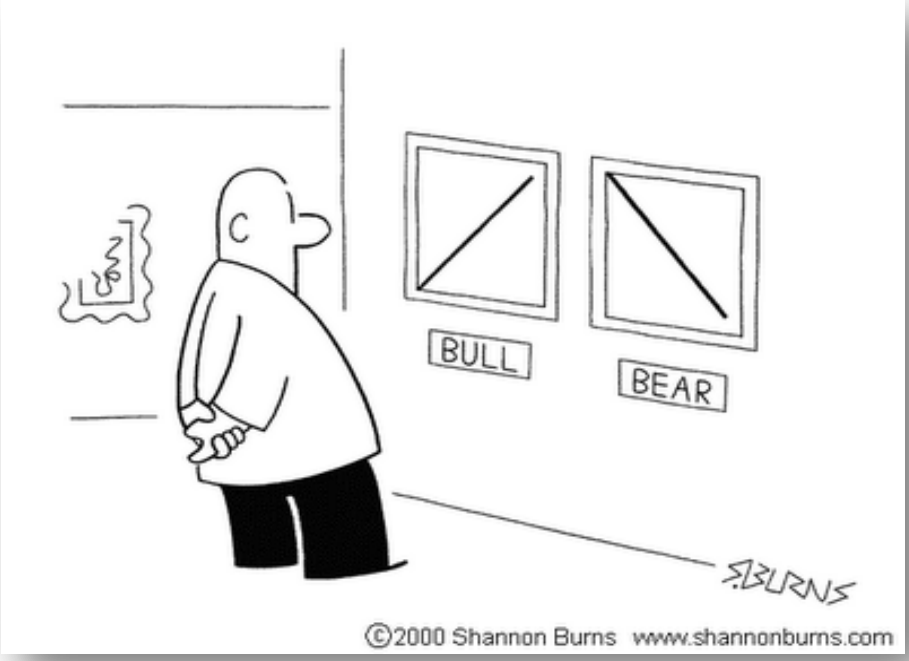
Predicting Changes in Quarterly Corporate Earnings Using Economic Indicators



[group A8]

prashant kumar bothra | piyush mathur | chandrakanth vasudev | harmanjit singh

term project report. December 28th, 2011

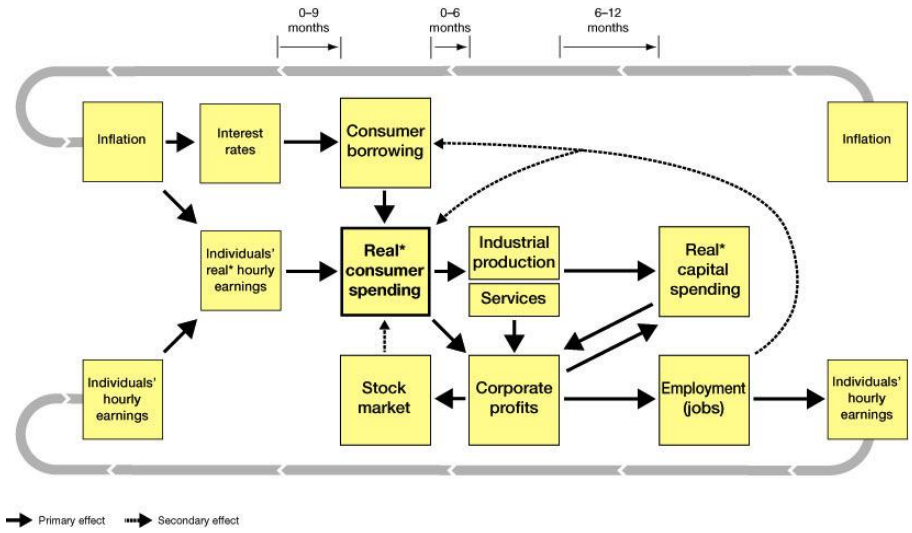
CONTENTS

Executive Summary

The purpose of our project is to check the validity and potentially strengthen an existing theory of business forecasting developed by Joseph H. Ellis (former research analyst at Goldman Sachs). Mr. Ellis' method looks at year over year percent changes in economic variables to predict trend reversals in corporate earnings (he uses S&P 500 EPS as a proxy) – purely from a visualization perspective. We have identified real interest rates, and annual percent changes in Inflation, Real Average Hourly Earnings, Real Personal Consumption Expenditures, Industrial Production, and Real Capital Spending as our potential predictor variables. Using data mining techniques discussed in this report, we have developed a mathematical model to predict annual percent changes in S&P 500 EPS (our dependent variable). Ultimately, this model can be used to create buy and sell signals for investors in the stock market.

The Joe Ellis Theory [in brief.]

There are four stages in economic downturns: 1) the peak, 2) modest slowing, 3) intensifying worrying by investors (a lot of panic selling occurs in this stage), and 4) the advent of recession. However, by the time a recession is officially announced by the National Bureau of Economic Research (official definition: two consecutive quarters of GDP growth), the damage has already been done! By then, economy is actually on an upturn, and yet investors are still selling off and panicking because of the media hype. **The key question, then, from an investor's perspective is: can we predict the economic slowdown in corporate earnings (note: from this point on, for consistency, we will refer to corporate earnings as: S&P 500 EPS) well in advance? In other words: when should an investor ideally sell his stocks? When should he start accumulating again?** After years of researching stocks and the financial markets at Goldman Sachs, Mr. Ellis found the following relationships between annual economic variables and their use in predicting swings in the S&P 500 EPS: Inflation and interest rates are leading indicators



of changes in real average hourly earnings. There is a 0-9 month lag between year-year changes in real average hourly earnings and its effect on year-year changes in real personal consumption expenditures. 0-6 months until changes in real personal consumption expenditures affects year-year changes in industrial production. Another 6-12 months between changes in industrial production and year-year changes in real capital spending. And finally, another 6-12 between changes in real capital spending and its effects on year-year changes in S&P 500 EPS. In summary, observing the above relationships allows us to be prepared for swings in S&P 500 EPS several quarters in advance.

The Data: Characteristics and Processing

Data Retrieval: Our first step was to collect the data from the websites / online databanks of various US agencies (see **Table 1** below). Of course, the downloaded sets of the different data items differed in start dates – Industrial Production data went back to 1919, while Real Average Hourly Earnings was available from 1964. To avoid data mining bias created from including more data for one variable and less for another, we used the 1st quarter of 1964 as the starting point for the entire raw dataset.

Calculating Annual Percentage Values: We calculated year over year percent changes for all variables (except interest rates). As discussed earlier, using annual percent changes versus quarter-quarter or month-month is preferred because the latter two methods produce too much volatility / noise (as observed in a time-series graph). Some data items came in monthly values. We first delegated them into their respective quarters (January through March was Q1), then calculated annual percentage changes, took their trailing three month averages, and finally averaged these values on a quarterly basis.

While shuffling through the data, we noticed only one extreme outlier and normalized it relative to its neighboring data points. This outlier was caused by an absolute EPS increase of $17.25 in Q4 2009 compared to Q4 2008, or approximately 19,200%!

Collected. Cleaned. No anomalies. And with 187 quarters of data values each for seven x variables and our y, we were now able to proceed with the visualization part of our analysis.

Table 1

Raw Data Sources [all publicly available data.]

