Forecasting Cement Prices for the Construction Industry

Team 4
Amit Tyagi
Harmanjit Singh
Mohini Jain
Nandini Chandrasekhar
Rahul Chakraborty

Executive Summary

This document attempts to forecast the prices of cement from changes in the price of crude oil. The analysis would help construction companies and other involved players manage their inventories effectively and influence their purchasing decisions.

The data for the study, namely oil and cement prices, were collected from syndicated agencies and extend over a 3 year period, from 2009-2011. The methodology included data cleansing and conformity wherein missing values were accounted for by linear extrapolation. The forecasting methods used included Naïve, Holt winter exponential and regression (based on multiple predictors like oil and time function) and were compared on accuracy measured by minimal errors in the forecasted values.

The data for oil was observed to have an increasing pattern over time while the cement prices had a periodic high every April. The key findings validated this higher price (April) which necessitates inventory replenishment during December when cement prices are comparatively lower. Another important finding was the price of cement was a function of the *1 month lag* in prices of oil.

The caveats which need to be highlighted are the limitations in data collection, wherein larger samples could have helped us in validating the model better. Ancillary variables like the seasonal rainfall or economic indicators could also have had an impact in forecasting cement prices.

Problem Description

The cost structure of the construction industry is dominated by the usage of raw materials such as sand, cement, steel etc. The price of raw materials fluctuates and this fluctuation could be driven by one of the external factors – crude oil price. The sudden surge in raw material price adversely impacts the industry's bottom line.

The objective of the project is to factor into the change in crude oil price to predict on a quarterly basis the price change in cement, which is one of the key raw materials used in construction. The forecasting model will serve as a valuable tool to builders/dealers in the construction industry or cement production companies to plan capacity expansion.

The benefit of the model will facilitate timing of purchase, inventory management, and budget allocation. Thus, the outcome of the model would include better resource planning and sourcing leading to increased savings.

Data

Data Collection

Monthly crude oil and cement prices were collected from the website (www.indiastat.com) for the time period January 2009 to May 2011,

Data Issues

The initial data had the following issues:

1. Missing values for cement for the months January to March 2011
2. Oil prices were available in the US dollar.

Issue Resolution Approach:

1. The former issue was solved by imputing missing data using moving average and straight line methods
2. The latter issue concerning with oil price in the US dollar value was converted to Indian Rupees using historic conversion rate corresponding to each data point sourced from www.oanda.com.

Data Exploration

Observing the data for cement over a 3 year period from 2009 to 2011, we can observe the following from the line chart (figure 3):

- The data does not reflect a clear trend.
- The annual data shows a peak around April indicating a clear seasonality
- Represents a stationary series

Observing the data for oil over the corresponding period, we can observe the following(figure 4):

- The data displays an increasing trend
- No significant seasonality since the line chart shows random spurts through the year
- Through 2010, the data displayed a stable pattern
- Represents a non-stationary series

Naïve Forecast

The naïve forecast uses a 12 month lag to predict values. The cement data shows an annual seasonality from the Jan-09 to Jan-11 values. For evaluating the prediction accuracy, we compared the forecasted and the actual values and examined the forecast errors measures, namely:

i) Average Error: -0.3
ii) Mean Absolute Error (MAE): 12.68
iii) Root Mean Square Error (RMSE): 16.04
iv) Mean Absolute Percentage Error (MAPE): -0.43%

Results

We can see that the Naïve method provides a decent model to forecast cement values since the errors are reasonably low.

Exponential Smoothing

Method

Given that our series of interest, cement prices, is characterized by lack of trend and 12 month seasonality, we decided to explore the method of exponential smoothing. In particular, we decided on using the Holt-Winter No Trend technique available in XL-Miner which is applicable for data that displays seasonality but no trend.

Data preparation

As first step, we partitioned the data into training data of 24 months and validation data of 5 months. This was in order to test the level of accuracy attained in using the Holt-Winter No Trend exponential smoothing on our dataset.

