

**CAL WORKS**

California Work Opportunity and Responsibility to Kids



July 11, 2016

UNIVERSITY of connecticut

MSBAPM

Table of Contents:

1. Purpose and Objective ………………………………………………………………………….3
   1. General Program Requirements…………………………………………………...3
   2. Eligibility…………………………………………………………………………………….3
   3. Timeline………………………………………………………………………………………3
   4. Background………………………………………………………………………………….4
2. Data set…………………………………………………………………………………………………..5
3. Models……………………………………………………………………………………………………6
   1. Los Angeles County Model…………………………………………………………….6
   2. Fresno County model…………………………………………………………………..17
   3. San Francisco model ……………………………………………………………………36
   4. Siskiyou County model…………………………………………………………………43
4. Business Insights…………………………………………………………………………………….64
5. References……………………………………………………………………………………………...65

Purpose and Objective:

**CalWORKs** is a welfare program that gives cash aid and services to eligible needy California families. The program serves all 58 counties in the state and is operated locally by county welfare departments. If a family has little or no cash and needs housing, food, utilities, clothing or medical care, it may be eligible to receive immediate short-term help. Families that apply and qualify for ongoing assistance receive money each month to help pay for housing, food and other necessary expenses. This is California's version of the [Federal Temporary Assistance to Needy Families program](http://www.acf.hhs.gov/programs/ofa/programs/tanf).

General Program Requirements

In order to qualify for this benefit program, you must be a resident of the state of California, either pregnant or responsible for a child under 19 years of age, a U.S. national, citizen, legal alien, or permanent resident, have low or very low income, and be either under-employed (working for very low wages), unemployed or about to become unemployed.

Eligibility:

There are several criteria on which the families qualify for the CalWORKs support. The factors are as follows:

* The income of the family [low income and very low income
* Applicant’s citizenship [must have California state citizenship
* Age
* Resources and assets
* Children must be deprived of parental support and care due to incapacity, death, absence of a parent or unemployment of principal wage earner.

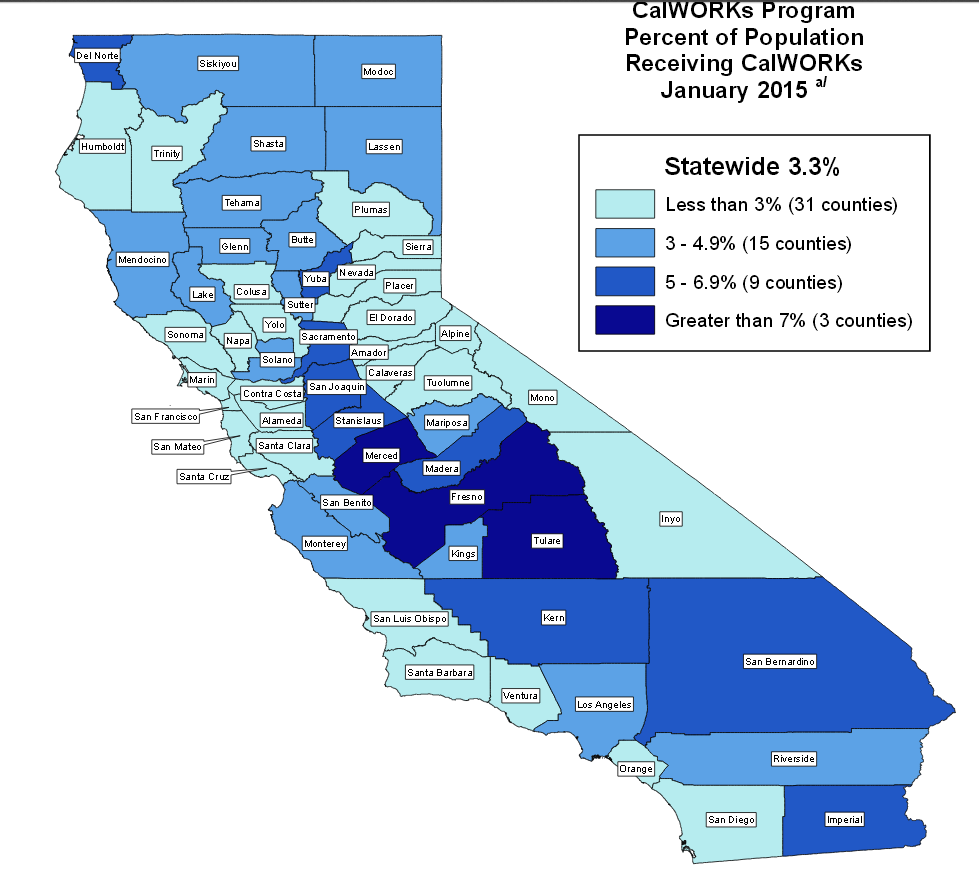
Time Limit:

For Adults- Cash aid is provided monthly up till a limit of 48 months.

For Children – Children may still be eligible for the cash aid after 48 months up to the age of 18.

Background

We decided to use a dataset from the state of California for our time series forecast. The dataset looks at the number of applications for state assistance. As the nation emerges from the Great Recession of 2008, some states are struggling more than others. California has the United States’ largest budget and has one of the largest budgets in the world. As of 2016, California’s budget is the eighth largest in the world (citation needed) and has a diverse economy. California’s budgetary income includes the entertainment industry, Silicon Valley, Wine Country and has a large amount of money coming in due to tourism. Despite California’s rich and diverse economy, they are also struggling with deficits and have a large amount of people drawing from funds for public assistance.

Source: <http://www.cdss.ca.gov/research/res/pdf/caltrends/CWPopRecJan15Map.pdf>

Dataset

The dataset we are using is quite large. It has data points for all of California’s counties from January 2010 until September 2016. For this exercise we decided to focus on four specific counties and selected them by the percentage of the population that applied for public assistance as well as their location in the state. The four counties we selected were:

Los Angeles County – located in famous Southern California, world renowned for its entertainment industry. It is also home to very wealthy people and famous movie stars.

San Francisco County – located on the pacific ocean, very wealthy part of the state and the country.

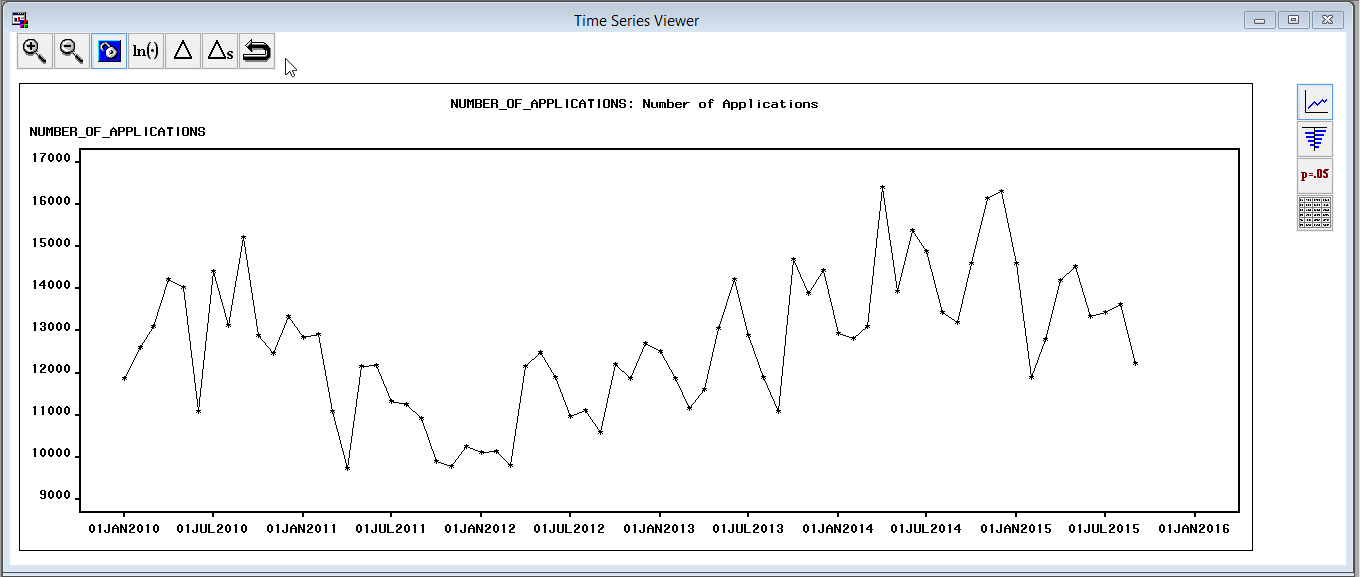
Fresno Country – located in central California, typically regarded as the less affluent part of the state.

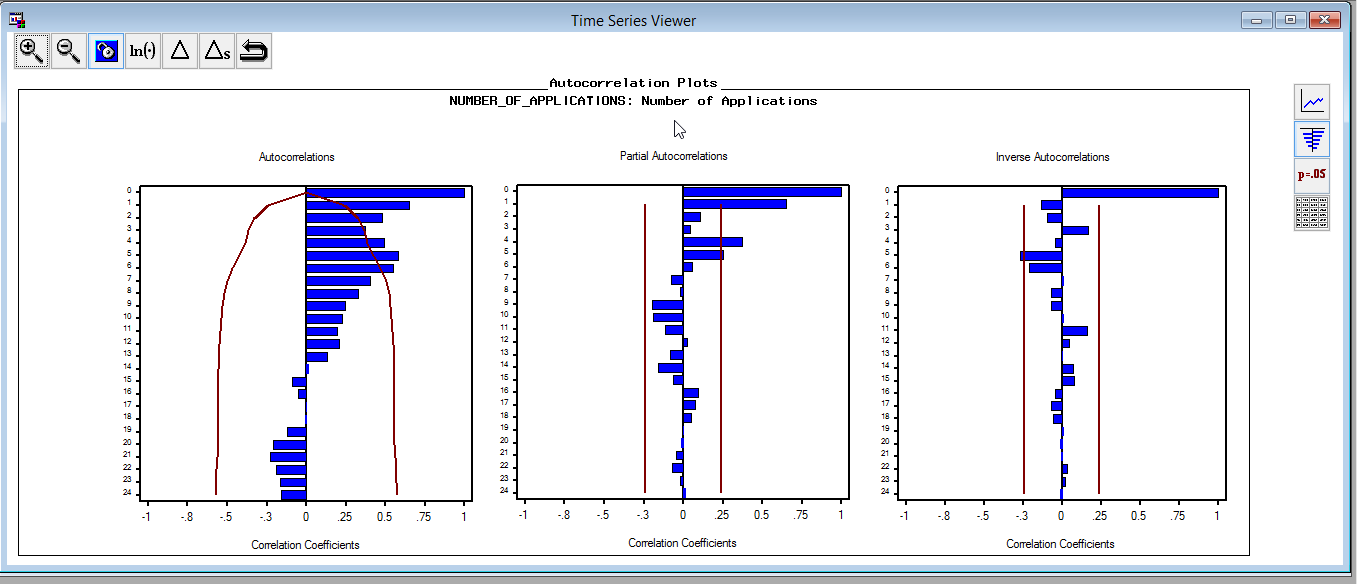
Siskiyou County – located in northern California, rich in farmland and is small and rural.

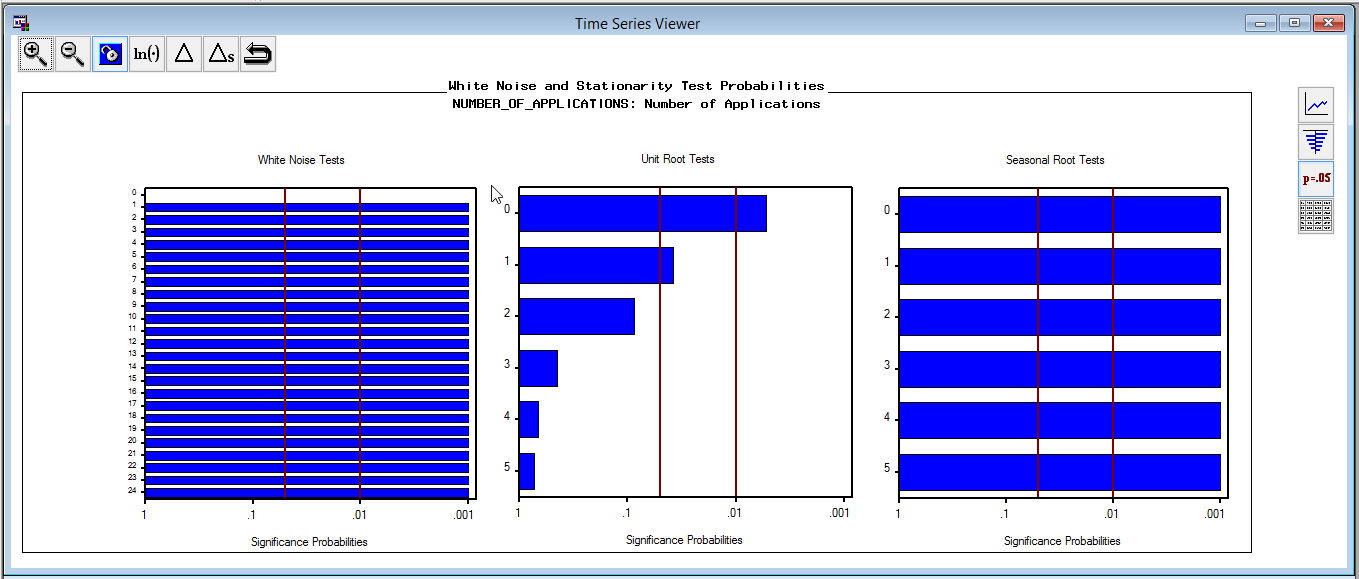
Models:

**Los Angeles County and Model**

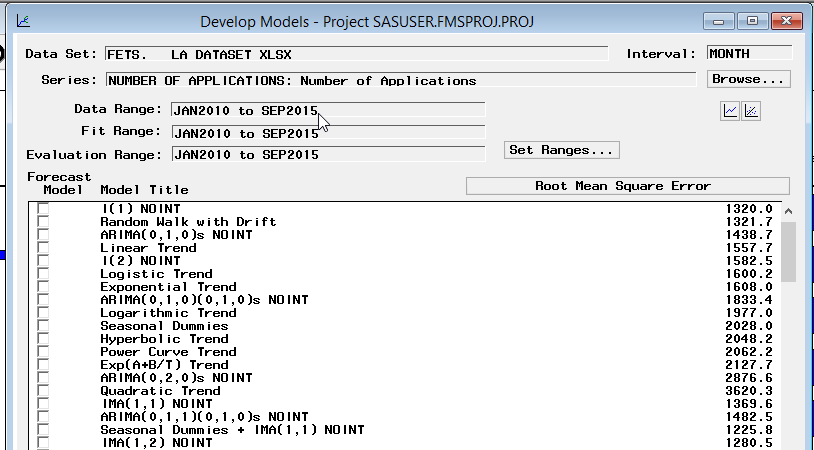
Below is how the raw data was shown on the SAS models. The team did not notice a trend with the data, however there were some seasonality component to the data when the team looked at it.

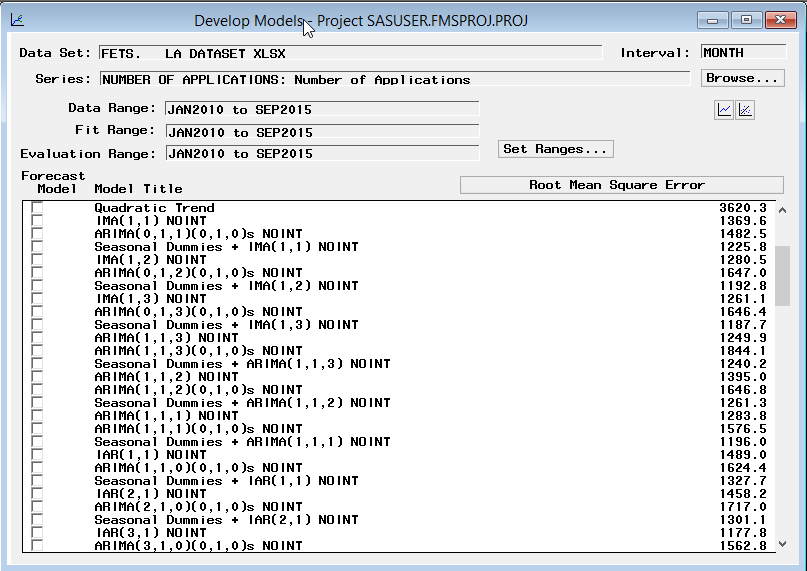






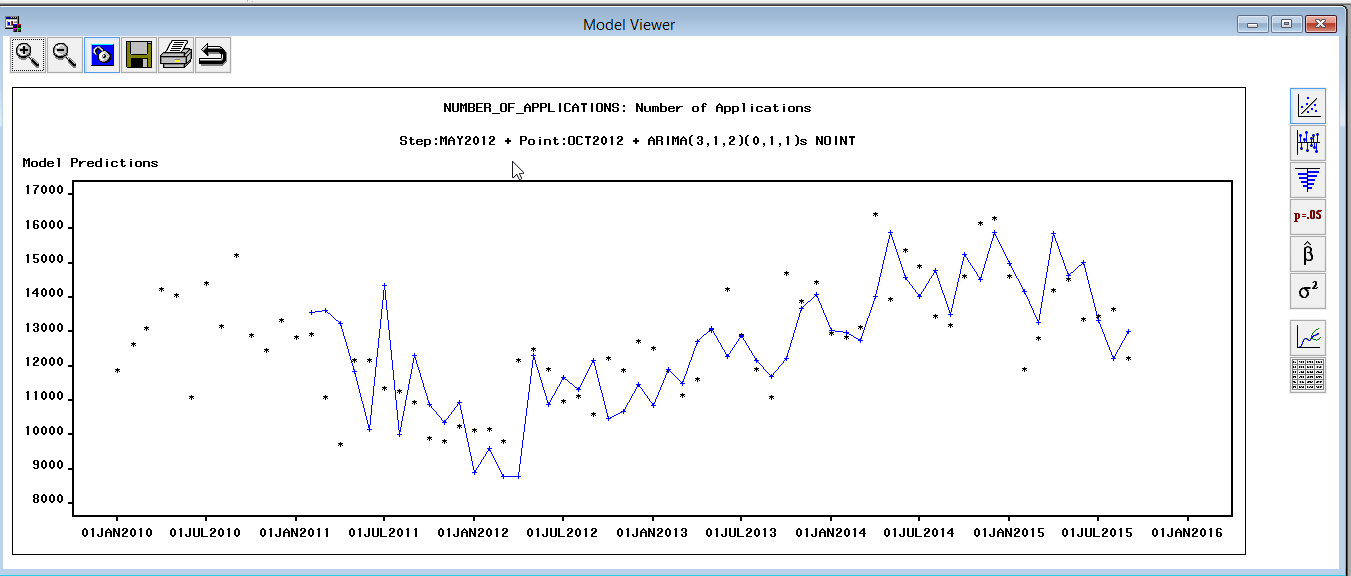
The data shown above shows that the errors are not random, but the ACF and PACF show signs of seasonality and auto regression.

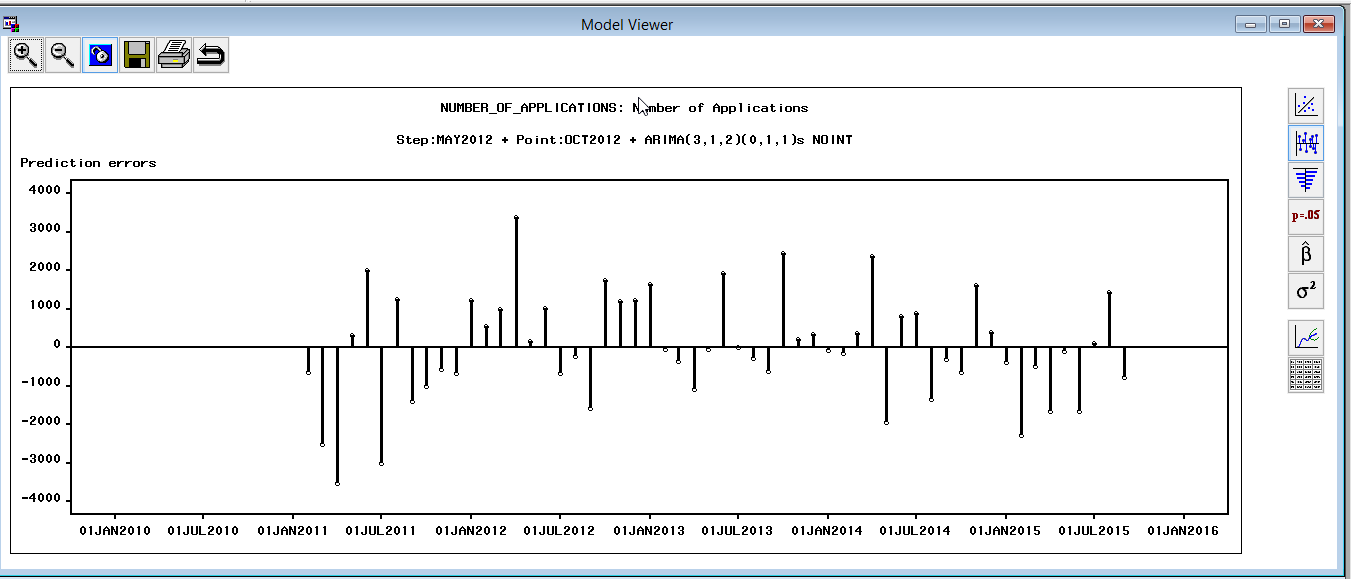


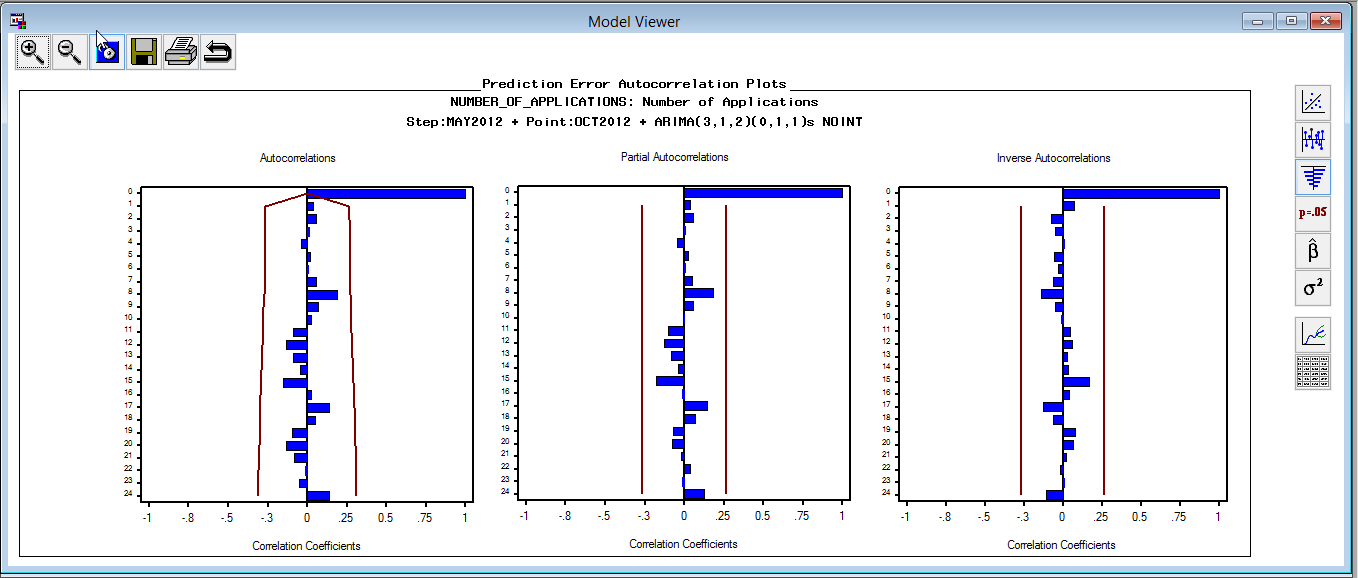


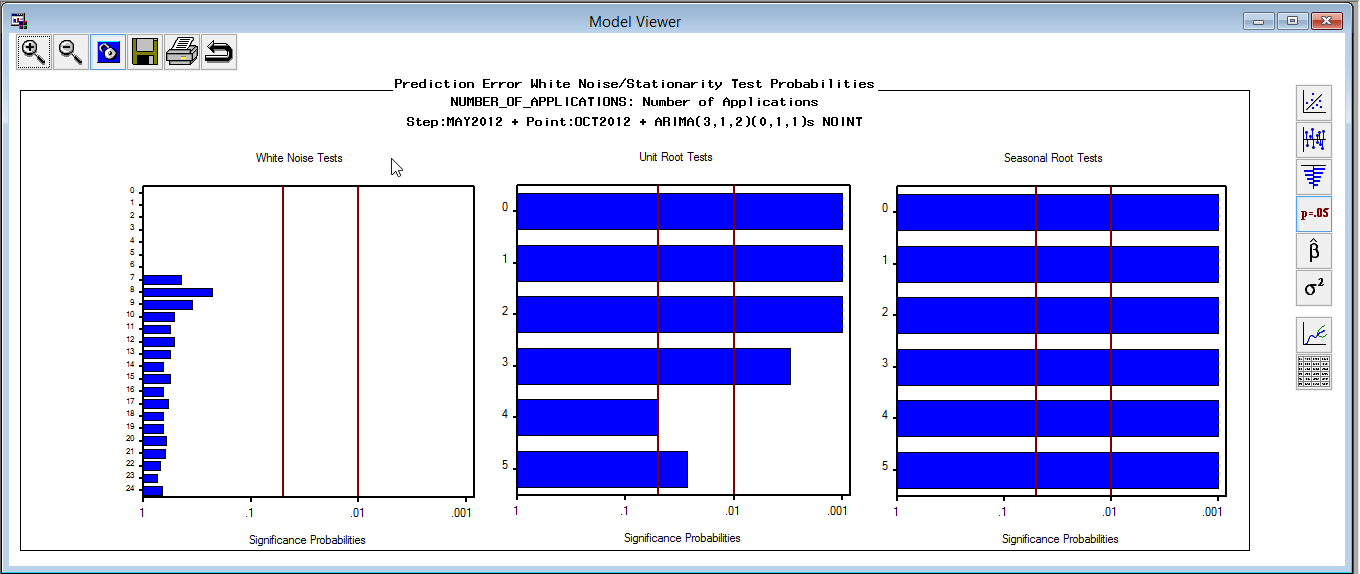
We ran a bunch of different types of models and found that models with seasonal difference are better in forecast and white noise, but models with seasonal dummies are better in RMSE, SBIC, AIC. One we narrowed down our model, we did an intervention of the model and used May2012 + Point: OCT2012 + as our intervention points because we noticed that area did not follow the pattern of the data.

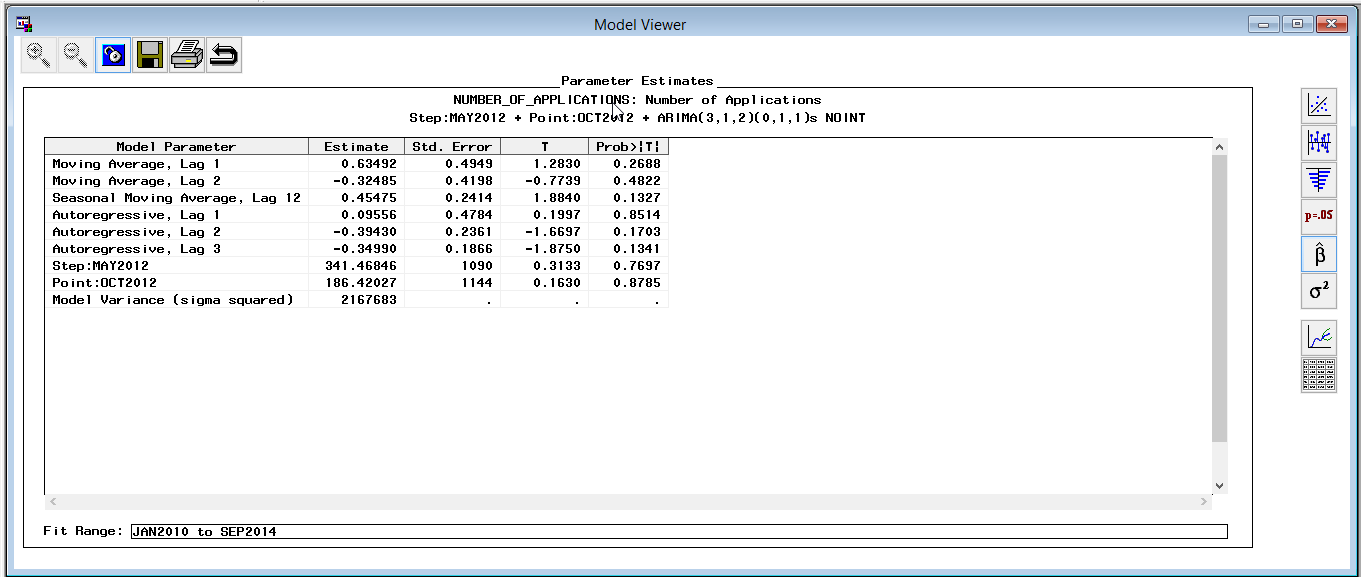
Once we did that the ARIMA (3, 1, 2) (0,1,1) became our final model.