VARIABLE	SOURCE	WEB LINK
✓ REAL AVERAGE HOURLY EARNINGS	BUREAU OF LABOR STATISTICS	www.bls.gov
✓ REAL INTEREST RATES ✓ INDUSTRIAL PRODUCTION	FEDERAL RESERVE	www.federalreserve.gov
✓ INFLATION ✓ REAL PERSONAL CONSUMPTION EXPENDITURES ✓ REAL CAPITAL SPENDING	BUREAU OF ECONOMIC ANALYSIS	www.bea.gov
✓ S&P 500 INDEX EPS	STANDARD & POORS	www.standardandpoors.com

For simplicity , we coded the variables used in our analysis according to their specific characteristics. Referring to Table 2 below, the prefix "**Q**" represents quarterly, "**R**" is for real (adjusted for inflation), followed by an abbreviation of the variable name, and finally, where appropriate, we added the suffix YY% to indicate the use of annual percentage change of the said variable.

Table 2

VARIABLE	ABBREVIATION
✓ INFLATION	QPCE_INFL_YY%
✓ INTEREST RATES (REAL)	QINTRATE
✓ REAL AVERAGE HOURLY EARNINGS	QRAHE_YY%
✓ REAL PERSONAL CONSUMPTION EXPENDITURES	QRPCE_YY%
✓ INDUSTRIAL PRODUCTION	QPROD_YY%
✓ REAL CAPITAL SPENDING	QRCAP_YY%
✓ S&P 500 INDEX EPS (T-1, Y/Y % CHANGE)	LAG1
✓ S&P 500 INDEX EPS	QEPS_YY%

Visualization

Having cleaned up the data, we needed to figure out if any underlying patterns existed. (For e.g., we needed to determine if there were any causal relationships between our chosen Y (S&P EPS) and the potential predictor variables). With this goal in mind, we ran the following visualization tools:

Line Graphs

For our model to be useful, we needed the data to demonstrate some kind of a causal relationship between the Y and the Xs. Plotting line graphs (against time in Quarters), therefore, seemed like a good idea. Since we were betting on a lead/lag relationship between many of the variables, we plotted all of them, pair-wise. The results (shown in Exhibits 1 through 9) confirmed some of our suspicions.

1. **RPCE vs EPS**: Changes in real personal consumption expenditures *leads* changes in S&P 500 EPS
2. **RPCE vs RPROD**: Changes in real personal consumption expenditures *leads* changes in industrial production
3. **RPROD vs RCAP**: Changes in industrial production *leads* changes in real capital spending
4. **RAHE vs RPCE**: Changes in real average hourly earnings *leads* changes in real personal consumption expenditures
5. **RCAP vs EPS**: real capital spending *leads* changes in S&P 500 EPS

Scatter Plots

To obtain further insights into the nature of the relationship between the variables involved, we proceeded to use scatter plots. Here too, as in line graphs, variables were plotted pair wise. Results are shown in Exhibits 2 through 10. The plots gave us an idea on what the trends were, whether the relationships were positive, negative, etc. and what kind of a trend line fits the pair. They also helped us in identifying outliers that could potentially be discounted when coming up with a predictive model.

Autocorrelation Function

When dealing with changes in the S&P 500 EPS, it made intuitive sense to us that there could be a correlation among indices form consecutive quarters. Before pursuing this path any further with our prediction models though, we needed to substantiate this. We used an ACF (Autocorrelation Function) plot (Exhibit 11) to determine if our assumption holds true or not. What we found was that there existed a definite correlation between the S&P EPS for any given time period (a quarter in this case) and quarters prior to t (t-1, t-2, t-3, etc.).

However, for our purposes, we needed take only (t-1) into consideration since it subsumed the effect of every quarter prior to it.

Prediction

Strategy for Prediction

For purposes of prediction we went beyond the quarter lags as recommended in the book. We considered the following scenarios as indicated below:

Scenario 1: RPCE lagged RAHE by 3 quarters, PROD lagged RPCE by 5, RCAP lagged RPCE by 7, and finally and EPS lagged RPCE by 9 quarters. This scenario was determined based on our visualizations.

Scenario 2: RPCE lagged RAHE by 2 quarters, PROD lagged RPCE by 4, RCAP lagged RPCE by 6, and finally and EPS lagged RPCE by 8 quarters.

For both the scenarios we ran the prediction techniques both with and without Quaterly Lag (referred to as Lag_1 henceforth) as one of the variables. The results were best with Scenario 1 and results for this scenario are the ones explained below.

Data Partitioning

We partitioned the data into 3 sets:

1. Training Data (50%)
2. Validation Set (30%)
3. Test Data (20%)

Modeling

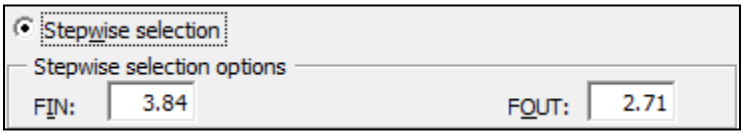
Prediction Trees

To identify the top predictors, we first ran Regression Trees (Both Full and Best Pruned) using XLMiner (CART). For the Full tree we set the max size of leaf nodes to 1. Exhibits 12 and 13 show the Prune Tree outputs (snapshot) we got both with/without Lag_1 as one of the main predictors.

The trees were really insightful as they revealed the potential top predictors. We followed this up with Multiple Linear Regression as described below:

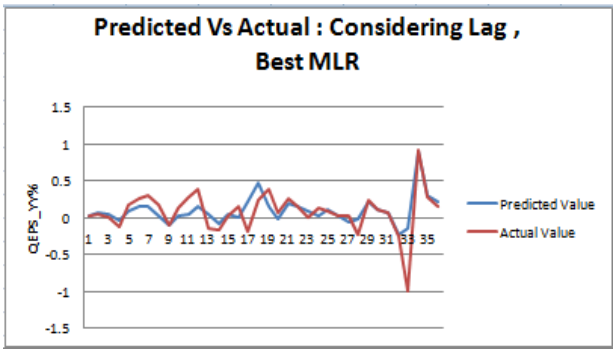
Multiple Linear Regression

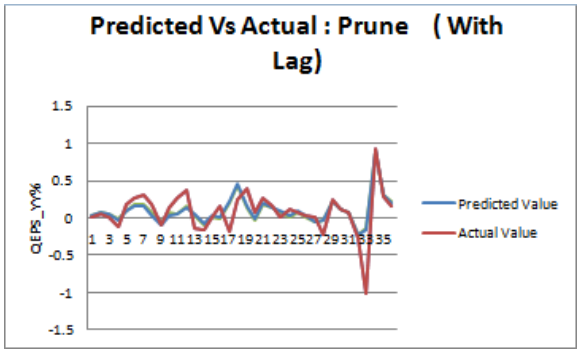
We ran MLR on our data partitions both with and without Lag_1 as one of the top predictors. We also included the 'Best Subset' option with 'Stepwise selection' as the algorithm of choice.

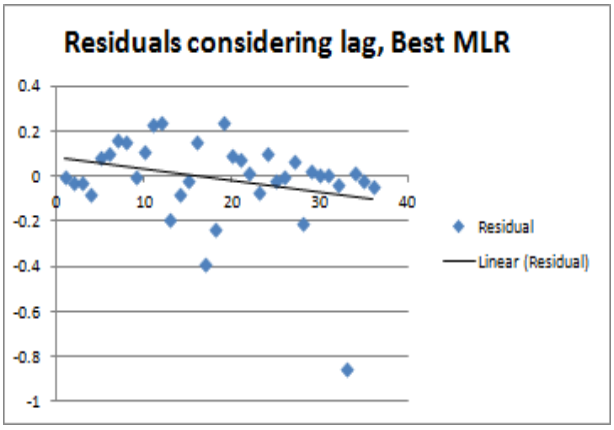


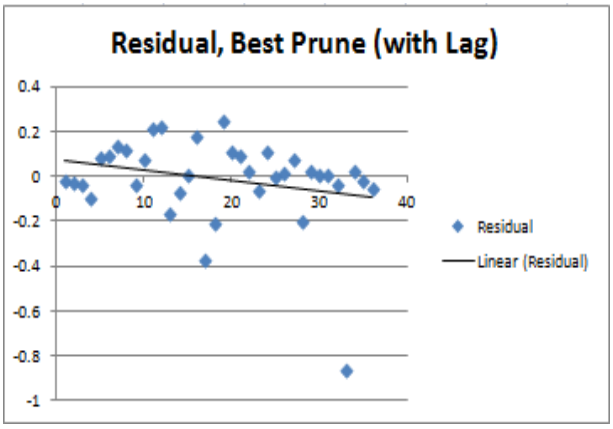
The revelations from MLR (Exhibits 14 and 15) were vastly different from what the Prediction Trees predicted.

