Lost Auto Sales

Matt Burns

Fall 2020

https://github.com/mattburns963/680.git

Problem and Hypothesis

In 2020 there was a sharp decline in auto sales in the United States. The conventional wisdom is that sales were lost due to the consumers' lack of ability to visit dealership and purchase a new car and the lack of availability of new cars due to COVID-related production shutdowns. These are likely two reasons sales were depressed, but I'd like to also quantify the lost sales due to the reduction in driving. The thought is that lower driving reduced the number of crashes (which require a new car to replace the totaled car) and reduced wear and tear.

This project will investigate the shortfall in auto sales year-to-date in 2020. It will specifically look to quantify the number of vehicles that would have been sold (in a non-pandemic circumstances) and then forecast when those sales will be made up. The idea is that sales lost during this year would eventually need to be made up in subsequent time periods. In addition to quantifying the lost sales, potential impacts of fewer miles driven (increasing the life span of the existing auto fleet) and the potential impacts of the sales of used cars (a substitute product) will be considered.

Data Retrieval Methodology

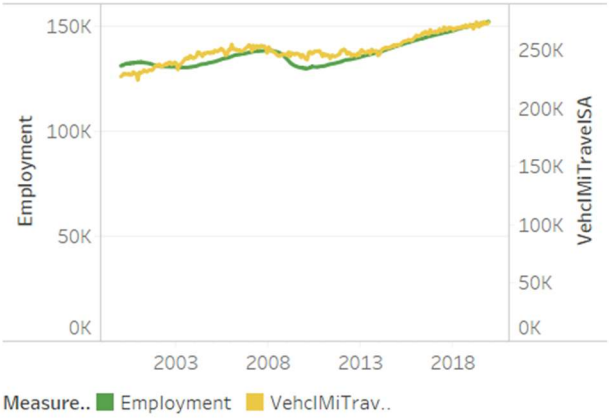
The data will include a timeseries of historical vehicle sales. Two decades of data is available and was used for EDA. The quantification of sales lost was based on more recent data.

Macroeconomic data was also be included in the analysis as macroeconomic parameters impact auto sales. These will at least include Consumer Price Index, Employment and Personal Consumption, Mortgage Debt Burden were included.

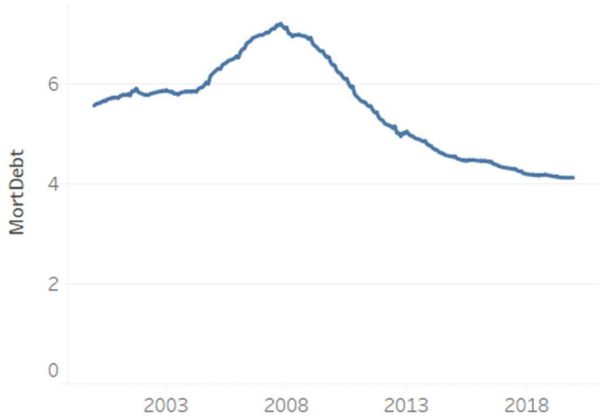
Additionally, I was able to obtain auto-specific data. This included auto miles traveled and auto production.

All of the data was obtained from the St. Louis Fed easily. Unlike the last project, this was fairly straightforward. Adjustments were made to adjust a few of the attributes from a quarterly native frequency to a monthly frequency.

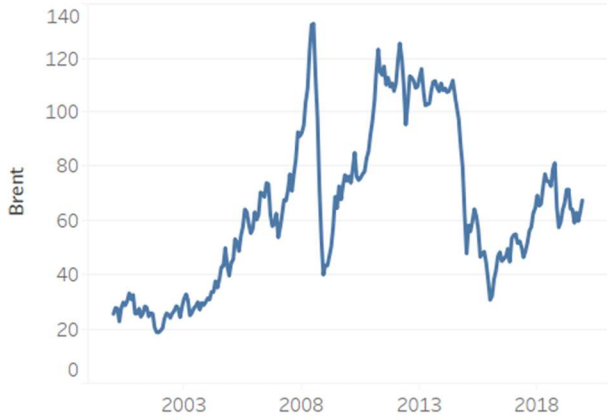
Mi Traveled vs Emp.



