University of Illinois at Urbana-Champaign
Econ 490 Economic Forecasting
Final Term Paper

Forecasting Price of Gasoline by Time Series Models

Jian Zhang

Master of Science in Policy Economics
December 2015

Introduction

Normally, I will drive from Champaign to St. Louis every single week, which makes me pay attention to gas price of gas station to find cheap gas. So, I want to use what I have learnt from class to estimate and forecast local gas price. In beginning of research, I was trying to collect historical data of gas price in Champaign-Urban, and I did, however, the data source asked for 0.6 dollar for daily data and 1.3 dollar for monthly data. Thus, I change my target location to Chicago and collected weekly data of regular gasoline retail price in Chicago area ranging from 12/01/2008 to 11/30/2015. I have used ARIMA, ARMA with seasonal dummy variables and VAR model to estimate future gasoline price. Compared among three models, the last two works relatively well with strong explanatory power, in the meantime, VAR model is more simple than ARMA with seasonal dummy variables, which makes sense because I have add Crude Oil Price as regressor to forecast gasoline price.

Data

First, plot the data of weekly Chicago regular all formulations retail gasoline prices in Figure 1.

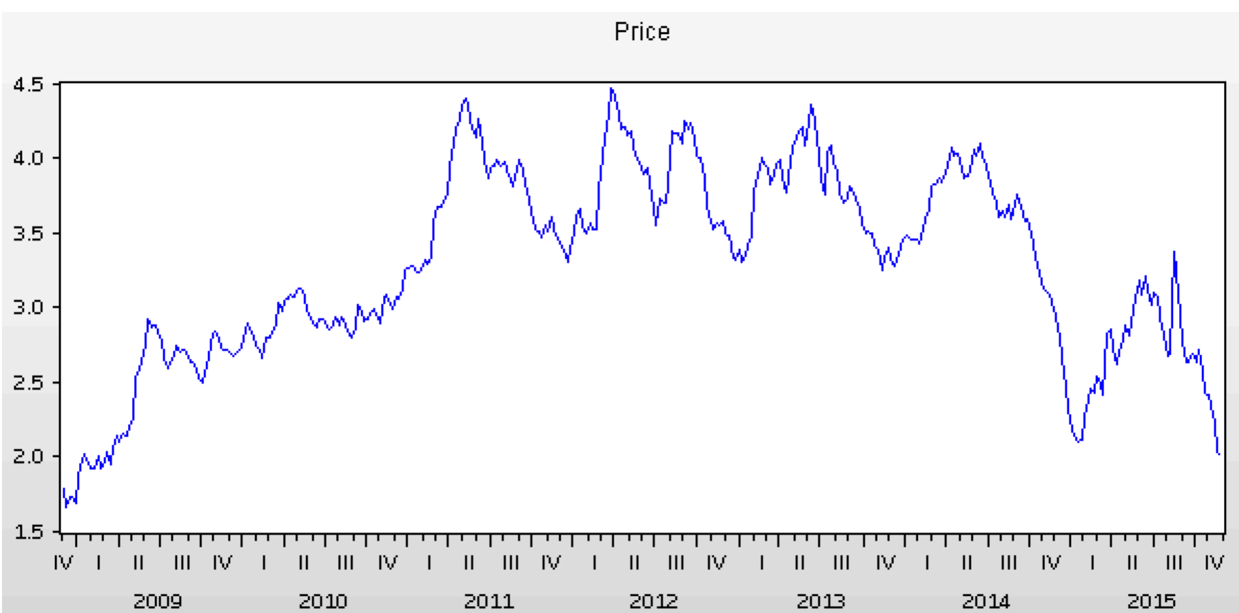


Figure 1

The graph shows that regular gasoline price increased during 2009 and remained stable in 2010. Seasonality can be observed during 2011 to 2014. In the end of 2014, gasoline price experienced huge fall due to crude oil price plummeted. It is apparent that price is relatively high in summer and relatively low in winter. In this case, it is reasonable to add seasonal dummy variables to capture property of this time series data.

Unit root test

Before plug in models, let's first take the unit root test by Eviews, we find that test favor the null hypothesis that this data have a unit root shown in Figure 2.

