Team - 6

HOUSE PRICE INDEX – ANALYSIS & FORECASTING

Anita Sivakumar
Divya Polavaram
Sriram Rajagopalan
Sunmeizi Huang
Suriyaa Sugumar Valantina

Contents

EXECUTIVE SUMMARY

Housing or Real estate is a key indicator of country's economy. On an average, it contributes roughly 5 percent of GDP while housing services contribute to 12 to 13 percent. Also, real estate is an investment and a mode of saving for many. It has added $1,898 billion to GDP in 2011. Decline in housing prices in 2008 is a good example of the extent of impact housing sector can have on economy. A prospering housing industry boosts employment opportunities. Also, real estate performance determines the interest rates, affecting the financial sector. In turn, housing prices are effected by several economic factors like employment, interest rate changes, country's economy etc.

All the key stakeholders like buyers, sellers, banks and real estate agent would be keen to know the trend of housing prices and growth prospects. Analysis and forecast of Housing price index will help them in strategizing business plans. But, real estate is a huge market and certain places are lucrative compared to others. Also, few markets are highly affected by changes in economic factors while others are relatively stable. Hence, it would be a wise choice to evaluate the markets which hold the major chunk of housing business.

In order to estimate the trend and see the future trends, top 5 states of US housing markets are chosen. Time series analysis of Quarterly housing price data from 1975 to 2016 has shown some key trends. Also, quarterly forecast for next three years from 217 to 2019 was obtained which serves as a substantial input for several stakeholders. Results show that housing prices in each state have a unique trend and are determined by several factors.

DATASET

The dataset was primarily sourced from Kaggle. It contains the FHFA house price index - the weighted, repeat-sale index which is subject to price changes in repeat sales and refinancing on the same properties. This data was collected by monitoring mortgage transactions on single family properties which were purchased by Fannie Mac or Fannie Mae.

The time series data explains the monthly and quarterly trend of House price index in all the 50 states in the US at both state and city level ranging between years 1975 and 2016. It also captures 3 HPI flavors and 3 HPI types. In this business problem, the non-seasonally adjusted index (index_nsa) was chosen as the time series variable.

Data summary:

- Total number of observations: 99,325
- Total number of attributes: 9
- Time Period: January 1975 - August 2016
- Time series variable: index_nsa
- Data Source: https://www.kaggle.com/PythonforSASUsers/hpindex

DATA PREPARATION

Some of the challenges we encountered are

- To choose the level of data for the exercise.
- To set the time period variable since the quarters were numbered 1-4.

Final data set was prepared at quarterly level for HPI type - “traditional” and HPI flavor - “All transactions” for the top 5 states - California, Florida, Illinois, New York and Texas. The second challenge was addressed by setting the end date of each quarter in place of the quarter number as follows

- 1 - March 31
- 2 - June 30
- 3 - September 30
- 4 - December 31

METHODOLOGY

Since the dataset spanned over 50 states of America, we decided to forecast the House Price Index(HPI) values for top 5 states based on the Home Price Appreciation(HPA).

Appreciation is an increase in the value of property over time. As an asset, real estate has the potential to appreciate based on several factors such as demand, inflation or improvements made to a society. The Federal Housing and Finance Agency(FHFA) of United states releases the top states based on Home price appreciation(HPA) for every quarter. The Home Price Appreciation(HPA) is generally determined by calculating the median Home value in each state in the United States. Considering the forecast period which ranges from 2017 to 2019, the top states which had the highest HPA by the end of Q3 2016 were selected for analysis. The following were the top states:

1. California
2. New York
3. Florida
4. Illinois
5. Texas

There are different types and flavors based on which the HPI is determined. Each flavor and type contributes to a different HPI value signifying the various factors which influences the index value. The different flavors are “All-transactions”, “expanded-data” and “purchase only”.

The “all-transactions” HPI is determined by adding prices from appraisal data obtained from the Enterprises. Chartered by Congress for the purpose of creating a reliable supply of mortgage funds for homebuyers, Fannie Mae and Freddie Mac also called as the Enterprises are the largest mortgage finance institutions in the U.S. representing a significant share of total outstanding mortgages. FHFA uses data supplied by Fannie Mae and Freddie Mac which contains details of weighted repeat-transactions index based on property matches within its own database.

Since the “all-transactions” HPI flavor is reliable and is based on all the mortgage transactions, it will provide a comprehensive value to the forecast for each state and hence we decided to use this flavor for forecasting HPI for the top 5 states.

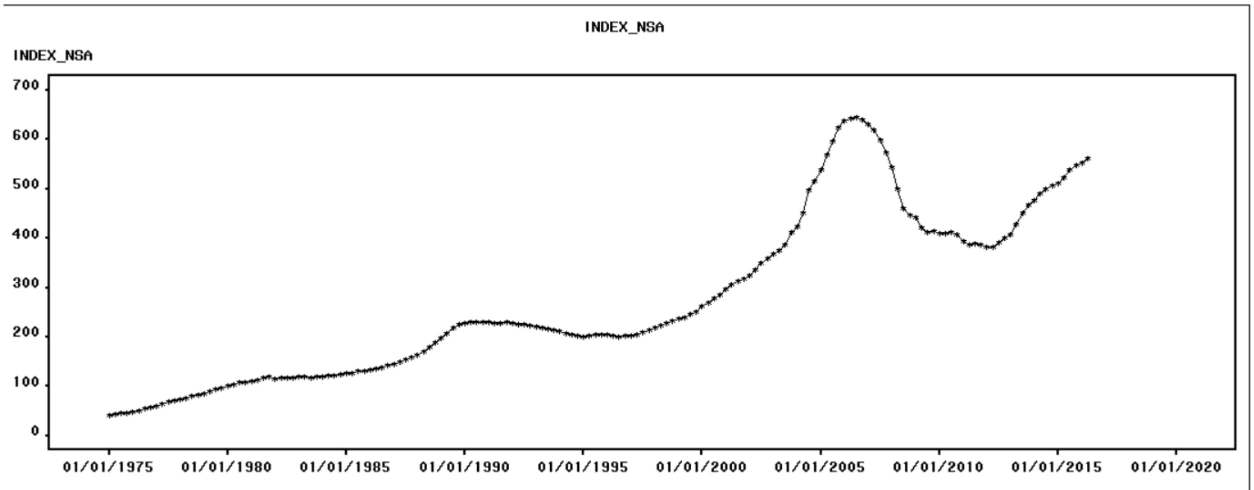
Similarly, we decided to forecast the HPI using the “Traditional” type measure. When incomes rise and/or mortgage rates fall, consumer house-buying power increases. The “Traditional” type HPI measures house price affordability which are dependent on the assumption of specific loan terms (down payment, Loan-to-value) and the choice of income level which broadly reflects the real-price experienced by customers.

The approach was to study the trends in “all-transactions: House Price Index of “traditional” type in the top 5 states and come up with an analysis on how the HPI values will turn out for the next 3 three years starting from 2017Q1. This will provide a buyer-seller level perspective on the best and worst places to invest on a house and helps to monitor long-term trends and events in the US housing market.

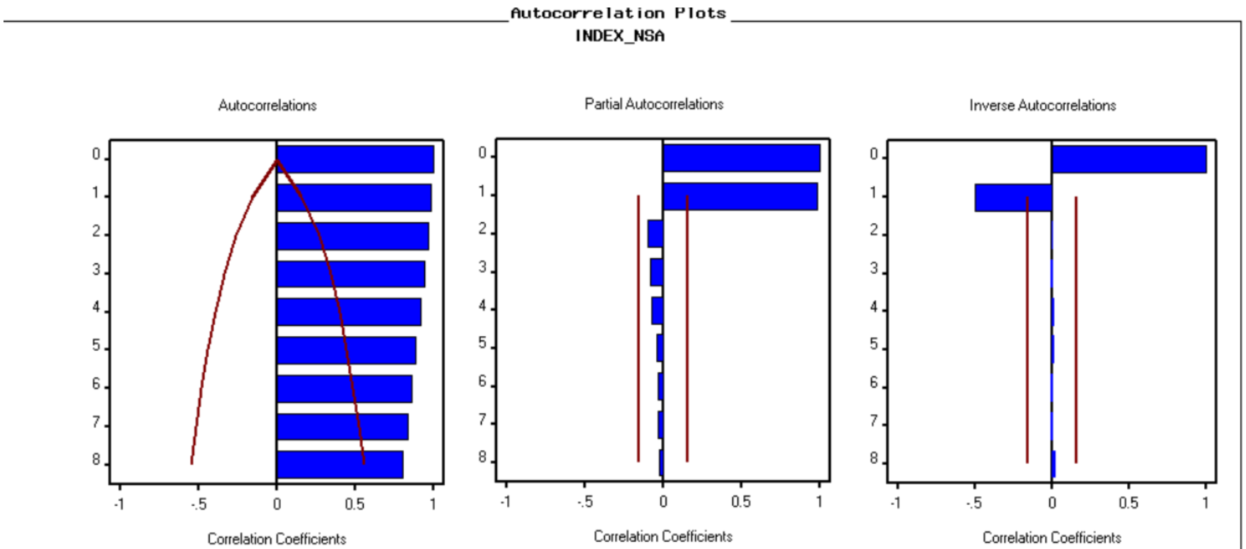
MODEL BUILDING AND SELECTION

California

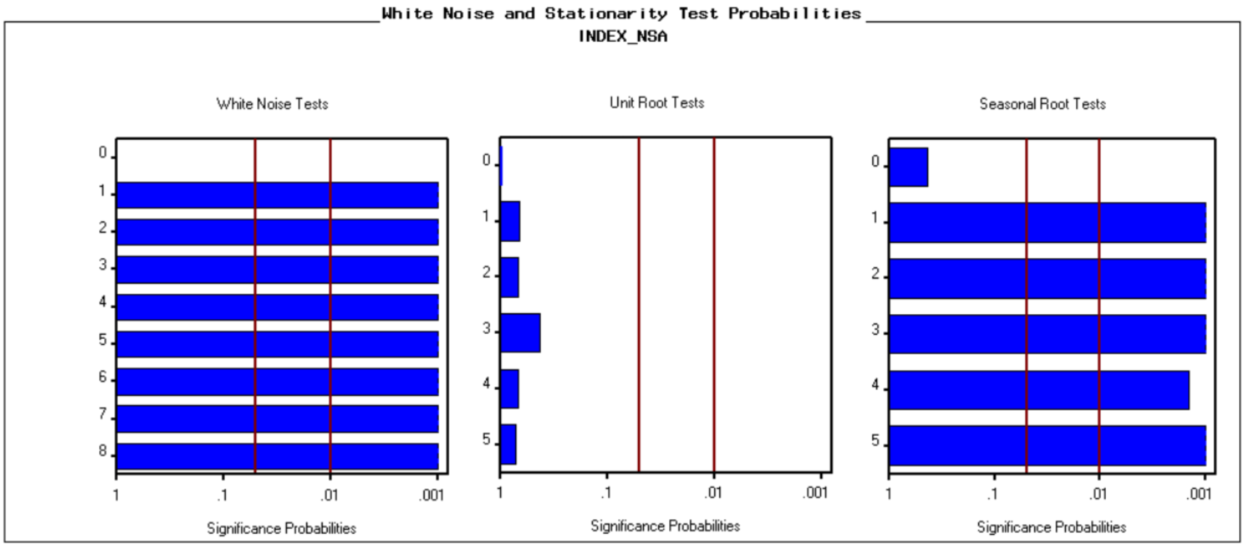
The following graph shows the quarterly trend of house price index in the state California.



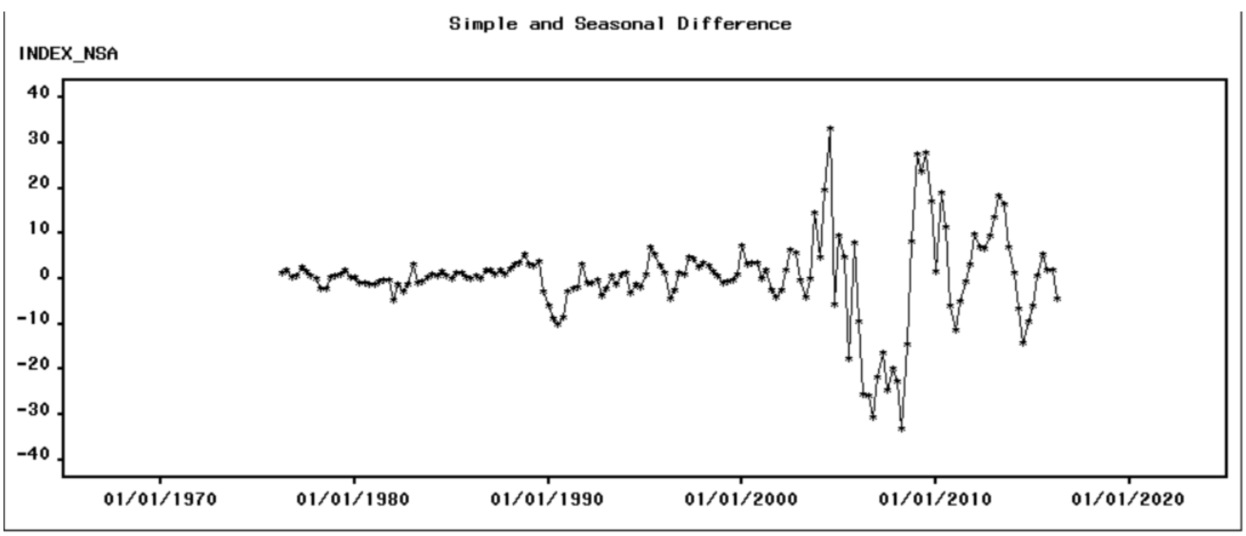
The autocorrelation plots indicate a strong trend component in the time series with evidence of significant number of slowly decaying ACF and many significant lags. ACF (1) is large compared to the subsequent lags.



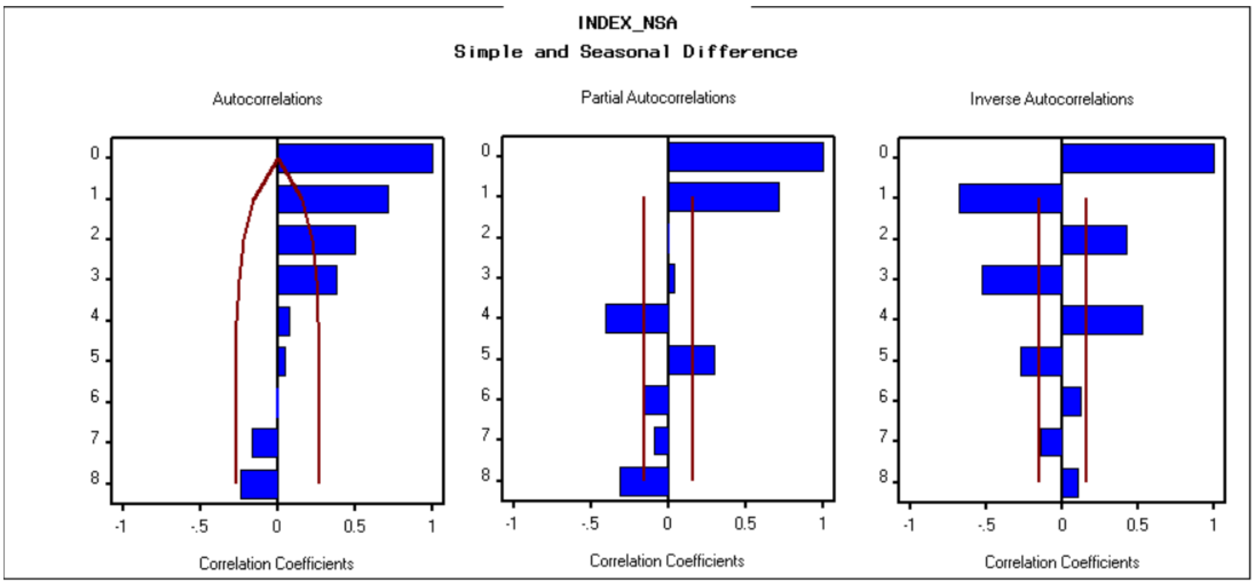
The unit root test strongly indicates the presence of trend component and need for first differencing. The seasonal root test indicates slight presence of seasonal component. However, the time series is clearly not stationary which is evident from the unit root tests.



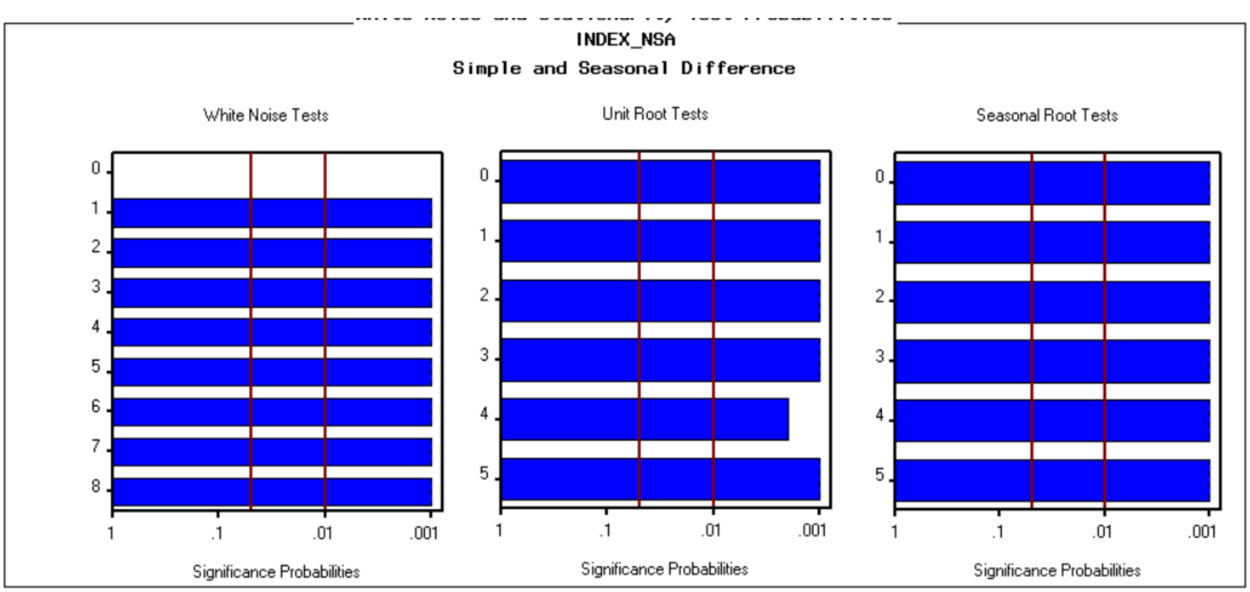
After applying first and seasonal differencing, we observed a stationary time series except for the spikes between 2009 and 2011.



After first and seasonal differencing, ACF (1) is different from 0 but there still seems to be slowly decaying components until ACF(3). However, it is significantly smaller when compared to the model without differencing.



Applying first and seasonal differencing indicates that the error term is stationary which is evident from the unit and seasonal root tests. However, the white noise tests still indicate the presence of autocorrelation in residuals.

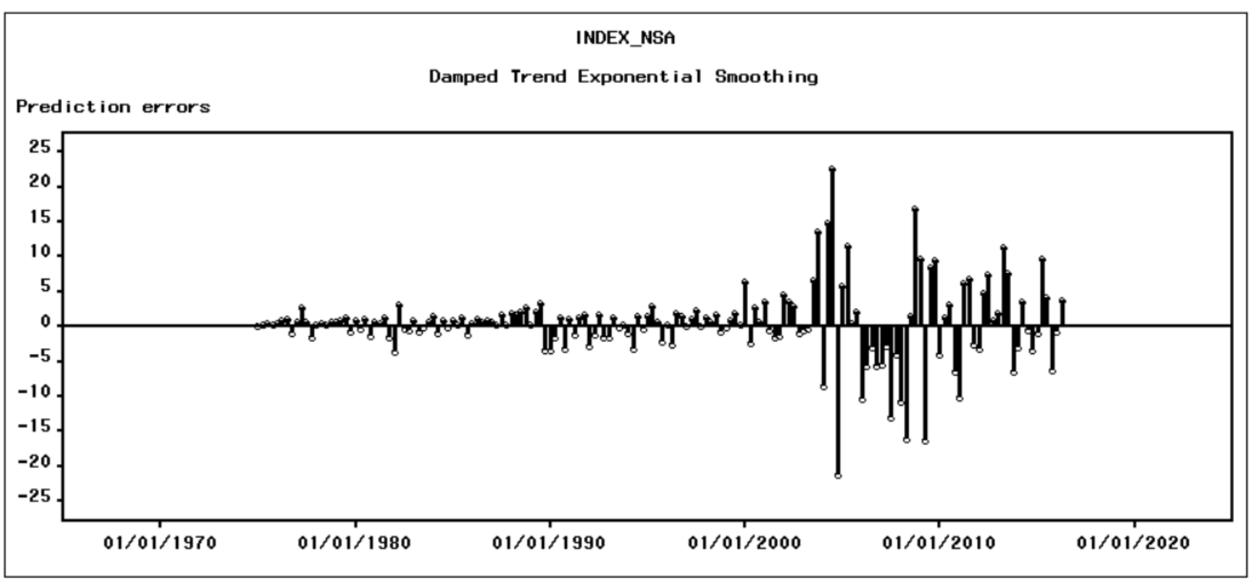


Models:

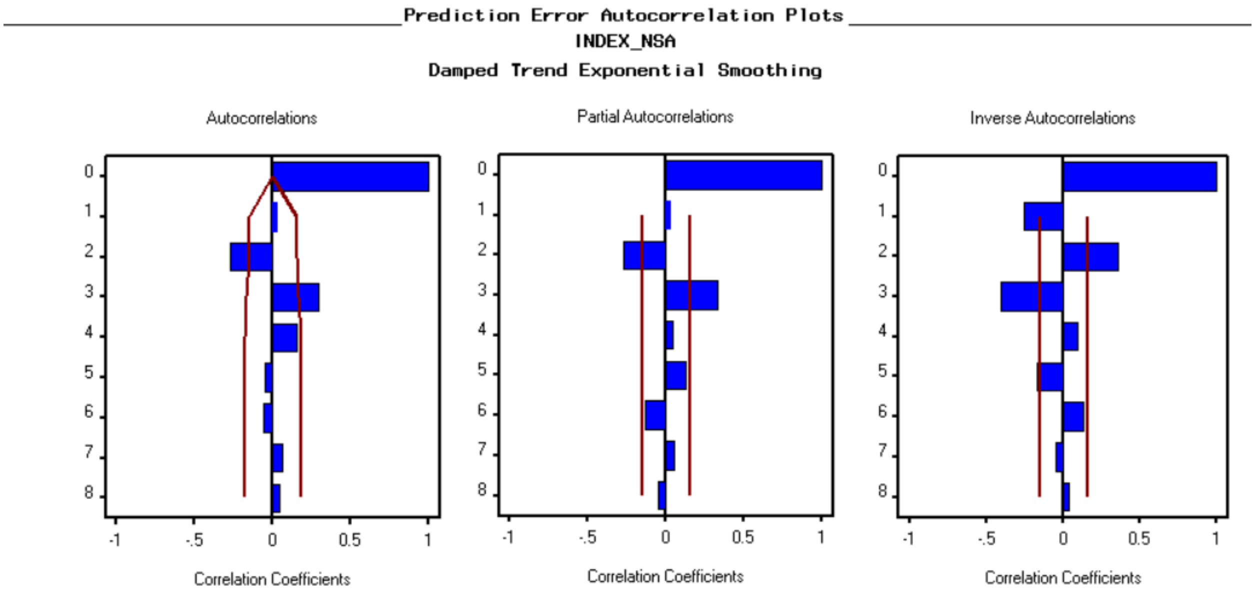
Since the time series has a strong trend component and the data ranges from the year 1975, we first tried building trend models to view the performance. However, the results did not turn out to be good indicating that the trend models are not the best fitting models. We tried to model the decaying weighted average of the past values and hence decided to try all the exponential smoothing models. The resulting RMSE values significantly improved and the Damped Trend Exponential Smoothing Model was the best fitting model. Further assessment of plots and model fit statistics for this model was carried out.

Forecast Model	Model Title	Root Mean Square Error
✔	Damped Trend Exponential Smoothing	5.13355
	Winters Method -- Additive	5.20303
	Linear (Holt) Exponential Smoothing	5.28666
	Double (Brown) Exponential Smoothing	5.28730
	Winters Method -- Multiplicative	5.48288
	Simple Exponential Smoothing	10.65854
	Seasonal Exponential Smoothing	10.99133
	Linear Trend	70.87829
	Quadratic Trend	70.31906
	Cubic Trend	67.83577
	Exponential Trend	83.47530

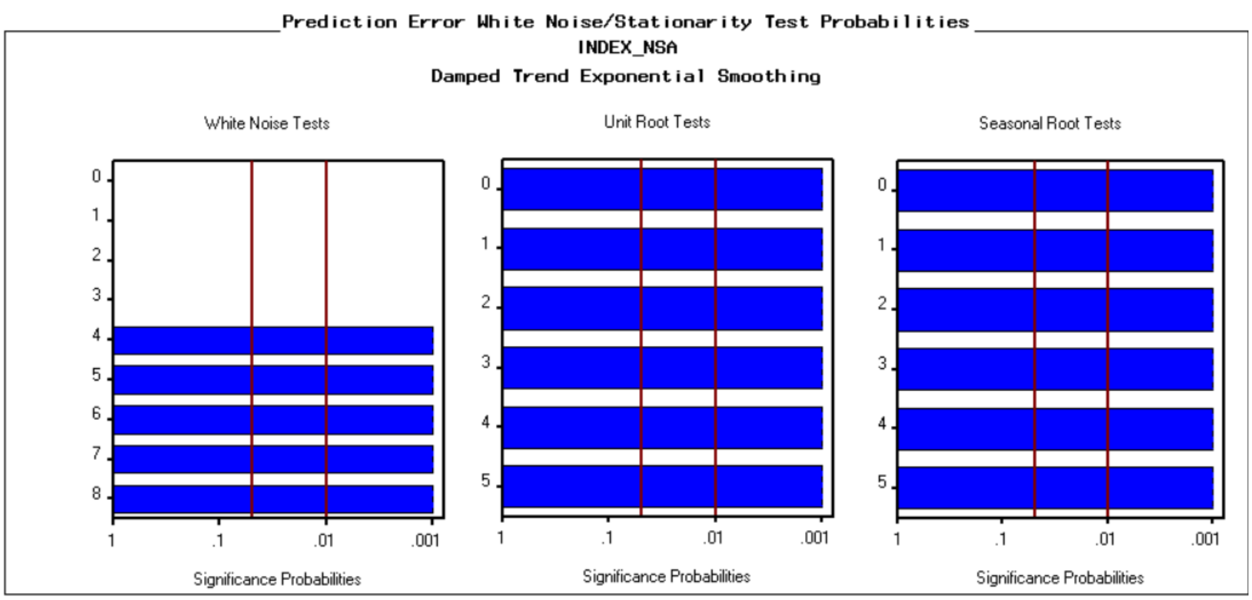
The residuals appear to be random with some spikes in between 2004 and 2009.