We used the following default values for smoothing constants: i) Alpha=0.2 ii) Gamma=0.05

Forecasts

We observed that forecasts made based on this method systematically under-forecast the prices for cement. Having checked the model accuracy on the validation data, we reran the analysis on the entire data to obtain forecasts for the required period Jun 2011 – Aug 2011. The forecast values and the corresponding charts are provided in the exhibit as Fig. Exponential Smoothing.

Comparing to Naïve forecast:

While both the Naïve forecast and the Exponential smoothing forecast systematically under-forecast the cement prices, we can say that exponential smoothing is still better based on the lower RMSE number it produces.

Next Step:

Given the suspected role of crude oil prices, we decided to explore the regression based methods of forecasting which have been described in the following section.

Regression

We conducted a series of regression analyses to explore the significance of crude oil prices in predicting the cement prices. Moreover, in order to account for the seasonality displayed in the data, we used indicator variables through which we explored the option of using both quarterly as well as monthly dummies.

In conducting the regression analysis, we used the adjusted R square of the model as an indication of the goodness of fit of the model for the data and hence its strength of prediction. However, while iterating for different models of regressions, we used the error measures such as RMSE and average error to compare across the models to find the better one.

Regression Model 1: Cement prices, time index and quarter

Dependent variable: cement prices
Independent variable: time index, 3 quarter dummies for Q1, Q2 and Q3 (Q4 as base)
The adjusted R square for the model is 0.3345
Validation data scoring RMSE = 24.60 and Average Error = 19.02

Regression Model 2: Cement prices, time index and months

Dependent variable: cement prices
Independent variable: time index, 11 monthly dummies with December month as base
The adjusted R square for the model is 0.5364
Validation data scoring RMSE = 14.58 and Average Error = 11.31

Regression Model 3: Cement prices, change in crude oil and months

Dependent variable: cement prices
Independent variable: change in crude oil prices, 11 monthly dummies with December month as base
The adjusted R square for the model is 0.5964
Validation data scoring RMSE = 15.16 and Average Error = 8.93

Results:

The forecast for one step June 2011 cement prices is

- Point Estimate: 244.67
- 95% confidence interval is (230.50,258.9)

Insights and Recommendations

Key Insights:
The key insights based on the analysis conducted are outlined below:

- April and May have at least 10% higher cement prices than December
- Cement Prices are affected by previous months change in Oil prices and not oil prices themselves
- Crude prices have a positive trend overall last two years.

Recommendations:

- Inventory should be brought in December for next fiscal year rather than April/May
- Track *changes* in crude prices on monthly basis to deduce future cement prices

Limitations

The key limitations of the analysis are given below:

- The analysis has been based on a limited number of data points available. To verify the outcome of the analysis, a more robust analysis with a longer time frame should be ideally used. This is difficult given cement is not publicity traded as a commodity in India
- The model can also include an economic variable such as rain forecasts to check if there is a correlation between cement prices and rain in the region which might stall construction activity.