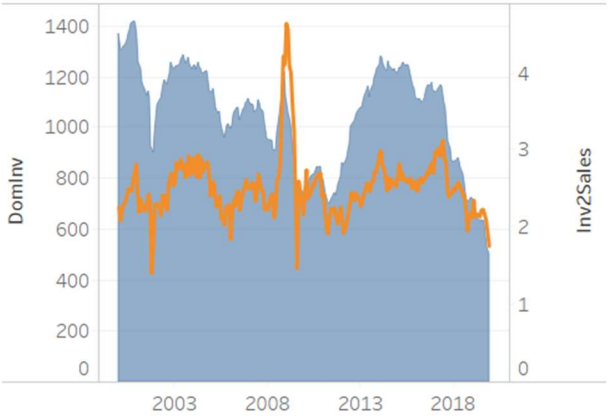
Mort DSB

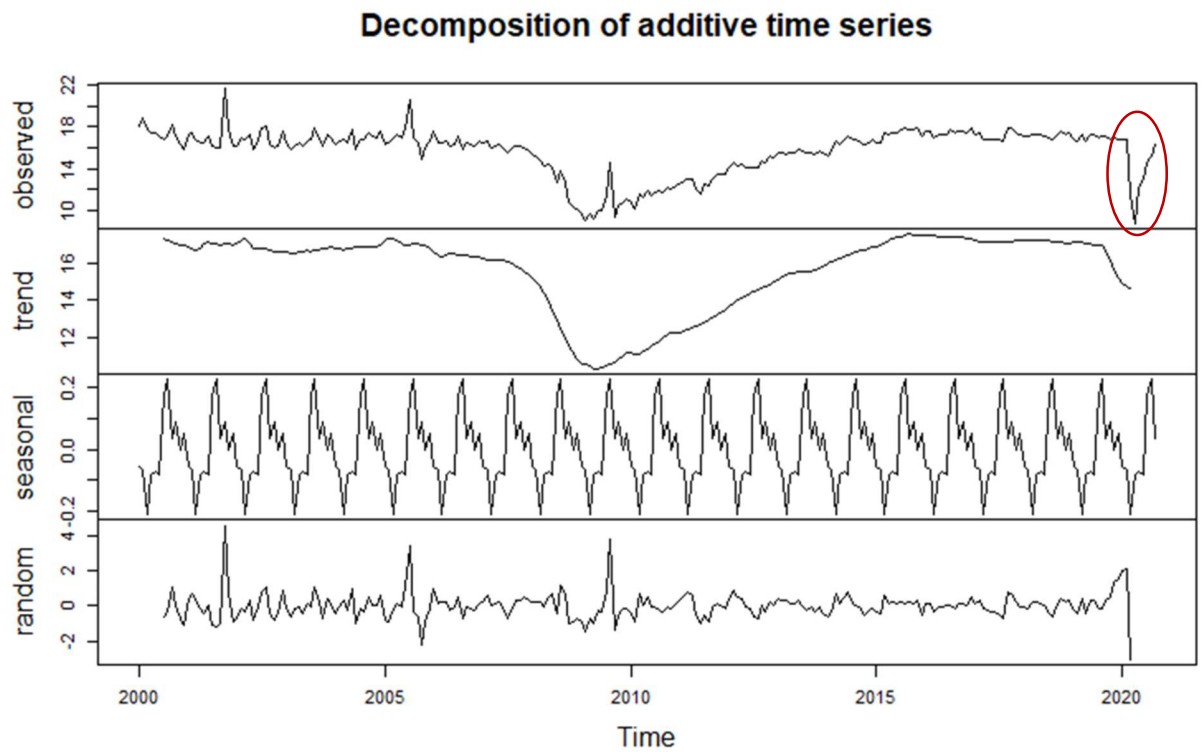


Brent Prices



Inventory to Sales





Analysis Methodology

First, I quantified the loss of sales by using a simple forecast of the auto sales based on data through 2019. After 2020 was forecasted out using pre-pandemic data it was compared to the actual sales data from the St. Louis fed via simple subtraction. From March to September 2.235 million light vehicle sales were lost.