Checking the autocorrelation plots of the residuals suggest that none of the spikes are significantly different from 0. We can infer that the model explains all the significant autocorrelation that was in original data.



The following plot clearly suggests that the residuals pass the unit and seasonal root tests, and the white noise test.



The parameter estimates suggest that all the parameters are significant with low p-value.

Parameter Estimates
INDEX_NSA
Damped Trend Exponential Smoothing

Model Parameter	Estimate	Std. Error	T	Prob>|T|
LEVEL Smoothing Weight	0.99900	0.0720	13.8691	<.0001
TREND Smoothing Weight	0.99900	0.1854	5.3877	<.0001
DAMPING Smoothing Weight	0.87838	0.0347	25.3085	<.0001
Residual Variance (sigma squared)	26.83836	.	.	.
Smoothed Level	561.67625	.	.	.
Smoothed Trend	8.90177	.	.	.

Inference from Statistics of Fit

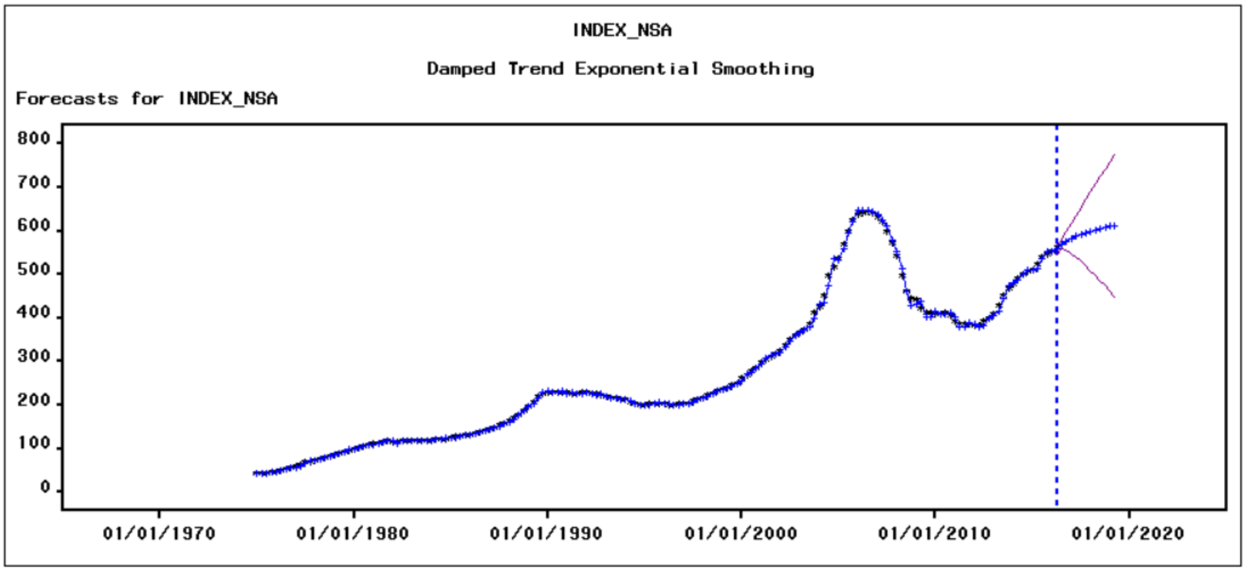
This model produces the lowest SBC, RMSE, MAPE and MAD compared to other exponential smoothing models.

Damped Trend Exponential Smoothing

Statistic of Fit	Value
Mean Square Error	26.35332
Root Mean Square Error	5.13355
Mean Absolute Percent Error	1.03987
Mean Absolute Error	3.12931
R-Square	0.999

ation Range: 1975:1 to 2016:2

The following graph shows the forecast for the selected model with tight 95% confidence intervals indicated by pink line.



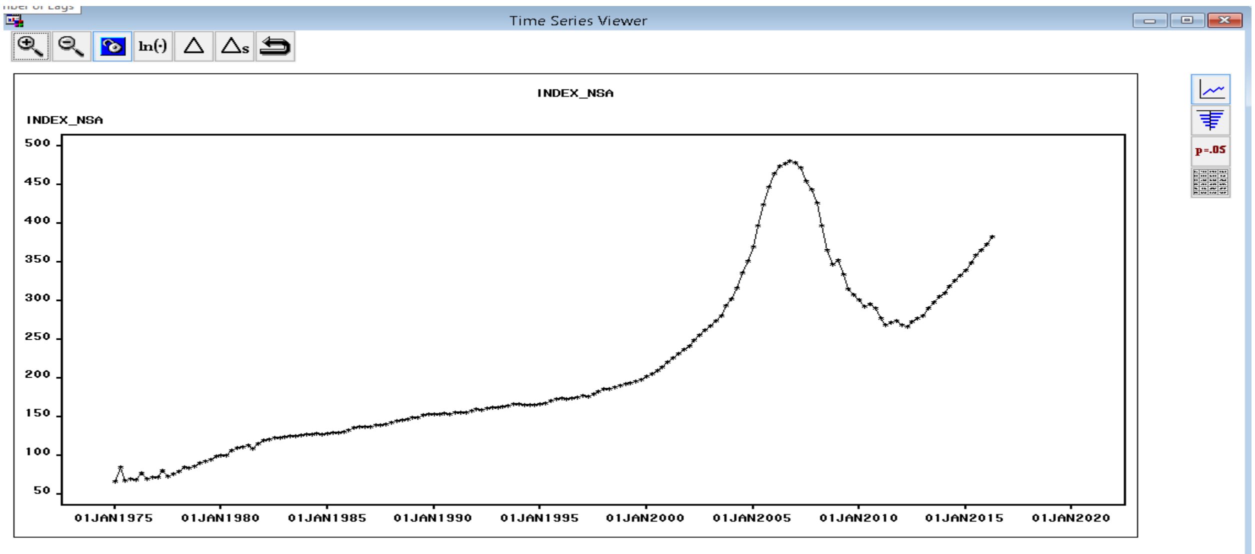
Forecast Results

INDEX_NSA
Damped Trend Exponential Smoothing

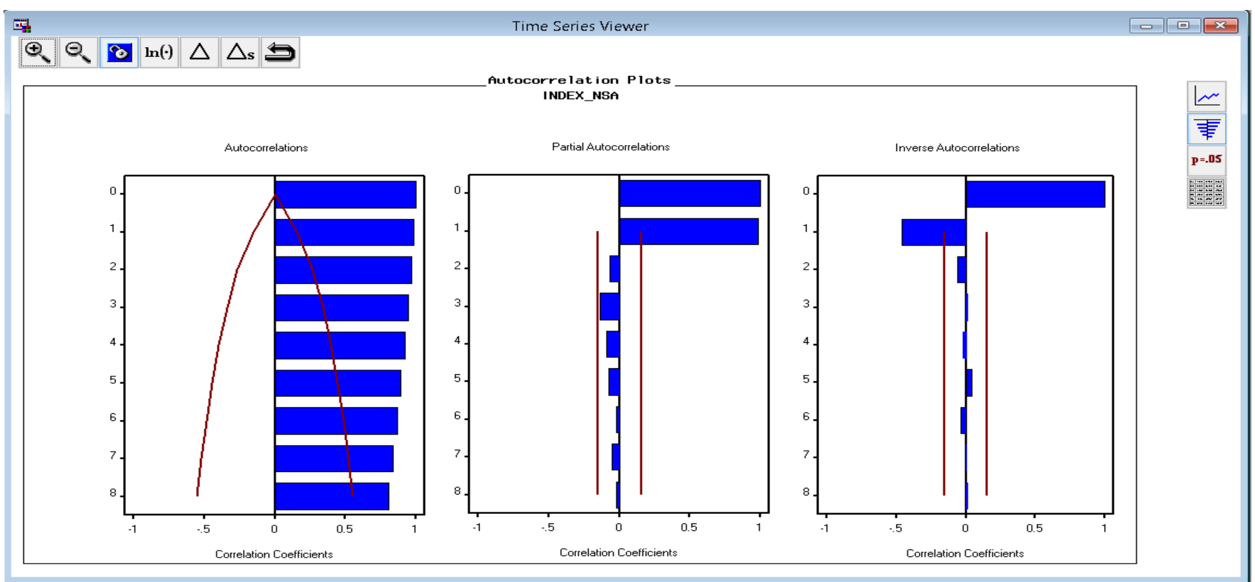
TIMEPERIOD	ACTUAL	PREDICT	U95	L95	ERROR
12/01/2016	.	576.3635	597.9457	554.7812	.
03/01/2017	.	582.3962	616.8551	547.9374	.
06/01/2017	.	587.6953	635.9152	539.4754	.
09/01/2017	.	592.3498	654.8497	529.8500	.
12/01/2017	.	596.4383	673.4886	519.3880	.
03/01/2018	.	600.0295	691.7266	508.3323	.
06/01/2018	.	603.1839	709.5003	496.8675	.
09/01/2018	.	605.9547	726.7739	485.1354	.
12/01/2018	.	608.3885	743.5307	473.2462	.
03/01/2019	.	610.5262	759.7667	461.2858	.
06/01/2019	.	612.4040	775.4871	449.3208	.

Florida

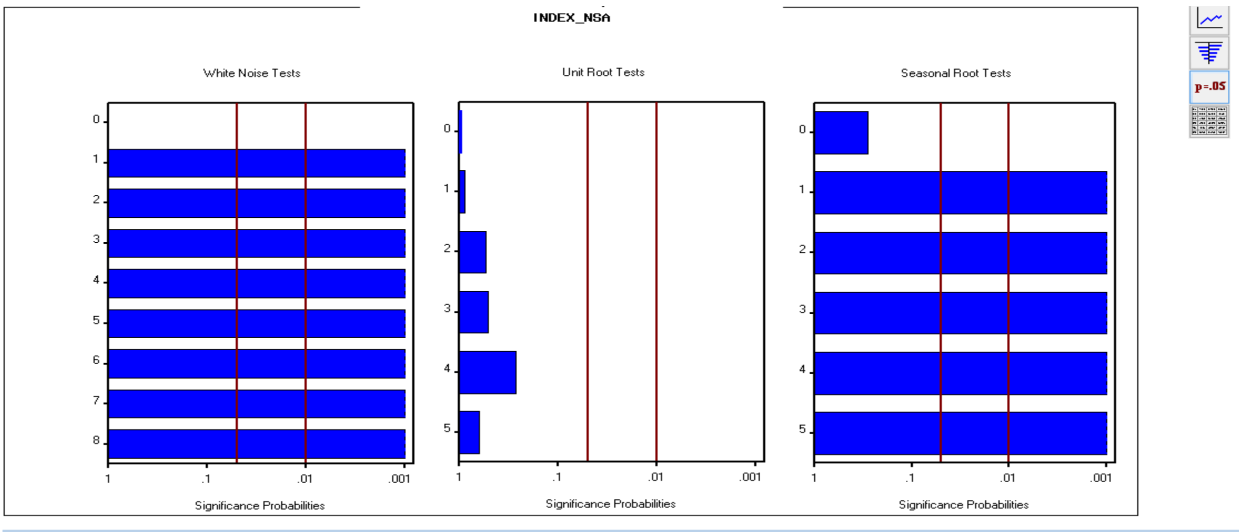
The following graph shows the quarterly trend of house price index in the state Florida.



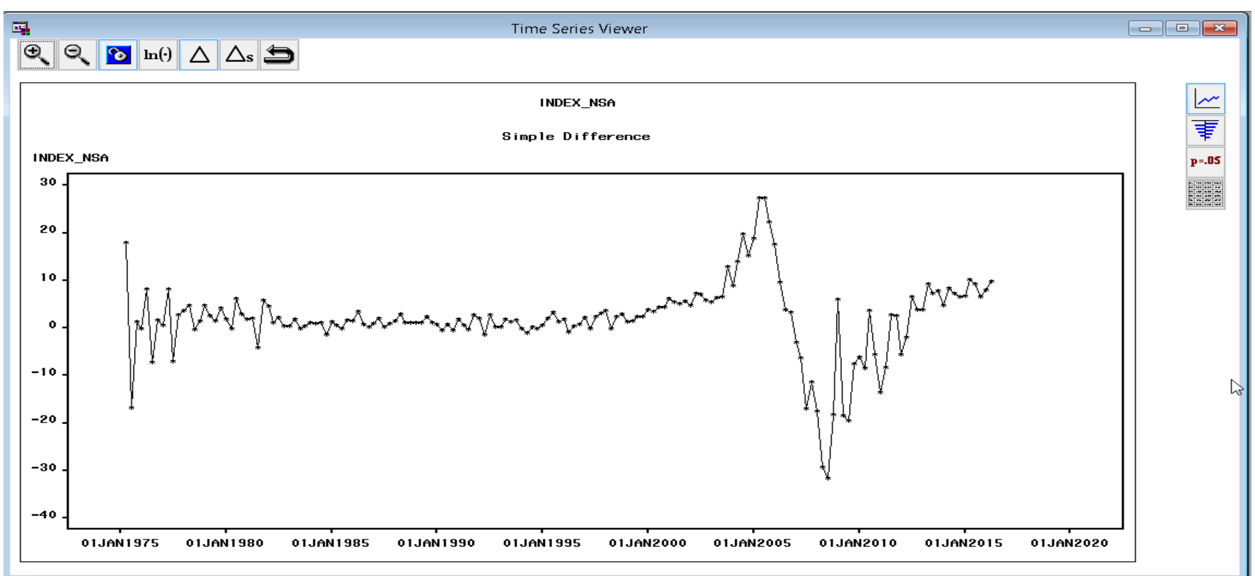
The autocorrelation plots indicate a strong trend component in the time series with evidence of significant number of slowly decaying ACF and many significant lags. ACF (1) is large compared to the subsequent lags.



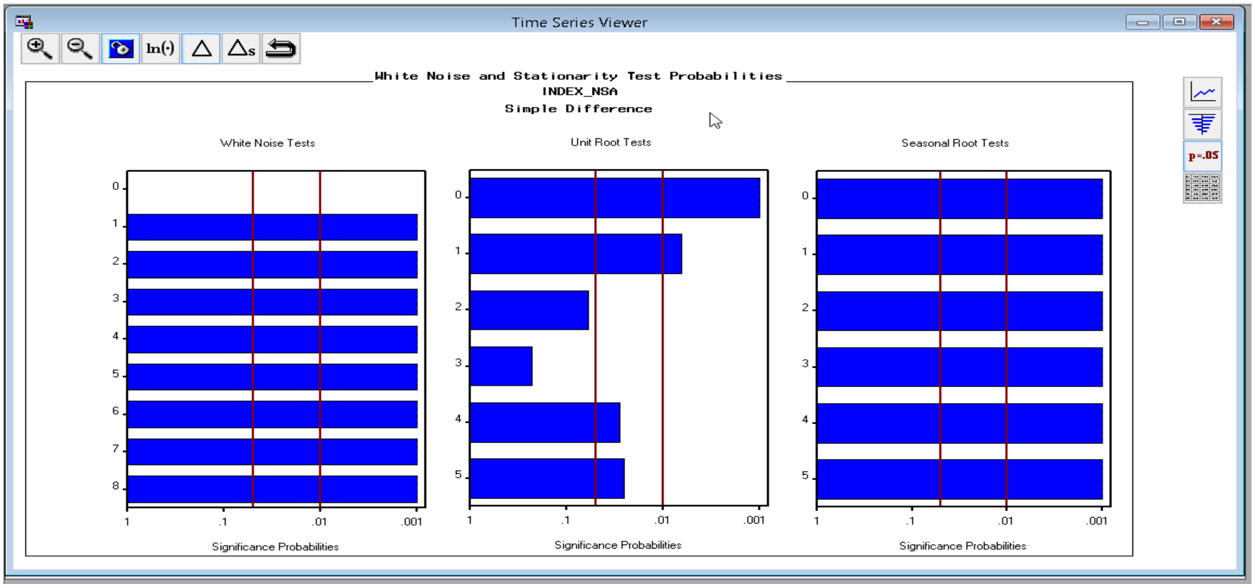
The unit root test strongly indicates the presence of trend component and need for second level differencing. The seasonal root test indicates presence of seasonal component. However, the time series is clearly not stationary which is evident from the unit root tests.

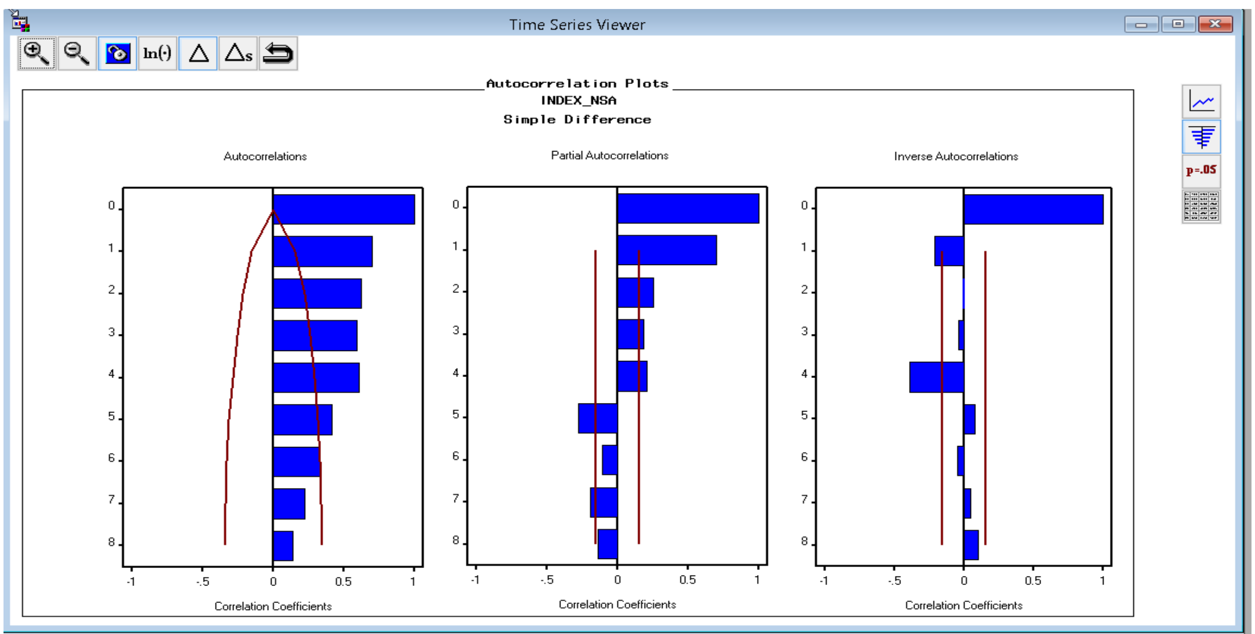


After applying first and seasonal differencing, we observed a stationary time series except for a potential intervention at 2008 which might account to the Housing Bubble impact.



Results appear to be better with simple difference.

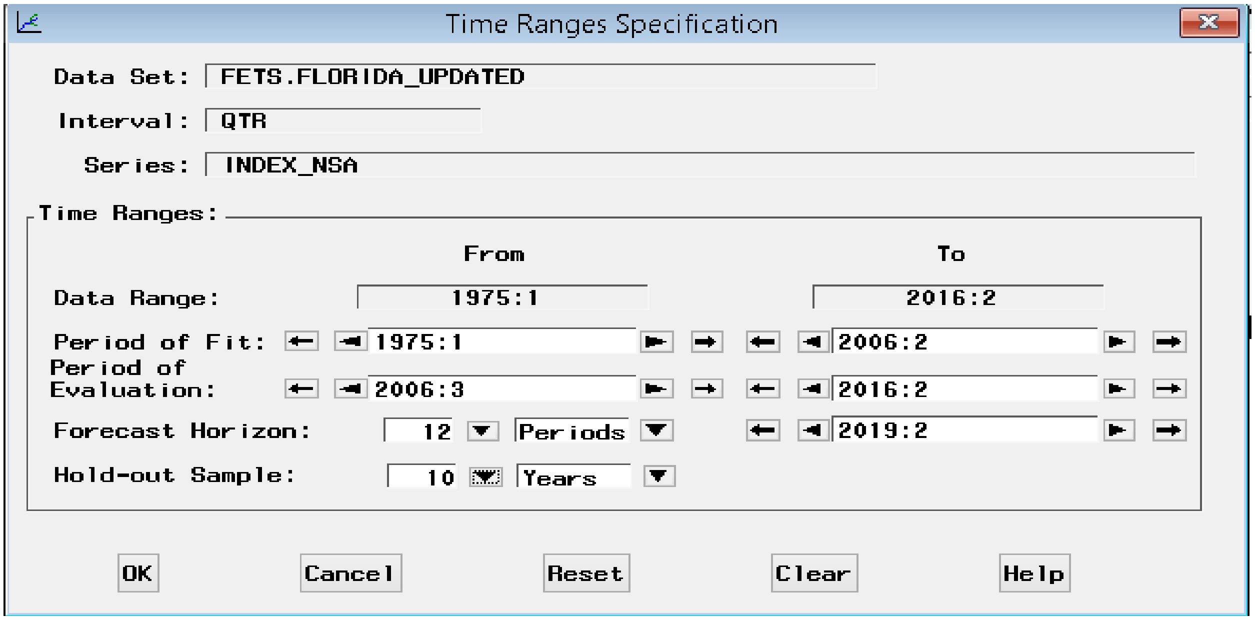




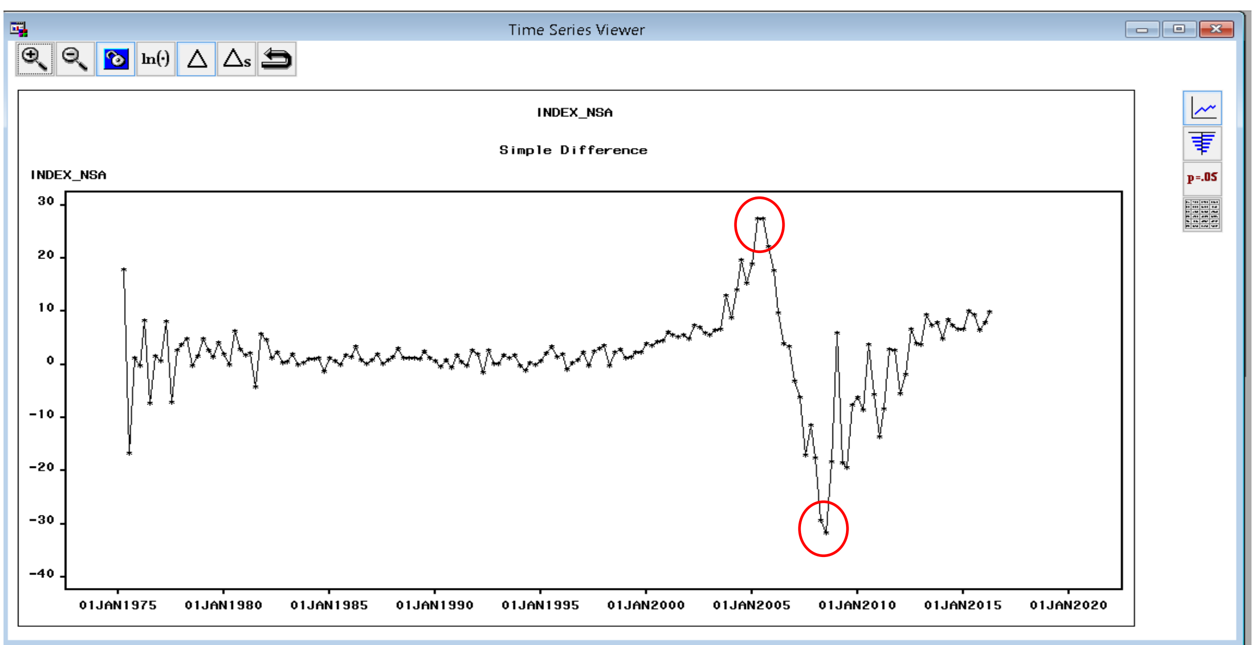
From above autocorrelation plots, following shall be the possible values of p, d, q.

(0,1,0), (0,1,1), (0,1,2), (0,1,3), (0,1,4), (1,1,0), (1,1,1), (1,1,2), (1,1,3), (1,1,4), (1,1,5), (1,1,6), (1,1,7), (1,1,8), (0,2,0), (0,2,1), (0,2,2), (0,2,3), (0,2,4), (0,2,5), (0,2,6), (0,2,7), (0,2,8), (1,2,0), (1,2,1), (1,2,2), (1,2,3), (1,2,4), (1,2,5), (1,2,6), (1,2,7), (1,2,8)

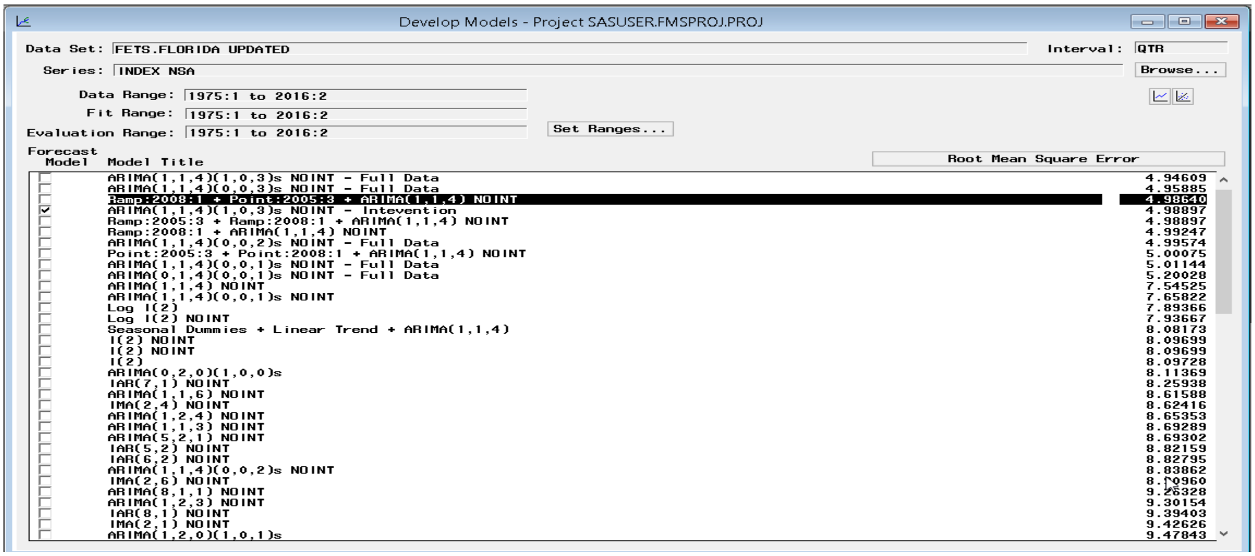
We set the hold-out sample by 25% of the whole data and built models.