Null Hypothesis: PRICE has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic - based on SIC, maxlag=16)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.383896	0.3873
Test critical values:	1% level	-3.984420	
	5% level	-3.422679	
	10% level	-3.134227	

*MacKinnon (1996) one-sided p-values.

Figure 2

Also, Eviews automatically computes following regression based on Augmented Dickey-Fuller test (Figure 3), all the coefficients are significant except trend. So I omit trend term to form new regression. (Figure 4)

Then, plot correlogram of residuals, and find out that accepting null hypothesis that residuals are white noise shown in Figure 5. However, declining possibility indicates weak evidence of potential serial correlation within more lags. This is a signal that MA or AR terms might be needed to improve model.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(PRICE)
Method: Least Squares
Date: 12/03/15 Time: 20:51
Sample (adjusted): 12/15/2008 8/31/2015
Included observations: 351 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PRICE(-1)	-0.021520	0.009027	-2.383896	0.0177
D(PRICE(-1))	0.232050	0.052445	4.424671	0.0000
C	0.078901	0.028033	2.814540	0.0052
@TREND("12/01/2008")	-2.87E-05	5.64E-05	-0.507621	0.6120

R-squared	0.073719	Mean dependent var	0.003684
Adjusted R-squared	0.065710	S.D. dependent var	0.101867
S.E. of regression	0.098463	Akaike info criterion	-1.786935
Sum squared resid	3.364172	Schwarz criterion	-1.742938
Log likelihood	317.6071	Hannan-Quinn criter.	-1.769424
F-statistic	9.205403	Durbin-Watson stat	1.936330
Prob(F-statistic)	0.000007		

Figure 3

Dependent Variable: D(PRICE)
Method: Least Squares
Date: 12/03/15 Time: 21:06
Sample (adjusted): 12/15/2008 8/31/2015
Included observations: 351 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.079689	0.027960	2.850072	0.0046
PRICE(-1)	-0.023296	0.008312	-2.802577	0.0054
D(PRICE(-1))	0.234604	0.052147	4.498901	0.0000

R-squared	0.073031	Mean dependent var	0.003684
Adjusted R-squared	0.067703	S.D. dependent var	0.101867
S.E. of regression	0.098358	Akaike info criterion	-1.791891
Sum squared resid	3.366670	Schwarz criterion	-1.758893
Log likelihood	317.4769	Hannan-Quinn criter.	-1.778758
F-statistic	13.70851	Durbin-Watson stat	1.936072
Prob(F-statistic)	0.000002		

Figure 4

Date: 12/03/15 Time: 21:10
Sample: 12/01/2008 8/31/2015
Included observations: 351

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.021	0.021	0.1532	0.695
		2	-0.077	-0.078	2.2631	0.323
		3	-0.019	-0.015	2.3869	0.496
		4	0.094	0.089	5.5371	0.236
		5	-0.013	-0.020	5.6022	0.347
		6	0.099	0.114	9.0883	0.169
		7	-0.041	-0.047	9.7012	0.206
		8	-0.035	-0.026	10.138	0.255
		9	0.067	0.071	11.766	0.227
		10	-0.057	-0.092	12.950	0.226
		11	0.041	0.071	13.553	0.259
		12	0.100	0.083	17.179	0.143

Figure 5

Next, using automatic ARIMA forecasting function by Eviews, I add “Price(-1)” and “D(Price(-1))” as exogenous variables and set model selection criteria as SIC. Eviews selects ARIMA(0,1,0), which implies the model above is the best.(Figure 6)

```
Automatic ARIMA Forecasting
Selected dependent variable: D(PRICE)
Date: 12/03/15   Time: 21:41
Sample: 12/01/2008 8/31/2015
Included observations: 351
Forecast length: 13
================================================
Number of estimated ARMA models: 100
Number of non-converged estimations: 0
Selected ARMA model: (0,0)(0,0)
SIC value: -1.73226015066
================================================
```

Figure 6

Forecasting

Based on the model $D(price) = c + \beta_1 price(\text{-}1) + \beta_2 D(price(\text{-}1)) + \epsilon$,
plot forecasting graph from 9/7/2015 to 11/30/2015 along with actual data. The graph seems like the model doesn’t capture the downward trend pretty well. (Figure 7)

