Kilger

MKT6971

Exercise #3

Name: Collin Real (yhi267)

Here is the third and final exercise. It lists the unemployment rate in the US from January 1948 to March 2020. Here is the plot:



The unit root tests suggest a non-constant mean so here is the plot of the first differenced data:



Next step was to run some ARIMA models and compare them. This led to the following ARIMA runs:

Model 1

Model 1: ARMA, using observations 1948:02-2020:03 (T = 866)

Dependent variable: d\_UNRATE

Standard errors based on Hessian

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Coefficient* | *Std. Error* | *z* | *p-value* |  |
| const | 0.00287270 | 0.0147910 | 0.1942 | 0.8460 |  |
| phi\_1 | 0.870665 | 0.0296668 | 29.35 | <0.0001 | \*\*\* |
| theta\_1 | −0.718031 | 0.0379465 | −18.92 | <0.0001 | \*\*\* |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mean dependent var | 0.001155 |  | S.D. dependent var | 0.209924 |
| Mean of innovations | −0.000378 |  | S.D. of innovations | 0.200521 |
| R-squared | 0.086522 |  | Adjusted R-squared | 0.085465 |
| Log-likelihood | 162.6270 |  | Akaike criterion | −317.2540 |
| Schwarz criterion | −298.1985 |  | Hannan-Quinn | −309.9612 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | *Real* | *Imaginary* | *Modulus* | *Frequency* |
| AR |  |  |  |  |  |
|  | Root 1 | 1.1485 | 0.0000 | 1.1485 | 0.0000 |
| MA |  |  |  |  |  |
|  | Root 1 | 1.3927 | 0.0000 | 1.3927 | 0.0000 |

Test for autocorrelation up to order 12

Ljung-Box Q' = 75.3636,

with p-value = P(Chi-square(10) > 75.3636) = 4.042e-012

Model 2

Model 2: ARMA, using observations 1948:02-2020:03 (T = 866)

Dependent variable: d\_UNRATE

Standard errors based on Hessian

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Coefficient* | *Std. Error* | *z* | *p-value* |  |
| const | 0.00298555 | 0.0148977 | 0.2004 | 0.8412 |  |
| phi\_1 | 0.555245 | 0.0625183 | 8.881 | <0.0001 | \*\*\* |
| phi\_2 | 0.238727 | 0.0373804 | 6.386 | <0.0001 | \*\*\* |
| theta\_1 | −0.538385 | 0.0583563 | −9.226 | <0.0001 | \*\*\* |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mean dependent var | 0.001155 |  | S.D. dependent var | 0.209924 |
| Mean of innovations | −0.000420 |  | S.D. of innovations | 0.196462 |
| R-squared | 0.123133 |  | Adjusted R-squared | 0.121101 |
| Log-likelihood | 180.2785 |  | Akaike criterion | −350.5570 |
| Schwarz criterion | −326.7375 |  | Hannan-Quinn | −341.4410 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | *Real* | *Imaginary* | *Modulus* | *Frequency* |
| AR |  |  |  |  |  |
|  | Root 1 | 1.1911 | 0.0000 | 1.1911 | 0.0000 |
|  | Root 2 | -3.5169 | 0.0000 | 3.5169 | 0.5000 |
| MA |  |  |  |  |  |
|  | Root 1 | 1.8574 | 0.0000 | 1.8574 | 0.0000 |

Test for autocorrelation up to order 12

Ljung-Box Q' = 36.8101,

with p-value = P(Chi-square(9) > 36.8101) = 2.845e-005

Model 3

Model 3: ARMA, using observations 1948:02-2020:03 (T = 866)

Dependent variable: d\_UNRATE

Standard errors based on Hessian

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Coefficient* | *Std. Error* | *z* | *p-value* |  |
| const | 0.00257941 | 0.0115202 | 0.2239 | 0.8228 |  |
| phi\_1 | 1.65561 | 0.0374836 | 44.17 | <0.0001 | \*\*\* |
| phi\_2 | −0.782771 | 0.0433592 | −18.05 | <0.0001 | \*\*\* |
| theta\_1 | −1.64177 | 0.0383751 | −42.78 | <0.0001 | \*\*\* |
| theta\_2 | 0.863215 | 0.0479172 | 18.01 | <0.0001 | \*\*\* |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mean dependent var | 0.001155 |  | S.D. dependent var | 0.209924 |
| Mean of innovations | −0.000443 |  | S.D. of innovations | 0.194870 |
| R-squared | 0.137289 |  | Adjusted R-squared | 0.134286 |
| Log-likelihood | 187.0535 |  | Akaike criterion | −362.1069 |
| Schwarz criterion | −333.5236 |  | Hannan-Quinn | −351.1678 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | *Real* | *Imaginary* | *Modulus* | *Frequency* |
| AR |  |  |  |  |  |
|  | Root 1 | 1.0575 | -0.3989 | 1.1303 | -0.0574 |
|  | Root 2 | 1.0575 | 0.3989 | 1.1303 | 0.0574 |
| MA |  |  |  |  |  |
|  | Root 1 | 0.9510 | -0.5041 | 1.0763 | -0.0776 |
|  | Root 2 | 0.9510 | 0.5041 | 1.0763 | 0.0776 |