Appendix

Figure 1: Initial Data of Crude Oil and Cement Prices

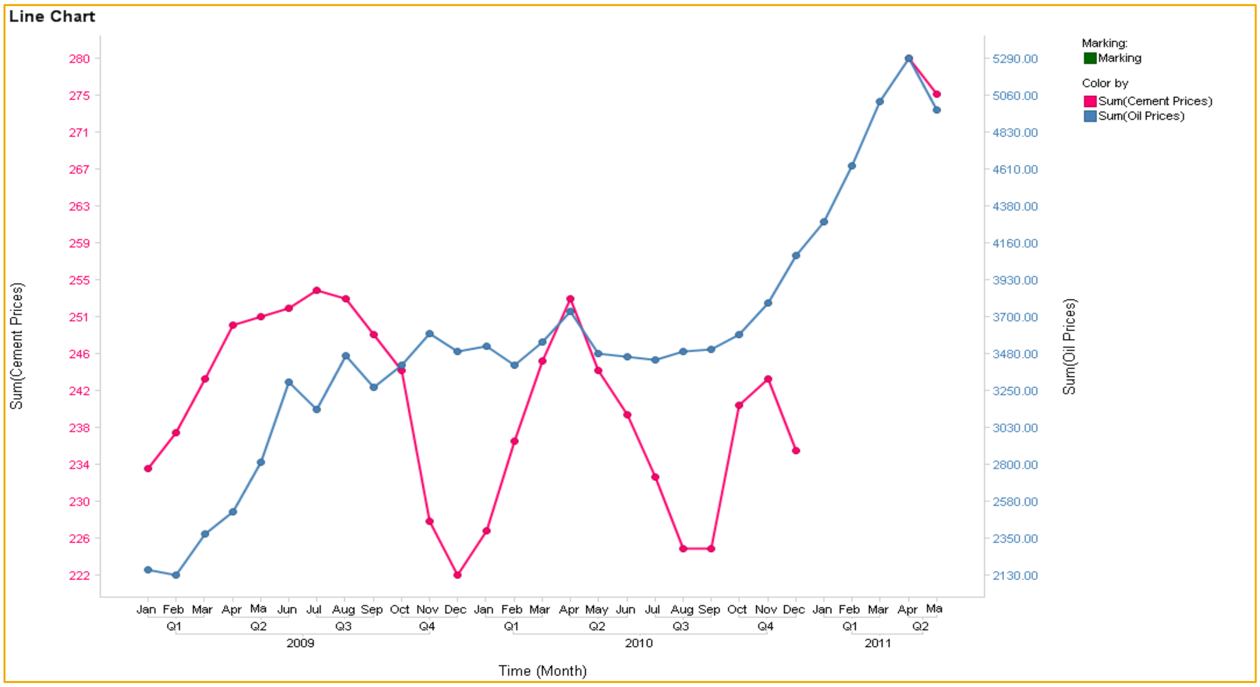


Figure 2: Cement and imputed Crude Oil prices as indicated by the box

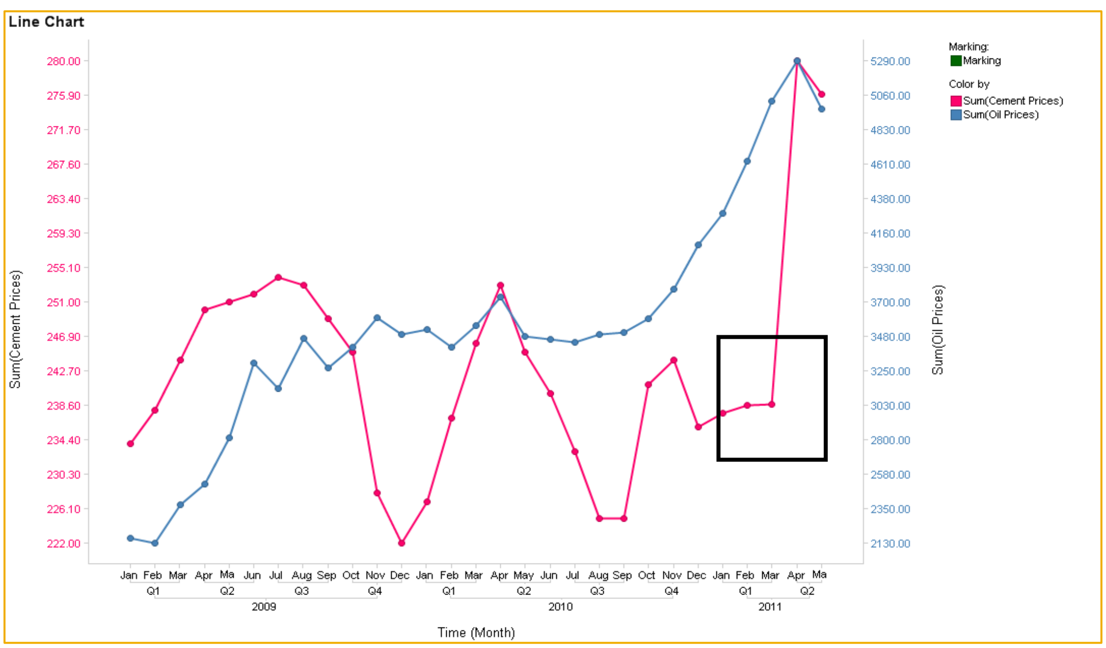