Note: The calculation below shows the annualized shortfall for each month. After the number for March to September are summed, they are divided by 12 to backout the annualization.

DATE	LtWgtSales	Expected	Annualized Shortfall
Jan-20	16.865		
Feb-20	16.765		
Mar-20	11.354	16.879	5.525
Apr-20	8.714	16.879	8.165
May-20	12.104	16.879	4.775
Jun-20	13.009	16.879	3.870
Jul-20	14.624	16.879	2.255
Aug-20	15.183	16.879	1.696
Sep-20	16.341	16.879	0.538
			2.235

Then I used a regression to determine the relationship between the following three factors and sales:

- Auto Production
- Auto Miles Travel
- Employment
- Reduced Travel (logged)

I used Inventory to Sales Ratio and Mortgage Debt to tighten up the regression. No sales impact will be attributed to them as they didn't dramatically change due to the pandemic.

```
Coefficients:
                    Estimate      Std. Error    t value       Pr(>|t|)
(Intercept)         -2.509e+02    4.153e+01     -6.041        6.00e-09 ***
Employment          9.955e-05     2.374e-05     4.193         3.91e-05 ***
log(Inv2Sales)      -3.914e+00    4.170e-01     -9.387        < 2e-16 ***
log(VehclMiTravelSA) 2.030e+01    3.537e+00     5.741         2.91e-08 ***
Brent               -2.626e-02    2.184e-03     -12.028       < 2e-16 ***
DomProdSA           2.168e-02     1.115e-03     19.455        < 2e-16 ***
MortDebt            -1.996e-01    8.397e-02     -2.377        0.0183 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.9115 on 233 degrees of freedom
Multiple R-squared:  0.8355,  Adjusted R-squared:  0.8313
F-statistic: 197.3 on 6 and 233 DF,  p-value: < 2.2e-16
```

These sensitivities allowed me to assign sales shortfalls to each of them. Please note that Mortgage Debt Service burden and Oil Prices (aka Brent) were used to tighten up the regression. They were not used to assign the lost sales.

I used estimates for Auto Production, Auto Miles Travel, Employment and Reduced Travel (logged) and compared them to actuals to assign lost auto sales to the four attributes. The coefficients in the regression above were used in the calculation to relate the attributes to the lost sales.

- 800,000 due to unemployment
- 1,400,000 due to less travel
- -486,000 due to lower oil prices
- 839,000 due to auto lower production
- 318,000 due to unexplained factors
- 2,235,000 Total reduction

Conclusions

1. Almost 83% of auto sales can be described by the following attributes:
   - Employment
   - Sales to Inventory Ratio
   - Vehicle Miles Traveled
   - Oil Prices
   - Production
   - Mortgage Debt Service Burden
2. Production, Travel and Unemployment ratios had the largest implicating sales
3. Almost 2.235 million vehicle sales were lost YTD in 2020
4. A majority of the lost sales will likely be recouped when employment and travel rebound

Q&A

1. How can we quantify the amount of used car sales that acted a as substitution product?
2. When will the pent-up demand impact the market?
3. Are there impacts to demand that are the result of permanent changes in driving habits?
4. Has reduced usage of mass transit and ride shares increase the demand for automobiles?
5. Why was the regression used for to explain the interactions of the variables, but not the forecast?
6. Can the R^2 of 83% be improved?
7. Can the inventory metrics and the Vehicle Miles Traveled be lagged to improve the regression?
8. How does the sales deficient compare to a straight-line forecast?
9. What are the advantageous of exponential smoothing over a regression for forecasting?
10. Can this methodology be used to quantify the lost retail sales?