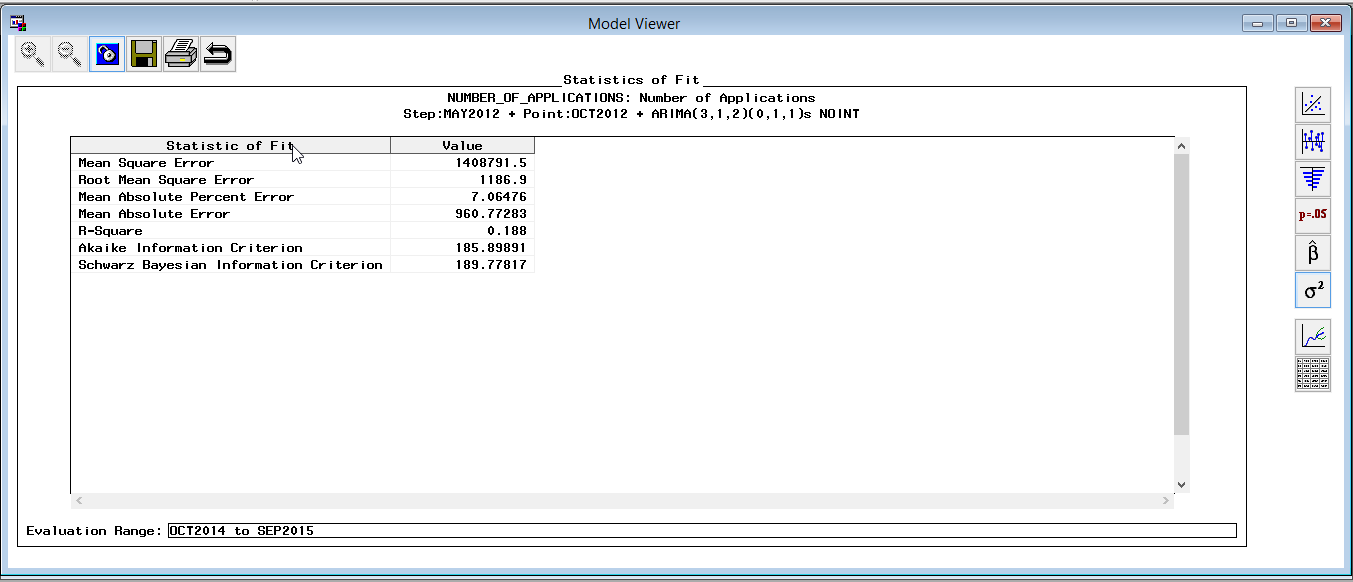


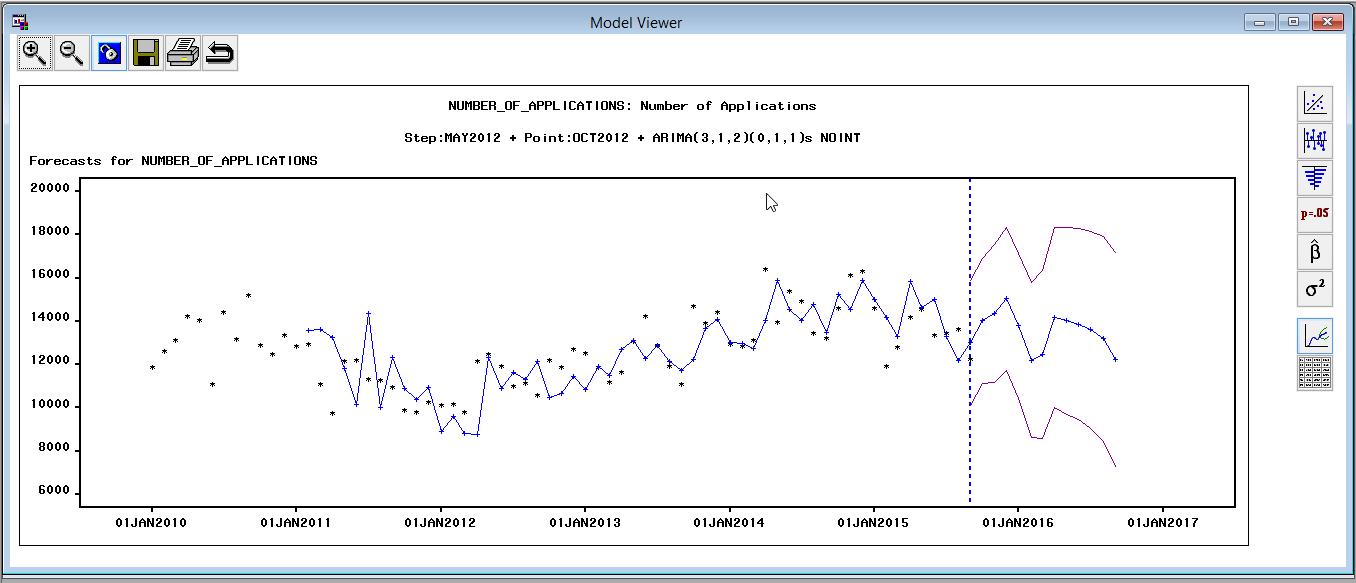




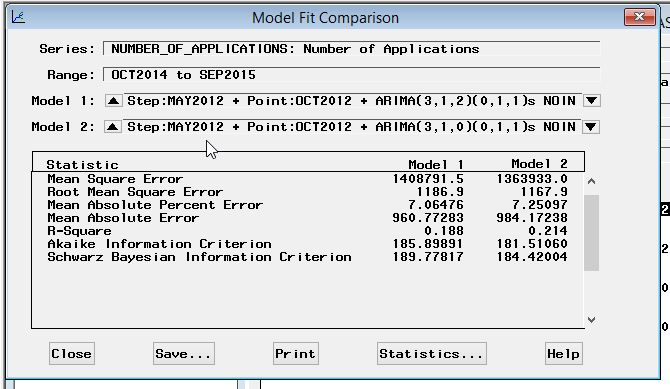




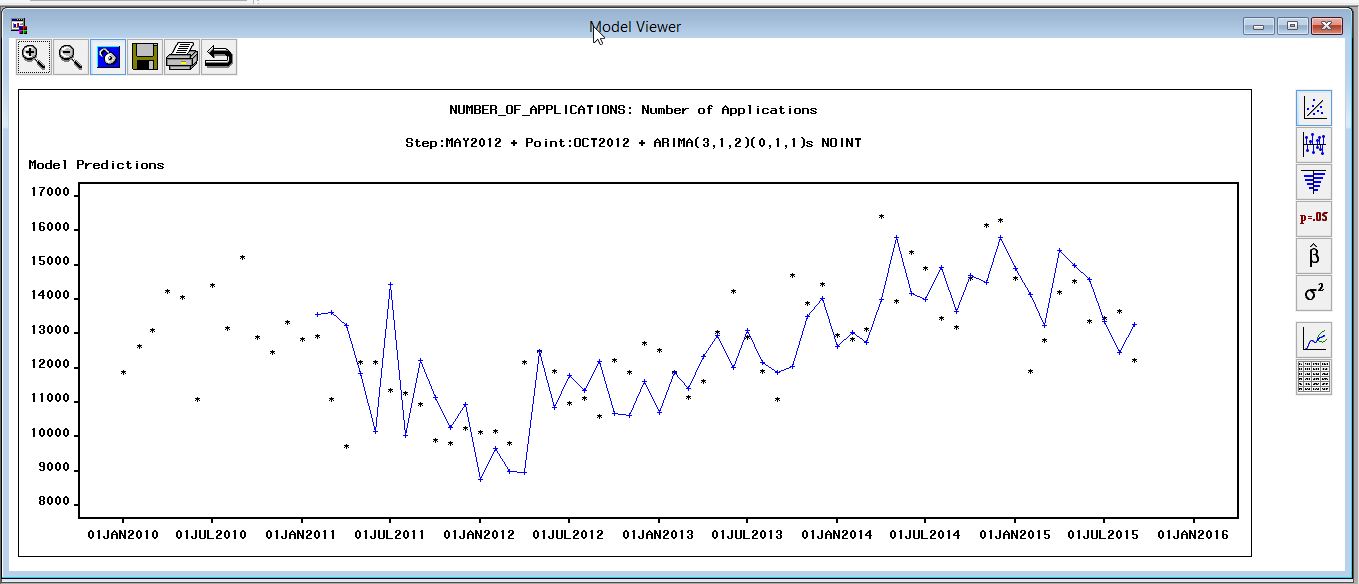


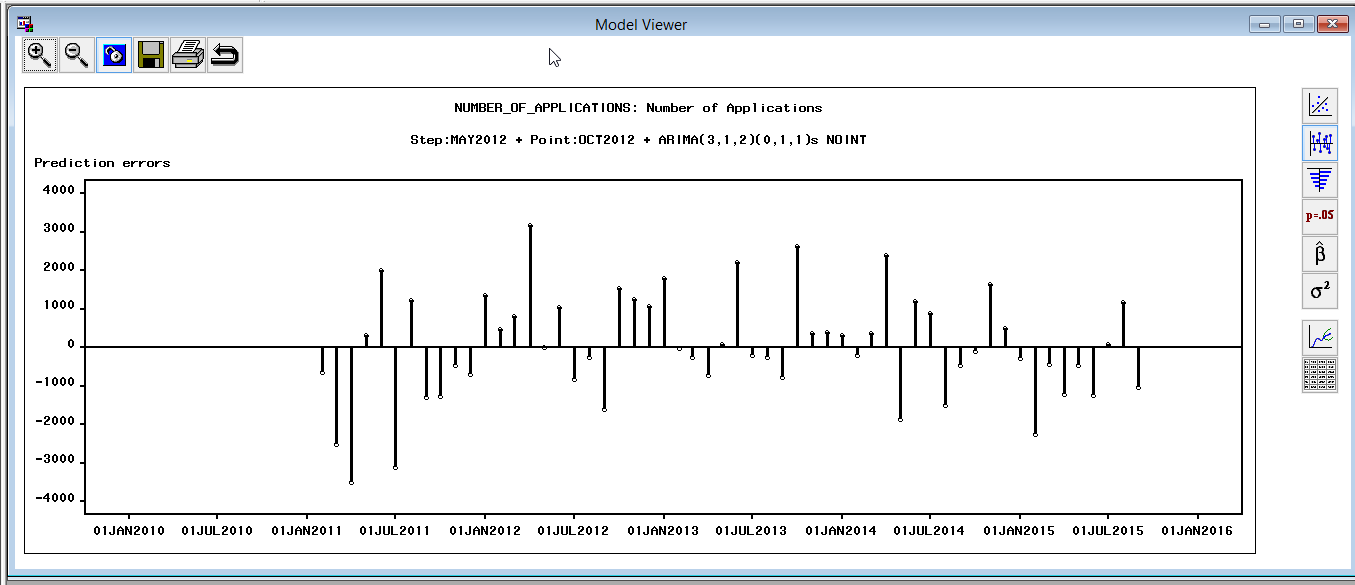


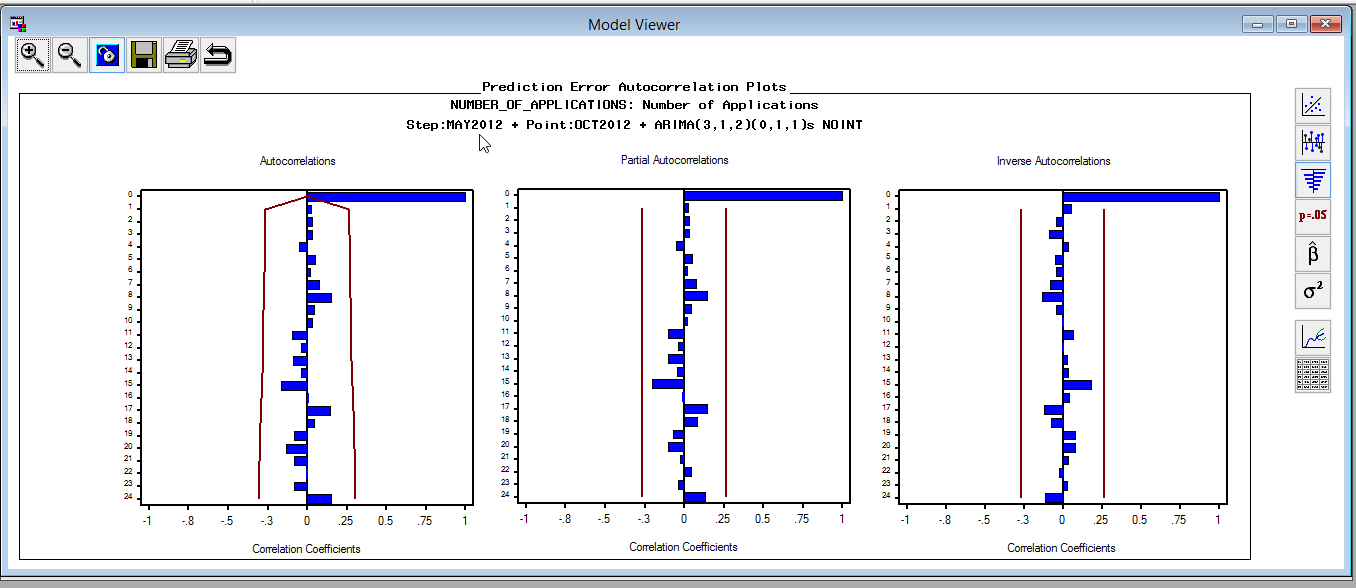
Looking at the result the team was happy with the randomness of the errors as well as the strong significance of the unit test and the seasonality test. The team also noticed that Lag 4, 5 in unit root tests are not that satisfied, but at least they meet 1% significant level. Forecast results seem reasonable compare to previous data, though confidential interval is relatively large. It’s already the best one compare to others.

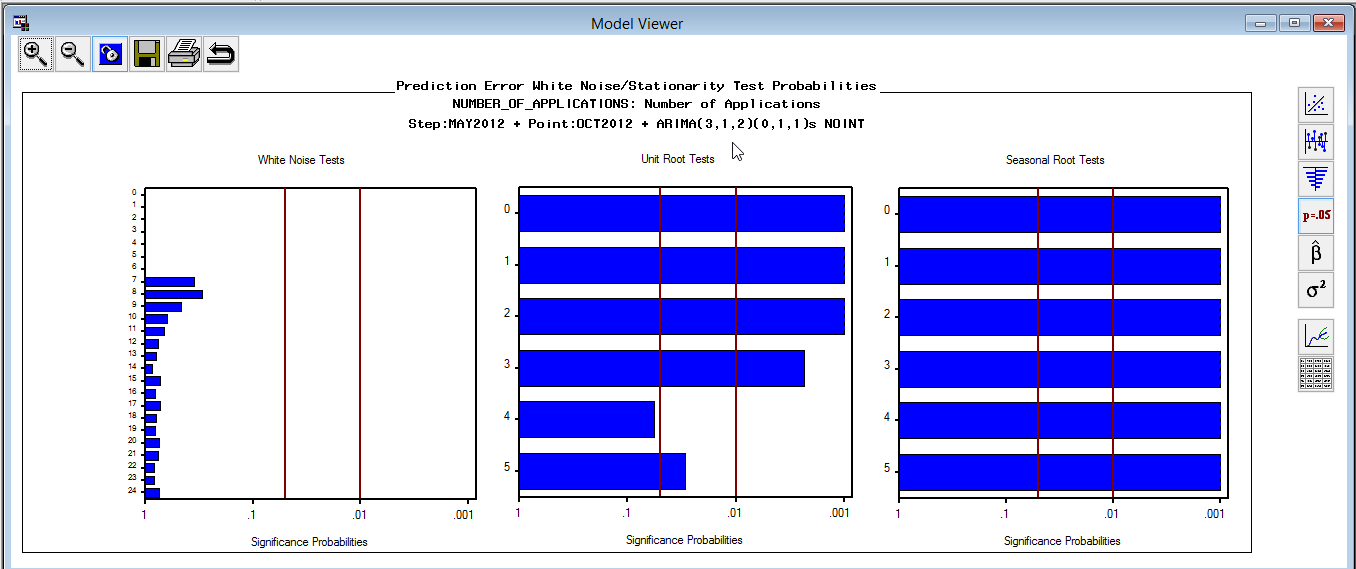


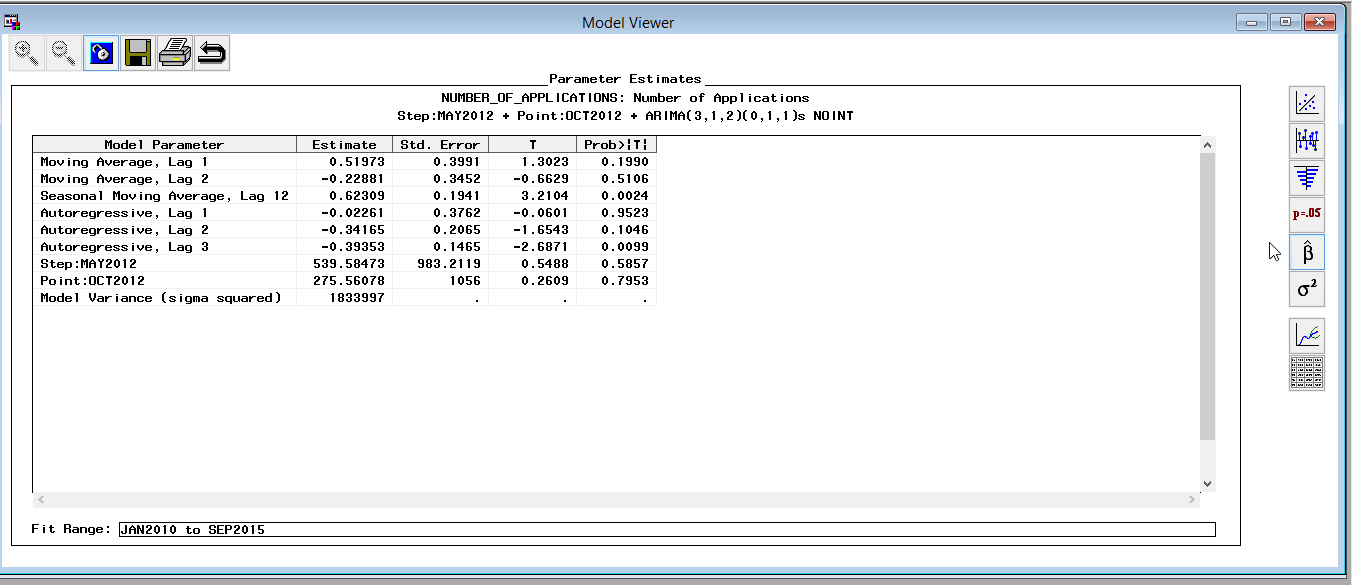
The team now refit the model with entire time span.

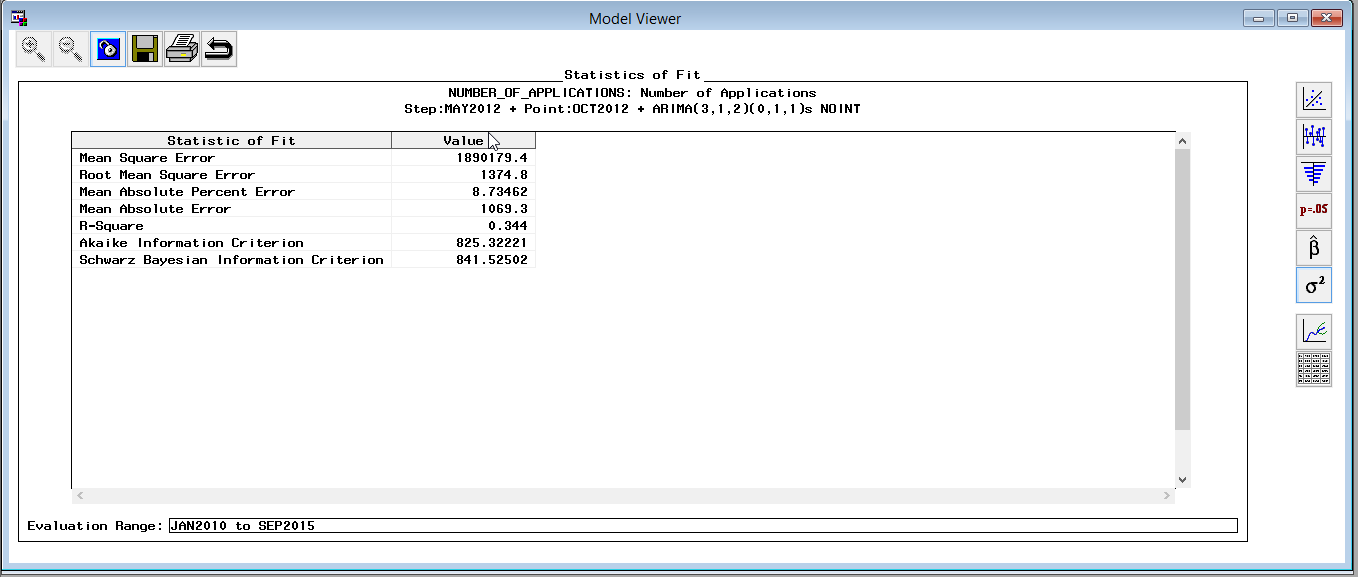


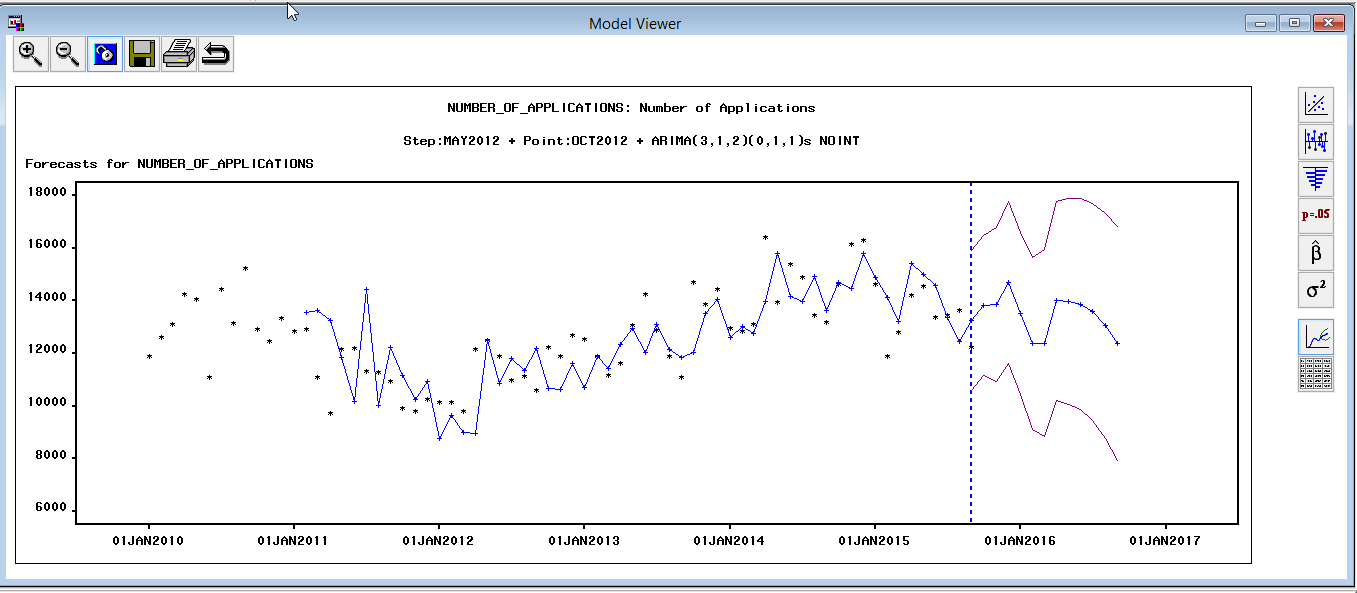




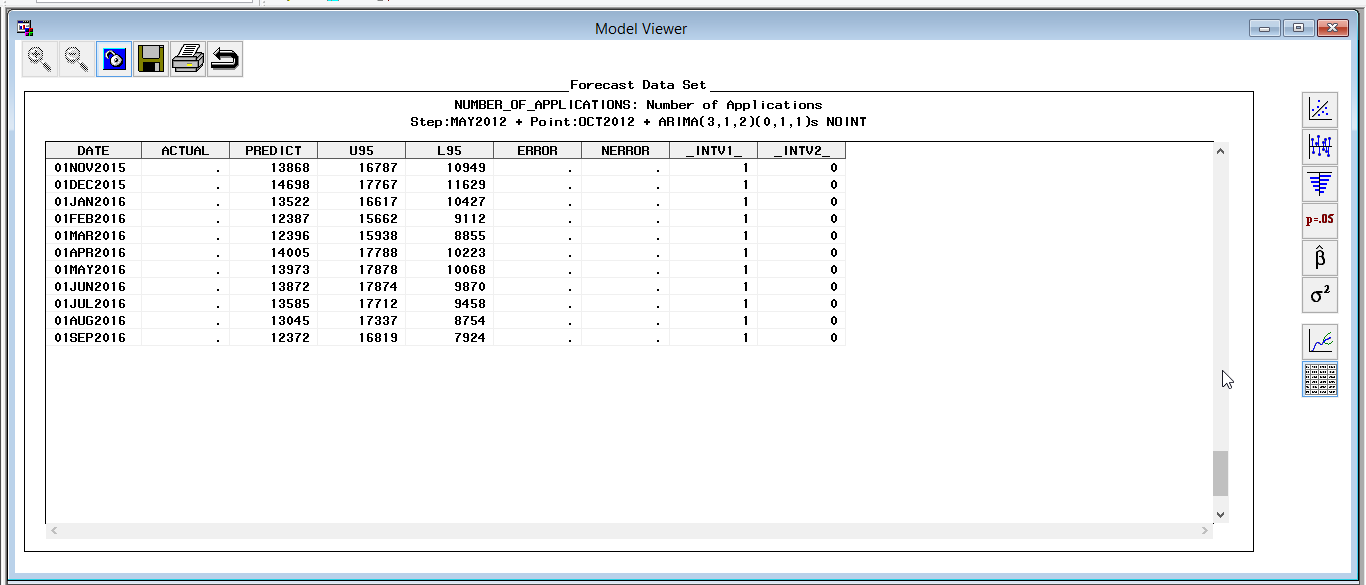


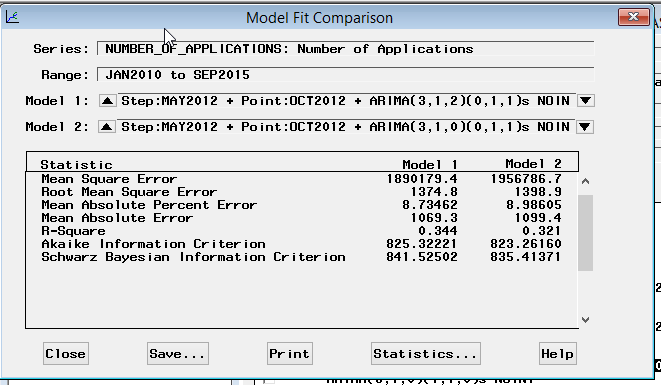






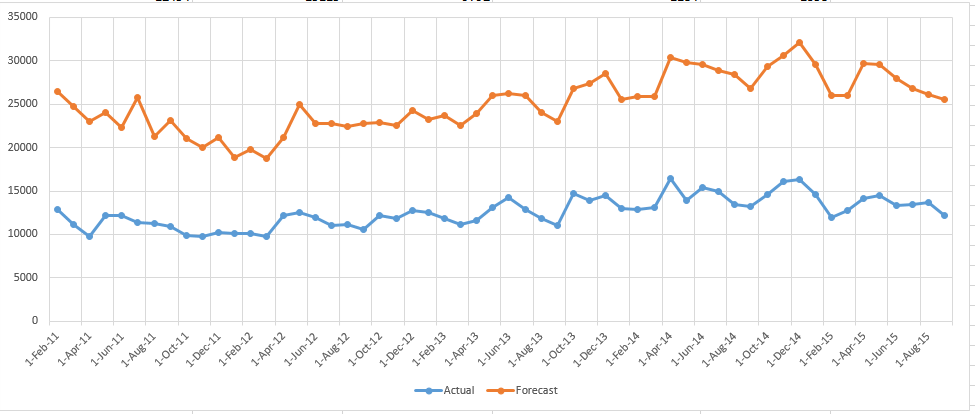
As we can see with the model refit, the forecast doesn’t vary too much, but out of all the models we felt this one performed the best.



****

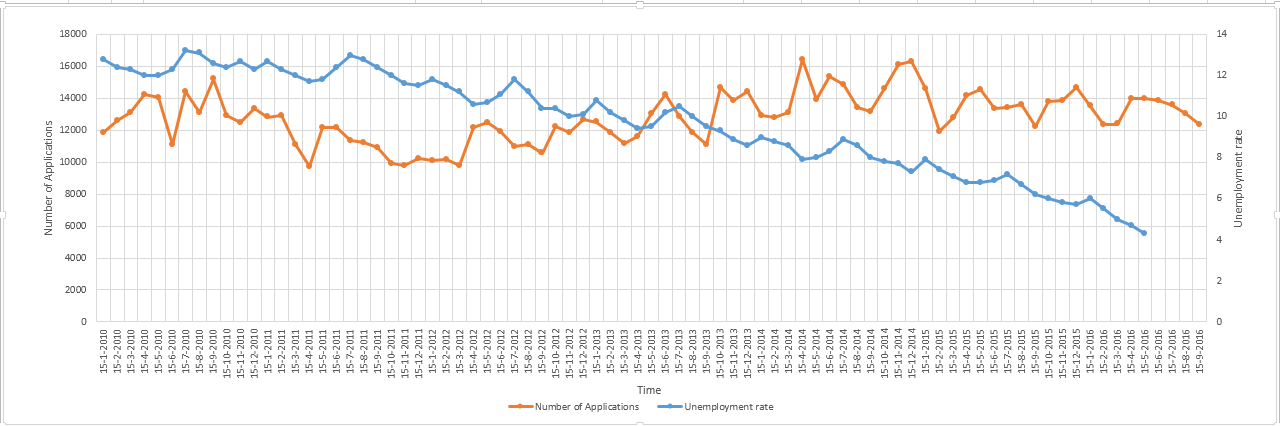
Comparing the Predicted vs. the Actual Trend

When comparing the predicted vs. the actual trend we were surprised to see that the forecast almost mirrored the actual values.

****

Comparing the model to external factors

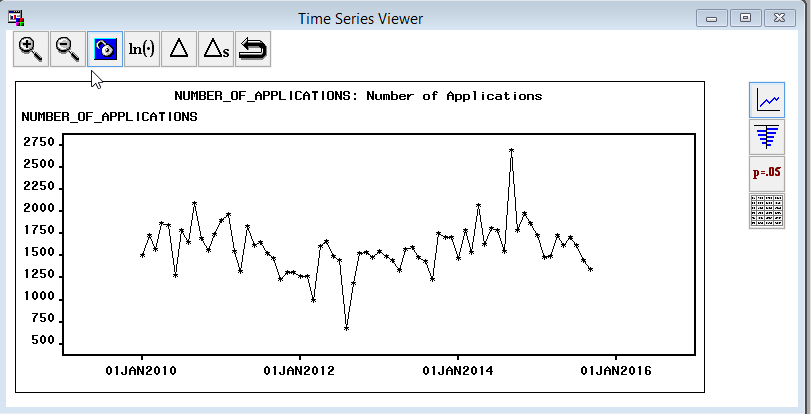
The team wanted to investigate if the forecast and data points had any correlation to any external factors. A natural candidate variable to use seemed to be unemployment. When we ran the data we saw a strong correlation to the number of applicants to state aid and to the unemployment rate over the same period.

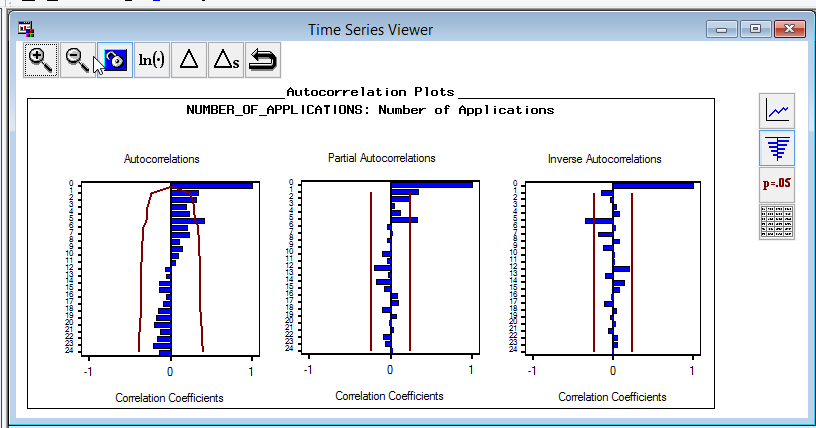
**Forecast range 01NOV2015-01SEP2016**

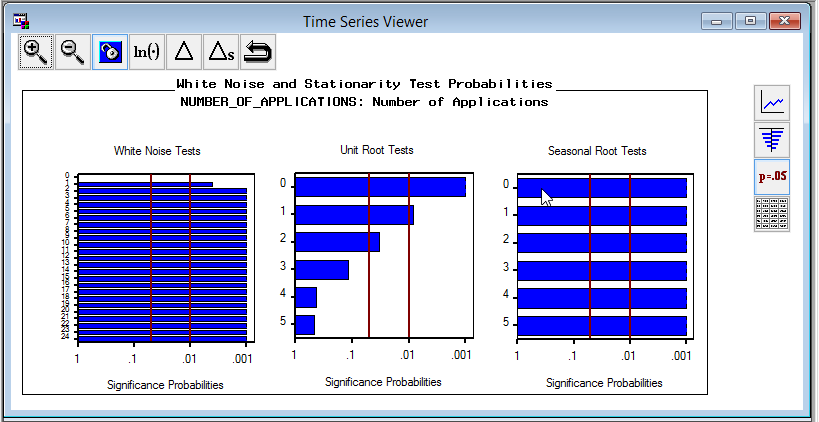
We can see there is some correlation between trend of unemployment rate and number of applications.

