STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

SYS 670: Forecasting and Demand Modeling

Final Project on
OPEC crude oil basket price data
Fall 2017

Professor
Dr. Zigh Teresa

Submitted by
Anshul Verma

TABLE OF CONTENTS:

1. INTRODUCTION:

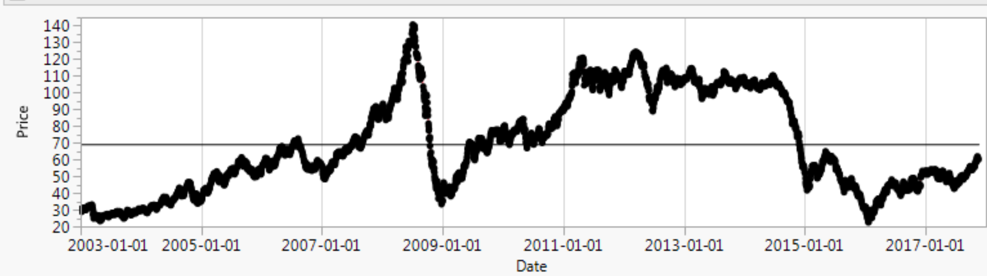
In this project, time series forecasting techniques are implemented to predict OPEC crude oil basket price. The dataset is obtained from OPEC official website and can be downloaded in four different forms which are original price, change in price, percentage of change in price and the cumulative price. Each of the four datasets is analyzed separately and a comparison of the four is done with regards to accuracy in prediction.

2. PRICE DATASET:

2.1 Basic Diagnostics

Plotting Time Series for the data:

Time Series Price



Mean	69.344087
Std	28.730419
N	3841
Zero Mean ADF	-0.15813
Single Mean ADF	-1.522475
Trend ADF	-1.200981

Time Series Basic Diagnostics

Lag	AutoCorr	Ljung-Box Q	p-Value
0	1.0000	.	.
1	0.9991	3836.88	<.0001*
2	0.9978	7665.08	<.0001*
3	0.9965	11484.5	<.0001*
4	0.9952	15295.0	<.0001*
5	0.9939	19096.2	<.0001*
6	0.9925	22887.9	<.0001*
7	0.9911	26670.0	<.0001*
8	0.9897	30442.3	<.0001*
9	0.9883	34204.7	<.0001*
10	0.9869	37957.2	<.0001*
11	0.9854	41699.4	<.0001*
12	0.9839	45431.4	<.0001*
13	0.9824	49152.7	<.0001*
14	0.9808	52863.2	<.0001*
15	0.9792	56562.5	<.0001*
16	0.9775	60250.3	<.0001*
17	0.9759	63926.2	<.0001*
18	0.9742	67590.3	<.0001*
19	0.9725	71242.9	<.0001*
20	0.9708	74883.6	<.0001*
21	0.9690	78512.1	<.0001*
22	0.9673	82128.3	<.0001*
23	0.9655	85732.5	<.0001*
24	0.9638	89324.4	<.0001*
25	0.9620	92903.8	<.0001*

Lag	Partial
0	1.0000
1	0.9991
2	-0.1821
3	0.0276
4	-0.0167
5	-0.0242
6	-0.0093
7	0.0039
8	-0.0267
9	0.0131
10	-0.0179
11	-0.0147
12	0.0055
13	-0.0313
14	-0.0067
15	-0.0195
16	-0.0242
17	-0.0081
18	0.0075
19	0.0100
20	-0.0224
21	-0.0193
22	0.0074
23	0.0010
24	-0.0065
25	-0.0175

Lag	Variogram
1	1.0000
2	2.3631
3	3.7375
4	5.1348
5	6.5873
6	8.0748
7	9.5619
8	11.0977
9	12.6270
10	14.1832
11	15.7785
12	17.3752
13	19.0288
14	20.7161
15	22.4502
16	24.2440
17	26.0740
18	27.9000
19	29.7020
20	31.5450
21	33.4415
22	35.3391
23	37.2320
24	39.1375
25	41.0855

Figure 2.1 Time Series for Price

Time Series Analysis:

- It displays graphs of the autocorrelations and partial auto correlations of the series. These indicate how and to what degree each point in the series is correlated with earlier values in the series.
- It helps to understand the underlying structure and function that produce the observations. It is assumed that a time series data set has at least one symmetric pattern. The most common patterns are trends and seasonality. There is no specific or impressive trend is followed in the proposed time series model for the given data. This emphasis that the given data is not standardized and it needs some smoothening procedure to standardize.
- Data is stationary and it showed mixed behavior.
- Visual inspection shows there is a steep rise in price of $140 is observed at the time period of 2008-05-01 and sudden fall in price of $35 in the year 2009. From the year 2011 to 2015 the oil

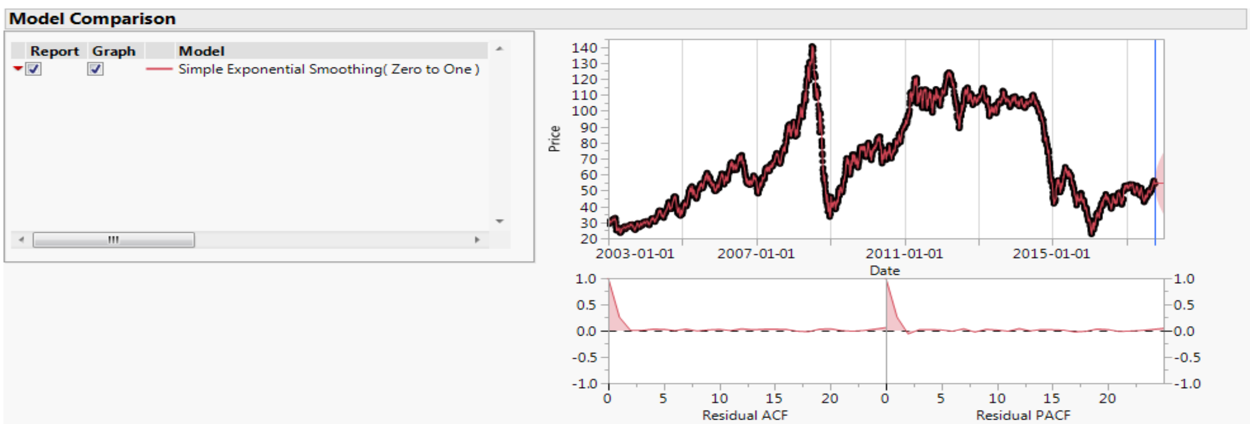
price is almost stabilized in the range of $100 to $120. Throughout the years from 2003 to 2017, there is huge fluctuations of oil price is observed between $25 and $120.

- The statistical tests like Standard deviation, Zero mean ADF, Single mean ADF and Trend ADF for the proposed models are analyzed. ADF stands for Augmented Dickey Fuller test, which is a way to check for a unit root within the time series and it is checking for an underlying time dependent trend within the data. The three metrics correspond to checking for no underlying time component (zero mean), constant increase over time (single mean) and acceleration over time (Trend). The Zero mean ADF, Single mean ADF and Trend ADF are in negative values which strongly emphasizes that the proposed model need to be standardized through smoothing
- By looking at the auto correlations and variogram functions, it is clearly evident that the proposed model data should be standardized.

2.2 MODELS:

Given the previous basic diagnostics, models are examined.

2.2.1 Simple Exponential Smoothing:



Model: Simple Exponential Smoothing(Zero to One)

Model Summary

DF	3809	Stable	Yes
Sum of Squared Errors	4227.9049	Invertible	Yes
Variance Estimate	1.10997766		
Standard Deviation	1.05355477		
Akaike's 'A' Information Criterion	11210.8465		
Schwarz's Bayesian Criterion	11217.0918		
RSquare	0.99866401		
RSquare Adj	0.99866401		
MAPE	1.21359921		
MAE	0.75824934		
-2LogLikelihood	11208.8465		

Parameter Estimates

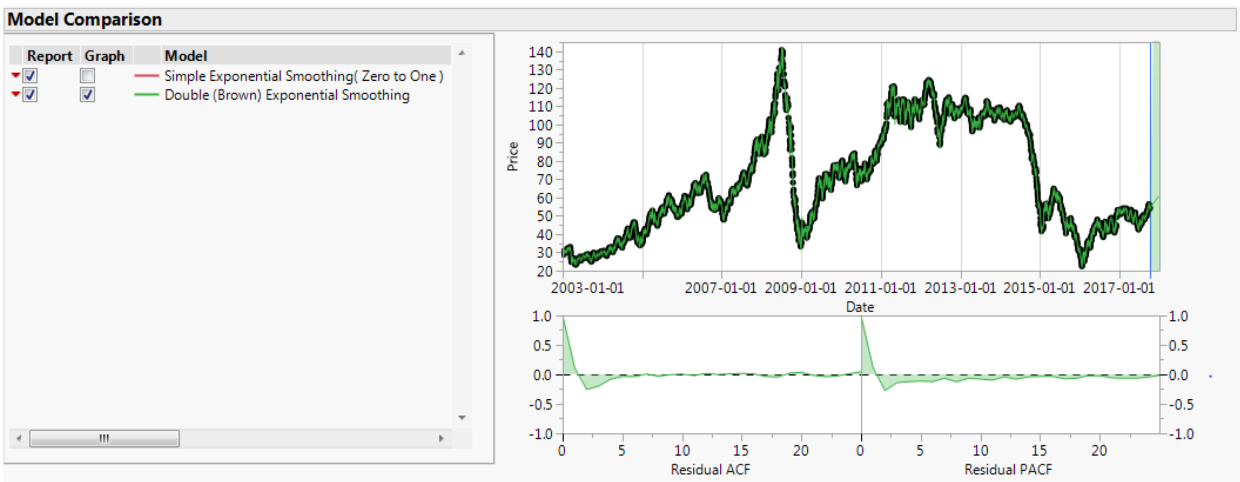
Term	Estimate	Std Error	t Ratio	Prob>|t|
Level Smoothing Weight	1.0000000	0.0172404	58.00	<.0001*

Figure 2.2 Simple Exponential Model Analysis

Interpretation:

- Accuracy measures like **MAPE, MAD and MSD** are used to compare the fitting of different time series models. Smaller values indicate a better fit.
- The model summary indicates that, MAPE = 1.21
  MAE = 0.7582
- And also the R-Square and R-Square adjusted values observed to be same (0.998) and it is approximately equal to 1. This clearly states that the model is highly stable.

2.2.2 Double Exponential Smoothing:



Model: Double (Brown) Exponential Smoothing

Model Summary

DF	3808	Stable	Yes
Sum of Squared Errors	4987.92324	Invertible	Yes
Variance Estimate	1.30985379		
Standard Deviation	1.14448844		
Akaike's 'A' Information Criterion	11839.1067		
Schwarz's Bayesian Criterion	11845.3518		
RSquare	0.99842297		
RSquare Adj	0.99842297		
MAPE	1.32578612		
MAE	0.83100523		
-2LogLikelihood	11837.1067		

Parameter Estimates

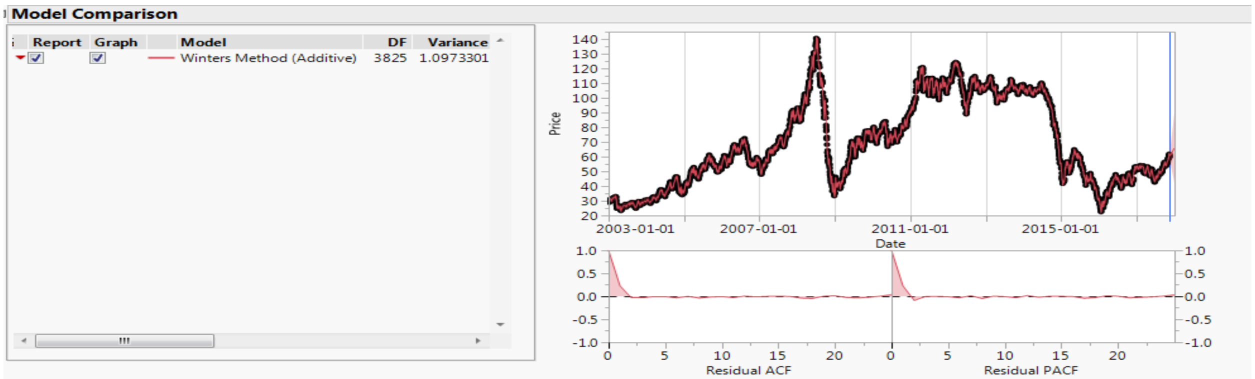
Term	Estimate	Std Error	t Ratio	Prob>|t|
Level Smoothing Weight	0.64934021	0.0115815	56.07	<.0001*

Figure 2.3 Double Exponential Model Analysis

Interpretation:

- Accuracy measures like **MAPE, MAD and MSD** are used to compare the fitting of different time series models. Smaller values indicate a better fit.
- The model summary indicates that, MAPE = 1.325
  - MAE = 0.831
- And also the R-Square and R-Square adjusted values observed to be same (0.998) and it is approximately equal to 1. This clearly states that the model is highly stable.

2.2.3 Winters method:



Model: Winters Method (Additive)

Model Summary

DF	3825	Stable	Yes
Sum of Squared Errors	4197.28746	Invertible	No
Variance Estimate	1.09733006		
Standard Deviation	1.04753523		
Akaike's 'A' Information Criterion	11286.024		
Schwarz's Bayesian Criterion	11304.7743		
RSquare	0.99865529		
RSquare Adj	0.99865459		
MAPE	1.21969815		
MAE	0.75975029		
-2LogLikelihood	11280.024		

Failed: Cannot Decrease Objective Function Hessian is not positive definite.

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Level Smoothing Weight	1.0000000	0.00000	.	.
Trend Smoothing Weight	0.0314833	.	.	.
Seasonal Smoothing Weight	0.0000000	14.51943	0.00	1.0000

Figure 2.4 Winter Model Analysis

Interpretation:

- On comparing the MAPE and MAE values of all the models we conclude that

	Simple Exponential Smoothing	Double Exponential Smoothing	Winters Method
MAPE	1.12135	1.325	1.2196
MAE	0.758	0.831	0.7597

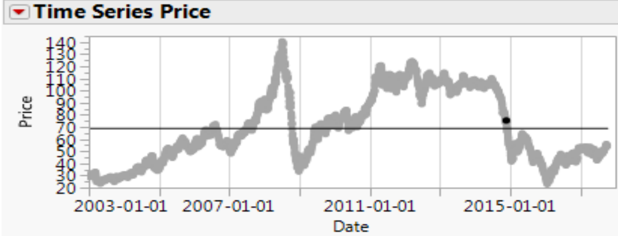
MAPE and MAE values got increased in double exponential smoothing when compared to simple exponential smoothing. This clearly indicates that 'Simple exponential smoothing model' is more fitted model than Double exponential method. Almost Simple Exponential and Winters Method are approximately same but there observed to be slight change in the error values, which clearly states that Simple Exponential smoothing is the best fitted model with less error values and more inclined towards its accuracy.

- Here also the R-square and R-Square adjusted observed to be same (0.998) and it is approximately equal to 1. This clearly states that Simple Exponential model is highly stable.
- Usually Double Exponential Smoothing model and winters method fits the data perfectly when there is a trend observed in the data. As there is no trend observed in the given data, it is clearly evident that Double exponential smoothing model won't give a standardized result.
- The t-ratio is estimate divided by the standard error. More the standard error less the t-ratio will be. On comparing the t-ratio results of Simple and Double exponential smoothing model, it is observed that simple exponential model is having the more value of 58 whereas Double Exponential smoothing model is having the value of 56.07.
- The estimate optimal alpha value is one, even though Simple exponential is highly stable, it is better to check for best fitted model through ARIMA
- As there is no trend is observed in the given data, there is no point in checking for other smoothing models like Holt's method etc.

2.2.4 ARIMA:

After a time series has been stationarized by differencing, the next step in fitting an ARIMA model is to determine whether AR or MA terms are needed to correct any autocorrelation that remains in the differenced series.