APA references

Articles relating to the auto industry sales are below. There are also articles that detail information regarding recent actual sales data as well sales prognostications. There are also articles to help inform my methodology.

Katje, Chris. (Downloaded Oct. 5, 2020) North American Car Sales Show Signs Of Recovery In Q3. Retrieved from https://www.benzinga.com/news/20/10/17758641/north-american-car-sales-show-signs-of-recovery-in-q3
This article provides recent data regarding auto sales at the company level. It will be used to adjust the near-term outlook of the sales volume.

Cox Automotive. (Downloaded Oct. 6, 2020) 13-Month Rolling Used-Vehicle SAAR. Retrieved from https://www.coxautoinc.com/newsroom/market-insights/
This article discusses used automotive sales over the last 13 months. Used cars are a substitute product of new cars and can potentially impact the subsequent demand for new cars.

Reuters. (Aug. 4, 2020) U.S. auto sales pandemic recovery continues as Toyota decline slows. Retrieved from https://www.autoblog.com/2020/08/04/us-july-auto-sales/
This article provides recent data regarding auto sales at the company level from July. It will be used to adjust the near-term outlook of the sales volume.

Associated Press. (Oct. 1 2020) U.S. auto sales fall 9.7% in Q3, but the news isn't all bad. Retrieved from https://www.marketwatch.com/story/u-s-auto-sales-fall-9-7-in-q3-but-the-news-isnt-all-bad-01601602903
This article provides recent data regarding auto sales at the company level from 3Q20. It will be used to adjust the near-term outlook of the sales volume.

Finlay, Steven. (Mar. 19, 2020) Worst Case: 3.1 Million Lost U.S. Vehicle Sales in 5 Months. Retrieved from https://www.wardsauto.com/industry/worst-case-31-million-lost-us-vehicle-sales-5-months
This article provides an initial view of the lost sales as a result of the pandemic. It will be referenced to identify any potential methodology improvements for my estimation.

Wayland, Michael. (Oct. 1, 2020) Fiat Chrysler's and GM's 3Q auto sales fall from last year, but demand improves from depths of coronavirus lockdowns. Retrieved from https://www.cnbc.com/2020/10/01/coronavirus-fiat-chrysler-and-gms-3q-auto-sales-fall-from-last-year-but-demand-improves.html
This article provides recent data regarding auto sales at the company level from 3Q20. It will be used to adjust the near-term outlook of the sales volume.

Narayanan, Aparna. (Oct. 2, 2020) Ford Beats GM, Fiat Chrysler In U.S. Auto Sales As Tesla Sets Record. Retrieved from https://www.investors.com/news/auto-sales-q3-2020-gm-tesla-ford/
This article provides recent data regarding auto sales at the company level from 3Q20. It will be used to adjust the near-term outlook of the sales volume.

National Automotive Dealers Association. (July 9, 2020) NADA Issues Second Quarter 2020 Auto Sales Analysis. Retrieved from https://www.nada.org/Q2-2020-Auto-Sales-Analysis/
This article analyzes data regarding auto sales at the company level from 2Q20. It will be used to help inform the decrease sales volume for the midst of the shutdown.

Qi, Zhang, Zhan, Hongfei, Yu, Junhe. (2017) Car Sales Analysis Based on the Application of Big Data. Retrieved from https://doi.org/10.1016/j.procs.2017.03.137
This academic paper details a big data methodology to analize car sales. It will be used to identify potential improvements to my methodology.

Isidore, Chris. (Oct. 1, 2020) Car sales are rebounding, but they're still not back to pre-pandemic levels. Retrieved from https://www.cnn.com/2020/10/01/business/car-sales-pandemic/index.html
This article discusses the outlook regarding the growth in vehicle sales in the near future. It will be used as a sanity check of my estimations.

