DA6823

Time Series Project

Name: Connor Allison

ID: fbu535

The objective of this project is for you to practice what you have learned about time series analysis and interpreting data. I suggest you use GRETL for this project. **Be sure that you cut and paste your answers to each of the questions for the project. If you talk about something in a table or plot, that table or plot needs to be in your report!!! If the question says plot something, cut and paste that plot into your report.** In previous semesters I have had students talk about the plot but not display it – that makes no sense.

1. Select a scientific, biomedical, business or other issue that appeals to you and go looking online for relevant time series data sets. The good news here is that there are tons of free and interesting time series data sets online. If you have problems locating them let me know and I will help. **Be sure that it looks like there is little or no seasonality to it.**

I have found a dataset that has a company’s monthly cost of production from 2015-2022.

1. Plot out your time series variable. Tell me using your Mark I eyeball whether or not you think the time series data set is stationary in terms of **constant mean** and also **constant variance**. Note that you should avoid time series data sets that have huge spikes in them (they are hard to model using GRETL) and also avoid data sets where the data plot looks like a straight line going up or down – those aren’t very interesting.



Using the Mark 1 eyeball I think that the data set does not have a constant mean over time. I also think that it does have constant variance over time.

1. Plot the ACF for the time series data set. Looking at ACF, does it look like there may be a trend or non-constant mean for each time series?



Looking at the ACF plot of the time series data, it seems like there is a trend or non-constant mean.

1. Now let’s examine the time series data set using unit root tests. First use the KPSS test for the time series data set and tell me if the test suggests if there is a constant mean or not. Then see if you can confirm your KPSS evaluation using the Augmented Dickey Fuller (ADF) or the ADF-GLS test and tell me what the ADF test suggests is the case.

KPSS test for Average\_cost

T = 64

Lag truncation parameter = 3

Test statistic = 0.586672

10% 5% 1%

Critical values: 0.351 0.462 0.729

Interpolated p-value 0.031

The p-value is < .05 so we can reject the null hypothesis that there is no trend (constant mean).

Augmented Dickey-Fuller test for Average\_cost

testing down from 10 lags, criterion AIC

sample size 63

unit-root null hypothesis: a = 1

test with constant

including 0 lags of (1-L)Average\_cost

model: (1-L)y = b0 + (a-1)\*y(-1) + e

estimated value of (a - 1): -0.692059

test statistic: tau\_c(1) = -5.75113

asymptotic p-value 4.644e-07

1st-order autocorrelation coeff. for e: -0.081

The p value is 4.644e-07 which allows us to reject the null hypothesis and conclude that there is no trend.

1. Summarize the results of steps 2 through 4 and tell what your decision is regarding constant mean in the time series data set.

The mark 1 eyeball, ACF and KPSS all indicated that there was a trend or non-constant mean. The KPSS test was very close and would have suggested no trend or a constant mean if we had set our critical value threshold at .01. On the other hand, the ADF test suggested no trend or constant mean. Overall, the results were three to one favoring non-constant mean or trend in the time series data.

1. Review the decision in step #5. If the test suggests that there is a non-constant mean then use differencing to create a new differenced variable for the time series **data set and proceed to the steps below (a,b,c). Be sure to cut and paste your supporting evidence (unit root tests, plots, etc.) below.** If you got luck and concluded that your data set already has a constant mean then you can skip all of step 6 and move on using your data set without differencing!
   1. Plot out the data for the new differenced data set. Tell me if it looks like the differencing got rid of the trend or non-constant mean.



After differencing the data, it looks like we were able to get rid of the trend.

* 1. Plot the ACF for the differenced time series. Tell me if this new ACF plot looks like there now is no trend.



The new ACF plot looks like there is now no trend.

* 1. Apply the KPSS test and the ADF or ADF-GLS test to the differenced data – does the trend disappear?

KPSS test for d\_Average\_cost

T = 63

Lag truncation parameter = 3

Test statistic = 0.0387766

10% 5% 1%

Critical values: 0.351 0.462 0.728

P-value > .10

Augmented Dickey-Fuller test for d\_Average\_cost

testing down from 10 lags, criterion AIC

sample size 61

unit-root null hypothesis: a = 1

test with constant

including one lag of (1-L)d\_Average\_cost

model: (1-L)y = b0 + (a-1)\*y(-1) + ... + e

estimated value of (a - 1): -1.95887

test statistic: tau\_c(1) = -9.08846

asymptotic p-value 2.537e-16

1st-order autocorrelation coeff. for e: -0.036

The ADF test still suggests no trend and the new results from the KPSS test also suggest no trend. Overall, any indication of trend has disappeared in all four of our tests.

