Four-Factor Affect The U.S. Retail Sales Indicator & Forecasting

Abdullah A. Aldrees

Abstract— Examining the significant of the following factors on Retail Sales Indicator of the U.S., fossil-fuel prices, Unemployemnt level, Inflation, and excess returns of the S&P 500. Applying multiregression method using STATA, statistical software,

Index Terms— Regression, Economic, Retail Sales Indicator, ARCH, Forecaste, Unemployment, CPI, Inflation

1 Introduction

KNOWLEDGE of Retail Sales indicator is an important piece of information for investors since Retail Sales is considered as one of the leading indicators. Leading indicators are gauges for economic condition or stage, and it more likely to change before the economic changes. According to (Barnes), "the indicator is released around the 13th of the month. It covers previous month's data. Also, it is released by Census Bureau and the U.S. Department of commerce". According to (ECONODAY, 2015), "Retail sales are a major indicator of consumer spending trends because they account for nearly one-half of total consumer spending and approximately one-third of aggregate economic activity".

However, the higher the Retailer Sales increases GDP, and vise versa. Knowing what affects the indicator could be beneficial for investors, retailers, or companies (cyclical companies specifically). Specifically, investors, companies, or portfolio managers could predict the performance of cyclical companies; they could eliminate their losses by either shoring or selling the stocks, it depends on their position. Moreover, the U.S. government could predict where the economy is heading in the future.

In my study, I am trying to test the Retailer Sales indicator as depended variable, and the factors that I think will affect it as independent variables are the following excess return of the S&P 500, percentage change in fossil-fuel prices, percentage change in Consumer Price Index (indicator), and percentage change in unemployment level.

I also examined which forecast model is the best for forecasting the effect of Brent oil prices on Retailer Sales. Since China is the top trader partner with the U.S. in February 2016, I wanted to test if there is any relationship among these three variables. Considering the marvelous piece of information, which is Retail Sales indicator, as a leading indicator for the economy. Also, oil prices significantly affect how consumers spend their money.

During the study, I used several models such as ARIMA to forecast both crude oil prices (Brent) and Retailer Sales, TGARCH (1,1,1) to forecast the volatility of oil prices, and Holt-Winters additive model for Retailer Sales. Finally, I used MGARCH to examine the volatility among Retailer Sales, Brent Prices, and U.S. imports China.

2 Literature Review

Unfortunately, Retail Sales indicator does not have sufficient studies. Still, I found one study of Retail Sales indicator called "Forecasting With Statistical Models and a Case Study of Retail Sales" by Dan M. Bechter and Jack L. Rutner. Their main focus was to forecast Retail Sales by controlling personal income, personal wealth. However, this study is insufficient due to the focus on individual's personal income and wealth. Also, it does not agree with my perspective, which is what factors impact Retail Sales indicator.

3 Rational

What is the relationship between Retail Sales and fossil-fuel prices, unemployment level, CPI, and excess S&P 500 returns? What impact they have on Retail Sales? Are they statistically significant? By identifying these factors and knowing the relationship among them, we could predict the movement of Retail Sales indicator. The first hypothesis is if Fossil-Fuel prices increased, Retail Sales decreases, and vise versa. The second hypothesis is if the CPI (inflation) increased, Retail Sales decrease. The third hypothesis is if the unemployment level increased, Retail Sales decrease, and vise versa. The fourth hypnosis is if the excess returns of the S&P 500 increased, Retail Sales increases.

1. H0: if Fossil-Fuel prices changed, Retail Sales change in the opposite direction.
   H1: if Fossil-Fuel prices changed, Retail Sales stays constant or move in the same direction.
2. H0: if the CPI (inflation) increased, Retail Sales decrease
   H1: if the CPI (inflation) increased, Retail Sales stays constant or move in the same direction.
3. H0: if the unemployment level increased, Retail Sales decrease.
   H_1: if the unemployment level increased, Retail Sales stays constant or move in the same direction.
4. H_0: if the excess returns of the S&P 500 increased, Retail Sales increases.
   H_1: if the excess returns of the S&P 500 increased, Retail Sales stays constant or move in the opposite direction.

If any hypothesis is confirmed, it will indicate that there is

a relationship between Retail Sales and that factor. For instance, if the fourth hypothesis is confirmed, we could watch the excess return of the S&P 500 for next month and compare it with last month if it increased we could predict that Retail Sales will increase. Specifically, since the GDP is released every quarter, and if the hypothesis is confirmed, we could predict the next GDP. On the other hand, if the hypnosis is disconfirmed, I suggest that either change the factor or do some transformation for the data for instance use the Logarithm of the factor.

