Used Car Price Prediction for Tesla Model 3 LR in California

BIOSTAT 234: Final Data Analysis Project Abstract

Name: Kuan-Hung (Peter) Yeh UID:705669455

March 7th, 2023

App Demonstration:

Please go to this URL for the Tesla Model3 Long Range Price Prediction app interface constructed based on R shiny app: Tesla Model3 Long Range Price Prediction.

Introduction:

In this study, I'll first focus on estimating the used car price for Tesla Model 3 Long Range AWD in California. Secondly, I'll determine which factors have an impact on the used car price and what is the impact for those predictors by summarizing their posterior. Lastly, I'll make the prediction of the estimated price based on some car conditions and the software capabilities to help people make decisions on whether the current price listed on the website is a good deal.

Dataset:

I created a web scraping tool using Python in Jupyter Notebook to extract information such as the used car price, vehicle age, mileage, color, damage history, and other variables from an official website. To initiate the search for used cars, a zip code is required, so we used the zip codes of the top five most populous cities in California: Los Angeles (90024), San Diego (92101), San Jose (95110), San Francisco (94102), and Fresno (93650). Since Tesla Model 3 LR is a rare used car on the website and each search yields at least 40 results, we found that choosing any zip code from each city gives the same search results. For instance, if we enter a zip code in LA and it displays cars from Fresno and Carlsbad (SD), it implies that there aren't many cars available in LA and hence the search results are from other cities.

On February 8th, after scraping the website, we collected the datasets for each city in CSV (Comma Separated Values) formatted files. These datasets serve as prior information, and on March 2nd, we scraped the website again to conduct a Bayesian analysis and predict the car price based on the prior information we collected in February.

Description of Dataset:

Because the vehicles in each city's dataset may appear multiple times, we deleted the observations that appear more than once in our dataset after we merged them. The dataset is labeled properly and there are 10 variables we scrapped from the website that help us predict the car price. The sample size for unique vehicles is 61 in the dataset from February and 67 from March. Due to the length of this article, the detailed variables description is attached in Appendix.

Prior and Model specification:

Denote the Price for the i^{th} vehicle by y_i. We will use multiple linear regression to model this data. The proposed model is:

$$
\begin{aligned}
y_i =& \beta_0 + \beta_1 Blue_i + \beta_2 Sliver_i + \beta_3 Black_i + \beta_4 Red_i + \beta_5 Interior_i + \beta_6 SoftwareEAP_i && (1)\\
&+ \beta_7 SoftwareFSD_i + \beta_8 Damage_i + \beta_9 Mileage_i + \beta_{10} Year'20_i + \beta_{11} Year'21_i && (2)\\
&+ \beta_{12} Year'22_i + \beta_{13} Wheel_i + \epsilon_i && (3)
\end{aligned}
$$

The error terms $\epsilon_i|\sigma^2 \sim N(0, \sigma^2)$ are a priori independent and identically distributed given σ^2

	Estimate	Std. Error	t value	$\Pr(\geq \lvert t \rvert)$
(Intercept)	41712.60	641.63	65.01	0.00
Deep Blue Metallic Paint	-1686.59	691.52	-2.44	0.02
Midnight Silver Metallic Paint	-129.51	395.38	-0.33	0.74
Red Multi-Coat Paint	433.22	513.86	0.84	0.40
Solid Black Paint	888.47	450.62	1.97	0.05
Black and White Premium Interior	521.07	401.08	1.30	0.20
Enhanced Autopilot	2212.74	413.26	5.35	0.00
Full Self-Driving Capability	4405.74	411.44	10.71	0.00
Previously Repaired	-1892.43	543.05	-3.48	0.00
Mileage	-0.12	0.02	-8.06	0.00
Year2020	1837.76	440.81	4.17	0.00
Year2021	6150.33	1004.46	6.12	0.00
Wheels19	294.10	305.02	0.96	0.34
σ^2	1,199,368			

Table 1: OLS estimators from the dataset in February.