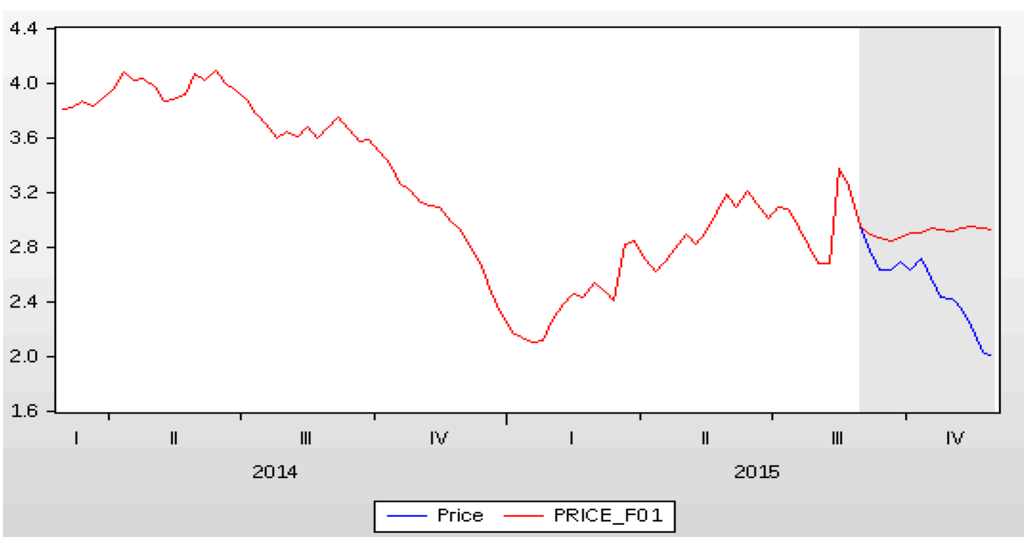
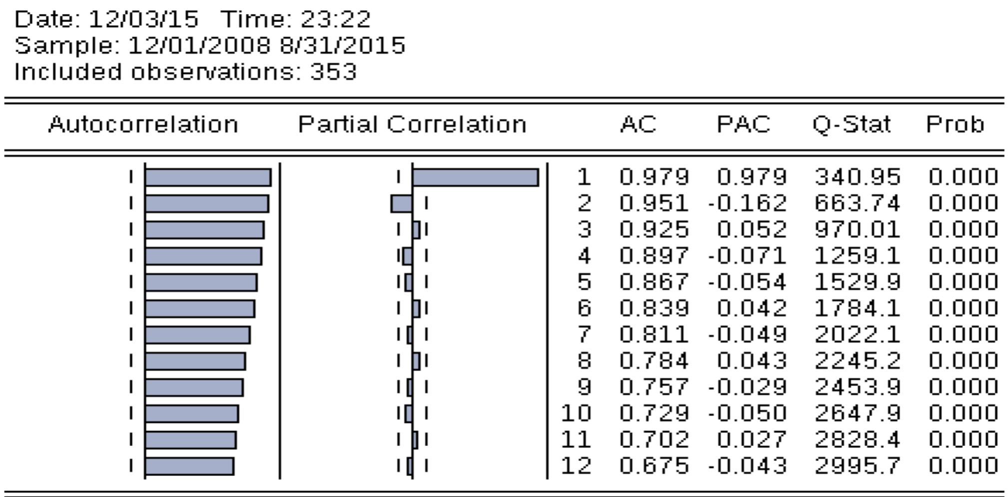


Figure 7

Revised model

Because of limited forecasting power of model I used above, I decided to try ARMA model to fit data. Correlogram in Figure 8 shows existence of serial correlation in the residuals, partial correlation approximately drops to zero after first lag, which is a signal that MA(1) or MA(2) terms might be fit, and latter SIC has favored this point.