Figure 3: Time Series of Cement Prices

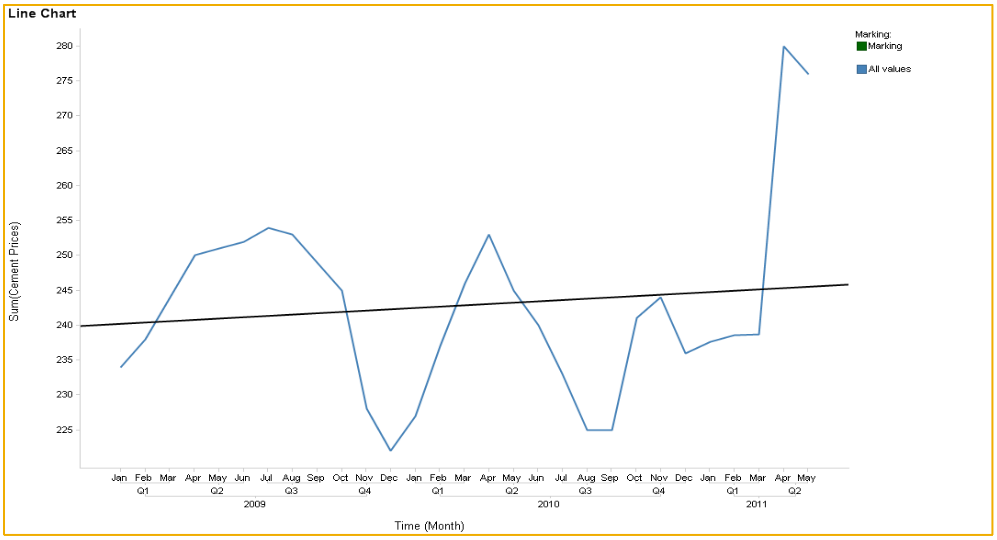


Figure 4: Time Series of Oil Prices

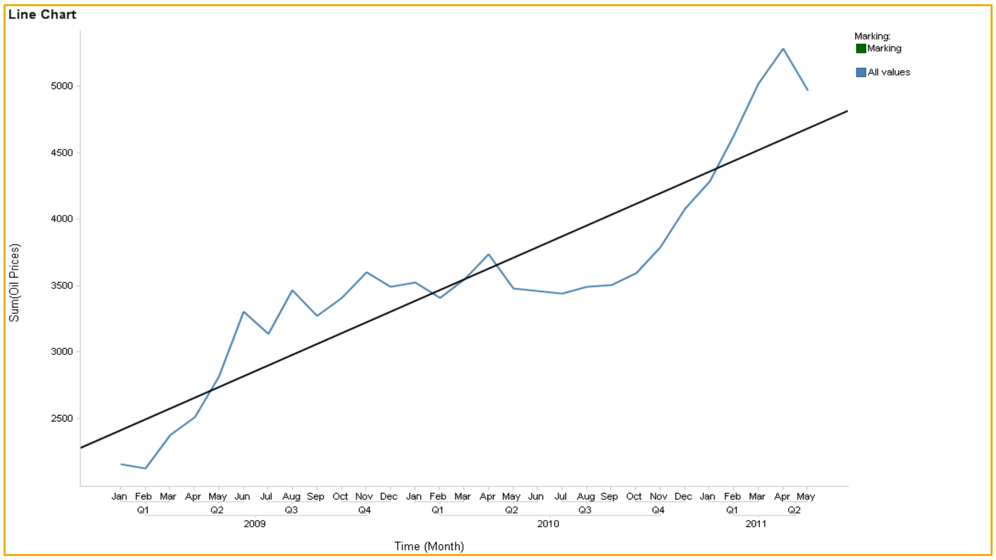
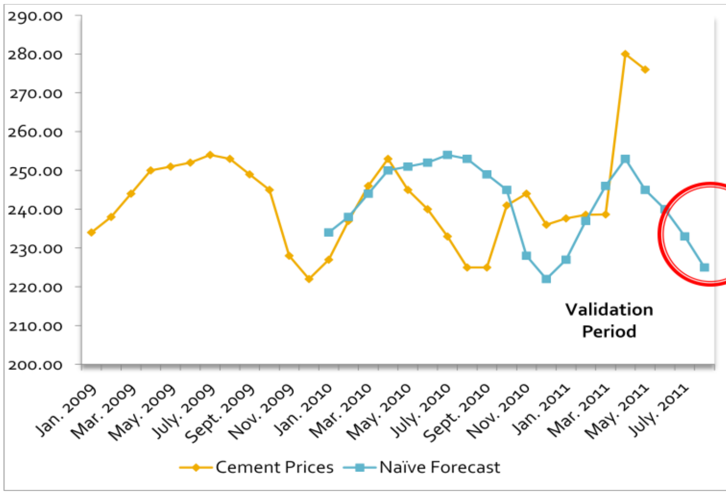