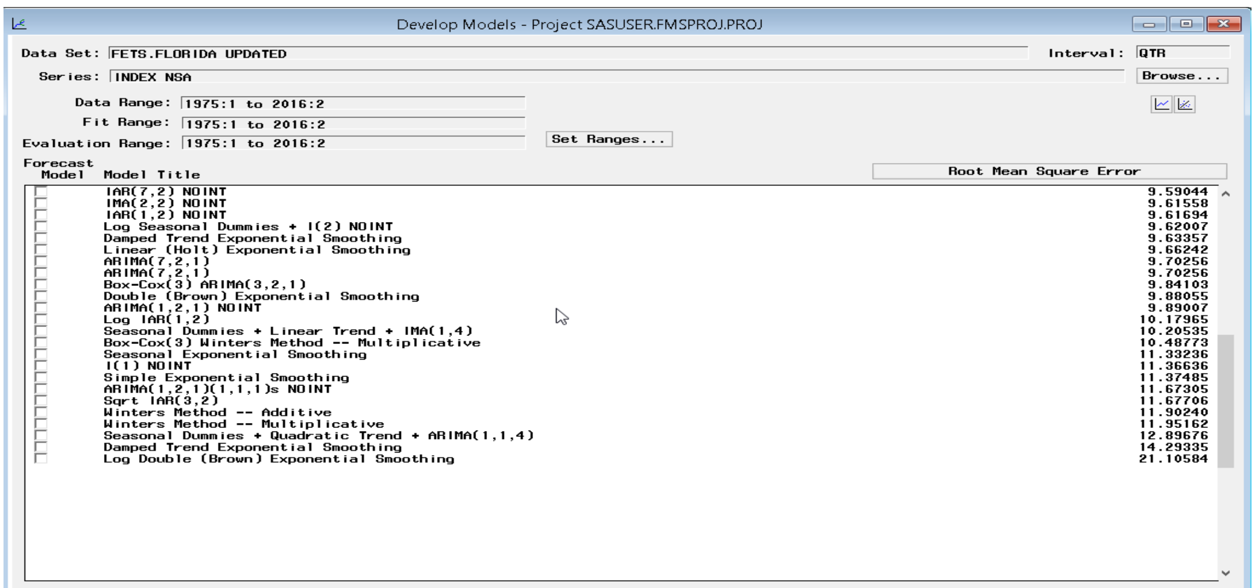


With difference, graph shows two prominent points - July 2005 and July 2008. Both are considered as ramp Interventions.



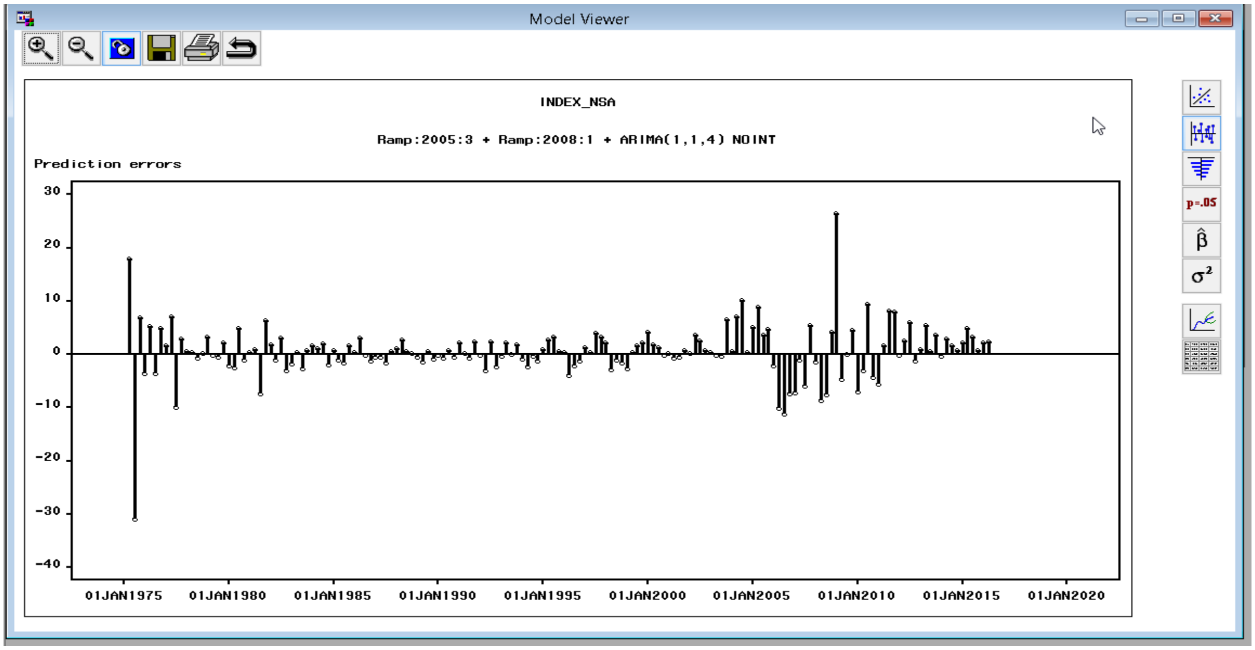
The following models were developed.



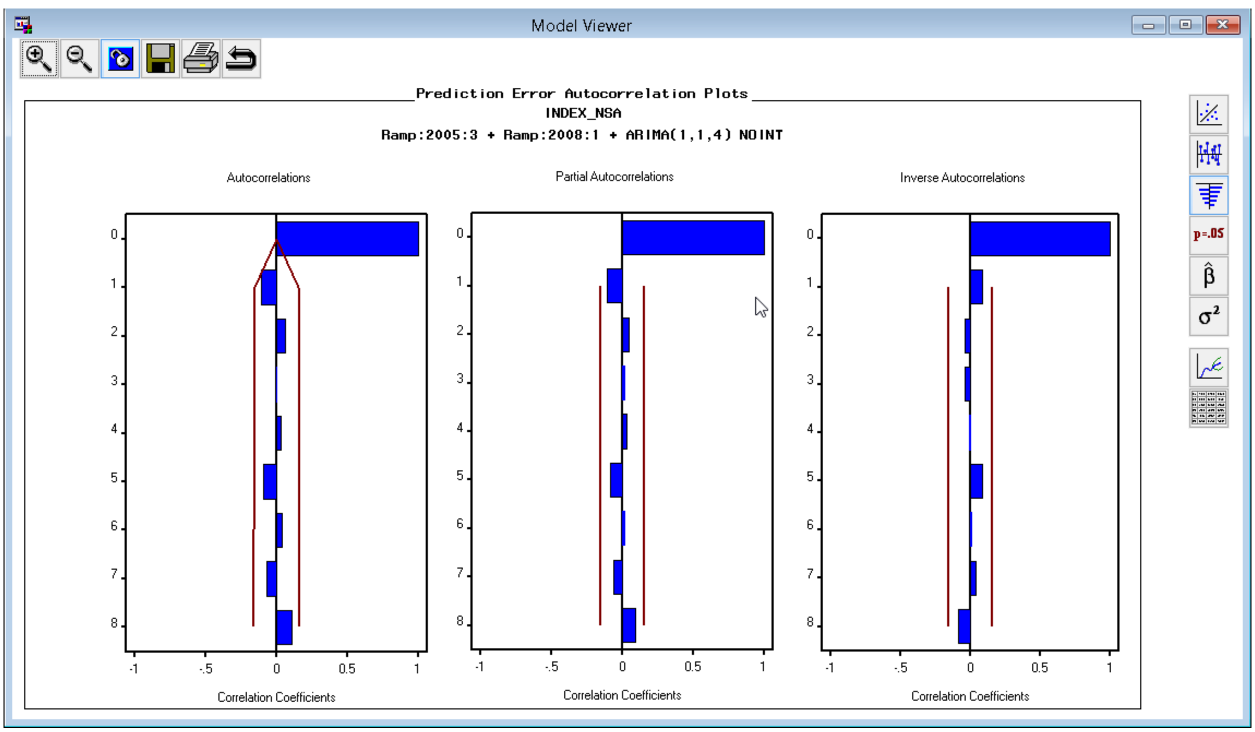


On analysis, the Ramp 2008:1 + Point 2005:3 + ARIMA (1,1,4) turned out to be the best performing model.

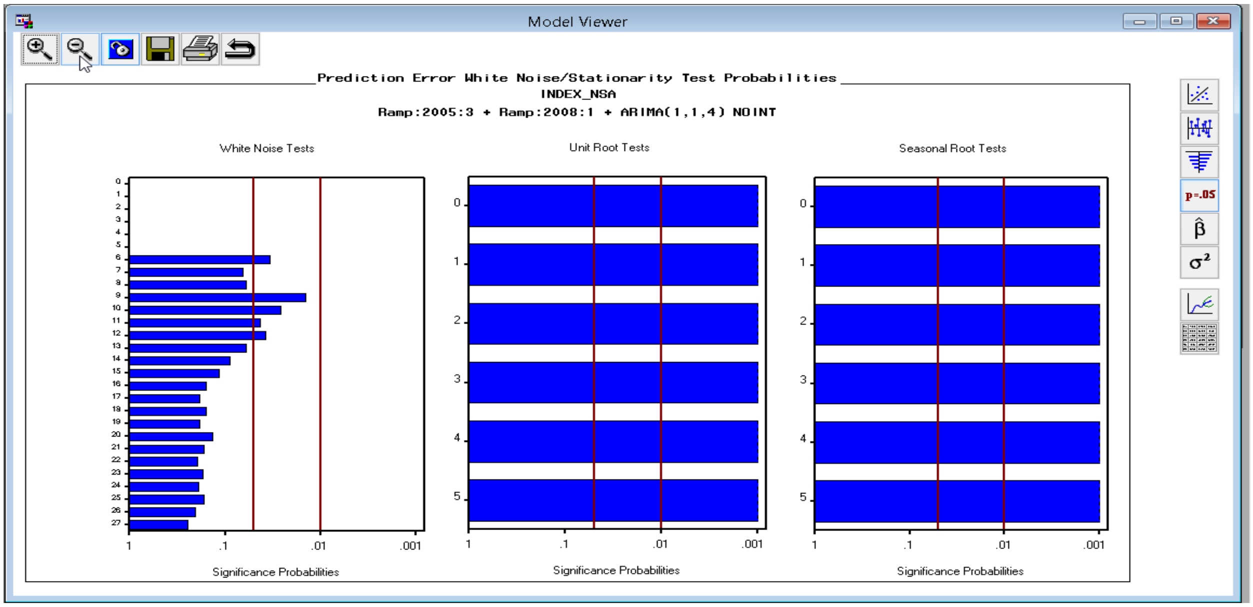
The residuals appear to be random with some spikes at the beginning and the rest being uniformly distributed.



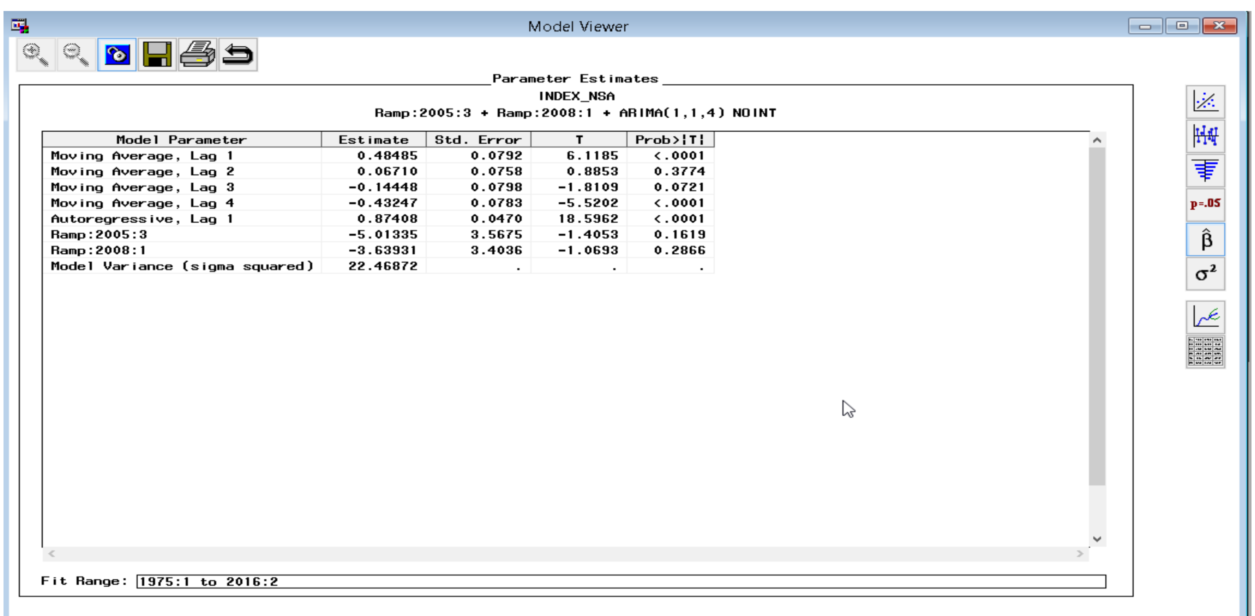
Checking the autocorrelation plots of the residuals suggest that none of the spikes are significantly different from 0. We can infer that the model explains all the significant autocorrelation that was in original data.



The following plot clearly suggests that the residuals pass the unit and seasonal root tests, and the white noise test.



The parameter estimates suggest that most of the parameters are significant with low p-value. However, the MA lag1 and lag 2 are not significant at 5% level



Model Viewer

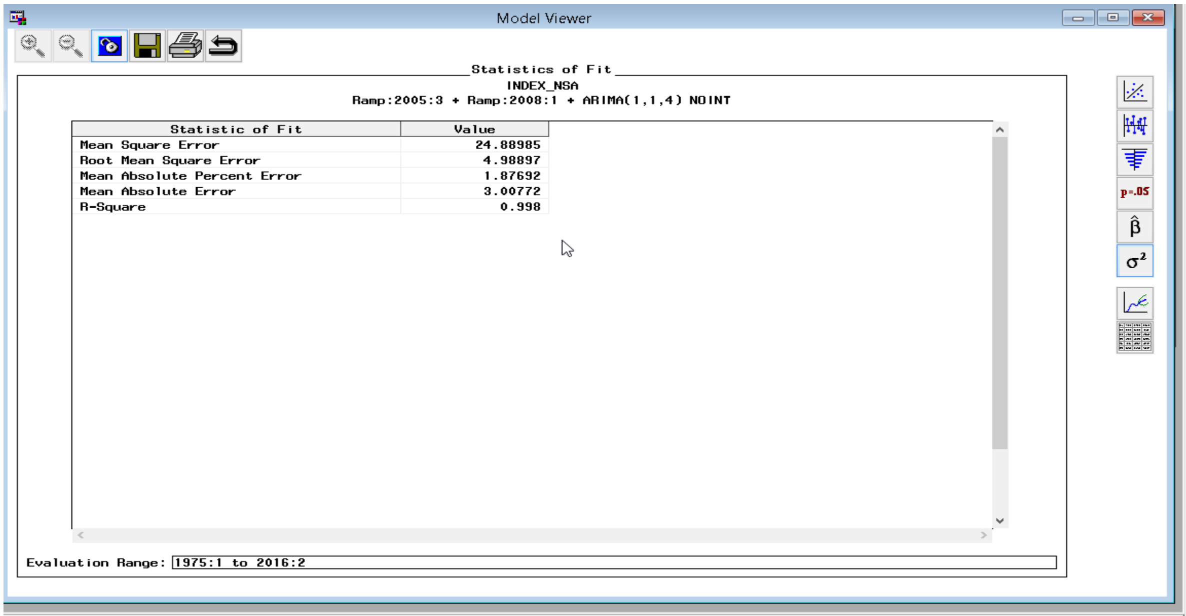
Parameter Estimates
INDEX_NSA
Ramp:2005:3 + Ramp:2008:1 + ARIMA(1,1,4) NOINT

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Moving Average, Lag 1	0.48485	0.0792	6.1185	<.0001
Moving Average, Lag 2	0.06710	0.0758	0.8853	0.3774
Moving Average, Lag 3	-0.14448	0.0798	-1.8109	0.0721
Moving Average, Lag 4	-0.43247	0.0783	-5.5202	<.0001
Autoregressive, Lag 1	0.87408	0.0470	18.5962	<.0001
Ramp:2005:3	-5.01335	3.5675	-1.4053	0.1619
Ramp:2008:1	-3.63931	3.4036	-1.0693	0.2866
Model Variance (sigma squared)	22.46872	.	.	.

Fit Range: 1975:1 to 2016:2

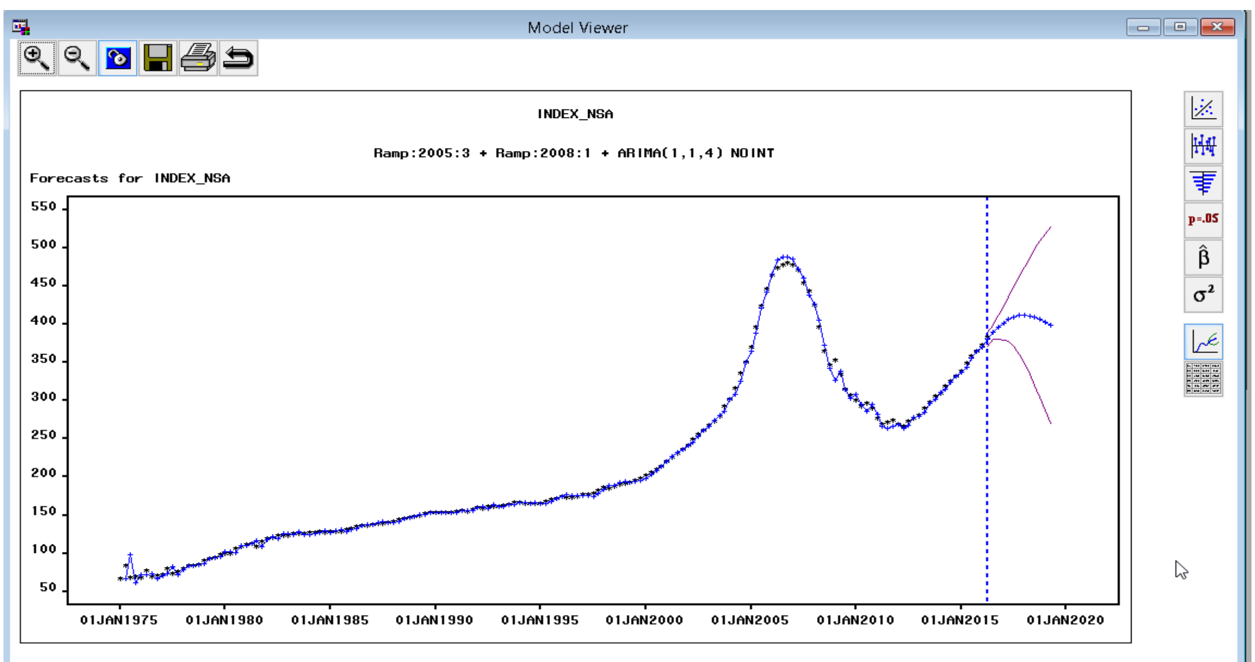
Statistics of Fit

This model produces the lowest SBC, RMSE, MAPE and MAD compared to other models that were built.

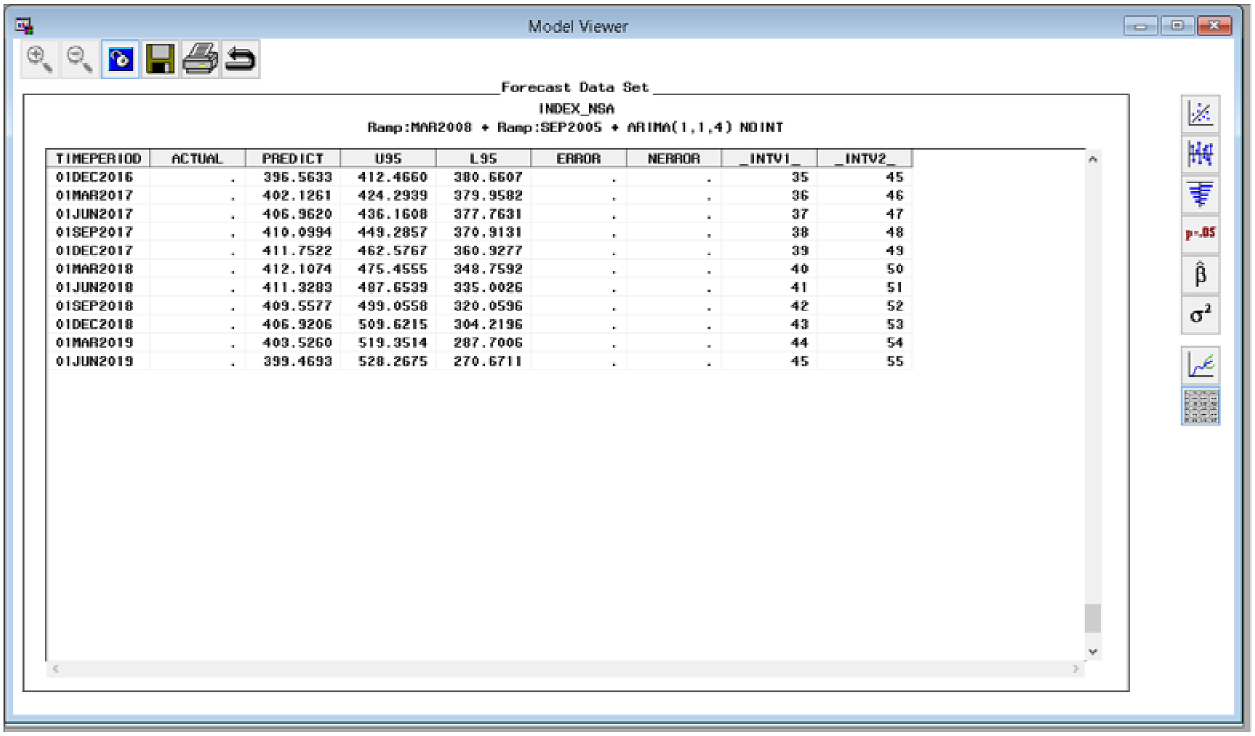


Statistic of Fit	Value
Mean Square Error	24.88985
Root Mean Square Error	4.98897
Mean Absolute Percent Error	1.87692
Mean Absolute Error	3.00772
R-Square	0.998

The following graph shows the forecast for the selected model with tight 95% confidence intervals indicated by pink line.



Forecasted Values


Model Viewer

Forecast Data Set

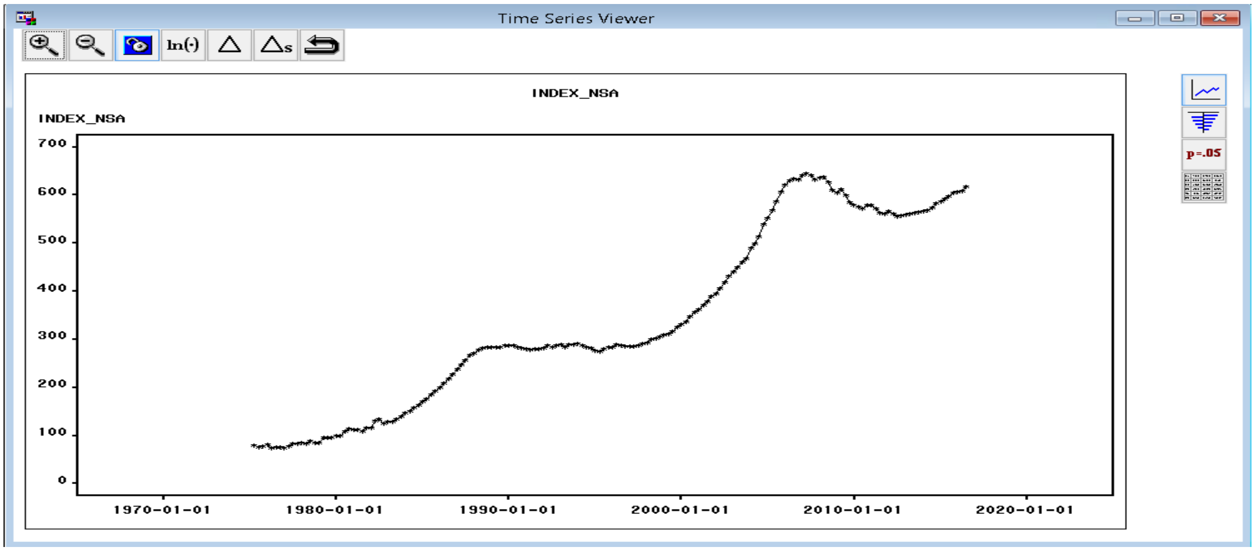
INDEX_NSA

Ramp:MAR2008 + Ramp:SEP2005 + ARIMA(1,1,4) NOINT

TIMEPERIOD	ACTUAL	PREDICT	U95	L95	ERROR	NERROR	_INTV1_	_INTV2_
01DEC2016	.	396.5633	412.4660	380.6607	.	.	35	45
01MAR2017	.	402.1261	424.2939	379.9582	.	.	36	46
01JUN2017	.	406.9620	436.1608	377.7631	.	.	37	47
01SEP2017	.	410.0994	449.2857	370.9131	.	.	38	48
01DEC2017	.	411.7522	462.5767	360.9277	.	.	39	49
01MAR2018	.	412.1074	475.4555	348.7592	.	.	40	50
01JUN2018	.	411.3283	487.6539	335.0026	.	.	41	51
01SEP2018	.	409.5577	499.0558	320.0596	.	.	42	52
01DEC2018	.	406.9206	509.6215	304.2196	.	.	43	53
01MAR2019	.	403.5260	519.3514	287.7006	.	.	44	54
01JUN2019	.	399.4693	528.2675	270.6711	.	.	45	55

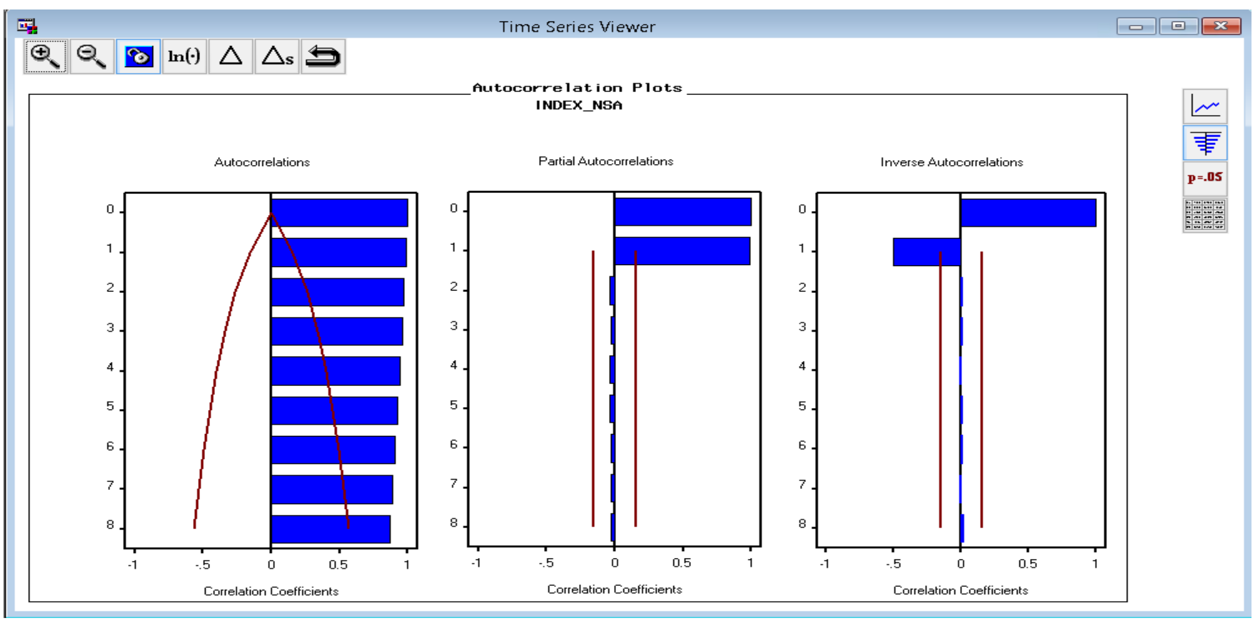
New York

The following graph shows the quarterly trend of house price index in the state New York.