Date: 12/03/15 Time: 23:22
Sample: 12/01/2008 8/31/2015
Included observations: 353

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.979	0.979	340.95	0.000
		2	0.951	-0.162	663.74	0.000
		3	0.925	0.052	970.01	0.000
		4	0.897	-0.071	1259.1	0.000
		5	0.867	-0.054	1529.9	0.000
		6	0.839	0.042	1784.1	0.000
		7	0.811	-0.049	2022.1	0.000
		8	0.784	0.043	2245.2	0.000
		9	0.757	-0.029	2453.9	0.000
		10	0.729	-0.050	2647.9	0.000
		11	0.702	0.027	2828.4	0.000
		12	0.675	-0.043	2995.7	0.000

Figure 8

Next, I use SIC table to choose ARMA model in Table 1, SIC chooses ARMA(1,1) model. In the beginning, seasonality can be observed in the data, so I add 52 seasonal dummy variables with ARMA(1,1) to fit data. Because I have used weekly data, not monthly data, 52 dummy variables represent 52 weeks of a single year.

SIC	MA				
AR		0	1	2	3
	0	1.9479	-0.7787	-0.1499	-0.5044
	1	-1.6735	-1.7192	-1.7026	-1.6883
	2	-1.7141	-1.7026	-1.6925	-1.6862
	3	-1.7031	-1.6867	-1.7059	-1.6948

Table 1

Residual plot of ARMA with seasonal dummy variables looks good, indicating that this model has strong explanatory power (Figure 9). Correlogram shows that residuals are white noise, although declining possibility has been noticed. (Figure 10)

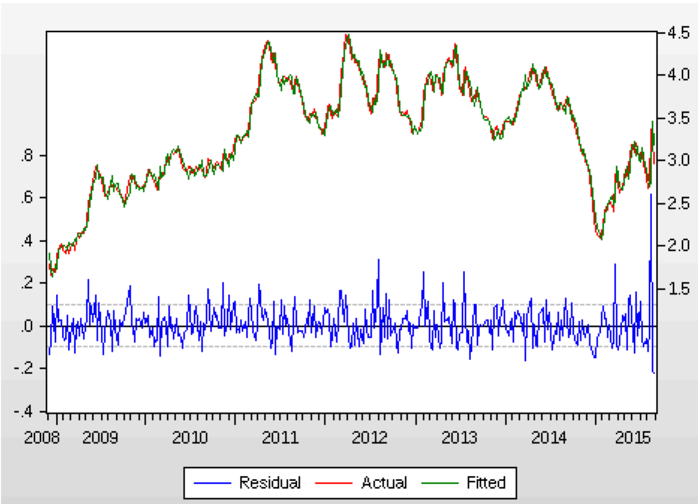


Figure 9

Date: 12/04/15 Time: 00:27
Sample: 12/01/2008 8/31/2015
Included observations: 353

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.007	-0.007	0.0195	0.889
		2	-0.048	-0.048	0.8296	0.660
		3	-0.064	-0.065	2.3081	0.511
		4	0.056	0.052	3.4180	0.490
		5	-0.101	-0.108	7.1147	0.212
		6	0.072	0.074	8.9937	0.174
		7	-0.068	-0.073	10.647	0.155
		8	-0.030	-0.039	10.980	0.203
		9	0.068	0.085	12.686	0.177
		10	-0.105	-0.144	16.697	0.081

Figure 10

Forecast revised model

Based on model $Price = c + \beta_1 AR(1) + \beta_2 MA(1) + \sum_{i=1}^{52} \gamma_i D_{it} + \epsilon$, I made the forecast graph from 9/7/2015 to 11/30/2015 along with actual data in Figure 11. This model seems to have captured the movement of gasoline price.