The rationale behind the study of forecasting is to investigate the relationship and impact of the volatility oil prices on Retailer Sales. Comparing the accuracy of ARIMA model and Holt-Winter is one of the main purpose of this study. In another word, which model has the lowest error of forecast, in terms of outlier. Are forecast errors normally distributed so that the forecast could be valid and the error will represent the actual uncertainty of the forecast? Also, applying TGARCH model to forecast the oil prices with volatility for the one year out of sample. MGARCH is applied in order to obtain the correlation between the three variables, and to see whether it is statistically significant or not.

4 DATA

The needed data were collected from several sources. Unemployment level, Consumer Price Index (CPI), and Retail Sales from Federal Reserve Bank of St. Louis www.research.stlouisfed.org. Fossil-Fuel prices from CFO Magazine. Market Return, risk free rate, and S&P 500 return from Wharton - University of Pennsylvania.

The sample size is based on monthly data since January 2004 until December 2014. The sample was specifically selected before the financial crisis (mid-2006), and within the oil prices failing during 2014. I wanted the sample to be until 2015, but due to the limited data of Fossil-Fuel prices was provided to me the sample was till end of 2014.

The forecasting the data were obtained from Federal Reserve Bank of St. Louis www.research.stlouisfed.org. Using freduse command in STATA with the following names: MCOILBRENTEU, RSXFS, IMPCH, for oil prices, Retailer Sales, and U.S. imports from China, respectively. The sample size is based on monthly data since February 1987 until March 2016. I wanted the sample to capture tremendous oil price movements, especially, after the steep drop in oil price during 2008 and 2014.

5 METHOD

The anticipated method is multiple regression to find out if there any relationship between Retail Sales (Y), which is the dependent variable with the following independent variables:

- Excess return of S&P 500 (X1)
- % Δ in Fossil-fuel Prices (X2)
- % Δ in CPI (X3)
- % Δ in unemployment level (X4)

Multiple regression equation:

$$Y = a + \beta 1\, X1 + \beta 2\, X2 + \beta 3\, X3 + \beta 4\, X4$$

Retail Sales = 0.0122 + 5.2737 (X1) + -3.8690 (X2) + 93.4860 (X3) + -6.7212 (X4)

5.1 Stata Output:

```
. reg RETAILMPCSMSA excsp ffpricechg chacpi chaunmp
```

Source	SS	df	MS	Number of obs	=	120
				F(4, 115)	=	23.48
Model	63.1082285	4	15.7770571	Prob > F	=	0.0000
Residual	77.2647715	115	.671867578	R-squared	=	0.4496
				Adj R-squared	=	0.4304
Total	140.373	119	1.17960504	Root MSE	=	.81968

RETAILMPCS~A	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
excsp	5.273735	1.898315	2.78	0.006	1.513538	9.033933
ffpricechg	-3.868954	1.168832	-3.31	0.001	-6.184185	-1.553723
chacpi	93.48604	26.75546	3.49	0.001	40.48862	146.4835
chaunmp	-6.721149	2.776163	-2.42	0.017	-12.22019	-1.222105
_cons	.0122144	.0871102	0.14	0.889	-.1603178	.1847466

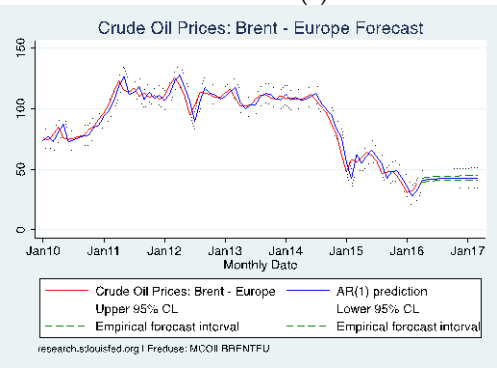
Moreover, excess return of S&P500 has a positive correlation with Retail Sales of 0.3494, Fossil-Fuel prices has a negative correlation of -0.5621, percentage change in CPI has a positive correlation of 0.5230, and percentage change in unemployment level has a negative correlation of -0.2651. Plus, all the factors have statistically significant at 5% of significance level. Each variable has 120 observations. Finally, in the future I will attempt to discover another model such as nonlinear regression that might explain or fits the data. Till this point, multi regression analysis is a marvelous model to predict the unknown coefficient of the independent variables from known values (past data)

