Econometric Project 3: The End is Here!...

Or is it?

Introduction/Background

Ever since the dawn of human communities, bartering has been a large part of the reason why we can work together. The exchange of goods and services allows for specialization, which in turn increases the quality and craftsmanship of the products that our civilizations have been able to create throughout time. The act of bartering itself, has changed several times throughout history. There was the invention of currency, which increased liquidity in markets and allowed for trades involving one good or service, rather than forcing both parties to be interested in the others' skill. Another economic invention that changed our society was the assembly line, allowing trades of larger scale for cheaper prices.

These economic inventions created changes in the way we as societies distributed resources and wealth and may have caused temporary inefficiencies as the people of a community learned to understand and use the tools given to them. As with any change, there were likely to be those speaking to how these alterations would destroy the fabric of their livelihoods. Just think about how frightened some Americans are with respect to automation. There are those who may well criticize the gradual move towards automation due to its impact on lower skilled labor, pointing towards the projections of the trucking industry given a rise in self-driving cars.

It is reasonable, therefore, to enter into the discussion of the "retail apocalypse" with an air of calm skepticism. There is always a reason to think the world is ending, the challenge comes in analyzing what kind of impact the event or phenomena will really have on real-world institutions and infrastructure. From the research that I have conducted, it appears that the retail market isn't being destroyed, or going out of business, or even declining at all. Instead, a new economic invention has changed the way that we consume.

To study the changes in the retail market, we will turn to an analysis done of the retail labor market. Essentially, while traditional 'brick and mortar' retail stores have been decreasing their employment since 2012, non-store employment has been increasing at a slightly quicker rate. For example, in 2013 the decline in traditional retail jobs was 80,000, but 100,000 new jobs were added by non-store retail employment[1]. This pattern year after year implies that the distribution of retail employment is changing from traditional department stores to online retailers. This also comes with a change in the average wage for the industry, with online retailers generally employing more high-skilled labor[1]. This in turn means that not only is job growth still a reoccurring theme in the retail sector, the increase is skewed towards higher-paying and higher-skilled jobs.

This is a good sign for the retail market for several reasons. Firstly, higher paying jobs means a larger cost for the employers, implying that companies have more financial resources capable of being

---

[1] "How Is Online Shopping Affecting Retail Employment?" *Liberty Street Economics*, libertystreeteconomics.newyorkfed.org/2017/10/how-is-online-shopping-affecting-retail-employment.html.

dedicated to their employees. Also, it implies that the firms are becoming more efficient with their distribution networks, with less need for conventional 'brick and mortar stores'. Instead, companies are beginning to use these locations as pickup locations for business done online, as a complement to their online retail[2].

With this in mind, we turn to the questions posed in this assignment. What was the impact of the "retail apocalypse" on sales? Keep in mind that for the purposes of this assignment, we will use the term to denote the rise of ecommerce in the period of 2012-2016. Why was the previous projection overly optimistic? How can I improve the forecast, and demonstrate its increased credibility? Finally, What does my model predict for the next four years in the retail sales market?

Methodology

I started by recreating the forecast given for the 2016-2020 period. Before I began my own analysis by logging the data, at first because that's what the assignment said but also due to the log trend present in the data. Then, I tested the series for a unit root with the Dickey-Fuller test. Having found one, I first differenced the data before choosing an ARMA combination, thereby creating an ARIMA model and removing the unit root (also somewhat helping stationarity? Right?). Using the ARIMA model I chose, I then reforecasted the 2016-2020 period. I compared the errors of the new model with those of the one given by the trend model using the Theil U statistic.

