Neil Strugnell

MSBA 8170

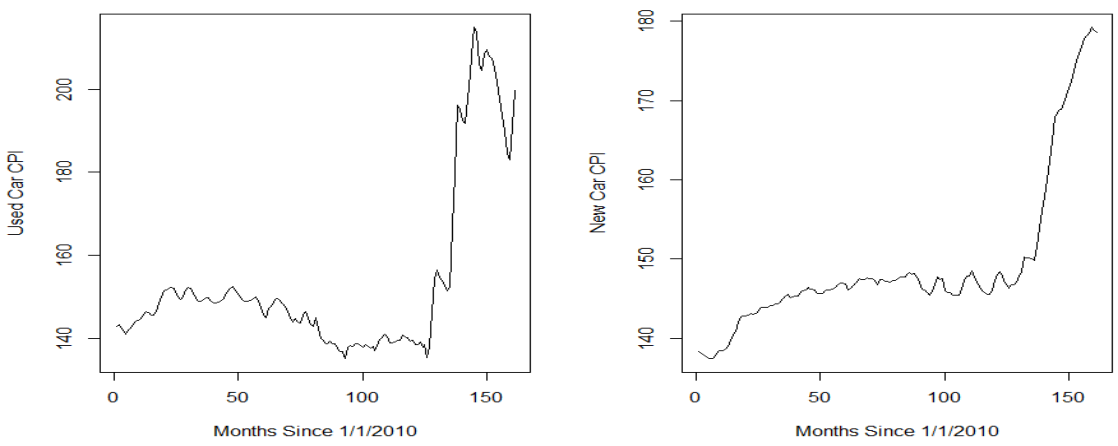
8/20/23

Case Study – United States Used Car Prices

Background Information

For my case study, I investigated potential factors affecting the prices of used cars in the United States for the past decade to build a predictive model identifying the most impactful. The question of the causes of used car prices demands particular interest due to a sudden upwards shift in price trends during and after 2020, compared to relatively static prices across the previous decade. These elevated prices impact people across the country whether or not they themselves are interested in purchasing a new or used vehicles, due to adjacent participation through activities such as ordering items for delivery or using ride share apps for transportation. In particular, I was interested in identifying the primary factors of used car prices to predict whether it is advisable to purchase a vehicle immediately or to wait.

Used and New Car Price Trends 1/1/2010 – 5/31/23



There are many potential factors that may impact used car prices, with many more possibilities that I could fully consider in the scope of this study. I sought to identify a manageable number of variables that would allow me to investigate influencing factors from global supply lines and the disruptive effects of COVID 19 and other shock effects around the world, demand for cars due to increased used of semi-autonomous delivery services and employment options such as Amazon delivery contractors and rideshare apps, and the effect of cost of repairs and long waits for mechanic services.

I gathered data from several sources to investigate price trends for new and used in the United States from 1/1/2010 – 5/31/2023. In particular, I found information on nine variables: monthly CPI data from the St. Louis FED for used car prices, new car prices, vehicles maintenance, and trends for the

overall U.S. economy, as well as data on finances rates for 24-month personal loans as a benchmark for the interest rate. To create a benchmark for movements in global supply lines and its impacts on domestic auto prices, I also found monthly Census data for U.S. auto imports and exports along with overall exports and imports for the same period.

Summary Statistics

Variable	CPI All	CPI Used Cars	CPI New Cars	CPI Maintenance	Finance Rate
Minimum	217.2	135	137.4	135.3	8.73
1st Quantile	232.4	140.2	145.4	143.7	9.65
Median	241.2	147.6	146.7	145.3	10.13
Mean	247.4	153.6	149.2	148.8	10.18
3rd Quantile	257.2	151.6	147.8	148.3	10.65
Max	303.3	214.9	179.2	182.4	11.48
Variance	4.73E+02	4.32E+02	9.45E+02	1.18E+02	3.83E-01

Variable	All Exports	Auto Exports	All Imports	Auto Imports
Minimum	95883	3431	143563	8928
1st Quantile	119358	12047	171446	24397
Median	125241	12562	188535	28705
Mean	124982	12281	190969	27201
3rd Quantile	133275	13002	205791	30397
Max	146627	14345	247936	35536
Variance	1.06E+08	2.17E+06	5.13E+08	2.12E+07