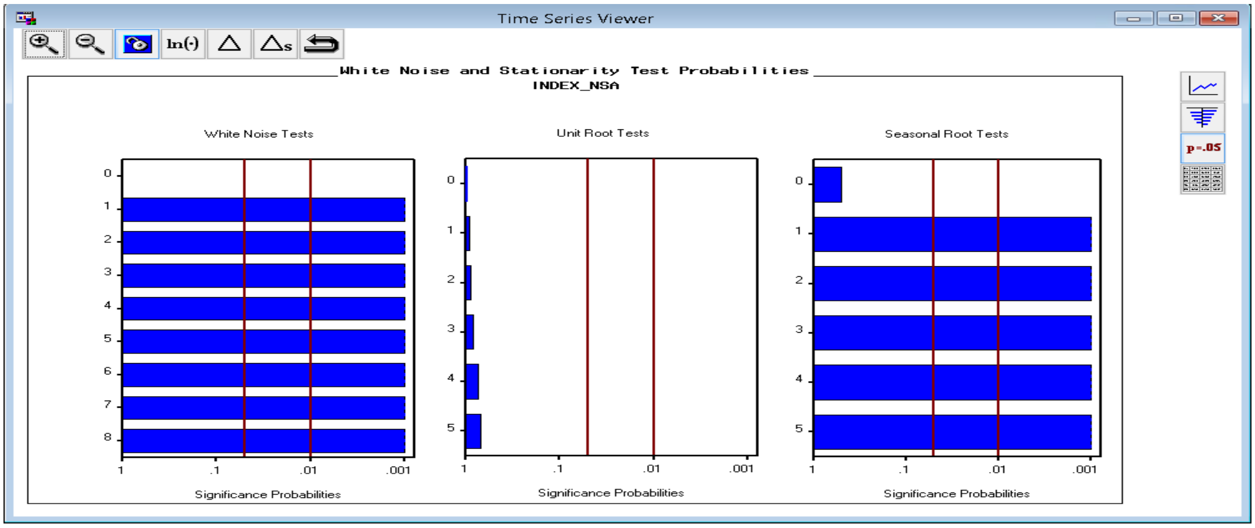
Inference from prediction error autocorrelation plots

The autocorrelation plots indicate a strong trend component in the time series with evidence of significant number of slowly decaying ACF and many significant lags. PACF (1)/IACF (1) is large compared to the subsequent lags.

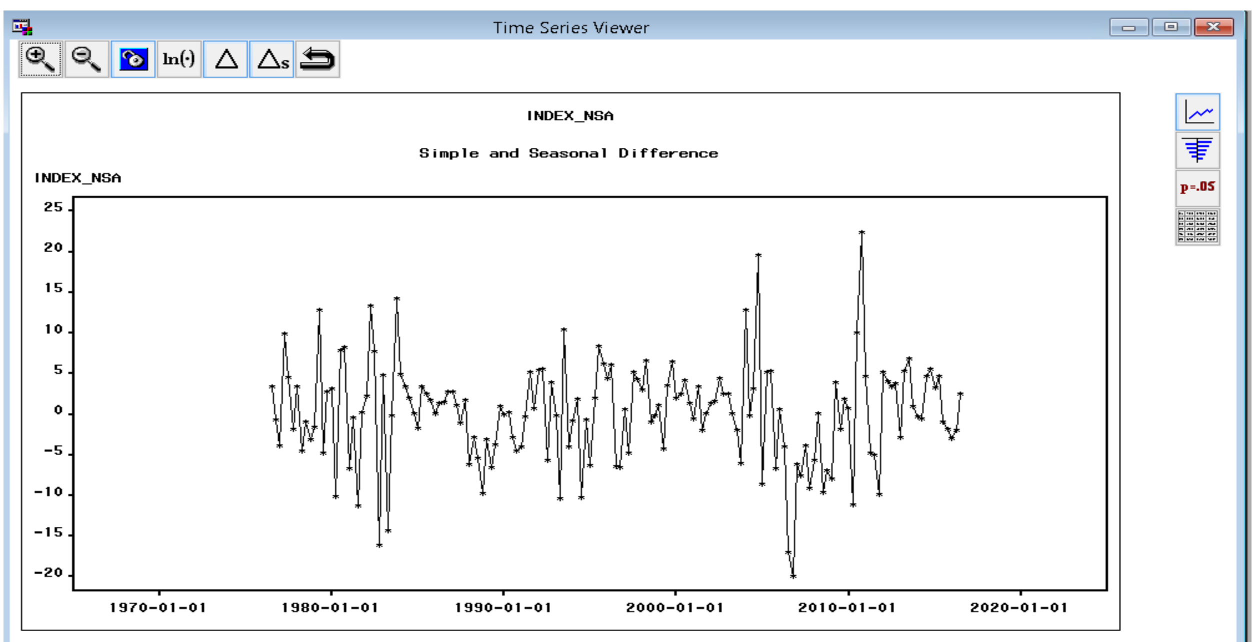


Interpretation from Prediction Error Tests

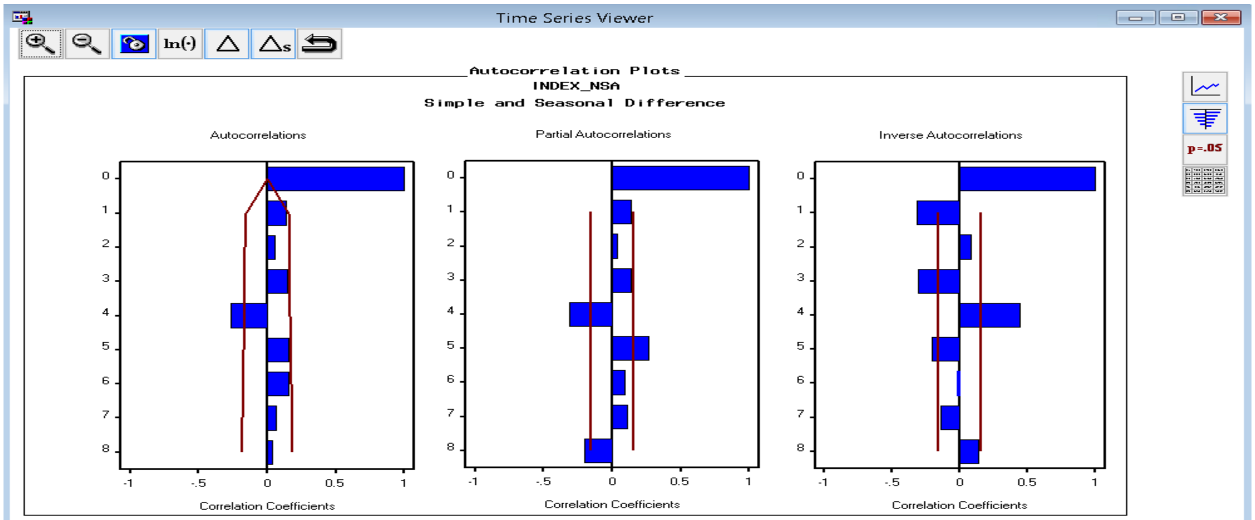
Dickey fuller unit root test strongly indicates the presence of trend component and need for first differencing. The seasonal root test indicates the presence of seasonal component. However, the time series is clearly not stationary which is evident from the unit root tests.



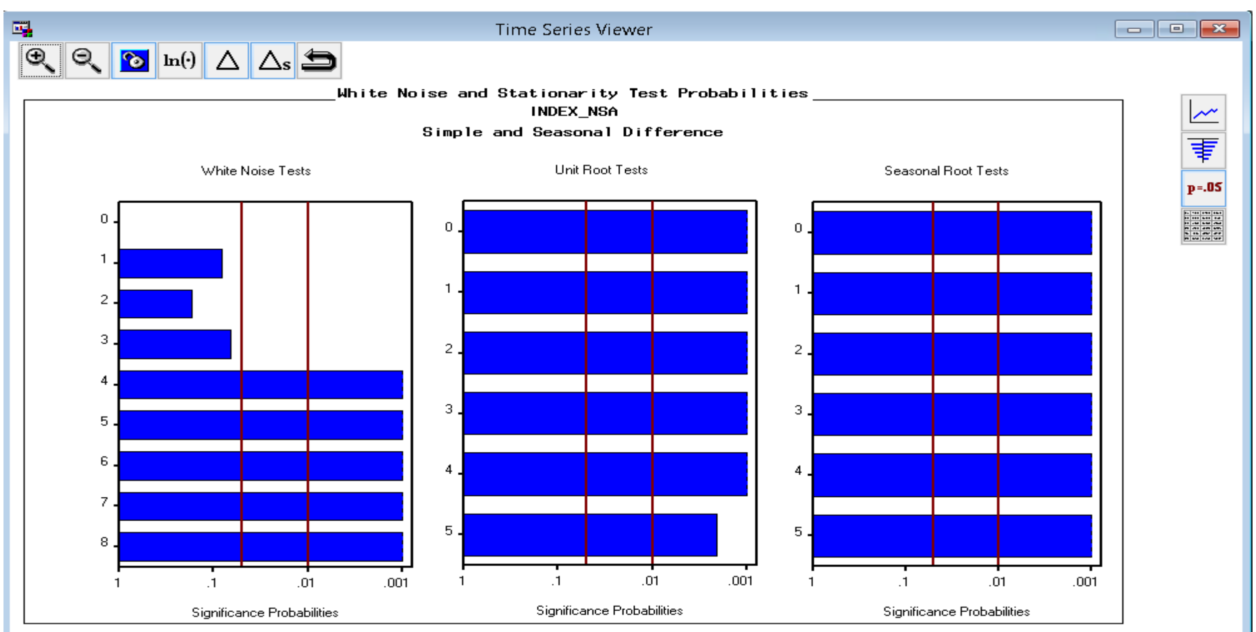
After applying first differencing and seasonal differencing, the graph seems without trend and seasonality as shown below.



After first and seasonal differencing, ACF (4) is still significantly different from 0 but smaller than the model without differencing.



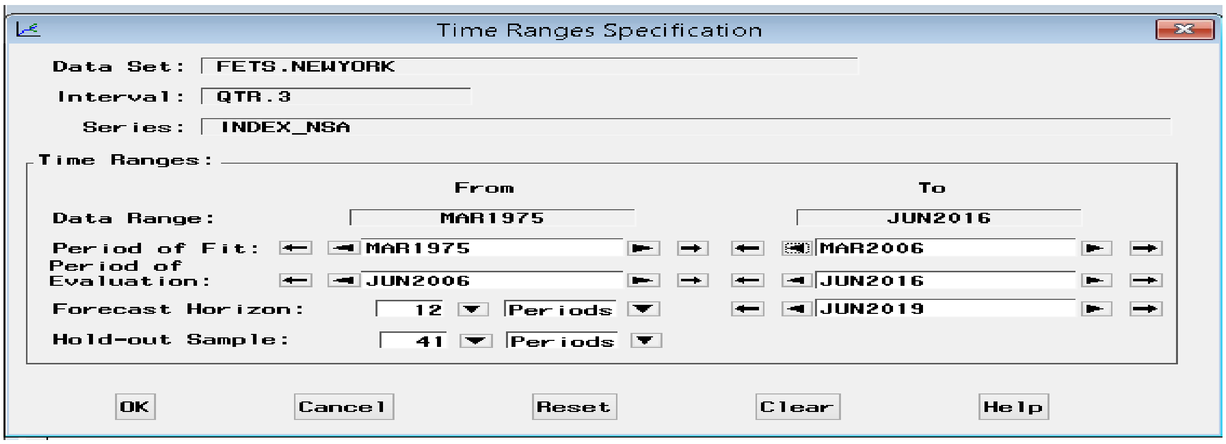
Applying first and seasonal differencing indicates that the error term is stationary which is evident from the unit and seasonal root tests. However, the white noise tests still indicate the presence of autocorrelation in residuals.



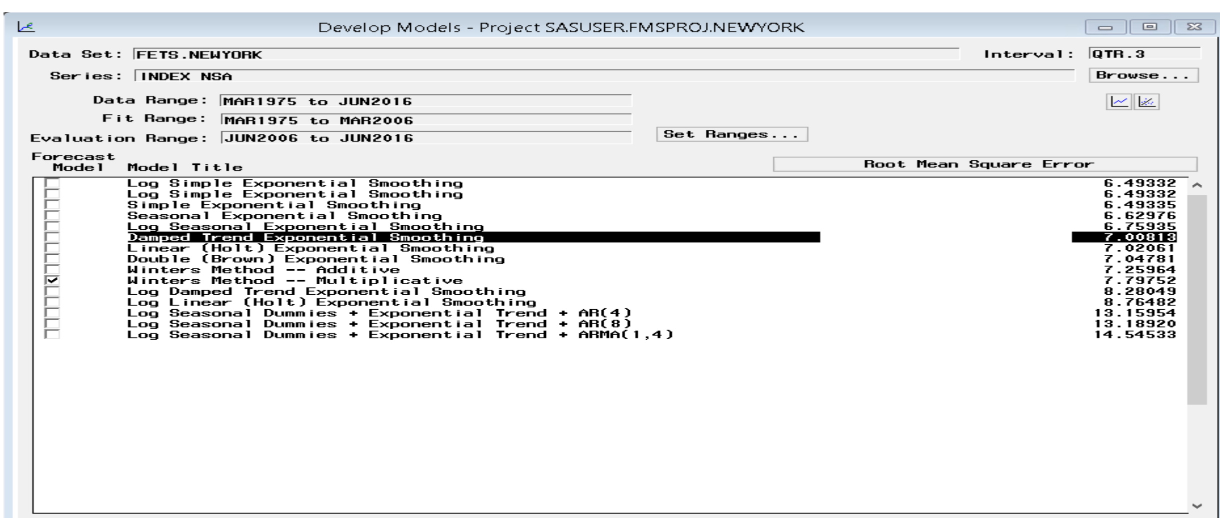
Models Built:

Since the time series has a strong trend component, we first tried building trend models and ARMA models to view the performance. However, the results did not turn out to be good indicating that the trend models are not the best fitting model. We tried to model the decaying weighted average of the past values and hence decided to try all the exponential smoothing models.

we set the hold-out sample by 25% of the whole data.



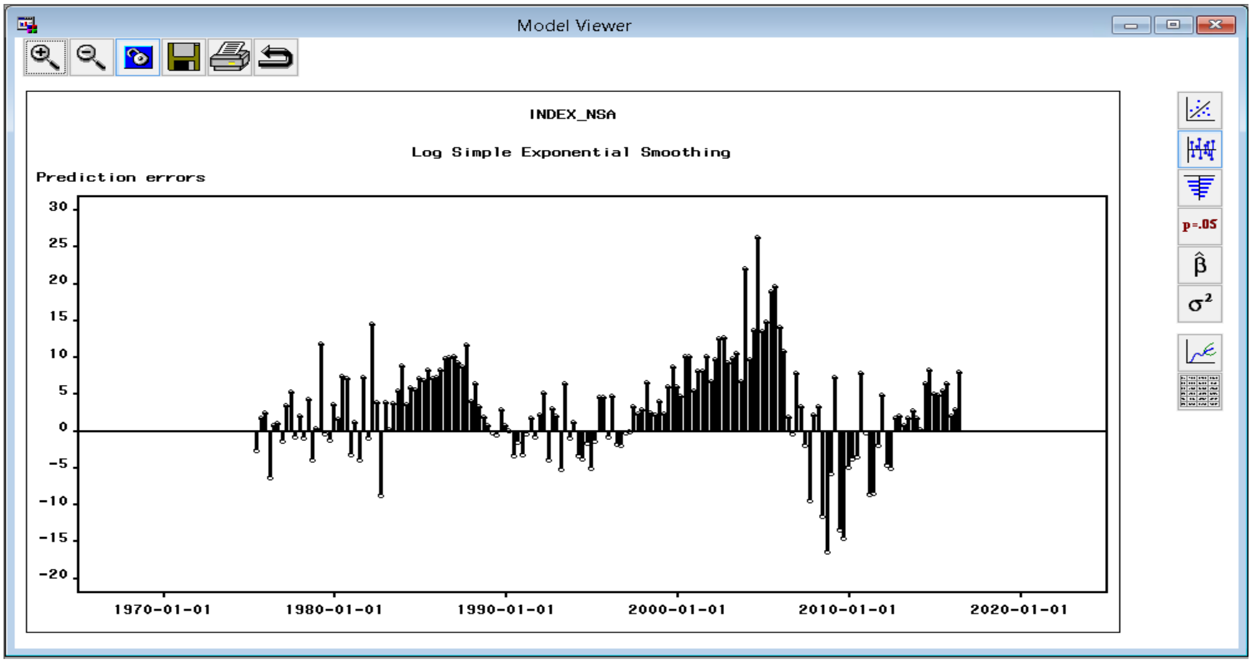
The following models were built and the Log Simple Exponential Smoothing Model had the lowest RMSE value. Further assessment of plots and model fit statistics for this model was carried out and we observed that it did not perform well as expected.



Log Simple Exponential Smoothing model

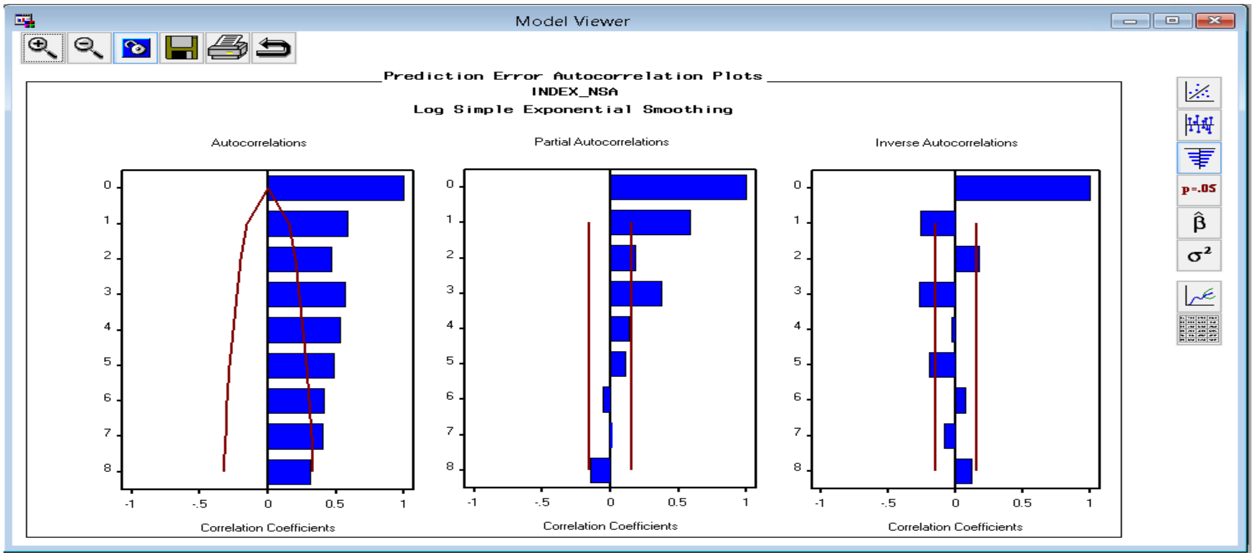
Interpretation from Prediction error plot

The residuals appear to be positive from 1980 to 2010.



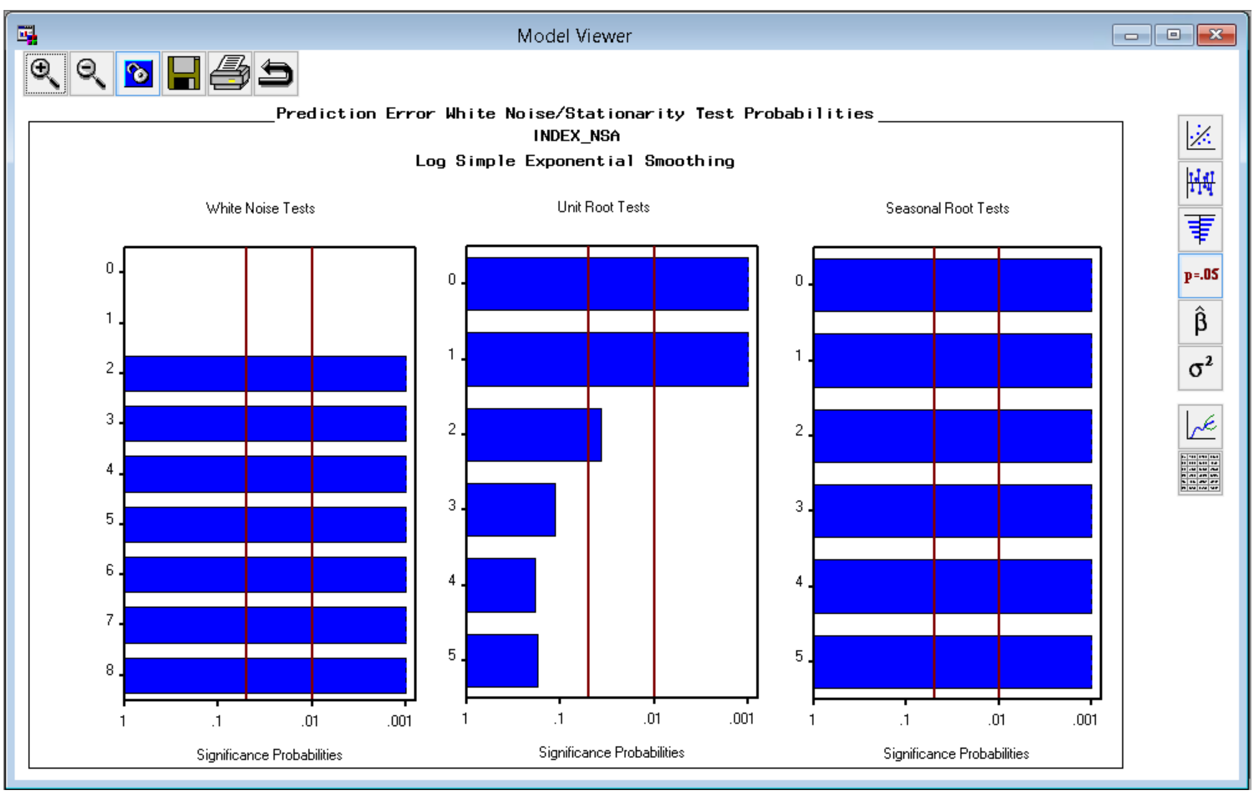
Inference from prediction error autocorrelation plots

Checking the autocorrelation plots of the residuals suggest that many spikes are significantly different from 0. We can infer that the model does not explain significant autocorrelation that was in original data.



Interpretation from Prediction Error Tests

The following plot clearly suggests that the residuals pass seasonal root tests, but not perform well in white noise test and unit root test.

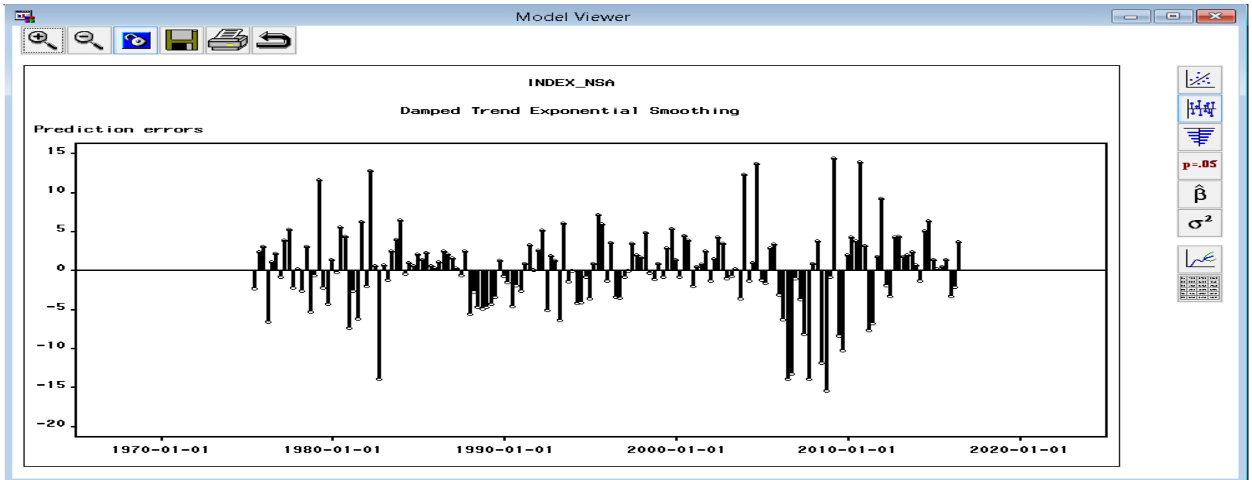


On analyzing further, the Damped Trend Exponential smoothing model performed better.

Further assessment of plots and model fit statistics for this model was carried out.