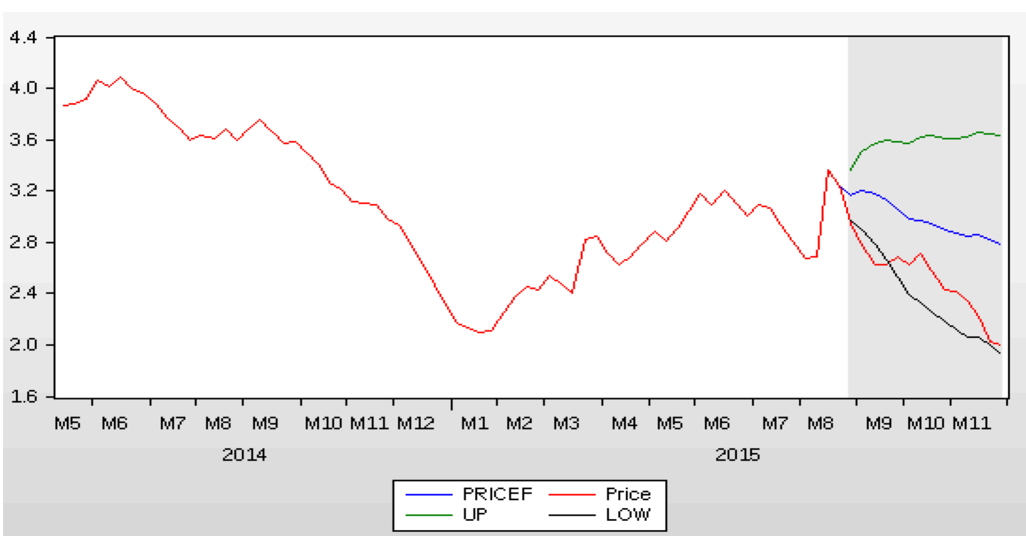


Figure 11

Alternative model

The revised model works relatively well in forecasting trend of gasoline price, however, forecasting error in December 2015 is still observable, so I decide to add Crude Oil Prices as extra regressor to improve current model, because movement of Crude Oil Prices directly cause gasoline price change, which causality test has confirm this point. (Figure 13)

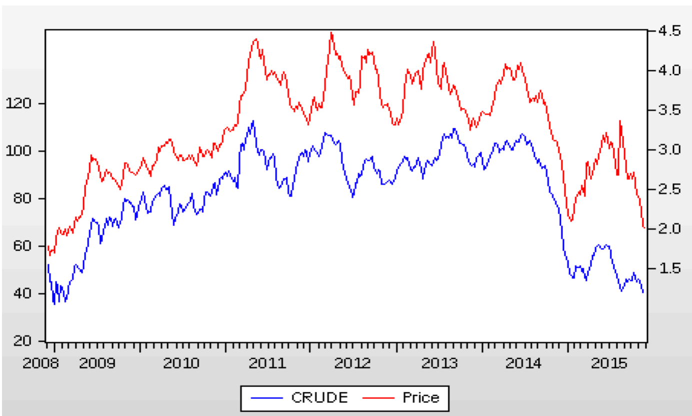


Figure 12

Pairwise Granger Causality Tests
Date: 12/04/15 Time: 01:15
Sample: 12/01/2008 8/31/2015
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
PRICE does not Granger Cause CRUDE	352	0.30329	0.5822
CRUDE does not Granger Cause PRICE		7.35081	0.0070

Figure 13

Cross-correlation graph in Figure 14 indicates that crude oil price is maximally correlated with gasoline price in the first 3 lags, which is about 3 weeks in real world.

Date: 12/04/15 Time: 01:20
Sample: 12/01/2008 8/31/2015
Included observations: 353
Correlations are asymptotically consistent approximations

CRUDE,PRICE(-i)	CRUDE,PRICE(+i)	i	lag	lead
		0	0.8664	0.8664
		1	0.8537	0.8566
		2	0.8382	0.8396
		3	0.8231	0.8204
		4	0.7986	0.8004
		5	0.7734	0.7796
		6	0.7536	0.7588
		7	0.7328	0.7367
		8	0.7109	0.7157
		9	0.6906	0.6938
		10	0.6697	0.6708

Figure 14

Then, estimate VAR model with 3 lags by Eviews, R^2 shows good fitness (Figure 15) and residual plot looks fine (Figure 16).