**Fresno County Data Set**

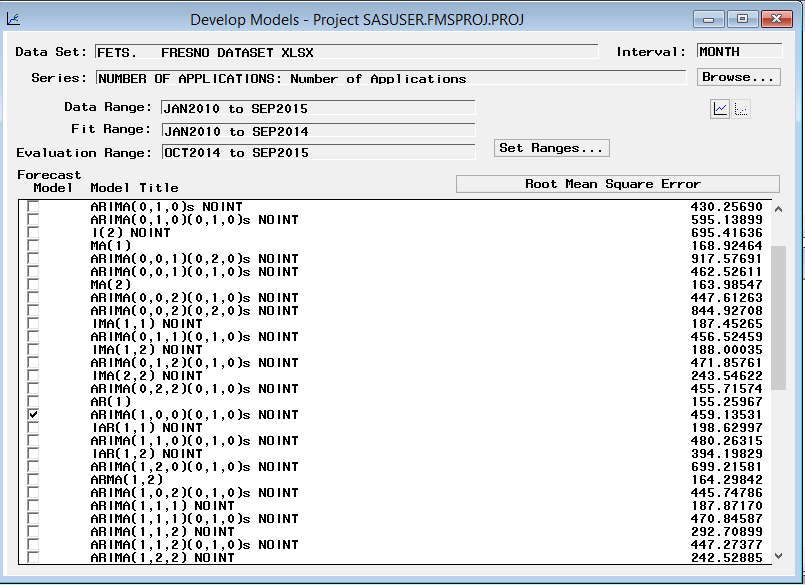
**Original graphs:**



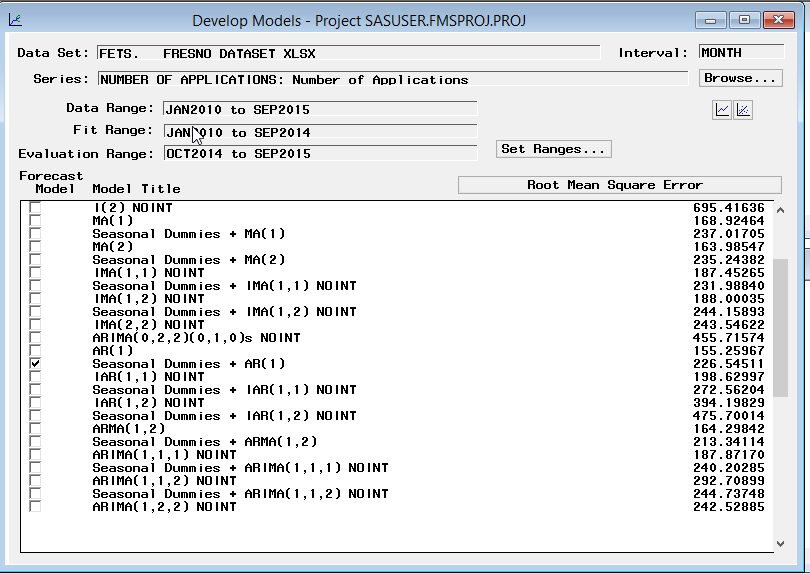




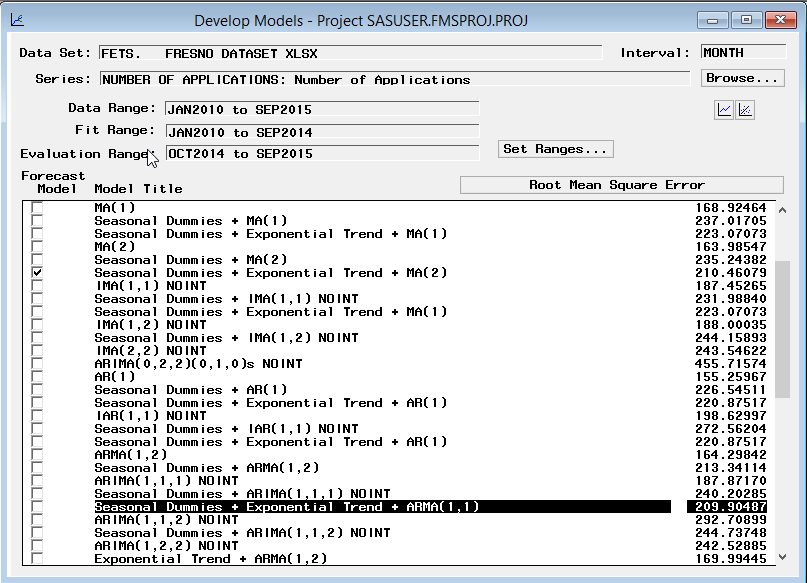
Looking at the models initially we determined that P at most was 2, q at most was 2 for a possible ARMA model. Another thing the team noticed that unlike LA, there appeared to be not seasonality, but similar to the LA data set, there was not trend.



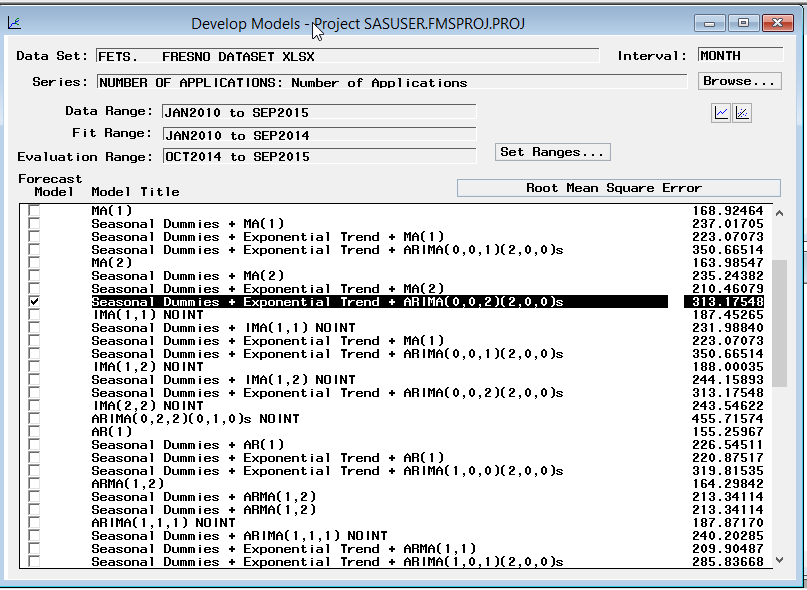
For all ARMA models, adding seasonal difference actually makes them worse, as e suspected that it might since seasonality was not shown strongly in the original data set.



Seasonal dummies are better than seasonal difference in our models.

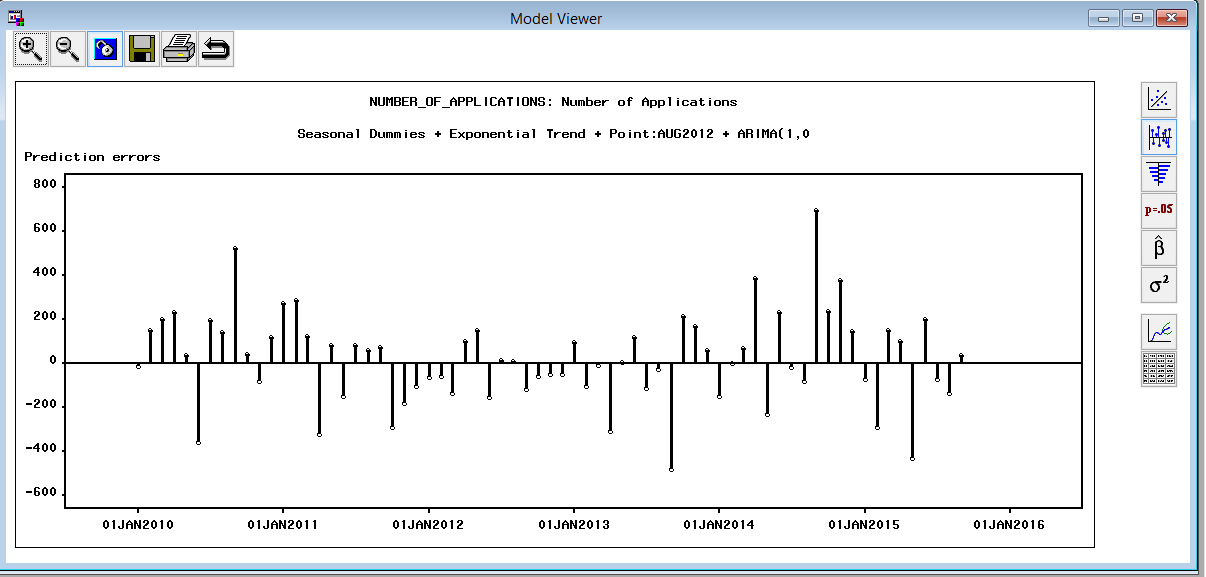


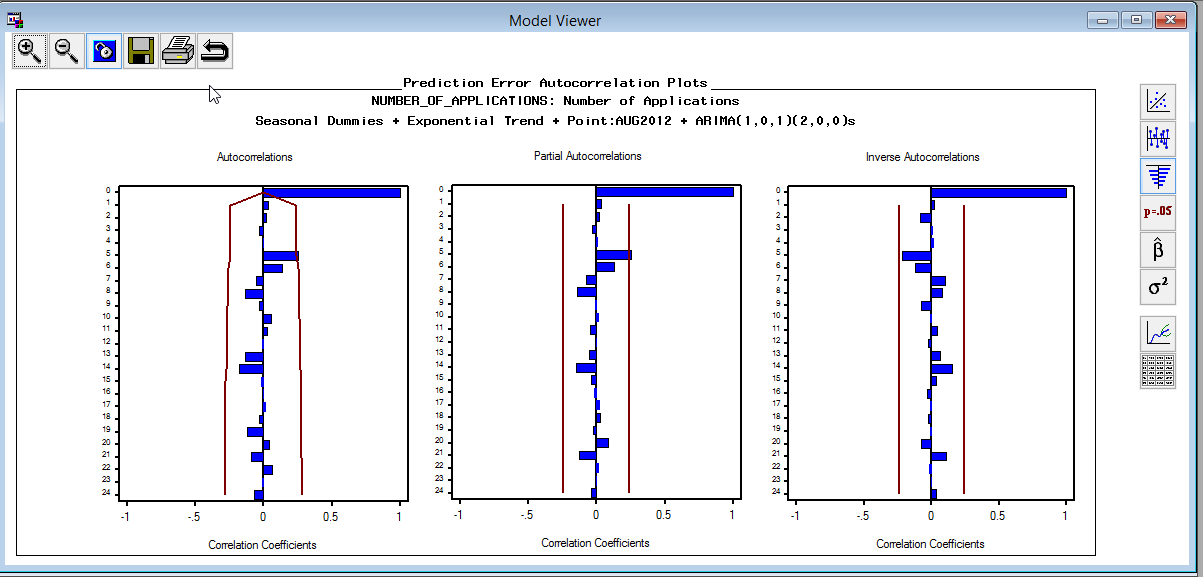
With exponential trend, RMSE is better than model with first difference.

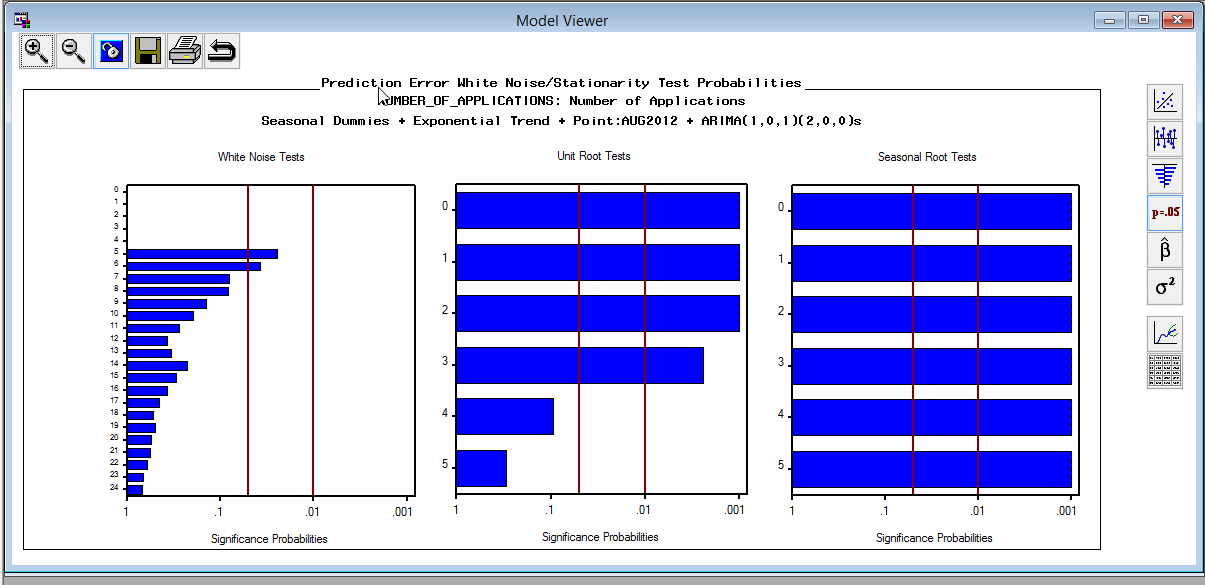


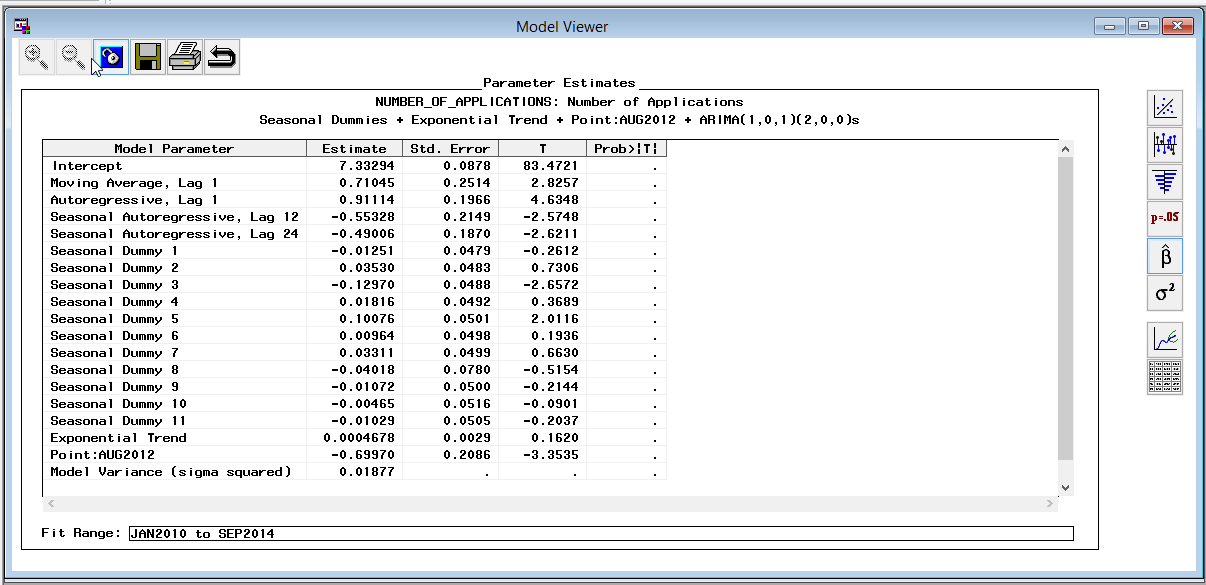
Although added seasonal p result in increase of RMSE, but it gives us far better accuracy. With SBIC, AIC, this difference is much smaller.

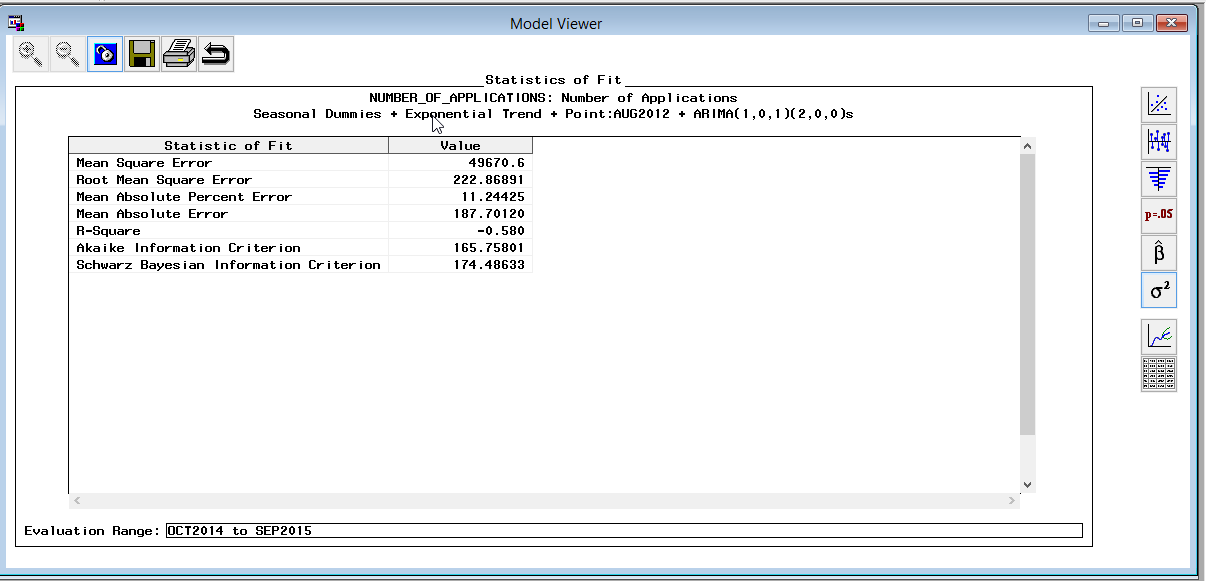
**Seasonal dummies+ exponential trend + Point: AUG2012 + ARIMA (1, 0, 1) (2,0,0):**

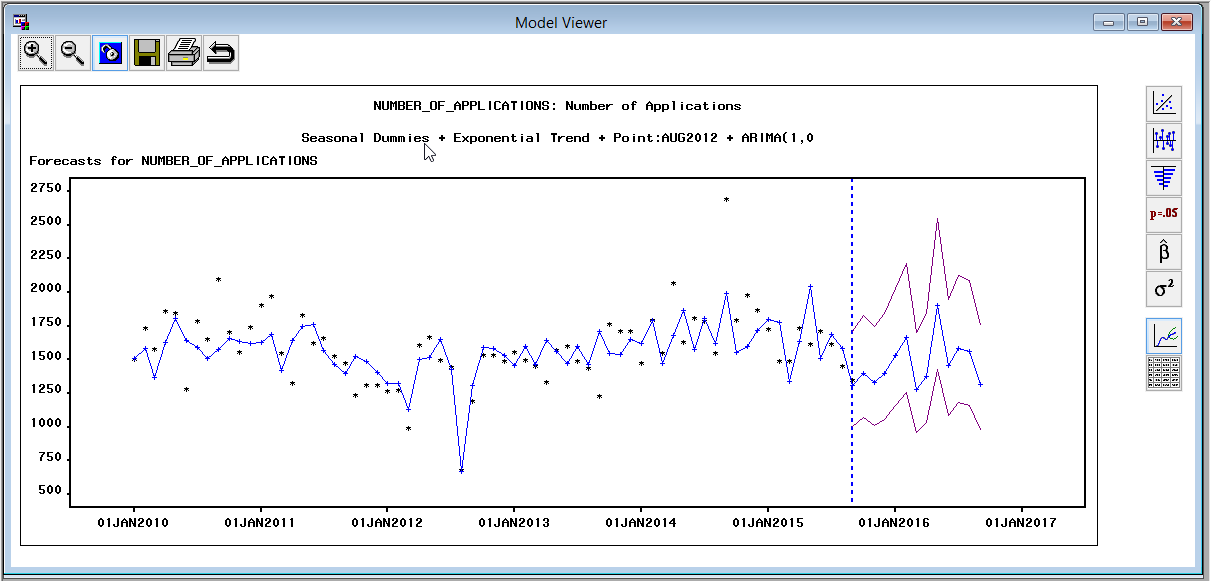


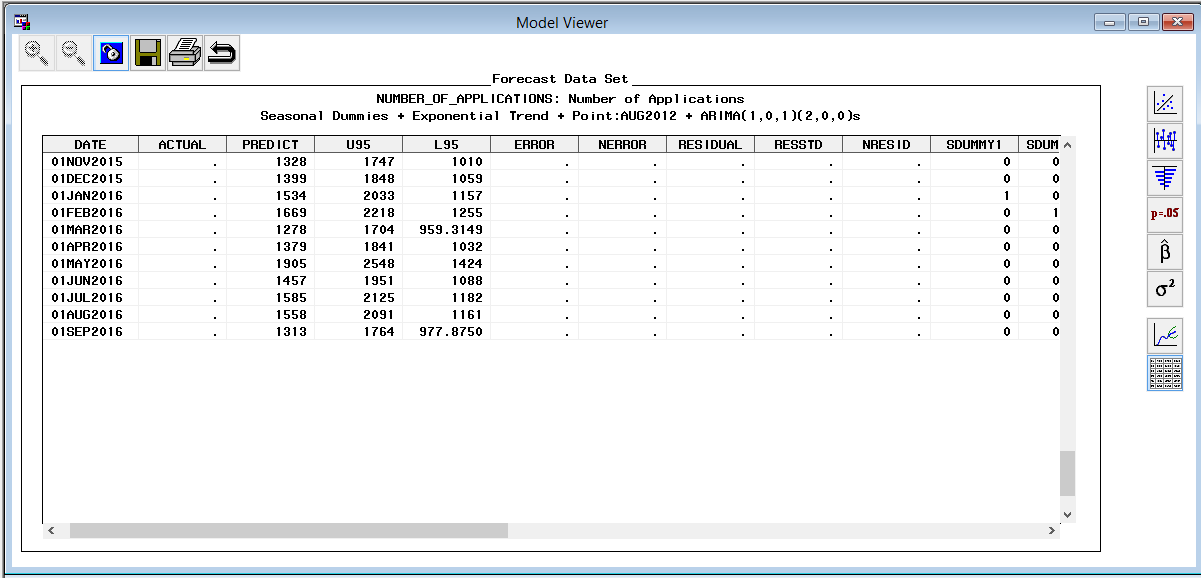




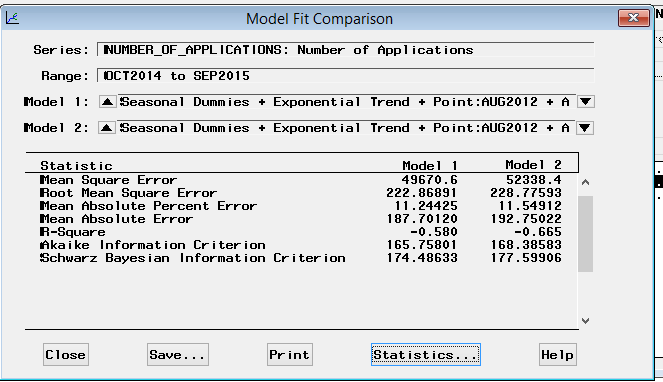




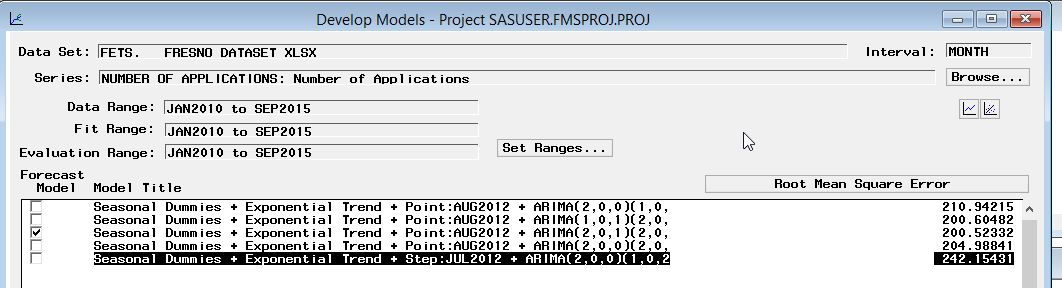
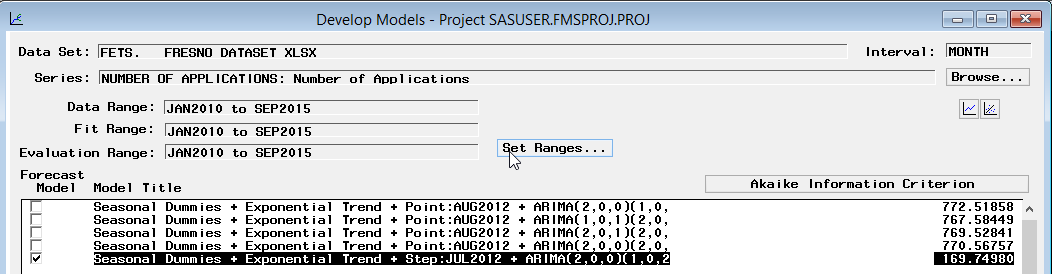
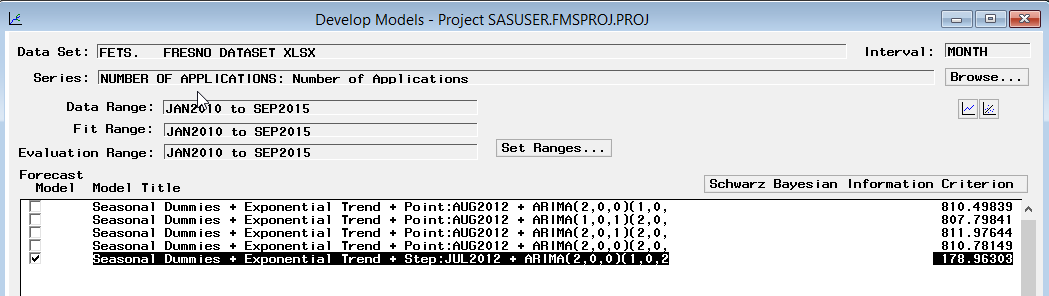




After our investigation with this data set we decided that Seasonal dummies+ exponential trend + Interventions + ARIMA (1, 0, 1) (2, 0, 0) would be the best model for this data.

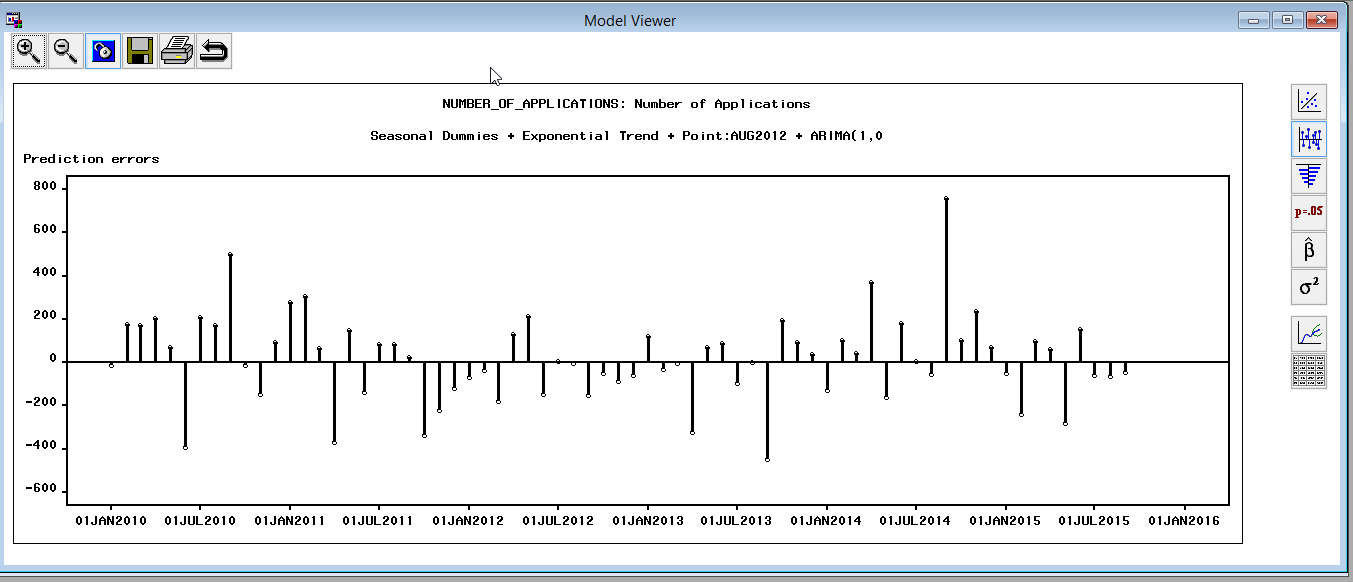
From model comparison, we can see seasonal dummies+ exponential trend + Point:Aug2012 + ARIMA(1,0,1)(2,0,0) has better SBIC,AIC and RMSE than seasonal dummies+ exponential trend + Point:Aug2012 + ARIMA(2,0,1)(2,0,0).

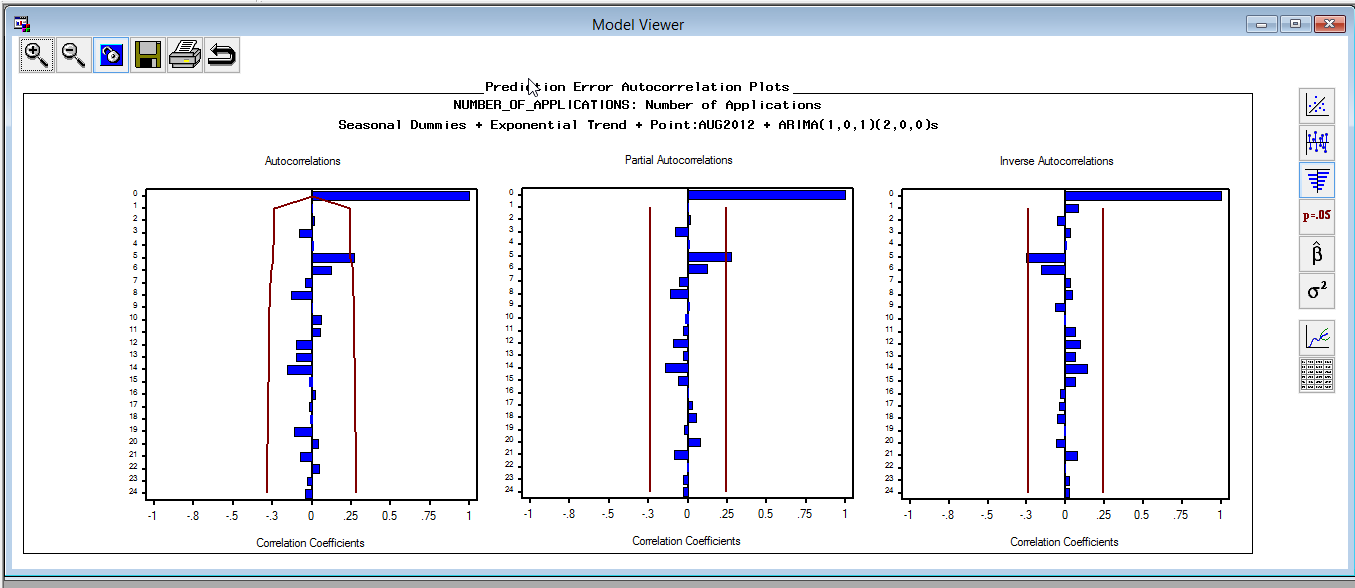
ACF, PACF, IACF is fine, only a spike at Lag 5 is suspective. White noise tests’ results of seasonal dummies+ exponential trend + Point: Aug2012 + ARIMA (1, 0, 1) (2, 0, 0) got only two spikes at 5% significant level. Unit root tests results are quite the same for most models, so we have to accept this. We will take seasonal dummies+ exponential trend + Point: Aug2012 + ARIMA (1, 0, 1) (2,0,0) as our final model.

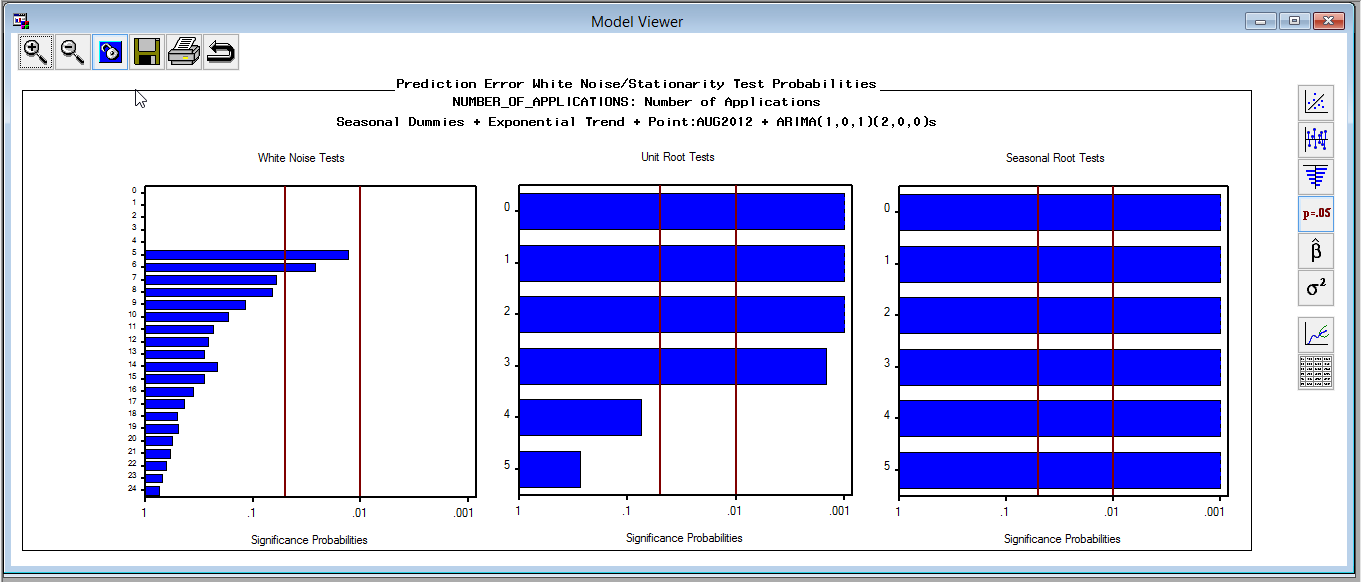
After refit with whole time span. (The last model fail to refit the whole time span)

Models’ RMSE are quite close to previous number, but there are somehow great some difference on SBIC and AIC.