Appendix: R Code

```
library(readxl)
library(caret)
library(Rmisc)
library(ggplot2)
library(forecastML)
library(effects)
library(DataCombine)
require("forecast")
require("expsmooth") # required for the data
library(smooth)

# pull in monthly economic data from Jan. 2000 thru Dec. 2019
econ <- read_excel("C:/Users/burns/OneDrive/Desktop/Matt/Grad School/DSC 680/Project
2/AutoData.xlsx")

# clean dataframe
econ[!complete.cases(econ),]

# Break apart the componets of auto sales
LtWgtSales <- read_excel("C:/Users/burns/OneDrive/Desktop/Matt/Grad School/DSC 680/Project
2/LtWgtSales2020.xlsx")
salestimeseries <- ts(LtWgtSales$LtWgtSales, frequency=12, start=c(2000,1))
salescomponents <- decompose(salestimeseries)
plot(salescomponents)

# Test 3 candidate models
multi.fit1 = lm(LtWgtSales~Employment+Inv2Sales+VehclMiTravelSA+Brent+DomProdSA+MortDebt,
data=econ)
summary(multi.fit1)

multi.fit2 = lm(LtWgtSales~Employment+Inv2Sales+VehclMiTravelSA+Brent+MortDebt, data=econ)
summary(multi.fit2)

multi.fit3 = lm(LtWgtSales~Employment+VehclMiTravelSA+Brent+DomProdSA+MortDebt, data=econ)
summary(multi.fit3)

# Investigate lagging Inventory
df_lead2 <- as.data.frame(econ)
df_lead2 <- slide(df_lead2, "Inv2Sales", NewVar = "I2SLead2", slideBy = 2)  # create lag1 variable
multi.fit1.lead2 = lm(LtWgtSales~Employment+I2SLead2+VehclMiTravelSA+Brent+DomProdSA+MortDebt,
data=df_lead2)
summary(multi.fit1.lead2)

df_lag2 <- as.data.frame(econ)
df_lag2 <- slide(df_lag2, "Inv2Sales", NewVar = "I2SLag2", slideBy = -2)  # create lag1 variable
multi.fit1.lag2 = lm(LtWgtSales~Employment+I2SLag2+VehclMiTravelSA+Brent+DomProdSA+MortDebt,
data=df_lag2)
summary(multi.fit1.lag2)
```

```
# Investigate Nonlinear adjustements to Inventory to Sales and Vehilcle Miles Traveled
multi.fit4 = lm(LtWgtSales~Employment+log(Inv2Sales)+VehclMiTravelSA+Brent+DomProdSA+MortDebt,
data=econ)
summary(multi.fit4)

multi.fit5 =
lm(LtWgtSales~Employment+log(Inv2Sales)+log(VehclMiTravelSA)+Brent+DomProdSA+MortDebt,
data=econ)
summary(multi.fit5)


eff.fit5 <- allEffects(multi.fit5, xlevels=50)
for(i in 1:6) {plot(eff.fit5[i])}


fit5 <- predict(multi.fit5)
econ <- cbind(econ, fit5)
ggplot(econ, aes(DATE, y = value, color = variable)) +
  geom_point(aes(y = LtWgtSales, col = "Lt Auto Sales Actual")) +
  geom_line(aes(y = fit5, col = "Lt Auto Sales Fitted")) +
  theme(legend.position="bottom")+ ggtitle("Fitted vs Actual")


saleses <- es(econ$LtWgtSales, "ZZZ", h=12)
plot(saleses)

empes <- es(econ$Employment, "ZZZ", h=12)
plot(empes)

mileses <- es(econ$VehclMiTravelSA, "ZZZ", h=12)
plot(mileses)

oiles <- es(econ$Brent, "ZZZ", h=12)
plot(oiles)

prodes <- es(econ$DomProdSA, "ZZZ", h=12)
plot(prodes)
```