Time Series Price



Mean	69.435943
Std	28.82358
N	3811
Zero Mean ADF	-0.219617
Single Mean ADF	-1.483198
Trend ADF	-1.104411

Time Series Basic Diagnostics

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.9991	3806.78	<.0001*	1	0.9991
2	0.9978	7604.76	<.0001*	2	-0.1791
3	0.9965	11393.9	<.0001*	3	0.0262
4	0.9952	15174.0	<.0001*	4	-0.0161
5	0.9938	18944.7	<.0001*	5	-0.0232
6	0.9924	22705.8	<.0001*	6	-0.0080
7	0.9910	26457.3	<.0001*	7	0.0045
8	0.9896	30198.9	<.0001*	8	-0.0261
9	0.9881	33930.6	<.0001*	9	0.0123
10	0.9867	37652.4	<.0001*	10	-0.0163
11	0.9852	41363.9	<.0001*	11	-0.0147
12	0.9837	45065.0	<.0001*	12	0.0026
13	0.9821	48755.5	<.0001*	13	-0.0296
14	0.9805	52435.1	<.0001*	14	-0.0068
15	0.9789	56103.4	<.0001*	15	-0.0181
16	0.9772	59760.2	<.0001*	16	-0.0221
17	0.9755	63405.1	<.0001*	17	-0.0099
18	0.9738	67038.3	<.0001*	18	0.0066
19	0.9721	70659.8	<.0001*	19	0.0091
20	0.9704	74269.5	<.0001*	20	-0.0194
21	0.9686	77866.9	<.0001*	21	-0.0208
22	0.9669	81452.1	<.0001*	22	0.0051
23	0.9651	85025.0	<.0001*	23	-0.0005
24	0.9633	88585.6	<.0001*	24	-0.0071
25	0.9615	92133.7	<.0001*	25	-0.0179

Figure 2.5 ACF & PACF Analysis

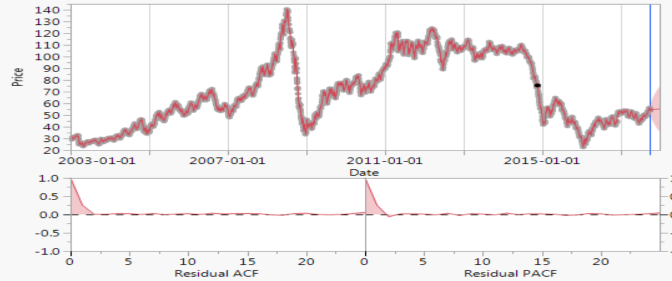
Interpretation:

From Time Series results, it is observed that Partial Auto correlation lag is 1. So, ARIMA model has been performed with the order (0, 1, 0)

ARIMA (0, 1, 0)

Model Comparison

Report	Graph	Model	DF	Variance	AIC	
☑	☐	ARI(1, 1)	3808	1.0362435	10950.024	10!
☑	☑	I(1)	3809	1.1099361	11210.704	11:



Model: I(1)

Model Summary

DF	3809	Stable	Yes
Sum of Squared Errors	4227.74645	Invertible	Yes
Variance Estimate	1.10993606		
Standard Deviation	1.05353503		
Akaike's 'A' Information Criterion	11210.7037		
Schwarz's Bayesian Criterion	11216.9491		
RSquare	0.99866406		
RSquare Adj	0.99866406		
MAPE	1.21325363		
MAE	0.75803608		
-2LogLikelihood	11208.7037		

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate	Mu
Intercept	0	0.00644884	0.0093940	0.69	0.4925	0.00644884	0.00644884

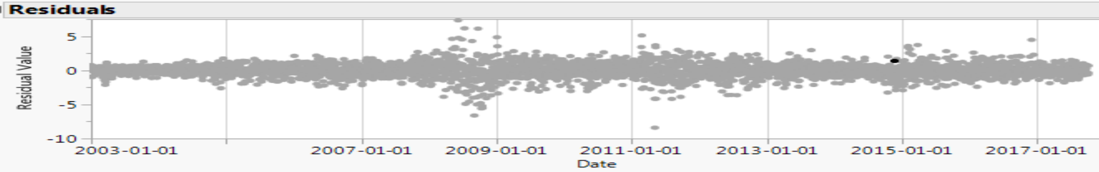
Figure 2.6 ARIMA(0,1,0) Model Analysis

Interpretation:

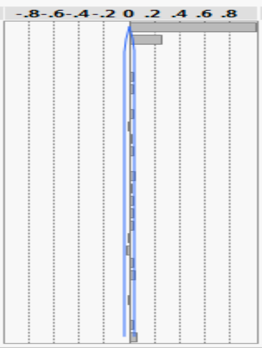
- Accuracy measures like **MAPE, MAD and MSD** are used to compare the fitting of different time series models. Smaller values indicate a better fit.
- The model summary indicates that, MAPE = 1.213
  MAE = 0.758
- And also the R-Square and R-Square adjusted values observed to be same (0.998) and it is approximately equal to 1. This clearly states that the model is highly stable.

Residuals:

Residuals



Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.2581	254.029	<.0001*	1	0.2581
2	0.0083	254.293	<.0001*	2	-0.0625
3	0.0077	254.520	<.0001*	3	0.0232
4	0.0315	258.303	<.0001*	4	0.0258
5	0.0296	261.646	<.0001*	5	0.0158
6	0.0026	261.671	<.0001*	6	-0.0084
7	0.0346	266.234	<.0001*	7	0.0400
8	-0.0035	266.280	<.0001*	8	-0.0262
9	0.0185	267.587	<.0001*	9	0.0292
10	0.0278	270.536	<.0001*	10	0.0153
11	0.0042	270.602	<.0001*	11	-0.0084
12	0.0403	276.811	<.0001*	12	0.0450
13	0.0201	278.351	<.0001*	13	-0.0025
14	0.0303	281.859	<.0001*	14	0.0259
15	0.0321	285.810	<.0001*	15	0.0202
16	0.0276	288.736	<.0001*	16	0.0132
17	-0.0079	288.975	<.0001*	17	-0.0226
18	-0.0194	290.413	<.0001*	18	-0.0106
19	0.0316	294.244	<.0001*	19	0.0351
20	0.0407	300.599	<.0001*	20	0.0227
21	0.0038	300.655	<.0001*	21	-0.0144
22	-0.0080	300.899	<.0001*	22	-0.0048
23	0.0055	301.014	<.0001*	23	0.0069
24	0.0333	305.264	<.0001*	24	0.0267
25	0.0614	319.728	<.0001*	25	0.0485



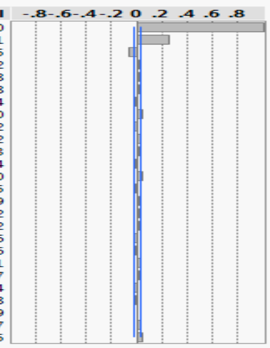


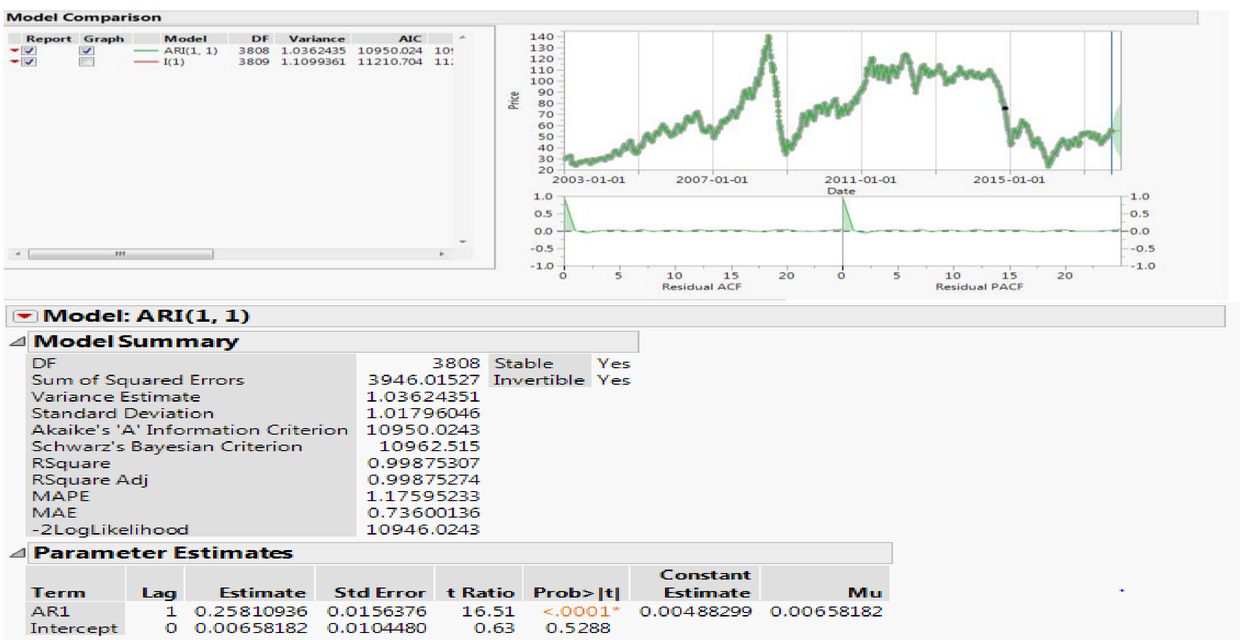
Figure 2.7 Residuals & ACF, PACF of ARIMA (0,1,0)

Interpretation:

- If the residuals appeared to behave randomly, it indicates that model fits the data well. On the other hand, if non-random structure is evident in the residuals, it clearly indicates that model fits the data poorly. So, if the model fit to the data correctly, that makes the statistical relationship between the predictive variable and the response variable.
- The graphical residual plot indicates that residuals scattered properly. As the residuals scattered completely, we can evident an impressive statistical relationship between the Price of the oil at respective time interval.
- Generally sum of residuals should be equal to zero and the residuals should be standardized among the mean line befitted in the middle of the plot. Here, all the residuals are scattered properly and retained around the mean line. This clearly states that the model fitter perfectly but up on careful visualization of PACF, it is observed to be of one significant spike. Still certain standard error retained in predicted model data. So, ARIMA need to be performed with the order (1, 1, 0).

ARIMA (1,1,0)

From the results of ARIMA (0, 1, 0) it is observed that, the PACF has only one significant spike, so consider now order as (1, 1, 0).



Model Summary

DF	3808	Stable	Yes
Sum of Squared Errors	3946.01527	Invertible	Yes
Variance Estimate	1.03624351		
Standard Deviation	1.01796046		
Akaike's 'A' Information Criterion	10950.0243		
Schwarz's Bayesian Criterion	10962.515		
RSquare	0.99875307		
RSquare Adj	0.99875274		
MAPE	1.17595233		
MAE	0.73600136		
-2LogLikelihood	10946.0243		

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate	Mu
AR1	1	0.25810936	0.0156376	16.51	<.0001*	0.00488299	0.00658182
Intercept	0	0.00658182	0.0104480	0.63	0.5288		

Figure 2.8 ARIMA (1,1,0) Model Analysis

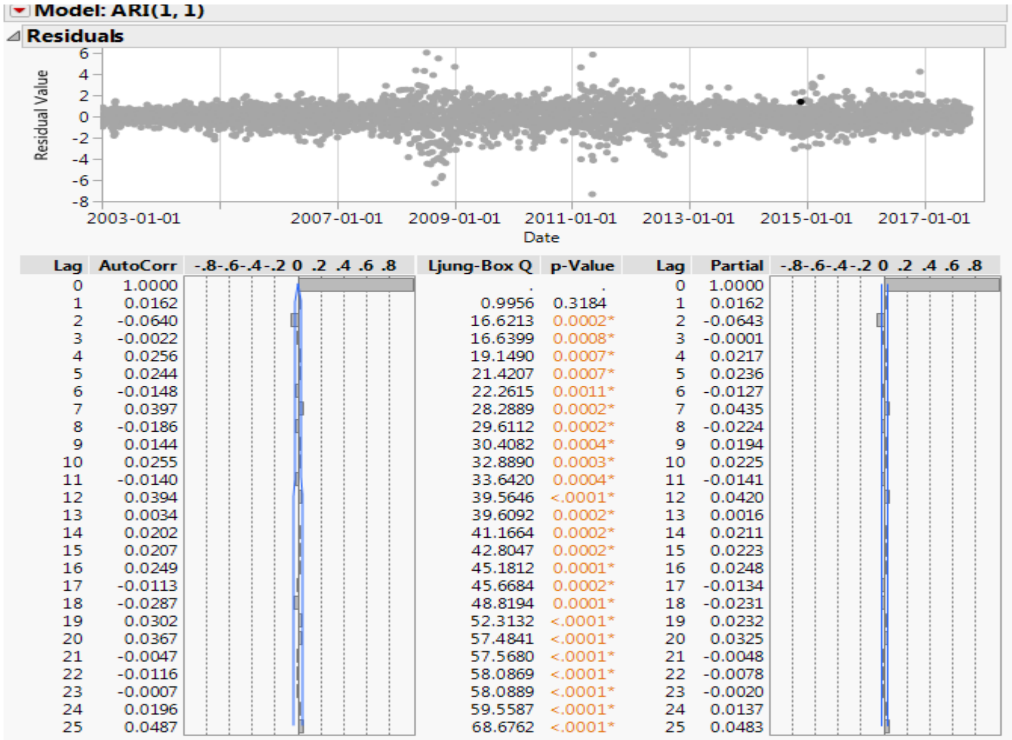
Interpretation:

- Accuracy measures like **MAPE, MAD and MSD** are used to compare the fitting of different time series models. Smaller values indicate a better fit.
- The model summary indicates that, MAPE = 1.17
  MAE = 0.73
- And also the R-Square and R-Square adjusted values observed to be same (0.998) and it is approximately equal to 1. This clearly states that the model is highly stable.
- On comparing the MAPE and MAE values of both the models we conclude that

	ARIMA(0, 1, 0)	ARIMA(1, 1, 0)
MAPE	1.213	1.17
MAE	0.758	0.73

- MAPE and MAE values got decreased in ARIMA (1, 1, 0) model when compared to ARIMA (0, 1, 0) model. This clearly indicates that 'ARIMA (1, 1, 0) model' is more fitted model with less error values and more inclined towards its accuracy. And perfect model for forecasting.

Residuals:



Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.0162	0.9956	0.3184	1	0.0162
2	-0.0640	16.6213	0.0002*	2	-0.0643
3	-0.0022	16.6399	0.0008*	3	-0.0001
4	0.0256	19.1490	0.0007*	4	0.0217
5	0.0244	21.4207	0.0007*	5	0.0236
6	-0.0148	22.2615	0.0011*	6	-0.0127
7	0.0397	28.2889	0.0002*	7	0.0435
8	-0.0186	29.6112	0.0002*	8	-0.0224
9	0.0144	30.4082	0.0004*	9	0.0194
10	0.0255	32.8890	0.0003*	10	0.0225
11	-0.0140	33.6420	0.0004*	11	-0.0141
12	0.0394	39.5646	<.0001*	12	0.0420
13	0.0034	39.6092	0.0002*	13	0.0016
14	0.0202	41.1664	0.0002*	14	0.0211
15	0.0207	42.8047	0.0002*	15	0.0223
16	0.0249	45.1812	0.0001*	16	0.0248
17	-0.0113	45.6684	0.0002*	17	-0.0134
18	-0.0287	48.8194	0.0001*	18	-0.0231
19	0.0302	52.3132	<.0001*	19	0.0232
20	0.0367	57.4841	<.0001*	20	0.0325
21	-0.0047	57.5680	<.0001*	21	-0.0048
22	-0.0116	58.0869	<.0001*	22	-0.0078
23	-0.0007	58.0889	<.0001*	23	-0.0020
24	0.0196	59.5587	<.0001*	24	0.0137
25	0.0487	68.6762	<.0001*	25	0.0483

Figure 2.9 Residuals & ACF, PACF of ARIMA (1,1,0)

Interpretation:

There observed to no lags in PCFA, which clearly infers that the fitted model is free of error values.

Forecast the data through ARIMA (1, 1, 0) model:

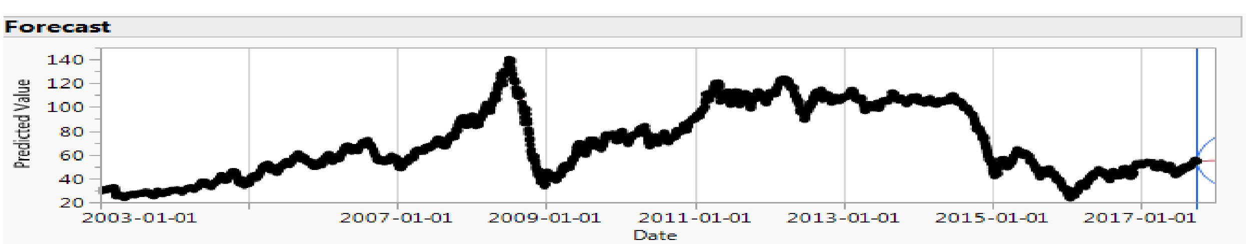


Figure 2.10 ARIMA (1,1,0) forecast

Forecasted Data:

	Actual Price	Date	Predicted Price	Std Err Pred Price	Residual Price	Upper CL (0.95) Price	Lower CL (0.95) Price
3811	54.62	2017-10-05	53.936448835	1.0535350303	0.6835511647	56.001339551	51.87155812
3812	•	2017-10-06	54.626448835	1.0535350303	•	56.691339551	52.56155812
3813	•	2017-10-07	54.632897671	1.4899235282	•	57.553094126	51.712701215
3814	•	2017-10-08	54.639346506	1.8247762	•	58.215842138	51.062850874
3815	•	2017-10-09	54.645795341	2.1070700605	•	58.775576773	50.51601391
3816	•	2017-10-10	54.652244176	2.3557759443	•	59.269480183	50.03500817
3817	•	2017-10-11	54.658693012	2.5806232503	•	59.71662164	49.600764383
3818	•	2017-10-12	54.665141847	2.7873916876	•	60.128329165	49.201954528
3819	•	2017-10-13	54.671590682	2.9798470565	•	60.511983592	48.831197772
3820	•	2017-10-14	54.678039517	3.1606050908	•	60.872711665	48.48336737
3821	•	2017-10-15	54.684488353	3.3315702904	•	61.214246134	48.154730572
3822	•	2017-10-16	54.690937188	3.4941803989	•	61.539404925	47.842469451
3823	•	2017-10-17	54.697386023	3.6495523999	•	61.850377287	47.54439476
3824	•	2017-10-18	54.703834859	3.7985745721	•	62.148904212	47.258765505
3825	•	2017-10-19	54.710283694	3.9419671282	•	62.436397293	46.984170094
3826	•	2017-10-20	54.716732529	4.0803236269	•	62.714019883	46.719445175
3827	•	2017-10-21	54.723181364	4.2141401211	•	62.982744227	46.463618501
3828	•	2017-10-22	54.7296302	4.3438362101	•	63.243392726	46.215867673
3829	•	2017-10-23	54.736079035	4.4697705847	•	63.4966684	45.97548967
3830	•	2017-10-24	54.74252787	4.5922527304	•	63.74317783	45.741877911
3831	•	2017-10-25	54.748976705	4.7115518887	•	63.983448719	45.514504692
3832	•	2017-10-26	54.755425541	4.8279040235	•	64.217943548	45.292907534
3833	•	2017-10-27	54.761874376	4.9415173095	•	64.447070332	45.07667842
3834	•	2017-10-28	54.768323211	5.0525765091	•	64.671191198	44.865455224
3835	•	2017-10-29	54.774772047	5.1612465006	•	64.890629303	44.65891479
3836	•	2017-10-30	54.781220882	5.2676751513	•	65.105674461	44.456767303
3837	•	2017-10-31	54.787669717	5.3719956776	•	65.31658777	44.258751664
3838	•	2017-11-01	54.794118552	5.4743285999	•	65.523605448	44.064631657
3839	•	2017-11-02	54.800567388	5.5747833752	•	65.726942024	43.874192751
3840	•	2017-11-03	54.807016223	5.6734597681	•	65.926793036	43.68723941
3841	•	2017-11-04	54.813465058	5.770449012	•	66.123337296	43.50359282
3842	•	2017-11-05	54.819913893	5.8658347965	•	66.316738834	43.323088953

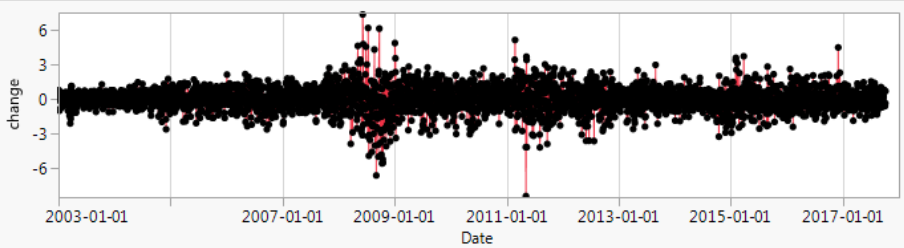
The above table is the forecasted data for 30 observations of fitted ARIMA (1,1,0) model.