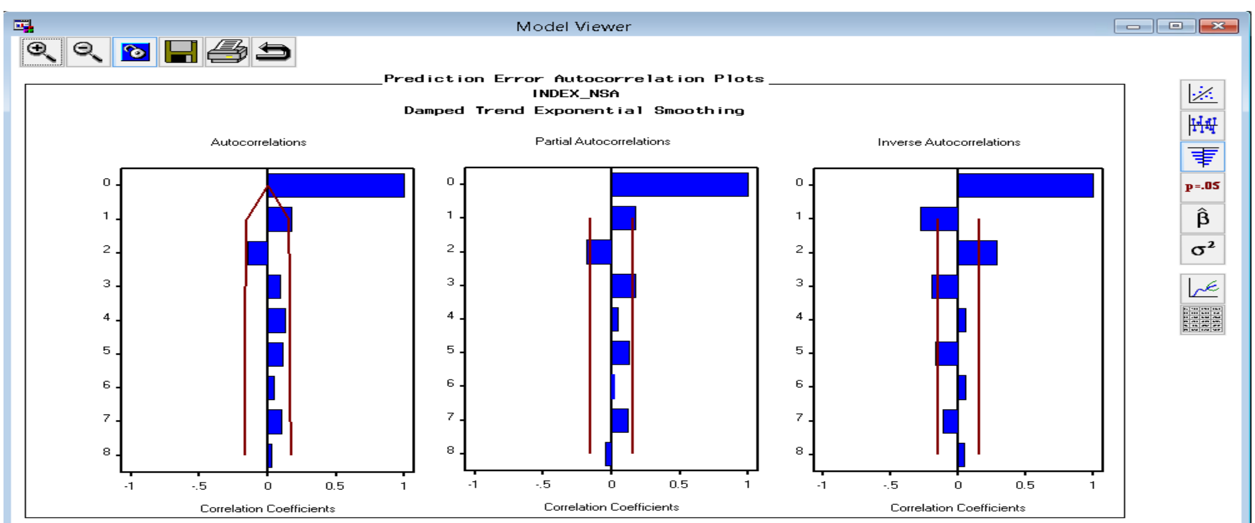
Interpretation from Prediction error plot

The residuals appear to be random and with some spikes around 2010



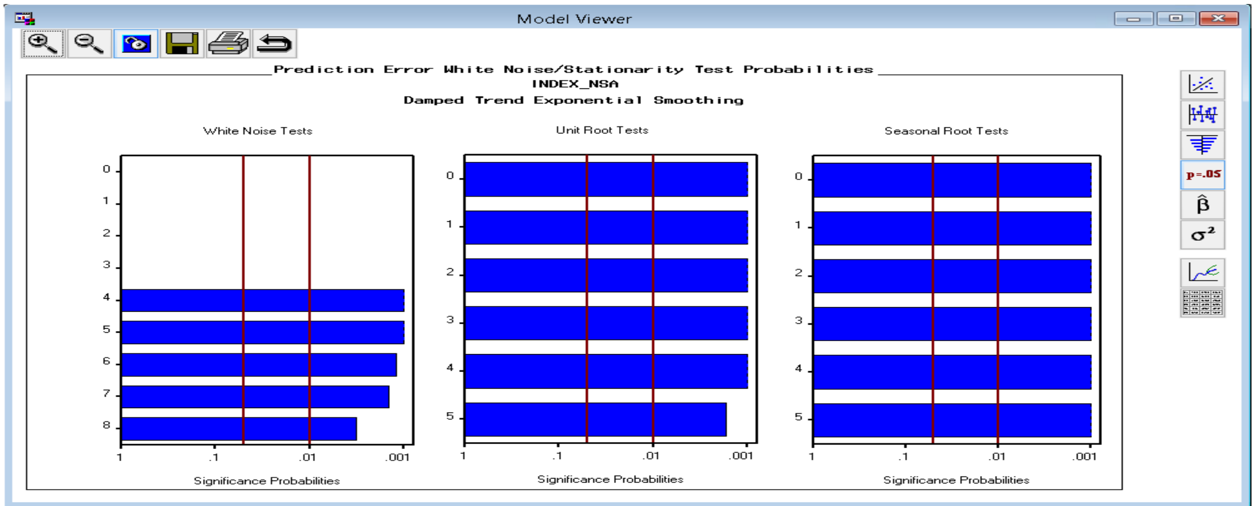
Inference from prediction error autocorrelation plots

Checking the autocorrelation plots of the residuals suggest that no spikes are significantly different from 0. We can infer that the model explains the significant autocorrelation that was in original data.



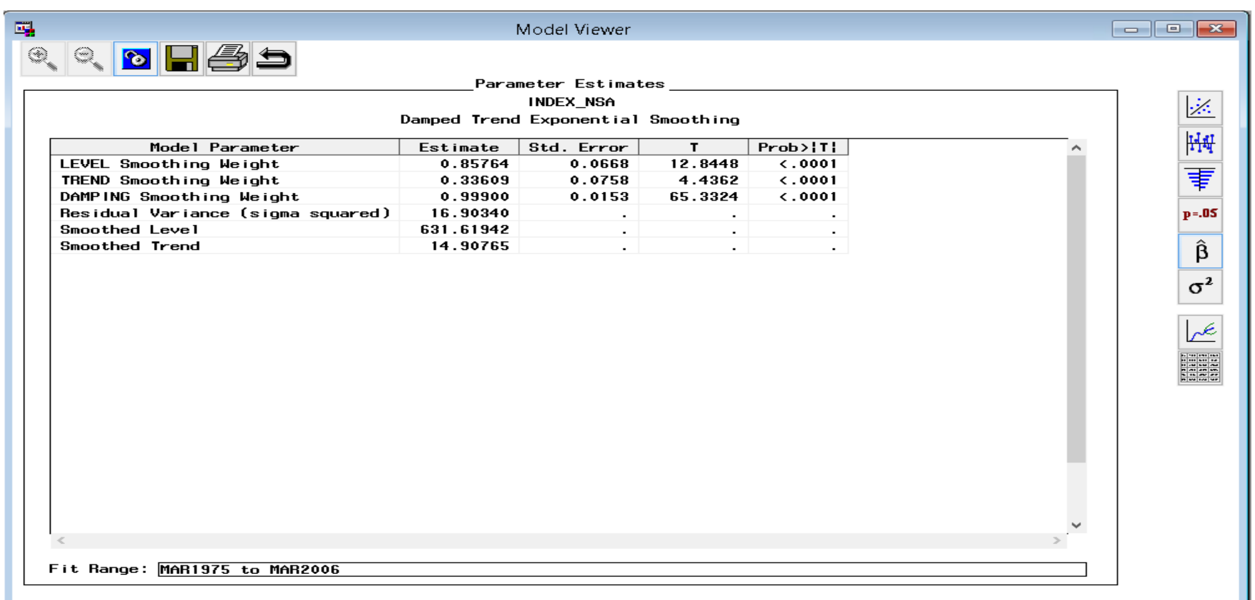
Interpretation from Prediction Error Tests

The following plot clearly suggests that the residuals pass seasonal root tests and unit root tests, but not perform well in white noise test after lag 4.



Inference from parameter estimates

The parameter estimates suggest that the level smoothing weight, trend smoothing weight and damping smoothing weight have low p-values which are significant at 1% level.


Model Viewer

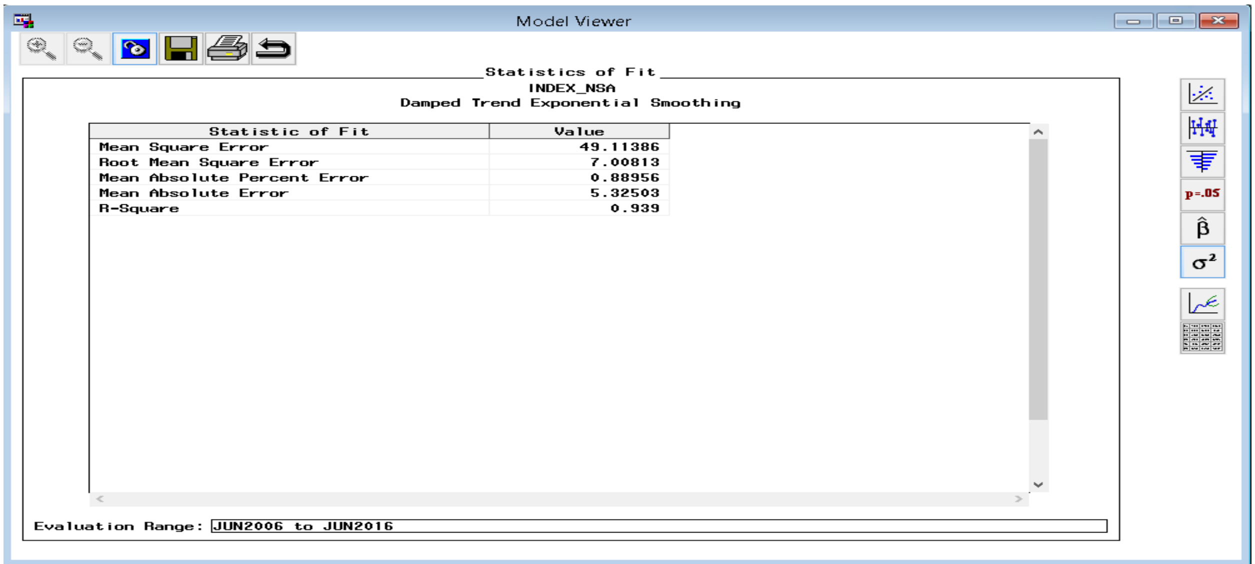
Parameter Estimates
INDEX_NSA
Damped Trend Exponential Smoothing

Model Parameter	Estimate	Std. Error	T	Prob>|T|
LEVEL Smoothing Weight	0.85764	0.0668	12.8448	<.0001
TREND Smoothing Weight	0.33609	0.0758	4.4362	<.0001
DAMPING Smoothing Weight	0.99900	0.0153	65.3324	<.0001
Residual Variance (sigma squared)	16.90340	.	.	.
Smoothed Level	631.61942	.	.	.
Smoothed Trend	14.90765	.	.	.

Fit Range: MAR1975 to MAR2006

Inference from Statistics of Fit

This model has a R-square of 0.939, which indicates that the model fits quite well. The other statistics has low values as well indicating that the model performs well.

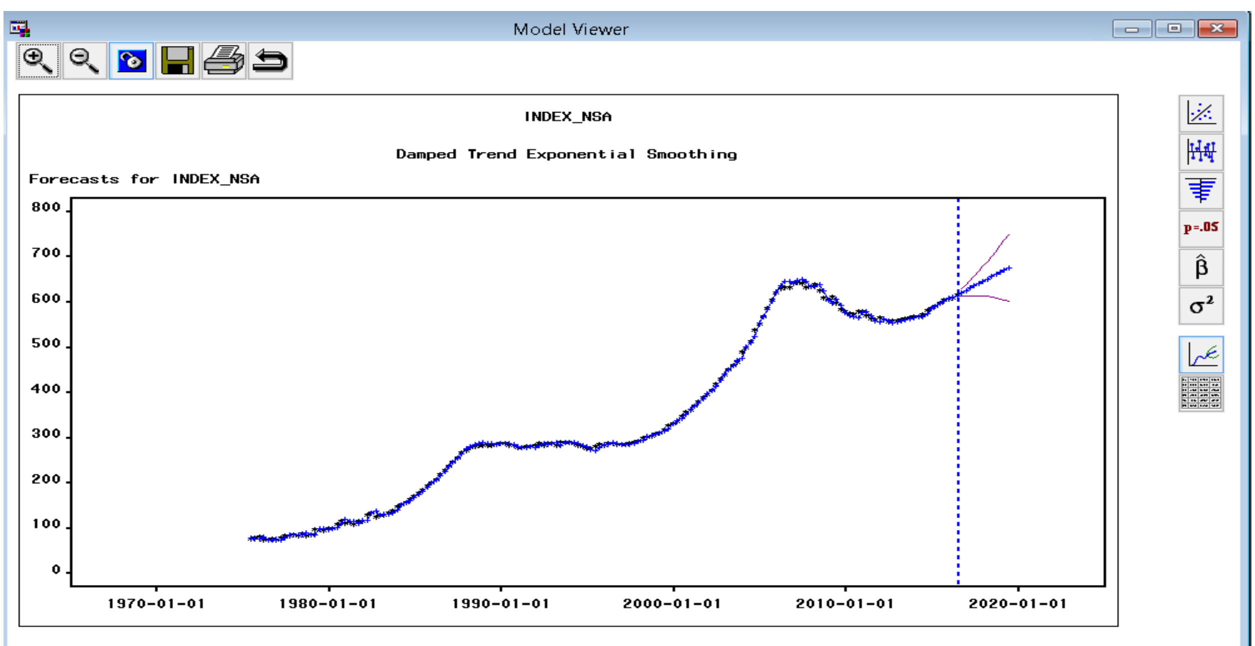

Model Viewer

Statistics of Fit
INDEX_NSA
Damped Trend Exponential Smoothing

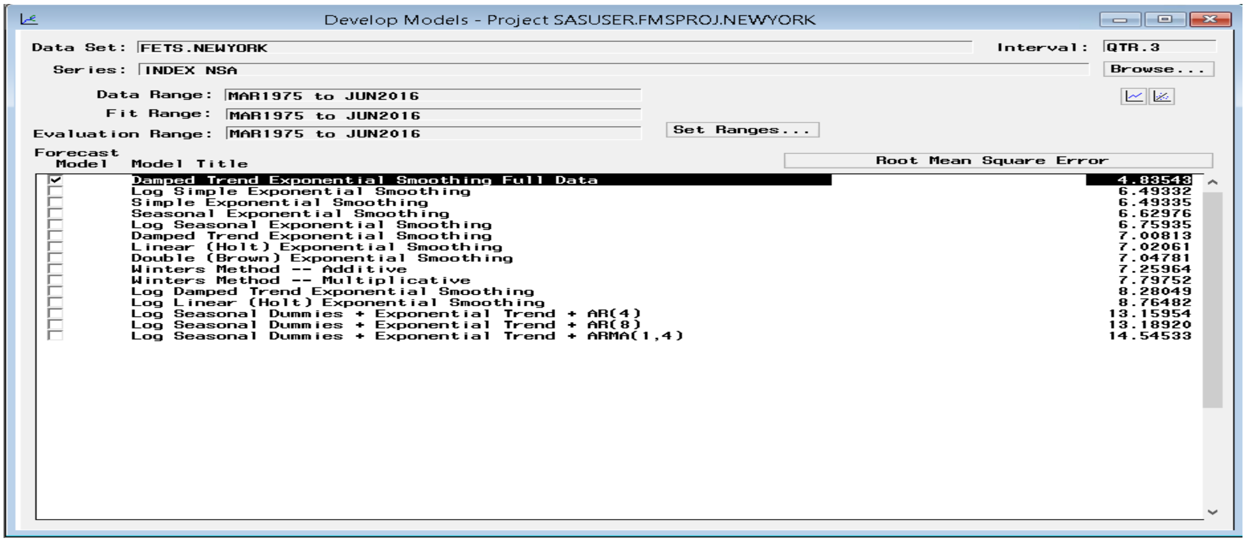
Statistic of Fit	Value
Mean Square Error	49.11386
Root Mean Square Error	7.00813
Mean Absolute Percent Error	0.88956
Mean Absolute Error	5.32503
R-Square	0.939

Evaluation Range: JUN2006 to JUN2016

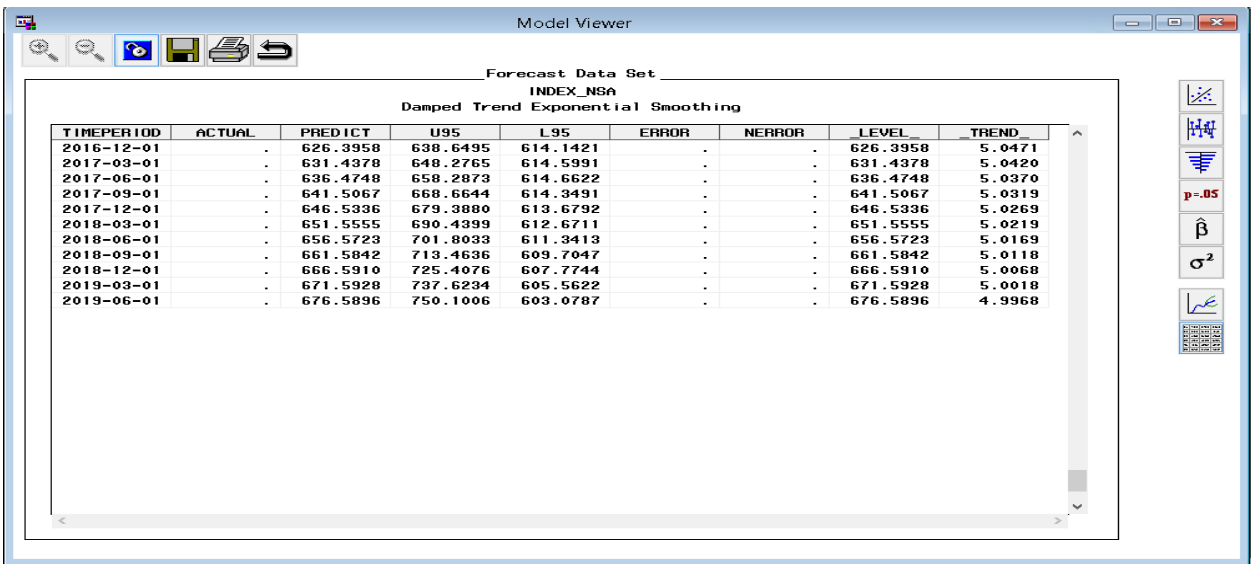
The following graph shows the forecast for the selected model with tight 95% confidence intervals indicated by pink line.



The same model was fit using the full data and results were forecasted.



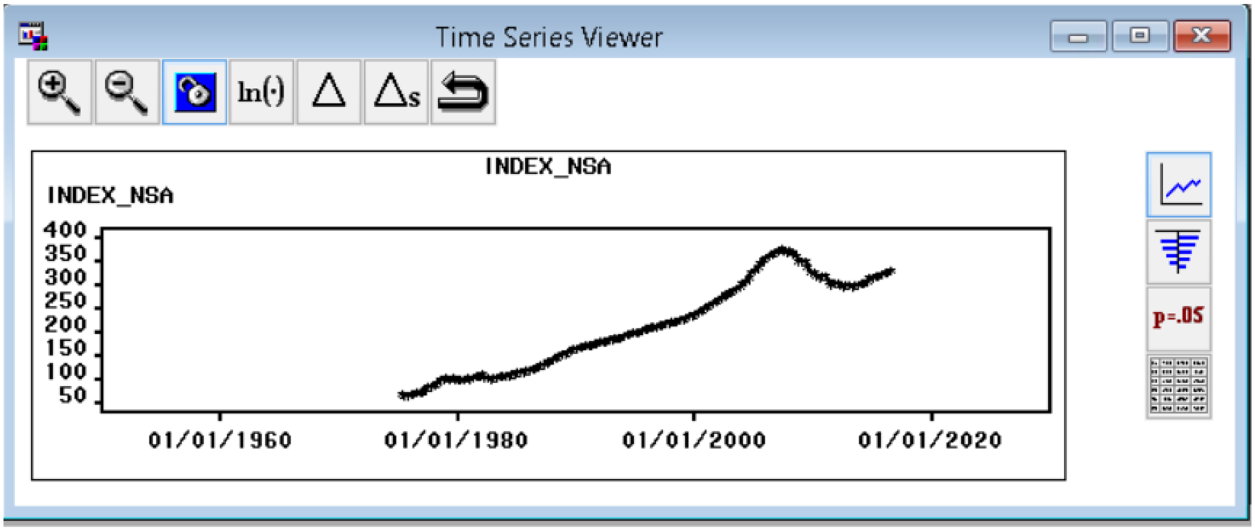
Forecast results:



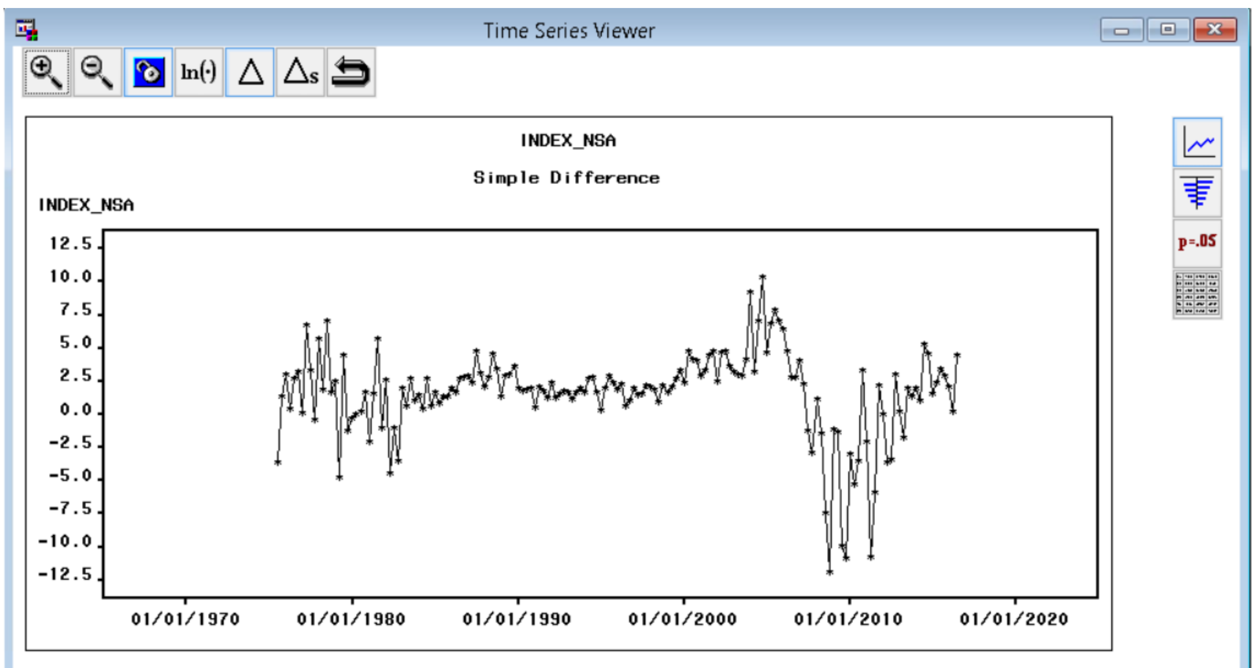
TIMEPERIOD	ACTUAL	PREDICT	U95	L95	ERROR	NERROR	_LEVEL_	_TREND_
2016-12-01	.	626.3958	638.6495	614.1421	.	.	626.3958	5.0471
2017-03-01	.	631.4378	648.2765	614.5991	.	.	631.4378	5.0420
2017-06-01	.	636.4748	658.2873	614.6622	.	.	636.4748	5.0370
2017-09-01	.	641.5067	668.6644	614.3491	.	.	641.5067	5.0319
2017-12-01	.	646.5336	679.3880	613.6792	.	.	646.5336	5.0269
2018-03-01	.	651.5555	690.4399	612.6711	.	.	651.5555	5.0219
2018-06-01	.	656.5723	701.8033	611.3413	.	.	656.5723	5.0169
2018-09-01	.	661.5842	713.4636	609.7047	.	.	661.5842	5.0118
2018-12-01	.	666.5910	725.4076	607.7744	.	.	666.5910	5.0068
2019-03-01	.	671.5928	737.6234	605.5622	.	.	671.5928	5.0018
2019-06-01	.	676.5896	750.1006	603.0787	.	.	676.5896	4.9968

Illinois

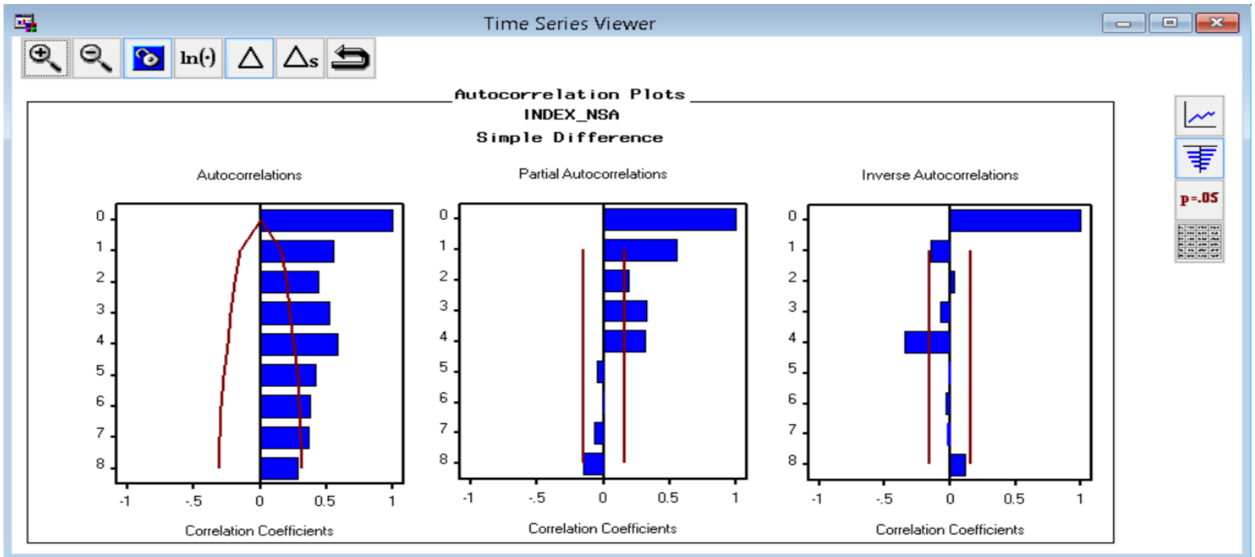
The following graph shows the quarterly trend of house price index in the state Illinois.



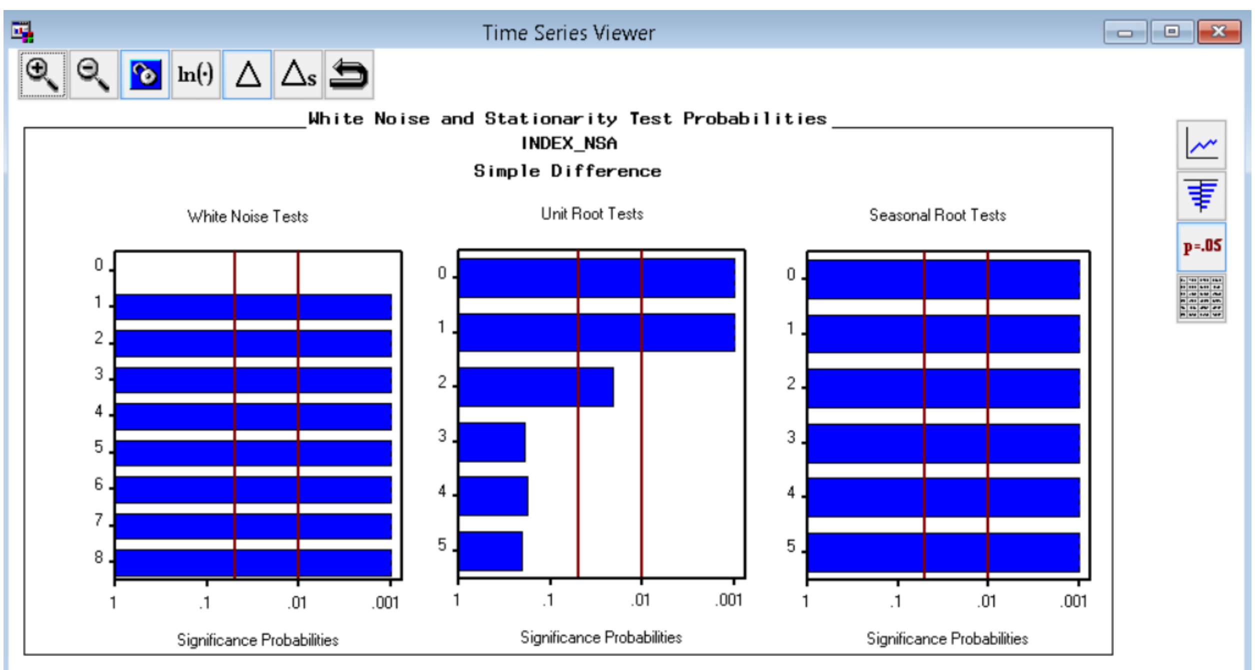
After differencing the data once, the series is almost closer to a stationary series except for the years from 2009 - 2010.