Results

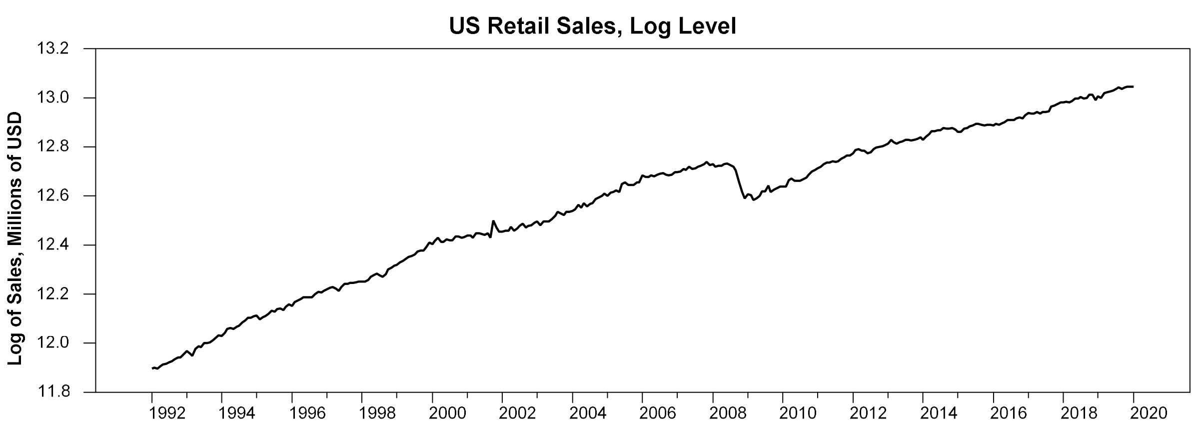

Figure 1: This is the series we worked with, the log version of this data: https://fred.stlouisfed.org/series/RSXFS, Advanced Retail Sales, Retail (Excluding Food services)

---

2 "How Is Online Shopping Affecting Retail Employment?" *Liberty Street Economics*, libertystreeteconomics.newyorkfed.org/2017/10/how-is-online-shopping-affecting-retail-employment.html.

To begin, I wanted to understand what the previous forecast was, and compare it with the data. So, I used the time-honored tradition of throwing a trend on the data to create the following model.

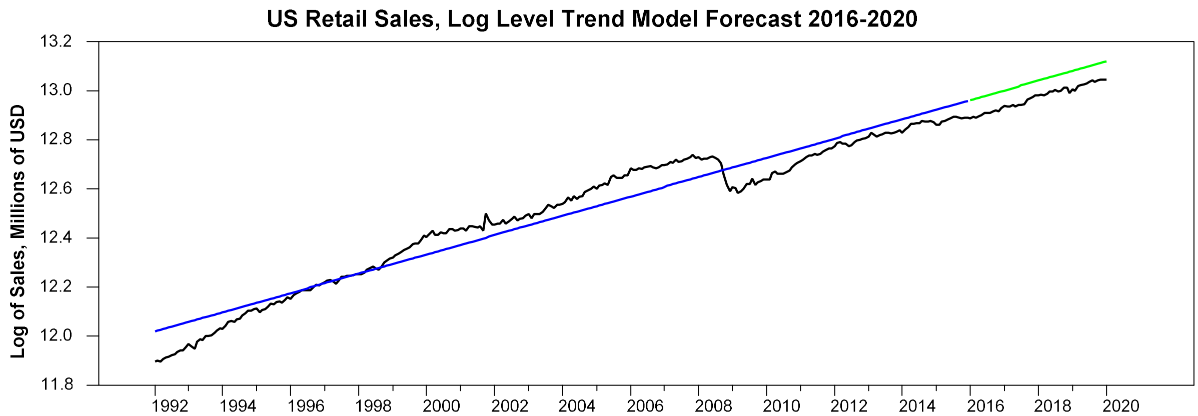

Figure 2 The blue line denotes the regression used to generate the forecast. The actual forecast is the green line.

From a visual standpoint, it is easy to see the reason for the overly optimistic. The forecast treats the data as deterministic or acts as if the recession of 2008-2009 was a predictable event that functioned as a cycle for the amount of retail sales. Essentially, retail growth up until that point had been so high that it was unsustainable in the economic sense, forcing a drop in the amount of retail sales such that the market could grow freely again.

While a trend will often be correct in terms of long-run general predictions, it often fails to accurately foresee short term fluctuations. So, I decided I would use another method to more accurately model the values of individual periods. To begin testing for my own analysis, I used the dickey-fuller test to comb the series for unit roots. Using the first difference of the data to remove a unit root can increase the predictability of future values. Having found one, we first differenced the data to be the "Log Difference" of the data (Log Difference = $\log(retail_t) - \log(retail_{t-1})$).

We chose the ARIMA model of (1,1,1) to describe the log series (the central 1 represents the differencing of the data) due to its low AIC value and lack of autocorrelated residuals. Also, it passed the mark for both having no further AR or MA unit roots, and for converging. With this new model, I reattempted the forecast for the 2016-2020 period you felt had been misrepresented.