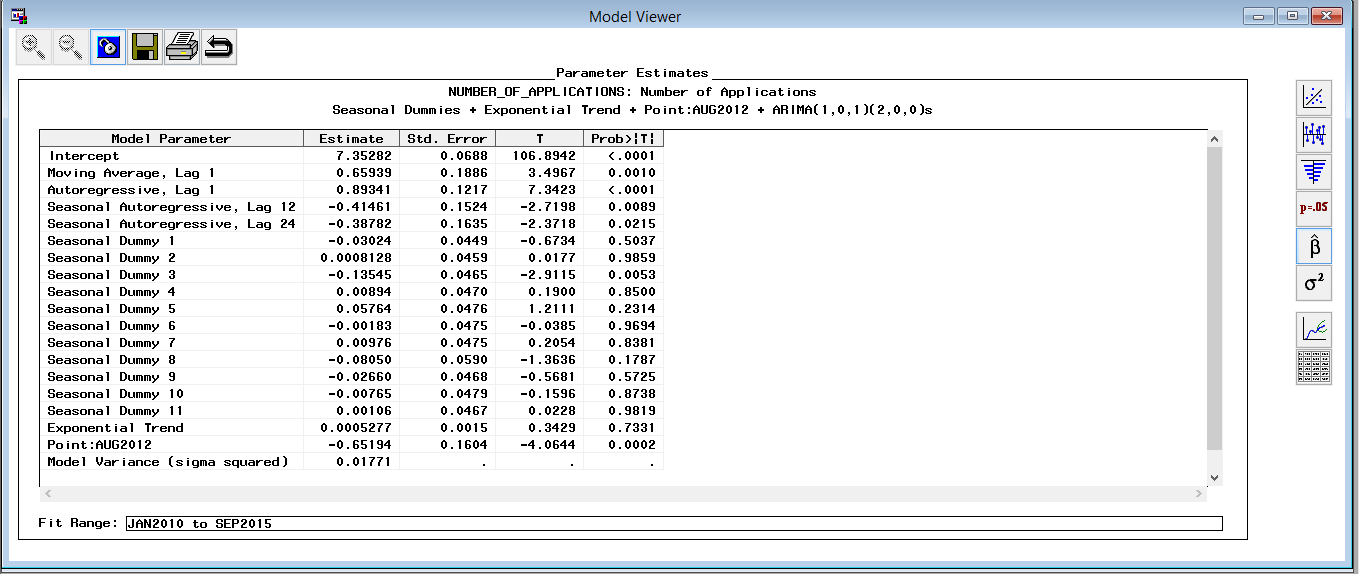
After refit with whole range time span.

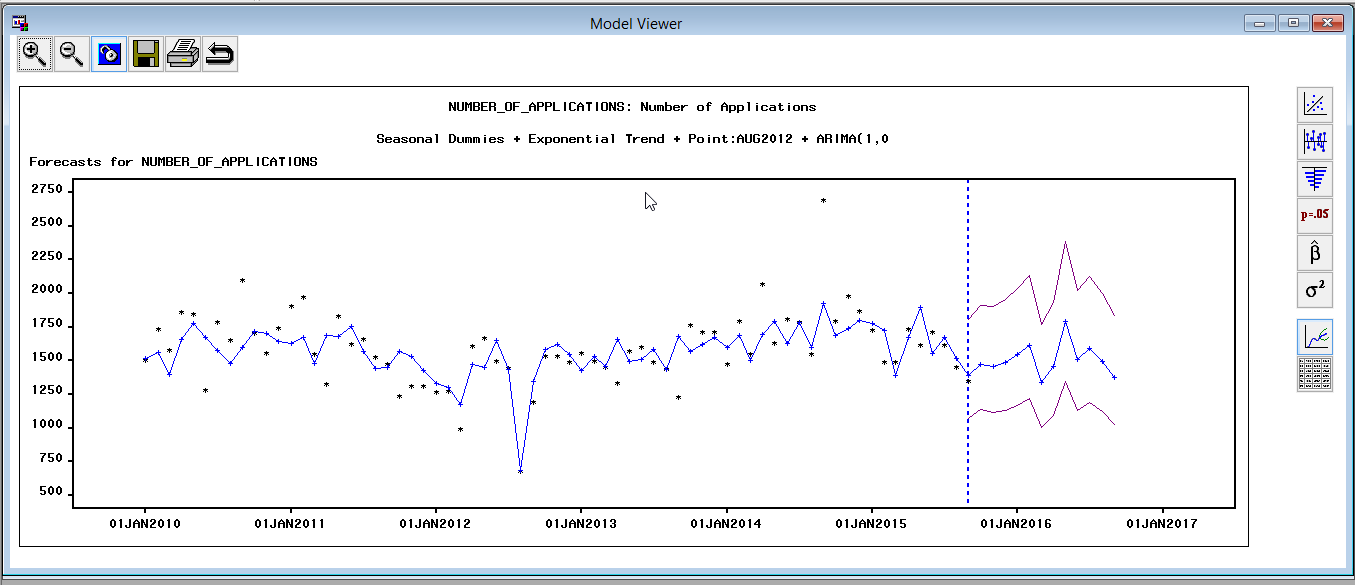


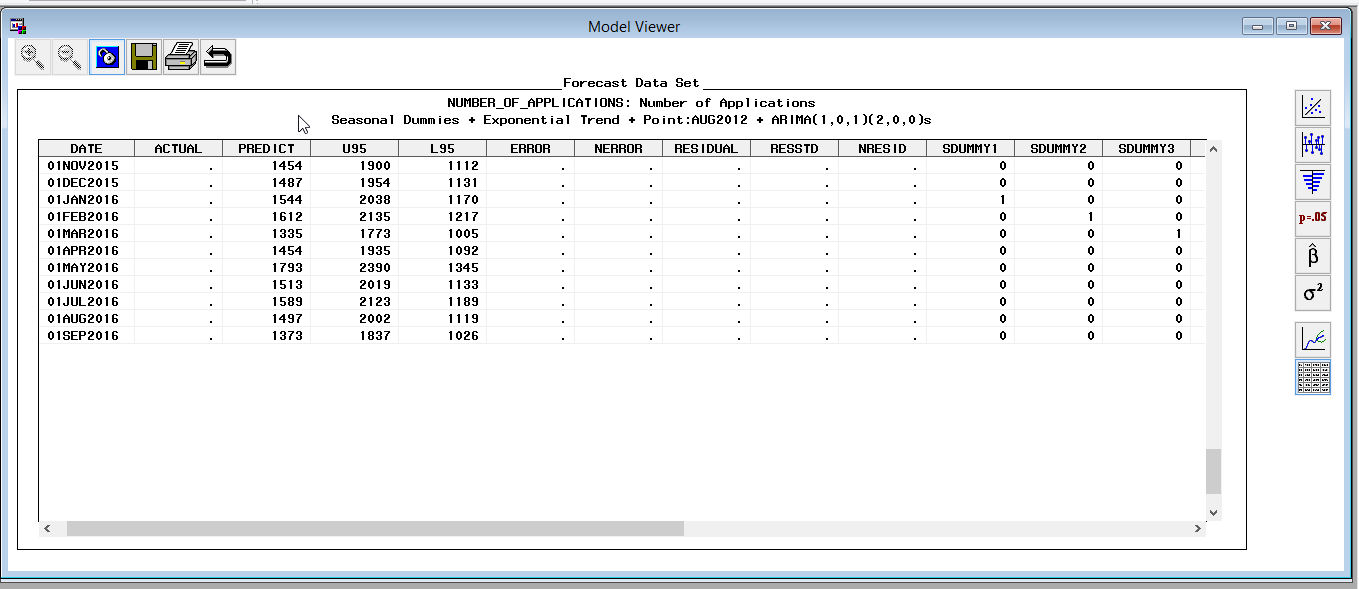




This model performed very well. It had a high forecast gradient, but the white noise and unit root tests did not perform as well as we would have hoped.

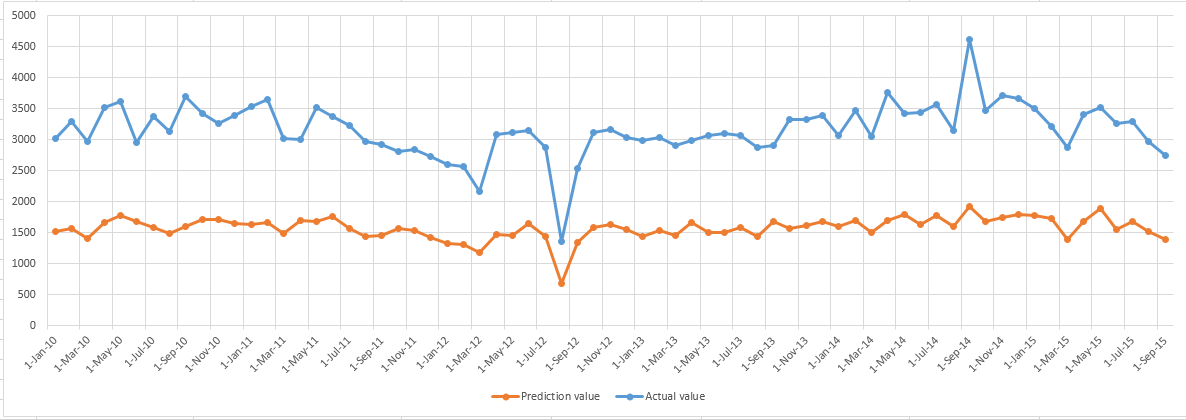






The results didn’t vary much.

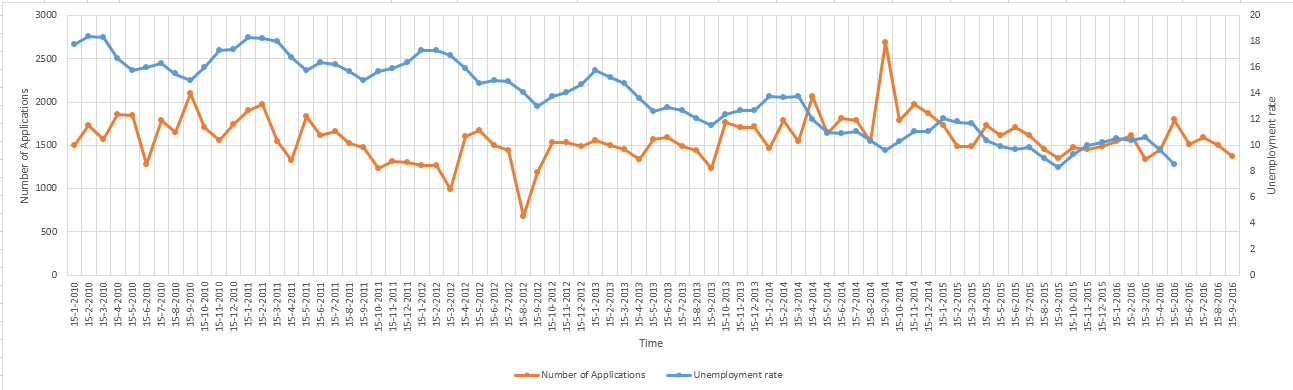
**Predicted vs. Actual**



As the model shows our model performs extremely well with the trends of the actual data, but it still appears to be slightly higher in comparison to the actual data.

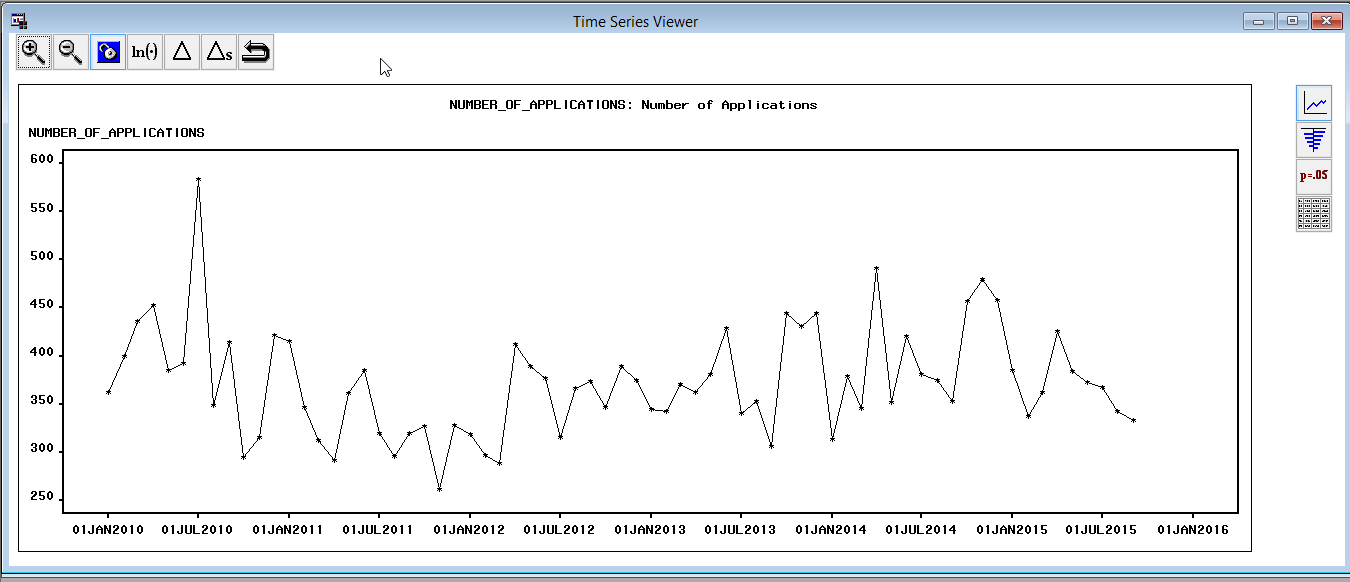
**Comparison with trend of unemployment rate.**

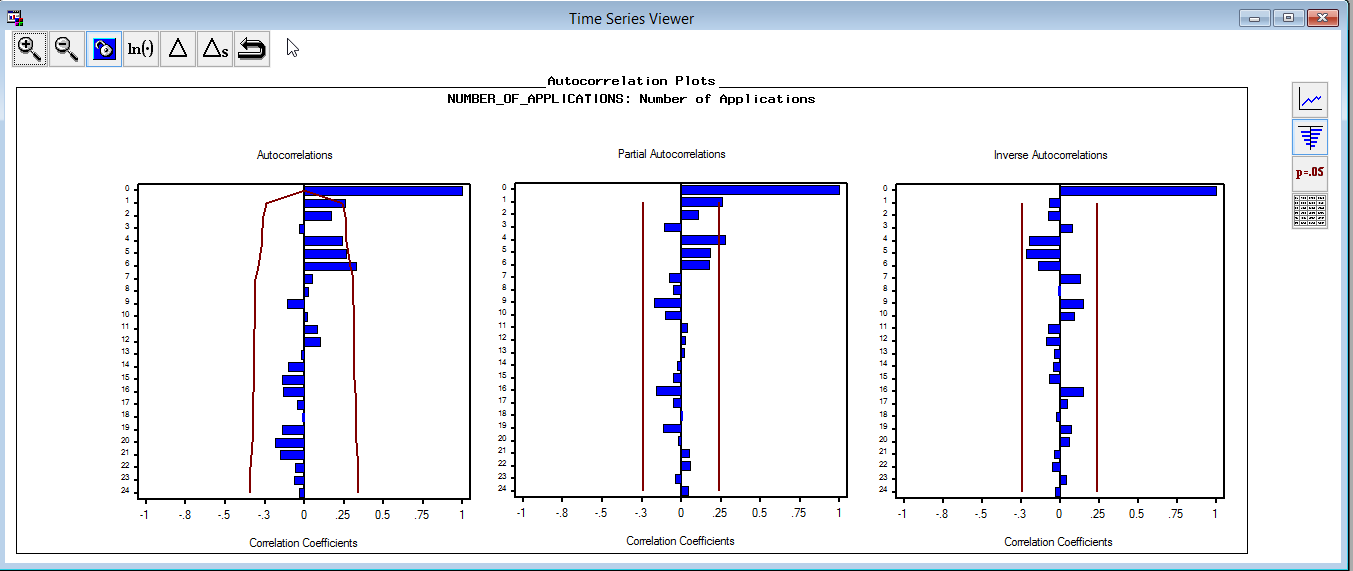
Forecast range 01NOV2015-01SEP2016

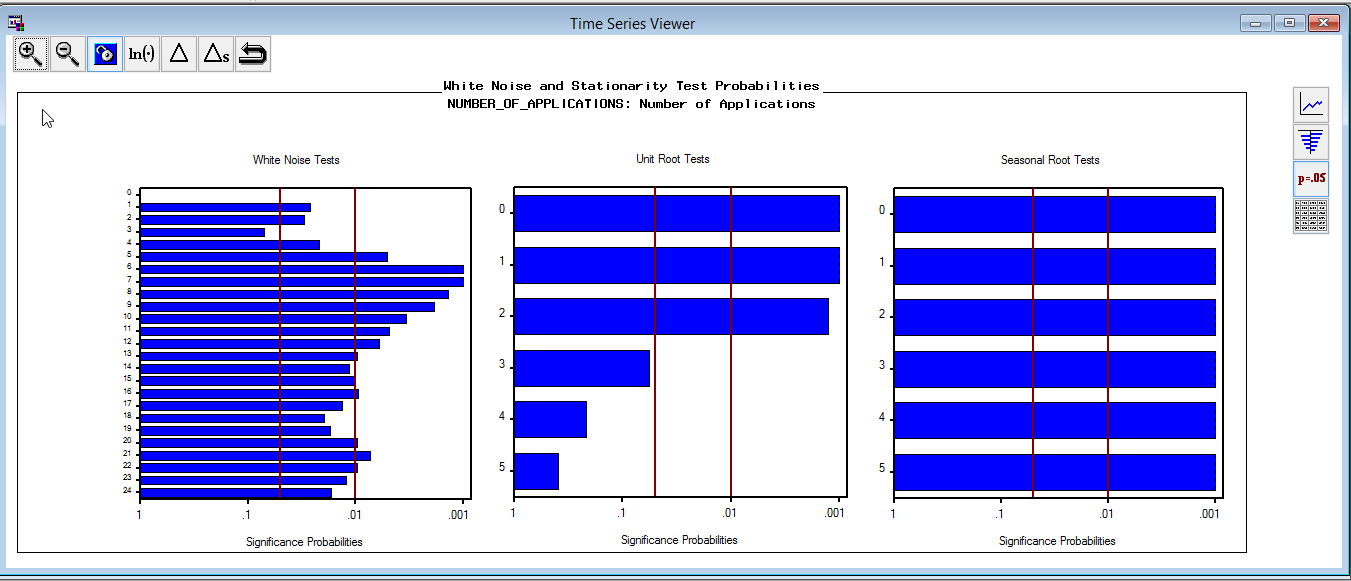


Our data shows a little bit of correlation between then unemployment rate and number of applications.

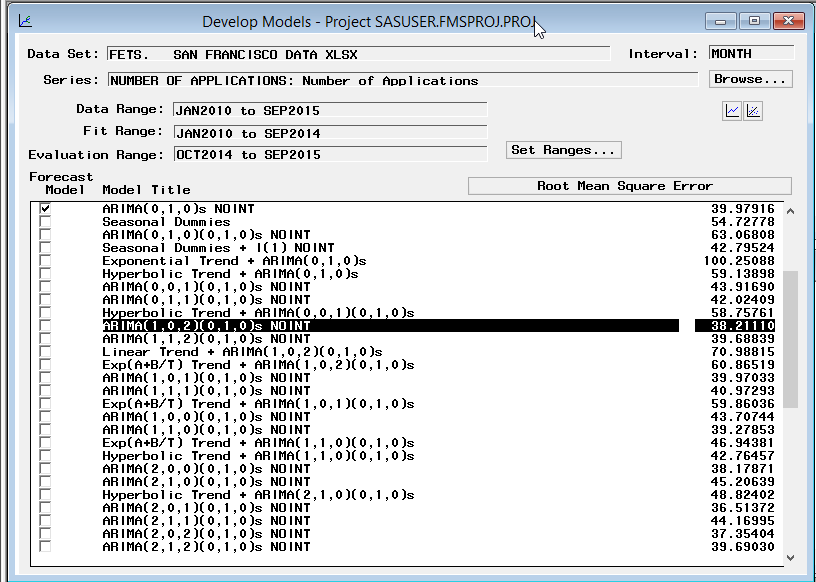
**San Francisco Data**

Original graphs:  


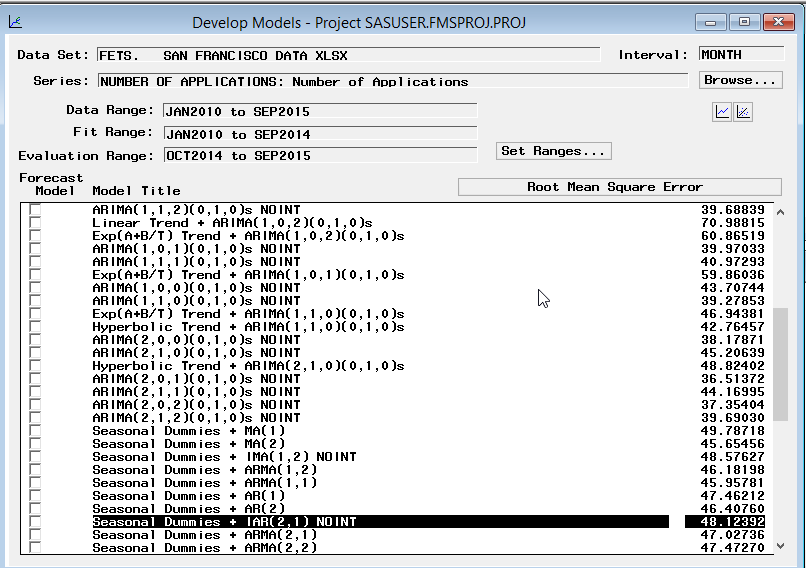




The data shows signs of seasonality and correlation.

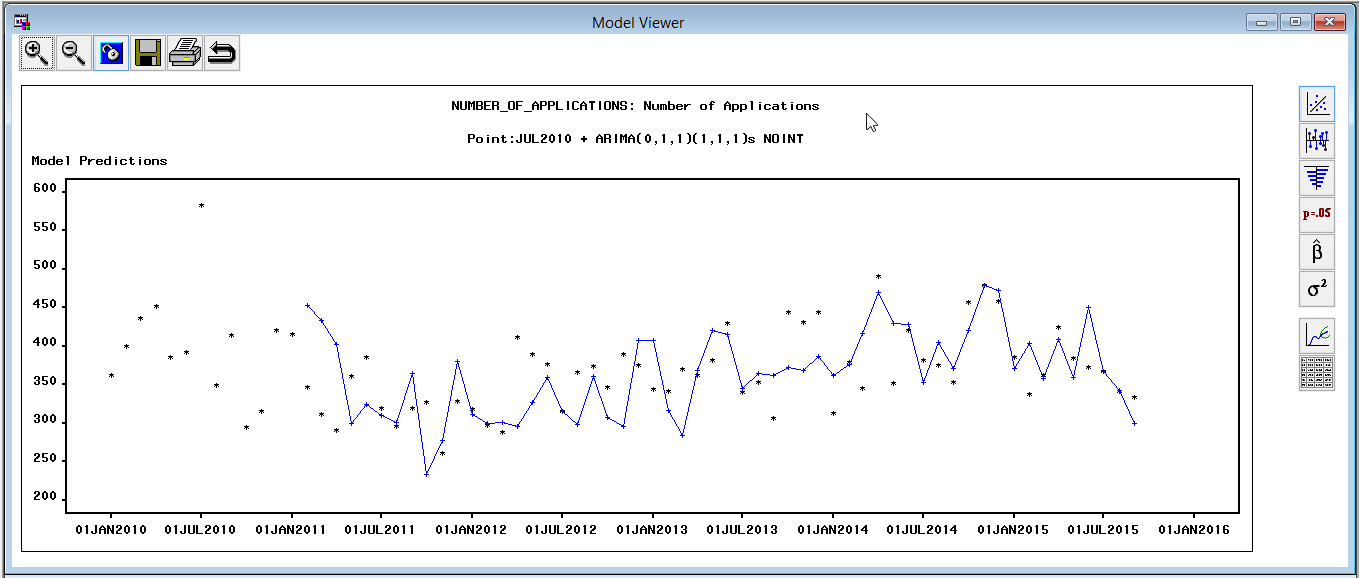
Models with seasonal difference combine with first difference give a better result than models with seasonal difference combine other trend.

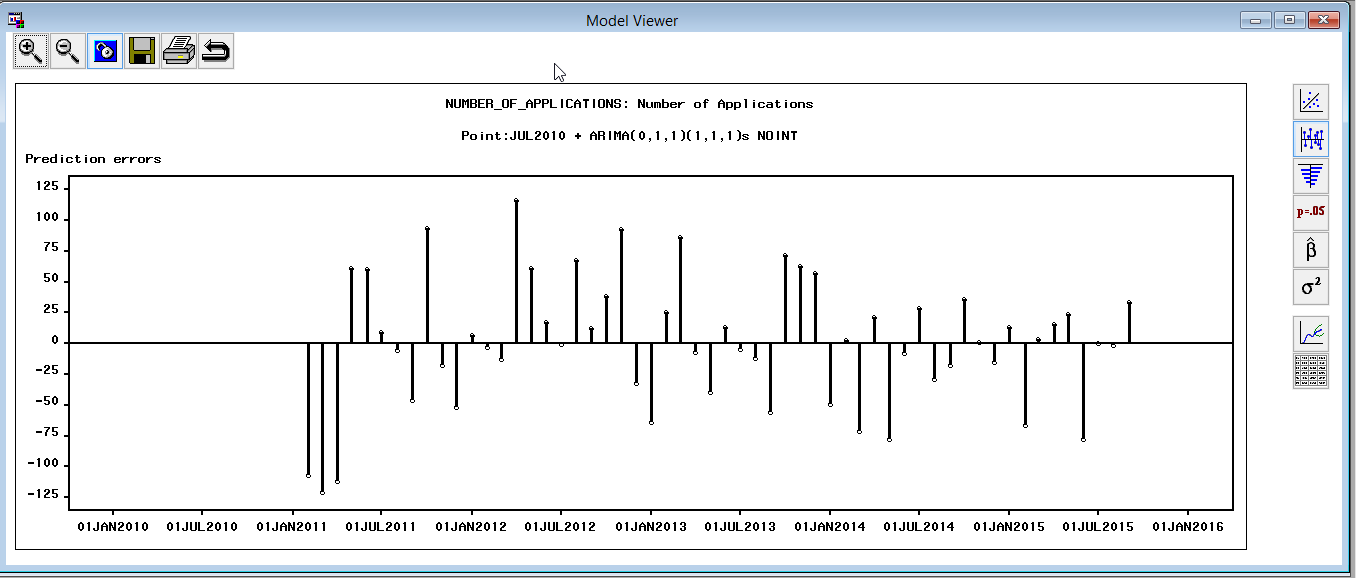
Models performed with both seasonal difference and first difference didn’t give a much better result than model with only seasonal difference.



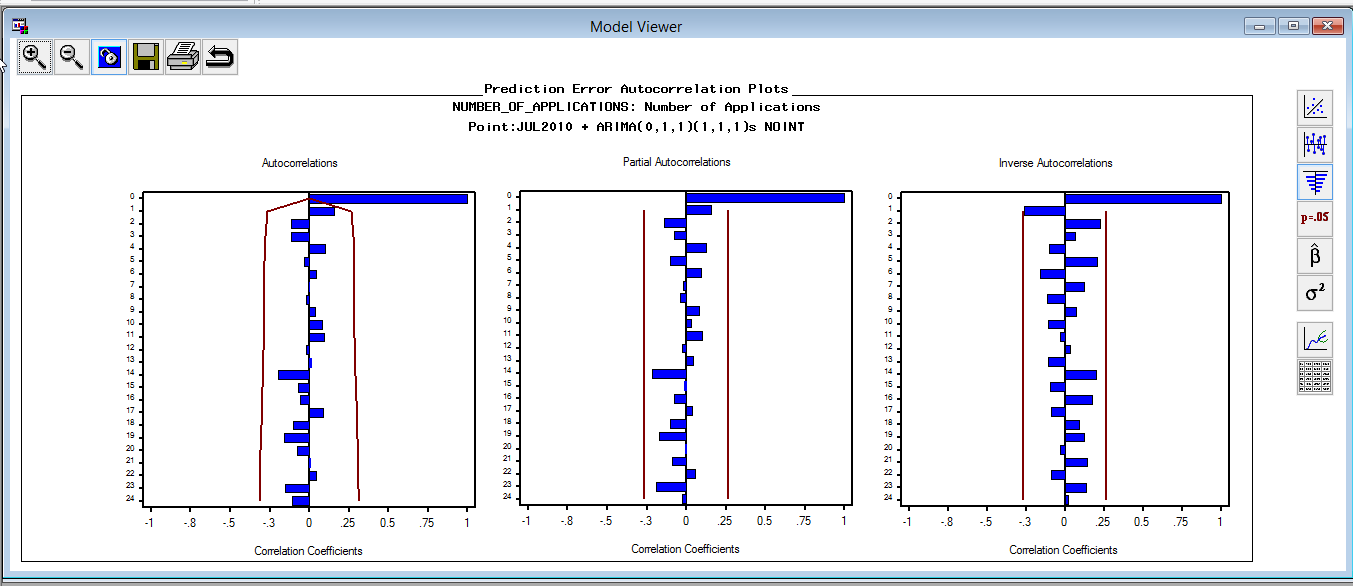
Seasonal dummies give a worse result than seasonal difference.

Using an intervention point of JUL2010+ARIMA(0,1,1)(1,1,1) gives the best model:

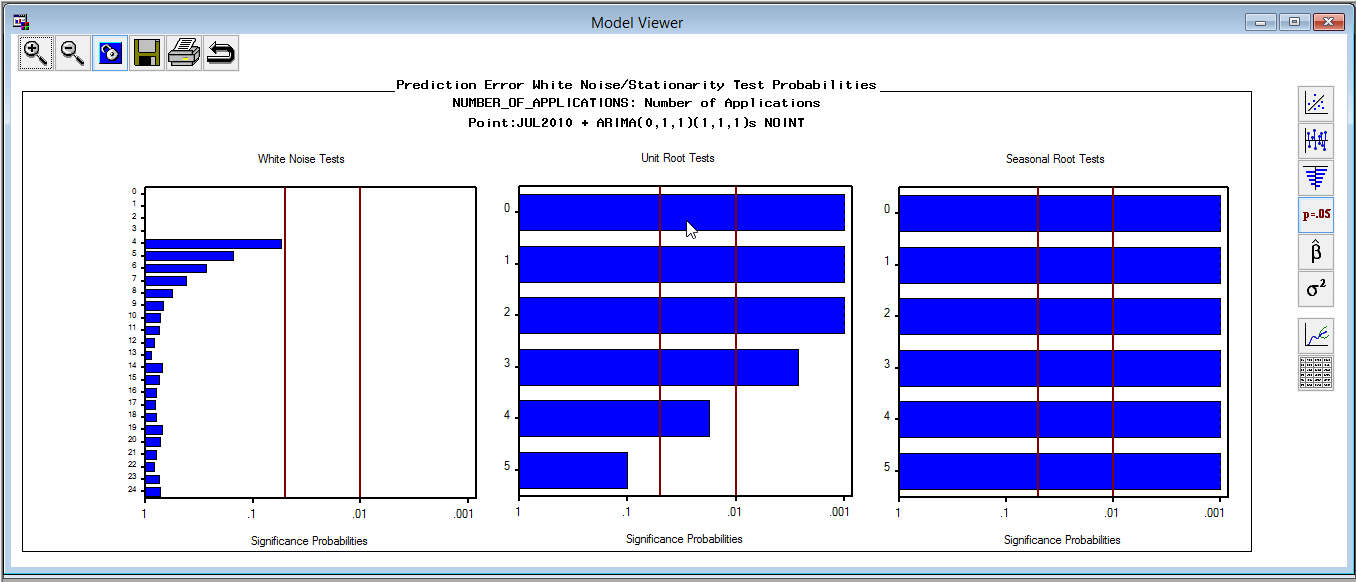




Residuals are not biased towards positive or negative side and we see that the errors are fairly distributed on either sides.