Since the results were different from the Prediction tree results, we decided to run MLR with the Best Subset predictors and Pruned Tree predictors. We then plotted the Actual Vs Predicted values from both the outputs as shown below.









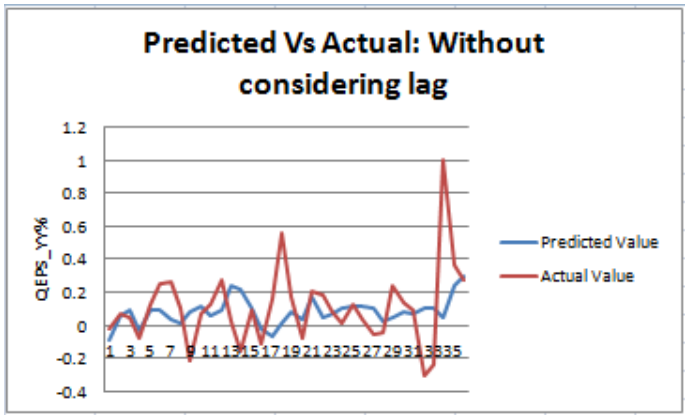
Total sum of squared errors	RMS Error	Average Error
1.3352591	0.192589	-0.0136721

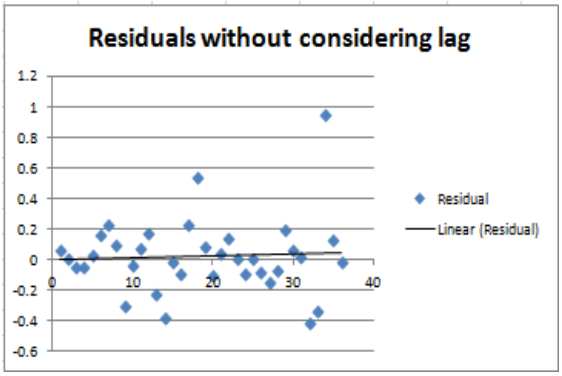
Total sum of squared errors	RMS Error	Average Error
1.2947012	0.1896416	-0.0131939

Following are the interesting observations from the charts above:

1. The predicted values are reasonably close to the actual values (except for the one extreme outlier)
2. Both MLR and Pruned Tree good pretty good results
3. RMSE is actually better with the predictor set recommended by Pruned Tree compared to that recommended by MLR.

The charts below show the results of running MLR (Best subset) without Lag_1 as one of the predictors.





Total sum of squared errors	RMS Error	Average Error
2.0732608	0.2399804	0.0227832

Clearly without Lag_1, the model was doing a poor job of predicting EPS. So we did not explore this option any further.

As can be seen from above, the best option was to consider Lag_1 as one of the predictors. In choosing between the Best Subset from MLR and Pruned Set predictors, we decided to be parsimonious since the other variables were not improving the prediction significantly. This was a crucial decision since fewer variables makes it easier for the user of our model to predict.

The final model we settled on was:

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.04853917	0.01487374	0.00158641	0.62015408
QRCAP_YY%	-0.5165928	0.17860588	0.0048546	0.0002402
Lag_1	0.74696863	0.07495416	0	1.40344453

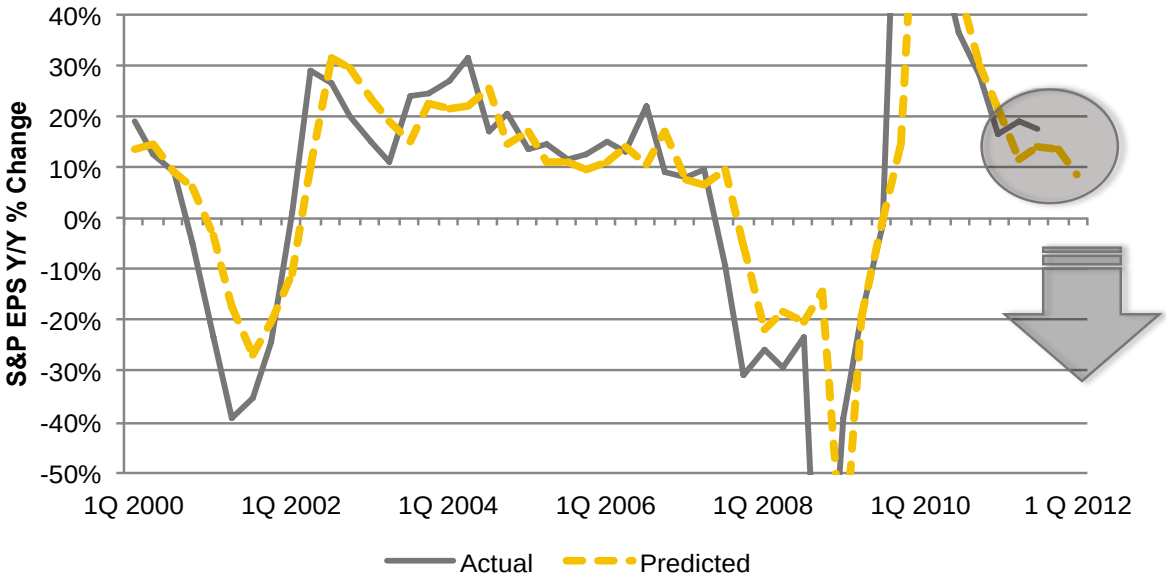
$$QEPS\_YY\%_{(t)} = 0.0486 + 0.747*QEPS\_YY\%_{(t-1)} - 0.517*QRCAP\_YY\%_{(t-2)}$$

The scores of the MLR tests are shown in Exhibit 16

Conclusion

Based on our MLR model, we are able to predict changes in S&P 500 EPS 1 quarter ahead. Why 1 quarter? Because Real Capital Spending is on a 2 quarter lag basis, and Lag_1, or S&P 500 EPS(t-1) on a 1 quarter lag basis. In numbers, this means we get to use 2Q 2011 Real Capital Spending YY% and 3Q 2011 Lag_1, yielding us 13.64% y/y% for 4Q S&P 500 EPS. However, we can strengthen our model if we were to use economic forecast estimates based on fundamentals from industry experts, economists, or estimates often published by the top investment houses. The graph below is our attempt at predicting changes in S&P 500 EPS up to Q1 2012. Lo and behold, we see that S&P 500 EPS is actually slowing down! Very much in line with the overarching theory advocated by Mr. Ellis that when changes in real capital spending slow down, S&P 500 EPS will slow down as well two to four quarters down the line. The following is an excerpt from Mr. Ellis' website:

"Slowing real-wage growth… indicates that Y/Y growth in real consumer spending will deteriorate over the next 1-2 years. This suggests that corporate-profit (S&P 500) earnings growth will also suffer, and raises a strong possibility that the stock market may be headed for another decline." October 14th, 2011



Our conclusion: Sell. Sell. Sell Now. Our model strengthens Mr. Ellis' claim that there is an economic downturn in the US approaching.