3. CHANGE IN PRICE DATASET

3.1 Basic Diagnostics:

Plotting Time Series for the data:

Time Series change



Mean	0.0064488
Std	1.0533968
N	3810
Zero Mean ADF	-47.38489
Single Mean ADF	-47.37998
Trend ADF	-47.39343

Time Series Basic Diagnostics

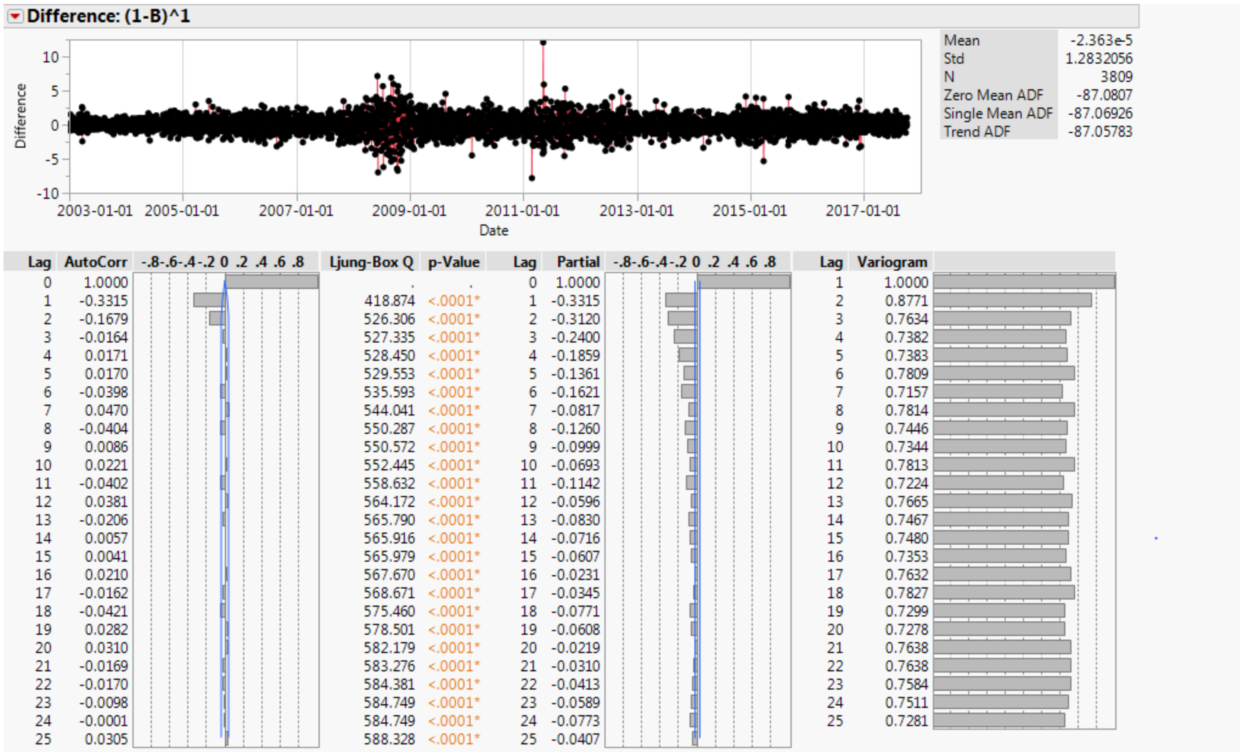
Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial	Lag	Variogram
0	1.0000	.	.	0	1.0000	1	1.0000
1	0.2581	254.029	<.0001*	1	0.2581	2	1.3367
2	0.0083	254.293	<.0001*	2	-0.0625	3	1.3375
3	0.0077	254.520	<.0001*	3	0.0232	4	1.3055
4	0.0315	258.303	<.0001*	4	0.0258	5	1.3080
5	0.0296	261.646	<.0001*	5	0.0158	6	1.3445
6	0.0026	261.671	<.0001*	6	-0.0084	7	1.3013
7	0.0346	266.234	<.0001*	7	0.0400	8	1.3526
8	-0.0035	266.280	<.0001*	8	-0.0262	9	1.3230
9	0.0185	267.587	<.0001*	9	0.0292	10	1.3105
10	0.0278	270.536	<.0001*	10	0.0153	11	1.3423
11	0.0042	270.602	<.0001*	11	-0.0084	12	1.2936
12	0.0403	276.811	<.0001*	12	0.0450	13	1.3209
13	0.0201	278.351	<.0001*	13	-0.0025	14	1.3071
14	0.0303	281.859	<.0001*	14	0.0259	15	1.3046
15	0.0321	285.810	<.0001*	15	0.0202	16	1.3106
16	0.0276	288.736	<.0001*	16	0.0132	17	1.3586
17	-0.0079	288.975	<.0001*	17	-0.0226	18	1.3740
18	-0.0194	290.413	<.0001*	18	-0.0106	19	1.3053
19	0.0316	294.244	<.0001*	19	0.0351	20	1.2930
20	0.0407	300.599	<.0001*	20	0.0227	21	1.3428
21	0.0038	300.655	<.0001*	21	-0.0144	22	1.3586
22	-0.0080	300.899	<.0001*	22	-0.0048	23	1.3405
23	0.0055	301.014	<.0001*	23	0.0069	24	1.3030
24	0.0333	305.264	<.0001*	24	0.0267	25	1.2652
25	0.0614	319.728	<.0001*	25	0.0485		

Figure 3.1 Time Series Analysis

Interpretation:

- If the residuals appeared to behave randomly, it indicates that model fits the data well. On the other hand, if non-random structure is evident in the residuals, it clearly indicates that model fits the data poorly. So, if the model fit to the data correctly, that makes the statistical relationship between the predictive variable and the response variable.
- The graphical residual plot indicates that residuals scattered properly. As the residuals scattered completely, we can evident an impressive statistical relationship between the change in Price of the oil at respective time interval.
- Here, all the residuals are scattered properly and retained around the mean line.

First Difference:



Mean	-2.363e-5
Std	1.2832056
N	3809
Zero Mean ADF	-87.0807
Single Mean ADF	-87.06926
Trend ADF	-87.05783

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial	Lag	Variogram
0	1.0000	.	.	0	1.0000	1	1.0000
1	-0.3315	418.874	<.0001*	1	-0.3315	2	0.8771
2	-0.1679	526.306	<.0001*	2	-0.3120	3	0.7634
3	-0.0164	527.335	<.0001*	3	-0.2400	4	0.7382
4	0.0171	528.450	<.0001*	4	-0.1859	5	0.7383
5	0.0170	529.553	<.0001*	5	-0.1361	6	0.7809
6	-0.0398	535.593	<.0001*	6	-0.1621	7	0.7157
7	0.0470	544.041	<.0001*	7	-0.0817	8	0.7814
8	-0.0404	550.287	<.0001*	8	-0.1260	9	0.7446
9	0.0086	550.572	<.0001*	9	-0.0999	10	0.7344
10	0.0221	552.445	<.0001*	10	-0.0693	11	0.7813
11	-0.0402	558.632	<.0001*	11	-0.1142	12	0.7224
12	0.0381	564.172	<.0001*	12	-0.0596	13	0.7665
13	-0.0206	565.790	<.0001*	13	-0.0830	14	0.7467
14	0.0057	565.916	<.0001*	14	-0.0716	15	0.7480
15	0.0041	565.979	<.0001*	15	-0.0607	16	0.7353
16	0.0210	567.670	<.0001*	16	-0.0231	17	0.7632
17	-0.0162	568.671	<.0001*	17	-0.0345	18	0.7827
18	-0.0421	575.460	<.0001*	18	-0.0771	19	0.7299
19	0.0282	578.501	<.0001*	19	-0.0608	20	0.7278
20	0.0310	582.179	<.0001*	20	-0.0219	21	0.7638
21	-0.0169	583.276	<.0001*	21	-0.0310	22	0.7638
22	-0.0170	584.381	<.0001*	22	-0.0413	23	0.7584
23	-0.0098	584.749	<.0001*	23	-0.0589	24	0.7511
24	-0.0001	584.749	<.0001*	24	-0.0773	25	0.7281
25	0.0305	588.328	<.0001*	25	-0.0407		

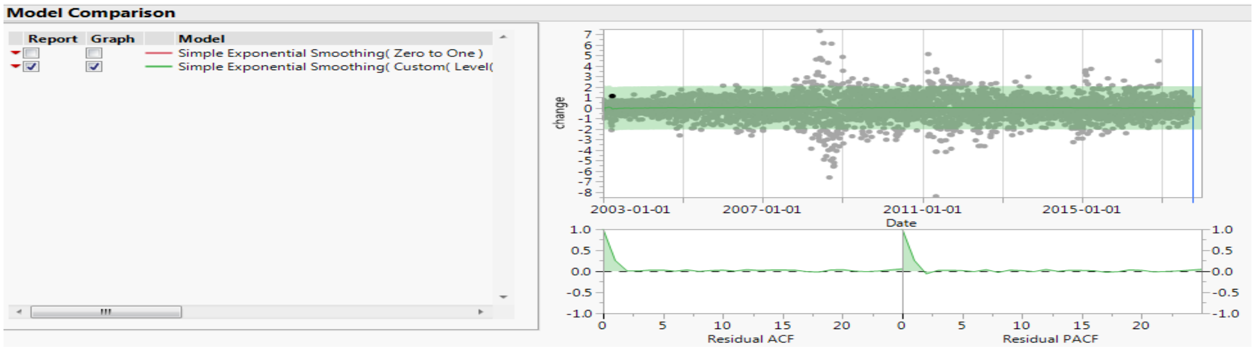
Figure 3.2 First Difference Analysis

Interpretation:

- It displays graphs of the autocorrelations and partial auto correlations of the series. These indicate how and to what degree each point in the series is correlated with earlier values in the series.
- It helps to understand the underlying structure and function that produce the observations. It is assumed that a time series data set has at least one symmetric pattern. The most common patterns are trends and seasonality. There is no specific or impressive trend is followed in the proposed time series model for the given data. This emphasis that the given data is not standardized and it needs some smoothening procedure to standardize.
- Visual inspection shows there is a huge fluctuation of change in price through the years from 2003 to 2017 but all the change in price values varied in the range of +5 to -5.
- Noticed the test statistics like Stand deviation, Zero mean ADF, Single mean ADF and Trend ADF for the proposed model. ADF stands for Augmented Dickey Fuller test, which is a way to check for a unit root within the time series and it is checking for an underlying time dependent trend within the data. The three metrics correspond to checking for no underlying time component (zero mean), constant increase over time (single mean) and acceleration over time (Trend). The Zero mean ADF, Single mean ADF and Trend ADF are in negative values which strongly emphasizes that the proposed model need to be standardized through smoothing
- By looking at the auto correlations and variogram functions, it is clearly evident that the proposed model data should be standardized.

3.2 MODELS:

3.2.1 Simple Exponential Smoothing:



Model: Simple Exponential Smoothing(Zero to One)

Model Summary

DF	3808	Stable	Yes
Sum of Squared Errors	4199.45161	Invertible	Yes
Variance Estimate	1.10279717		
Standard Deviation	1.0501415		
Akaike's 'A' Information Criterion	11185.9633		
Schwarz's Bayesian Criterion	11192.2084		
RSquare	0.00621717		
RSquare Adj	0.00621717		
MAPE	.		
MAE	0.75722133		
-2LogLikelihood	11183.9633		

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Level Smoothing Weight	0.03152203	0.0057755	5.46	<.0001*

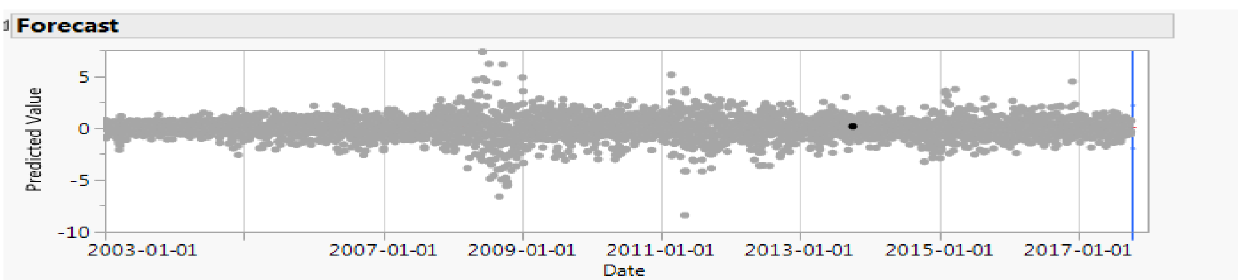
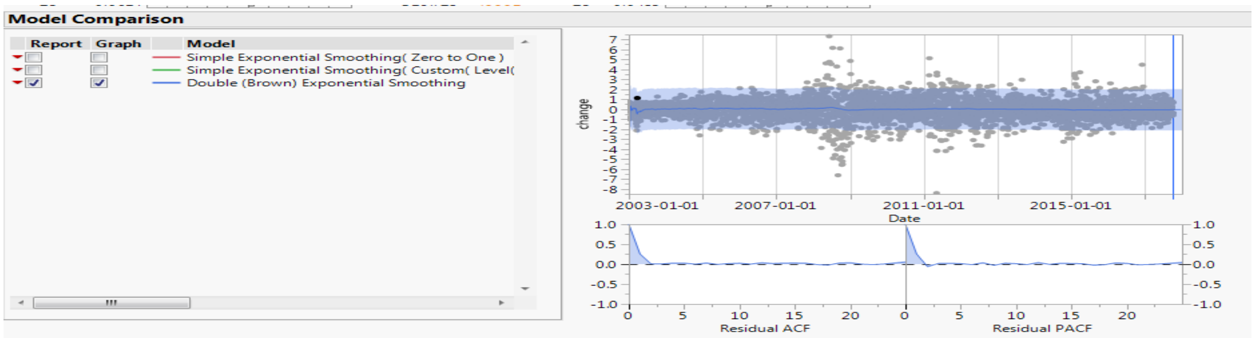


Figure 3.3 Simple Exponential Smoothing Analysis

Interpretation:

- Accuracy measures like **MAPE, MAD and MSD** are used to compare the fitting of different time series models. Smaller values indicate a better fit.
- The model summary indicates that, MAE = 0.7582
- And also the R-Square and R-Square adjusted values observed to be negative. It is possible to get a negative R-square for equations that do not contain a constant term because R-square is defined as the proportion of variance explained by the fit, if the fit is actually worse than just fitting a horizontal line then R-square is negative. As R-Square is negative it is not a stable model.

3.2.2 Double Exponential Smoothing:



Model: Double (Brown) Exponential Smoothing

Model Summary

DF	3807	Stable	Yes
Sum of Squared Errors	4230.26389	Invertible	Yes
Variance Estimate	1.11118043		
Standard Deviation	1.05412543		
Akaike's 'A' Information Criterion	11223.6216		
Schwarz's Bayesian Criterion	11229.8665		
RSquare	-0.002298		
RSquare Adj	-0.002298		
MAPE	.		
MAE	0.75997782		
-2LogLikelihood	11221.6216		

Parameter Estimates

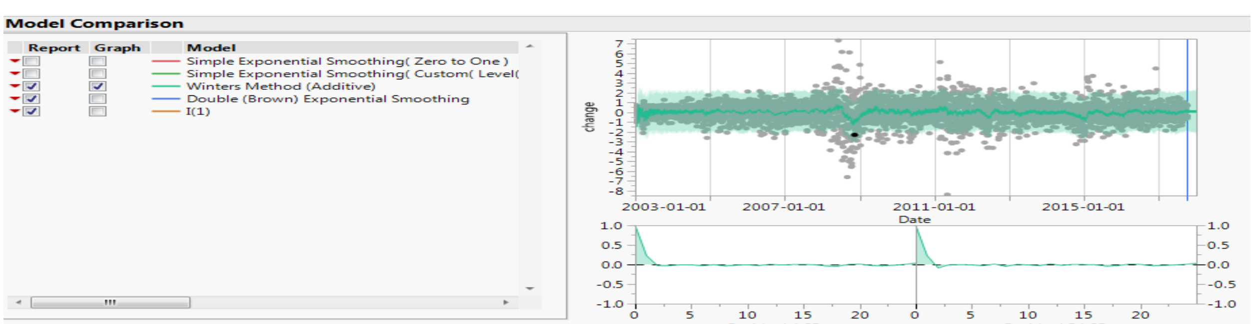
Term	Estimate	Std Error	t Ratio	Prob>|t|
Level Smoothing Weight	0.01329371	0.0004129	32.20	<.0001*

Figure 3.4 Double Exponential Smoothing Analysis

Interpretation:

- On comparing the MAE values of both the models we conclude that they are observed to be same. The results are not sufficient to predict the most fitted model.
- And also the R-Square and R-Square adjusted values observed to be negative. It is possible to get a negative R-square for equations that do not contain a constant term because R-square is defined as the proportion of variance explained by the fit, if the fit is actually worse than just fitting a horizontal line then R-square is negative. Here also R-Square is negative it is not a stable model.
- Usually Double Exponential Smoothing model fits the data perfectly when there is a trend observed in the data. As there is no trend observed in the given data, it is clearly evident that Double exponential smoothing model won't give a standardized result.
- As there is no trend is observed in the given data, there is no point in checking for other smoothing models like Holt's method etc.