Based on the information and the pricing listing for buying a new car from Tesla's official website and the OLS estimators' information from the old dataset in February, we have some prior information for setting the prior like the different colors, software capacities, wheel size, etc. We set the prior mean and standard errors similar to the reported estimation from the old dataset. We use independent priors to do the analysis and assume all β_i, $i = 0 \sim 12$ to be normal and σ^2 be Inverse Gamma. The priors are:

- $\beta_0 \sim N(40,000, 2.5 * 10^{-6})$ (Intercept)

- $\beta_1 \sim N(-1600, 4*10^{-6})$ (The average price difference between Blue and Pearl White(ref))
- $\beta_2 \sim N(-130, 10^{-5})$ (The average price difference between Sliver and Pearl White(ref))
- $\beta_3 \sim N(900, 10^{-6})$ (The average price difference between Black and Pearl White(ref))
- $\beta_4 \sim N(450, 10^{-6})$ (The average price difference between Red and Pearl White(ref))
- $\beta_5 \sim N(500, 6{*}10^{-6})$ (The average price difference between White Interior and All Black(ref))
- $\beta_6 \sim N(2200, 6{*}10^{-6})$ (Average price difference for EnhancedAutopilot compare to Autopilot)
- $\beta_7 \sim N(4400, 6*10^{-6})$ (Average price difference for Full Self-Driving compare to Autopilot)
- $\beta_8 \sim N(-2000, 4*10^{-6})$ (The average price difference for previously repaired vehicle)
- $\beta_9 \sim N(-0.12, 2, 500)$ (Average price change when increasing one mileage in odometer)
- $\beta_{10} \sim N(1800, 4*10^{-4})$ (The average price difference for 2020 compared to 2019)
- $\beta_{11} \sim N(6150, 10^{-6})$ (The average price difference for 2021 compared to 2019)
- $\beta_{12} \sim N(6500, 10^{-6})$ (The average price difference for 2022 compared to 2019)
- $\beta_{13} \sim N(300, 10^{-5})$ (The average price difference for previously repaired vehicle)
- $\tau \sim Gamma(24, 2, 800, 000)$

For τ, we assumed that it is Gamma(24, 2,800,000), with a mean 24/2,800,000 $\approx$ 1/1,200,000. The number 24 = 48/2 is one-half the degrees of freedom for estimating. This prior weights equally from the information in the old dataset and the analysis dataset.

Results and discussion:

The model was executed in JAGS, with 10000 iterations, out of which the first 5000 were discarded as burn-in. It was created over five chains and applied a thinning parameter of 5 to remove some of the autocorrelations at adjacent lags. In the prior setting, it can be observed that I did not include the Year and Range variables in the model simultaneously. This is because I discovered that the EPA-estimated range for Tesla vehicles on the website is based on the year of the vehicle. Hence, the Year variable has a linear relationship with Range, and I chose to include the Year in the model alone.

In the Exploring Data Analysis section (Figure 1 - 8), scatter plots or box plots were used to depict the relationship between price and other variables such as mileage, color, and software, among others. The trends for different variables on price are logical and are not significantly different from our expectations. For instance, we anticipated that vehicles with higher mileage, a history of repairs, older versions, and common interior and exterior colors (not upgraded) would have lower prices on average.

Table 1 shows the posterior coefficients of the Bayesian model mentioned earlier. As revealed by the results, most of the variables are significant predictors of the price of used cars. For example, the estimated price for blue and black exterior colors is on average \$294.89 (95% CI: -960.12 - 357.07) and \$38.69 (95% CI: -768.97 - 705.15) lower than pearl white (the basic color). For the red and silver exterior, the average price is \$708.79 (95% CI: -427.61 - 1840.09) and \$191.54 (95% CI: -331,67 - 712.07) higher than pearl white. This phenomenon is intriguing because, when purchasing a new Tesla, buyers must pay an additional \$1,000 to \$2,000 to select different color options instead of pearl white. However, in the used car market, it appears that people prefer to buy the most common color, and other color options tend to be slightly cheaper than pearl white, except for silver and red (which adds \$1,000 and \$2,000 when bought new).