APPENDIX

Exhibit 1: Changes in Real Personal Consumption Expenditures ultimately leads to changes in S&P 500 EPS

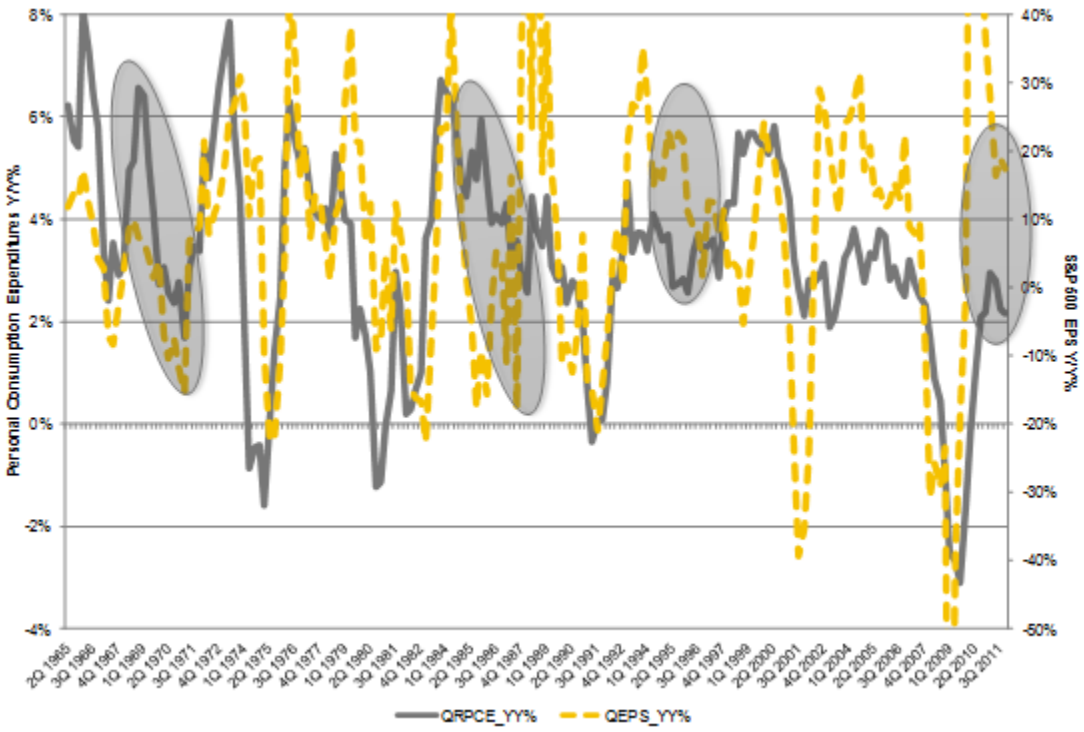


Exhibit 2: Scatter Plot of a) Changes in Real Consumption Expenditures and b) Changes in S&P 500 EPS

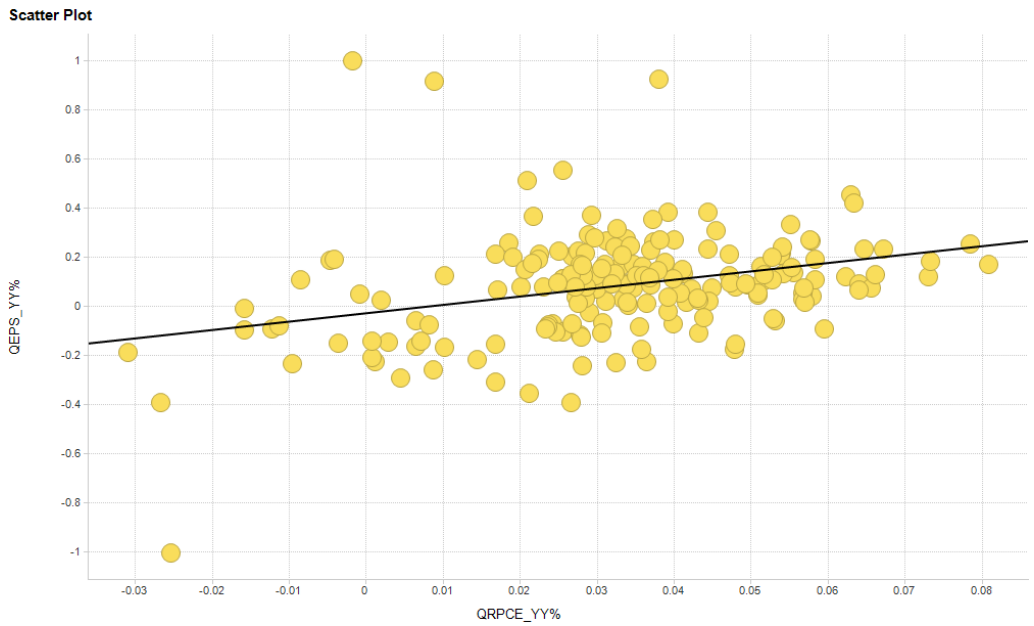


Exhibit 3: Changes in Real Average Hourly Earnings leads to changes in Real Personal Consumption Expenditures

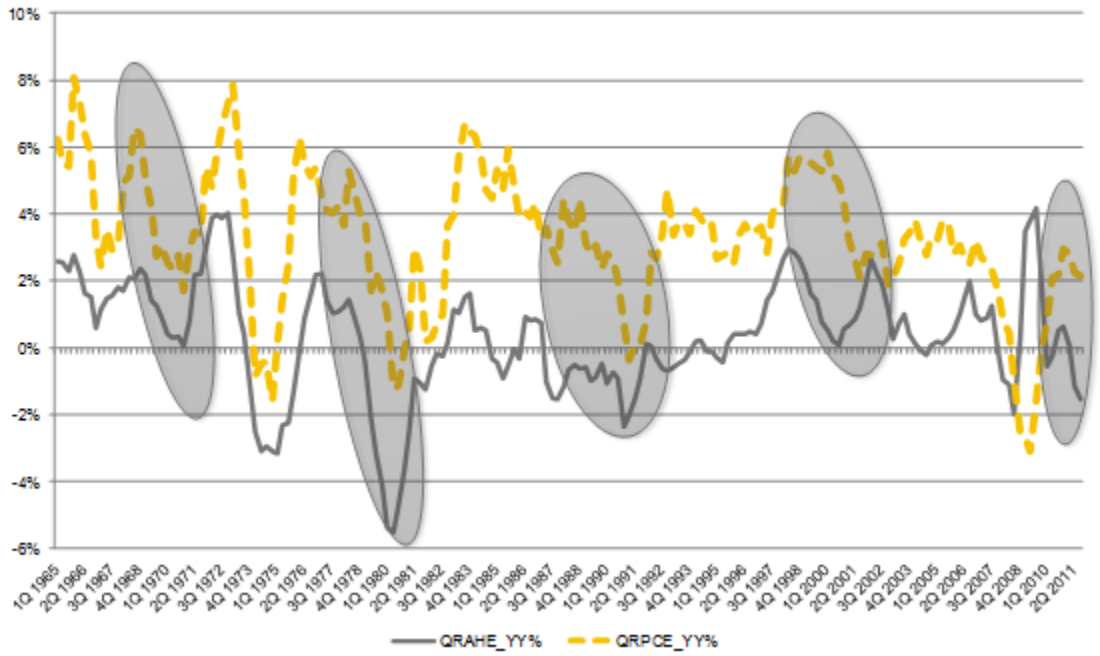


Exhibit 4: Scatter Plot of a) Changes in Real Average Hourly Earnings and b) Changes in Real Personal Consumption Expenditures

