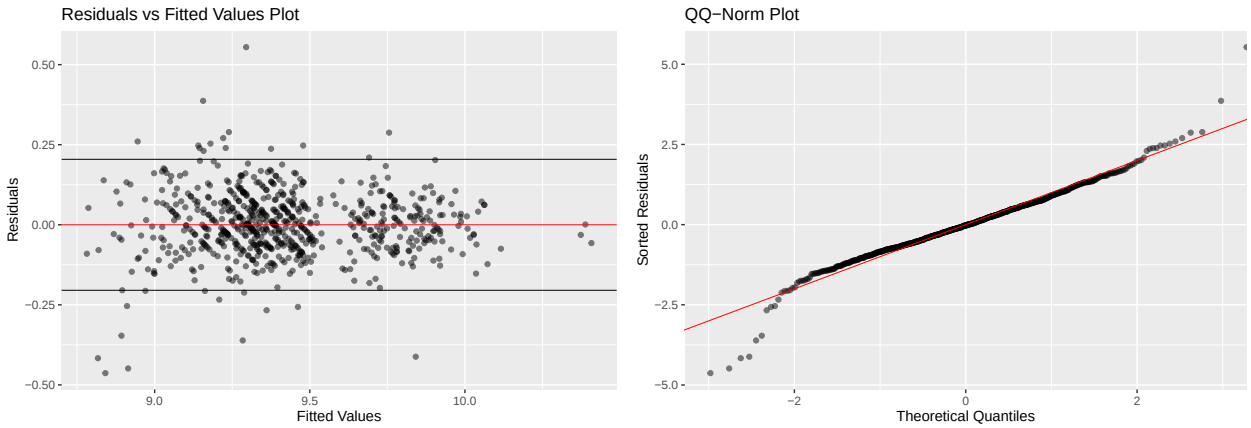
The final fitted model should look like our data. By drawing from the predictive distribution and comparing it to the distribution of the response variable, we can determine if our model is fitting appropriately. In the plot below, we see our predictive distribution tracks the response distribution well increasing our confidence in the model.

```
fit4_rep = posterior_predict(fit4)
ppc_dens_overlay(data_scale$Price, fit4_rep) + scale_y_continuous(breaks=NULL)
```



Check your assumptions

Checking the assumptions of our final model, we find the model includes all the relevant predictors as we chose these using the horseshoe prior and forcing predictors to be highly related with the response. The outcome measure accurately reflects our prediction interest and is generalized to all Toyota Corollas. Looking at the residual plots, we see there are no patterns and trends within the residuals vs fitted values plot. Most values are within 2 standard residuals of 0 and the values are spread across the 0 line. There might be some clumping but nothing big enough to question the model. There aren't any heavy tails in the qq-norm plot and the values closely align with the red line.



Others results (as appropriate)

Because the goal of this study is prediction, our final model has great observed predictive powers using leave one out cross validation and also k-fold cross validation. Using

```
# Put smaller table cv results here
```

Refer back to the purpose of the study

In all, the purpose of this study was to predict the selling price of used Toyota Corollas and ensure a small profit based on their new purchase and trade-in promotion. Based on several variables, we were able to fit this model to help the dealership closely estimate the final selling price for their used cars. With this model, the dealership can now ensure a reasonable profit by plugging in the characteristics of each individual car into this model and output a predicted selling price. This will result in more accurate selling prices and higher profits for the dealer.