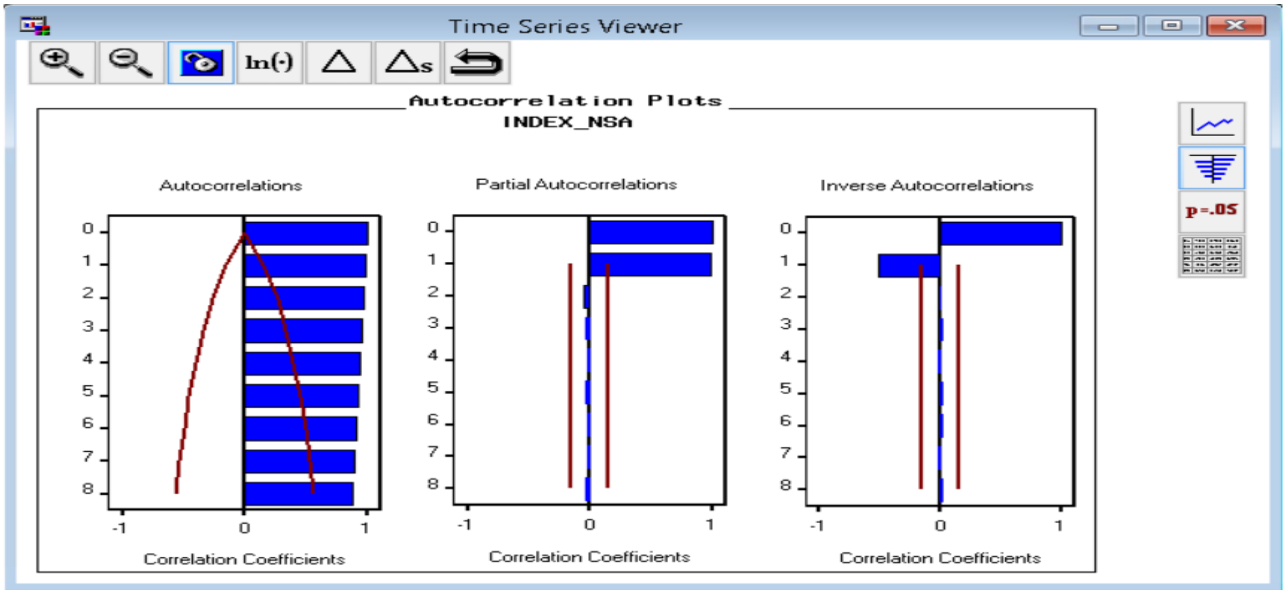
The autocorrelation plots indicate a strong trend component in the time series with slowly decaying Autocorrelation and many significant lags. AR component seems to be significant till the fourth lag.



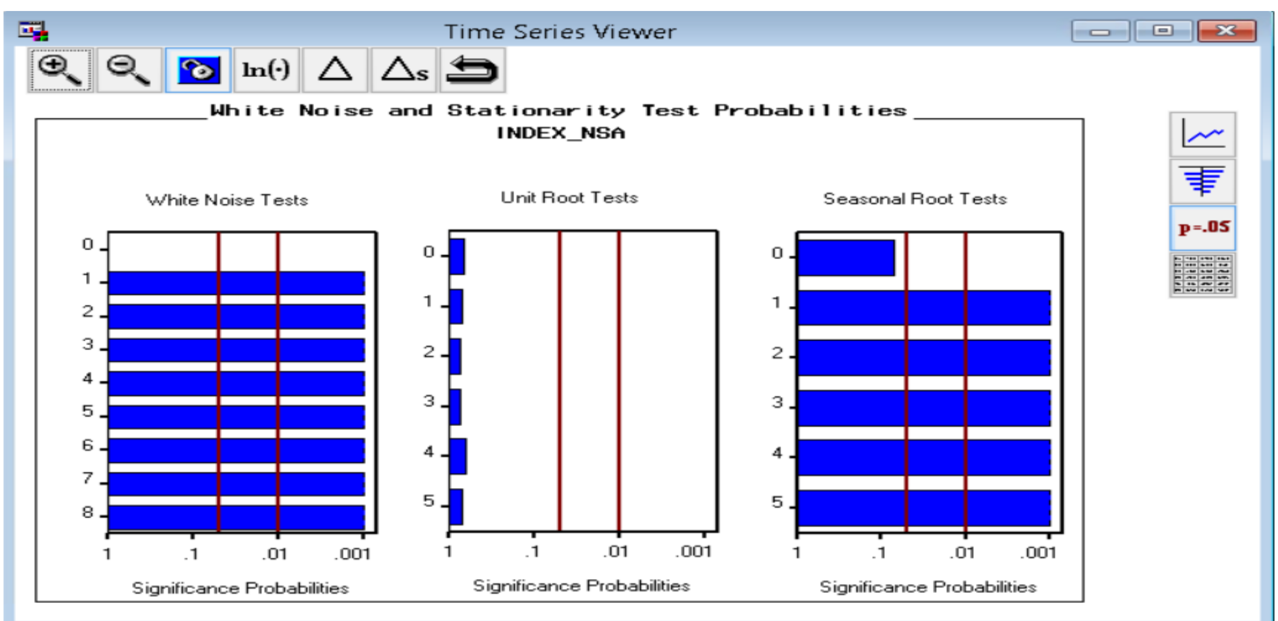
The unit root test strongly indicates the presence of trend component and need for first differencing. The seasonal root test indicates that there is no seasonal component. However, the time series is clearly not stationary which is evident from the unit root tests.



After first differencing, AR (1) component seems to be the significant lag and the Autocorrelation plot seems to decay more uniformly.



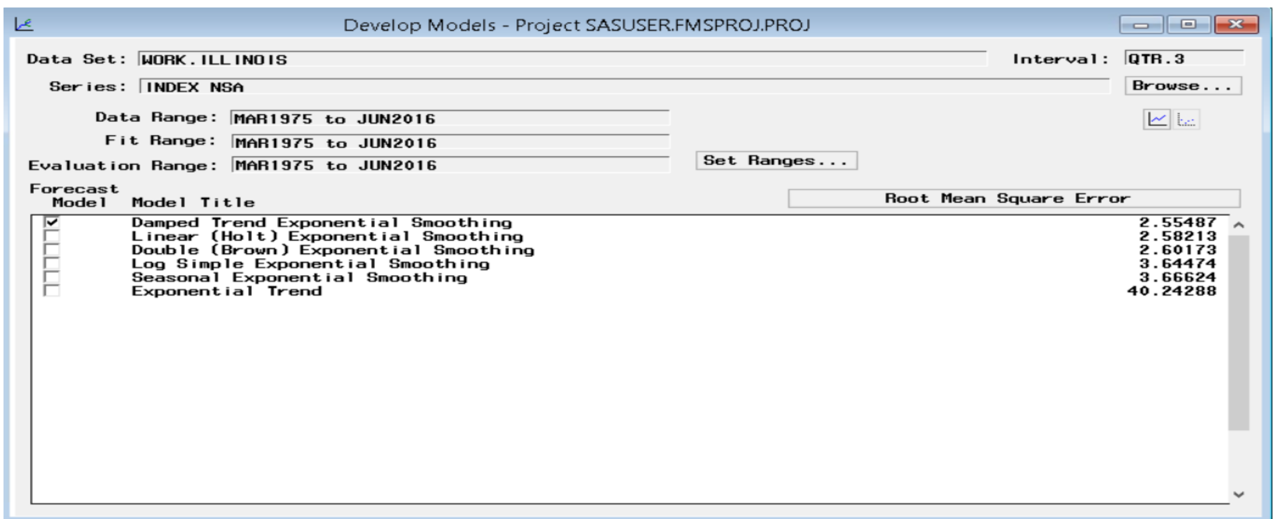
Even after differencing, the trend component seems to exist and the seasonal component seems to have been introduced. However, the white noise tests still indicate the presence of autocorrelation in residuals.



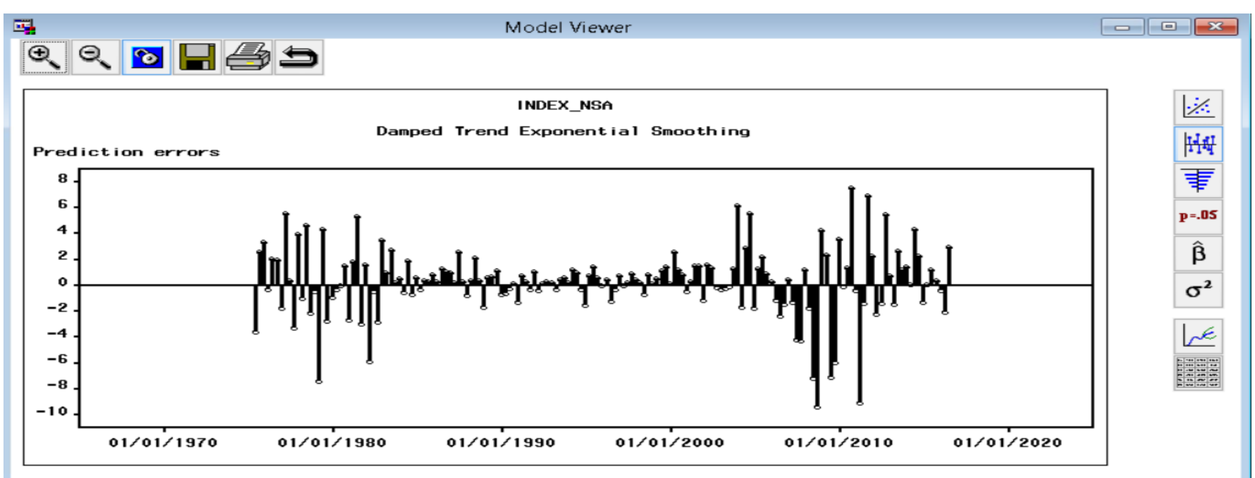
Models:

Since the time series has a strong trend component and the data ranges from the year 1975, we first tried building trend models to view the performance. However, the results did not turn out to be good indicating that the trend models is not the best fitting model. We tried to model the decaying weighted average of the past values and hence decided to try all the exponential smoothing models. The resulting RMSE values significantly improved and the Damped Trend

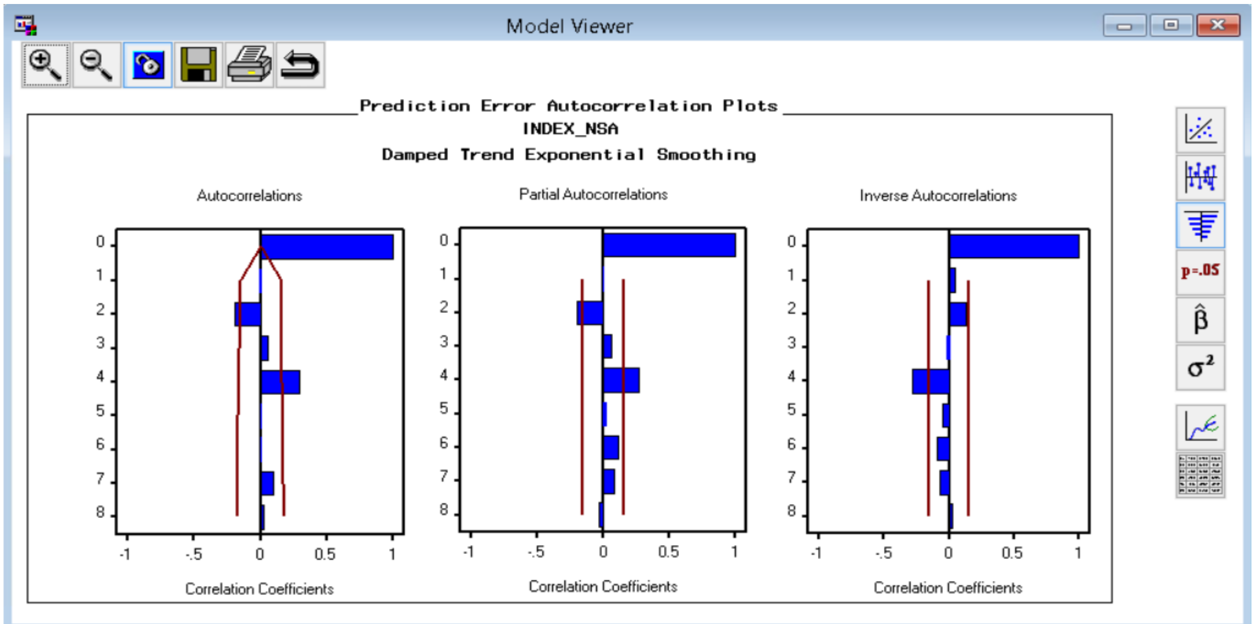
Exponential Smoothing Model was the best fitting model. Further assessment of plots and model fit statistics for this model was carried out.



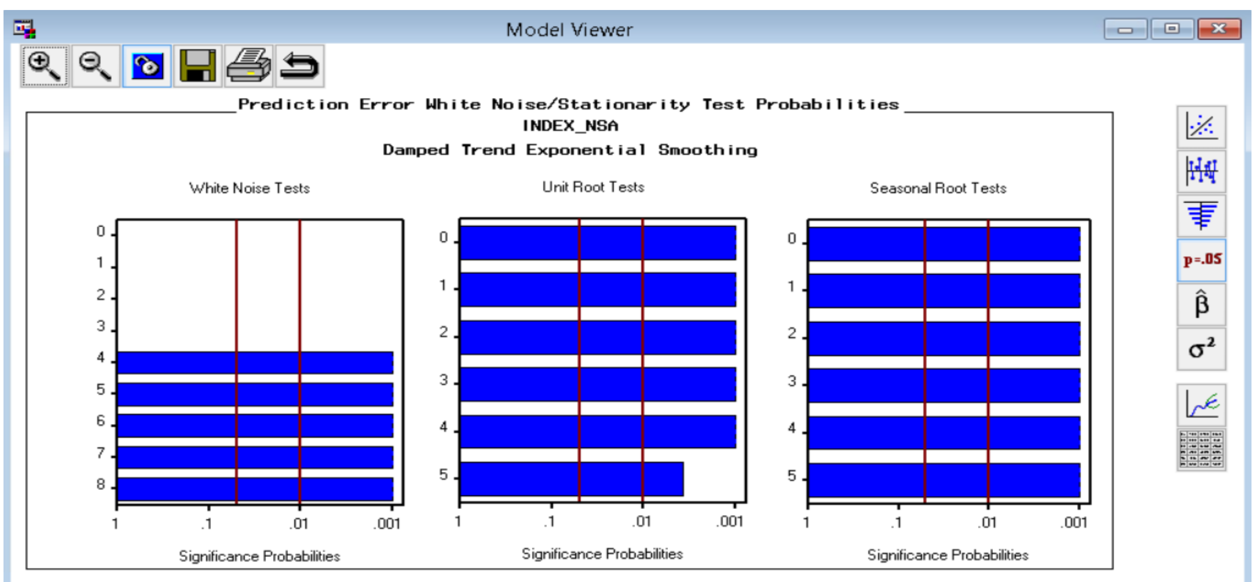
There seems to be spikes in residuals in years - 1978, 2009, 2010 and 2011.



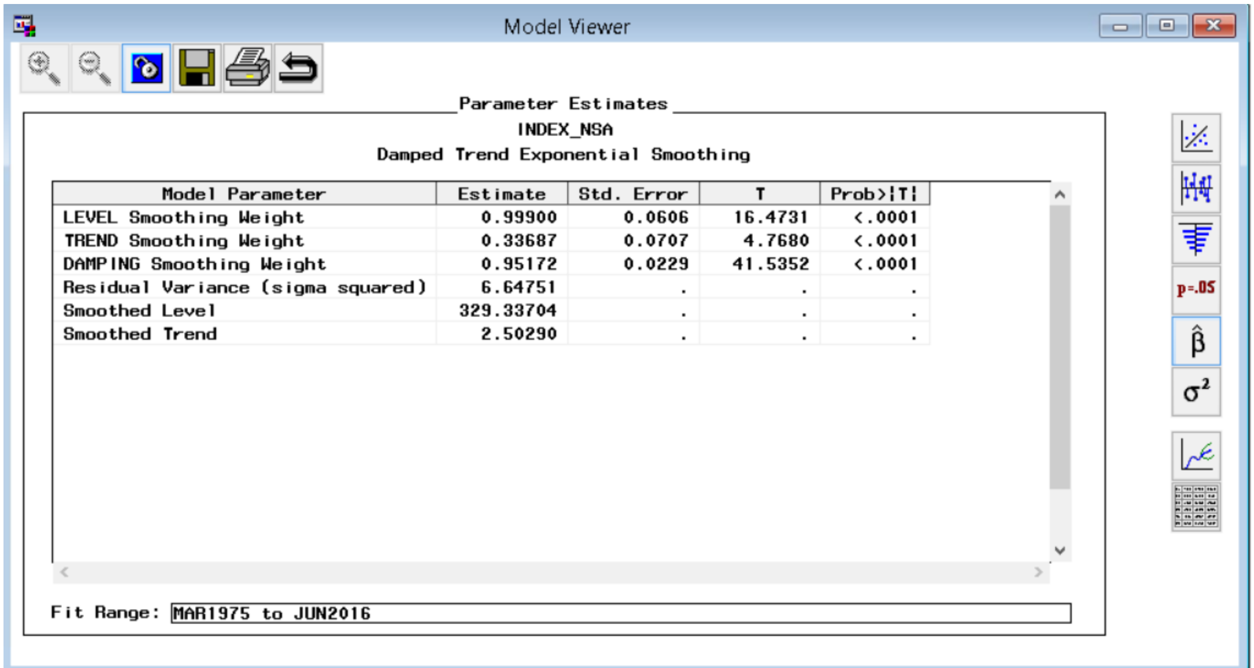
Checking the autocorrelation plots of the residuals suggest that none of the spikes are significantly different from 0. We can infer that the model explains all the significant autocorrelation that was in original data.



The following plot clearly suggests that the residuals pass the unit and seasonal root tests, and the white noise test.



The parameter estimates suggest that all the parameters are significant with low p-value.


Model Viewer

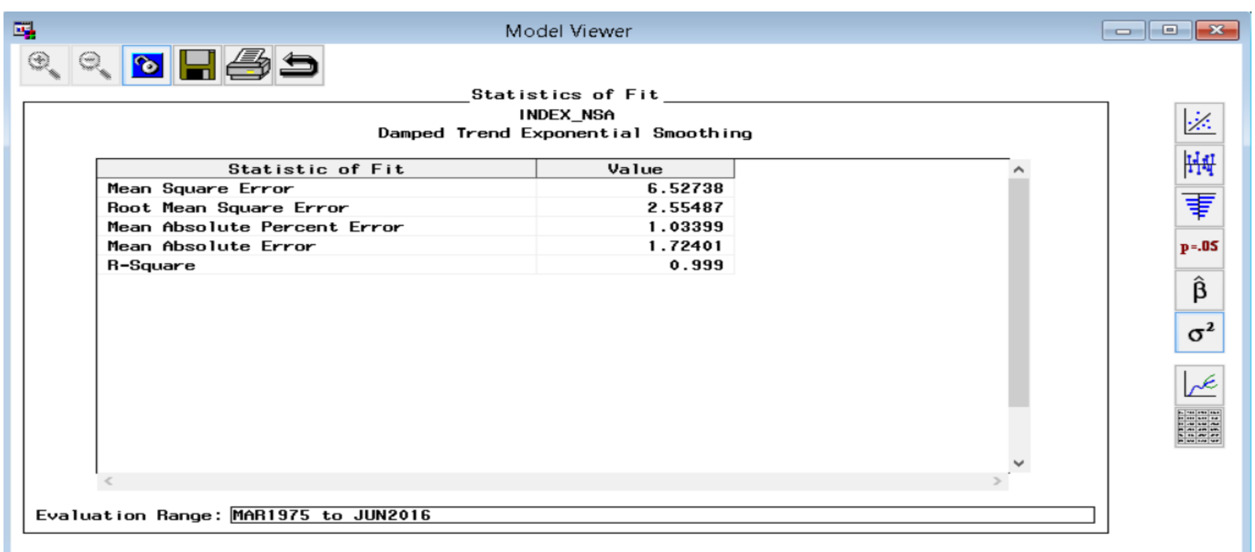
Parameter Estimates
INDEX_NSA
Damped Trend Exponential Smoothing

Model Parameter	Estimate	Std. Error	T	Prob>|T|
LEVEL Smoothing Weight	0.99900	0.0606	16.4731	<.0001
TREND Smoothing Weight	0.33687	0.0707	4.7680	<.0001
DAMPING Smoothing Weight	0.95172	0.0229	41.5352	<.0001
Residual Variance (sigma squared)	6.64751	.	.	.
Smoothed Level	329.33704	.	.	.
Smoothed Trend	2.50290	.	.	.

Fit Range: MAR1975 to JUN2016

Inference from Statistics of Fit

This model produces the lowest SBC, RMSE, MAPE and MAD compared to other exponential smoothing models.

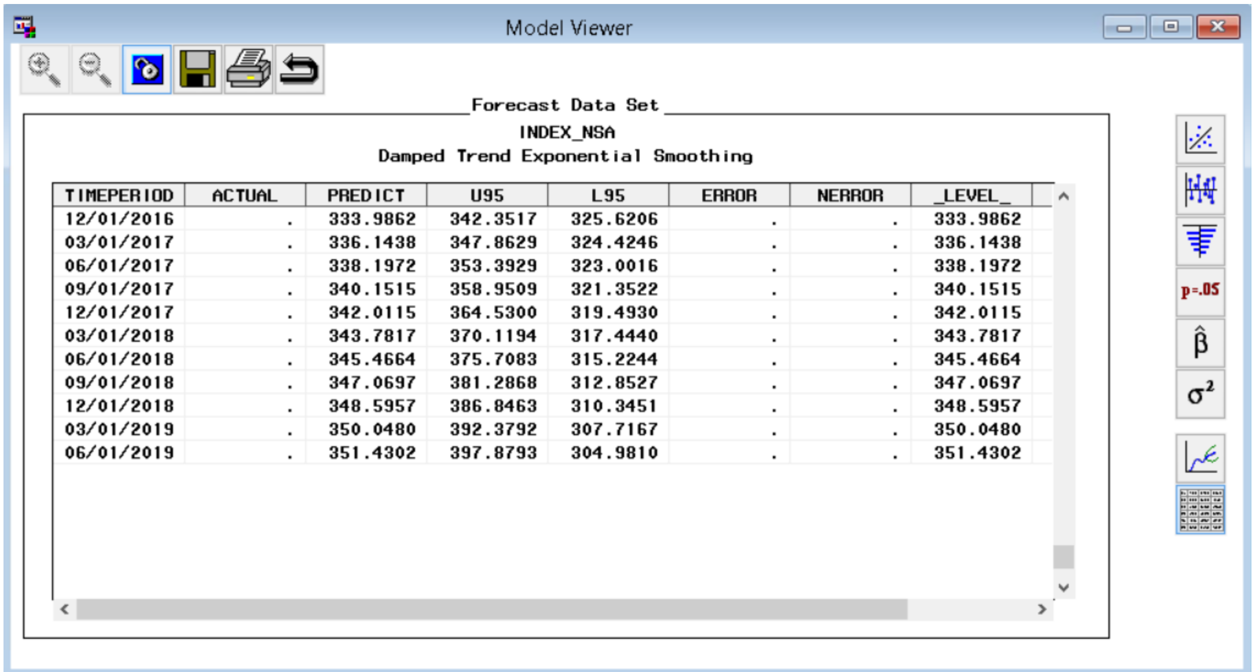

Model Viewer

Statistics of Fit
INDEX_NSA
Damped Trend Exponential Smoothing

Statistic of Fit	Value
Mean Square Error	6.52738
Root Mean Square Error	2.55487
Mean Absolute Percent Error	1.03399
Mean Absolute Error	1.72401
R-Square	0.999

Evaluation Range: MAR1975 to JUN2016

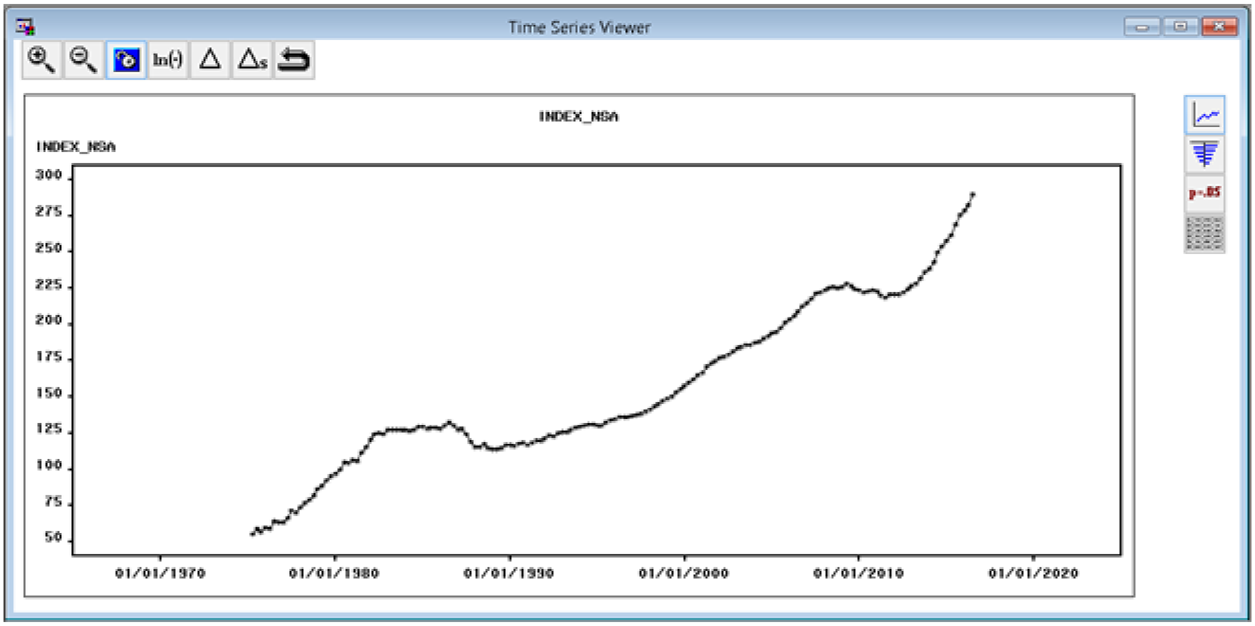
Forecast Values



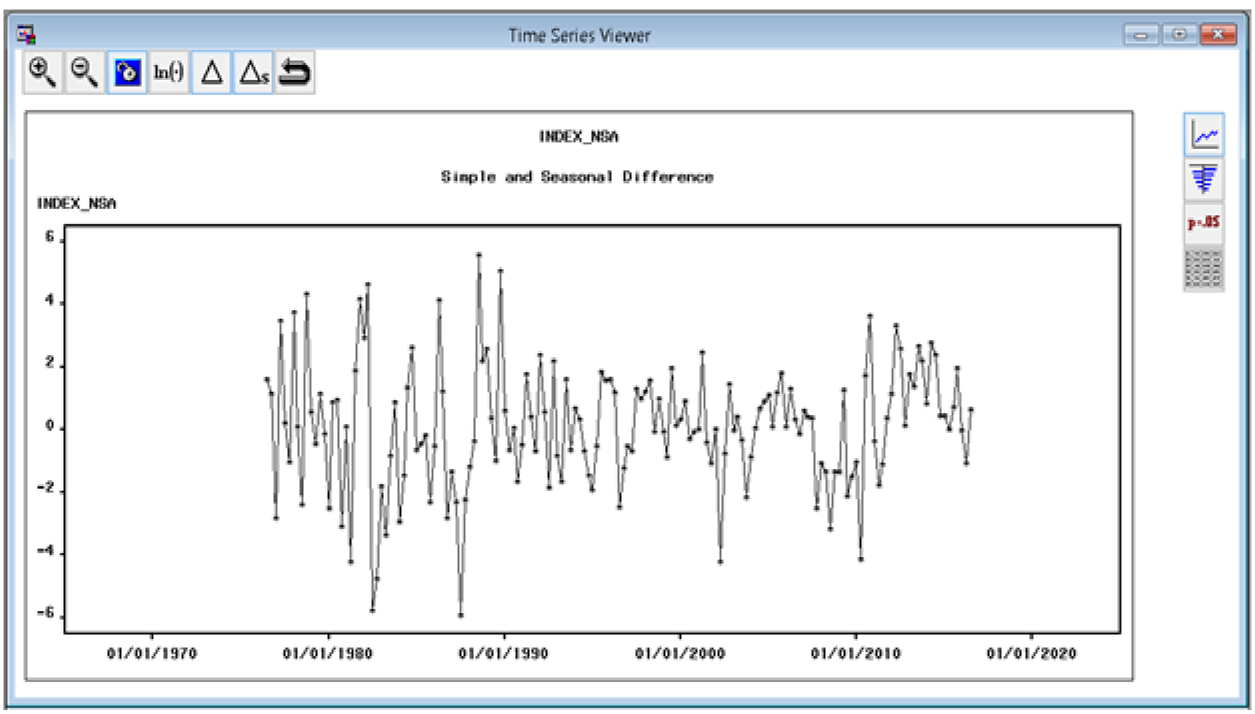
TIMEPERIOD	ACTUAL	PREDICT	U95	L95	ERROR	NERROR	_LEVEL_
12/01/2016	.	333.9862	342.3517	325.6206	.	.	333.9862
03/01/2017	.	336.1438	347.8629	324.4246	.	.	336.1438
06/01/2017	.	338.1972	353.3929	323.0016	.	.	338.1972
09/01/2017	.	340.1515	358.9509	321.3522	.	.	340.1515
12/01/2017	.	342.0115	364.5300	319.4930	.	.	342.0115
03/01/2018	.	343.7817	370.1194	317.4440	.	.	343.7817
06/01/2018	.	345.4664	375.7083	315.2244	.	.	345.4664
09/01/2018	.	347.0697	381.2868	312.8527	.	.	347.0697
12/01/2018	.	348.5957	386.8463	310.3451	.	.	348.5957
03/01/2019	.	350.0480	392.3792	307.7167	.	.	350.0480
06/01/2019	.	351.4302	397.8793	304.9810	.	.	351.4302

Texas

The following graph shows the quarterly trend of house price index in the state Texas.