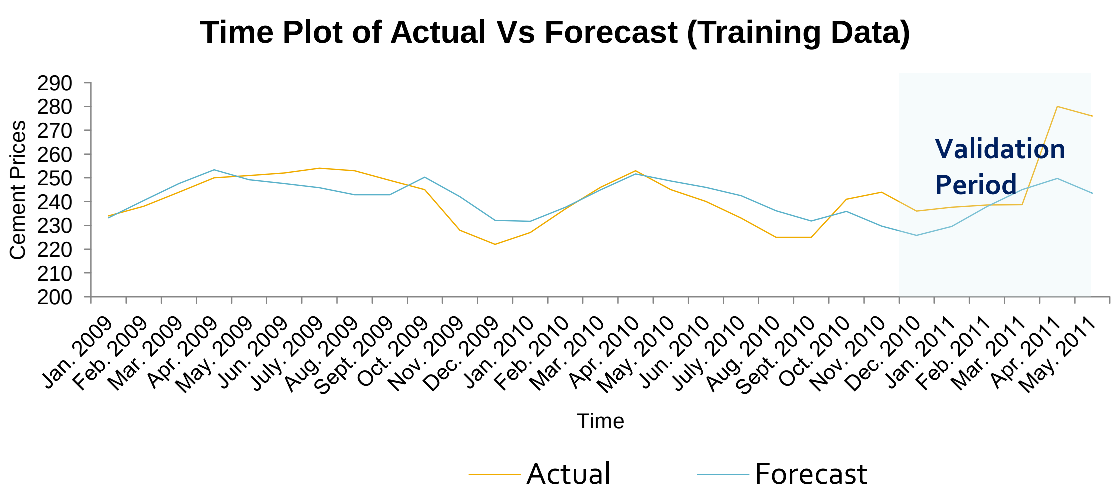
Figure 5: Naïve Forecasts



Time	Forecasted Cement Prices
June. 2011	240.00
July. 2011	233.00
Aug. 2011	225.00