Regarding other results, the price of a car with a black or white premium interior is \$284.69 (95% CI: -237.92 - 819.76) higher than all black interiors. When upgrading to different levels of autonomous driving software, which Tesla is most famous for, the price of a vehicle equipped with Enhanced Autopilot and Full Self-Driving increases by \$1336.59 (95% CI: 818.92 - 1869.19) and \$4399.93 (95% CI: 3600.11 - 5195.61) more, respectively, compared to the basic autopilot software. This is also interesting because, when purchasing a new Tesla, buyers must pay an additional \$6,000 and \$15,000 to upgrade their software to Enhanced Autopilot and Full Self-Driving, respectively, instead of basic Autopilot. Based on this result, people who would like to experience autonomous driving software capabilities with a limited budget are highly recommended to get a used Tesla vehicle since you can pay less to get the same driving experience.

The vehicle's price which is previously repaired is on average \$2030.91 (95% CI: -2825.17 - 1232.80) lower than the vehicles without repair history. Similarly, when the used car mileage increase by one mile, the vehicle's price is on average \$0.09 (95% CI: -0.11 - -0.07) lower. Moreover, the car price with the more recent version of Model 3 LR is on average higher compared to the 2019 version. For example, the year version of 2020, 2021, and 2022 are on average \$1788.44 (95% CI: 1692.00 - 1884.46), \$7968.07 (95% CI: 6726.07 - 9175.70), and \$8876.54 (95% CI: 7467.54 - 10259.39) higher than 2019's version. Lastly, the price of a car with 19" Sport Wheels is \$680.83 (95% CI: 241.16 - 1121.63) higher than 18" Aero Wheels. From the ACF and time-series plots for fixed effects, it seems like all the parameters converge well and autocorrelation drops extremely quickly for only 50 lags (Figure 9 - 14).

Prediction:

Based on the poseterior results in Table 1, I made a visualization app interface via R Shiny. Please go to the this URL: Tesla Model3 Long Range Price Prediction. My model enables people to determine their trade-in value or assess whether a currently listed vehicle on Tesla's website is a good deal based on their preferences, such as color or software. Additionally, I've included a price range estimate using a 10% to 90% credible interval, a similar feature to other well-known used car prediction websites like Kelly Blue Book or Carmax. I am confident that my website can serve as a dependable resource for those interested in owning a used Tesla Model 3 Long Range Version. In addition to providing valuable information on pricing and trade-in value, it offers a comprehensive range of features for potential buyers to compare with the new car price and used car price and to see which makes more sense to them when it comes to the limited budget.

Conclusion:

In conclusion, car exterior color, software, mileage, and all other variables in my model were strongly associated with the price of the Tesla Model 3 LR in California. We also found out that Blue and Black are less popular and more depreciated in the used car market since the price is lower than the basic pearl white. Moreover, people who would like to experience autonomous driving software capabilities like Enhanced Autopilot and Full-Self Driving with a limited budget are highly recommended to get a used Tesla vehicle since you can pay less (approx. $4,000 & 10,000 less than buying a new one) to get the same driving experience. Lastly, the R shiny Interface also served as a reliable resource for those interested in owning a used Tesla Model 3 Long Range Version in California.

Appendix:

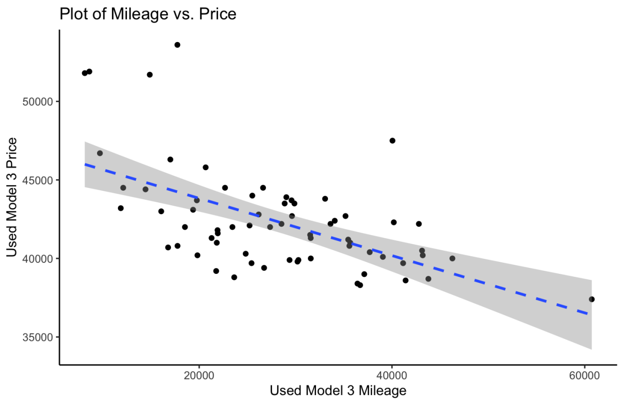

Figure 1: Scatterplot for Mileage v.s Price

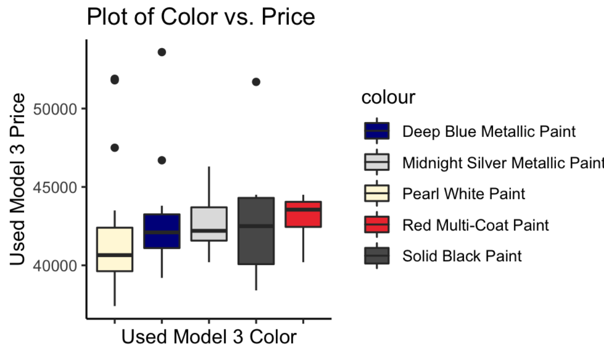

Figure 2: Boxplot for Color v.s Price

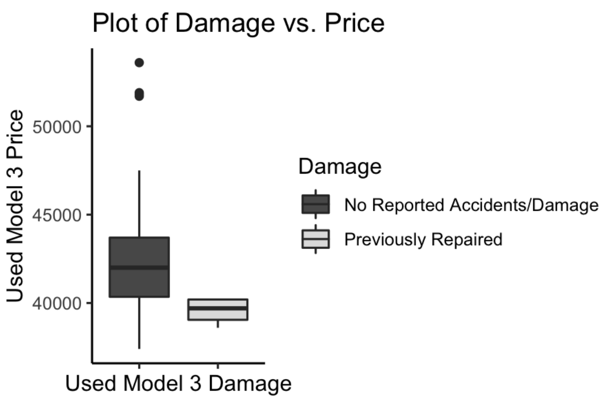

Figure 3: Boxplot for Damage v.s Price

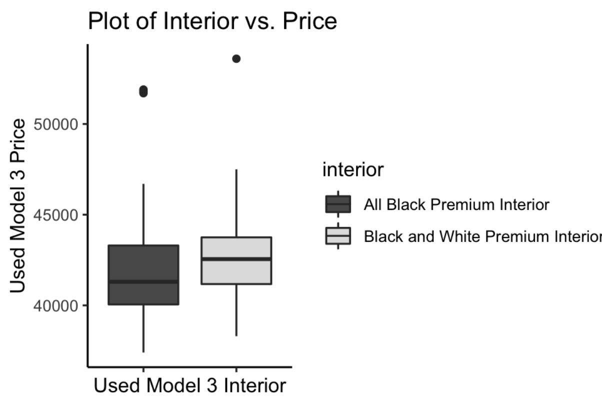

Figure 4: Boxplot for Interior v.s Price

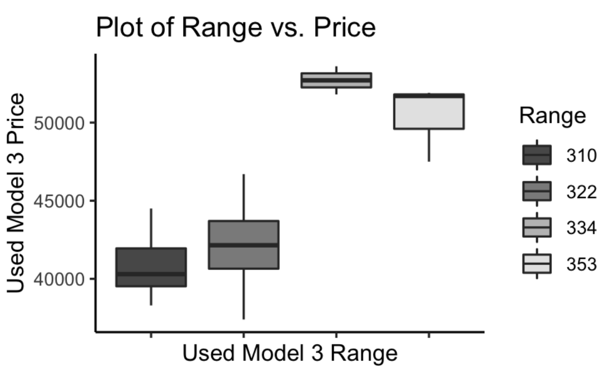

Figure 5: Boxplot for Range v.s Price

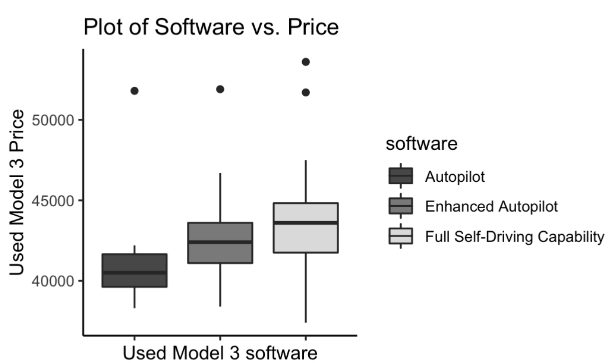

Figure 6: Boxplot for Software v.s Price

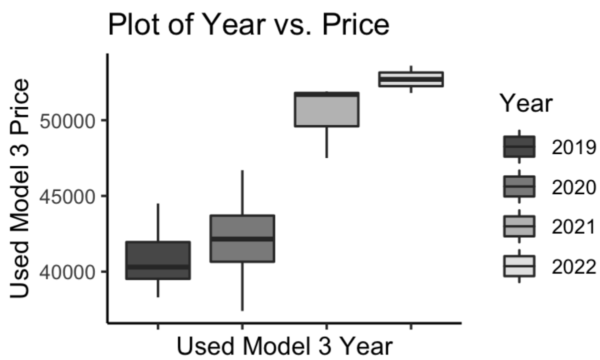

Figure 7: Boxplot for Year v.s Price

Figure 8: Boxplot for Wheel v.s Price

	Mean	SD	2.50%	97.5%	P(≥0|Y)
Intercept	42252.31	382.63	41497.58	42996.14	1.00
Blue	-294.89	335.63	-960.12	357.07	0.19
Sliver	191.54	265.99	-331.67	712.07	0.76
Black	-38.69	377.18	-768.97	705.15	0.46
Red	708.79	575.21	-427.61	1840.09	0.89
Interior	284.69	268.80	-237.92	819.76	0.86
EAP	1336.59	268.27	818.92	1869.19	1.00
FSD	4399.93	407.56	3600.11	5195.61	1.00
Repair	-2030.91	406.72	-2825.17	-1232.80	0.00
Mile	-0.09	0.01	-0.11	-0.07	0.00
Year20	1788.44	49.42	1692.00	1884.46	1.00
Year21	7968.07	623.68	6726.07	9175.70	1.00
Year22	8876.54	714.84	7467.54	10259.39	1.00
Wheel19	680.83	224.39	241.16	1121.63	1.00