Test for autocorrelation up to order 12

Ljung-Box Q' = 39.2977,

with p-value = P(Chi-square(8) > 39.2977) = 4.328e-006

Model 4

Model 15: ARMA, using observations 1948:02-2020:03 (T = 866)

Dependent variable: d\_UNRATE

Standard errors based on Hessian

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Coefficient* | *Std. Error* | *z* | *p-value* |  |
| const | 0.00250730 | 0.0113898 | 0.2201 | 0.8258 |  |
| phi\_1 | 0.578072 | 0.0624914 | 9.250 | <0.0001 | \*\*\* |
| phi\_2 | 0.117027 | 0.0739480 | 1.583 | 0.1135 |  |
| phi\_3 | 0.611279 | 0.108845 | 5.616 | <0.0001 | \*\*\* |
| phi\_4 | −0.695650 | 0.0557809 | −12.47 | <0.0001 | \*\*\* |
| theta\_1 | −0.585967 | 0.0671052 | −8.732 | <0.0001 | \*\*\* |
| theta\_2 | 0.0631790 | 0.0740003 | 0.8538 | 0.3932 |  |
| theta\_3 | −0.595233 | 0.107839 | −5.520 | <0.0001 | \*\*\* |
| theta\_4 | 0.766918 | 0.0693611 | 11.06 | <0.0001 | \*\*\* |
| theta\_5 | 0.0305044 | 0.0709625 | 0.4299 | 0.6673 |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mean dependent var | 0.001155 |  | S.D. dependent var | 0.209924 |
| Mean of innovations | −0.000422 |  | S.D. of innovations | 0.192210 |
| R-squared | 0.160680 |  | Adjusted R-squared | 0.152845 |
| Log-likelihood | 198.7941 |  | Akaike criterion | −375.5881 |
| Schwarz criterion | −323.1854 |  | Hannan-Quinn | −355.5330 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | *Real* | *Imaginary* | *Modulus* | *Frequency* |
| AR |  |  |  |  |  |
|  | Root 1 | 1.0508 | 0.4052 | 1.1262 | 0.0586 |
|  | Root 2 | 1.0508 | -0.4052 | 1.1262 | -0.0586 |
|  | Root 3 | -0.6114 | -0.8715 | 1.0646 | -0.3474 |
|  | Root 4 | -0.6114 | 0.8715 | 1.0646 | 0.3474 |
| MA |  |  |  |  |  |
|  | Root 1 | 0.9450 | 0.5028 | 1.0704 | 0.0778 |
|  | Root 2 | 0.9450 | -0.5028 | 1.0704 | -0.0778 |
|  | Root 3 | -0.5661 | -0.8856 | 1.0511 | -0.3405 |
|  | Root 4 | -0.5661 | 0.8856 | 1.0511 | 0.3405 |
|  | Root 5 | -25.8989 | 0.0000 | 25.8989 | 0.5000 |

LM test for autocorrelation up to order 12 -

Null hypothesis: no autocorrelation

Test statistic: Chi-square(3) = 17.9674

Test for autocorrelation up to order 12

Ljung-Box Q' = 17.9674,

with p-value = P(Chi-square(3) > 17.9674) = 0.0004467

1. What kind of metrics are the Akaike (AIC), Schwartz (BIC) and Hannan-Quinn statistics?

AIC, BIC, and Hannan-Quinn statistics are goodness of fit metrics used for model selection. The smaller the metric, the better the fit. The idea is to produce various ARIMA models and compare these metrics.

1. Which two are the most conservative in terms of penalizing the model for degrees of freedom?

BIC and Hannan-Quinn are more conservative because they penalize the model more for additional variables.

1. What does the Ljung Box Q test test for ?

The Ljung-Box Q statistic measures the amount of autocorrelation left in the residuals of the model. It is looking to see if there is any variance left after the model is done that might be better explained by another model.

Null hypothesis: No evidence of serial autocorrelation in the residuals.

Alternative hypothesis: There is serial autocorrelation in the residuals.

Ideally, the p-value should be > 0.05. Failing to reject the null hypothesis suggests that any variance left is unexplainable.

1. Create a table with the ARIMA model designation, adjusted R square, AIC, BIC and Ljung Box values for the four models. What looks like the best model of the four? How do you tell? **Be sure to paste your table into this exercise!**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ARIMA Model | Adj. R square | AIC | BIC | Ljung Box Q |
| Model 1 | 0.085465 | −317.2540 | −298.1985 | 75.3636 |
| Model 2 | 0.121101 | −350.5570 | −326.7375 | 36.8101 |
| Model 3 | 0.134286 | −362.1069 | −333.5236 | 39.2977 |
| Model 4 | 0.152845 | −375.5881 | −323.1854 | 17.9674 |

**Model 4** looks like it’s the best model of the four because it has the highest adjusted R-square and lowest AIC values. Also, six of the nine independent variables in Model 4 are statistically significant, the most of any model.

1. Examining the Ljung Box test statistic, do you think that there is more variance in the residuals that you might be able to find with some additional ARIMA models?

Yes, the p-values of all four ARIMA models are less than 0.05, indicating that variance in these models might be better explained by another model.