Fall
2019

Forecasting Pepsi Quarterly Earnings with Explanatory Variables

ST 534 FINAL REPORT

YUGUO WU
DANIEL PETERS
ARUN KAKANI

Goal

In this project our group wanted to model Pepsi's quarterly revenue as a time series with the help of an explanatory variable and forecast its values over the next 12 quarters. Using stockpup.com we were able to find a dataset of Pepsi's quarterly revenue from Q4 1994 through Q2 2019 and 39 other related variables including share price, ROE, capital expenditures etc.

Discussion of Approach

To begin our analysis we started by graphing the quarterly revenues to gain a better understanding of the data. We then looked for explanatory variables candidates. That process can be summarized in the following steps:

1. Look at data of the explanatory variable.
    a. Does it need a difference
    b. Does there seem to be seasonality?
2. Take a difference(s) if necessary.
    a. Answer questions from step 1
    b. Is white noise a good fit for the residuals?
3. Estimate model if white noise is a good fit for the residuals?
    a. Is white noise a good fit for the residuals?
4. Identify a model relating revenue to capital expenditures.
5. Estimate a model relating revenue to capital expenditures (undifferenced).
6. Estimate a model relating revenue to capital expenditures with appropriate differences.
    a. Are crosscorrelation behaviors appropriate, near zero for negative lags?
7. Look at outliers.
8. Forecast revenue for 12 quarters.

This process was repeated until we found a variable that produced satisfactory results at each step. That variable was capital expenditures. Due to the volume of output from these steps we have only included output related to capital expenditures. This output is annotated and appended at the end of this report, in the results section it is referenced when appropriate.

Results

Look at data of the explanatory variable

- Looking at the results, we can see that, on page (1), the observations show no deterministic trend, so looking at the single mean row for the Dickey-Fuller Test, we see that the P-Value is <0.0001 showing that we don't have to to take a difference at lag 1.
- On page (1), looking at the ACF, we see that the decay in the ACF is slow and the spikes in the ACF are at 4 lags apart; as the data is quarterly. Given this, we take a difference at lag 4 at first.

Take a difference(s) if necessary

- Given the above, we difference the explanatory variable at lag 4. Doing this shows, from the Autocorrelation Check for White Noise on page (2), given the low P-Values that the differenced explanatory variable is not white noise.

Estimate a model if white noise is a good fit for the residuals

- Looking at the ACF, PACF and IACF on page (2), it looks like the residuals might fit an **ARMA(1,1) x $(0,1)_1$.**
- Now on page 4, looking at the Autocorrelation check, we see that the column Pr > ChiSq shows that the p-values are adequate and an OK fit, so we can move onto the next step.
- Applying the same fit to the dependent variable, Revenue, we can get the cross correlation; page (8).
- Looking at the cross correlations, we see that the first spike is at lag 0, so b = 0 and they also begin to decay at lag 0, so u = 0. As the decay of the cross correlations is in the form of damped exponential function, r = 1.

Identify a model relating revenue to capital expenditures

- First, we used:

  estimate input=(0$(0)/(1)capital_expenditures)

  This gave us ACF, PACF and IACF that taper off exponentially and the white noise probability is very high; page (10) . Also, looking at the Conditional Least Squares Estimation on page (9) shows that the p-value of MU is not statistically significant. Given this, we have to use a different model.

- Secondly, we used:

  estimate q=(1,4) p=(1,4) input=(0$(0)/(1)capital_expenditures)