R-squared	0.976679	0.979733
Adj. R-squared	0.976271	0.979378
Sum sq. resids	3.148445	2495.142
S.E. equation	0.095808	2.697122
F-statistic	2394.099	2763.465
Log likelihood	327.8008	-840.3579
Akaike AIC	-1.833147	4.842045
Schwarz SC	-1.755989	4.919204
Mean dependent	3.312737	83.05966
S.D. dependent	0.621955	18.78178

Figure 15

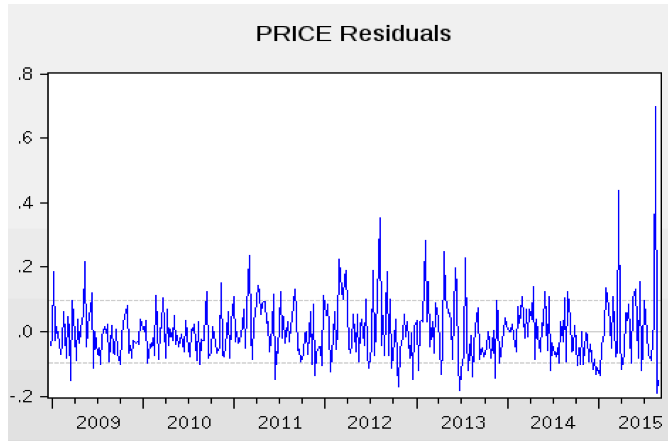


Figure 16

Forecasting VAR with actual data, this model works as good as our former model shown in Figure 17. However, SIC of ARMA model is -1.060182 and SIC of VAR model is -1.755989. We would like to choose VAR model because of its simplicity with accuracy, which makes sense since crude oil price is the best proxy for retail gasoline price.

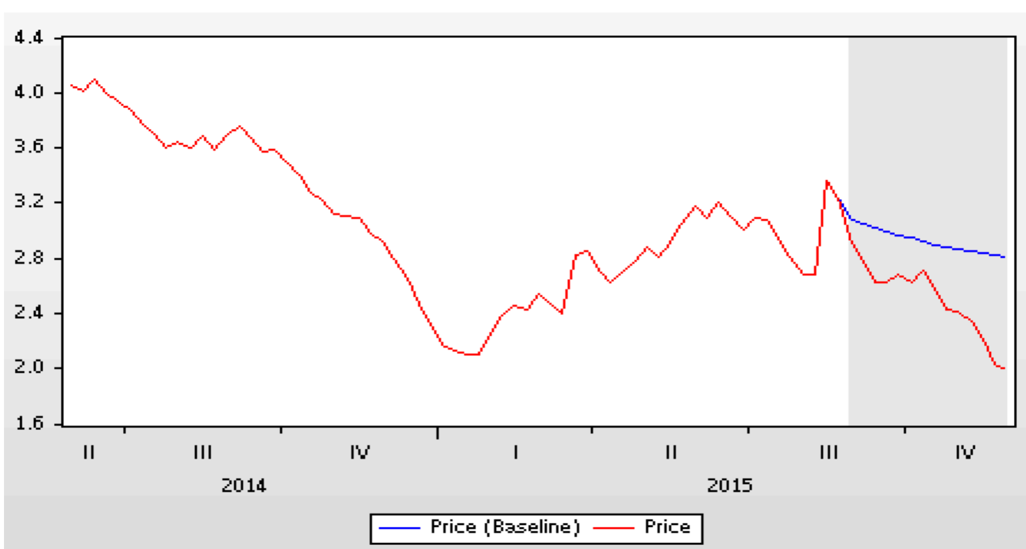


Figure 17

Stability test of model

Since stability test cannot be used to estimate MA models, we cannot perform stability test to our second model consisted of seasonal dummy variables with ARMA model. Here is recursive parameter estimate and CUSUM plot of VAR model shown in Figure 18 and 19.