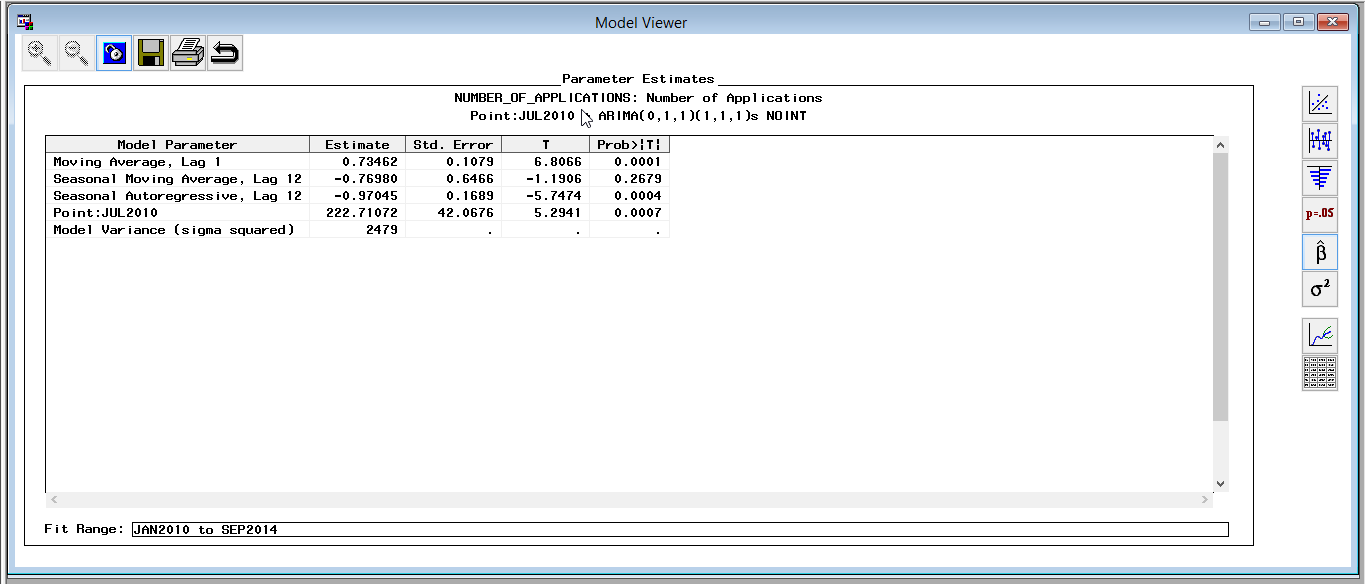


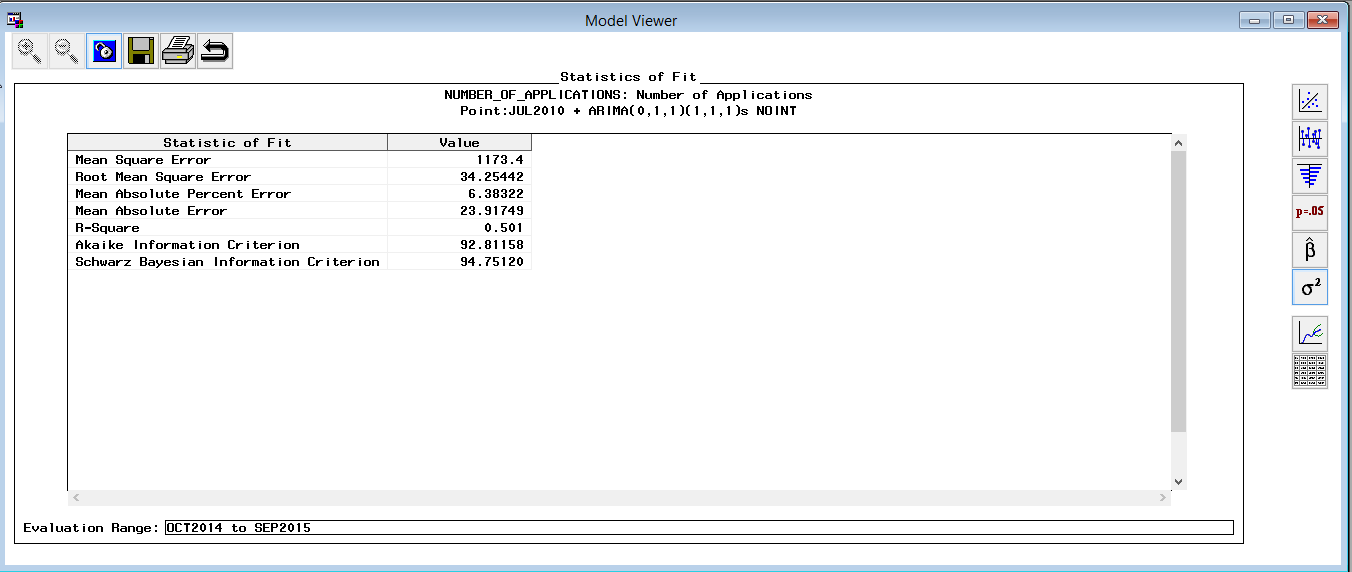
We know that model’s ACF, PACF and IACF are calculated based on the residuals and the residuals should be random in order to forecast in an efficient way. In the above plots, we observe all the lags to be within 95% confidence interval, which says that there is little correlation between the residuals and also the residuals are random.

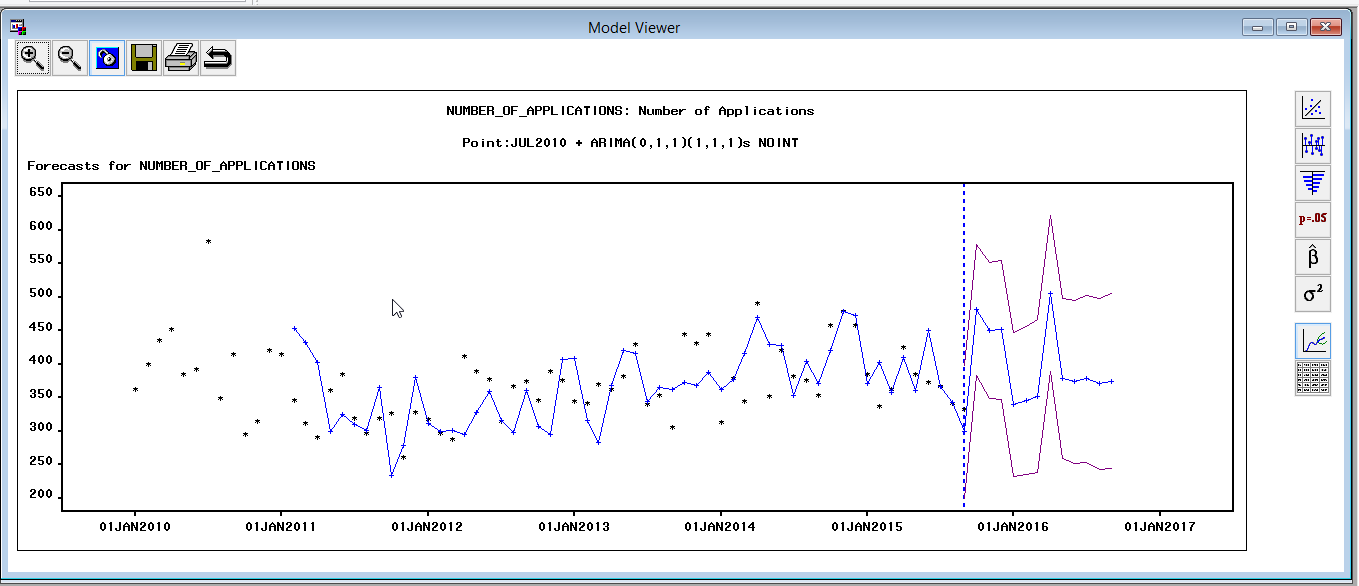


The white noise test is passed as all the lags are within the threshold value of 1%.

The unit root test explains the coefficient for the autoregressive lag and test states that the coefficient should not be 1 in any lag. This determines that our outcome is stationary. In the above plot, we observe that the unit test for all lags is significant except for lag 5. This states that our model is almost stationary.

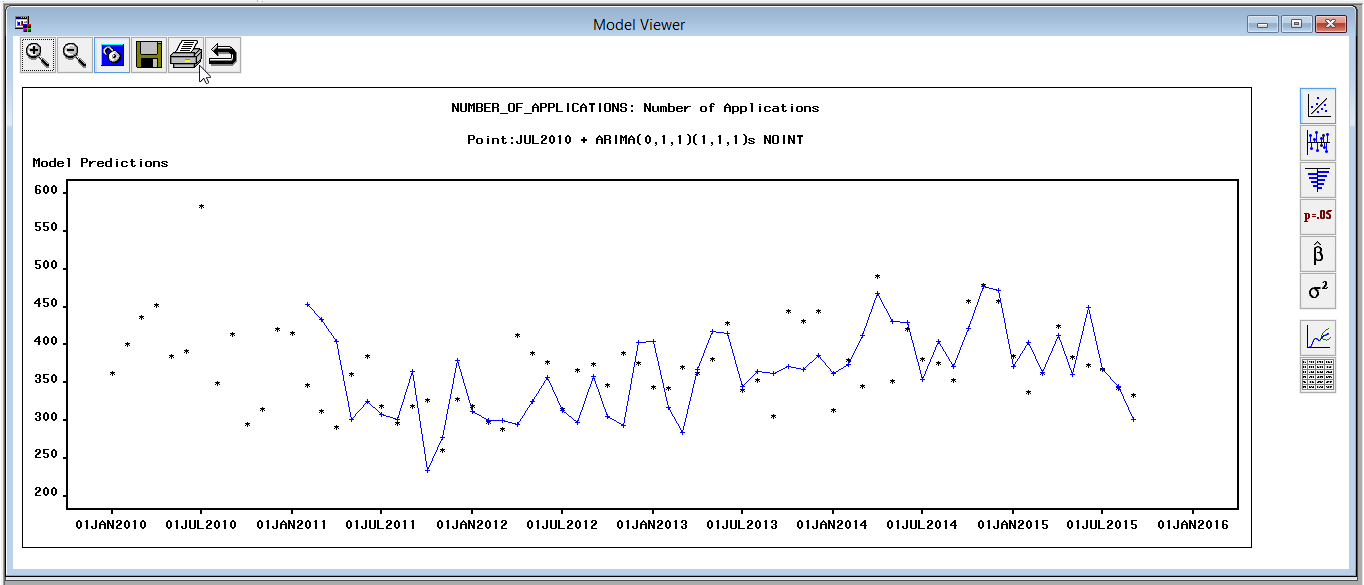


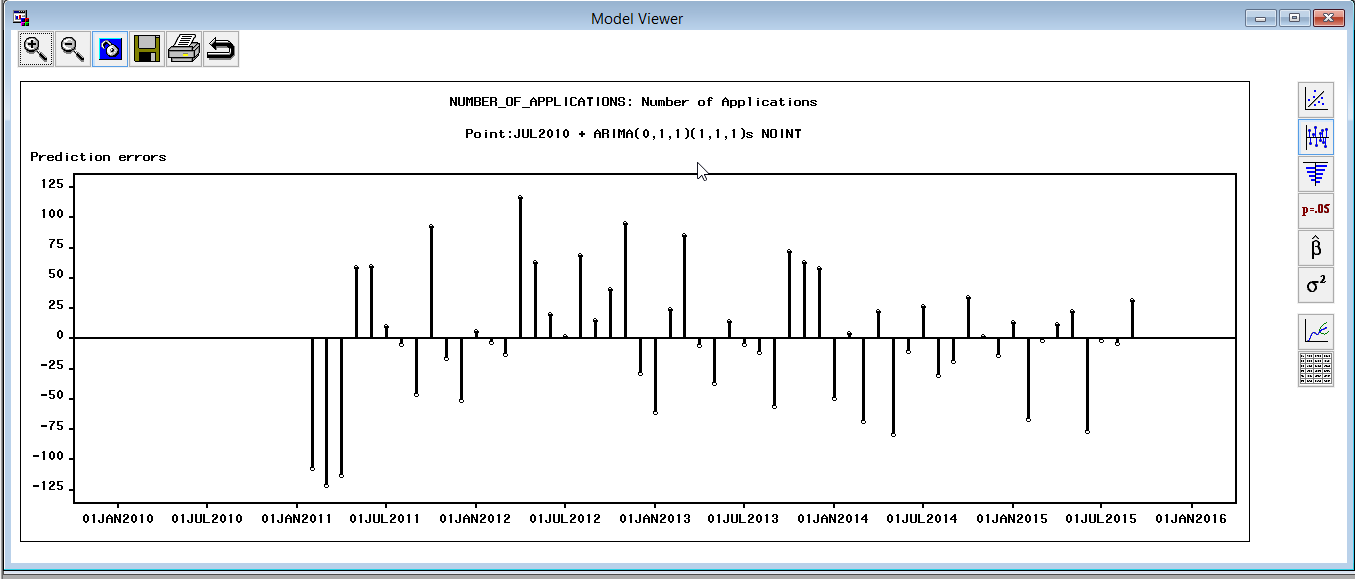




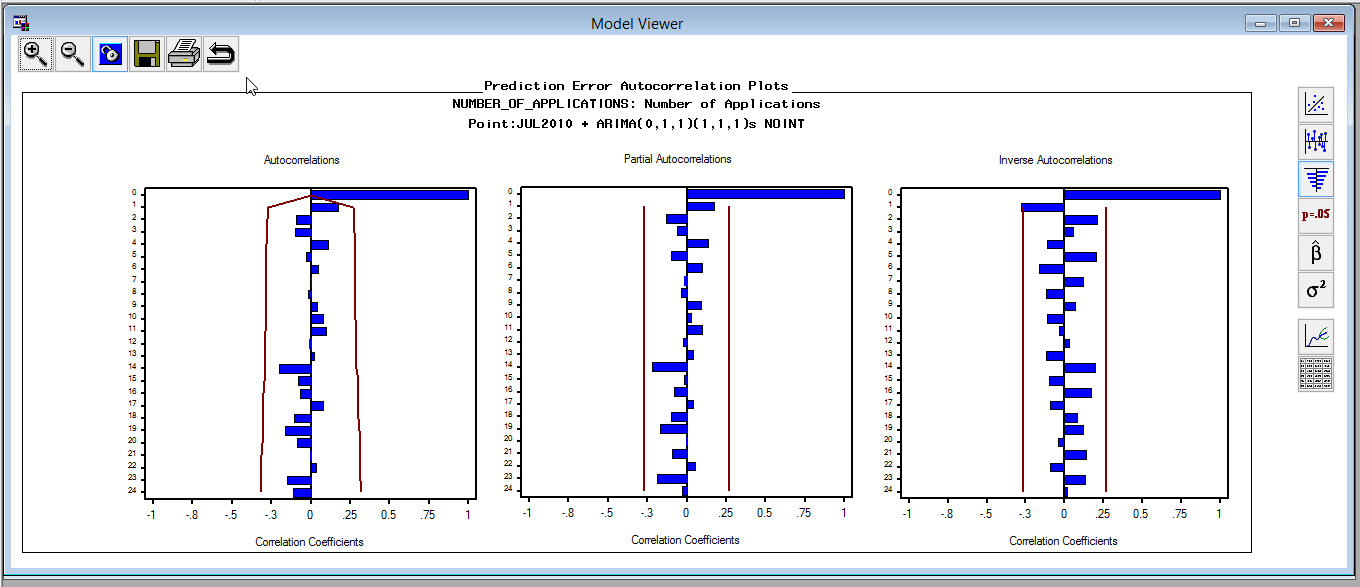
This model shows that the model passes the white noise test and that unit root test and seasonal root test. The forecast is also strong and model accuracy is high with this model.

**After refit with entire time span:**

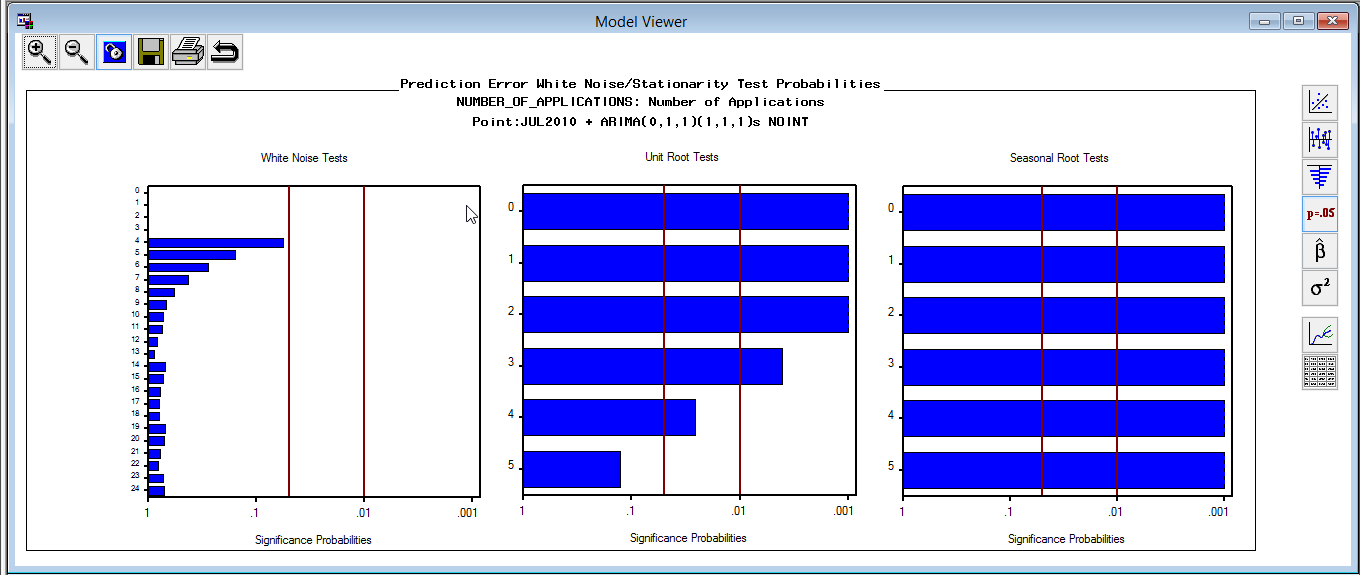




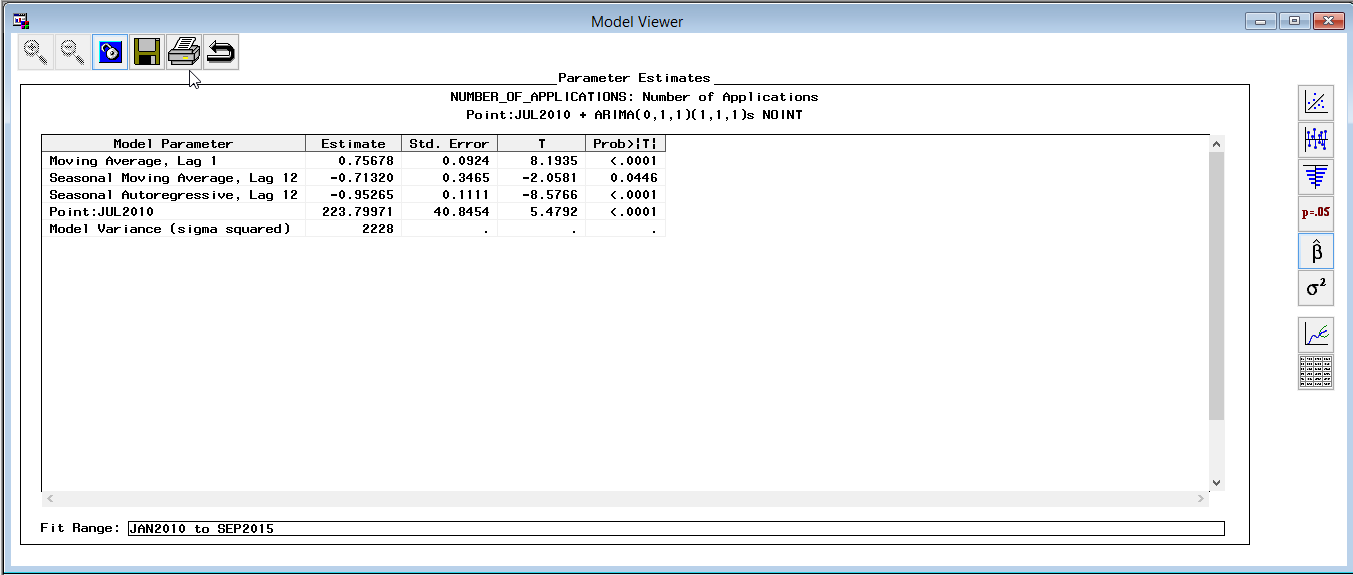
With the entire time span taken, we observe that the residuals are random on either sides.

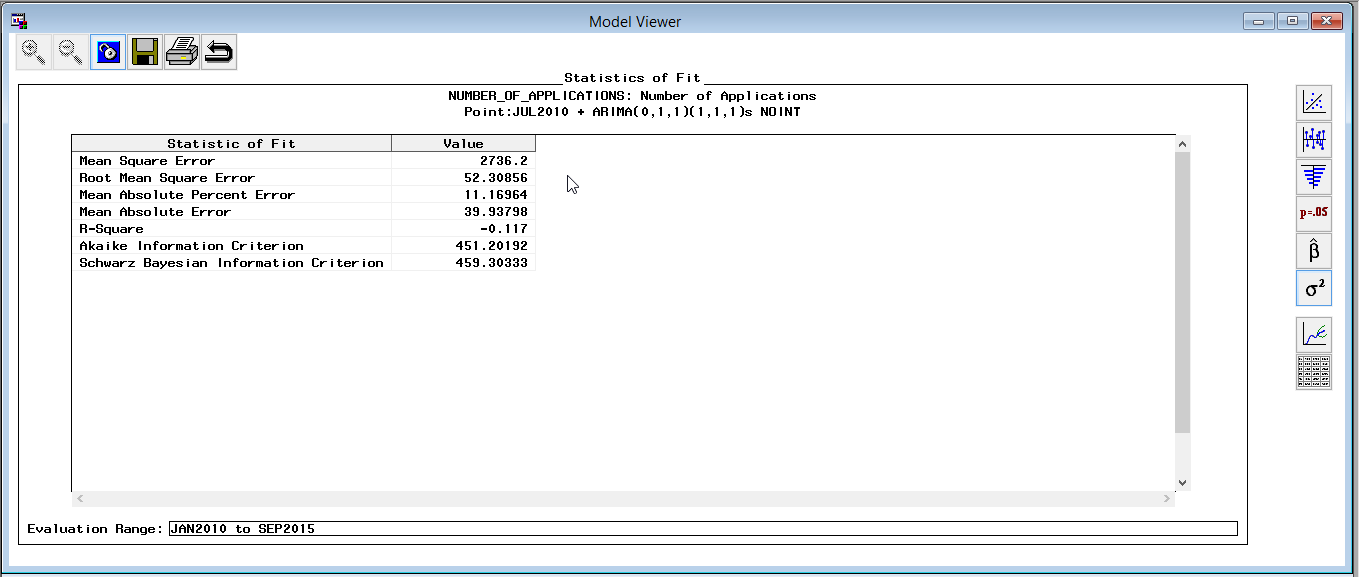


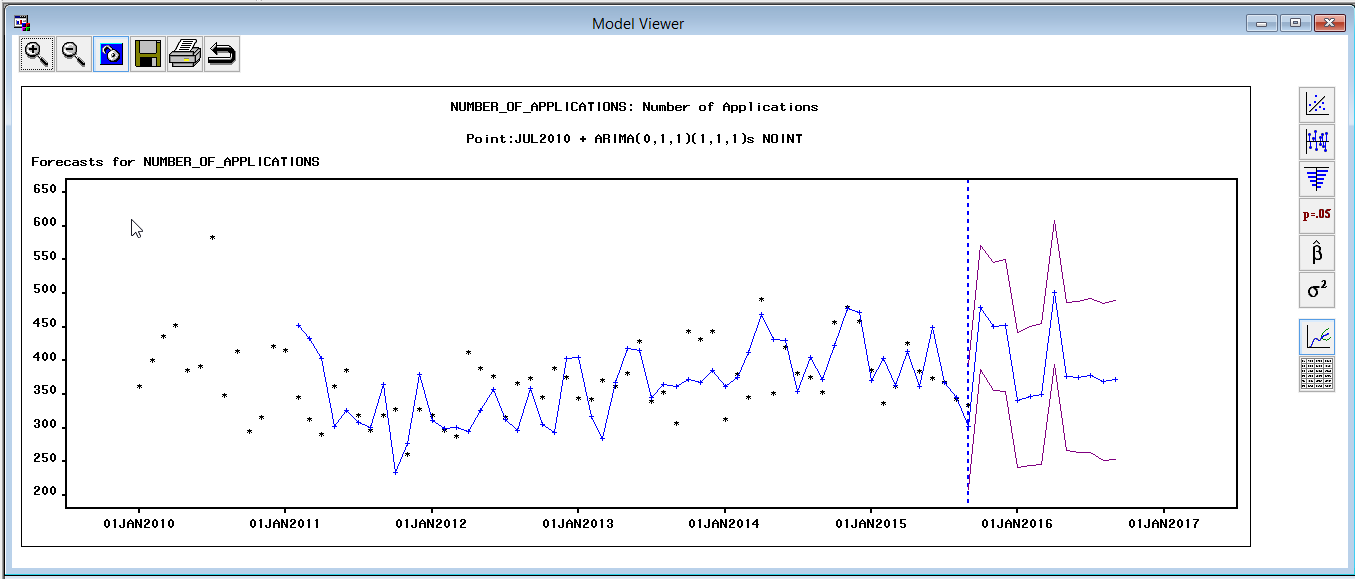
The above plots look similar to those taken with holdout sample. The ACF, PACF and IACF plots state that the residuals are randomly distributed.



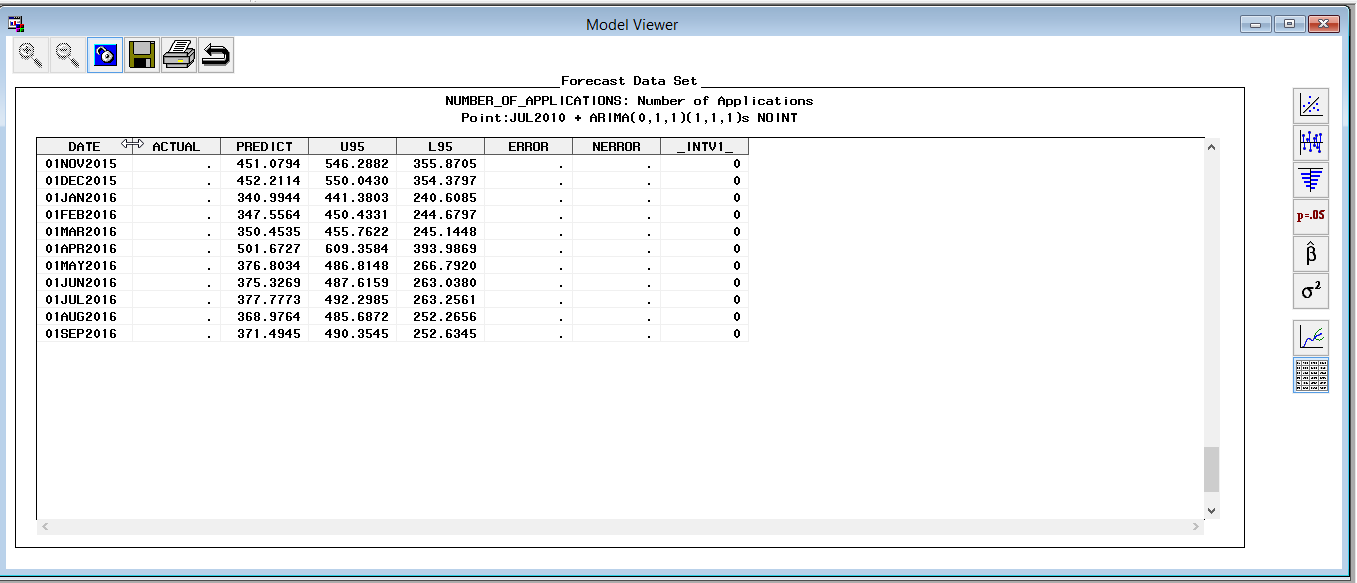
The white noise test is passed and the residuals are random.



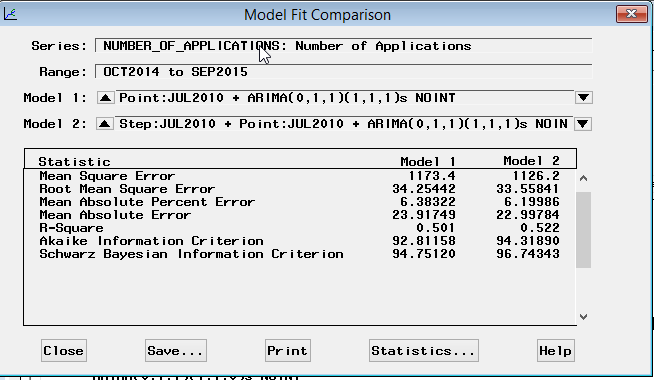
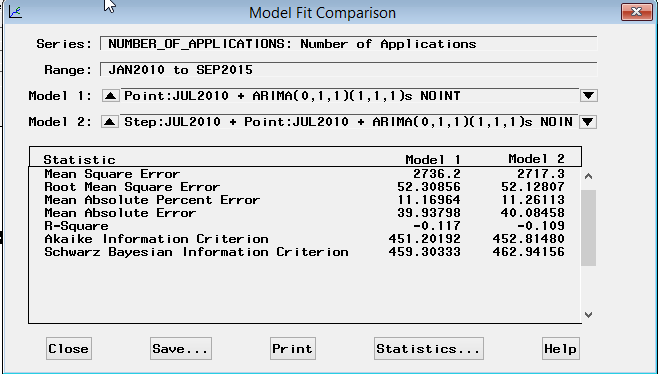




The forecast looks reasonable.

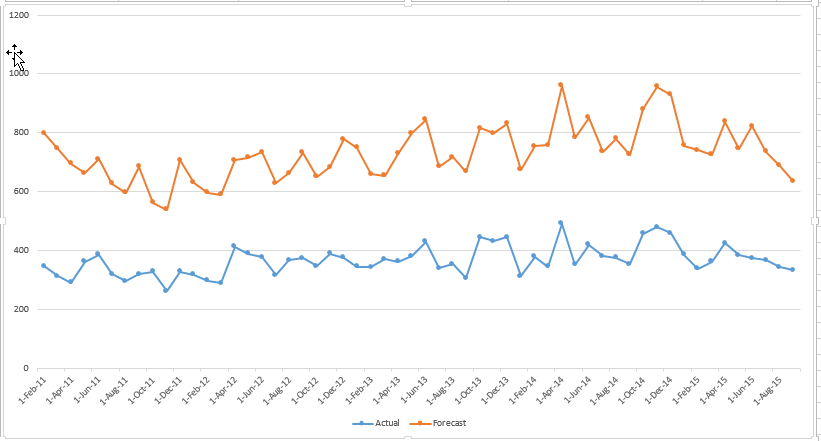


Overall, ACF , PACF, IACF looks fine. Despite there are some spikes in white noise tests, none of them excess 1% level, which is acceptable. Lag 4, 5 in unit root test are not very satisfied. Forecast results are quite precise. Compare all aspects with other models, it is the best one.