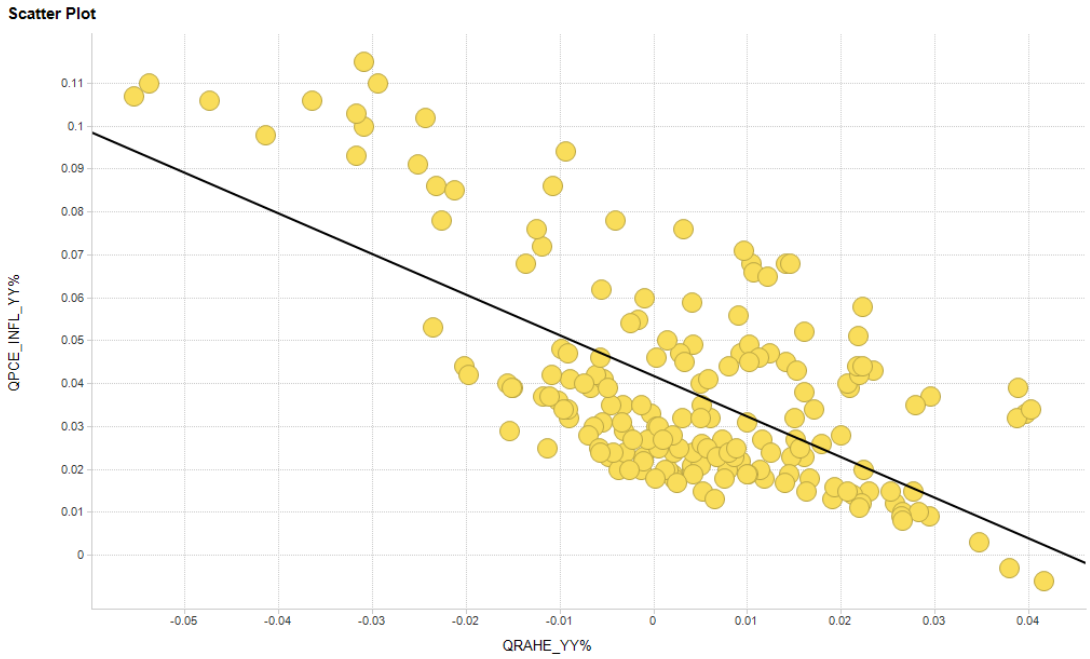


Exhibit 5: Changes in Real Personal Consumption Expenditures leads to changes in Industrial Production

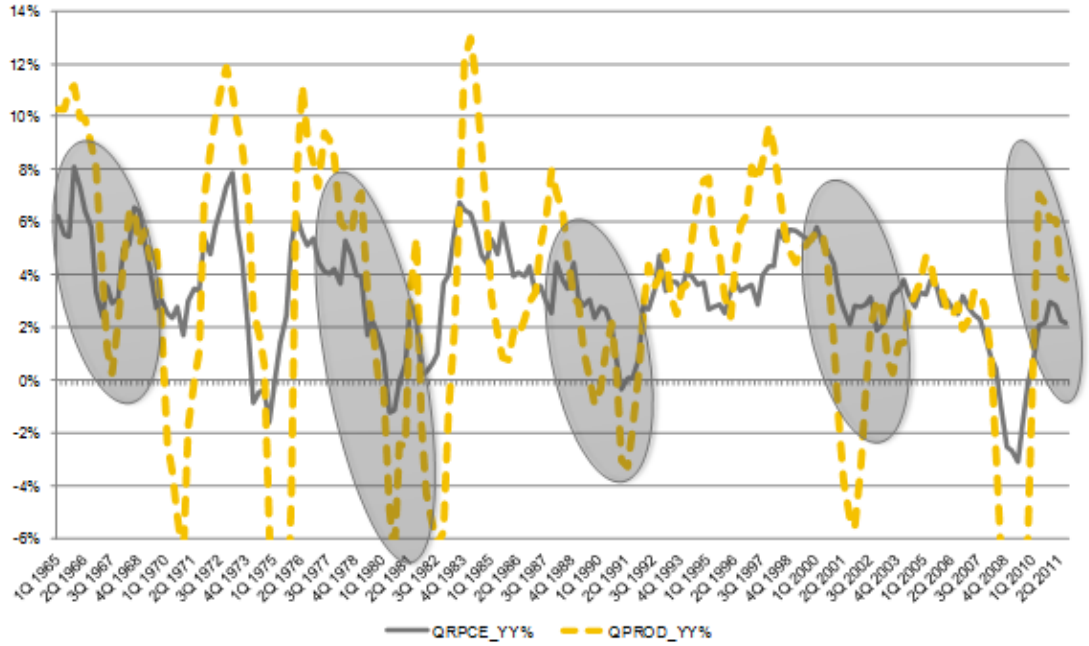


Exhibit 6: Scatter Plot of a) Changes in Real Personal Consumption Expenditures and b) Changes in Industrial Production

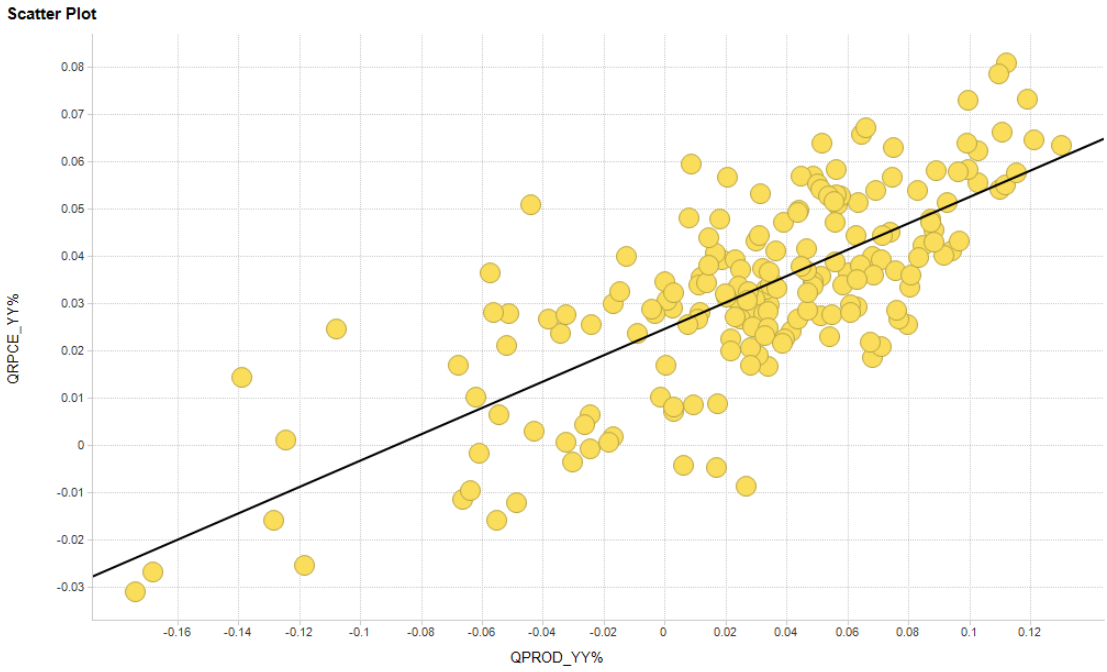


Exhibit 7: Changes in Industrial Production leads to changes in Real Capital Spending