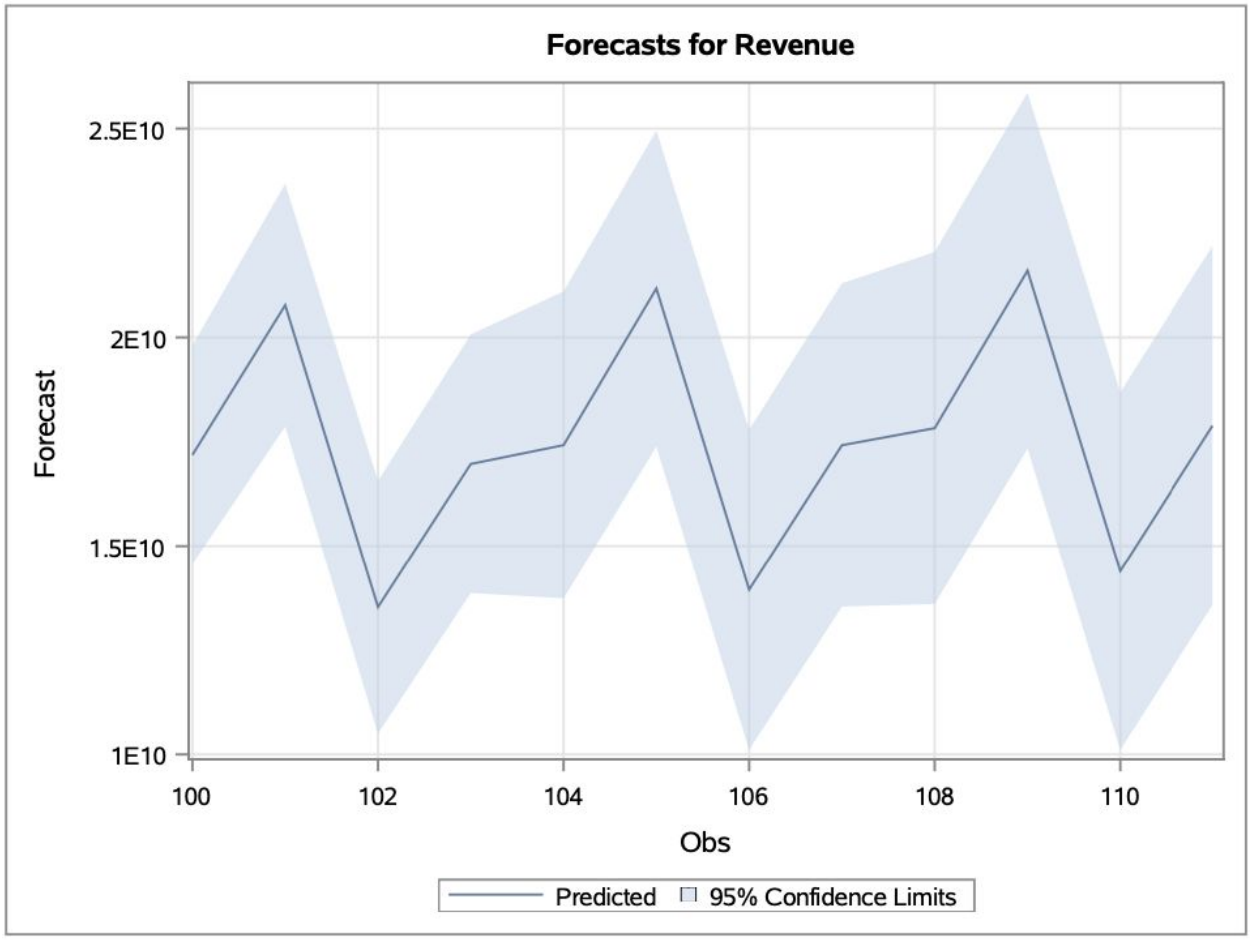
  This gave us, on page (14), ACF, PACF and IACF seem to fit and looking at the Autocorrelation Check of Residuals, the Pr > ChiSq column shows that the model is adequate given the p-values. Now the Conditional Least Squares Estimation on page (13) shows that all the p-values are statistically significant. Therefore, this model is the one that we are going to use for the forecasts.

Look at outliers

- Looking at the Outlier Details on page (16), there are two outliers detected, but as the p-values for both is <0.0001, we can reject their effect on the model.

Forecast revenue for 12 quarters

- page (17)



Referencing the SAS output in relation to the steps previously outlined

Step 1 - Page 1

Step 2 - Page 2

Step 3 - Page 3 - Page 5 (Ends at moving average factors)

Step 4 - Page 5 (Begins at variable name) - Page 8 (Ends at cross correlation graph)

Step 5 - Page 8 (Begins at preliminary analysis) - Page 11(Ends at cross correlation table)pg

Step 6 - Page 11(Begins at model for revenue) - Page 15

Step 7 - Page 16 (Outlier tables)

Step 8 - Page 16-17