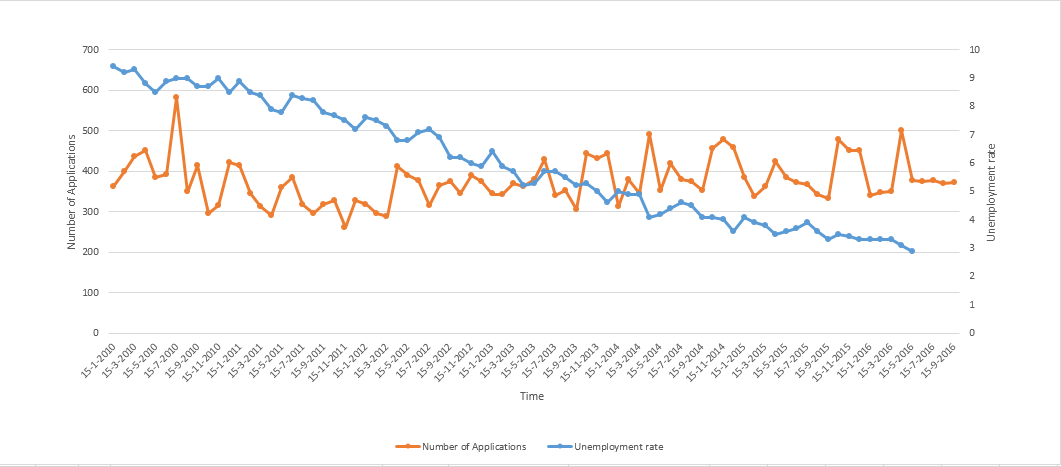
After refit, there are some huge variances in AIC, SBIC. Change in RMSE is relatively small when comparing to variances of AIC, SBIC.

**Predicted vs. Actual Forecast**

****

The trends of the predicted and actual show a strong correlation

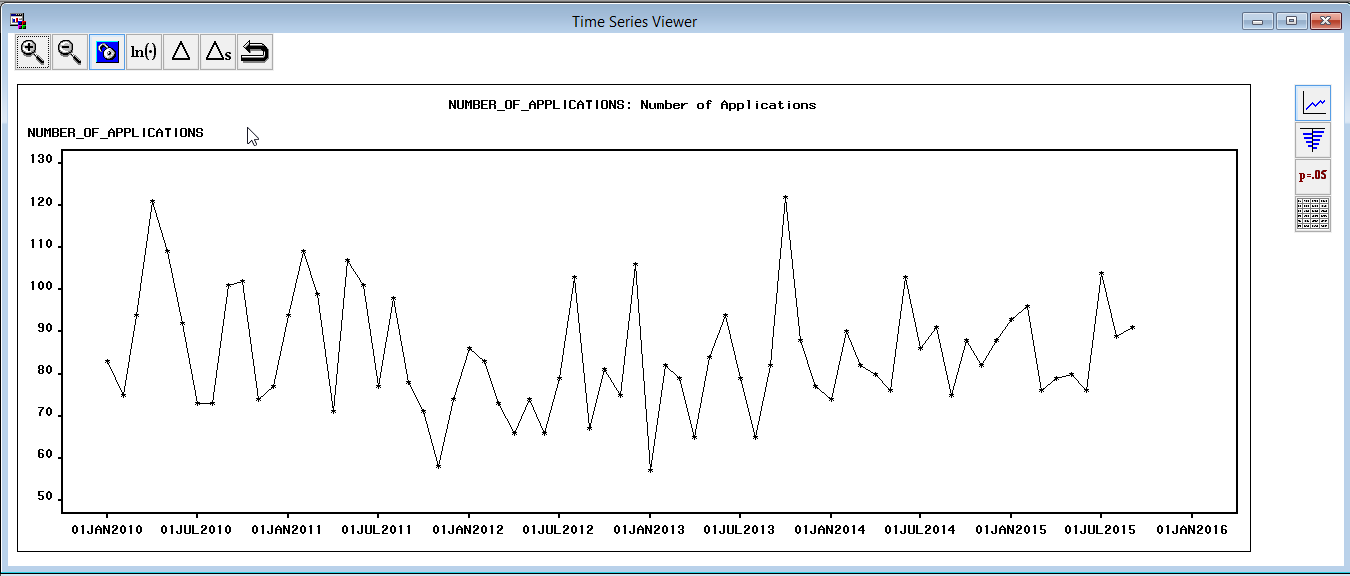
**Comparison with trend of unemployment rate.**

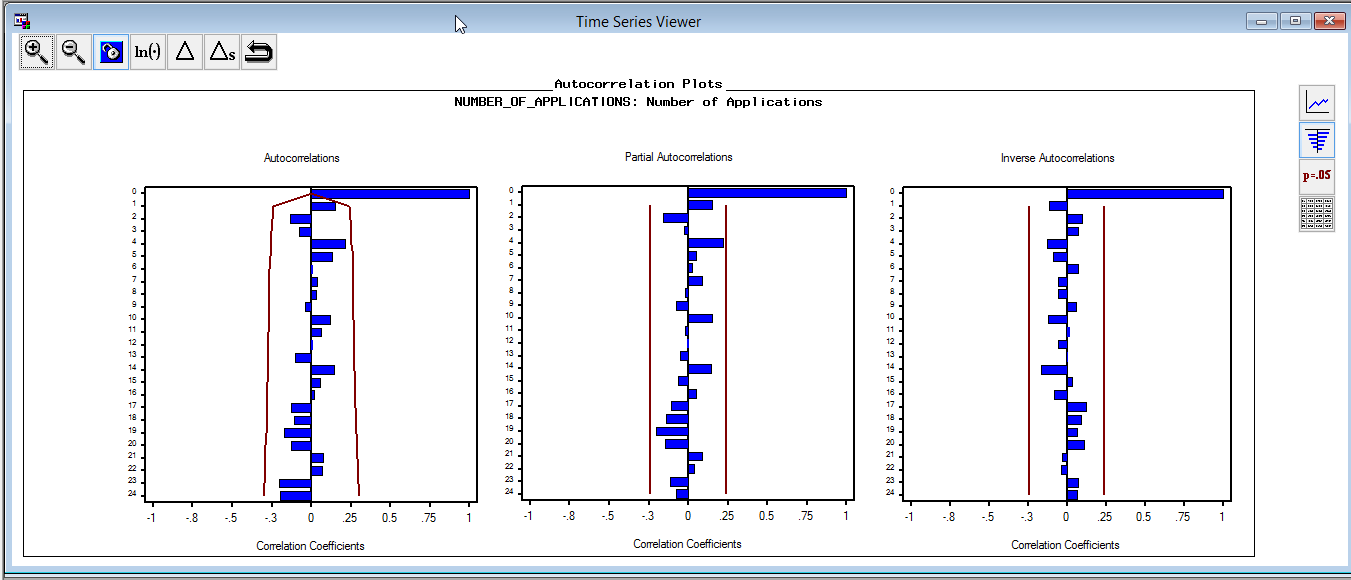
**Forecast range 01NOV2015-01SEP2016**

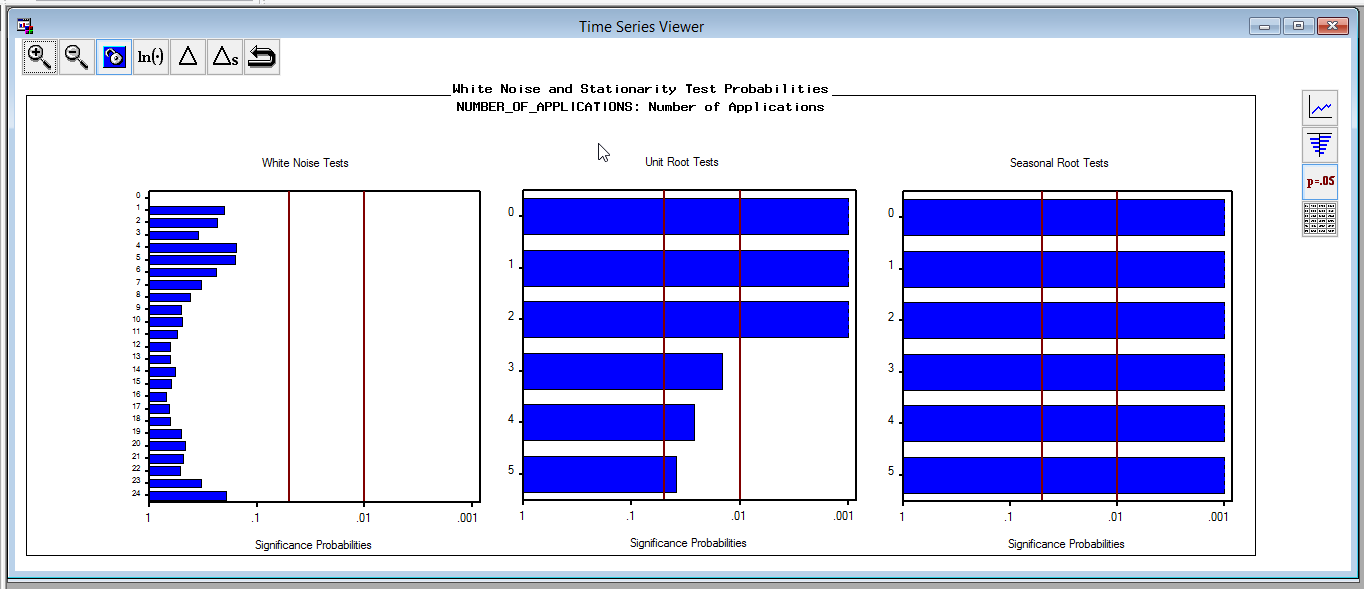
The number of unemployment is decreasing, yet applications for assistance still fluctuates. The team speculates that this could be due to the high cost of living in San Francisco.

**Siskiyou County**

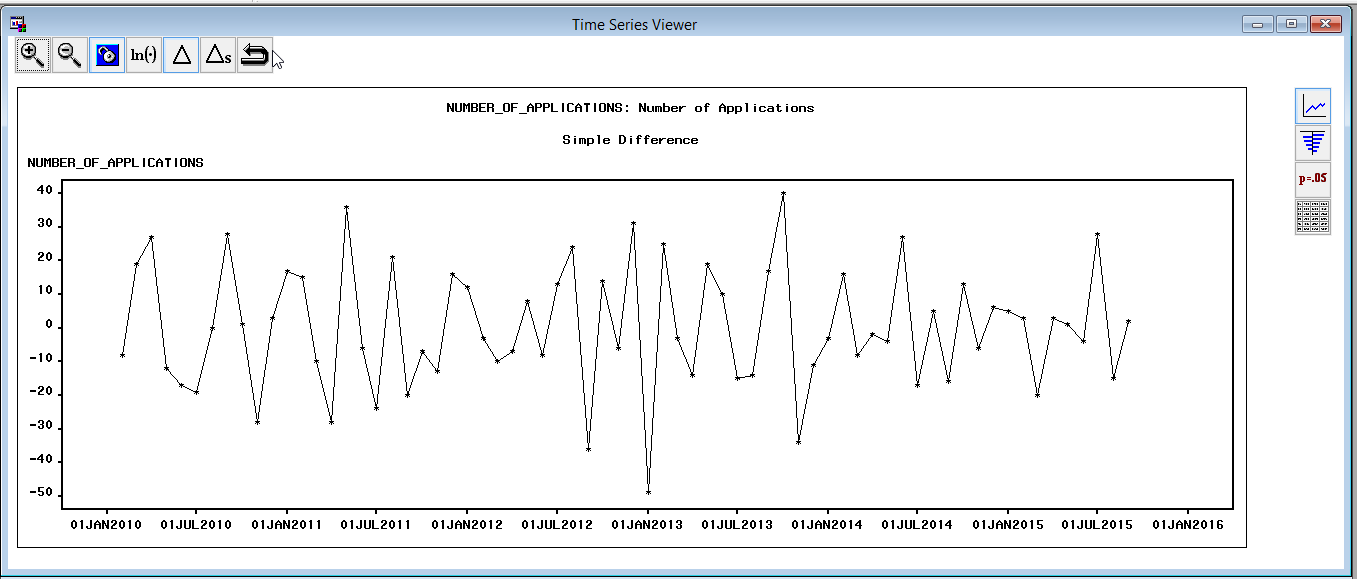
Original graphs:

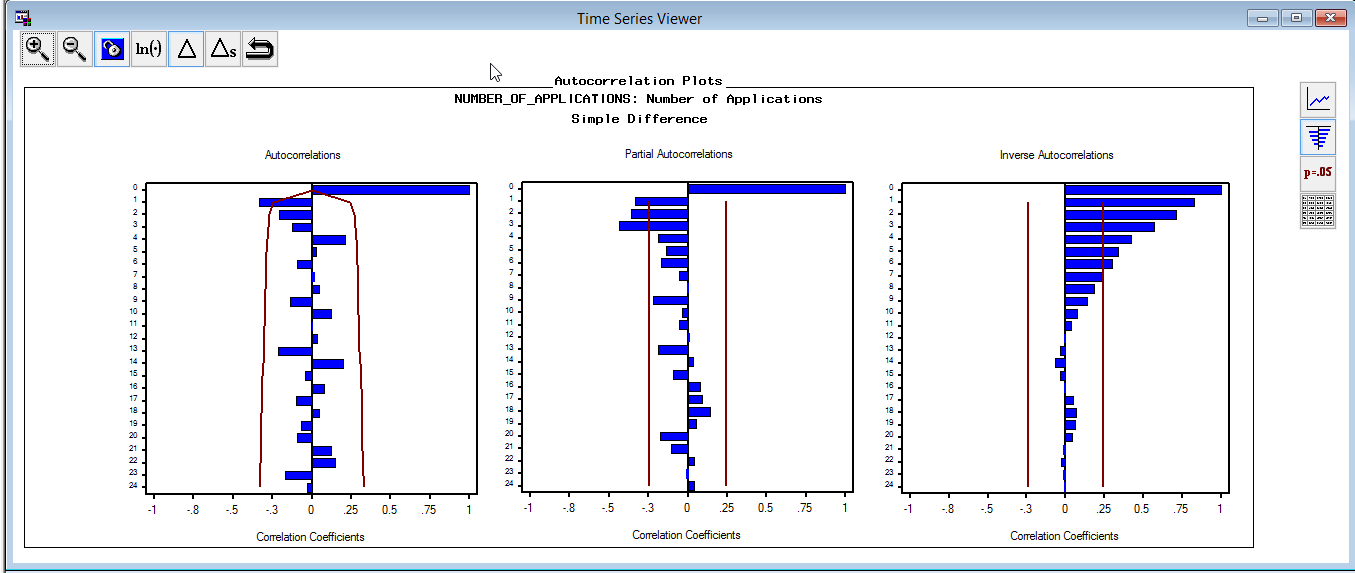


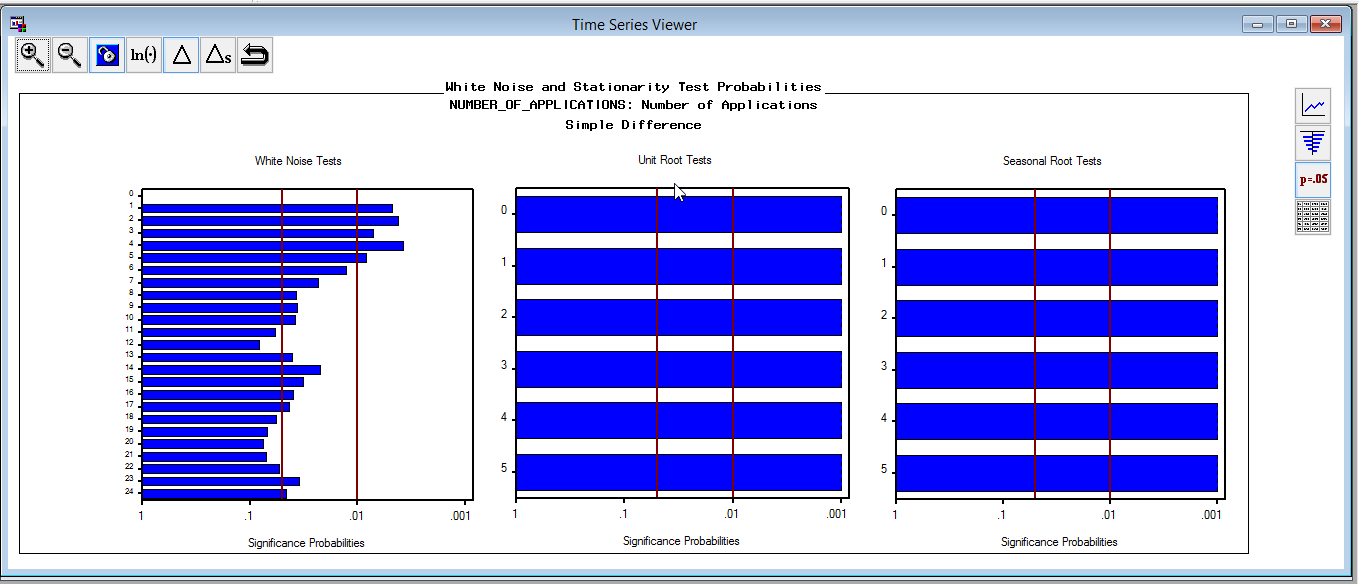




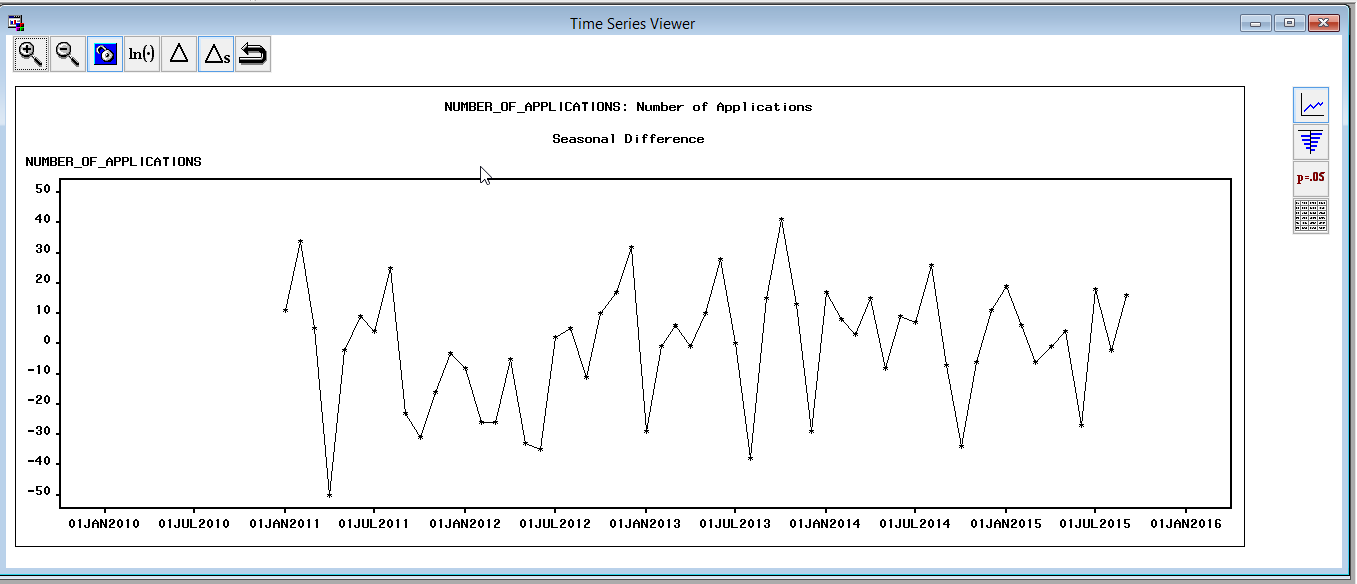
After first difference: The white noise decreases and seasonality becomes evident. Just as the other three counties, there is no trend, but there is some degree of seasonality with the data.

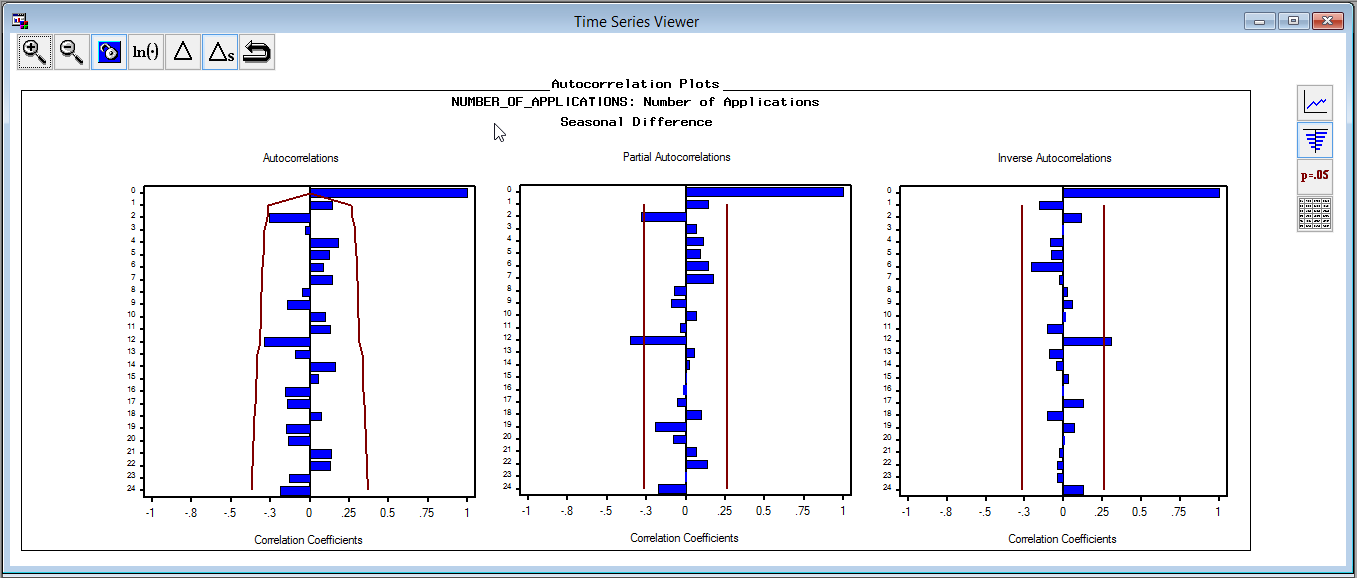


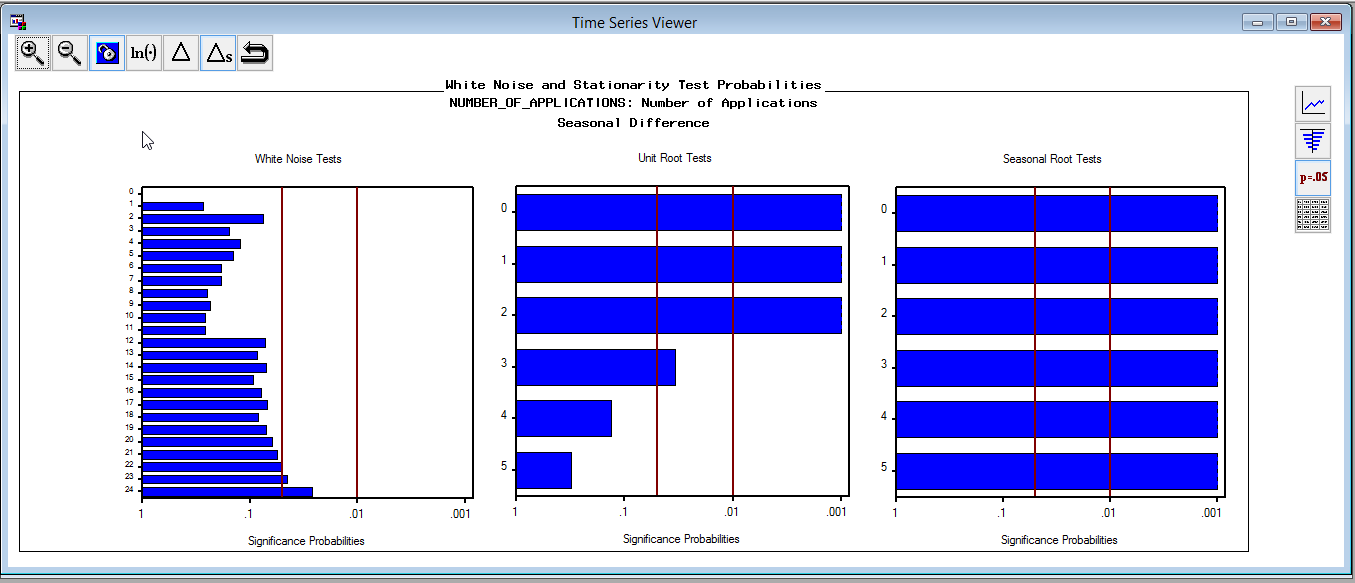


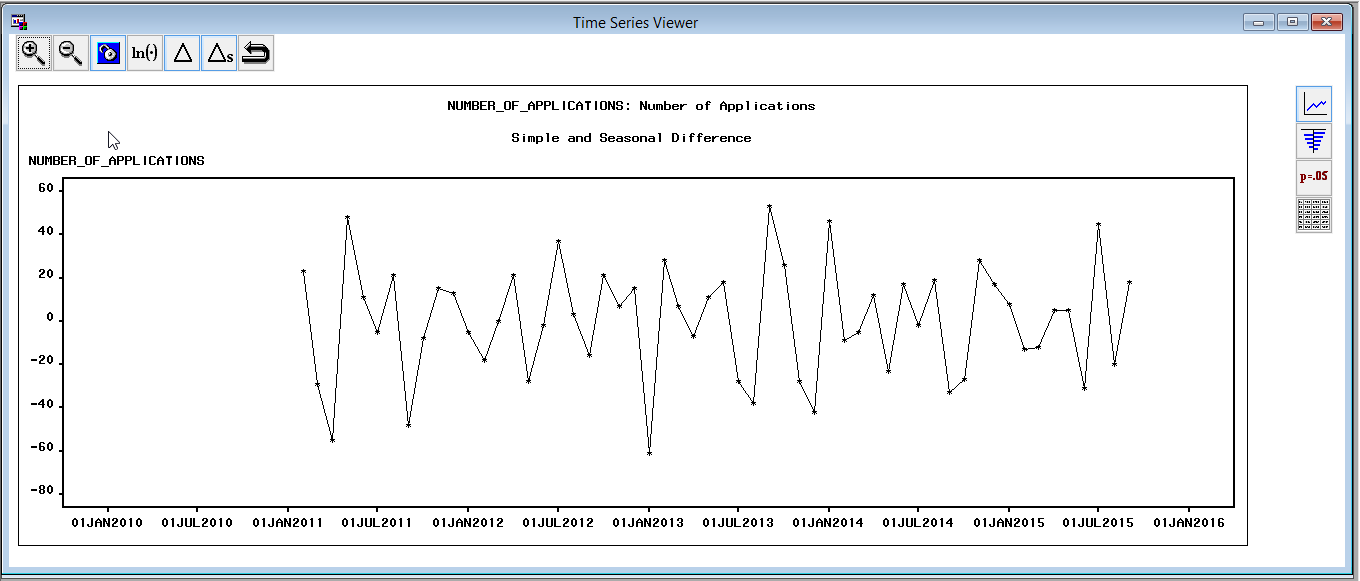


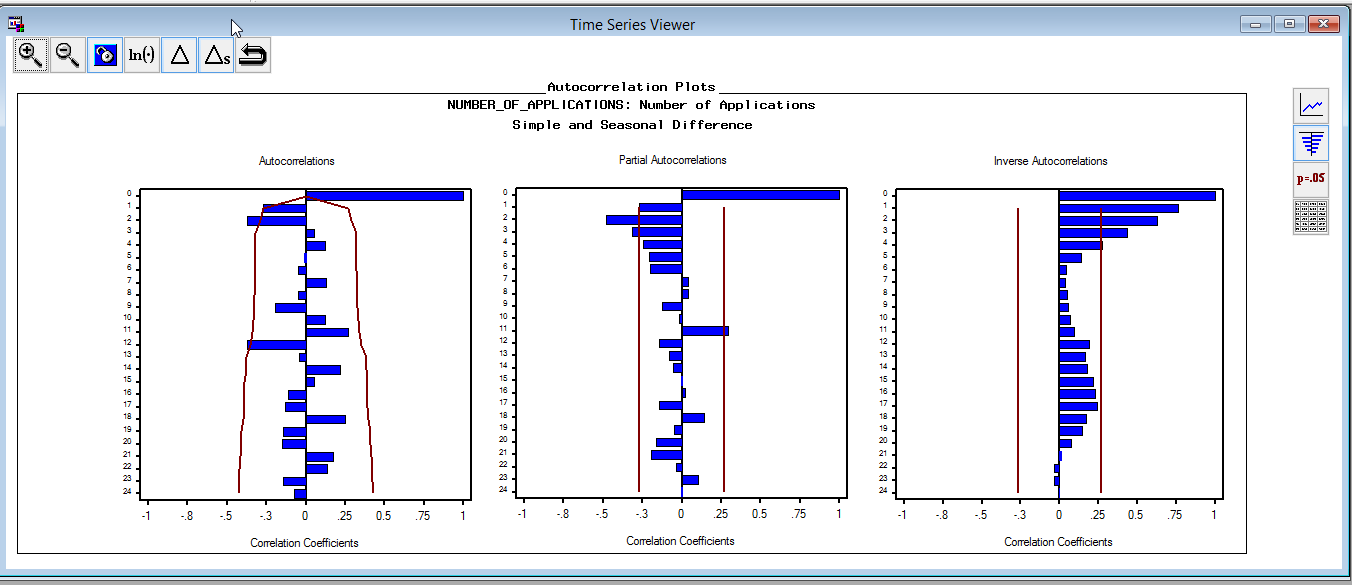
After seasonal difference: Seasonal difference improved white noise, but the unit root test and seasonality tests were not as strong.