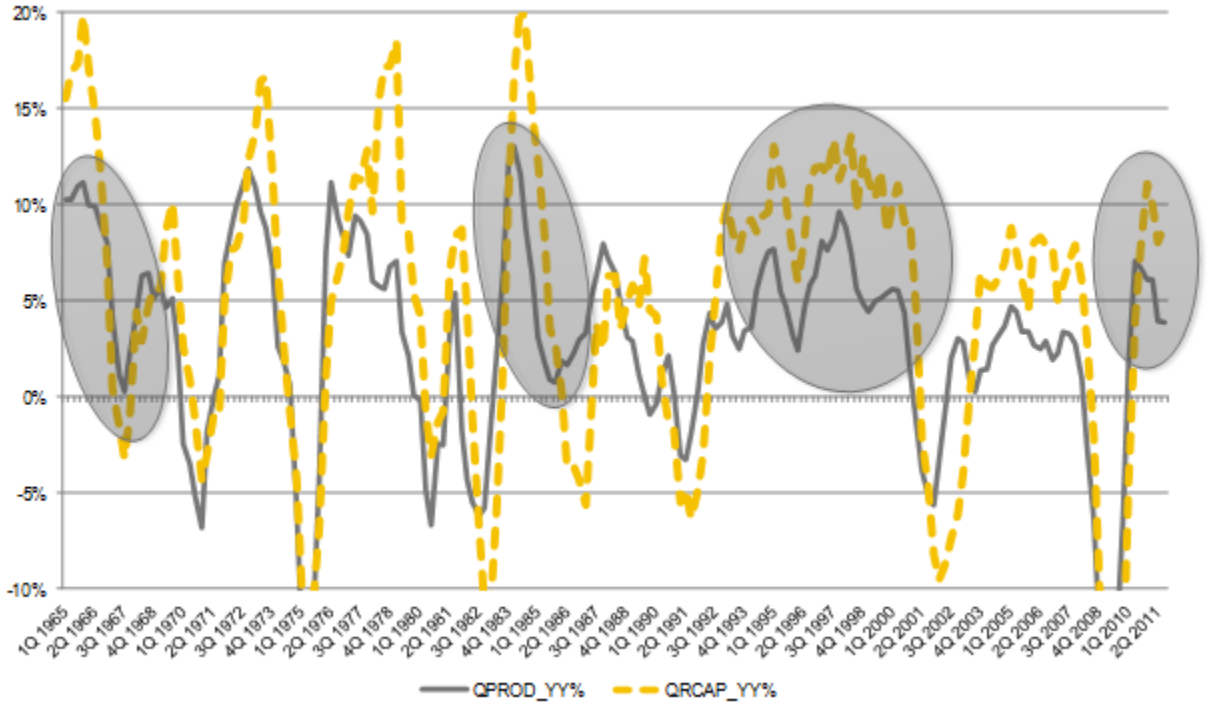


Exhibit 8: Scatter Plot of a) Changes in Industrial Production and b) Changes in Real Capital Spending

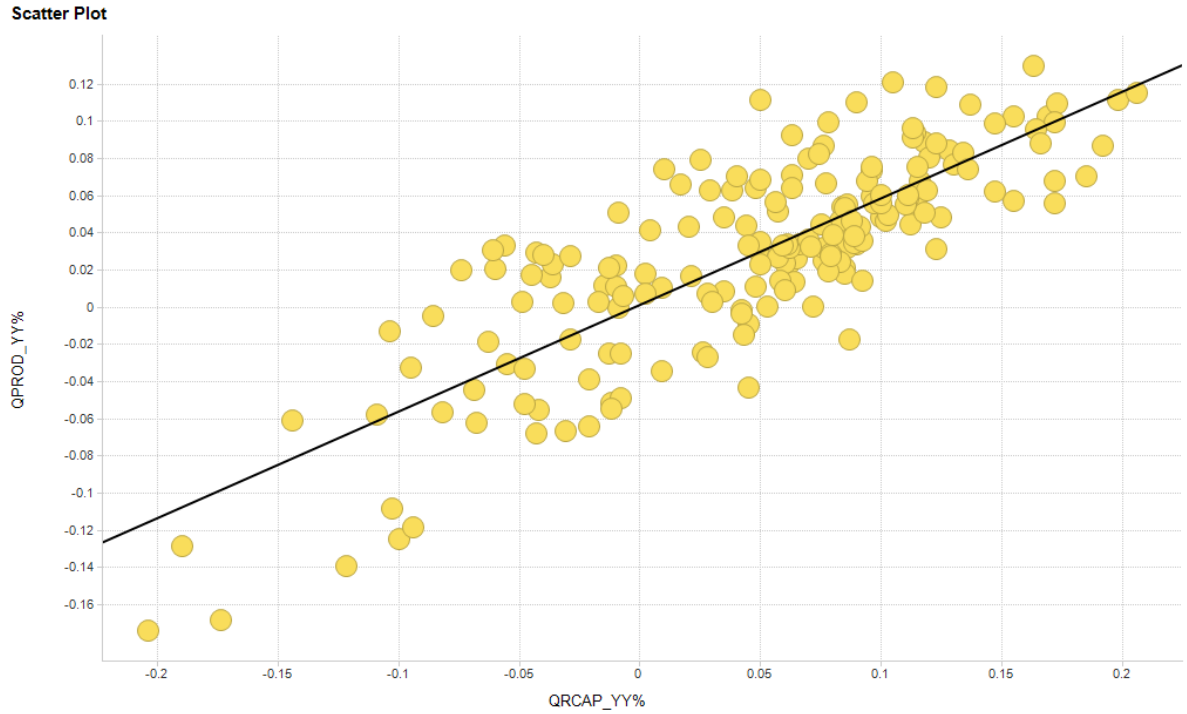


Exhibit 9: Changes in Real Capital Spending leads to changes in S&P 500 EPS

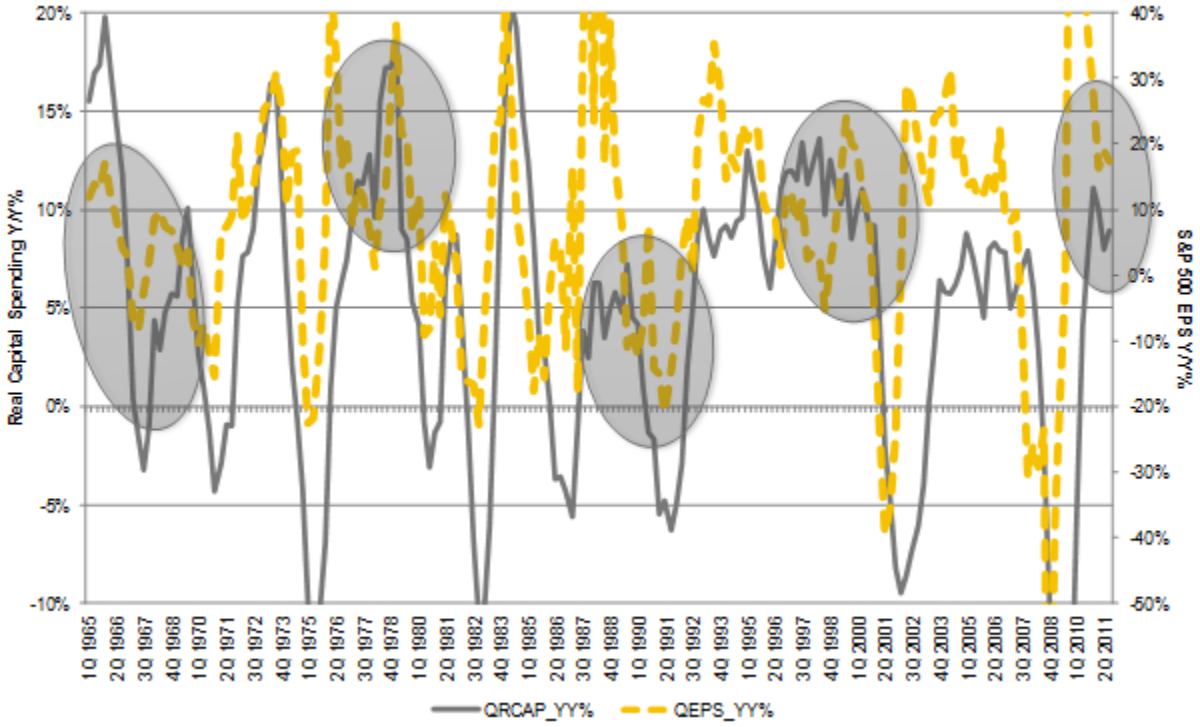


Exhibit 10: Scatter Plot of a) Changes in Real Capital Spending and b) Changes in S&P 500 EPS