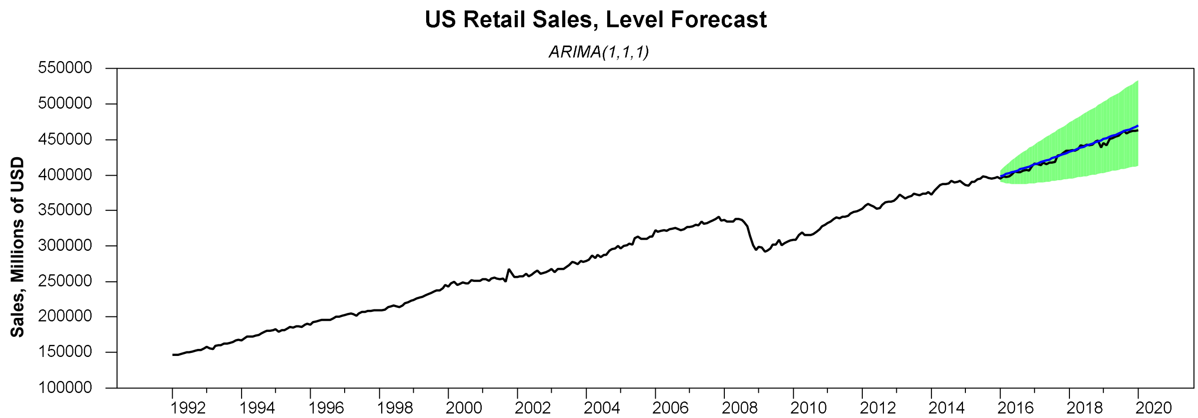

Figure 3 Here is the projection given by the new model, where the blue line is the projection, and the green area is the 95% confidence interval.

As you can see when translated into the raw level forecast, the new ARIMA model precisely predicts the amount of retail sales over the 2016-2020 period. To add a numeric metric by which to judge the success of this model, I tested both the original trend model and the new ARIMA model for their Theil U coefficients. These values give a sense of how well each model predicted the series. Specifically, the Theil U value is calculated via the errors of the model when compared to the actual data, and the errors of a “null model” (a model that predicts this periods values as the same as last period’s). The lower the coefficient, the more accurately a model predicted the real world data.

Since each model had to predict more than just one period into the future, it was appropriate to calculate the Theil U values for several periods ahead in the future. As such, I calculated the Theil U values for the forecasts of each model up until 24 steps into the future. Essentially, I was testing their ability to predict values up until 2 years in the future.

Steps Ahead	Theil U (Original Trend Model)		Theil U (ARIMA (1,1,1) model)		Steps Ahead
1	6.5618	1.1524	0.8975	0.2586	13
2	4.9139	1.0972	0.7780	0.2553	14
3	3.7788	1.0285	0.6899	0.2355	15
4	3.0842	0.9646	0.6029	0.1986	16
5	2.6062	0.9113	0.5328	0.1915	17
6	2.2456	0.8554	0.4623	0.1719	18
7	1.9809	0.8031	0.4183	0.1601	19
8	1.7844	0.7528	0.3624	0.1605	20
9	1.6090	0.7076	0.3215	0.1343	21
10	1.4607	0.6813	0.2988	0.1383	22
11	1.3336	0.6582	0.2897	0.1450	23

Figure 5 Theil U values for the original trend model

Figure 4 Theil U values for the new ARIMA model

12	1.2323	0.6393	0.2713	0.1446	24

There's only one column to look at in these tables, and that's the Theil U categories. As displayed above, the new ARIMA model was more accurate in predicting future data, regardless of how long in the future it was. Notice, however, that the trend model does drop significantly in its Theil U component as time goes on. This means that the trend model did in fact do somewhat well in estimating future data, so long as it was from the perspective of a long timeframe (1.5 years in the future or more). However, even 2 years into the future, the ARIMA model still had a significantly more accurate estimate.