3.2.3 Winters Model:



Model: Winters Method (Additive)

Model Summary

DF	3794	Stable	Yes
Sum of Squared Errors	4188.15232	Invertible	Yes
Variance Estimate	1.10388833		
Standard Deviation	1.0506609		
Akaike's 'A' Information Criterion	11225.0266		
Schwarz's Bayesian Criterion	11243.7525		
RSquare	-0.0015272		
RSquare Adj	-0.0020552		
MAPE	.		
MAE	0.76276886		
-2LogLikelihood	11219.0266		

Figure 3.5 Winters Model Analysis

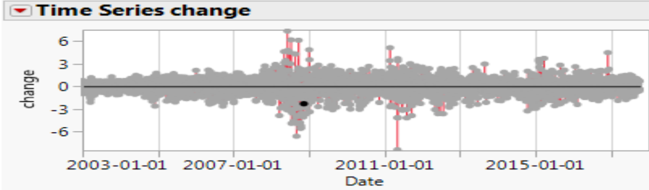
Interpretation:

- The model summary indicates that, MAE = 0.762. So the MAE value observed to be more for the winters method, this indicates it is not a fitted model.
- And also the R-Square and R-Square adjusted values observed to be negative. It is possible to get a negative R-square for equations that do not contain a constant term because R-square is defined as the proportion of variance explained by the fit, if the fit is actually worse than just fitting a horizontal line then R-square is negative.
- It is difficult to identify the stability of change in price through smoothing model, it is better to check for ARIMA.

3.2.4 ARIMA:

After a time series has been stationarized by differencing, the next step in fitting an ARIMA model is to determine whether AR or MA terms are needed to correct any autocorrelation that remains in the differenced series.

Time Series change



Mean	0.0064488
Std	1.0533968
N	3810
Zero Mean ADF	-47.38489
Single Mean ADF	-47.37998
Trend ADF	-47.39343

Time Series Basic Diagnostics

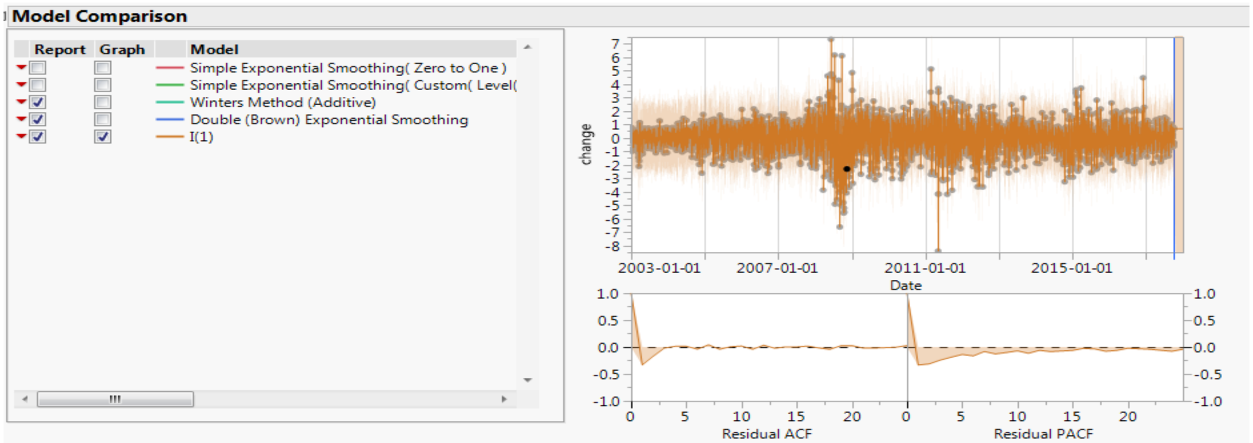
Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.2581	254.029	<.0001*	1	0.2581
2	0.0083	254.293	<.0001*	2	-0.0625
3	0.0077	254.520	<.0001*	3	0.0232
4	0.0315	258.303	<.0001*	4	0.0258
5	0.0296	261.646	<.0001*	5	0.0158
6	0.0026	261.671	<.0001*	6	-0.0084
7	0.0346	266.234	<.0001*	7	0.0400
8	-0.0035	266.280	<.0001*	8	-0.0262
9	0.0185	267.587	<.0001*	9	0.0292
10	0.0278	270.536	<.0001*	10	0.0153
11	0.0042	270.602	<.0001*	11	-0.0084
12	0.0403	276.811	<.0001*	12	0.0450
13	0.0201	278.351	<.0001*	13	-0.0025
14	0.0303	281.859	<.0001*	14	0.0259
15	0.0321	285.810	<.0001*	15	0.0202
16	0.0276	288.736	<.0001*	16	0.0132
17	-0.0079	288.975	<.0001*	17	-0.0226
18	-0.0194	290.413	<.0001*	18	-0.0106
19	0.0316	294.244	<.0001*	19	0.0351
20	0.0407	300.599	<.0001*	20	0.0227
21	0.0038	300.655	<.0001*	21	-0.0144
22	-0.0080	300.899	<.0001*	22	-0.0048
23	0.0055	301.014	<.0001*	23	0.0069
24	0.0333	305.264	<.0001*	24	0.0267
25	0.0614	319.728	<.0001*	25	0.0485

Figure 3.6 ACF & PACF Analysis

Interpretation:

From Time Series results, it is observed that Partial Auto correlation lag is 1.
So, ARIMA model has been performed with the order (0, 1, 0)

ARIMA (0, 1, 0)



Model: I(1)

Model Summary

DF	3808	Stable	Yes
Sum of Squared Errors	6271.9629	Invertible	Yes
Variance Estimate	1.64704908		
Standard Deviation	1.2833741		
Akaike's 'A' Information Criterion	12711.1084		
Schwarz's Bayesian Criterion	12717.3536		
RSquare	-0.483734		
RSquare Adj	-0.483734		
MAPE	.		
MAE	0.93722993		
-2LogLikelihood	12709.1084		

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate	Mu
Intercept	0	-0.0000236	0.0000386	-0.61	0.5419	-2.3571e-5	-2.3571e-5

Figure 3.7 ARIMA (0,1,0) Analysis

Interpretation:

Accuracy measures like **MAPE, MAD and MSD** are used to compare the fitting of different time series models. Smaller values indicate a better fit.

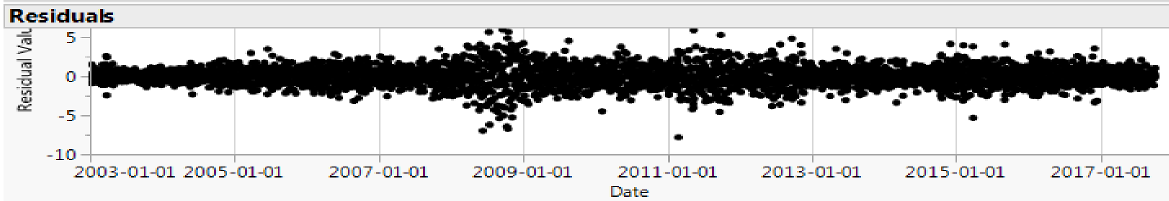
The model summary indicates that, MAE = 0.937

Here also the R-Square and R-Square adjusted values observed to be negative and even estimate value is also negative. It is not a best fitted model.

Residuals:

Model: I(1)

Residuals



Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	-0.3315		418.874	<.0001*	1	-0.3315	
2	-0.1679		526.306	<.0001*	2	-0.3120	
3	-0.0164		527.335	<.0001*	3	-0.2400	
4	0.0171		528.450	<.0001*	4	-0.1859	
5	0.0170		529.553	<.0001*	5	-0.1361	
6	-0.0398		535.593	<.0001*	6	-0.1621	
7	0.0470		544.041	<.0001*	7	-0.0817	
8	-0.0404		550.287	<.0001*	8	-0.1260	
9	0.0086		550.572	<.0001*	9	-0.0999	
10	0.0221		552.445	<.0001*	10	-0.0693	
11	-0.0402		558.632	<.0001*	11	-0.1142	
12	0.0381		564.172	<.0001*	12	-0.0596	
13	-0.0206		565.790	<.0001*	13	-0.0830	
14	0.0057		565.916	<.0001*	14	-0.0716	
15	0.0041		565.979	<.0001*	15	-0.0607	
16	0.0210		567.670	<.0001*	16	-0.0231	
17	-0.0162		568.671	<.0001*	17	-0.0345	
18	-0.0421		575.460	<.0001*	18	-0.0771	
19	0.0282		578.501	<.0001*	19	-0.0608	
20	0.0310		582.179	<.0001*	20	-0.0219	
21	-0.0169		583.276	<.0001*	21	-0.0310	
22	-0.0170		584.381	<.0001*	22	-0.0413	
23	-0.0098		584.749	<.0001*	23	-0.0589	
24	-0.0001		584.749	<.0001*	24	-0.0773	
25	0.0305		588.328	<.0001*	25	-0.0407	

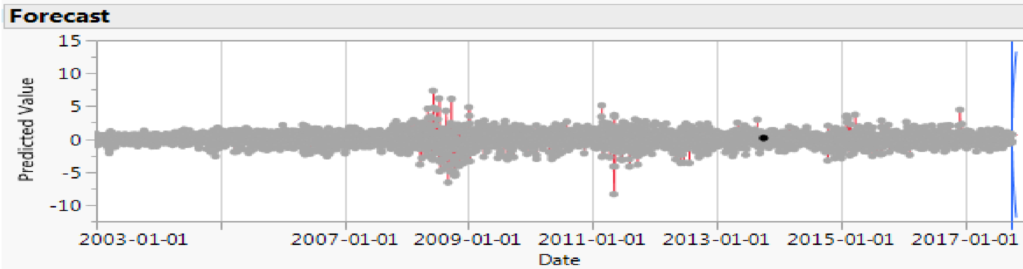
Figure 3.8 ARIMA (0,1,0) Residual Analysis

Interpretation:

- If the residuals appeared to behave randomly, it indicates that model fits the data well. On the other hand, if non-random structure is evident in the residuals, it clearly indicates that model fits the data poorly. So, if the model fit to the data correctly, that makes the statistical relationship between the predictive variable and the response variable.
- The graphical residual plot indicates that residuals scattered properly. As the residuals scattered completely, we can evident an impressive statistical relationship between the change in Price of the oil at respective time interval.
- Generally sum of residuals should be equal to zero and the residuals should be standardized among the mean line befitted in the middle of the plot. Here, all the residuals are scattered properly and retained around the mean line.

Forecast:

Forecast



	Actual change	Date	Predicted change	Std Err Pred change	Residual change	Upper CL (0.95) change	Lower CL (0.95) change
3807	-0.39	2017-09-29	-0.480023571	1.2833740999	0.0900235709	2.0353434436	-2.995390585
3808	-0.6	2017-10-02	-0.390023571	1.2833740999	-0.209976429	2.1253434436	-2.905390585
3809	-0.32	2017-10-03	-0.600023571	1.2833740999	0.2800235709	1.9153434436	-3.115390585
3810	-0.35	2017-10-04	-0.320023571	1.2833740999	-0.029976429	2.1953434436	-2.835390585
3811	0.69	2017-10-05	-0.350023571	1.2833740999	1.0400235709	2.1653434436	-2.865390585
3812	•	2017-10-06	0.6899764291	1.2833740999	•	3.2053434436	-1.825390585
3813	•	2017-10-07	0.6899528582	1.8149650577	•	4.2472190045	-2.867313288
3814	•	2017-10-08	0.6899292874	2.2228691462	•	5.0466727562	-3.666814181
3815	•	2017-10-09	0.6899057165	2.5667481998	•	5.7206397455	-4.340828313
3816	•	2017-10-10	0.6898821456	2.869711728	•	6.3144137784	-4.934649487
3817	•	2017-10-11	0.6898585747	3.1436116939	•	6.8512242761	-5.471507127
3818	•	2017-10-12	0.6898350038	3.3954887074	•	7.3448705803	-5.965200573
3819	•	2017-10-13	0.6898114329	3.6299301154	•	7.8043437255	-6.42472086
3820	•	2017-10-14	0.6897878621	3.8501222997	•	8.2358889056	-6.856313181
3821	•	2017-10-15	0.6897642912	4.0583852458	•	8.6440532083	-7.264524626
3822	•	2017-10-16	0.6897407203	4.2564703551	•	9.0322693175	-7.652787877
3823	•	2017-10-17	0.6897171494	4.4457382923	•	9.4032040871	-8.023769788
3824	•	2017-10-18	0.6896935785	4.6272711228	•	9.758978326	-8.379591169
3825	•	2017-10-19	0.6896700076	4.8019461809	•	10.101311578	-8.721971563
3826	•	2017-10-20	0.6896464368	4.9704865159	•	10.431620994	-9.05232812
3827	•	2017-10-21	0.6896228659	5.1334963996	•	10.751090924	-9.371845192
3828	•	2017-10-22	0.689599295	5.2914869711	•	11.060723183	-9.681524593
3829	•	2017-10-23	0.6895757241	5.4448951731	•	11.361374163	-9.982222715
3830	•	2017-10-24	0.6895521532	5.5940980083	•	11.653782775	-10.27467847
3831	•	2017-10-25	0.6895285824	5.7394234559	•	11.938591848	-10.55953468
3832	•	2017-10-26	0.6895050115	5.8811589578	•	12.216364756	-10.83735473
3833	•	2017-10-27	0.6894814406	6.019558104	•	12.487598527	-11.10863565
3834	•	2017-10-28	0.6894578697	6.1548459646	•	12.752734291	-11.37381855
3835	•	2017-10-29	0.6894342988	6.2872233878	•	13.012165702	-11.6332971
3836	•	2017-10-30	0.6894107279	6.4168704996	•	13.266245801	-11.88742434

Figure 3.9 Forecasted data

The above table is forecasted data of 30 observations based on the ARIMA model (0,1,0).

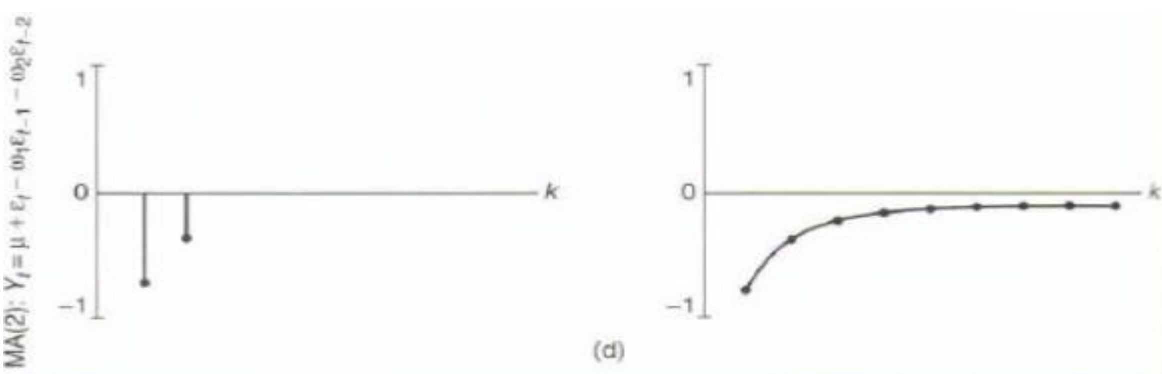
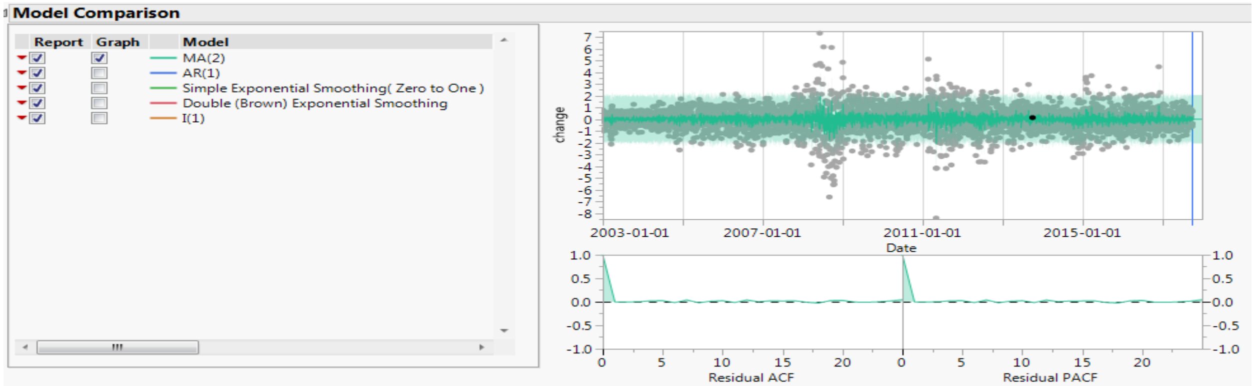


FIGURE 9-3 Autocorrelation and Partial Autocorrelation Coefficients of MA(1) and MA(2) Models

We observed that the ACF and PACF layouts are similar to that of the MA (2) model. Therefore, we will try to fit the MA (2) model to this dataset.