Table. 1: Posterior coefficients from Bayesian regression

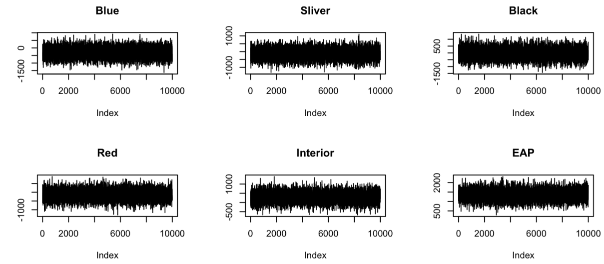

Figure 9: Trace plot for first six coefficient

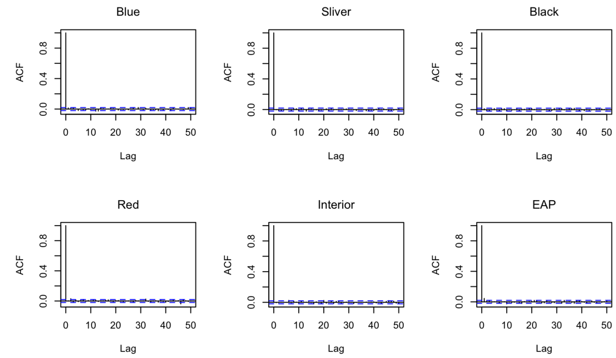

Figure 10: Autocorrelation plot for first six coefficient

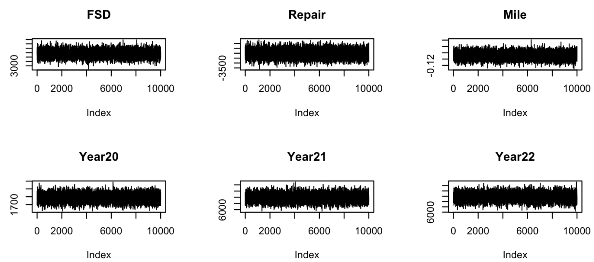

Figure 11: Trace plot for other six coefficient

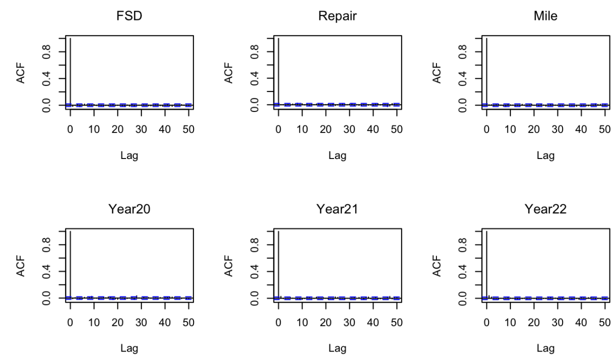

Figure 12: Autocorrelation plot for first six coefficient

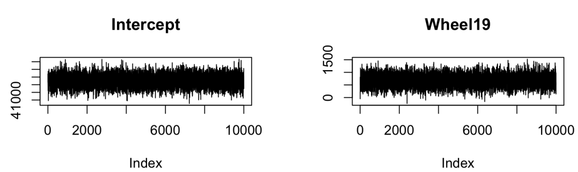

Figure 13: Trace plot for Intercept and Wheel

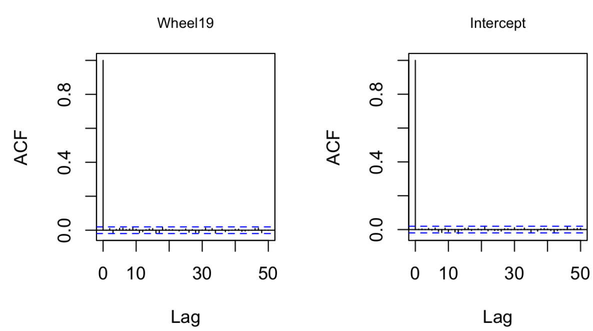

Figure 14: Autocorrelation plot for Intercept and Wheel

Detailed Description of Dataset:

- **Price**: This is the car price in US dollars listed on the Tesla official website. The price for new Tesla Model 3 back in 2020 is $51,990.
- **Color**: This is the exterior color for Tesla Model 3 LR. The following color options are available for the Model 3 LR: Pearl White, Deep Blue, Midnight Silver, Red Multi-Coat, and Solid Black. For buying new model 3 in those color options, some color options need additional prices which are:
  - Pearl White - Included (Reference)
  - Deep Blue Metallic - $1,000
  - Midnight Silver Metallic - $1,000
  - Solid Black - $1,500
  - Red Multi-Coat - $2,000
- **Interior**: This is the interior color design for Tesla Model 3 LR. The available interior color designs are Black and White Premium Interior and All Black Premium Interior. The interior color options are:
  - All Black Premium Interior - Included (Reference)
  - Black and White Premium Interior - $1,000
- **Software**: These is the different levels of autonomous driving capabilities that are available in Tesla vehicles. The available software options are Autopilot (AP), Enhanced Autopilot (EAP), and Full Self-Driving (FSD). The price for different levels of software are:
  - Autopilot (AP) - Included (Reference)
  - Enhanced Autopilot (EAP) - $6,000
  - Full Self-Driving (FSD)- $15,000
- **Damage**: This is a binary variable indicates that whether this vehicle is no reported accidents/damage before or previously repaired.
- **Mileage**: This is the mileage from the vehicle's odometer.
- **Range**: This is an EPA-estimated range for this vehicle on a single charge. It's also a variable that indicates battery degradation. The lower the EPA-estimated range is, the lower the battery capacity relative to when it was a new vehicle. The EPA-estimated range for a new Tesla Model 3 LR is 358 miles.