This leads to the final operational question for this assignment, what is the forecast of the next four years for the retail sales market? Here, I present my forecast for the 2020-2024 retail sales figure:

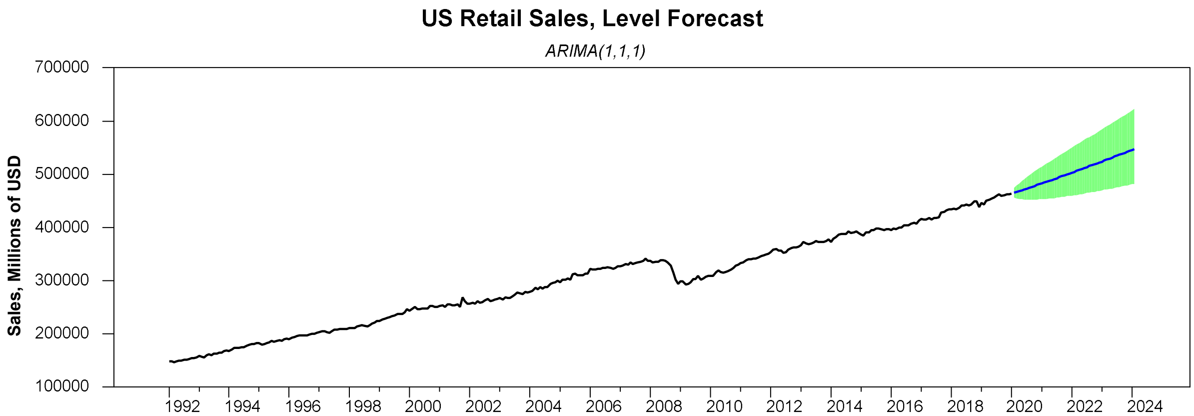

Figure 6 The ARIMA model's forecast of the Retail sales market for the next 4 years. The raw estimate for each year is shown in blue, while the 95% confidence interval is shown in green.

Conclusion

It's tough to answer the question of how certain I am in my forecast. Trying to estimate future scenarios always implies assuming many circumstances will remain the way they are indefinitely. Had the 2008-2009 recession not happened, the original trend model may well have given an accurate estimate of the real data. Without another structural change or shock to the retail industry, I am extremely confident that the ARIMA model will give accurate predictions, even accounting for a slight exponential trend in the series. I will argue in its favor as well by pointing out that the Theil U values for up to 2 years prediction are below 1, implying a forecast that is better than guessing.

All of these forecasts, including the trend model, forecast a bright future for the retail market. While traditional brick and mortar stores are closing, a new wave of online retailers have begun operations. Employment in the retail industry is increasing, and online shopping has not only changed the way people consume, but also revolutionized how chains use their store locations. In fact, amazon has begun to open local retail stores to complement their online retail.

Every numeric and subjective indicator seems to point to a strong retail market today, and a growing one for tomorrow. The retail apocalypse isn't the destruction of our industry, but the revitalization of it.

Focus on using brick and mortar locations as complements for online retail, not substitutes. Think about how to more efficiently and quickly service customers, because the era of e-retail is here.

NOTE: I got here and then I read some stuff about the coronavirus, the projections are not assuming any impact on the retail market, which I know is unrealistic.