ARIMA – MA (2) Model:



Model: MA(2)

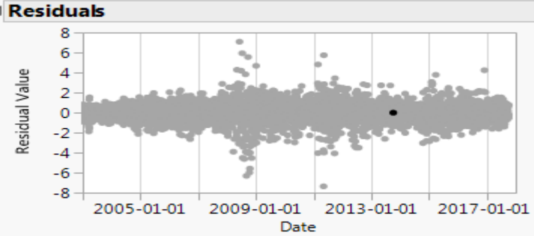
Model Summary

DF	3807	Stable	Yes
Sum of Squared Errors	3929.39201	Invertible	Yes
Variance Estimate	1.0321492		
Standard Deviation	1.01594744		
Akaike's 'A' Information Criterion	10935.9487		
Schwarz's Bayesian Criterion	10954.6849		
RSquare	0.07056042		
RSquare Adj	0.07007214		
MAPE	.		
MAE	0.73380441		
-2LogLikelihood	10929.9487		

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate	Mu
MA1	1	-0.2754215	0.0160766	-17.13	<.0001*	0.00656229	0.00656229
MA2	2	-0.0091164	0.0142969	-0.64	0.5237		
Intercept	0	0.0065623	0.0172730	0.38	0.7040		

Residuals



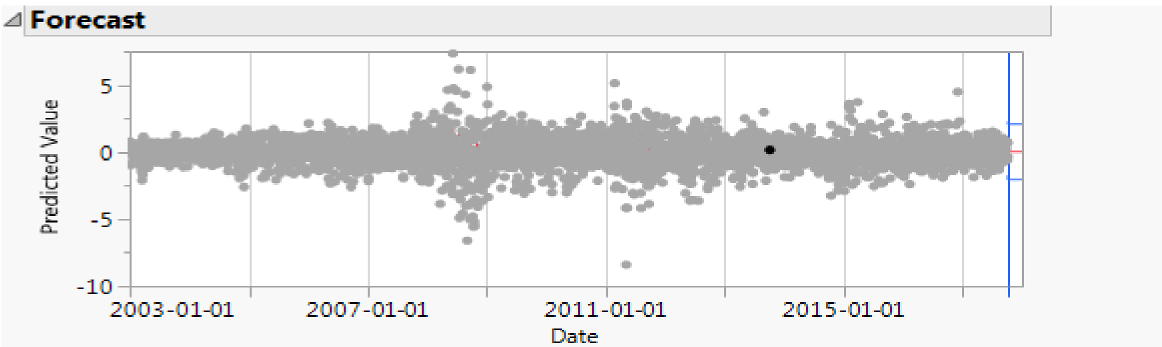
Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.0000	0.0000	0.9999	1	0.0000
2	-0.0007	0.0020	0.9990	2	-0.0007
3	0.0014	0.0093	0.9998	3	0.0014
4	0.0243	2.2576	0.6885	4	0.0243
5	0.0270	5.0318	0.4120	5	0.0270
6	-0.0156	5.9616	0.4275	6	-0.0156
7	0.0431	13.0483	0.0709	7	0.0431
8	-0.0188	14.4022	0.0719	8	-0.0196
9	0.0162	15.4011	0.0805	9	0.0151
10	0.0270	18.1920	0.0518	10	0.0270
11	-0.0140	18.9403	0.0622	11	-0.0153
12	0.0428	25.9350	0.0110*	12	0.0415
13	0.0028	25.9648	0.0172*	13	0.0042
14	0.0239	28.1558	0.0136*	14	0.0196
15	0.0196	29.6268	0.0133*	15	0.0212
16	0.0247	31.9665	0.0101*	16	0.0227
17	-0.0081	32.2182	0.0141*	17	-0.0125
18	-0.0254	34.6811	0.0104*	18	-0.0232
19	0.0294	37.9897	0.0060*	19	0.0226
20	0.0342	42.4816	0.0024*	20	0.0339
21	-0.0032	42.5203	0.0036*	21	-0.0054
22	-0.0081	42.7696	0.0050*	22	-0.0089
23	0.0020	42.7848	0.0073*	23	0.0009
24	0.0202	44.3518	0.0069*	24	0.0150
25	0.0480	53.1899	0.0008*	25	0.0481

Figure 3.10 MA(2) Analysis

Interpretation:

- Accuracy measures like **MAPE, MAD and MSD** are used to compare the fitting of different time series models. Smaller values indicate a better fit.
- The model summary indicates that, MAE = 0.733
- Here also the R-Square and R-Square adjusted values observed to be 0.0705 and positive. Of all the proposed models, ARIMA MA (2) best fitted model.
- There observed to be no lags, this clearly evident that, this model is fitted with less error and more accuracy.

Forecast:



	Actual change	Date	Predicted change	Std Err Pred change	Residual change	Upper CL (0.95) change	Lower CL (0.95) change
3811	0.69	2017-10-05	-0.078307191	1.0159474404	0.7683071909	1.9129132025	-2.069527584
3812	•	2017-10-06	0.2154158321	1.0159474404	•	2.2066362255	-1.775804561
3813	•	2017-10-07	0.0135664605	1.053776518	•	2.0789304836	-2.051797563
3814	•	2017-10-08	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3815	•	2017-10-09	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3816	•	2017-10-10	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3817	•	2017-10-11	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3818	•	2017-10-12	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3819	•	2017-10-13	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3820	•	2017-10-14	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3821	•	2017-10-15	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3822	•	2017-10-16	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3823	•	2017-10-17	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3824	•	2017-10-18	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3825	•	2017-10-19	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3826	•	2017-10-20	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3827	•	2017-10-21	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3828	•	2017-10-22	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3829	•	2017-10-23	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3830	•	2017-10-24	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3831	•	2017-10-25	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3832	•	2017-10-26	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3833	•	2017-10-27	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3834	•	2017-10-28	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3835	•	2017-10-29	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3836	•	2017-10-30	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3837	•	2017-10-31	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3838	•	2017-11-01	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3839	•	2017-11-02	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3840	•	2017-11-03	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3841	•	2017-11-04	0.006562291	1.0538172185	•	2.0720060854	-2.058881504
3842	•	2017-11-05	0.006562291	1.0538172185	•	2.0720060854	-2.058881504

Figure 3.11 Forecasted data

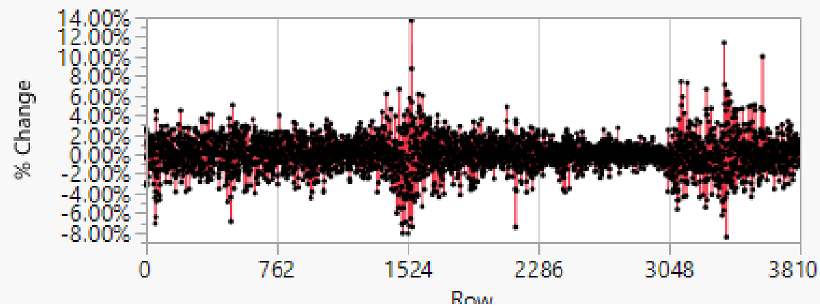
The above table is forecasted data of 30 observations based on the best fitted MA(2)model.

4. PERCENTAGE OF CHANGE DATASET:

4.1 BASIC DIAGNOSTICS:

The time series model of the percentage change in oil prices is given by:

Time Series % Change



Mean	0.0002985
Std	0.0168351
N	3810
Zero Mean ADF	-47.70698
Single Mean ADF	-47.71207
Trend ADF	-47.72537

Time Series Basic Diagnostics

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.2518	241.739	<.0001*	1	0.2518
2	0.0190	243.111	<.0001*	2	-0.0474
3	0.0336	247.412	<.0001*	3	0.0434
4	0.0205	249.019	<.0001*	4	0.0019
5	0.0040	249.081	<.0001*	5	-0.0008
6	-0.0278	252.029	<.0001*	6	-0.0310
7	0.0123	252.603	<.0001*	7	0.0286
8	-0.0204	254.194	<.0001*	8	-0.0346
9	-0.0014	254.201	<.0001*	9	0.0163
10	0.0170	255.301	<.0001*	10	0.0125
11	0.0060	255.441	<.0001*	11	-0.0001
12	0.0151	256.313	<.0001*	12	0.0143
13	0.0353	261.092	<.0001*	13	0.0308
14	0.0547	272.519	<.0001*	14	0.0381
15	0.0429	279.569	<.0001*	15	0.0226
16	0.0380	285.106	<.0001*	16	0.0230
17	-0.0156	286.043	<.0001*	17	-0.0364
18	-0.0178	287.256	<.0001*	18	-0.0051
19	0.0106	287.687	<.0001*	19	0.0149
20	0.0082	287.946	<.0001*	20	0.0037
21	-0.0270	290.741	<.0001*	21	-0.0295
22	-0.0218	292.558	<.0001*	22	-0.0051
23	0.0073	292.761	<.0001*	23	0.0102
24	0.0277	295.695	<.0001*	24	0.0260
25	0.0603	309.658	<.0001*	25	0.0510
26	0.0437	317.001	<.0001*	26	0.0145
27	0.0017	317.012	<.0001*	27	-0.0172
28	0.0035	317.060	<.0001*	28	0.0024
29	0.0295	320.400	<.0001*	29	0.0222
30	0.0057	320.524	<.0001*	30	-0.0106

Figure 4.1 Time series Analysis

Interpretation:

- By analyzing the time series of this data, we find that the data is stationary. The data is distributed evenly around the mean line. Most of the values lie within interval of +4% and -4%. There are very few values outside of this interval.
- We find that the data is non-seasonal and non-cyclic because there is no repeating pattern in the data. There is no trend in this data because the data is distributed along a flat horizontal line.
- Since the data is stationary, we will try to fit the following models to this dataset:

4.2 MODELS

4.2.1 Simple Exponential Smoothing:

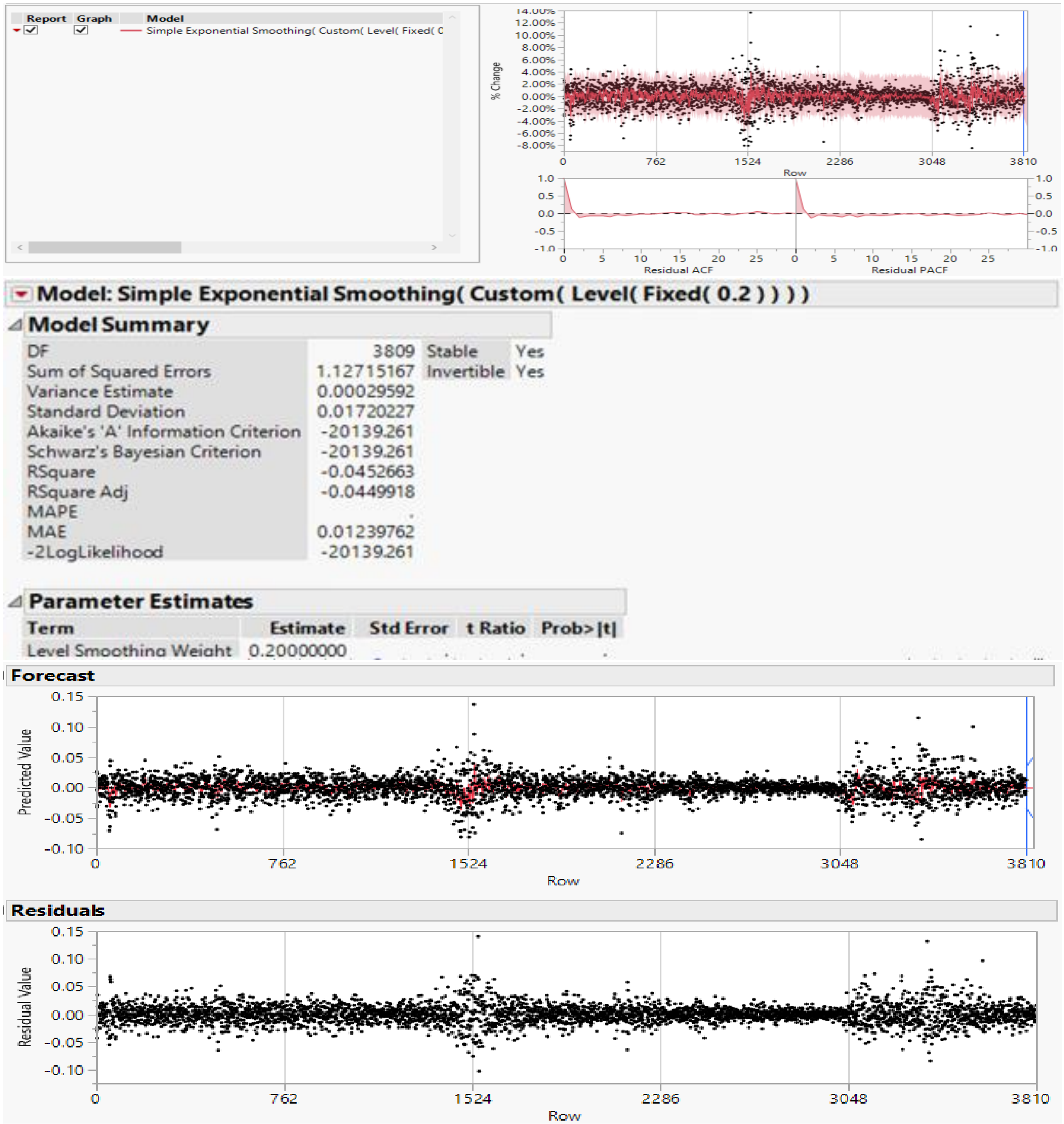


Figure 4.2 Simple Exponential Smoothing Analysis

Interpretation:

- This model has MAE value of 0.0123 and standard deviation of 0.0172.
- This model has RSquare value of -0.0452 and RSquare Adjusted value of -0.0449.
- Since the RSquare value is negative, this model is unstable and not a good fit for this data.

4.2.2 Double Exponential Smoothing:

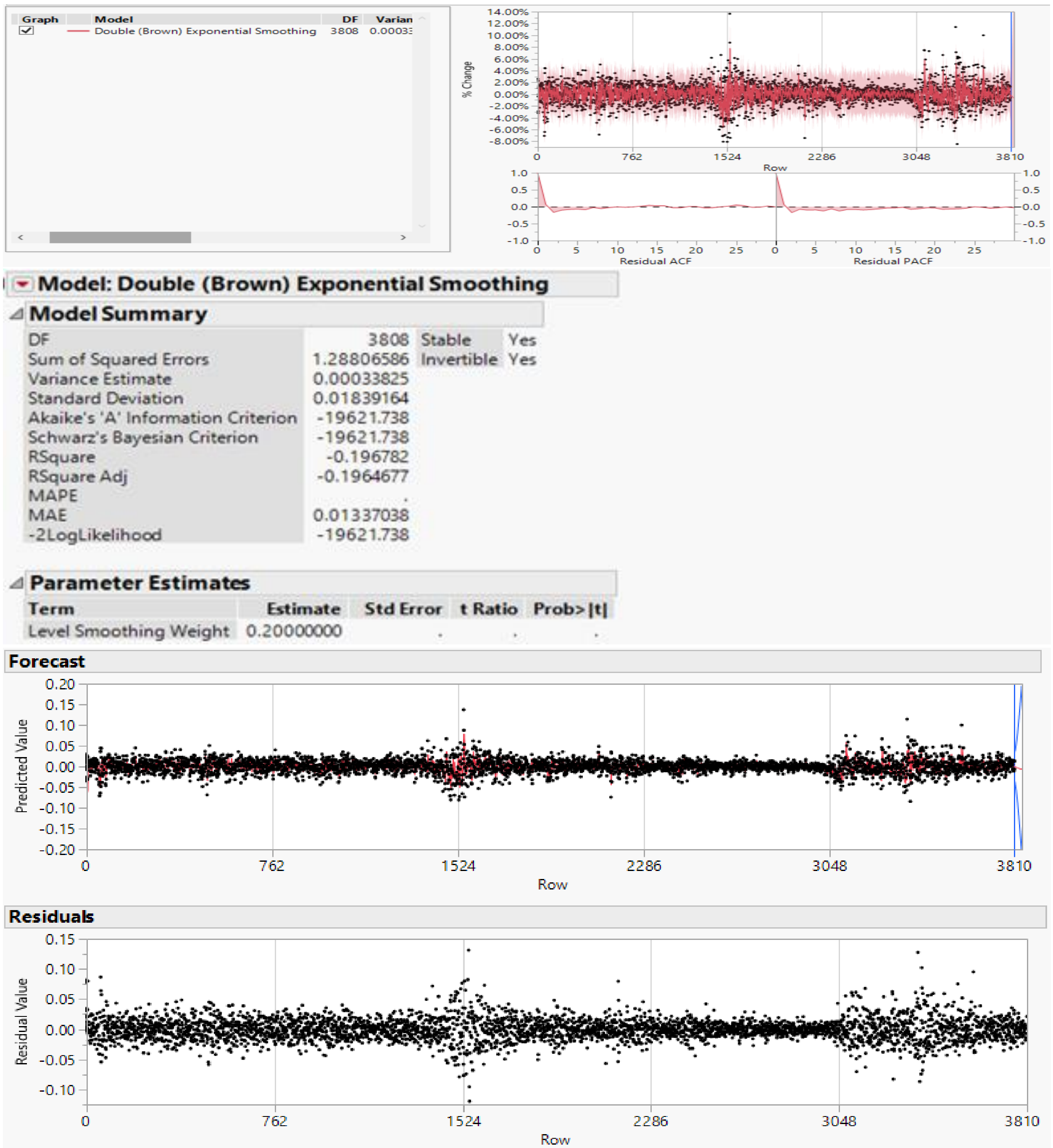


Figure 4.3 Double Exponential Smoothing Analysis

Interpretation:

- This model has MAE value of 0.0133 and standard deviation of 0.0183.
- This model has R-Square value of -0.1967 and R-Square Adjusted value of -0.1964.
- Since the R-Square value is negative, this model is unstable and not a good fit for this data.

4.2.3 Linear Exponential Smoothing:

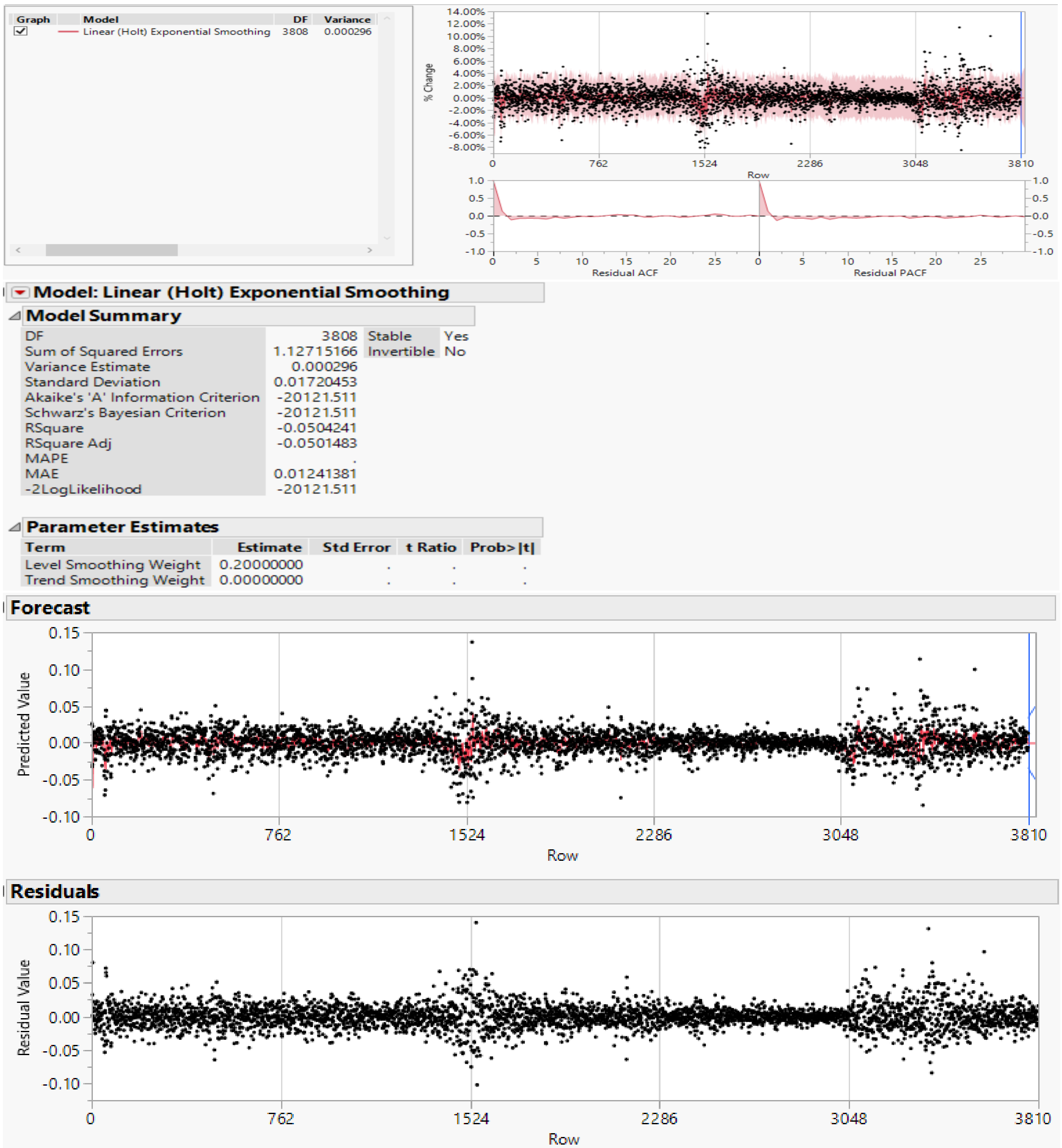


Figure 4.4 Linear Exponential Smoothing Analysis

Interpretation:

This model has MAE value of 0.0124 and standard deviation of 0.0172.
This model has R-Square value of -0.0504 and R-Square Adjusted value of -0.0501.
Since the R-Square value is negative, this model is unstable and not a good fit for this data.