Parameter	Value
Average Error	-0.30
MAE	12.68
RMSE	16.04
MAPE	-0.43%

Figure 6: Exponential Smoothing

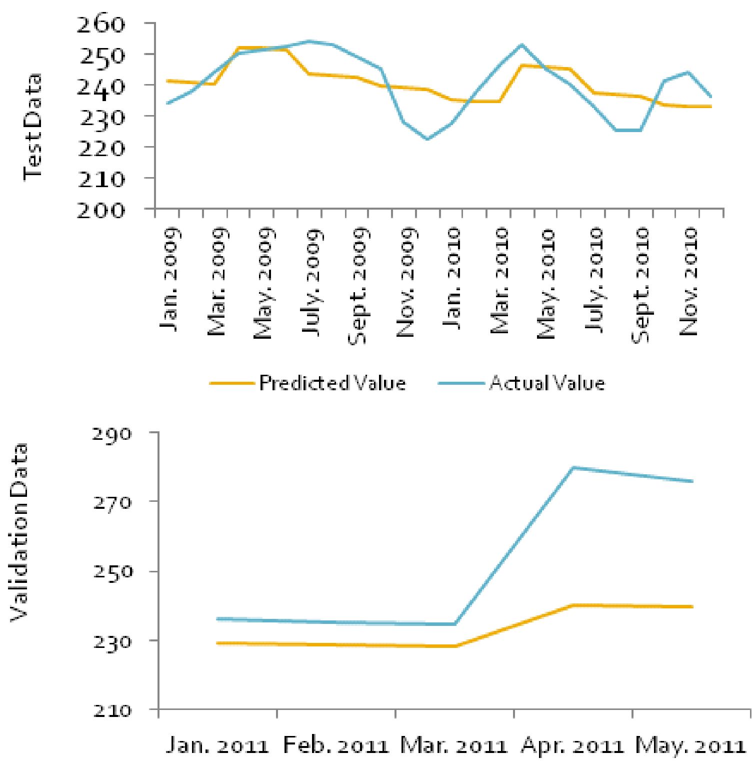


MAPE	-0.51%
MAE	19.24
RMSE	10.78

Time	Forecast	LCI	UCI
Jun. 2011	240.10	219.09	261.11
July. 2011	237.66	216.65	258.67
Aug. 2011	233.27	212.26	254.27

Time	Actual	Forecast	Error	LCI	UCI
Jan. 2011	237.67	240.10	-2.43	219.09	261.11
Feb. 2011	238.56	237.66	0.90	216.65	258.67
Mar. 2011	238.69	233.27	5.42	212.26	254.27
Apr. 2011	280.00	231.33	48.67	210.33	252.34
May. 2011	276.00	237.24	38.76	216.23	258.24

Figure 7: Times Series Regression (Model 1)



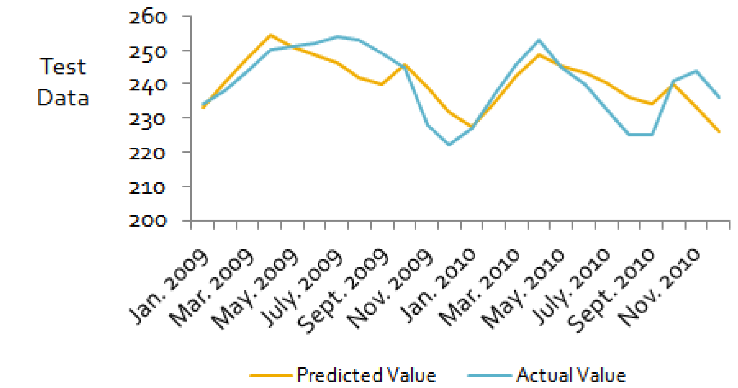
Training Data scoring - Summary Report

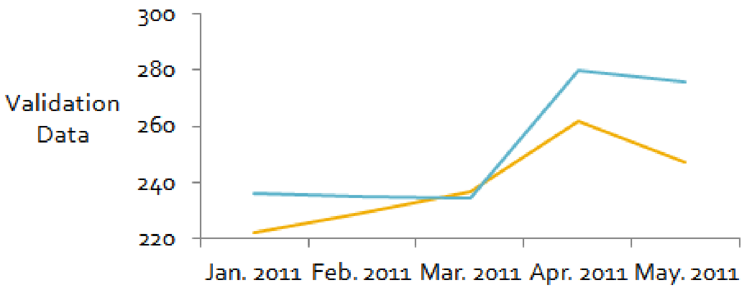
Total sum of squared errors	RMS Error	Average Error
1526.574621	7.975416972	-4.6425E-06

Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
3025.38032	24.59829392	19.02459087

Figure 8: Times Series Regression (Model 2)