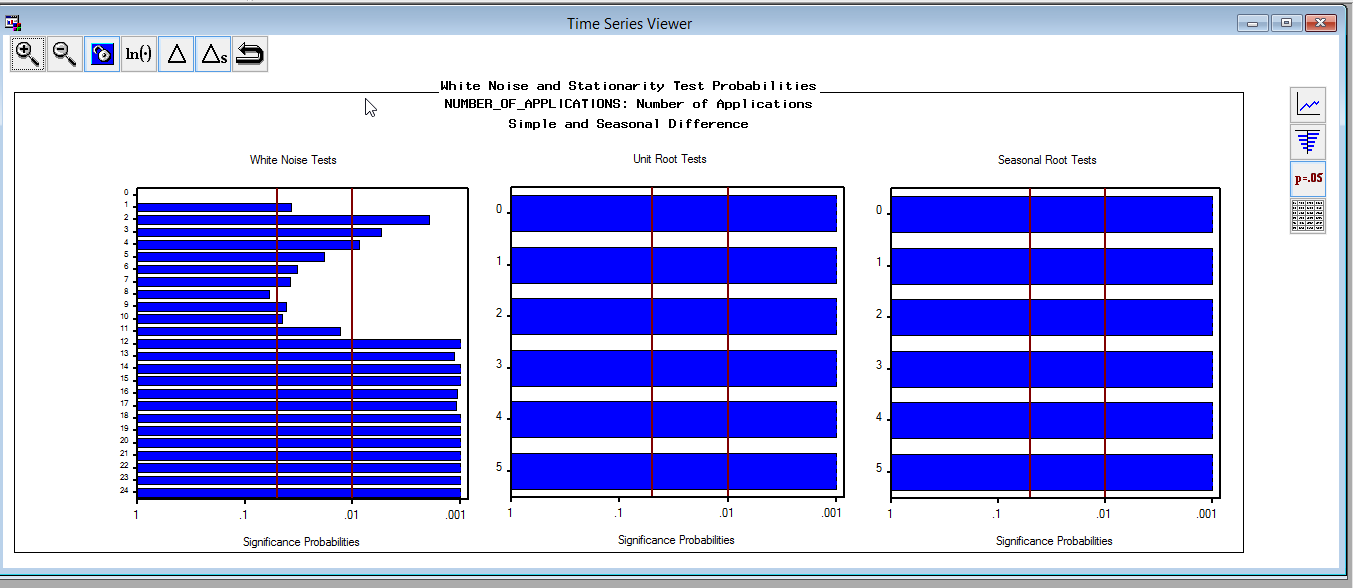




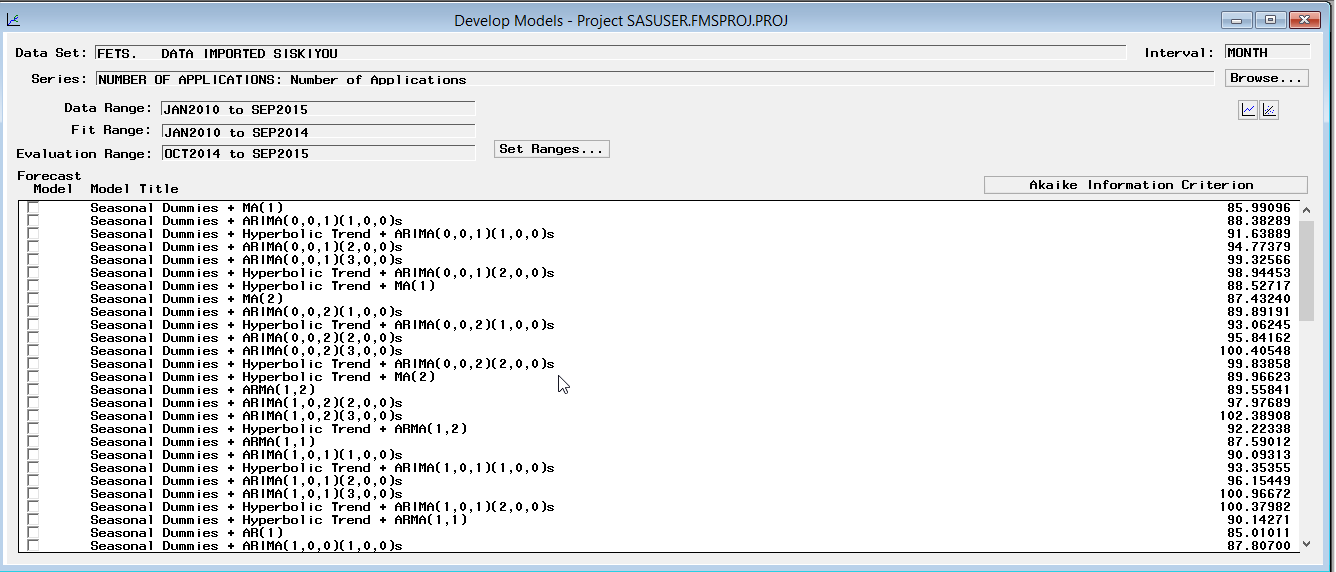


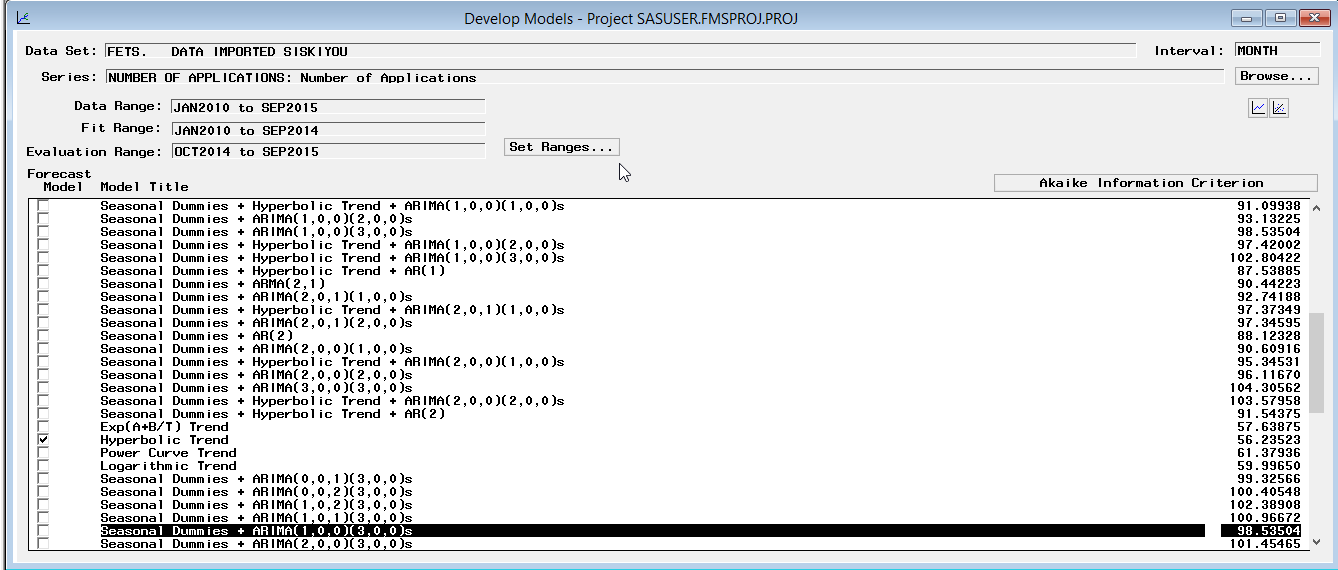
After seasonal difference and first difference:  






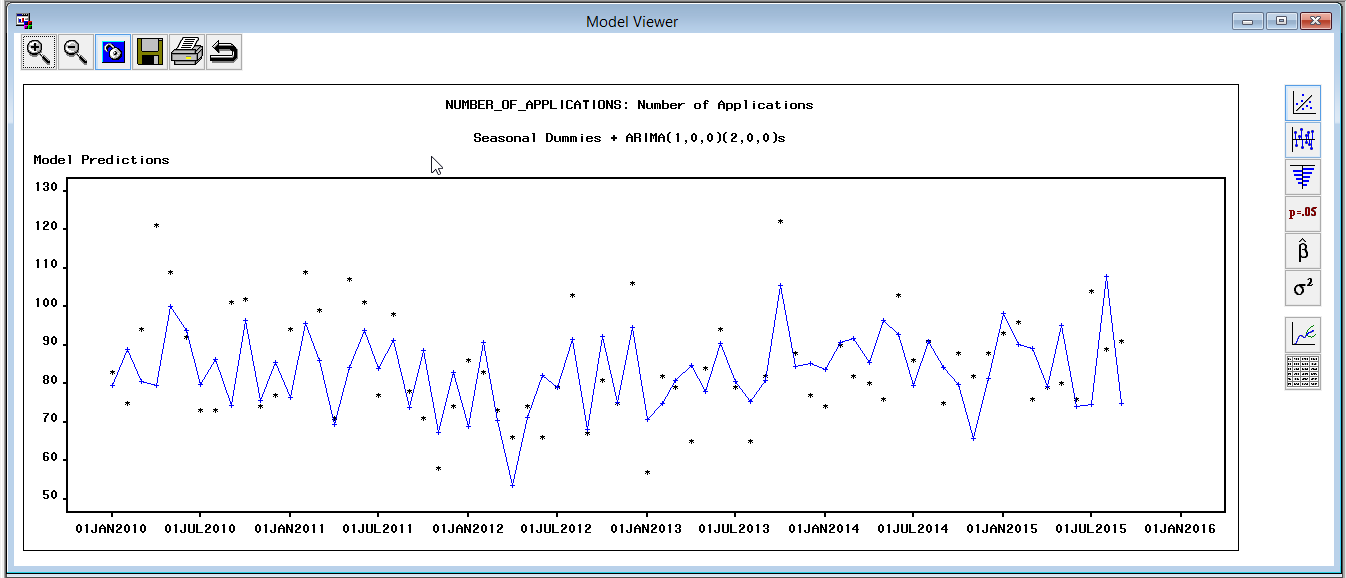
Combining seasonal difference and first difference actually gives us a worse result than only pick one of them.

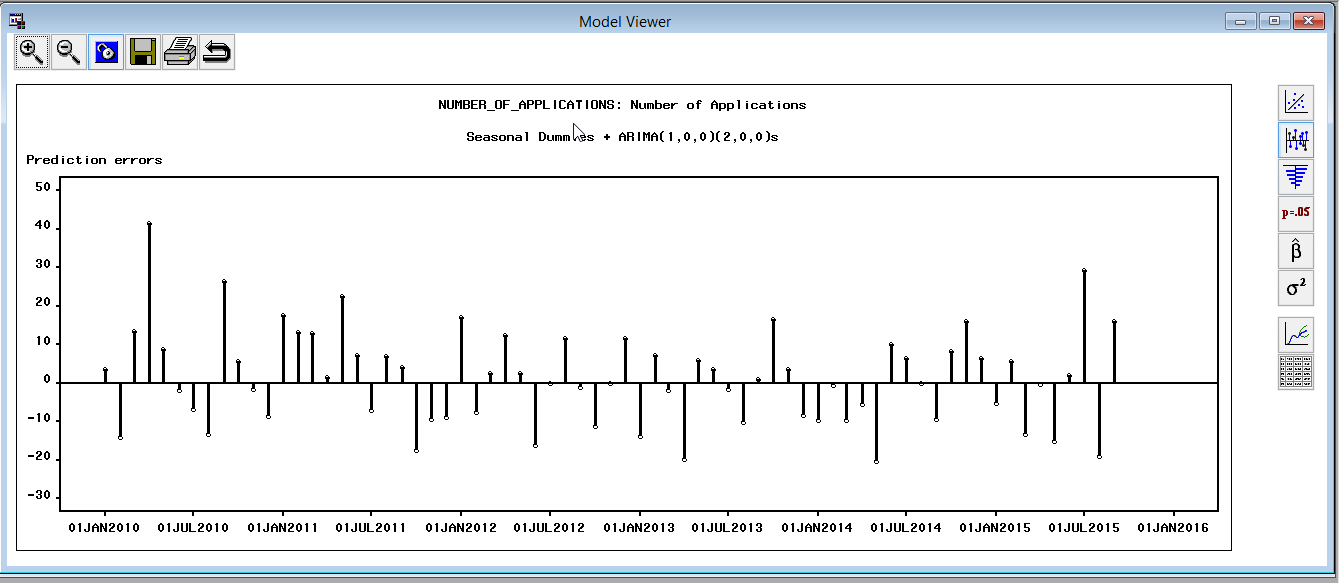


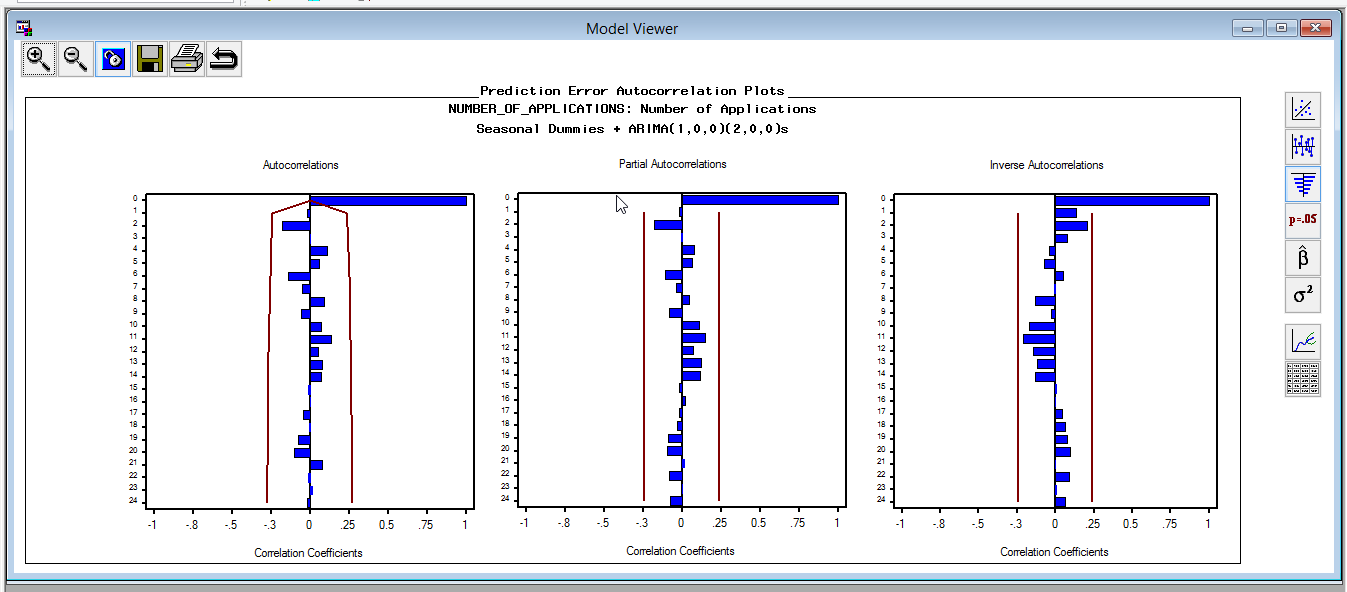


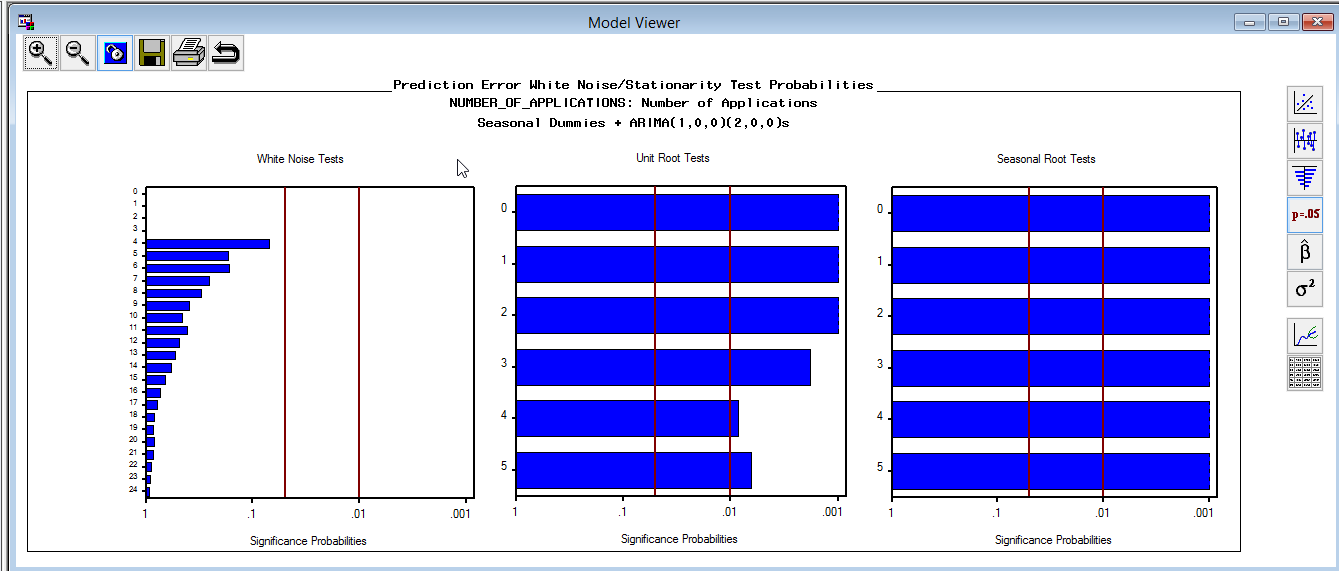
The result is better with only seasonal dummies but not trend.

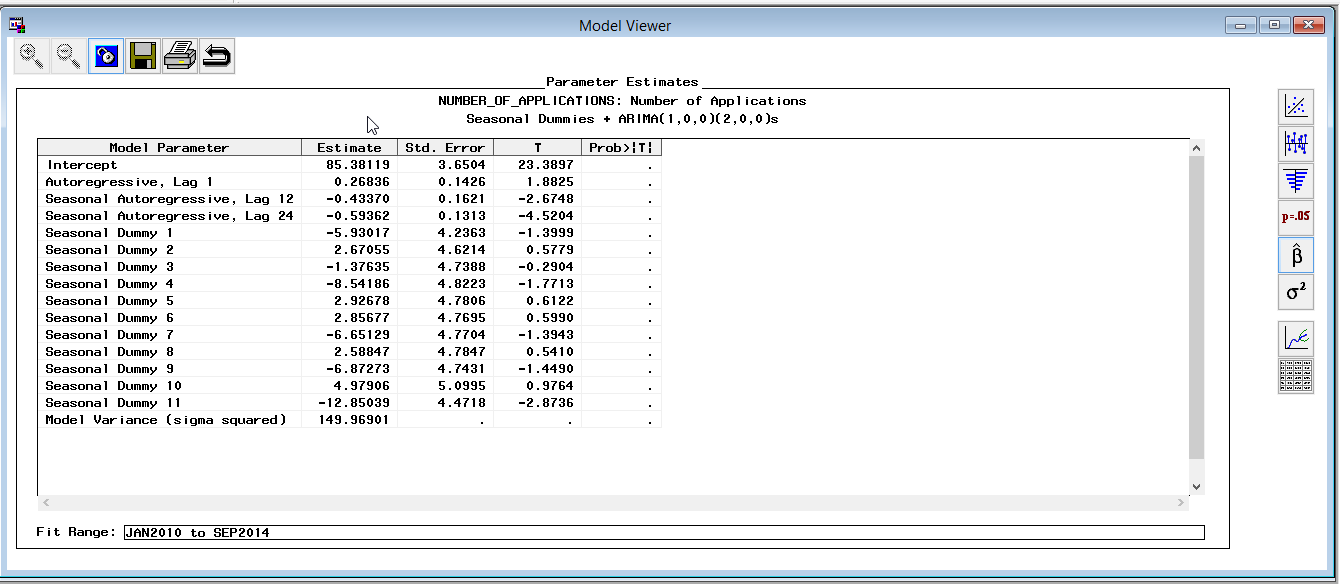
**Seasonal dummies+ ARIMA(1,0,0,)(2,,0,0)**