4.2.4 ARIMA – I (1) Model:

By analyzing the initial time series model, we see that there is one spike in ACF and PACF lags. Therefore, we try the I (1) model for this dataset:

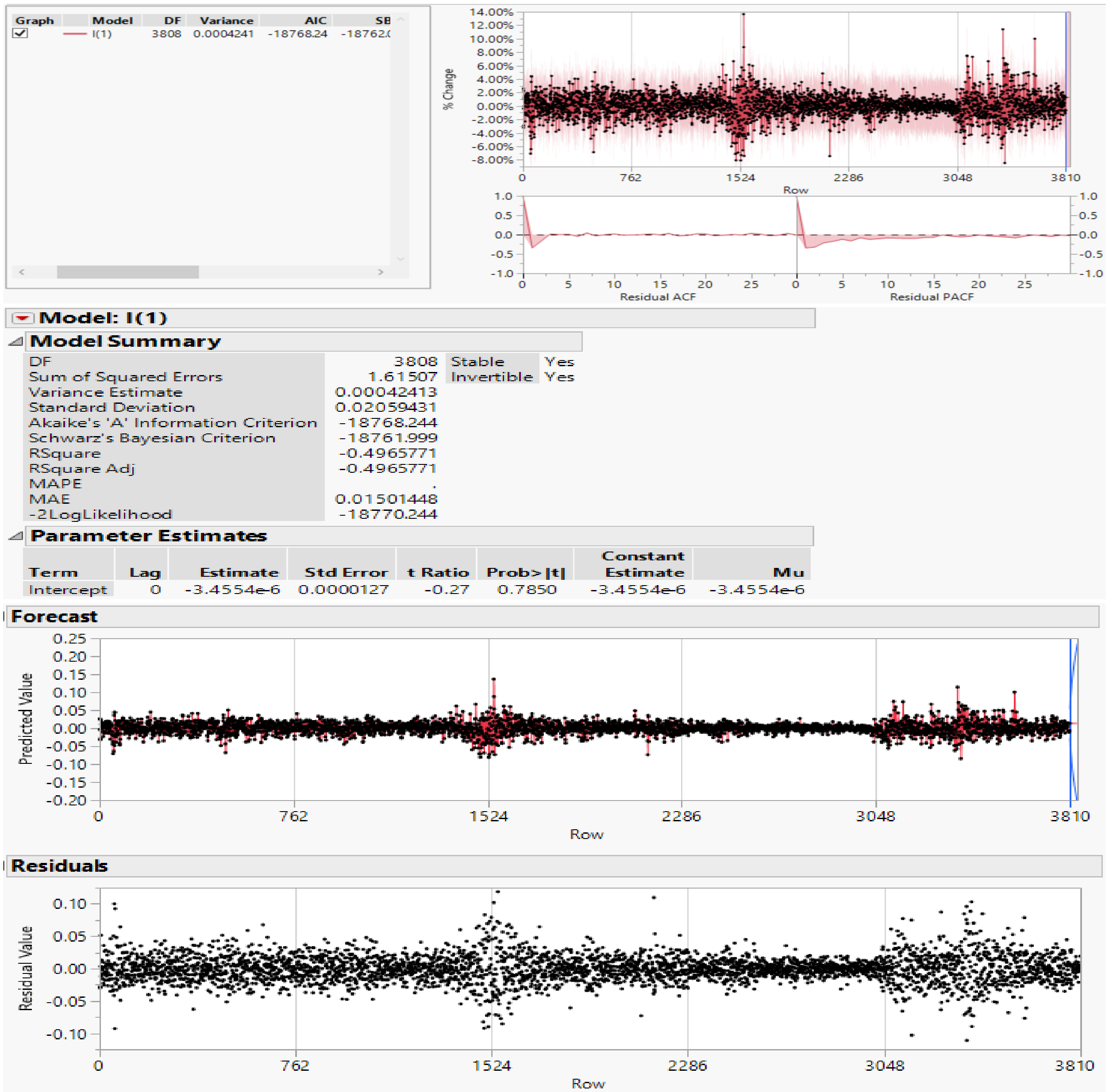


Figure 4.5 ARIMA (0,1, 0) Model Analysis

Interpretation:

- This model has MAE value of 0.0150 and standard deviation of 0.0205.
- This model has R-Square value of -0.4965 and R-Square Adjusted value of -0.4965.
- Since the R-Square value is negative, this model is unstable and not a good fit for this data.

By analyzing the ACF and PACF lags of the residuals of I (1) model, we get:

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	-0.3440	451.003	<.0001*	1	-0.3440
2	-0.1650	554.796	<.0001*	2	-0.3213
3	0.0185	556.100	<.0001*	3	-0.2112
4	0.0014	556.108	<.0001*	4	-0.1727
5	0.0105	556.527	<.0001*	5	-0.1212
6	-0.0479	565.288	<.0001*	6	-0.1603
7	0.0482	574.174	<.0001*	7	-0.0830
8	-0.0346	578.740	<.0001*	8	-0.1233
9	0.0006	578.741	<.0001*	9	-0.1062
10	0.0196	580.204	<.0001*	10	-0.0846
11	-0.0133	580.883	<.0001*	11	-0.0910
12	-0.0072	581.081	<.0001*	12	-0.0984
13	0.0005	581.082	<.0001*	13	-0.0971
14	0.0211	582.789	<.0001*	14	-0.0741
15	-0.0052	582.894	<.0001*	15	-0.0704
16	0.0330	587.063	<.0001*	16	-0.0089
17	-0.0344	591.592	<.0001*	17	-0.0401
18	-0.0207	593.234	<.0001*	18	-0.0578
19	0.0206	594.856	<.0001*	19	-0.0439
20	0.0220	596.706	<.0001*	20	-0.0100
21	-0.0269	599.471	<.0001*	21	-0.0342
22	-0.0159	600.441	<.0001*	22	-0.0474
23	0.0053	600.550	<.0001*	23	-0.0608
24	-0.0080	600.795	<.0001*	24	-0.0815
25	0.0329	604.939	<.0001*	25	-0.0421
26	0.0172	606.079	<.0001*	26	-0.0095
27	-0.0293	609.365	<.0001*	27	-0.0288
28	-0.0161	610.364	<.0001*	28	-0.0471
29	0.0329	614.518	<.0001*	29	-0.0137
30	-0.0094	614.857	<.0001*	30	-0.0271

Figure 4.6 ACF & PACF Analysis

Comparing these values with the ACF and PACF layouts of various ARIMA models, we get:

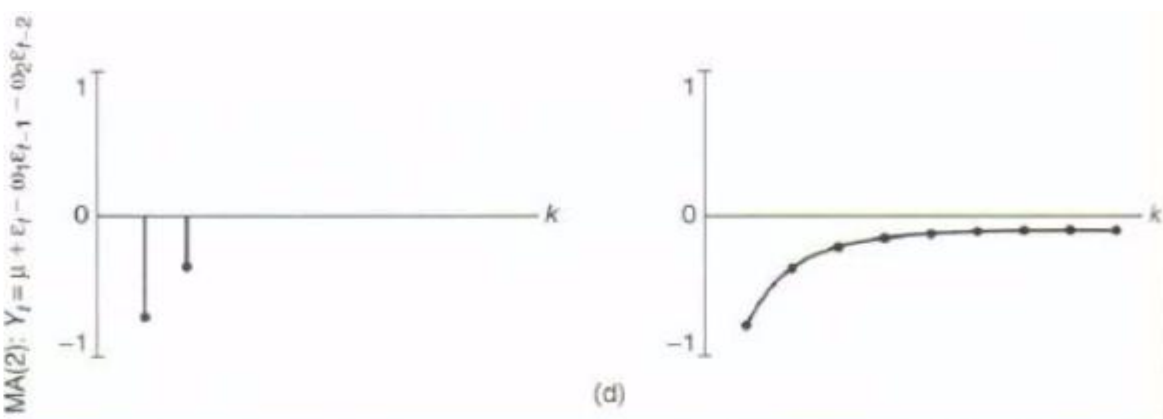
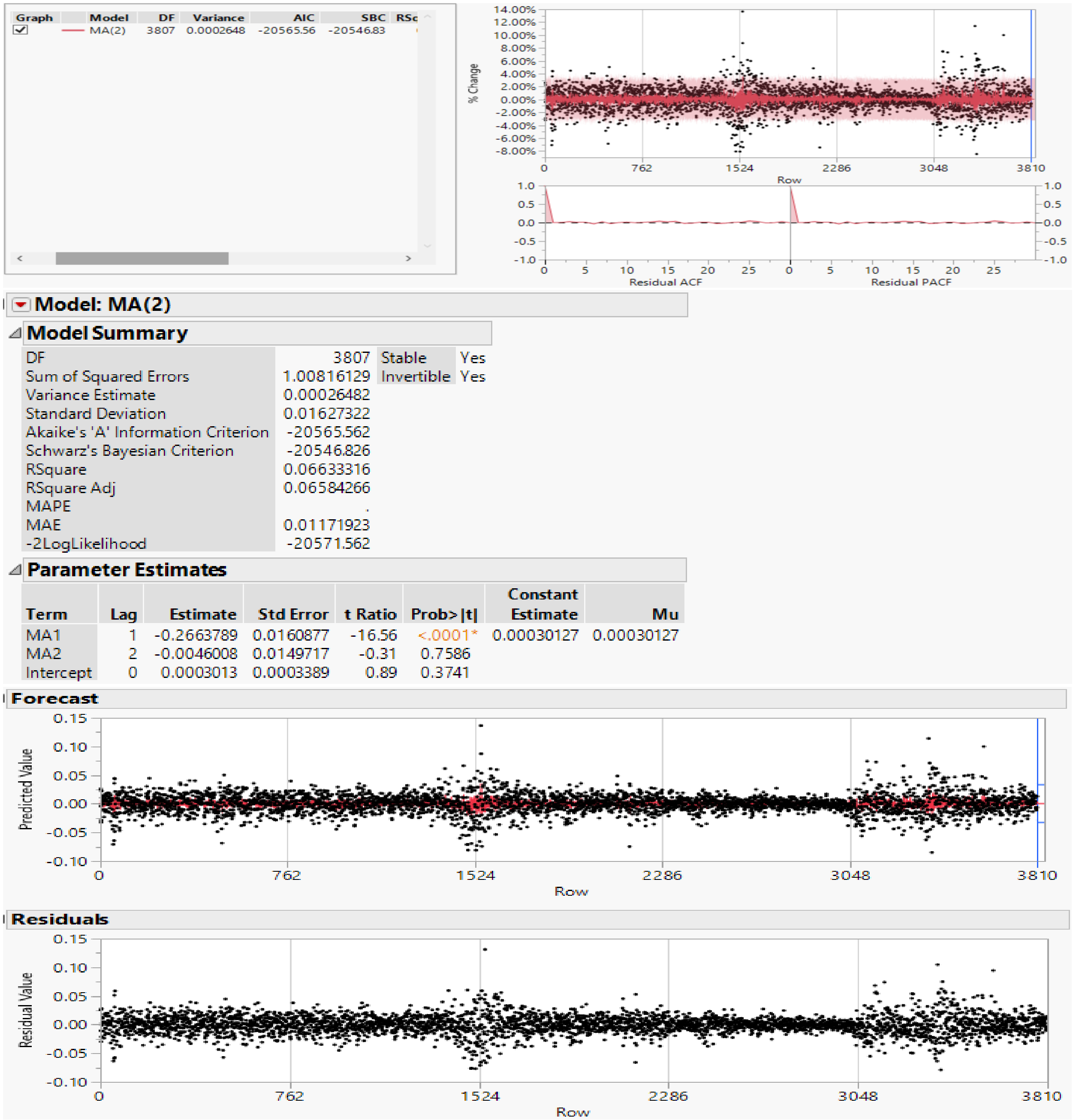


FIGURE 9-3 Autocorrelation and Partial Autocorrelation Coefficients of MA(1) and MA(2) Models

We find that the ACF and PACF layouts are similar to that of the MA (2) model.

Therefore, we will try to fit the MA (2) model to this dataset.

ARIMA – MA (2) Model:



Graph	Model	DF	Variance	AIC	SBC	RSq
☑	MA(2)	3807	0.0002648	-20565.56	-20546.83	

Model: MA(2)

Model Summary

DF	3807	Stable	Yes
Sum of Squared Errors	1.00816129	Invertible	Yes
Variance Estimate	0.00026482		
Standard Deviation	0.01627322		
Akaike's 'A' Information Criterion	-20565.562		
Schwarz's Bayesian Criterion	-20546.826		
RSquare	0.06633316		
RSquare Adj	0.06584266		
MAPE	.		
MAE	0.01171923		
-2LogLikelihood	-20571.562		

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate	Mu
MA1	1	-0.2663789	0.0160877	-16.56	<.0001*	0.00030127	0.00030127
MA2	2	-0.0046008	0.0149717	-0.31	0.7586		
Intercept	0	0.0003013	0.0003389	0.89	0.3741		

Figure 4.7 ARIMA MA(2) Model Analysis

Interpretation:

- This model has MAE value of 0.0117 and standard deviation of 0.0162.
- This model has R-Square value of 0.0663 and R-Square Adjusted value of 0.0658.
- Since the R-Square value is positive, this model is more stable and a good fit for this data.
-

By analyzing the ACF and PACF lags of the residuals of MA (2) model, we get:

Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.0000		0.0000	0.9991	1	0.0000	
2	0.0073		0.2056	0.9023	2	0.0073	
3	0.0290		3.4142	0.3321	3	0.0290	
4	0.0108		3.8556	0.4259	4	0.0107	
5	0.0104		4.2723	0.5109	5	0.0100	
6	-0.0375		9.6322	0.1410	6	-0.0385	
7	0.0285		12.7278	0.0790	7	0.0277	
8	-0.0277		15.6567	0.0476*	8	-0.0280	
9	0.0012		15.6625	0.0743	9	0.0029	
10	0.0169		16.7553	0.0800	10	0.0164	
11	-0.0006		16.7568	0.1153	11	0.0011	
12	0.0095		17.1021	0.1458	12	0.0078	
13	0.0221		18.9630	0.1243	13	0.0238	
14	0.0433		26.1293	0.0249*	14	0.0400	
15	0.0227		28.1032	0.0209*	15	0.0233	
16	0.0376		33.5104	0.0063*	16	0.0362	
17	-0.0212		35.2347	0.0058*	17	-0.0256	
18	-0.0156		36.1678	0.0067*	18	-0.0173	
19	0.0117		36.6927	0.0087*	19	0.0097	
20	0.0118		37.2306	0.0110*	20	0.0143	
21	-0.0257		39.7564	0.0080*	21	-0.0247	
22	-0.0174		40.9230	0.0084*	22	-0.0141	
23	0.0082		41.1825	0.0113*	23	0.0045	
24	0.0134		41.8727	0.0133*	24	0.0153	
25	0.0487		50.9817	0.0016*	25	0.0497	
26	0.0331		55.1836	0.0007*	26	0.0311	
27	-0.0063		55.3349	0.0010*	27	-0.0110	
28	-0.0027		55.3625	0.0015*	28	-0.0086	
29	0.0306		58.9493	0.0008*	29	0.0242	
30	-0.0015		58.9575	0.0012*	30	-0.0054	

Figure 4.8 ACF & PACF Analysis

We see this model has eliminated the spikes in the ACF and PACF lags.
Therefore, this model is a good fit for this dataset.

By comparing the different models for this dataset, we get:

Model Comparison

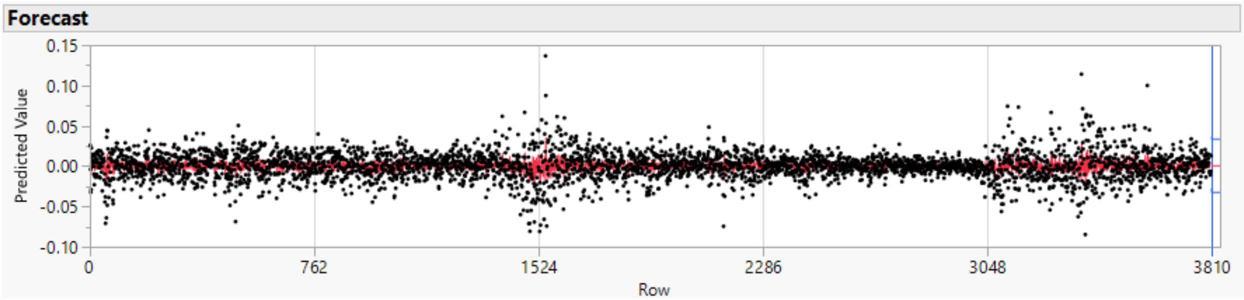
Model	DF	Variance	RSquare	MAPE	MAE
MA(2)	3837	0.0002636	0.066	.	0.011696
Simple Exponential Smoothing(Custom(Level(Fixed(0.2))))	3839	0.0002946	-0.05	.	0.012373
Linear (Holt) Exponential Smoothing	3838	0.0002947	-0.05	.	0.012389
Double (Brown) Exponential Smoothing	3838	0.0003367	-0.20	.	0.013339
I(1)	3838	0.0004223	-0.50	.	0.014989

Figure 4.9 Comparison Model

Interpretation:

- We observed that MA (2) is the best model for this dataset.
- It has the largest R-Square value of 0.066.
- It has the lowest MAE value of 0.011.
- It has also eliminated all the spikes from the ACF and PACF lags.