**Note: From this point onward through step 9, if the time series was differenced, use the differenced time series data set for all the rest of the questions. Otherwise you can use the undifferenced data set.**

1. Plot the PACF for the time series data set. Using the combined information from the ACF you plotted earlier along with the information in the PACF, tell me if you see any autoregressive and/or moving average processes in the data set and what they are. Use the discussion in class as well as online resources – here is a decent resource from Duke University [**https://people.duke.edu/~rnau/411arim3.htm**](https://people.duke.edu/~rnau/411arim3.htm) or Penn State <https://onlinecourses.science.psu.edu/stat510/node/64>



Looking at the ACF and PACF, we see a clear first-order moving average process. We see no signs of autoregressive processes. There is also an apparent 11th order moving average when we look just at the PACF. Additionally, when we compare the two plots, we see a possible 7th order moving average, but it doesn’t seem to be very strong. It seems like a (0,0,1) model would be a good starting point.

1. For your time series data set, experiment with different ARIMA models for them. Try at least four models. As you try them, list out the results of the various models and

**Model 1: (0,0,1)**

Function evaluations: 45

Evaluations of gradient: 23

Model 1: ARMA, using observations 2015:02-2020:04 (T = 63)

Estimated using AS 197 (exact ML)

Dependent variable: d\_Average\_cost

Standard errors based on Hessian

coefficient std. error z p-value

-------------------------------------------------------

const 7.88734 6.59058 1.197 0.2314

theta\_1 −0.823507 0.0977337 −8.426 3.58e-17 \*\*\*

Mean dependent var 11.23923 S.D. dependent var 346.8955

Mean of innovations 3.737395 S.D. of innovations 273.2958

R-squared 0.371221 Adjusted R-squared 0.371221

Log-likelihood −443.4249 Akaike criterion 892.8499

Schwarz criterion 899.2793 Hannan-Quinn 895.3786

Real Imaginary Modulus Frequency

-----------------------------------------------------------

MA

Root 1 1.2143 0.0000 1.2143 0.0000

-----------------------------------------------------------

Test for autocorrelation up to order 12

Ljung-Box Q' = 11.4387,

with p-value = P(Chi-square(11) > 11.4387) = 0.4073

Test for ARCH of order 12

coefficient std. error t-ratio p-value

-------------------------------------------------------------

alpha(0) 212330 68490.3 3.100 0.0036 \*\*\*

alpha(1) −0.163569 0.163365 −1.001 0.3230

alpha(2) −0.0635483 0.159610 −0.3981 0.6927

alpha(3) 0.0425218 0.151223 0.2812 0.7801

alpha(4) −0.0782549 0.151225 −0.5175 0.6078

alpha(5) −0.236564 0.147377 −1.605 0.1167

alpha(6) −0.232292 0.150508 −1.543 0.1310

alpha(7) −0.124981 0.153876 −0.8122 0.4217

alpha(8) 0.0598500 0.149876 0.3993 0.6919

alpha(9) −0.207447 0.213187 −0.9731 0.3367

alpha(10) −0.460560 0.213810 −2.154 0.0376 \*\*

alpha(11) −0.274648 0.222347 −1.235 0.2243

alpha(12) −0.0807948 0.220224 −0.3669 0.7157

Null hypothesis: no ARCH effect is present

Test statistic: LM = 10.7739

with p-value = P(Chi-square(12) > 10.7739) = 0.548385

**Model 2: (1,0,1)**

Function evaluations: 44

Evaluations of gradient: 21

Model 2: ARMA, using observations 2015:02-2020:04 (T = 63)

Estimated using AS 197 (exact ML)