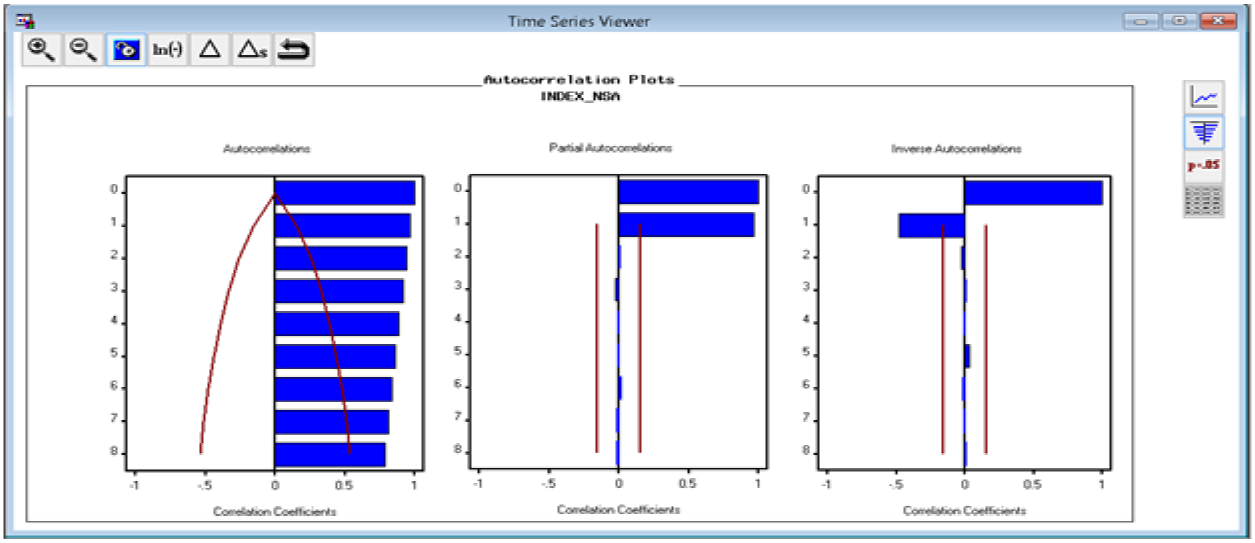


After applying first and seasonal differencing, we observed a flattened-out series.

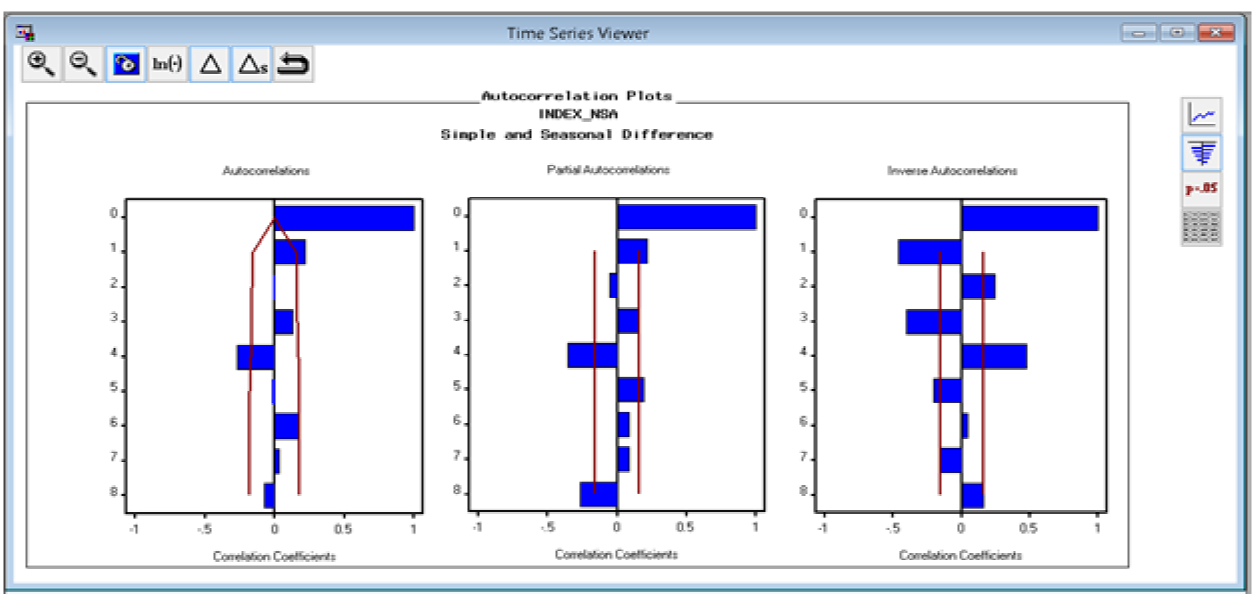


Inference from prediction error autocorrelation plots

The autocorrelation plots indicate a strong trend component in the time series with evidence of significant number of slowly decaying ACF and many significant lags. ACF (1) is large compared to the subsequent lags.

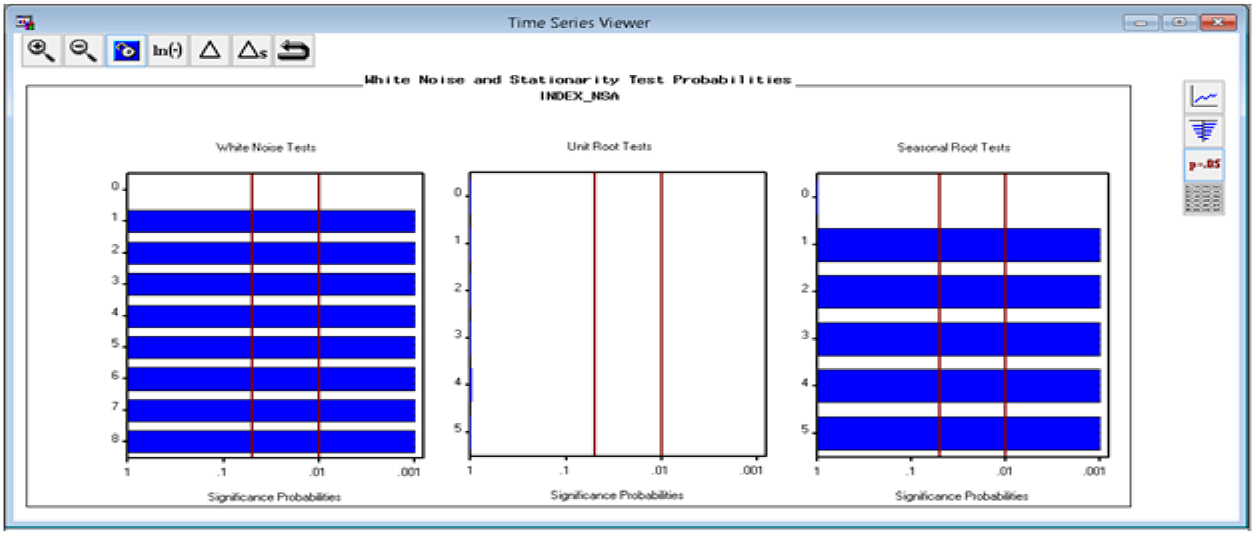


After first and seasonal differencing, ACF (1) is still significantly different from 0 but smaller than the model without differencing.

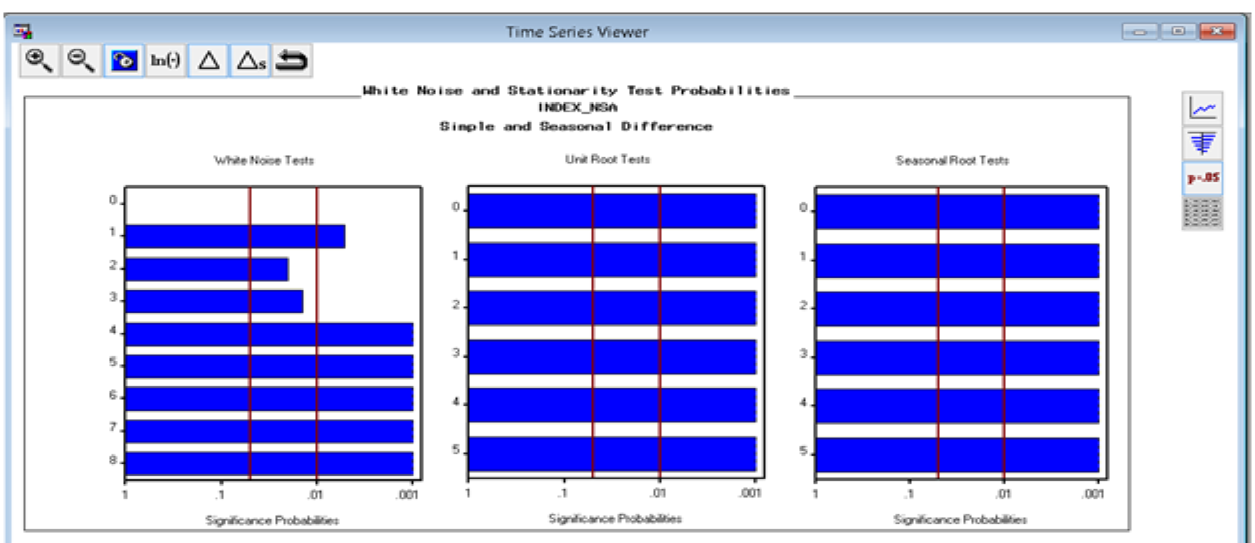


Interpretation from Prediction Error Tests

Dickey fuller unit root test strongly indicates the presence of trend component and need for first differencing. The seasonal root test indicates the presence of seasonal component. However, the time series is clearly not stationary which is evident from the unit root tests.

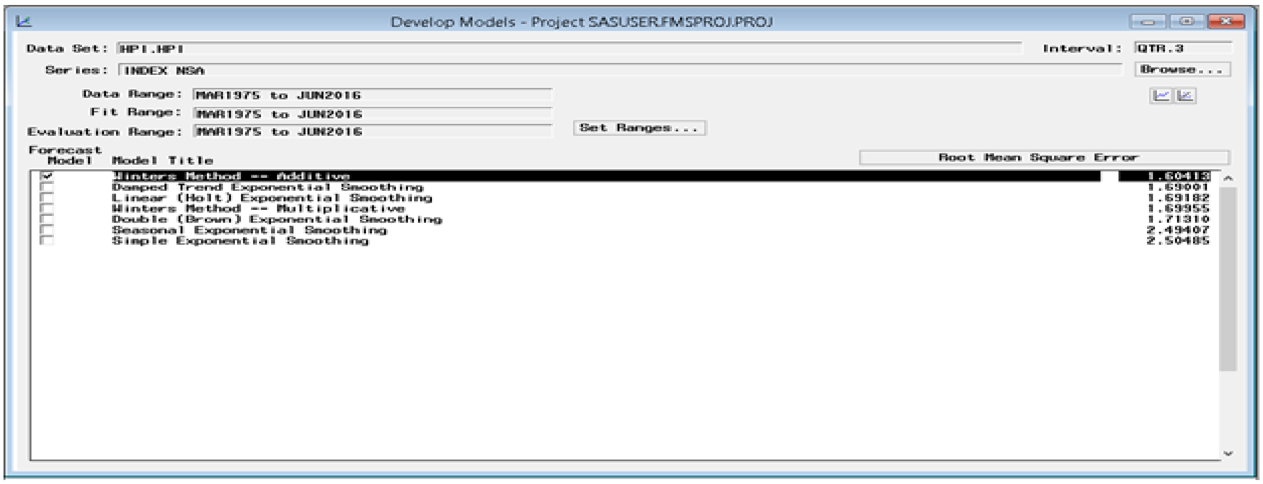


Applying first and seasonal differencing indicates that the error term is stationary which is evident from the unit and seasonal root tests. However, the white noise tests still indicate the presence of autocorrelation in residuals.



Models

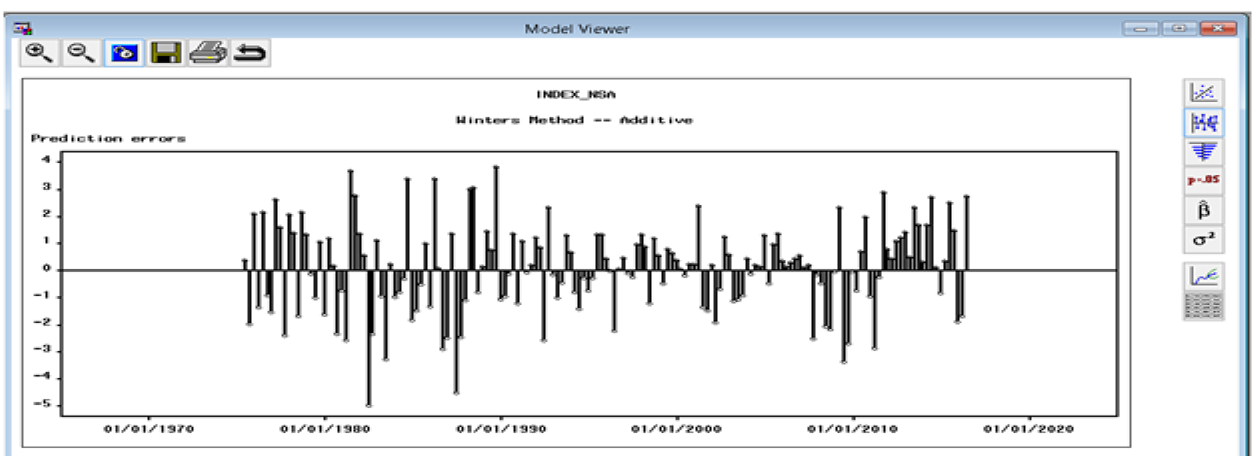
Since the time series has both trend and seasonal components and the data ranges from the year 1975, we wanted to model the decaying weighted average of the past values and hence decided to try all the exponential smoothing models. We observed that Winters Method – Additive exponential smoothing model had the lowest RMSE value. Further assessment of plots and model fit statistics for this model was carried out.



Winters – Additive model

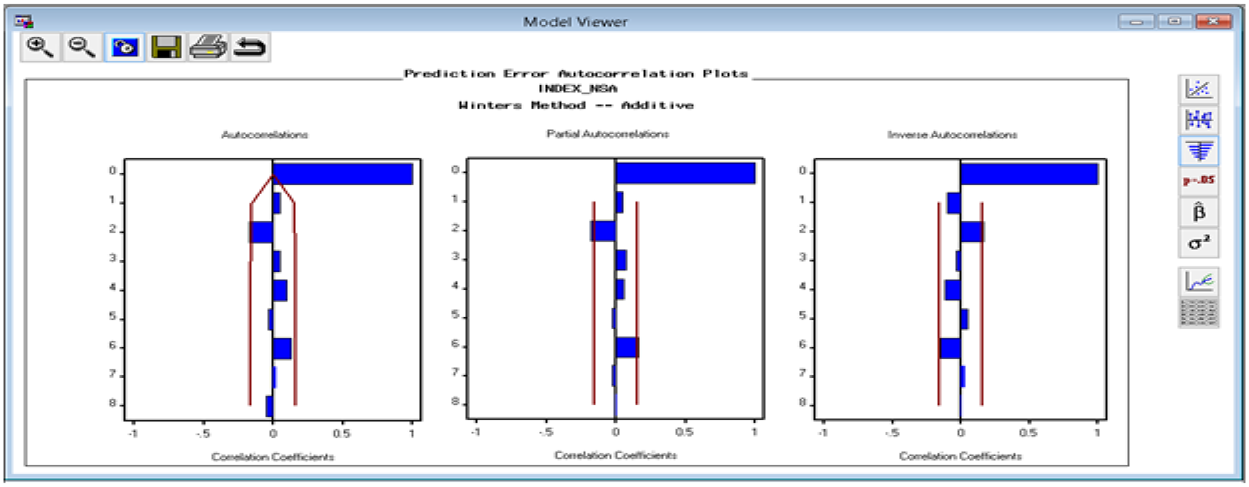
Interpretation from Prediction error plot

The residuals appear to be random with some strong spikes in the beginning and end of 80’s.



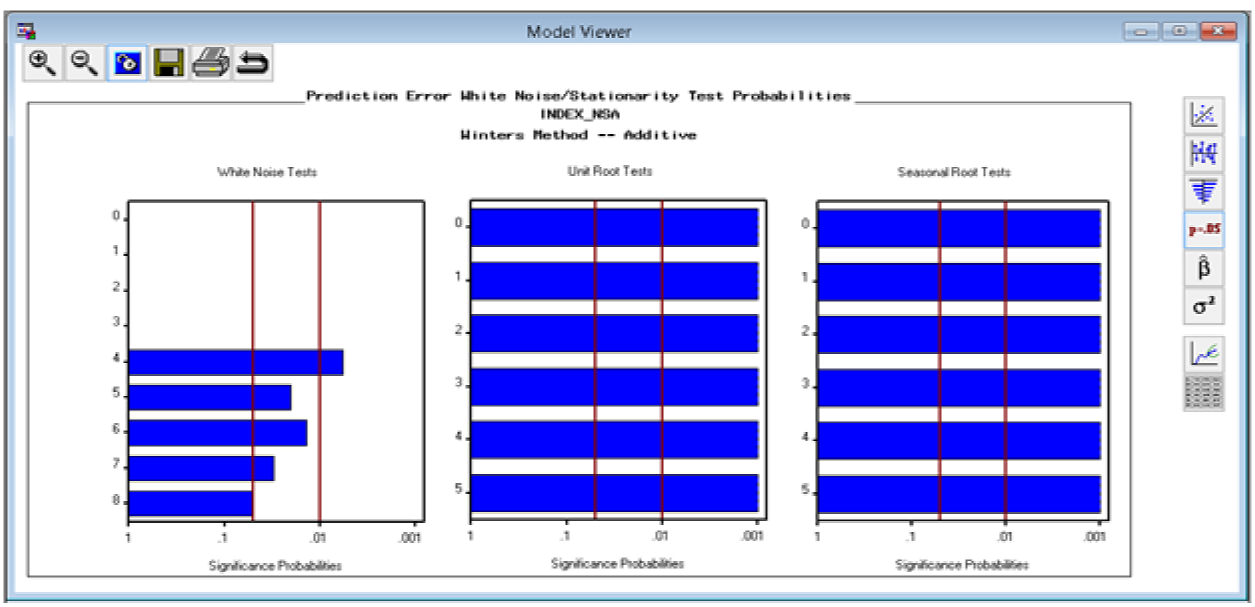
Inference from prediction error autocorrelation plots

Checking the autocorrelation plots of the residuals suggest that none of the spikes are significantly different from 0. We can infer that the model explains all the significant autocorrelation that was in original data.



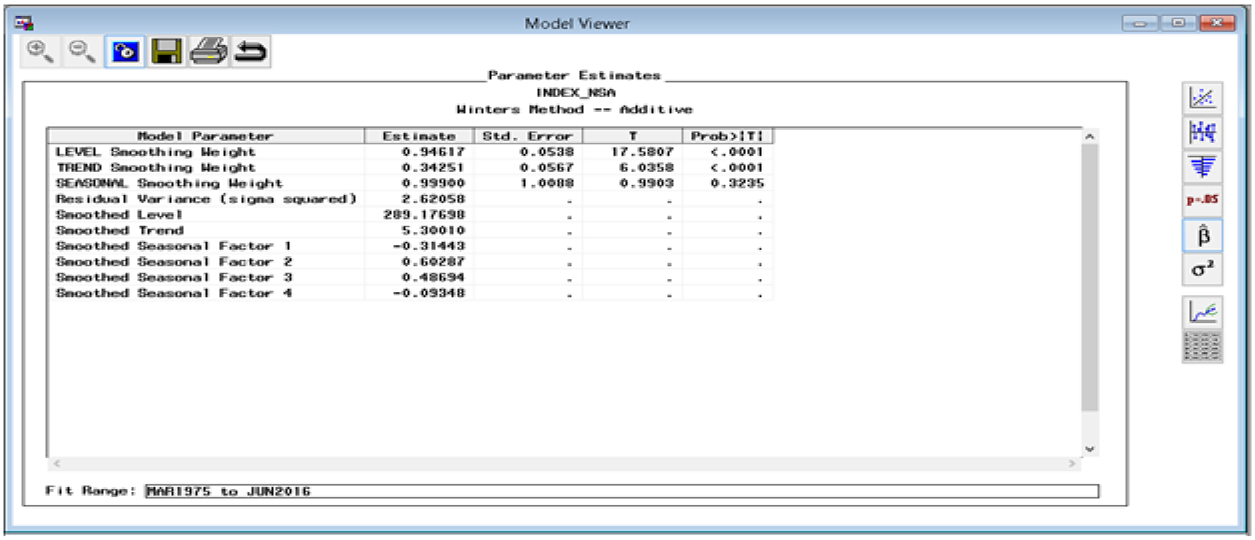
Interpretation from Prediction Error Tests

The following plot clearly suggests that the residuals pass the unit and seasonal root tests, and the white noise test at 5% significance level.



Inference from parameter estimates

The parameter estimates suggest that the seasonal smoothing weight has a high p-value which is not significant at 5% level.

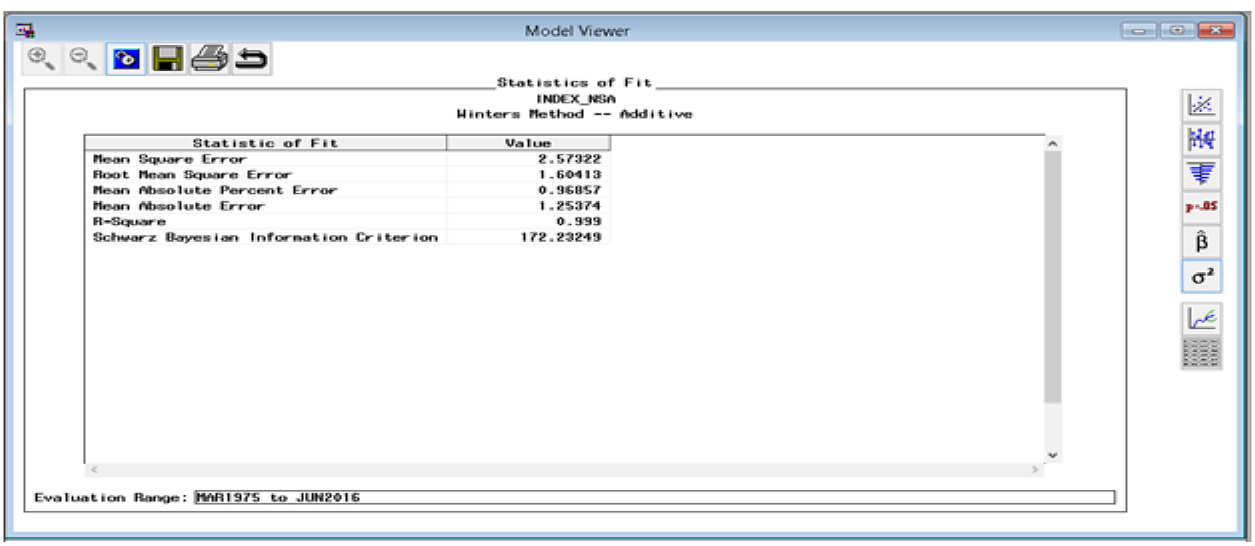


Model Parameter	Estimate	Std. Error	T	Prob>|T|
LEVEL Smoothing Weight	0.94617	0.0538	17.5807	<.0001
TREND Smoothing Weight	0.34251	0.0567	6.0358	<.0001
SEASONAL Smoothing Weight	0.99900	1.0088	0.9903	0.3235
Residual Variance (sigma squared)	2.62058	.	.	.
Smoothed Level	289.17698	.	.	.
Smoothed Trend	5.30010	.	.	.
Smoothed Seasonal Factor 1	-0.31443	.	.	.
Smoothed Seasonal Factor 2	0.60287	.	.	.
Smoothed Seasonal Factor 3	0.48694	.	.	.
Smoothed Seasonal Factor 4	-0.09348	.	.	.