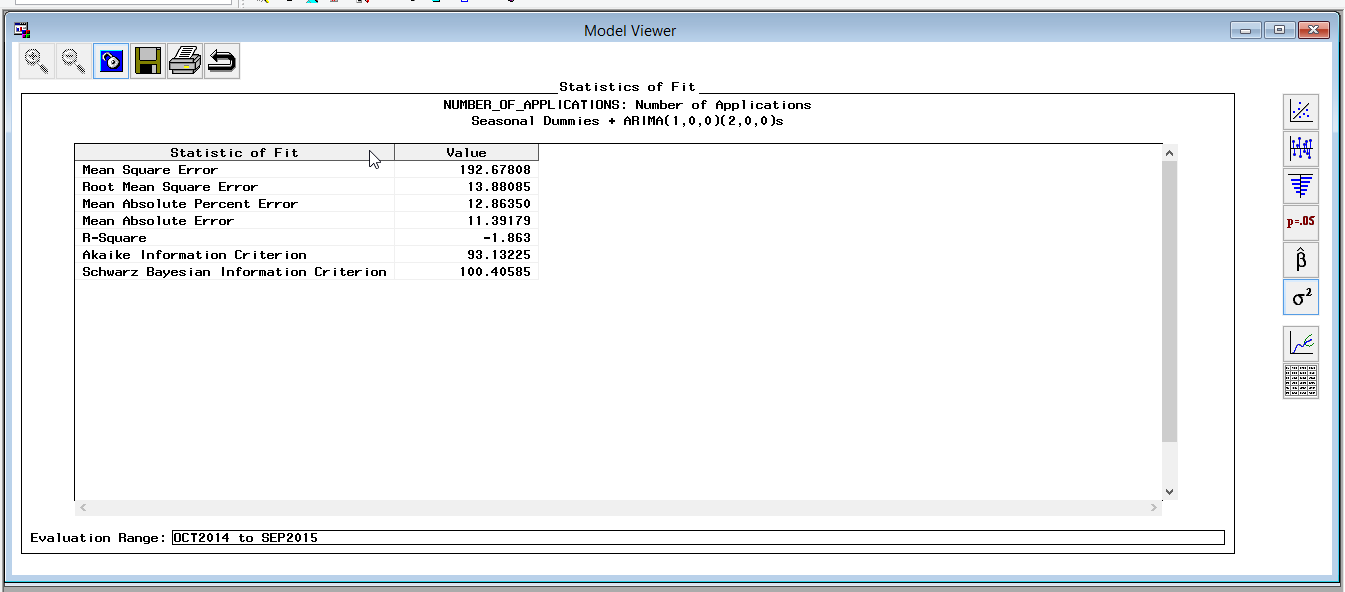


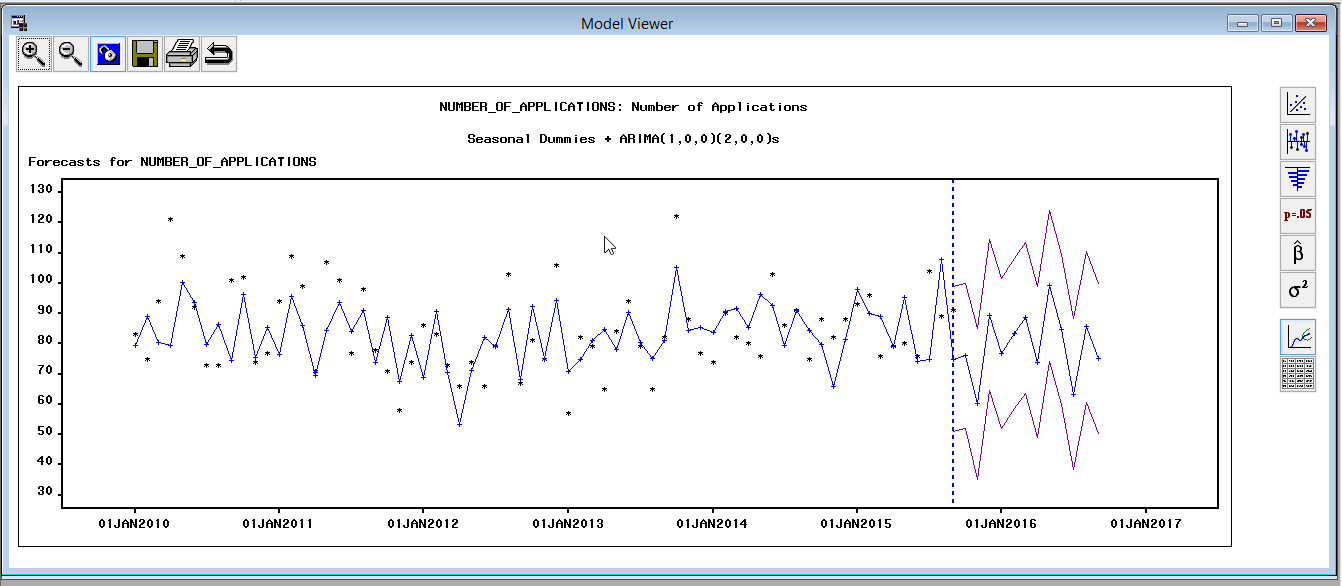


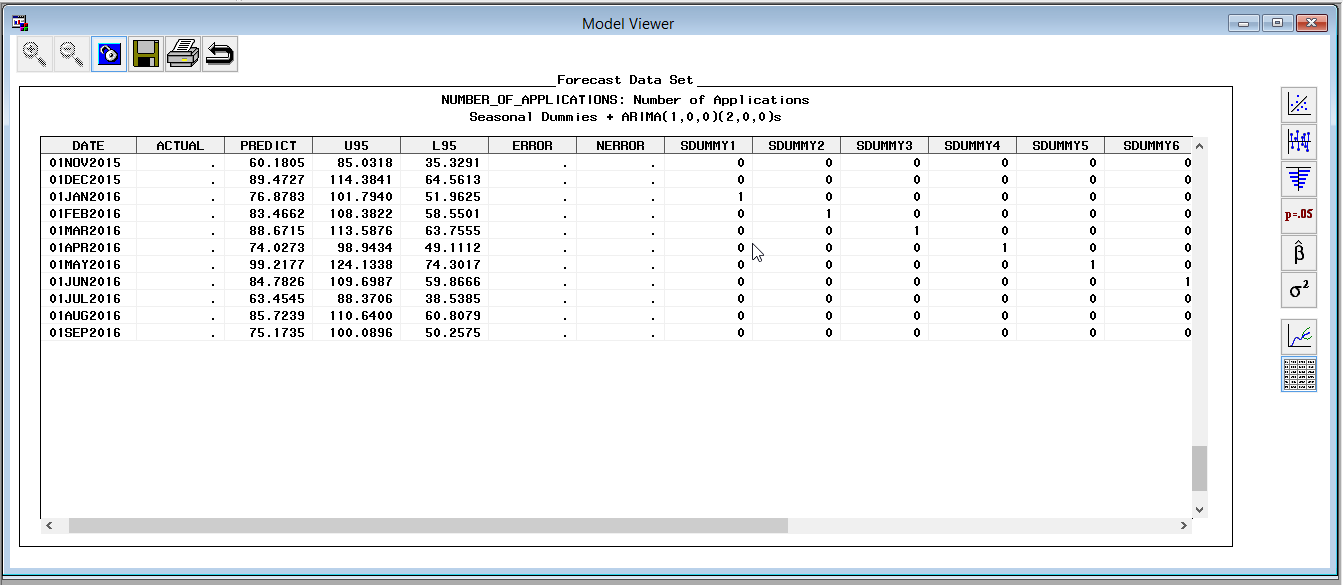




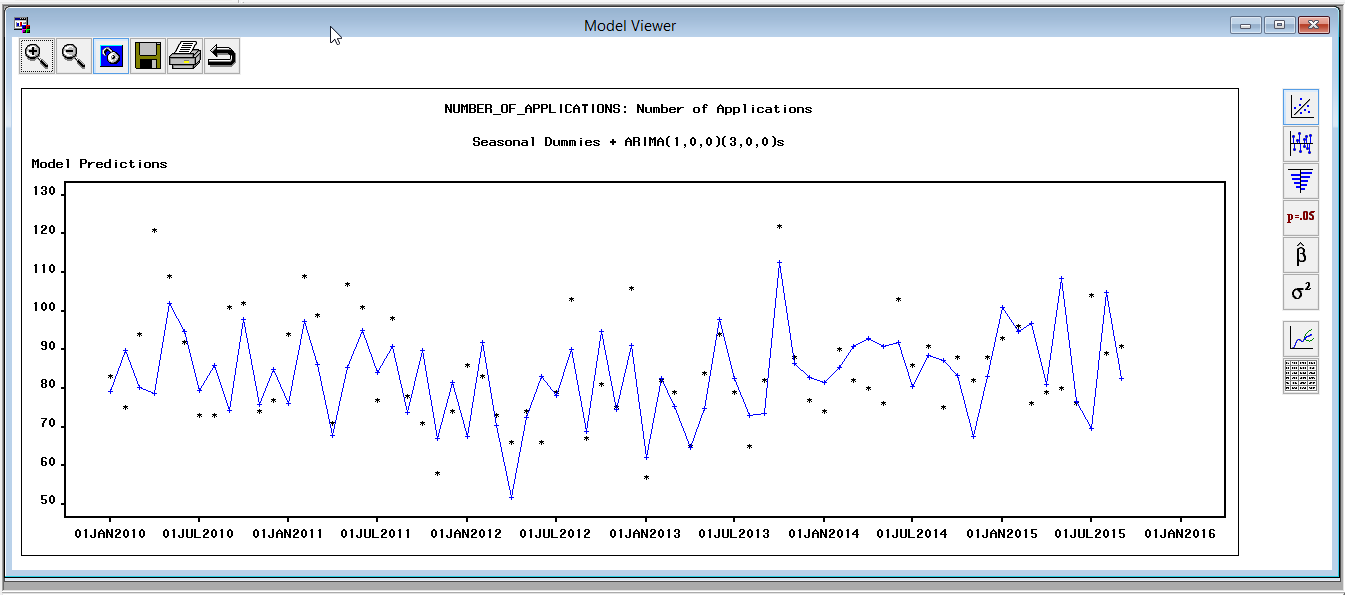


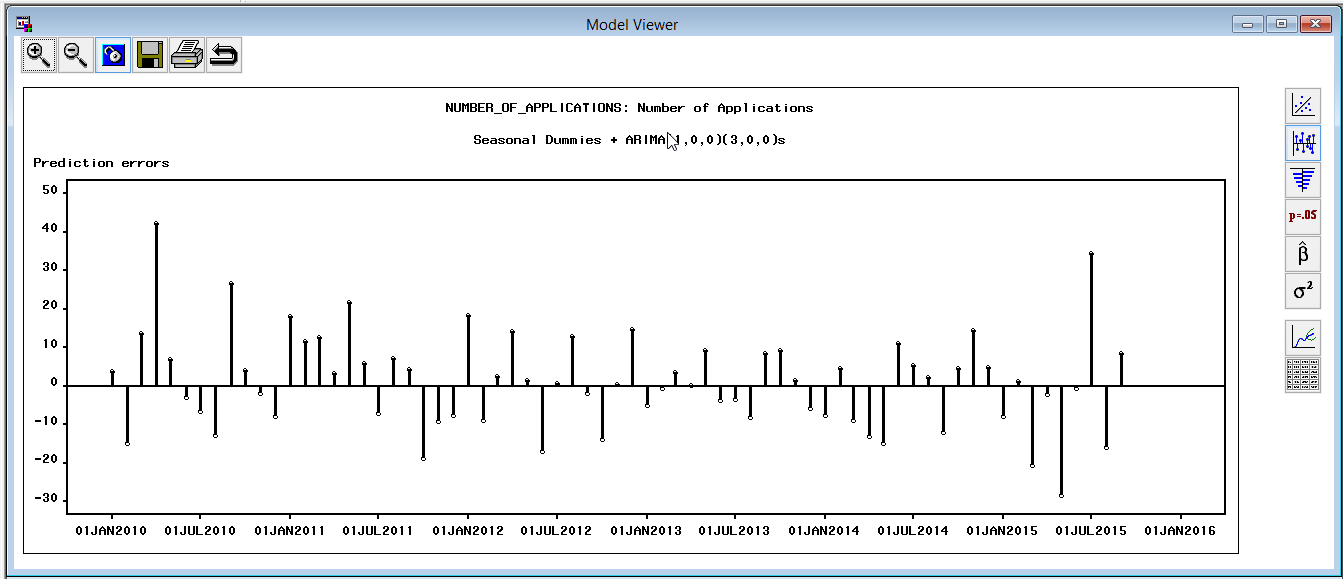


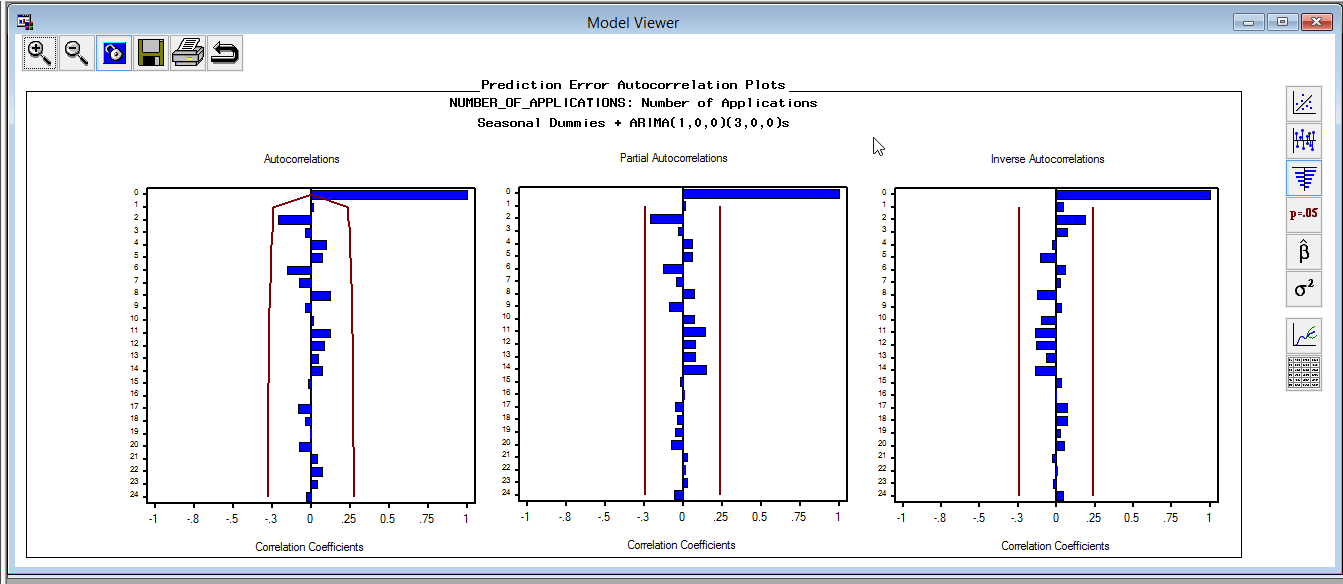


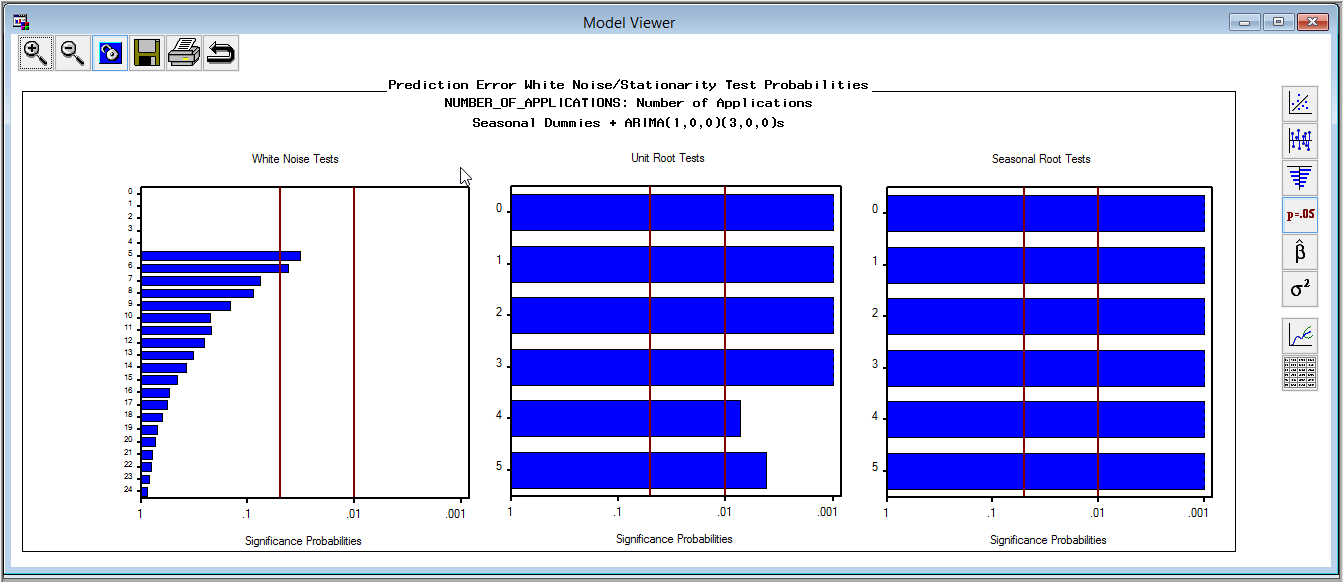


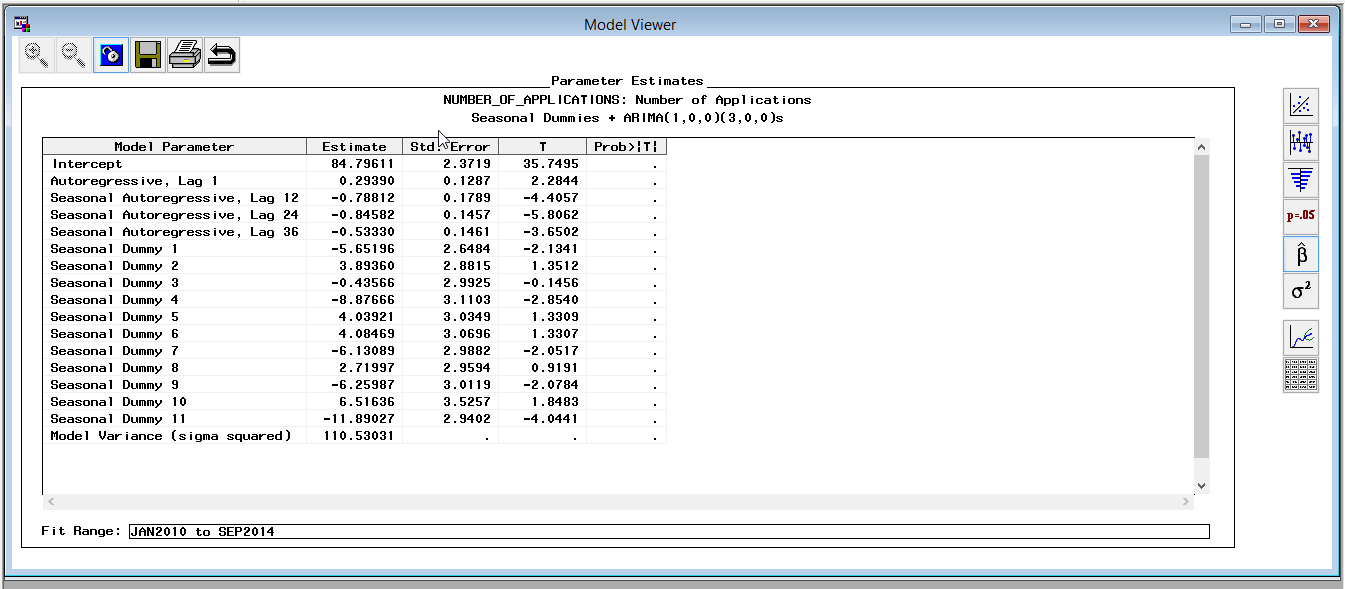
**Seasonal dummies+ ARIMA(1,0,0,)(3,,0,0)**

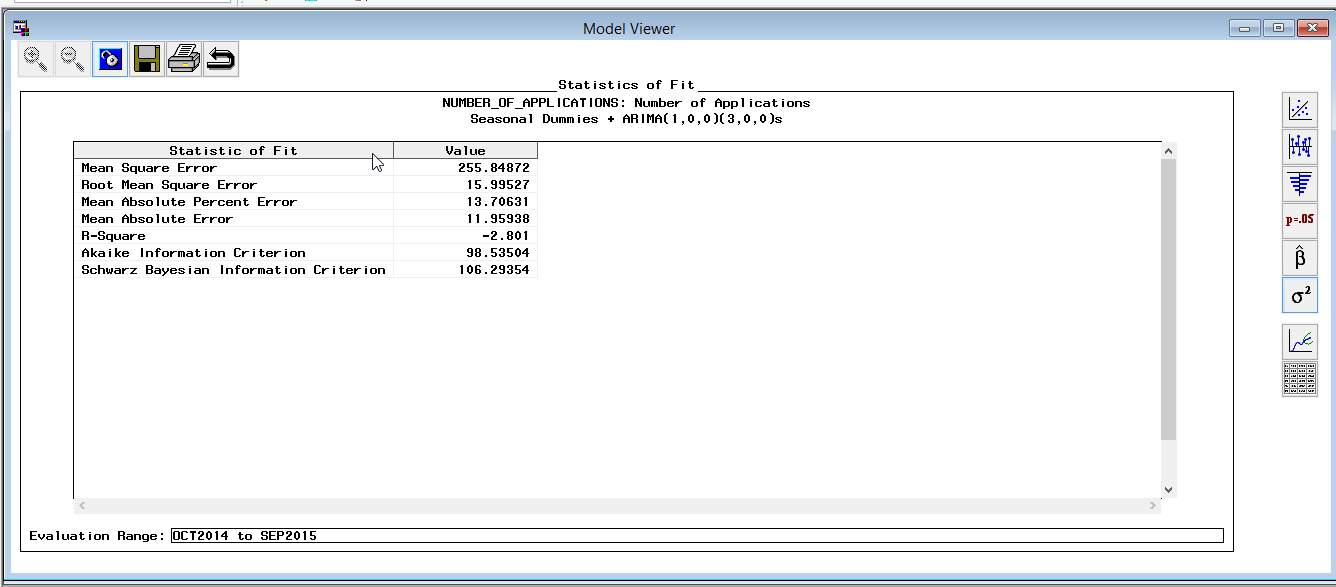


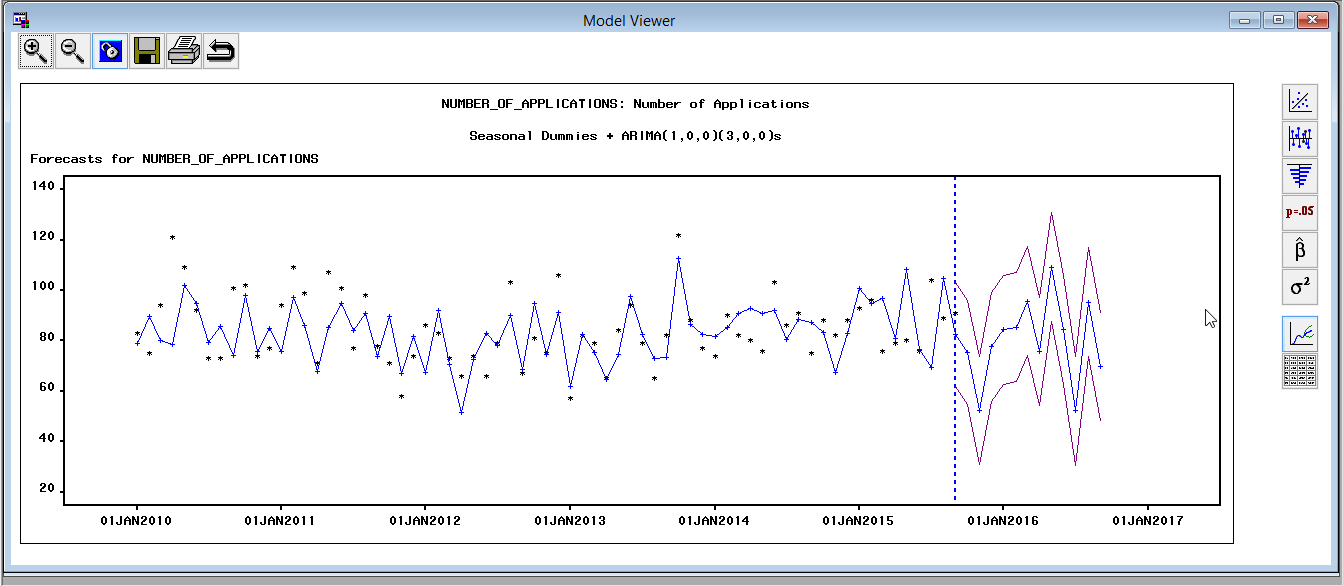


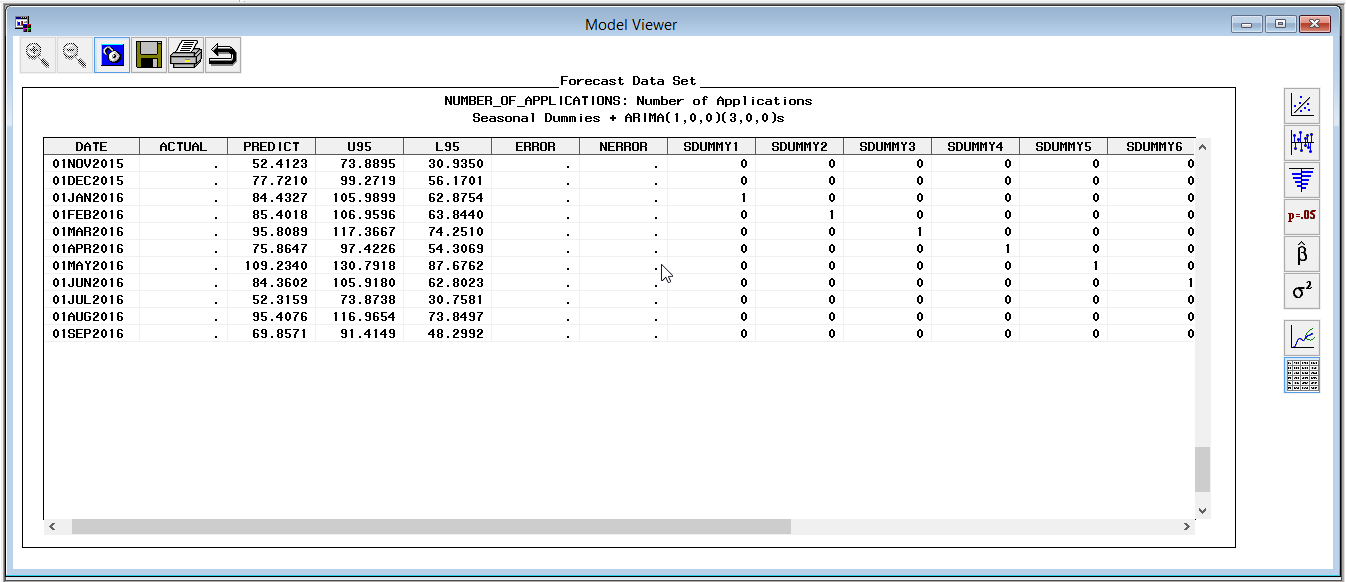


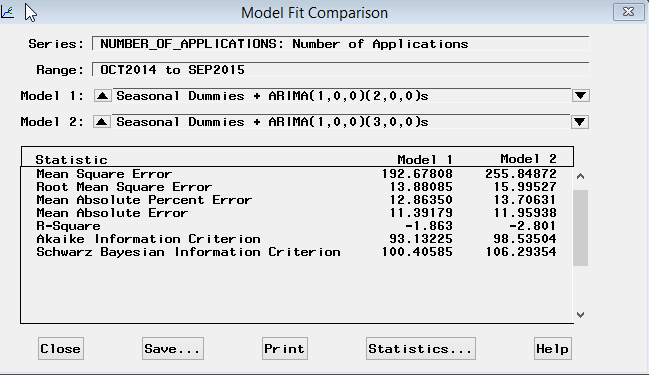








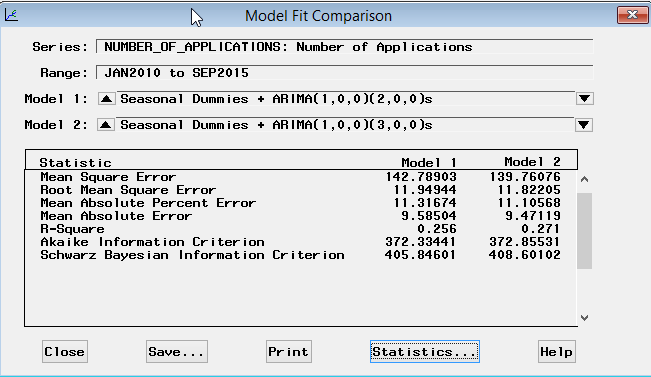


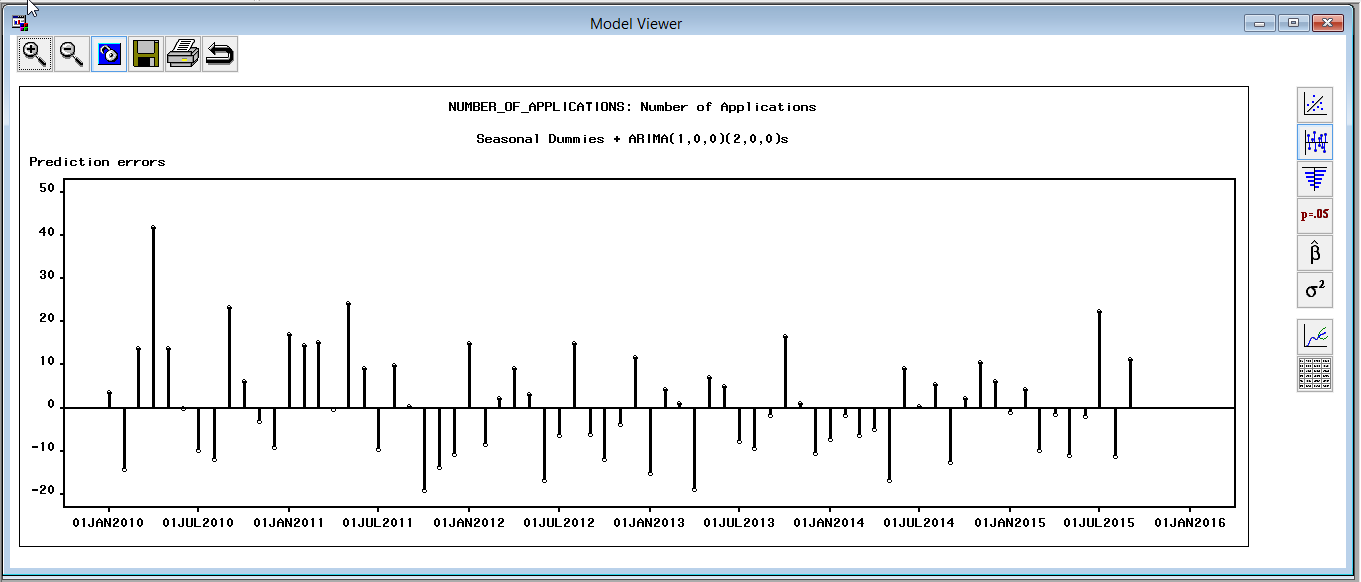


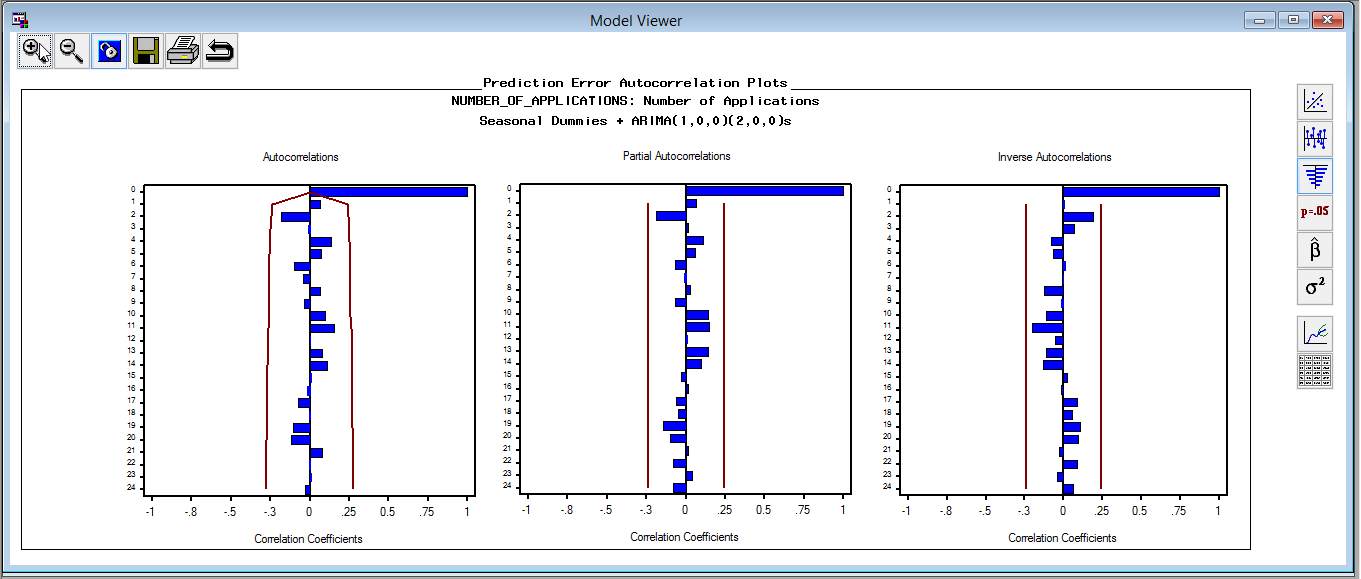
Compare Seasonal dummies+ ARIMA(1,0,0,)(2,,0,0) and Seasonal dummies+ ARIMA(1,0,0,)(3,,0,0),

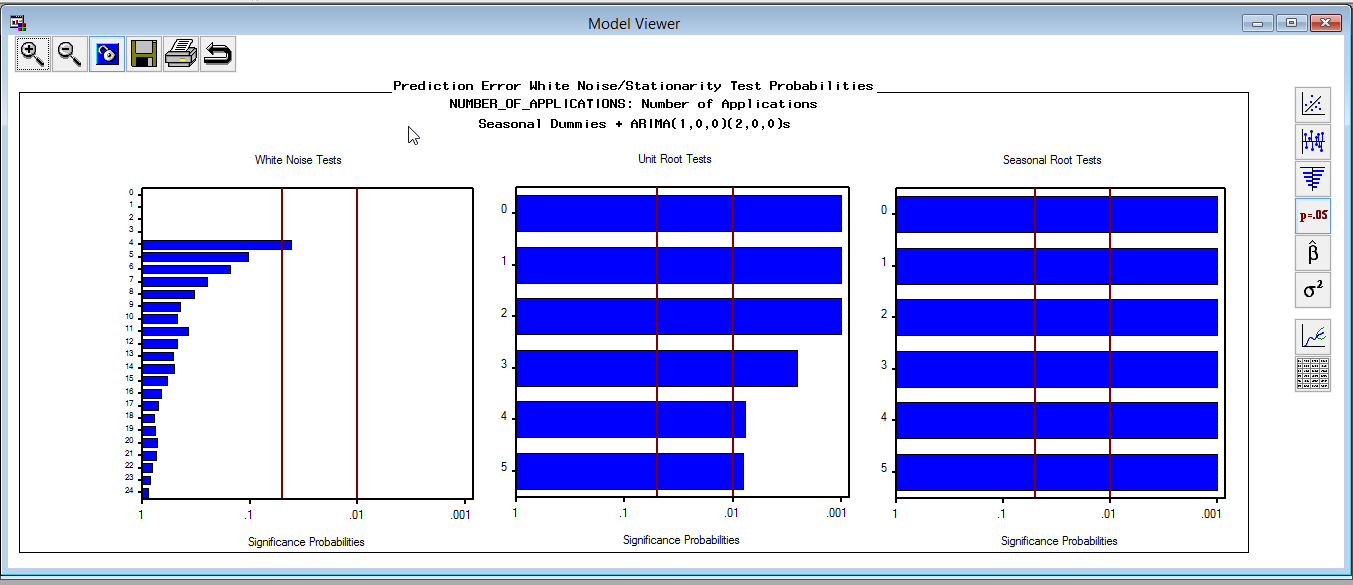
Although Seasonal dummies+ ARIMA(1,0,0,)(3,,0,0) gives a slightly more precise forecast, Seasonal dummies+ ARIMA(1,0,0,)(2,,0,0) has better SBIC, RMSE and AIC, also its white noise tests results are better. For ACF, PACF, IACF, two models are more or less the same. Considering there is huge variance from each date point to another, we will take Seasonal dummies+ ARIMA (1,0,0,)(2,,0,0) as our final model.

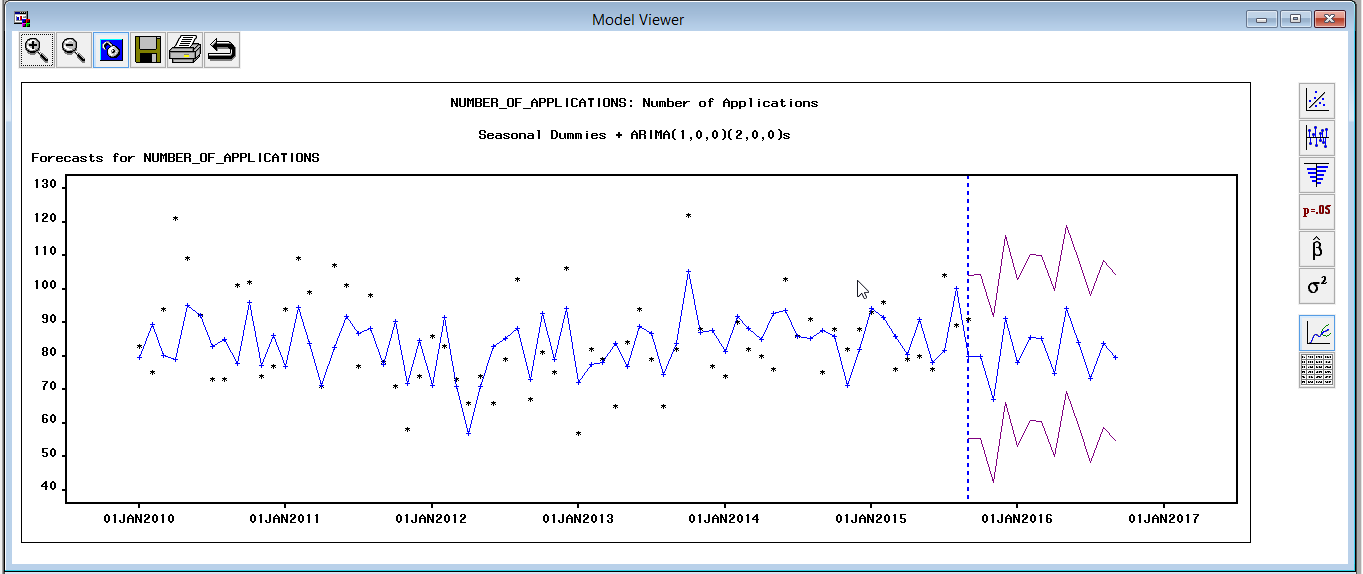
After refit with whole time span:

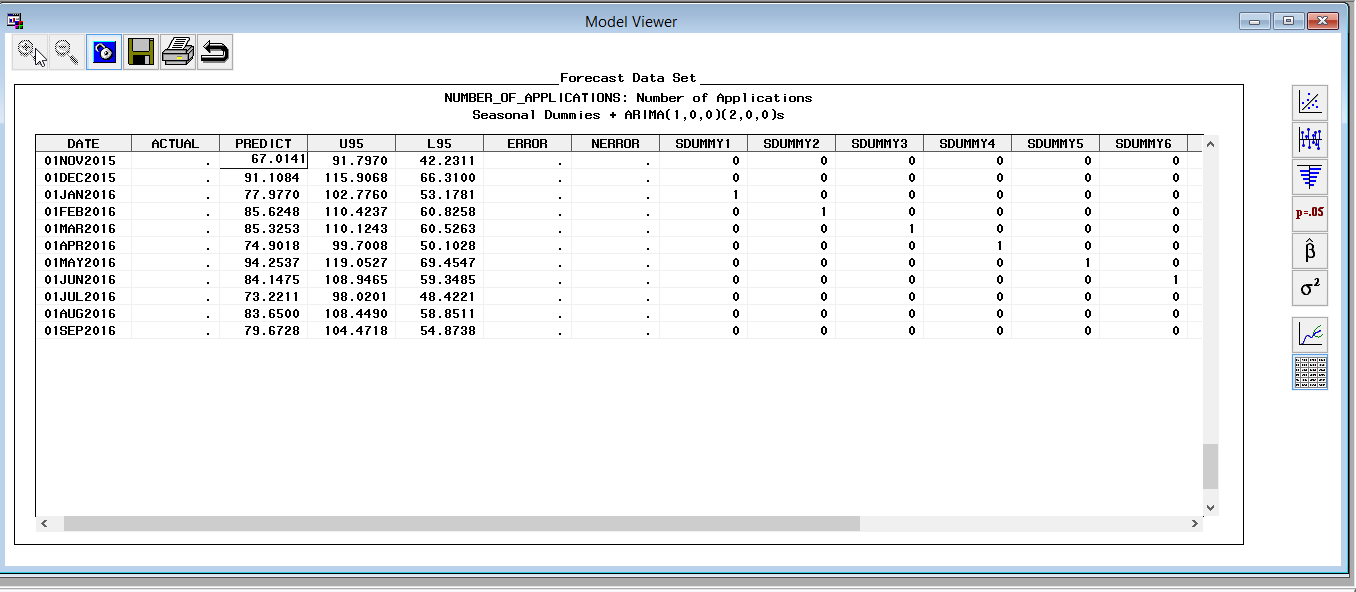
 RMSE becomes better, but huge variances occur on AIC and SBIC.





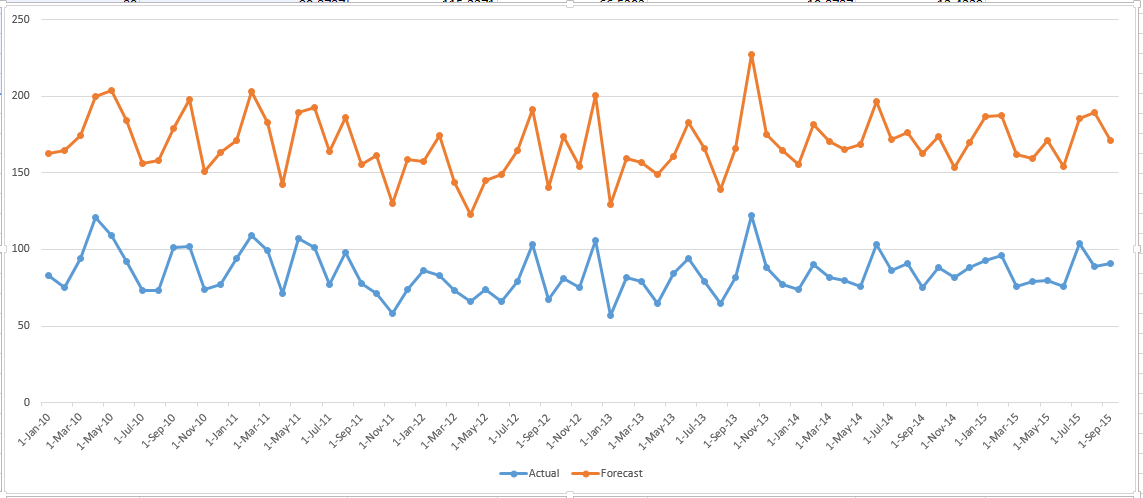






This model performs well. The white noise test is reduced in the randomness after the final model is introduced, while the seasonality and unit root test remain strong. The forecast also looks good and performs well.

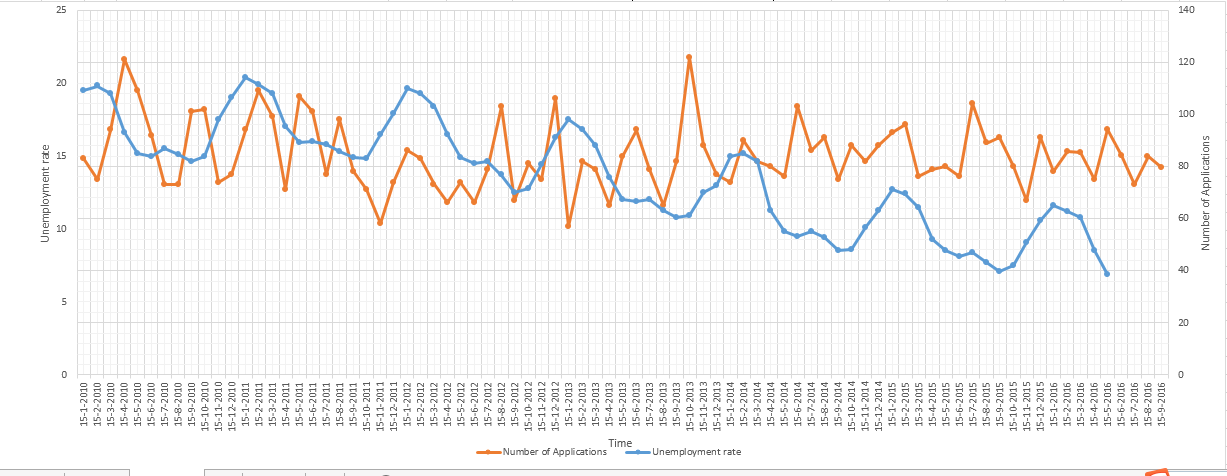
**Predicted vs Actual**

****

Just like the other four models, the forecast trend remains high but mirrors the actual trend.

**Comparison with trend of unemployment rate.**

**Forecast range 01NOV2015-01SEP2016**



We can see there is some correlation between trend of unemployment rate and number of applications.

Business Insight

**Budget allocation :**

Forecasting the application count would help in allocating budget for particular counties.

**Child leaving there family:**

The trend line of forecast can give us information about children leaving who can possibly leave there family, as major portion of the application request is from zero-parent family.

**Umployment trend :**

Knowing the trend of unemployment would help the government in creating more employment opportunities.

**Education and Health Benefits:** Depending on the age of the candidates who have applied, necessary education or health benefits must be given accordingly.

References :

* <http://www.cdss.ca.gov/calworks/>
* <http://www.latimes.com/business/la-fi-california-world-economy-20150702-story.html>
* <http://www.nationalmemo.com/the-5-best-ideas-from-californias-progressive-resurgence/>