Dependent variable: d\_Average\_cost

Standard errors based on Hessian

coefficient std. error z p-value

--------------------------------------------------------

const 7.63674 5.39270 1.416 0.1567

phi\_1 0.119647 0.172382 0.6941 0.4876

theta\_1 −0.880747 0.119144 −7.392 1.44e-13 \*\*\*

Mean dependent var 11.23923 S.D. dependent var 346.8955

Mean of innovations 7.540762 S.D. of innovations 271.8844

R-squared 0.376414 Adjusted R-squared 0.366192

Log-likelihood −443.1752 Akaike criterion 894.3503

Schwarz criterion 902.9229 Hannan-Quinn 897.7219

Real Imaginary Modulus Frequency

-----------------------------------------------------------

AR

Root 1 8.3579 0.0000 8.3579 0.0000

MA

Root 1 1.1354 0.0000 1.1354 0.0000

-----------------------------------------------------------

Null hypothesis: no ARCH effect is present

Test statistic: LM = 10.7739

with p-value = P(Chi-square(12) > 10.7739) = 0.548385

Test for autocorrelation up to order 12

Ljung-Box Q' = 10.9552,

with p-value = P(Chi-square(10) > 10.9552) = 0.3610

Test for ARCH of order 12

coefficient std. error t-ratio p-value

-------------------------------------------------------------

alpha(0) 230511 74158.4 3.108 0.0036 \*\*\*

alpha(1) −0.229736 0.164530 −1.396 0.1707

alpha(2) −0.165972 0.162197 −1.023 0.3126

alpha(3) −0.0336619 0.157769 −0.2134 0.8322

alpha(4) −0.111103 0.157166 −0.7069 0.4839

alpha(5) −0.246029 0.153275 −1.605 0.1167

alpha(6) −0.221445 0.155389 −1.425 0.1623

alpha(7) −0.181658 0.157369 −1.154 0.2556

alpha(8) 0.0600408 0.154872 0.3877 0.7004

alpha(9) −0.186888 0.200544 −0.9319 0.3573

alpha(10) −0.347519 0.201065 −1.728 0.0920 \*

alpha(11) −0.269782 0.202687 −1.331 0.1911

alpha(12) −0.110380 0.201388 −0.5481 0.5868

Null hypothesis: no ARCH effect is present

Test statistic: LM = 10.9987

with p-value = P(Chi-square(12) > 10.9987) = 0.52903

**Model 3: (0,0,7)**

Function evaluations: 144

Evaluations of gradient: 54

Model 3: ARMA, using observations 2015:02-2020:04 (T = 63)

Estimated using AS 197 (exact ML)

Dependent variable: d\_Average\_cost

Standard errors based on Hessian

coefficient std. error z p-value

--------------------------------------------------------

const 6.60257 3.23897 2.038 0.0415 \*\*

theta\_1 −0.754177 0.150602 −5.008 5.51e-07 \*\*\*

theta\_2 −0.0980574 0.182652 −0.5369 0.5914

theta\_3 0.0761863 0.193906 0.3929 0.6944

theta\_4 −0.118240 0.133880 −0.8832 0.3771

theta\_5 −0.0419306 0.172721 −0.2428 0.8082

theta\_6 0.100584 0.177178 0.5677 0.5702

theta\_7 −0.164364 0.130798 −1.257 0.2089

Mean dependent var 11.23923 S.D. dependent var 346.8955

Mean of innovations 20.22053 S.D. of innovations 260.9068

R-squared 0.435393 Adjusted R-squared 0.374900

Log-likelihood −441.5365 Akaike criterion 901.0730

Schwarz criterion 920.3612 Hannan-Quinn 908.6592

Real Imaginary Modulus Frequency

-----------------------------------------------------------

MA

Root 1 1.0000 0.0000 1.0000 0.0000

Root 2 -1.1930 0.5932 1.3324 0.4266

Root 3 -1.1930 -0.5932 1.3324 -0.4266

Root 4 1.0157 0.8880 1.3491 0.1143

Root 5 1.0157 -0.8880 1.3491 -0.1143

Root 6 -0.0167 -1.3721 1.3722 -0.2519

Root 7 -0.0167 1.3721 1.3722 0.2519

-----------------------------------------------------------

Test for autocorrelation up to order 12

Ljung-Box Q' = 9.46073,

with p-value = P(Chi-square(5) > 9.46073) = 0.09204

0,0,Test for ARCH of order 12

coefficient std. error t-ratio p-value

---------------------------------------------------------------

alpha(0) 132471 52906.8 2.504 0.0167 \*\*

alpha(1) −0.153551 0.155159 −0.9896 0.3286

alpha(2) −0.00762041 0.149112 −0.05111 0.9595

alpha(3) 0.106508 0.143458 0.7424 0.4624

alpha(4) −0.0252473 0.143573 −0.1758 0.8613

alpha(5) −0.167450 0.139457 −1.201 0.2373

alpha(6) −0.198294 0.139946 −1.417 0.1646

alpha(7) −0.144266 0.142023 −1.016 0.3162

alpha(8) 0.123903 0.142125 0.8718 0.3888

alpha(9) −0.0614660 0.158177 −0.3886 0.6997

alpha(10) −0.201323 0.157669 −1.277 0.2094

alpha(11) −0.284344 0.160552 −1.771 0.0846 \*

alpha(12) 0.107140 0.165182 0.6486 0.5205

Null hypothesis: no ARCH effect is present

Test statistic: LM = 12.2874

with p-value = P(Chi-square(12) > 12.2874) = 0.422879

**Model 4: (0,0,2)**

Function evaluations: 70

Evaluations of gradient: 31

Model 4: ARMA, using observations 2015:02-2020:04 (T = 63)

Estimated using AS 197 (exact ML)