Fit Range: MAR1975 to JUN2016

Inference from Statistics of Fit

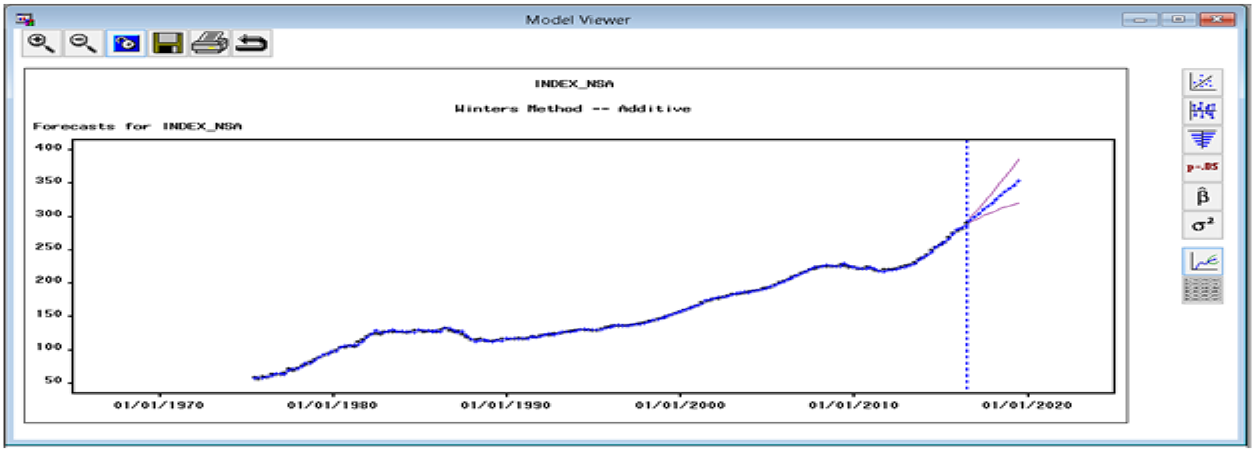
This model produces the lowest SBC, RMSE, MAPE and MAD compared to other exponential smoothing models.



Statistic of Fit	Value
Mean Square Error	2.57322
Root Mean Square Error	1.60413
Mean Absolute Percent Error	0.96857
Mean Absolute Error	1.25374
R-Square	0.999
Schwarz Bayesian Information Criterion	172.23249

Evaluation Range: MAR1975 to JUN2016

The following graph shows the forecast for the selected model with tight 95% confidence intervals indicated by pink line.



Forecast results:

Model Viewer

Forecast Data Set

INDEX_NSA

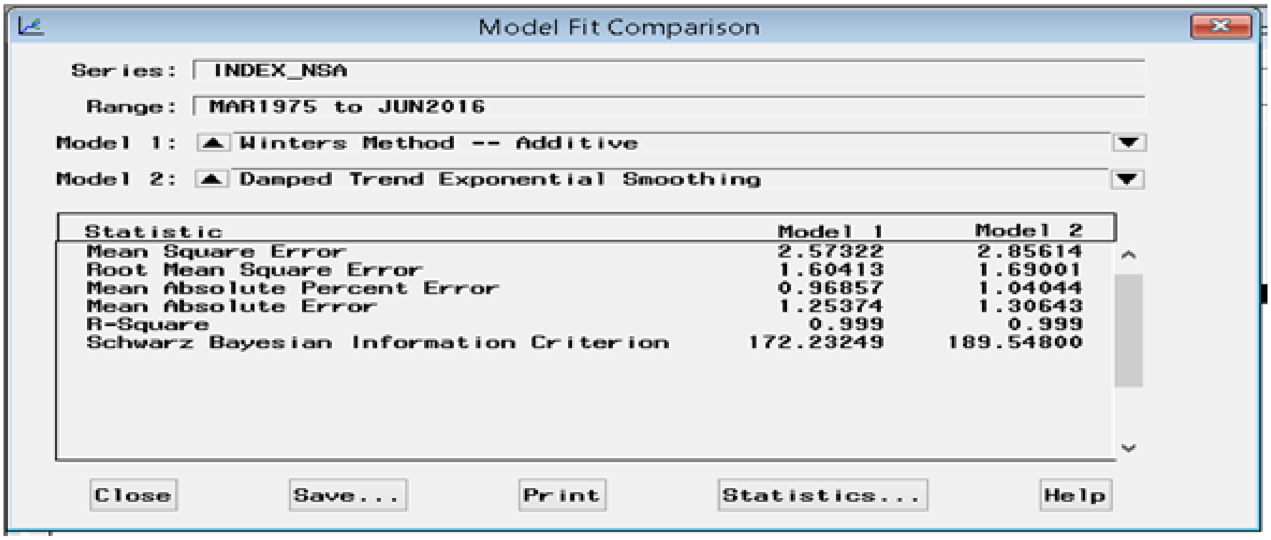
Winters Method -- Additive

TIMEPERIOD	ACTUAL	PREDICT	U95	L95	ERROR	NERROR	_LEVEL_	_TREND_	_SFACTOR
12/01/2016	.	299.6837	304.8130	294.5544	.	.	299.7772	5.3001	-0.0935
03/01/2017	.	304.7629	311.9669	297.5588	.	.	305.0773	5.3001	-0.3144
06/01/2017	.	310.9803	320.4114	301.5491	.	.	310.3774	5.3001	0.6029
09/01/2017	.	316.1644	328.0819	304.2470	.	.	315.6775	5.3001	0.4869
12/01/2017	.	320.8841	335.3180	306.4502	.	.	320.9776	5.3001	-0.0935
03/01/2018	.	325.9633	343.0644	308.8621	.	.	326.2777	5.3001	-0.3144
06/01/2018	.	332.1807	352.0925	312.2688	.	.	331.5778	5.3001	0.6029
09/01/2018	.	337.3648	360.3084	314.4213	.	.	336.8779	5.3001	0.4869
12/01/2018	.	342.0845	368.0964	316.0726	.	.	342.1780	5.3001	-0.0935
03/01/2019	.	347.1637	376.3714	317.9560	.	.	347.4781	5.3001	-0.3144
06/01/2019	.	353.3811	385.9070	320.8551	.	.	352.7782	5.3001	0.6029

Model Comparison:

Since the seasonal smoothing weight in Winters method – additive model was not significant, we compared the results of the next best model. Though Damped Trend exponential smoothing model was equally good in terms of prediction and error plots as that of the Winters Method – Additive model, head on head comparison of the model fit statistics suggest that Winters method—Additive model has produced the best results for all model fit statistics. So, we disregarded the only disqualification criteria of Winters Additive model.

Hence, we chose Winters method – Additive model for the state of Texas.



Model Fit Comparison

Series: INDEX_NSA

Range: MAR1975 to JUN2016

Model 1: Winters Method -- Additive

Model 2: Damped Trend Exponential Smoothing

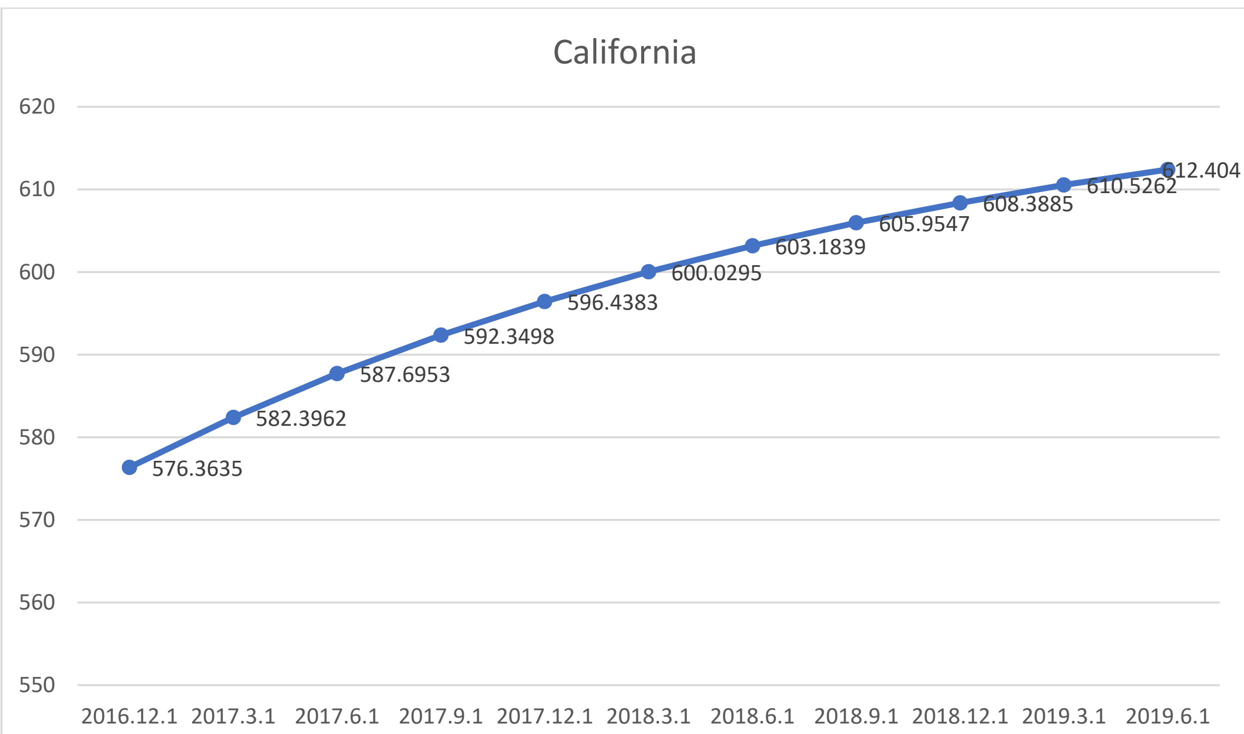
Statistic	Model 1	Model 2
Mean Square Error	2.57322	2.85614
Root Mean Square Error	1.60413	1.69001
Mean Absolute Percent Error	0.96857	1.04044
Mean Absolute Error	1.25374	1.30643
R-Square	0.999	0.999
Schwarz Bayesian Information Criterion	172.23249	189.54800

Close Save... Print Statistics... Help

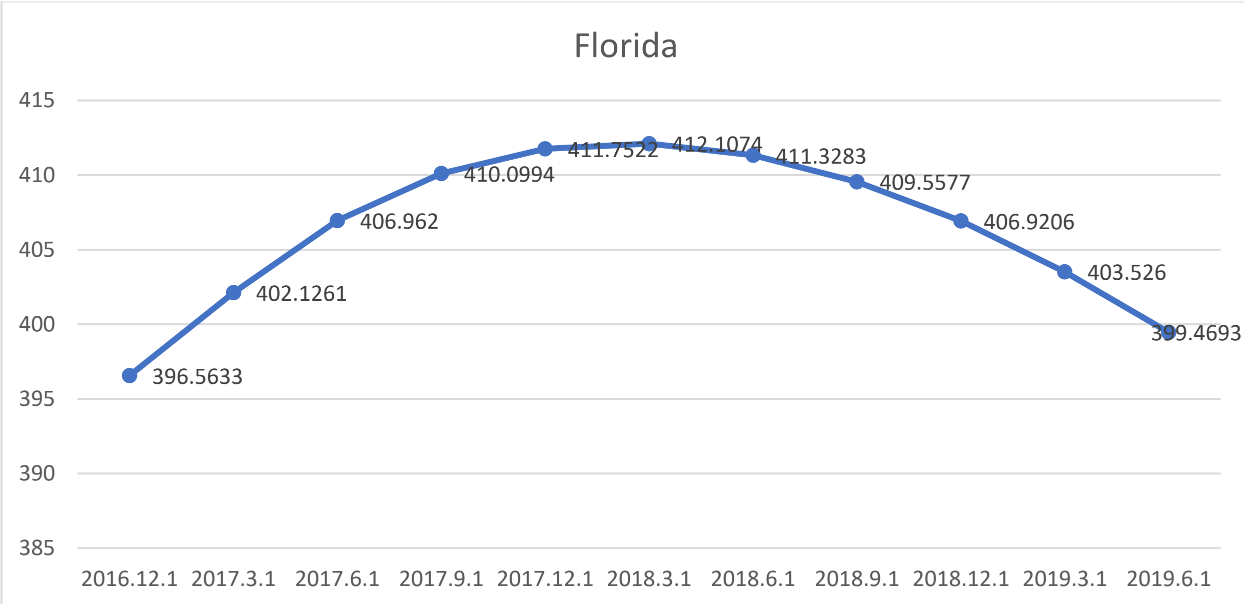
INFERENCES

States	California	Florida	New York	Illinois	Texas
Dec.2016	576.3635	396.5633	626.3958	333.9862	299.6937
Mar.2017	582.3962	402.1261	631.4378	336.1438	304.7629
Jun.2017	587.6953	406.962	636.4748	338.1972	310.9803
Sep.2017	592.3498	410.0994	641.5067	340.1515	316.1644
Dec.2017	596.4383	411.7522	646.5336	342.0115	320.8841
Mar.2018	600.0295	412.1074	651.5555	343.7817	325.9633
Jun.2018	603.1839	411.3283	656.5723	345.4664	332.1807
Sep.2018	605.9547	409.5577	661.5842	347.0697	337.3648
Dec.2018	608.3885	406.9206	666.591	348.5957	342.0845
Mar.2019	610.5262	403.526	671.5928	350.048	347.1637
Jun.2019	612.404	399.4693	676.5896	351.4302	353.3811

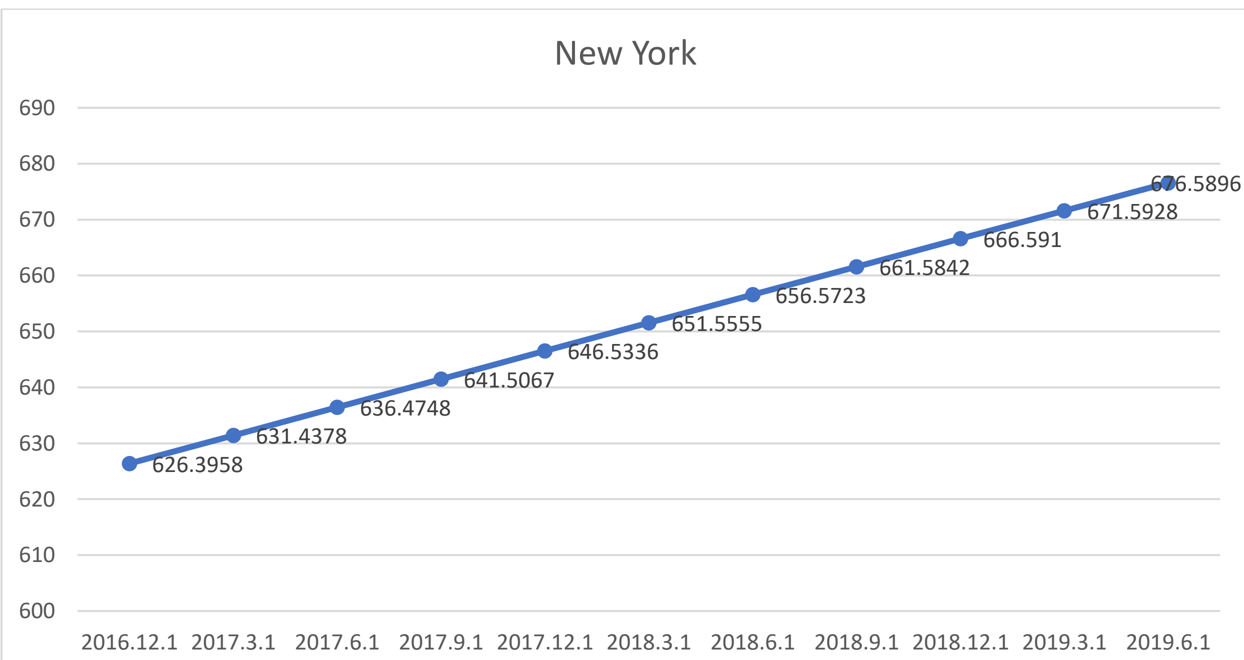
predictions of 5 states



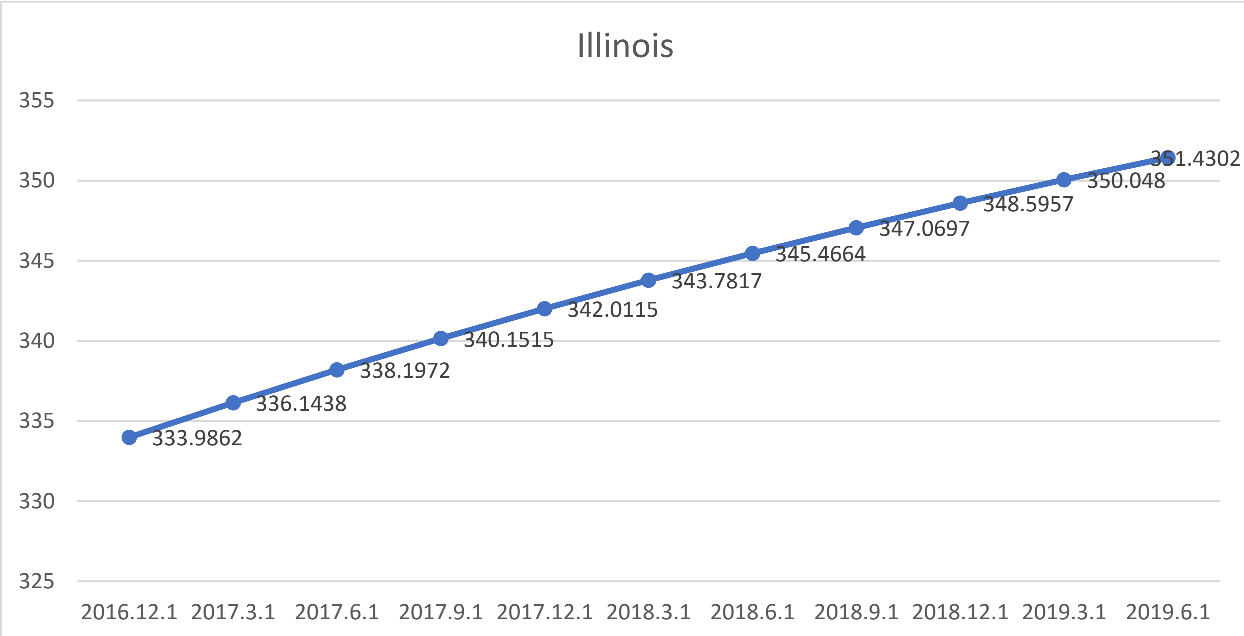
HPI of California has a rising trend that creases by 6.25%



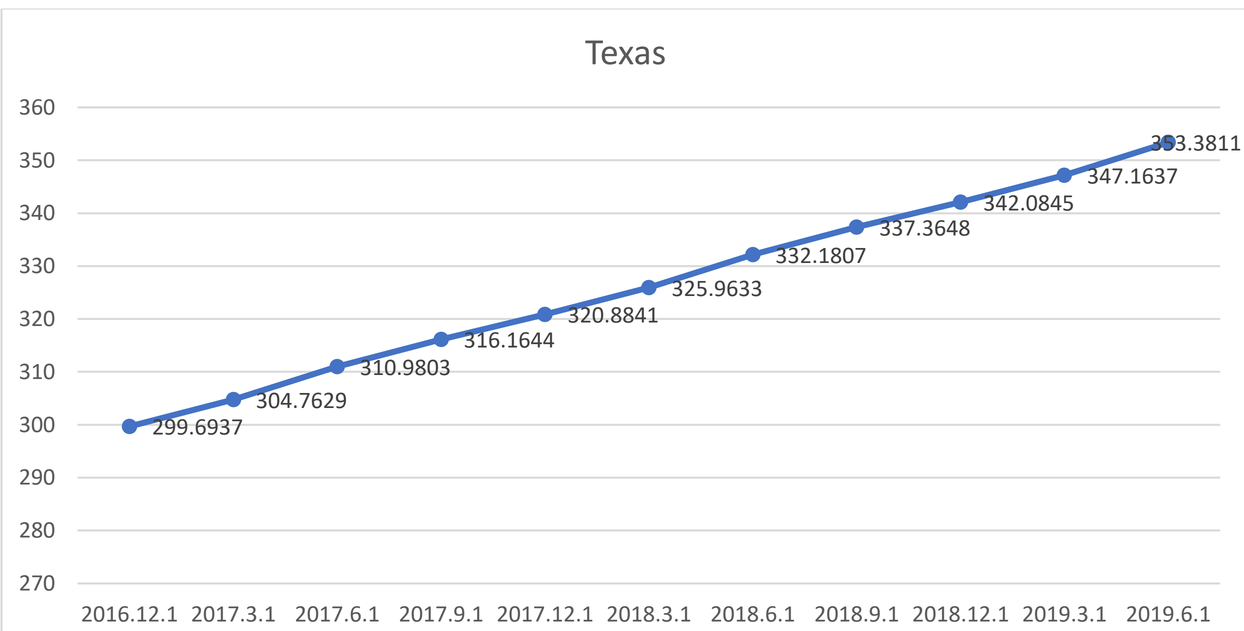
HPI of Florida has an increasing trend and it reaches the peak at Jan.1st, 2018 with HPI of 412.1074. Then the line falls down, and at Apr.1st, 2019 the prediction is almost the same as the Oct,1st, 2016.



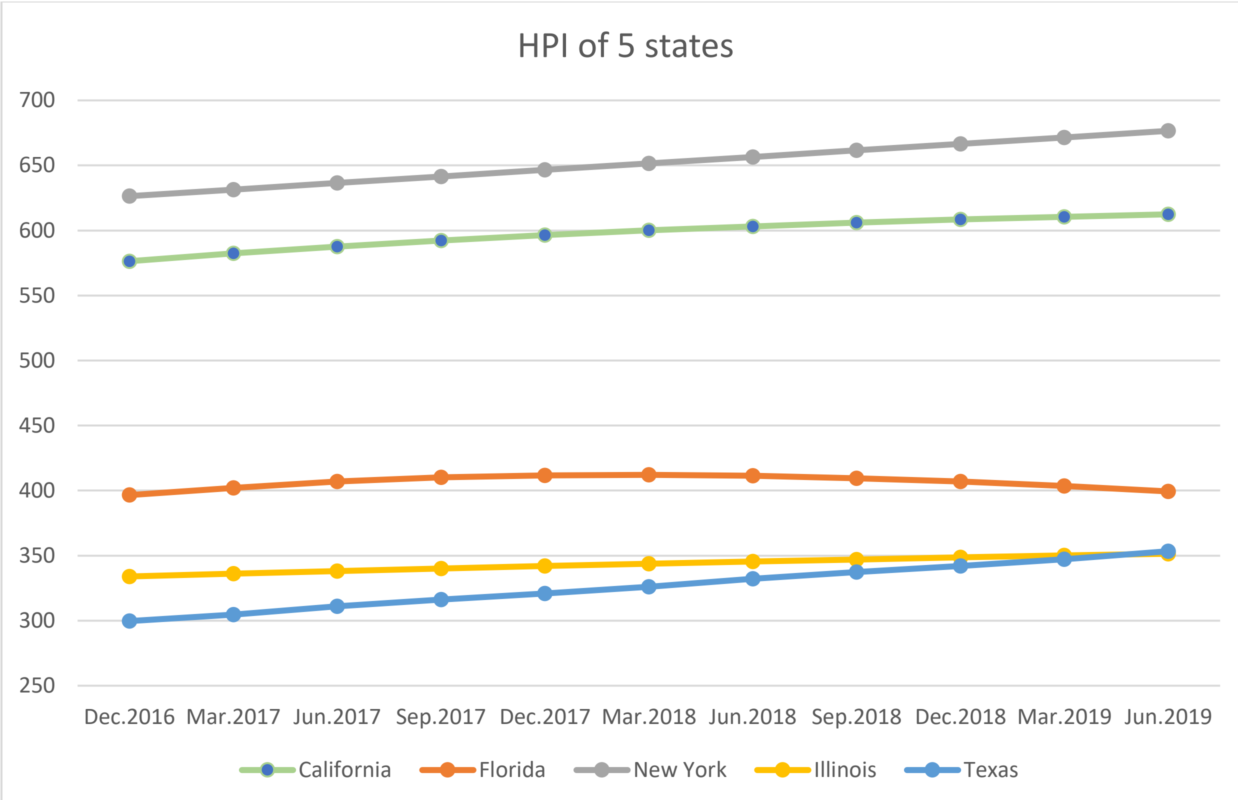
It can be clear that the HPI of New York creases 8.01% from 626.3958 to 676.5896



Illinois might have the lowest increasing in HPI, which is only 5.22%



The graph of Texas shows that the HPI of this state rises up rapidly. And it is 17.91% increasing and also is the highest figure of 5 states.



Comparing 5 states together, we can find that New York has the highest HPI than any other states. California ranks the second place both in HPI and HPI increasing rate (6.25%). These two states have HPI around 600, which is obviously higher than other 3 states.

HPI of Florida, Illinois and Texas are between 300 and 400. Florida has a rise and drop, and it is the only state that has a decrease trend. Texas has the lowest HPI which is only half of New York, but it increases the most quickly and at Apr.1st, 2019, it will catch up with Illinois.

RECOMMENDATIONS

1. House Price Index's forecasts indicate that housing prices shall increase year on year for all states except Florida.
2. Florida's housing price index shows an increase up to first quarter of 2018 and reduction post that. This trend appears to be similar to the housing price bubble where housing prices started falling from 2005 and reached lowest value in 2008.
3. As an investment, it is advised to buy properties now as all the states show a trend for increase in prices.
4. But, investors should sell the property before 2018 in Florida as the prices shall drop
5. In Texas, Illinois, Florida and California, it is advised to buy property and not to sell them for next 3 years to reap higher benefits.
6. Texas has the highest probable return on investment in 3 years – Approximately 17%

REFERENCES

1. House Price Index, Extracted on 21st April 2017.

   https://www.fhfa.gov/DataTools/Downloads/pages/house-price-index.aspx

2. Trends in National House Prices, Extracted on 20th April 2017.

   http://www.freddiemac.com/finance/house_price_index.html

3. All-Transactions House Price Index for the United States, Extracted on 22nd April, 2017.

   https://fred.stlouisfed.org/series/USSTHPI