6 ANALYSIS

Firstly, F (4, 115) = 24.17, P < 0.001. Thus there is a statistically significant relationship among Retail Sales, Excess return of S&P 500, % Δ in Fossil-fuel Prices, % Δ in CPI, and % Δ in unemployment level; the r-squared is 44.96%, which means the model explains 44.96% of Retail Sales movement by the other factors. Likewise, the coefficient of the factors (S&P 500, FFP, % ΔI n CPI, % Δ in unemployment level) are the following 5.2737, -3.8690, 93.4860, and -6.7212, respectively. All the factors are statistically significant at 5% level of significance, which means they confirmed the mentioned hypothesis above. On the other hand, the constant is 0.0122, but it is not statistically significant. To illustrate, when the excess return of the S&P 500 increases by 1%, Retailer Sales increases on average by 5.2737 units; when the gas prices increases by 1%, Retail

Sales decreases by 3.8690. Also, if CPI increases by 1%, Retail Sales increases by 93.4860 units. Finally, if the unemployment level increases by 1%, Retail Sales decreases by 6.7212 units.

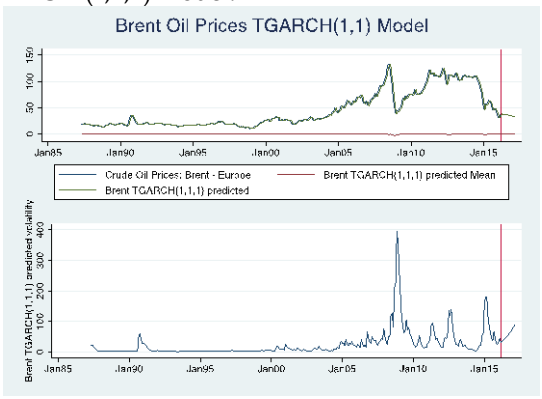
For the Crude Oil forecast models, Box Jenkins: Auto-regressive model of order one AR(1)



The best model in the class of ARIMA models is AR(1) on the first difference since it has the lowest Schwarz's Bayesian information criteria (BIC). The original time series has a unit-root due to the Dicky-Fuller test is not rejected at any acceptance confidence level. This is an indicator that the time series is not weakly stationary. The model predicts an increase of oil price for the following year, about 30% or $10. See the table below for price forecasting with 95% upper and lower confidence intervals and empirical forecast intervals (25th, 75th percentiles of residuals)

In terms of accuracy, the model has 23 outliers, 6.63%. Mean Squared Error, Mean Absolute Error and Percentage Error of the predicted values from the model are 17.32, 2.85, and 6.82%, respectively. However, the errors are not normally distributed. The model and lag is statistically significant since p-value is less than 0.001, Wald chi2(2) is 211.51. To conclude, if oil prices continue to rise, we might see a decrease in Retailer Sales since individuals will have to pay more for gas and less for other products.

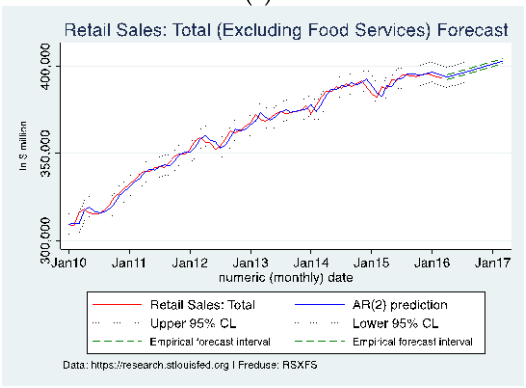
TGARCH (1,1,1) model:



The best model in ARCH family is TGARCH (1,1,1) on the first difference. Comparing TGARCH with other models, it has the lowest information criteria. As mentioned before, the time series contains unit root. The model predicts a rise in oil prices for the next five months about 20% or $6 then decrease about 15% or ($5). Also, the volatility increases for the next year.

From accuracy's standpoint, Mean Squared Error, Mean Absolute Error and Percentage Error for the forecasted values are 19.70, 2.90, and 6.85%, respectively. The lags are statistically significant, and the errors are not normally distributed, so robust option was included in the model to solve this problem. In conclusion, AR(1) is better than TGARCH in terms or accuracy. However, TGARCH has one advantage over AR(1) model, which is the volatility forecast.