Therefore,
We will use the MA (2) model for predicting the percentage change in prices.
Using this MA (2) model to predict the last 30 values of the dataset, we get:



Actual % Change	Predicted % Change	Standard Error	Residuals
-0.27%	0.40%	1.63%	-0.68%
-1.41%	0.04%	1.68%	-1.45%
0.99%	0.03%	1.68%	0.96%
0.57%	0.03%	1.68%	0.54%
-0.18%	0.03%	1.68%	-0.21%
1.25%	0.03%	1.68%	1.22%
1.12%	0.03%	1.68%	1.09%
0.16%	0.03%	1.68%	0.13%
0.52%	0.03%	1.68%	0.49%
-1.07%	0.03%	1.68%	-1.10%
-0.20%	0.03%	1.68%	-0.23%
0.65%	0.03%	1.68%	0.62%
-0.32%	0.03%	1.68%	-0.35%
1.33%	0.03%	1.68%	1.30%
0.21%	0.03%	1.68%	0.18%
1.95%	0.03%	1.68%	1.92%
1.25%	0.03%	1.68%	1.22%
0.45%	0.03%	1.68%	0.42%
1.16%	0.03%	1.68%	1.13%
-1.28%	0.03%	1.68%	-1.31%
1.15%	0.03%	1.68%	1.12%
3.21%	0.03%	1.68%	3.18%
1.72%	0.03%	1.68%	1.69%
-0.74%	0.03%	1.68%	-0.77%
0.15%	0.03%	1.68%	0.12%
0.34%	0.03%	1.68%	0.31%
-1.03%	0.03%	1.68%	-1.06%
-1.22%	0.03%	1.68%	-1.25%
-1.21%	0.03%	1.68%	-1.24%
0.32%	0.03%	1.68%	0.29%

Figure 4.10 Forecasted data

Forecasted values for 30 observations of percentage change in oil prices using best fitted MA(2) model is shown above.

5. CUMULATIVE DATASET:

5.1. BASIC DIAGNOSTICS:

Analyzing the cumulative data is quite tricky because a significant drop or increase in the original price cannot be observed in the short-term analysis of the accumulative graph. For example, the original prices experienced a drastic drop from 140.73 to 33.36 between 7/3/2008 and 12/24/2008. This can be clearly seen in the price graph (figure 5.1). On the contrary, this huge drop needs time to be identified in the cumulative graph. Thus, using cumulative data solely to study the behavior of a time series is not favored.

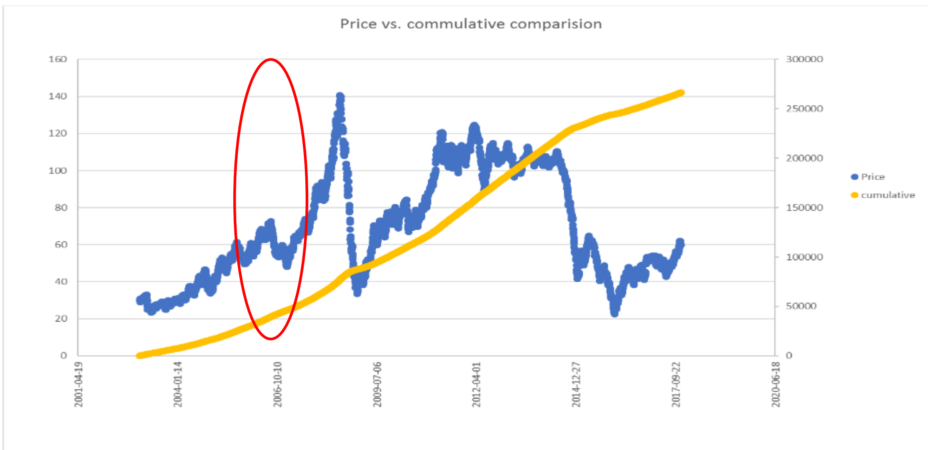


Figure 5.1. The behavior of the cumulative vs the original price

The cumulative data shows an increasing trend. It is, also, highly autocorrelated by nature. The basic diagnostics in figure 5.2 demonstrates that as all the readings in the autocorrelation function exceed the 5% significance limit while the partial autocorrelation function cuts off after lag 1. Thus, the perfect move would be to take the first difference. This is going to be discussed in detail in section 5.5.

Four models are examined to predict the cumulative which are Linear Regression, Simple Exponential, Double Exponential and ARIMA.

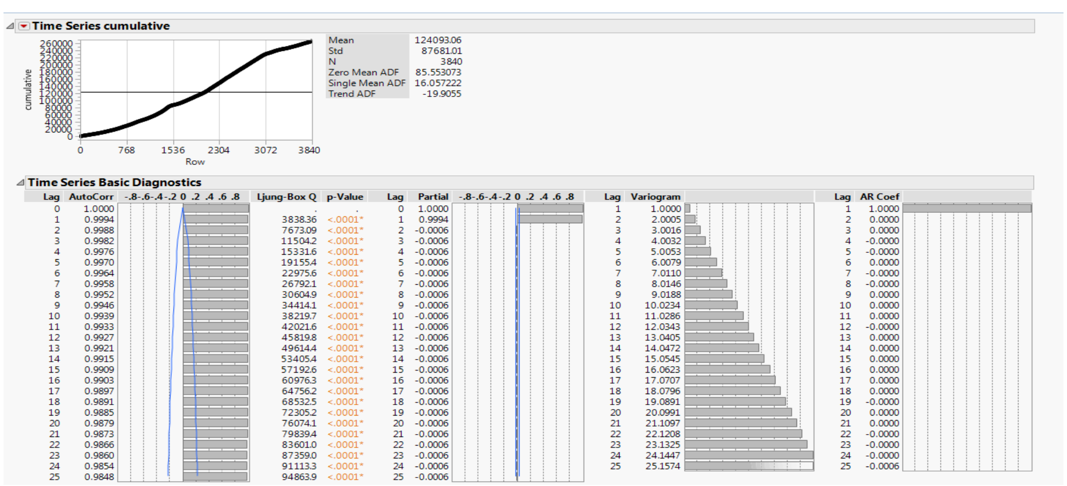


Figure 5.2. ACF and PACF

5.2 MODELS:

5.2.1. Linear Regression

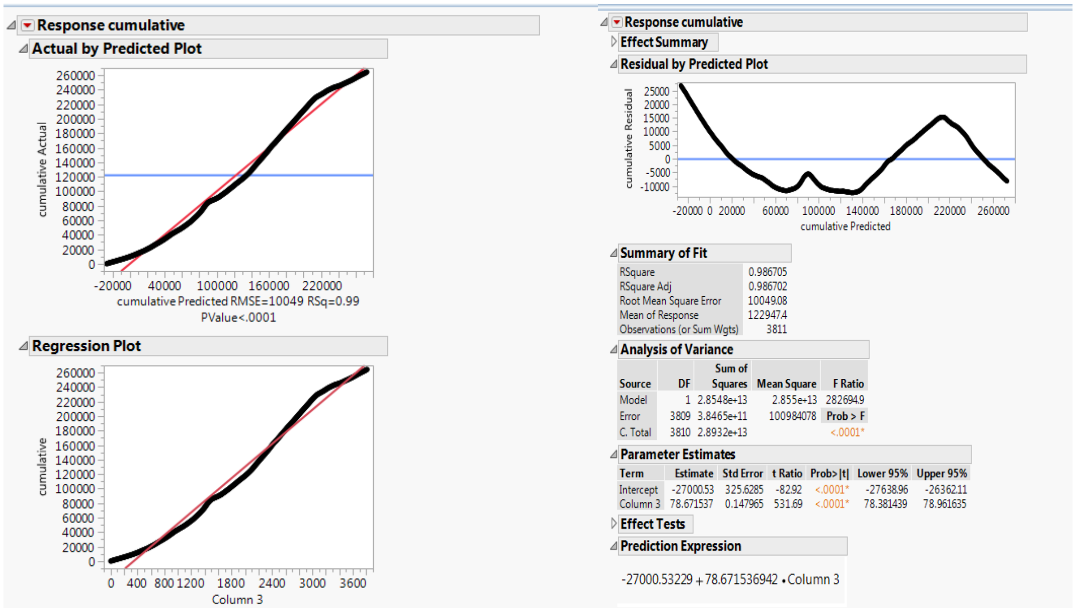


Figure 5.3. Linear Regression Model Statistics

The first model assumes that the data accepts a linear relationship with time. The linear regression fit statistics are found to be good with 0.986 R^2-adjusted value, high F-Ratio of 282694.9 and high t Ratios (figure 5.3). Also, the 95% confidence and prediction intervals are quite good as the range is not that large between the upper and the lower 95%. However, the residual analysis shows violation of the regression model assumptions. The residuals are not normally distributed, the errors variation is inconstant and the residual strangely shows a trend. (figure 5.4)

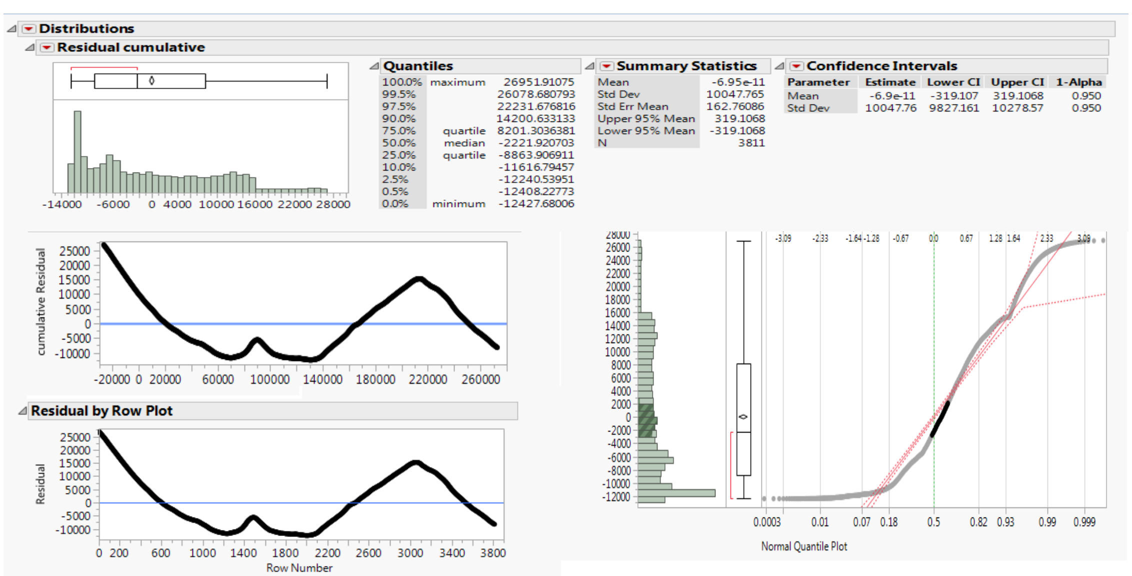


Figure 5.4. Residual Plots

In order to investigate potential outliers, R-student was considered and all data proved to be within the range. (figure 5.5)

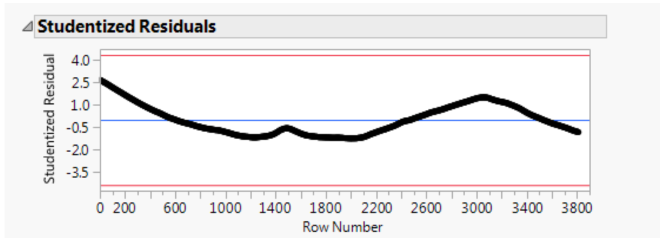


Figure 5.5. Studentized Residuals

To test the performance of the model in prediction the cumulative values, the last 30 values were predicted. As shown in table 5.1, the sum of squared error is too large even though the percentage of errors averages around 3%. This supports the initial analysis that using cumulative is sometimes misleading.

Table 5.1. Forecasting the last 30 values using the Linear Regression Model

Linear Regression Model (SS_E = 2,123,002,651)				
Forecasted	Actual	Error	Squared Error	% Error
272974.0381	264728.6	8245.488	67988073.51	3.11%
273052.7096	264782.8	8269.93	68391735.7	3.12%
273131.3811	264837.3	8294.061	68791450.26	3.13%
273210.0527	264891.8	8318.293	69193993.12	3.14%
273288.7242	264946.9	8341.844	69586364.95	3.15%
273367.3958	265002.6	8364.776	69969473.42	3.16%
273446.0673	265058.5	8387.617	70352123.83	3.16%
273524.7388	265114.6	8410.169	70730939.72	3.17%
273603.4104	265170.1	8433.32	71120892.39	3.18%
273682.0819	265225.5	8456.582	71513777.47	3.19%
273760.7534	265281.3	8479.483	71901639.4	3.20%
273839.425	265336.9	8502.565	72293611.17	3.20%
273918.0965	265393.2	8524.907	72674031.06	3.21%
273996.7681	265449.6	8547.128	73053397.9	3.22%
274075.4396	265507.2	8568.25	73414900.99	3.23%

274154.1111	265565.5	8588.651	73764928.13	3.23%
274232.7827	265624	8608.793	74111311.08	3.24%
274311.4542	265683.2	8628.254	74446770.5	3.25%
274390.1257	265741.7	8648.476	74796132.54	3.25%
274468.7973	265800.8	8668.027	75134696.78	3.26%
274547.4688	265861.8	8685.679	75441016.37	3.27%
274626.1403	265923.9	8702.28	75729683.21	3.27%
274704.8119	265985.5	8719.342	76026922.87	3.28%
274783.4834	266047.2	8736.313	76323172.16	3.28%
274862.155	266109.1	8753.075	76616321.19	3.29%
274940.8265	266170.4	8770.476	76921257.92	3.30%
275019.498	266230.9	8788.628	77239982.66	3.30%
275098.1696	266290.7	8807.51	77572224.78	3.31%
275176.8411	266350.6	8826.201	77901825.93	3.31%

Another scenario is considered which is removing the dataset before the sudden drop in 7/3/2008 and forecasting 90 values into the future. The model is shown in figure 5.6.

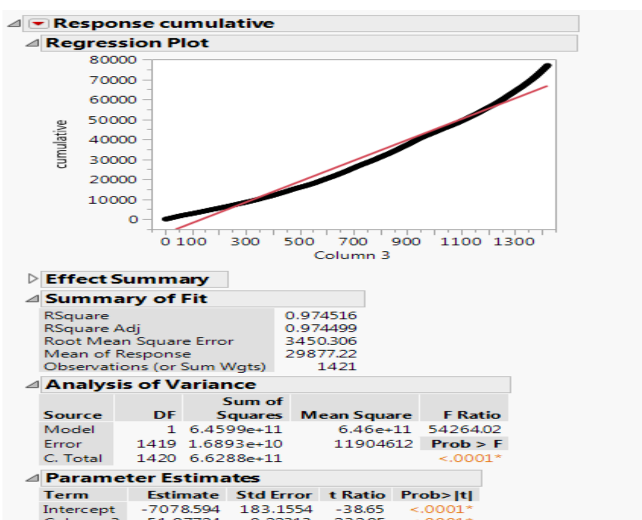


Figure 5.6. Linear Regression Model (7/3/2008)

In figure 5.7, the analysis shows that the percentage of error slightly grows from 1% to only 4.6% even though the actual price dropped by 76%. The point here again is that a sudden drop cannot be observed using the cumulative data as the cumulative is not sensitive enough to such a change.

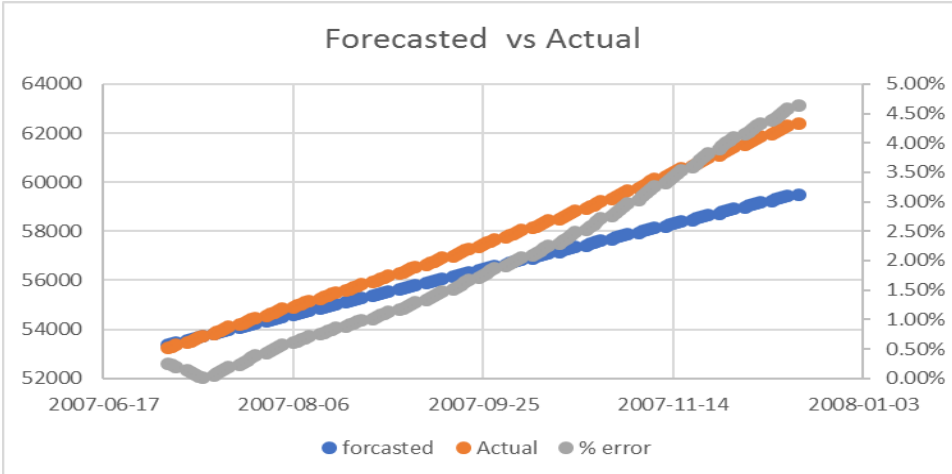


Figure 5.7. Forecasting using Linear Regression Model (7/3/2008)

5.2.2. Simple Exponential Smoothing

Several simple exponential models were tried to find the optimum exponential factor λ. However, given that the data is highly correlated, it was expected to see the sum of squared error keeps decreasing with increasing λ as shown in figure 5.8. Also, the simple exponential does not do well when the data has trends. Therefore, the simple exponential smoothing was found inadequate.