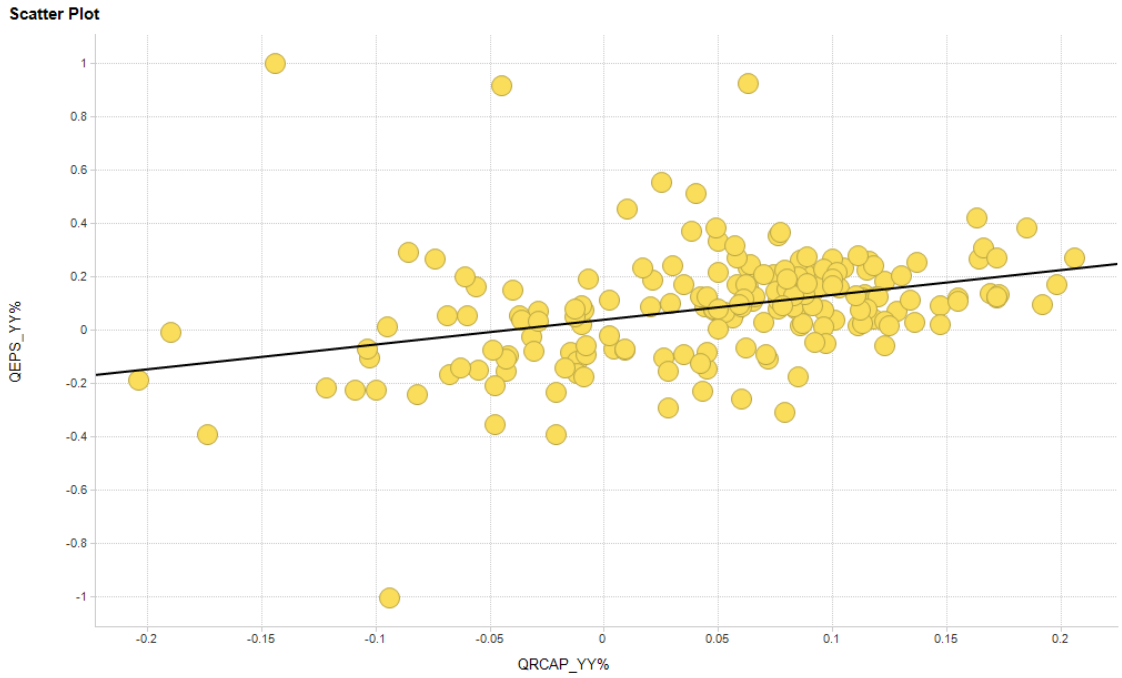


Exhibit 11: Autocorrelation results of EPS_YY%

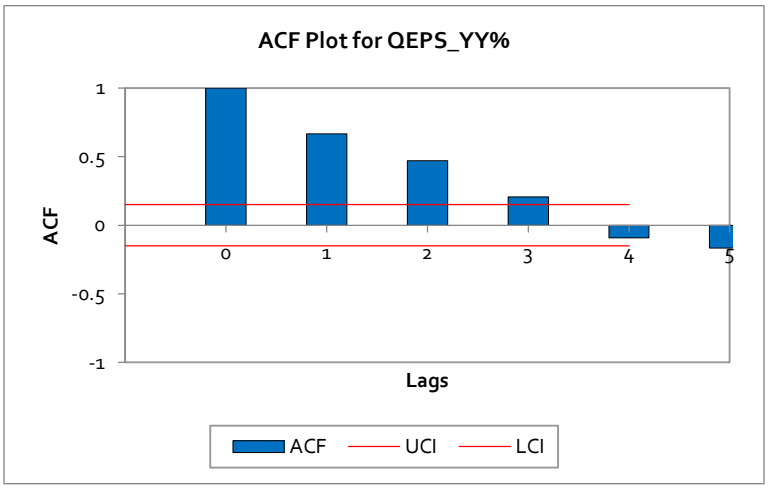


Exhibit 11: Prune Tree without Lag_1 as a predictor

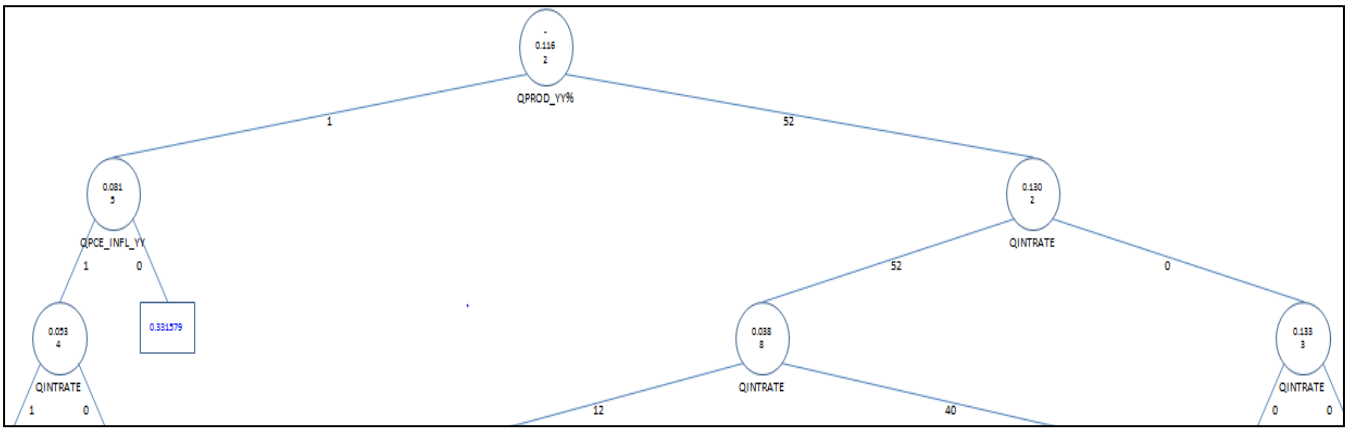


Exhibit 12: Prune Tree with Lag_1 as one of the predictors

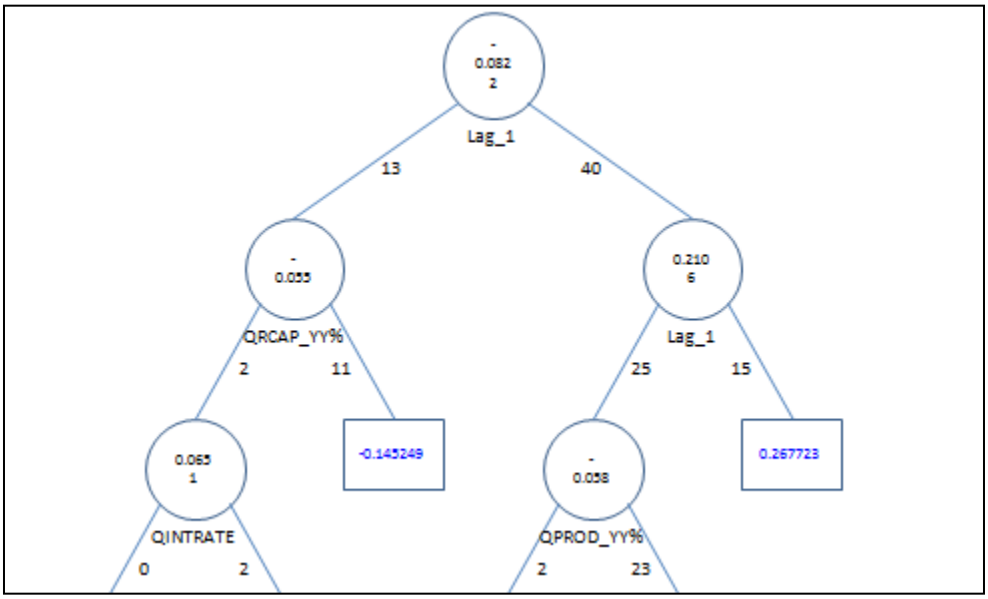


Exhibit 14: Best Subset Results (MLR with Lag_1)

