**PROJECT 3: EXECUTIVE REPORT**

1. **PROJECT DESCRIPTION**

This project is to analyze data3 and chose the model for this dataset, and make predictions for the future 13 points based on chosen model. Fig 3-1 shows the plot of this time series.

 

Figure 3-1 PROJ 2 Time Series Figure 3-2 PROJ 2 Forecast

1. **CHOSEN MODEL**

For this dataset the final chosen model is **ARIMA(3,1,1),** which means the data has a trend, and observation on the current time depends on observations of 3 time units before as well as the noise of previous time unit and the current time unit.

1. **PARAMETER ESTIMATION**

,

1. **MODEL EVALUATION**

The model is chosen and evaluated in many ways including checking stationary, correlogram, periodogram, MSE, Parsimony, residual diagnostic etc. See Appendix. For the finalized model, Mean Squared Error is 2269.789; Akaike Information Criterion is 6170.79.

1. **FORECAST**

The forecast result can be seen from Figure 3-2 above. We can see the prediction could match with the data pattern well. As the prediction time goes further away from the original data, the prediction interval would be larger.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Time | 585 | 586 | 587 | 588 | 589 | 590 | 591 | 592 | 593 | 594 | 595 | 596 | 597 |
| Pred. Lowerbound | -10505 | -10676 | -10999 | -11258 | -11617 | -11865 | -12173 | -12344 | -12565 | -12655 | -12808 | -12848 | -12969 |
| Prediction | -10411 | -10417 | -10503 | -10483 | -10542 | -10500 | -10543 | -10493 | -10537 | -10494 | -10548 | -10519 | -10587 |
| Pred.Upperbound | -10318 | -10159 | -10007 | -9708 | -9467 | -9134 | -8914 | -8642 | -8508 | -8332 | -8289 | -8189 | -8205 |

Table 3-1 PROJ 3 Forecast Result

**PROJECT 3: TECHNICAL APPENDIX**

1. **Data Plot and Exploration**

Firstly, plot the data, and it looks have a downward trend and not stationary. After first differentiation, there is process turns to be stationary. Augmented Dickey–Fuller test shows that the process is stationary after differentiation. For the differentiated data, plot the sample ACF, pACF, and periodogram separately and see the pattern.

* 1. **Sample Autocorrelation and Partial Autocorrelation**

Obviously, the original process acf drops down slowly indicating non-stationary. From the differenced sample ACF and pACF plot, we can see that the ACF has some sinusoidal decay pattern; pACF cuts off at around lag 8 or also have some decay exponentially. So the models could be some ARIMA model, like ARIMA(2,1,X), ARMA(3,1,X), or ARIMA (8,1,0), ARIMA (10,1,0) etc.





Figure 3-3 Proj 3 Sample ACF and pACF before and after differentiation

* 1. **Periodogram**



Figure 3-4 Proj 3 Periodogram before and after differenciation

We can see there is a peak between frequency (0-0.1), which imply an AR(2), and a trough around (0.3,0.4), which means a MA(2) or just some transition; also the beginning and end point maybe some indication of 1 or 2 MA or AR elements. So we based on this plot, we can choose some models like ARIMA(2,1,1), ARIMA(3,1,1), ARIMA(2,1,2), ARIMA(3,1,2), ARIMA(2,1,0)

1. **Choose Candidate Models**

After exploration on the data above, I initially chose the ARIMA(2,1,0), ARIMA(3,1,0), ARIMA(5,1,0), ARIMA(8,1,0), ARIMA(10,1,0), ARIMA(2,1,1), ARIMA(3,1,1), ARIMA(2,1,2), ARIMA(3,1,2) to try to fit the data.

1. **Candidates models Comparison**

First, data3 is split to training data (90%) and test data (10%). Use the training data to fit various candidate models, and use test data for prediction SSE test.

* 1. **Coefficients significance test**

In the coefficients significance test, coefficients in ARIMA(10,1,0), ARIMA(3,1,2) models are not significant. So these two models will not be considered and tested later.

* 1. **AIC**

AIC values are compared in these candidate models. ARIMA(5,1,0), ARIMA(8,1,0), ARIMA(3,1,1) has smaller AIC values compared with other models. See Table3-2

* 1. **MSE**

Calculating the MSE, ARIMA(5,1,0), ARIMA(8,1,0), ARIMA(3,1,1) have larger MSE compared with others. See Table3-2.

* 1. **Test data Prediction SSE**

Use the models to make prediction of test data. Calculate the summation of squared error. ARIMA(2,1,1), ARIMA(3,1,1), ARIMA(3,1,2) have better performance in this criteria. See Table3-2.

* 1. **Residual Diagnostics**

After running the residual diagnostics, ARIMA(8,1,0), ARIMA(3,1,1) have random residuals, while other models residuals show some patterns or fail the randomness test. Below is the residual diagnostic of the model ARIMA(3,1,1).

 

Figure 3-5 Proj 3 Residual Diagnostics

* 1. **Periodogram match**

ARIMA(5,1,0), ARIMA(8,1,0), ARIMA(3,1,1) have good fit of spectrum with original training data periodogram. See ARIMA(3,1,1) fit in Figure 3-6 below.

 

Figure 3-6 Proj 3 Sample and Theoretical Graphs Match Result

* 1. **Sample ACF and pACF match**

The ARIMA(5,1,0), ARIMA(8,1,0), ARIMA(3,1,1) models acf, pacf have good match with original sample acf and pacf. See ARIMA(3,1,3) fit in figure 3-6 above.

* 1. **Parsimony**

Out of parsimony concern, the higher order AR processes are not preferred.

* 1. **Model Comparison Summary**

Summarize all the criterions above. Green cell means that the model performs good in the particular criterion.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **COMPARE Models** | **ARIMA(2,1,0)** | **ARMA(3,1,0)** | **ARIMA(5,1,0)** | **ARIMA(8,1,0)** | **ARMA(2,1,1)** | **ARMA(3,1,1)** | **ARMA(2,1,2)** |
| Coef Signif Test | Y | Y | Y | Y | Y | Y | Y |
| AIC | 5894.14 | 5676.26 | 5551.9 | 5508 | 5802.65 | 5476.92 | 5787 |
| MSE | 4943.859 | 3241.017 | 2544.902 | 2334.824 | 4141.066 | 2194.403 | 4012.452 |
| Test prediction SSE | 149301363 | 247473822 | 263065901 | 318838335 | 206697931 | 291306905 | 98310842 |
| Residual Diagnostics | N | N | N | Y | N | Y | N |
| Periodogram match | N | N | Y | Y | N | Y | N |
| Correlogram match | N | N | Y | Y | N | Y | N |
| Parsinomy | Y | Y | Y | N | Y | Y | Y |

Table 3-2 Model Candidates Comparison

1. **Finalize Model and Forecast**

**a.** Overall, ARIMA(3,1,1) performs well in all the criterions in Table3-2. Choose as the final model and fit to all the original data and make 13 points forecast as shown in the report.

,

**b.** Calculate the MSE, AIC and residual diagnostics of the final model. All results are as well as training model above.