- **Year**: This is the calendar year that this vehicle was made. The typical years that Tesla Model 3 LR made are from 2019 to 2022. The current used vehicle on the website is from 2019 to 2021, mostly 2019 and 2020. (2019, 2020, 2021 & 2022)
- **Wheels**: This is the wheel size of the vehicle. The available wheels size options and prices are:
  - 18" Aero Wheels – Included (Reference)
  - 19" Sport Wheels- $1,500

FDAP_jags

Kuan-Hung Yeh

2023-03-06

```
#Model
sink("proj1model.txt")
cat("
model
{
   for(i in 1:N_obs) {
         y[i] ~ dnorm( mu[i] , tau)
         mu[i] <- beta0 + inprod(x[i,] , beta[] )
        }

     beta0 ~ dnorm(mbeta0 , precbeta0)

for (j in 1:N_betas) {
     beta[j] ~ dnorm( beta_mean[j] , beta_prec[j] )
        }
       tau ~ dgamma( tau.a , tau.b )
       sigma <- 1 / sqrt( tau )

       BasicModel <- beta0 + beta[9]*30000

    }
  ",fill = TRUE)
sink()

Blue = ifelse(Model3$colour == "Deep Blue Metallic Paint", 1, 0)
Sliver = ifelse(Model3$colour == "Midnight Silver Metallic Paint", 1, 0)
Black = ifelse(Model3$colour == "Solid Black Paint", 1, 0)
Red = ifelse(Model3$colour == "Red Multi-Coat Paint", 1, 0)
Interior = ifelse(Model3$interior == "Black and White Premium Interior", 1, 0)
EAP = ifelse(Model3$software == "Enhanced Autopilot", 1, 0)
FSD = ifelse(Model3$interior == "Full Self-Driving Capability", 1, 0)
Repair = ifelse(Model3$Damage == "Previously Repaired", 1, 0)
Mile = Model3$Mileage
Year20 = ifelse(Model3$Year == 2020, 1, 0)
Year21 = ifelse(Model3$Year == 2021, 1, 0)
Year22 = ifelse(Model3$Year == 2022, 1, 0)
Wheel19 = ifelse(Model3$Wheels == 19, 1, 0)

x = matrix(data=c(Blue, Sliver, Black, Red, Interior, EAP, FSD, Repair, Mile, Year20, Year21, Year22, Wh
y = Model3$price

prior_data <- list(N_obs=67, N_betas=13,
```

```r
                    mbeta0=40000, precbeta0=2.5e-06,
                    beta_mean=c(-1600, -130, 900, 450, 500, 2200, 4400, -2000, -0.12, 1800, 6150, 6500, 3
                    beta_prec=c(4e-06, 1e-05, 1e-06, 1e-06, 6e-06, 6e-06, 6e-06, 4e-06, 2500, 4e-04, 1e-0
                    tau.a=24, tau.b=2800000, x=x, y=y)

inits <- rep(list(list(beta=rep(0, 13), beta0=0, tau=1)), 5)

parameter <- c("beta0", "beta", "sigma", "BasicModel")

run1 = jags(data = prior_data, inits = inits, parameters.to.save = parameter, "proj1model.txt",
            n.chains=5, n.iter=11000, n.burnin=0, n.thin=1)
Output1=AddBurnin(run1$BUGSoutput$sims.array,burnin=1000,n.thin=1)
temp1 = Output1$Burnin.sims.matrix

#table
mysummary = function(invector) {
  c(mean(invector), sd(invector), quantile(invector, .025),
    quantile(invector,.975),
    length(invector[invector>0])/length(invector))
}

beta = cbind(temp1[,2:15], temp1[,17])

betaout = t(apply(beta, 2, mysummary))
rownames(betaout) = colnames(x)
colnames(betaout) = c("Mean", "SD", " 2.50%", "97.5%", "P(>0|Y)")
rownames(betaout) = c("Blue", "Sliver", "Black", "Red", "Interior", "EAP", "FSD", "Repair", "Mile", "Yea
knitr::kable(round(betaout, 2), "latex", escape = TRUE)
```