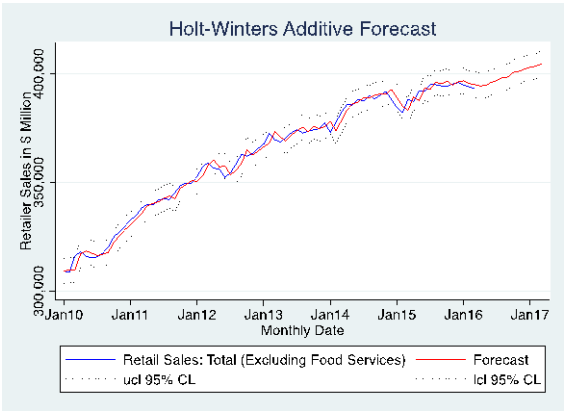
For the Retailer Sale Indicator forecast models, Auto-regressive model of order two AR(2)



AR(2) on the first difference is the best model among ARIMA family, and it has the lowest BIC value. The original time series contains a unit-root according to Dicky-Fuller test for unit-root. The model is statistically significant at 10% confidence level, and predicts upward trending. It could indicate that the economy will grow about 2%. The following table has the forecasted values.

In terms of accuracy, the model has 14 outliers, 4.81%. Percentage Error of the predicted values from the model is 0.73%. However, the errors are not normally distributed1. The model and lag is statistically significant at 10% Level, Wald chi2(2) is 3.24.

Holt-Winters Additive Model:



Since the time series is seasonally adjusted, in this case, Holt-Winters Additive model is the best to use. The errors are not normally distributed. The model has 0.74% percentage error and 4.12% outliers. The model predicts upward increasing for Retailer Sales. From an accuracy perspective, AR(2)

model is better than Holt-Winters model.

4 Conclusion

To sum up, understanding what affects Retail Sales could lead to a better economic condition of a country or it might warn the individuals, merchandisers, investors, companies, or even the government about what will happen to the GDP or where is the economic cycle is heading to. Moreover, it can, also, be cost saving for companies who produces goods such as cars or electronic devices. If the company would know that the Retail Sales would decrease, they would either reduce their production or predict lower sales level compared with last month.

In addition, Autoregressive model is better for oil prices than TGARCH in accuracy manner. AR(2) is better model than Holt-Winter due to its accuracy in forecasting. Also, increase in volatility of Retailer Sales increases volatility in oil prices. All the variables are correlated with each other, however, just oil prices and Retailer Sales is statistically significant.

For Further research's ideas, the researcher could test Retail Sales with GDP, and what percent change of Retail Sales could affect GDP change. Furthermore, the researcher could examine inflation (CPI) as dependent variable and Retail Sales as independent and see the relationship and what happen. Thus the researcher could test Retailer Sales as dependent variables and specific industry's excess return as independent variable. Which industry is more sensitive to Retailer Sales? Finally, I would identify which industry contributes toward the Retail Sales indicator, by doing so, we could do more analysis to know how and why this industry is more volatile to Retail Sales, and how they react toward the listed factors. Additionally, the researcher could test Retail Sales with China's currency, and what are the volatility in the currency and Retail Sales. Additionally, the researcher could examine inflation (CPI), supply and demand of oil prices both domestic and international. Also, I would suggest that test and see the relationship between market return, S&P 500, and DJIA using ARCH family to acquire the volatility of each index. And examine if increasing in RS, could increase the dividend payout.

References

[1] Barnes, R. (n.d.). Economic Indicators: Retail Sales Report. Retrieved from Investopedia: http://www.investopedia.com/university/releases/retailsales.asp

[2] Bechter, D. M., & Rutner, J. L. (1978, March 1). Forecasting With Statistical Models and a Case Study of Retail Sales. Retrieved November 28, 2015, from Kansas City Fed: https://www.kansascityfed.org/PUBLICAT/ECONREV/econrevarchive/1978/1q78bech-2.pdf

[3] ECONODAY. (2015, Nov 11). Retail Sales. Retrieved from 2015 Economic Calender: http://mam.econoday.com/byshoweventfull.aspx?event_id=274&cust=mam

[4] Mintz, S. (2008, 12 1). What Goes Down Will Come Up. Retrieved 11 18, 2015, from CFO Magazine: http://ww2.cfo.com/strategy/2008/12/what-goes-down-will-come-up/