Dependent variable: d\_Average\_cost

Standard errors based on Hessian

coefficient std. error z p-value

--------------------------------------------------------

const 7.74213 5.67566 1.364 0.1725

theta\_1 −0.771152 0.122645 −6.288 3.22e-10 \*\*\*

theta\_2 −0.0823908 0.130845 −0.6297 0.5289

Mean dependent var 11.23923 S.D. dependent var 346.8955

Mean of innovations 6.240276 S.D. of innovations 272.1788

R-squared 0.374784 Adjusted R-squared 0.364535

Log-likelihood −443.2194 Akaike criterion 894.4388

Schwarz criterion 903.0113 Hannan-Quinn 897.8104

Real Imaginary Modulus Frequency

-----------------------------------------------------------

MA

Root 1 1.1544 0.0000 1.1544 0.0000

Root 2 -10.5141 0.0000 10.5141 0.5000

-----------------------------------------------------------

Test for autocorrelation up to order 12

Ljung-Box Q' = 11.1039,

with p-value = P(Chi-square(10) > 11.1039) = 0.3495

Test for ARCH of order 12

coefficient std. error t-ratio p-value

--------------------------------------------------------------

alpha(0) 225691 72303.0 3.121 0.0034 \*\*\*

alpha(1) −0.209276 0.164501 −1.272 0.2110

alpha(2) −0.147590 0.161404 −0.9144 0.3663

alpha(3) −0.0140074 0.156116 −0.08972 0.9290

alpha(4) −0.103995 0.155402 −0.6692 0.5074

alpha(5) −0.235887 0.151451 −1.558 0.1276

alpha(6) −0.215586 0.153739 −1.402 0.1689

alpha(7) −0.165540 0.155905 −1.062 0.2950

alpha(8) 0.0595581 0.153352 0.3884 0.6999

alpha(9) −0.197715 0.203371 −0.9722 0.3371

alpha(10) −0.372774 0.204133 −1.826 0.0757 \*

alpha(11) −0.275253 0.207028 −1.330 0.1916

alpha(12) −0.105994 0.206022 −0.5145 0.6099

Null hypothesis: no ARCH effect is present

Test statistic: LM = 10.7921

with p-value = P(Chi-square(12) > 10.7921) = 0.546818

* 1. Construct a table with the identity of the model, the R square, the AIC, BIC(Schwartz), the Hannan-Quinn, Lejune-Box and a final column that notes the terms that are significant in the model. **Be sure to paste that table into your project report!**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | adj R square | AIC | BIC | HQ | Lijung Box | ARCH | sig terms |
| 1: (0,0,1) | 0.371 | 892.8 | 899.2 | 895.3 | 0.407 | 0.548 | MA1 |
| 2: (1,0,1) | 0.366 | 894.3 | 902.9 | 897.7 | 0.361 | 0.529 | MA1 |
| 3: (0,0,7) | 0.374 | 901.0 | 920.36 | 908.6 | 0.092 | 0.422 | MA1 |
| 4: (0,0,2) | 0.364 | 894.4 | 903 | 897.8 | 0.349 | 0.546 | MA1 |

* 1. Plot the observed versus fitted data for the time series data set **for each model.**

**Model 1 (0,0,1)**

****

**Model 2 (1,0,1)**

****

**Model 3 (0,0,7)**

****

**Model 4 (0,0,2)**

****

* 1. Pick one of the models as your favorite and tell me why you like that one the best.

The first model is my favorite because it accounts for the most variance. We know this because it has the highest Lijung Box, and the lowest AIC, BIC and HQ.

* 1. Forecast your model out 6 time periods and graph the time series including the forecast. How well does the forecast seem to work?

****

The forecast is not very good. The first forecasted data point might be good, but the rest are just in a straight line which suggests that we do not have a high enough Adj R-Squared. Bummer ☹

1. Test the time series data set you select for constant variance using the ARCH test (GRETL does this nicely). Note that we will not do anything about this issue for the moment, but it’s good to know.

**Model 1 (0,0,1):**

Test for ARCH of order 12

coefficient std. error t-ratio p-value

-------------------------------------------------------------

alpha(0) 212330 68490.3 3.100 0.0036 \*\*\*

alpha(1) −0.163569 0.163365 −1.001 0.3230

alpha(2) −0.0635483 0.159610 −0.3981 0.6927

alpha(3) 0.0425218 0.151223 0.2812 0.7801

alpha(4) −0.0782549 0.151225 −0.5175 0.6078

alpha(5) −0.236564 0.147377 −1.605 0.1167

alpha(6) −0.232292 0.150508 −1.543 0.1310

alpha(7) −0.124981 0.153876 −0.8122 0.4217

alpha(8) 0.0598500 0.149876 0.3993 0.6919

alpha(9) −0.207447 0.213187 −0.9731 0.3367

alpha(10) −0.460560 0.213810 −2.154 0.0376 \*\*

alpha(11) −0.274648 0.222347 −1.235 0.2243

alpha(12) −0.0807948 0.220224 −0.3669 0.7157

Null hypothesis: no ARCH effect is present

Test statistic: LM = 10.7739

with p-value = P(Chi-square(12) > 10.7739) = 0.548385