Covariance Matrix

	cpi_all	cpi_used_cars	cpi_new_cars	fin_rate
cpi_all	473.486745	333.164810	1.947727e+02	-3.3838055
cpi_used_cars	333.164810	431.693314	1.738843e+02	-2.0932056
cpi_new_cars	194.772742	173.884343	9.461901e+01	-0.6534943
fin_rate	-3.383806	-2.093206	-6.534943e-01	0.3828880
real_ex_all	176790.635084	95340.232926	6.969591e+04	-1374.2590012
real_ex_cars	8092.356565	1795.452211	3.581079e+03	-26.4912628
real_imp_all	460715.536423	300198.252299	1.794679e+05	-5026.7799868
real_imp_cars	66397.024300	26771.387916	2.540001e+04	-698.2272244
cpi_maintain	207.344418	197.864047	1.026153e+02	-0.1129225

	real_ex_all	real_ex_cars	real_imp_all	real_imp_cars
cpi_all	176790.635	8.092357e+03	460715.54	66397.0243
cpi_used_cars	95340.233	1.795452e+03	300198.25	26771.3879
cpi_new_cars	69695.907	3.581079e+03	179467.95	25400.0135
fin_rate	-1374.259	-2.649126e+01	-5026.78	-698.2272
real_ex_all	106149132.580	1.093399e+07	206359845.17	43796194.7776
real_ex_cars	10933987.739	2.168059e+06	13848537.89	5373475.7849
real_imp_all	206359845.169	1.384854e+07	512938842.02	84274078.9184
real_imp_cars	43796194.778	5.373476e+06	84274078.92	21231321.5303
cpi_maintain	68014.173	2.945128e+03	179650.25	22652.6067

	cpi_maintain
cpi_all	2.073444e+02
cpi_used_cars	1.978640e+02
cpi_new_cars	1.026153e+02

```
fin_rate      -1.129225e-01
real_ex_all    6.801417e+04
real_ex_cars   2.945128e+03
real_imp_all   1.796502e+05
real_imp_cars  2.265261e+04
cpi_maintain   1.183082e+02
```

Analysis

I started with a simple linear regression including all the response variables to see if any of the predictor variables immediately stood out compared with the rest. This model found that almost all of the variables that I chose to include were strongly predictive of used car prices, except for total U.S. exports by month, with an extremely high R-squared response for the initial model of 0.8922. These results immediately speak to the complexity of predicting U.S. car prices when there are so many potential factors influencing price trends, but does validate that the variables included in the model already strongly predictor trends in car prices.

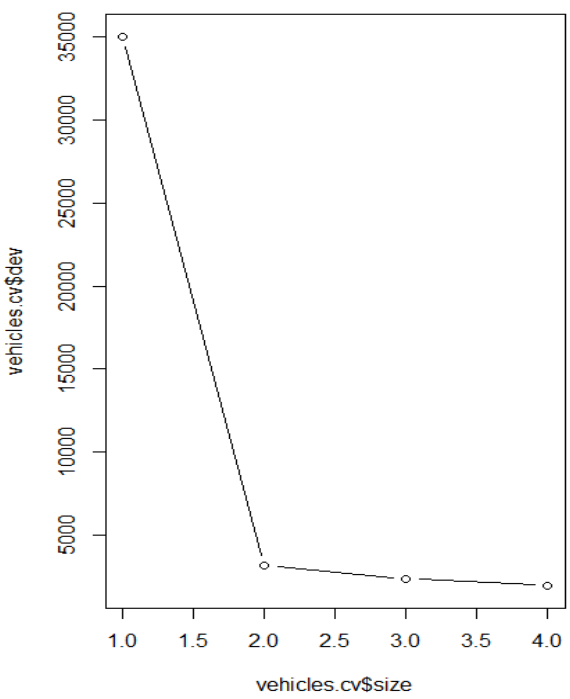
Coefficients of the linear model:

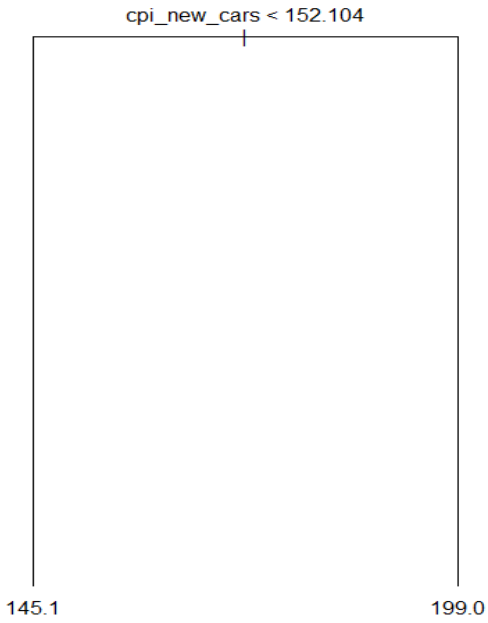
```
                Estimate Std. Error t-value Pr(>|t|)
(Intercept)   -7.496e+01  1.755e+01  -4.272 3.40e-05 ***
cpi_all       -8.816e-01  1.420e-01  -6.211 4.81e-09 ***
cpi_new_cars   1.490e+00  3.207e-01   4.646 7.27e-06 ***
fin_rate      -4.345e+00  1.100e+00  -3.950 0.000119 ***
real_ex_all   -4.113e-04  2.257e-04  -1.822 0.070365 .
real_ex_cars   3.408e-03  1.082e-03   3.148 0.001977 **
real_imp_all   9.773e-04  1.185e-04   8.244 7.24e-14 ***
real_imp_cars -3.052e-03  4.421e-04  -6.904 1.29e-10 ***
cpi_maintain   1.173e+00  2.782e-01   4.217 4.24e-05 ***
R-Squared: 0.8922, Adjusted R-Squared: 0.8865
```

I included overall U.S. exports as a variable to compare against real exports and imports of cars and overall U.S. imports, but I am not too surprised that it was immediately indicated as not significant – the U.S. exports such a vast set of products every month that the main way real exports can predict car prices would be as a barometer for the overall health of the economy that month, and this aspect of overall domestic economic health is better encapsulated in overall real imports. This is because a drop in exports may indicate economic difficulties abroad, which is less likely to immediately impact domestic prices, whereas a drop in imports indicated less domestic purchasing power during the period.

In terms of the statistically significant variables, several notably trend in the opposite direction of used car prices. These are overall domestic CPI, the interest rate, and real car imports. Of these trends in predictor variables, only the negative effect of overall CPI increases surprised me, but it made more sense when I reflected on the historical constancy of used car prices for most of the period of the data I used, as overall CPI trended upward for the entire 2010 decade while used car prices were flat.

After looking over this initial model, I split the data into a testing set and a training set. I fit a regression tree using the training data, then used the testing data to calculate an initial Mean Square Error of 47.50. I used cross-validation to determine the best level of tree complexity, and found that the best size estimate was 2. However, after pruning to tree complexity 2 I found that the MSE had jumped up to 54.76, which strongly suggests that reducing to complexity 2 over pruned the regression tree.





I tried the bagging approach next with mtry = 10, along with the importance() function to find the most important variables for the regression. After bagging I found that my test MSE was 15.12, which is a substantially improved result compared with the MSE I found from pruning. The most important variable affecting used car prices were found to be maintenance costs, real U.S imports, new car prices and real U.S. exports, in that order. The first three of these variables all match with the predictions of the initial linear regression model, but after the bagging approach real U.S. exports were again a statistically significant predictor variable, which brings into question the accuracy of the initial model, especially given how high the R-squared was for it.

Conclusions

Looking over the results of my analysis, I teased out some trends in the factors of United States used car prices, but I feel that there is a lot of potential depth for further analysis. From my results I can consistently see that maintenance costs, real U.S. imports, and new car prices are all significant predictors for used car prices, but I did not find precise magnitudes for these predictors, and the significance of real U.S. exports remains a question as it is shown to have significance in some models but not others. I would also like to deepen the analysis by finding a variable to represent domestic demand for cars, perhaps in terms of overall miles driven every month. It would also be fascinating to dig into trends in prices by brand, for instance, to see what effects, if any, the recently exposed vulnerabilities of certain Kia and Hyundai models to easy theft had the prices of those models and on the overall financial well-being of the brand.

Data Sets

CPI Data: https://fred.stlouisfed.org/release/tables?rid=10&eid=34483#snid=34526

Finance Rate on Personal Loans: https://fred.stlouisfed.org/series/TERMCBPER24NS

Real Exports & Imports: https://www.census.gov/foreign-trade/statistics/historical/real.html