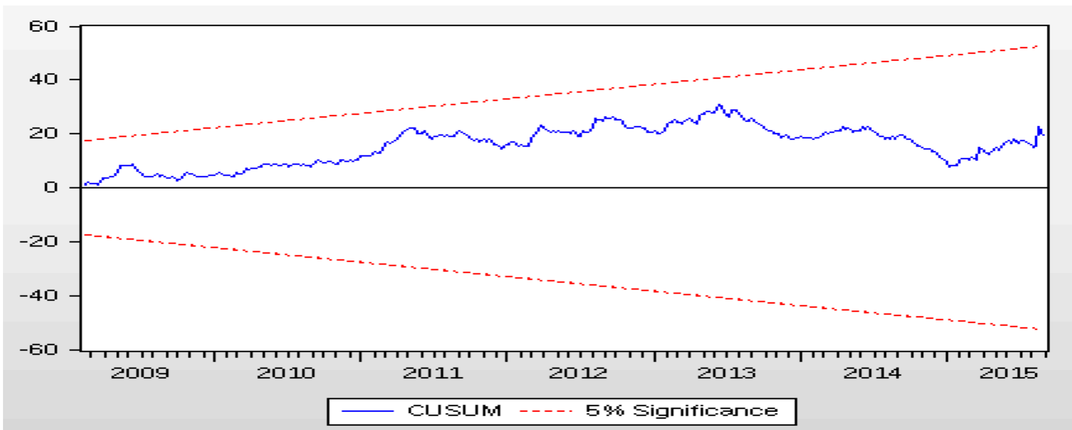


Figure 18

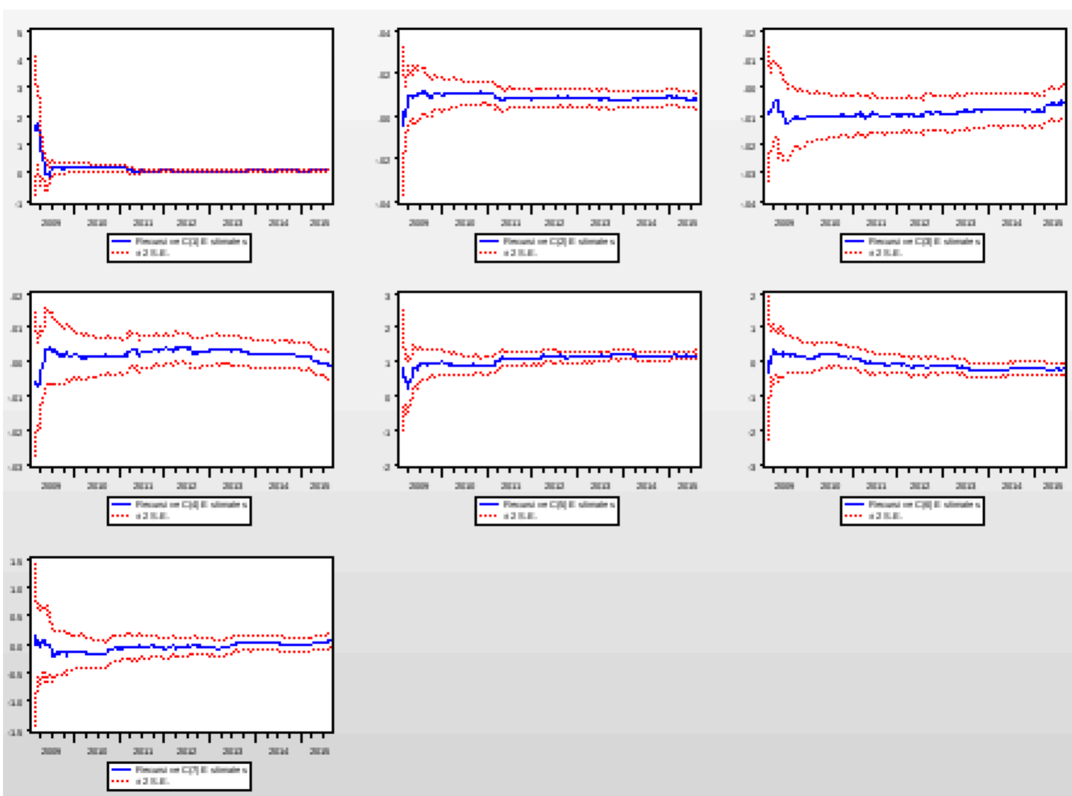
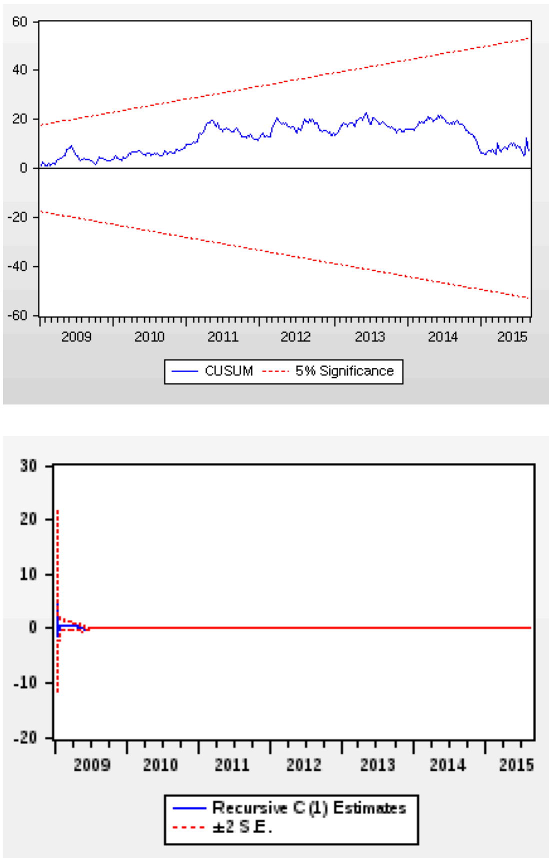