	#Coeffs	RSS	Cp	R-Squared	Adj. R-Squared	Probability	Model (Constant present in all models)							
							1	2	3	4	5	6	7	8
Choose Subset	2	1.31937933	5.74338007	0.493490298	0.48760065	0.15025397	Constant	Lag_1	*	*	*	*	*	*
Choose Subset	3	1.20116031	-0.29780576	0.538874578	0.528024568	0.88702148	Constant	QRCAP_YY%	Lag_1	*	*	*	*	*
Choose Subset	4	1.19836199	1.51185429	0.539948853	0.523518455	0.8237828	Constant	QRAHE_YY%	QRCAP_YY%	Lag_1	*	*	*	*
Choose Subset	5	2.42864203	87.19475555	0.067644366	0.022711565	0	Constant	QINTRATE	E_INFL_YY%	QRAHE_YY%	QRPCE_YY%	*	*	*
Choose Subset	6	2.42304039	88.81373596	0.069794836	0.013075009	0	Constant	QINTRATE	E_INFL_YY%	QRAHE_YY%	QRPCE_YY%	QPROD_YY%	*	*
Choose Subset	7	2.29662395	82.21496582	0.118326105	0.053016928	0	Constant	QINTRATE	E_INFL_YY%	QRAHE_YY%	QRPCE_YY%	QPROD_YY%	QRCAP_YY%	*
Choose Subset	8	1.17613518	8	0.548481725	0.508973876	1	Constant	QINTRATE	E_INFL_YY%	QRAHE_YY%	QRPCE_YY%	QPROD_YY%	QRCAP_YY%	Lag_1

Exhibit 15: Best Subset Results (MLR without Lag)

	#Coeffs	RSS	Cp	R-Squared	Adj. R-Squared	Probability	Model (Constant present in all models)							
							1	2	3	4	5	6	7	8
Choose Subset	2	1.31937933	5.74338007	0.493490298	0.48760065	0.15025397	Constant	Lag_1	*	*	*	*	*	*
Choose Subset	3	1.20116031	-0.29780576	0.538874578	0.528024568	0.88702148	Constant	QRCAP_YY%	Lag_1	*	*	*	*	*
Choose Subset	4	1.19836199	1.51185429	0.539948853	0.523518455	0.8237828	Constant	QRAHE_YY%	QRCAP_YY%	Lag_1	*	*	*	*
Choose Subset	5	2.42864203	87.19475555	0.067644366	0.022711565	0	Constant	QINTRATE	E_INFL_YY%	QRAHE_YY%	QRPCE_YY%	*	*	*
Choose Subset	6	2.42304039	88.81373596	0.069794836	0.013075009	0	Constant	QINTRATE	E_INFL_YY%	QRAHE_YY%	QRPCE_YY%	QPROD_YY%	*	*
Choose Subset	7	2.29662395	82.21496582	0.118326105	0.053016928	0	Constant	QINTRATE	E_INFL_YY%	QRAHE_YY%	QRPCE_YY%	QPROD_YY%	QRCAP_YY%	*
Choose Subset	8	1.17613518	8	0.548481725	0.508973876	1	Constant	QINTRATE	E_INFL_YY%	QRAHE_YY%	QRPCE_YY%	QPROD_YY%	QRCAP_YY%	Lag_1

Exhibit 16: Results of MLR on the final model

Training Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
1.201160342	0.116831286	-4.06454E-09

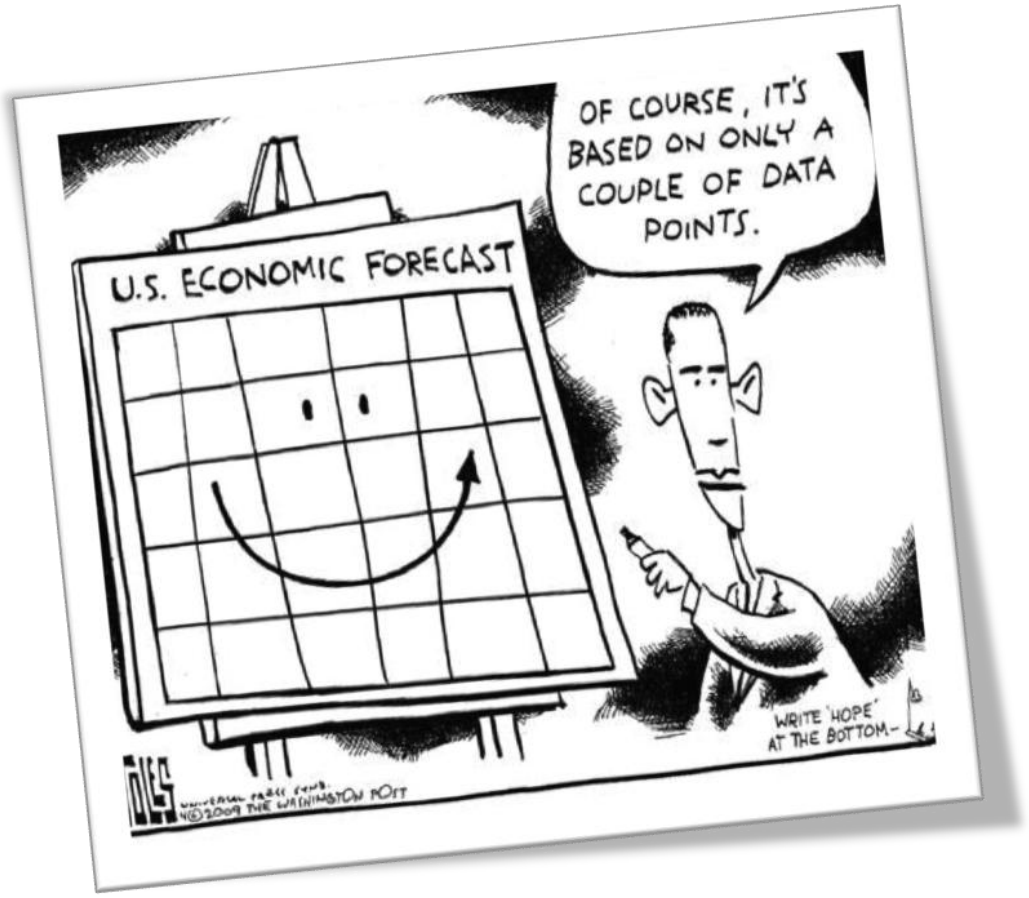
Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
2.059123974	0.197107574	-0.005765275

Test Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
1.294701248	0.189641566	-0.013193876

Fin.



the just-in-case disclosure: the authors of this report are not liable for any financial decisions made as a result of the findings discussed above.