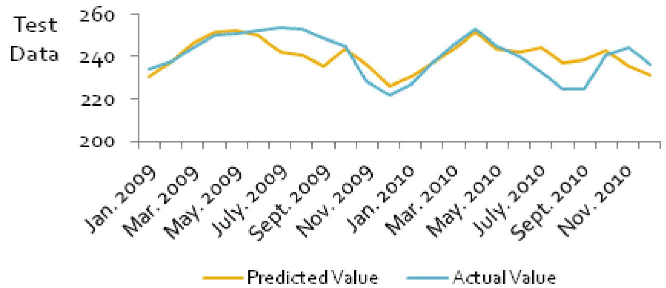
Training Data scoring - Summary Report

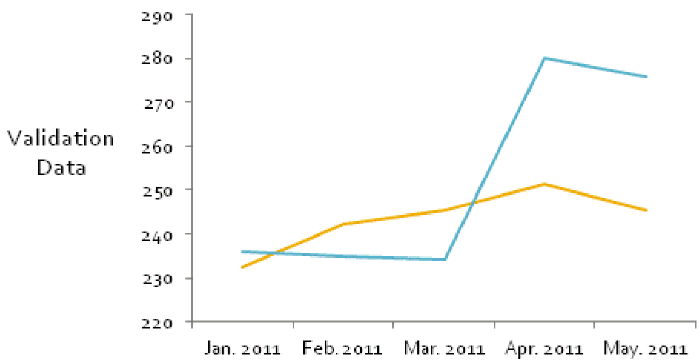
Total sum of squared errors	RMS Error	Average Error
1063.333333	6.656241849	7.25833E-07

Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
1062.446432	14.57701226	11.30664186

Figure 8: Times Series Regression (Model 3)



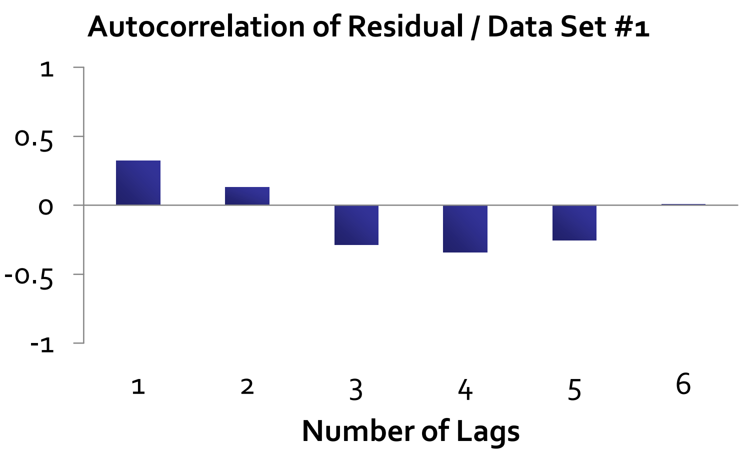


Training Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
1153.23117	6.931904409	7.05229E-06

Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
1149.759154	15.16416271	8.935974365



The Regression Model

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	227.9784088	7.31329346	0	1388166
Change Oil Prices	47.3692855 8	47.84405518	0.08434155	236.9483643
Jan	2.30851197	10.27161407	0.82629663	170.8402252
Feb	10.6656694 4	10.47013855	0.3302393	4.03690577
Mar	13.4651508 3	10.55434036	0.22831073	3.62213135
Apr	20.9184341 4	10.36295986	0.06858356	181.3603821
May	19.0390167 2	10.2391777	0.08990059	146.835495
Jun	14.3844404 2	10.57441139	0.20095661	62.94096756
Jul	16.8527965 5	10.51124954	0.13716948	127.4580536
Aug	8.34284115	10.37500381	0.43835571	2.26122379
Sep	10.2819099 4	10.49529457	0.34830493	22.82782364
Oct	13.4179601 7	10.25596333	0.21744627	151.9502106
Nov	5.50732994	10.3494997	0.60521376	29.6870594

Residual df	11
Multiple R-squared	0.597283706
Std. Dev. estimate	10.23910141
Residual SS	1153.231201

ANOVA

Source	df	SS	MS	F-statistic	p-value
Regression	12	1140.768842	95.06407019	0.906760735	0.09678779 3
Error	11	1153.231201	104.8392001		
Total	23	2294.000043			