Figure 19

Both plots show no evidence of parameter instability. Also, we have test the first ARIMA model shown in Figure 20 and 21, which also indicates that parameters are stable in this model, however, this ARIMA model doesn't have strong forecasting power.



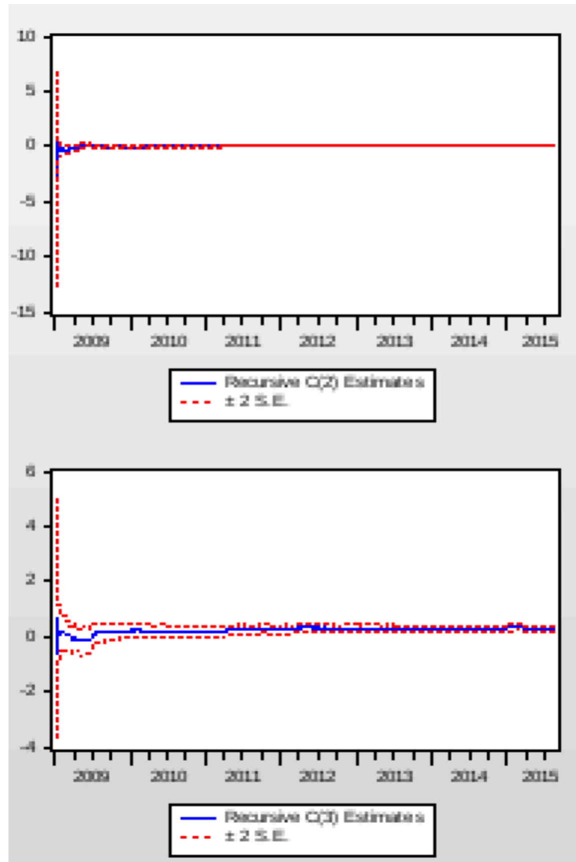


Figure 20

Figure 21

Conclusion

ARIMA model has the ability to capture basic direction of gasoline price movement, but low R^2 and forecast plot indicates limited explanatory power with less capability to forecast future gasoline price. ARMA model with seasonal dummy variables capture the main trend of gasoline price, it has strong explanatory power and forecasting ability, the drawback of this model is that it consists of too many independent variables, seasonal dummy variables based on 12 months or 4 seasons should be enough to capture seasonality. The last model basically uses Crude Oil Price as proxy to indicate gasoline price, but it is not possible for every data to find a perfect proxy like this.

Reference

Diebold, F. (2006). Elements of forecasting (4th ed.). Mason, Ohio: Thomson/South-Western.

US. Energy Information Administration, Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma [MCOILWTICO], retrieved from FRED, Federal Reserve Bank of St. Louis. https://research.stlouisfed.org/fred2/series/MCOILWTICO/, December 5, 2015.

US. Energy Information Administration ,Weekly Chicago Regular All Formulations Retail Gasoline Prices, retrieved from, https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EMM_EPMR_PTE_YORD_DPG&f=W. November 30, 2015.