Model Comparison

Report	Graph	Model	DF	Variance	AIC	SBC	RSquare	-2LogLH	Weights	.2 .4 .6 .8	MAPE	MAE
✓		Simple Exponential Smoothing(Zero to One)	3809	5654.8797	43732.756	43739.001	1.000	43730.756	1.000000		0.227171	69.446281
✓		Simple Exponential Smoothing(Custom(Level(Fixed(0.9))))	3810	6979.2137	44533.456	44533.456	1.000	44533.456	0.000000		0.250838	77.160758
✓		Simple Exponential Smoothing(Custom(Level(Fixed(0.8))))	3810	8832.6301	45430.805	45430.805	1.000	45430.805	0.000000		0.280081	86.803299
✓		Simple Exponential Smoothing(Custom(Level(Fixed(0.6))))	3810	15700.201	47622.530	47622.530	1.000	47622.53	0.000000		0.365646	115.72699
✓		Simple Exponential Smoothing(Custom(Level(Fixed(0.4))))	3810	35315.463	50711.368	50711.368	1.000	50711.368	0.000000		0.529054	173.55625
✓		Simple Exponential Smoothing(Custom(Level(Fixed(0.2))))	3810	141138.89	55990.407	55990.407	1.000	55990.407	0.000000		0.981066	346.89883

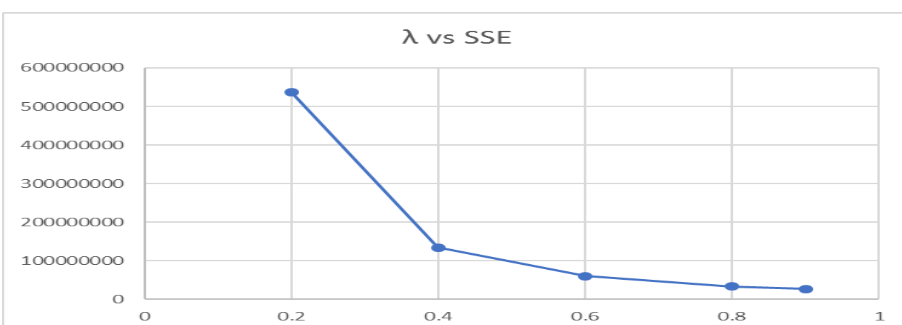


Figure 5.8. Simple Exponential Smoothing - Models Comparison

5.2.3. Double Exponential Smoothing

The double exponential models showed similar behavior to the single exponential models. the SS_E gets lower as λ increases. This also suggest that the double exponential smoothing is not a good fit for the data. The high value of λ is because the data is highly autocorrelated and more weight is placed on recent values. Figure 5.9 shows the comparison between different double exponential models.

Model Comparison

Report	Graph	Model	DF	Variance	AIC	SBC	RSquare	-2LogLH	Weights	.2 .4 .6 .8	MAPE	MAE
✓	✓	Double (Brown) Exponential Smoothing	3808	1.1101094	11208.356	11214.601	1.000	11206.356	1.000000		0.002755	0.758244
✓		Double (Brown) Exponential Smoothing	3809	1.2740478	11732.050	11732.050	1.000	11732.05	0.000000		0.002939	0.812283
✓		Double (Brown) Exponential Smoothing	3809	1.5675904	12521.934	12521.934	1.000	12521.934	0.000000		0.003196	0.904258
✓		Double (Brown) Exponential Smoothing	3809	2.9886006	14980.284	14980.284	1.000	14980.284	0.000000		0.004256	1.264770
✓		Double (Brown) Exponential Smoothing	3809	8.8067629	19097.814	19097.814	1.000	19097.814	0.000000		0.006869	2.204645
✓		Double (Brown) Exponential Smoothing	3809	68.921362	26936.929	26936.929	1.000	26936.929	0.000000		0.016525	6.125195

Figure 5.9. Double Exponential Smoothing - Models Comparison

5.2.4. ARIMA

Recalling the ACF and the PACF in section 5.1, the PACF cuts off after lag 1, suggesting an MA (1) or MA (2), and the ACF exceed the significance limit at all the lags, which can be taken care of by taken the 1st or the 2nd difference. Thus, the ARIMA Model Group function in JMP was used to find the best fit for this data. 36 different models were compared as shown in figure 5.10.

Time Series cumulative

Model Comparison

Report	Graph	Model	DF	Variance	AIC	SBC	RSquare	-2LogLH	Weights	.2 .4 .6 .8	MAPE	MAE
✓	✓	ARIMA(1, 2, 2)	3805	1.0286943	10921.357	10946.337	1.000	10913.357	0.476020		0.002615	0.732688
✓		ARIMA(1, 2, 3)	3804	1.0288813	10923.050	10954.276	1.000	10913.05	0.204075		0.002613	0.732490
		ARIMA(2, 2, 2)	3804	1.0288934	10923.095	10954.320	1.000	10913.095	0.199599		0.002613	0.732519
		ARIMA(2, 2, 3)	3803	1.0290711	10924.751	10962.222	1.000	10912.751	0.087198		0.002614	0.732619
✓		ARIMA(2, 1, 2)	3805	1.0275859	10927.816	10959.043	1.000	10917.816	0.018836		0.017621	0.741790
		ARIMA(2, 1, 3)	3804	1.0277685	10929.494	10966.967	1.000	10917.494	0.008138		0.017620	0.741588
		IMA(2,1)	3807	1.0320856	10931.845	10944.335	1.000	10927.845	0.002513		0.002608	0.733558
		IMA(2,2)	3806	1.0322695	10933.523	10952.259	1.000	10927.523	0.001086		0.002610	0.733757
		ARIMA(1, 2, 1)	3806	1.0322711	10933.529	10952.264	1.000	10927.529	0.001082		0.002610	0.733752
		ARI(2, 2)	3806	1.0325905	10934.706	10953.442	1.000	10928.706	0.000601		0.002616	0.734084
		IMA(2,3)	3805	1.0325319	10935.490	10960.471	1.000	10927.49	0.000406		0.002610	0.733772
		ARIMA(2, 2, 1)	3805	1.0325387	10935.515	10960.496	1.000	10927.515	0.000401		0.002610	0.733760
		ARIMA(1, 1, 1)	3807	1.0317937	10941.313	10960.049	1.000	10935.313	0.000022		0.015359	0.741383
		ARIMA(1, 1, 2)	3806	1.0319681	10942.957	10967.939	1.000	10934.957	0.000010		0.015410	0.741586
		ARIMA(2, 1, 1)	3806	1.0319695	10942.962	10967.944	1.000	10934.962	0.000010		0.015411	0.741582
		ARIMA(1, 1, 3)	3805	1.032233	10944.934	10976.161	1.000	10934.934	0.000004		0.015398	0.741592
		ARI(1, 2)	3807	1.036345	10947.524	10960.014	1.000	10943.524	0.000001		0.002635	0.735944
		ARI(2, 1)	3807	1.0359876	10956.759	10975.495	1.000	10950.759	0.000000		0.015686	0.743871
		I(2)	3808	1.1100704	11208.222	11214.467	1.000	11206.222	0.000000		0.002755	0.758039
		ARI(1, 1)	3808	1.1099383	11218.293	11230.784	1.000	11214.293	0.000000		0.014438	0.765156
		IMA(1,3)	3806	27.467508	23445.676	23470.658	1.000	23437.676	0.000000		0.053675	4.421076
		IMA(1,2)	3807	68.065747	26900.759	26919.495	1.000	26894.759	0.000000		0.076373	7.029353
		IMA(1,1)	3808	215.50089	31288.636	31301.127	1.000	31284.636	0.000000		0.123190	12.741158
		ARMA(1, 2)	3807	355.3785	35462.186	35487.169	0.983	35454.186	0.000000		610.45838	206.70009
		I(1)	3809	830.82762	36425.741	36431.986	1.000	36423.741	0.000000		0.222178	25.116311
		ARMA(1, 1)	3808	1516.6052	38755.996	38774.732	0.996	38749.996	0.000000		293.39134	122.88885
		ARMA(1, 3)	3806	4426.1515	42857.333	42888.561	0.999	42847.333	0.000000		151.71207	97.201707
		AR(1)	3809	5654.8797	43745.234	43757.725	1.000	43741.234	0.000000		0.227111	69.428058
		ARMA(2, 1)	3807	5657.8505	43749.234	43774.216	1.000	43741.234	0.000000		0.227111	69.428058
		AR(2)	3808	22622.407	49029.887	49048.624	1.000	49023.887	0.000000		0.447907	138.84178
		MA(3)	3807	120192377	81751.472	81776.454	0.983	81743.472	0.000000		225.32847	9795.0702
		MA(2)	3808	476027645	86985.885	87004.622	0.936	86979.885	0.000000		307.96349	19413.712
		MA(1)	3809	1.9001e+9	92248.067	92260.558	0.749	92244.067	0.000000		467.90500	38668.663
		ARIMA(0, 0, 0)	3810	7.5938e+9	97518.672	97524.917	-0.00	97516.672	0.000000		803.79029	77196.714
		ARMA(2, 2)	3806	7.6645e+9	97557.980	97589.208	-0.01	97547.98	0.000000		751.69779	76754.242
		ARMA(2, 3)	3805	7.6741e+9	97563.763	97601.237	-0.01	97551.763	0.000000		752.00065	76792.660

Figure 5.10. ARIMA - Models Comparison

JMP suggested ARIMA (1,2,2) as the best fit since it has the lowest MAE. It also has high t-statistics except for the intercept. The residuals are stationary and within the 5% limit. The summary is shown in figure 5.11.

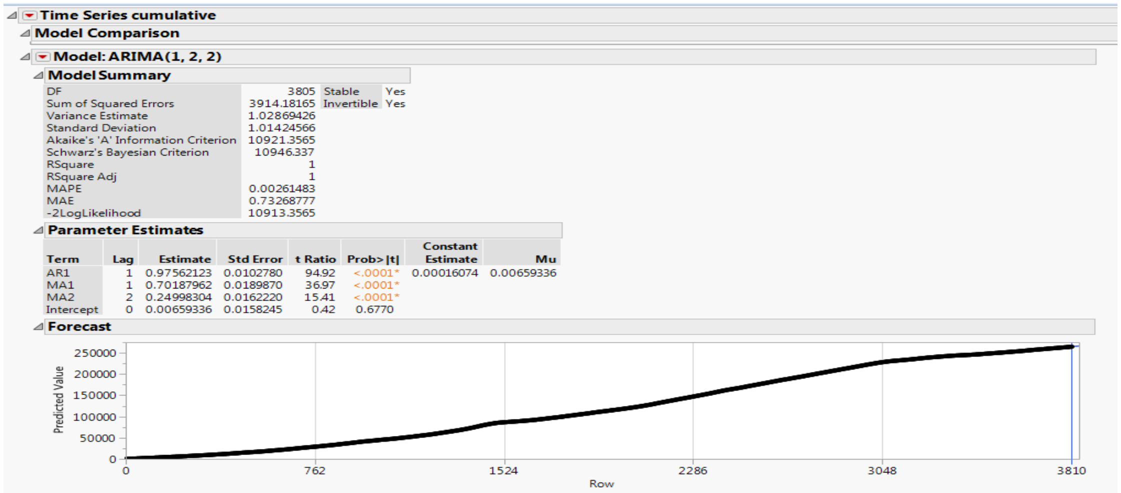


Figure 5.11. ARIMA (1,2,2)

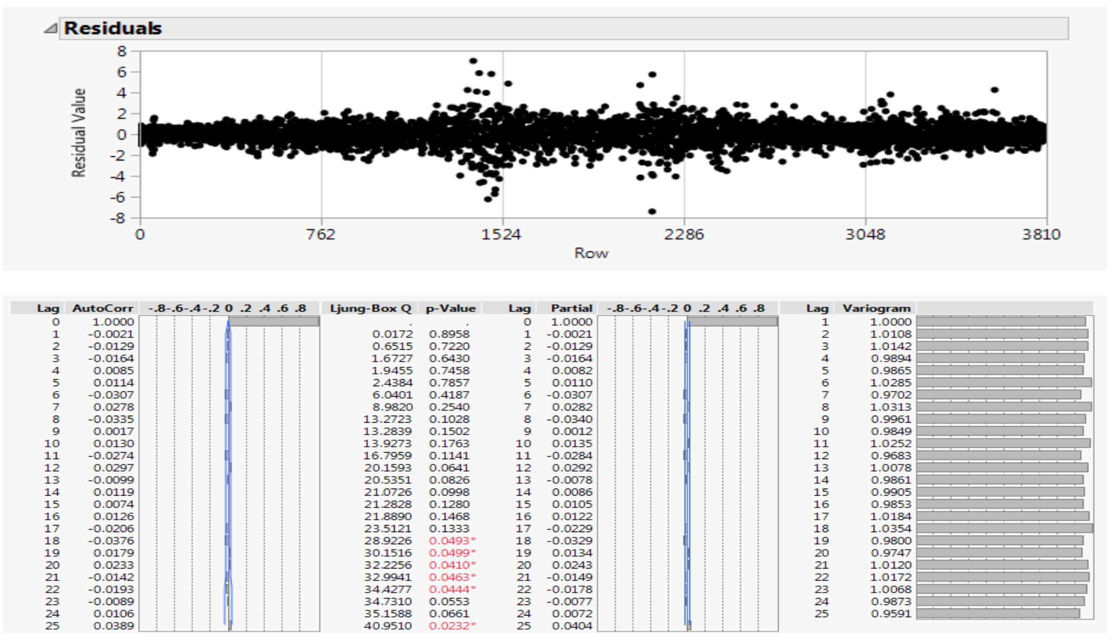


Figure 5.12. ARIMA (1,2,2) - Residual

The last 30 values were forecasted using ARIMA (1,2,2). The results are shown in table 5.2 and figure 5.13.

Table 5.2. Forecasting the last 30 values using ARIMA (1,2,2)

ARIMA (1,2,2) -- (SS_E = 24131.35)					
Actual	Forecasted	Error	Squared Error	Upper CL	Lower CL
264674.9	264675.2	-0.38586	0.14888794	264677.2	264673.2
264728.6	264730.1	-1.58646	2.516855333	264735.1	264725.2

264782.8	264785.1	-2.30087	5.294002759	264793.7	264776.4
264837.3	264840.1	-2.74818	7.552493315	264853	264827.1
264891.8	264895.1	-3.33751	11.138973	264913	264877.2
264946.9	264950.2	-3.288	10.810944	264973.5	264926.8
265002.6	265005.3	-2.6588	7.069217442	265034.6	264976
265058.5	265060.4	-1.97909	3.91679723	265096.1	265024.8
265114.6	265115.6	-1.04807	1.098450726	265158.1	265073.1
265170.1	265170.8	-0.75496	0.569964602	265220.7	265121
265225.5	265226.1	-0.60899	0.370868821	265283.6	265168.6
265281.3	265281.4	-0.13942	0.019437937	265347	265215.8
265336.9	265336.7	0.11448	0.01310567	265410.8	265262.6
265393.2	265392.1	1.07342	1.152230496	265475.1	265309.1
265449.6	265447.5	2.11808	4.486262885	265539.7	265355.3
265507.2	265503	4.22915	17.88570972	265604.8	265401.1
265565.5	265558.4	7.02729	49.38280474	265670.2	265446.6
265624	265613.9	10.05312	101.0652217	265736.1	265491.8
265683.2	265669.5	13.72729	188.4384907	265802.3	265536.7
265741.7	265725	16.6104	275.9053882	265868.8	265581.2
265800.8	265780.6	20.13305	405.3397023	265935.8	265625.5
265861.8	265836.3	25.52582	651.5674867	266003.1	265669.5
265923.9	265891.9	31.93927	1020.116968	266070.7	265713.1
265985.5	265947.6	37.86397	1433.680224	266138.7	265756.5
266047.2	266003.3	43.85045	1922.861965	266207.1	265799.6
266109.1	266059.1	50.01924	2501.92437	266275.8	265842.3
266170.4	266114.8	55.52086	3082.565895	266344.8	265884.8
266230.9	266170.6	60.2458	3629.556418	266414.2	265927.1
266290.7	266226.4	64.21455	4123.508432	266483.9	265969
266350.6	266282.3	68.3476	4671.394426	266553.9	266010.7

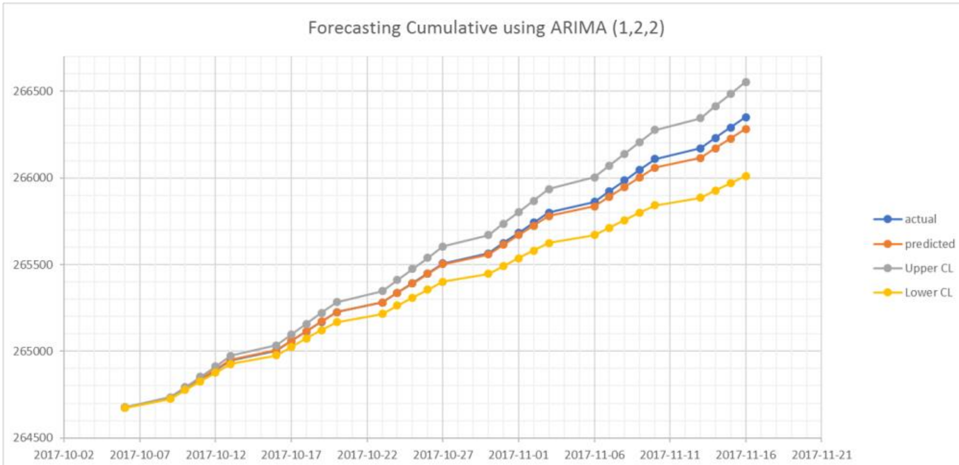


Figure 5.13. Forecasting using ARIMA (1,2,2)

Interpretation:

Out of the four forecasting techniques that we used to predict the cumulative values, ARIMA (1,2,2) found to be the best fit as it has.

6. CONCLUSION:

Up on performing analysis with all the four different datasets, we obtained best fitted model with less error and more accuracy for individual dataset respectively.

To predict the best fitted model , of all the datasets, collective comparison interpreted that,

DATASET	Best Fitted Model	MAPE	MAE	R-Square	R-Square adjusted
Price	ARIMA(1,1,0)	1.17	0.73	0.998	0.998
Change	MA(2)	-	0.73	0.0705	0.07
Percentage of Change	MA(2)	-	0.0166	0.066	0.065
Cummulative	ARIMA(1,2,2,)	0.00261	0.732	0.92	0.92

On observing all the Accuracy measures like MAPE, MAE, R-Square and R-Square adjusted values, we observed that **"Oil Price dataset"** ARIMA (1,1,0) model is the best model with more accuracy and less error. Even forecasted 30 observations with this model are more approximately near to the actual values.