RATS Coding

```
*** Intro Code ***
 **Memory/Data**
cal(m) 1900:1
all 2030:1
data(format=fred) * * RSXFS

 **Variables**
set DRSXFS = RSXFS - RSXFS{1}
set LRSXFS = Log(RSXFS)
set LDRSXFS = LRSXFS - LRSXFS{1}
set trend = t

 ** Dates **
compute startdate = 1992:1
compute enddate = 2015:12
compute forestart = 2016:1
compute foreend = 2020:1
compute forestart2 = 2020:2
compute foreend2 = 2024:2

compute counter = 1

****************************************************************************************
graph(vlabel="Log of Sales, Millions of USD", header="US Retail Sales, Log Level") 1
#LRSXFS * *

*** Try recreate previous forecast for 2012-2016 period to find reasons for over optimism ***
** Linear Trend **
linreg LRSXFS startdate enddate resids1
# constant trend
prj fitted1

* Recreate Forecast *
uforecast(from=2016:1,to=2020:1,stderrs=sterr1) fore1
set upper1 enddate foreend = fore1+1.96*(sterr1)
set lower1 enddate foreend = fore1-1.96*(sterr1)

graph(vlabel="Log of Sales, Millions of USD", header="US Retail Sales, Log Level Trend Model Forecast
2016-2020",ovcount=2, $
  overlay=fan) 5
#LRSXFS startdate foreend
#fitted1 startdate enddate
#fore1 * * 3
#upper1 * * 4
#lower1 * * 4
```

```
***************************************************************************************
********
*** Now, Lets Test the Series for a unit root ***
@dfunit(det=trend,method=aic) LRSXFS startdate enddate
*Fail to reject the Null. There is sufficient evidence for us to suspect a Unit Root*
@dfunit(det=trend,method=aic) LDRSXFS startdate enddate

graph 1
#LDRSXFS * *

***Now, lets try to forecast the first difference model ***
** ARIMA Model **
@bjautofit(pmax=10, qmax=10) LDRSXFS startdate enddate

* Trial 1: (1,1,1)
boxjenk(constant,ar=1,ma=1,diffs=1,define=arima1) LRSXFS startdate enddate resids2
@armaroots(equation=arima1) LRSXFS
@autocorr(header="Retail Sales, Log difference, ARIMA Model 1,1,1(residuals)") resids2 startdate
enddate

uforecast(equation=arima1,nostatic,stderrs=Lerror2,from=enddate+1,to=foreend) Lfore2
set Lupper2 enddate+1 foreend = Lfore2+1.96*Lerror2
set Llower2 enddate+1 foreend = Lfore2-1.96*Lerror2

set fore2 = exp(Lfore2)
set upper2 = exp(Lupper2)
set lower2 = exp(Llower2)
set error2 = exp(Lerror2)

graph(style=line,vlabel="Sales, Millions of USD", header="US Retail Sales, Level
Forecast",ovcount=2,ovsame, $
  overlay=fan, subheader="ARIMA(1,1,1)") 4
# RSXFS startdate foreend
# fore2 enddate+1 foreend 2
# upper2 enddate+1 foreend 3
# lower2 enddate+1 foreend 3

***************************************************************************************
********
*** Forecast evaluation ***
*Theil U: Squared mean errors/Squared mean errors (no change forecast [last periods value=this periods
forecast])

** Trend Evaluation **
linreg(define=trnd) LRSXFS startdate enddate resids1
# constant trend
theil(setup,steps=24,to=foreend)
#trnd
```

```
do j = enddate+1, foreend
  compute counter = counter+1
            theil(noprint) j
            linreg(noprint,define=trnd) LRSXFS startdate j resids1
# constant trend
end do j
theil(dump)

** ARIMA (1,1,1) Evaluation **
boxjenk(constant,ar=1,ma=1,diffs=1,define=arima1) LRSXFS startdate enddate resids2
theil(setup,steps=24,to=foreend)
#arima1
do j = enddate+1, foreend
  compute counter = counter+1
            theil(noprint) j
            boxjenk(noprint,constant,ar=1,ma=1,diffs=1,define=arima1) LRSXFS startdate j resids2
end do j
theil(dump)

*************************************************************************************
********
*** Future Forecast ***
** ARIMA (1,1,1) Model **
boxjenk(constant,ar=1,ma=1,diffs=1,define=arima1a) LRSXFS startdate foreend Lerror2

uforecast(equation=arima1a,nostatic,stderrs=Lerror2,from=foreend+1,to=foreend2) Lfore3
set Lupper3 forestart2 foreend2 = Lfore3+1.96*(Lerror2)
set Llower3 forestart2 foreend2 = Lfore3-1.96*(Lerror2)

set fore3 = exp(Lfore3)
set upper3 = exp(Lupper3)
set lower3 = exp(Llower3)
set error3 = exp(Lerror3)

graph(style=line,vlabel="Log of Sales, Millions of USD", header="US Retail Sales, Log Level
Forecast",ovcount=2,ovsame, $
   overlay=fan, subheader="ARIMA(1,1,1)") 4
# LRSXFS startdate foreend
# Lfore3 forestart2 foreend2 2
# Lupper3 forestart2 foreend2 3
# Llower3 forestart2 foreend2 3


graph(style=line,vlabel="Sales, Millions of USD", header="US Retail Sales, Level
Forecast",ovcount=2,ovsame, $
   overlay=fan, subheader="ARIMA(1,1,1)") 4
# RSXFS startdate foreend
# fore3 forestart2 foreend2 2
```

upper3 forestart2 foreend2 3
lower3 forestart2 foreend2 3