Time Series Midterm Summer 2020

In-Class Portion

Please put your answers directly on this Word document. When completed, please submit the completed Word doc to the “Midterm” assignment on 2DS and email the test back to me at [bsadler@smu.edu](mailto:bsadler@smu.edu). The take home portion will be due by 11:59pm on Saturday, June 27 and that should be submitted on 2DS under the Midterm assignment as well (and emailed to me.) You can find the take home portion on Github in the Unit 8 folder**. By taking this test you are agreeing to not communicate with any human being (except for Bivin Sadler) about the in-class or take home portion of this test. You may use your notes, the book, the internet, etc. …. Just no live human or AI (unless you have a question for me.)**

Note throughout the test: “By hand” simply means show your work as if you were writing it by hand. You may actually write it by hand and take a pic or you may type it.

The following data are quarterly sales (in units sold) for a make-believe company:

|  |  |
| --- | --- |
| Time | Units |
| 1 | 1 |
| 2 | 5 |
| 3 | 10 |
| 4 | 9 |
| 5 | 4 |
| 6 | 6 |
| 7 | 12 |

1. (4pts) Compute by hand.

Appendix

1. (4pts) Compute by hand.

Appendix

1. (4 pts) Compute by hand.

Appendix

Using the data above, consider the Model: where the phis are estimated and is unknown but not assumed to be zero.

Appendix

1. (4 pts) Compute (1) by hand showing your work. You may work it on paper and submit a pic / scan of your work or you may simply type what you would have written by hand.

Appendix

1. (6 pts) Compute a 95% probability interval for (1) by hand. “By hand” means show your calculation of the margin of error as we did in live session. However, in your calculation, you should assume that your client has asked you to use = 3.0. A value that they calculated in another study that they prefer to continue to use in this study.

1. (4pts) Write this AR(1) model as a GLP (just the first 4 terms).

Appendix

1. (4pts) Will this model have stronger (more dominant) wandering behavior, oscillating behavior or periodic behavior with period greater than 2? Why?

factor.wge(phi = c(.2,.08))

This has more of a wandering behavior because the abs reciprocal value at frequency of 0 is larger than that of the one in 0.5

1. Matching: Match the letter to the number below the labeled ACFs and realizations. (2pts each )

|  |  |
| --- | --- |
| 1. | A. |
| 2. | B. |
| 3. | C. |
| 4. | D. |
| 5. | E. |

1. B 2. E 3. C 4. A 5. D
2. (4pts) Below is a realization of sales data of swimsuits in Dallas, Texas. Talking with a the senior vice president (SVP) of the company that recorded the data, she has indicated that over the past 30 years they have seen below average sales in October and continuing to decline until March in which the sales steadily increase till August in which they begin to decrease through which they fall below average in October and the process repeats. The SVP also stated that she and the industry fully expect these trends to continue for the foreseeable future as swimsuit sales tend to be correlated with the temperature in Dallas, Texas. Below the plot, answer the question, is this data likely generated from a stationary or non-stationary process? Why? (One to three sentences should do!)



Swimsuit Sales

Months

This is a non-stationary model because mean is dependent on time. This is also seems to be seasonal data.

1. Use the bank of models below to assign possible models to each visual representation of forecasts.(2pts each)

|  |  |
| --- | --- |
| 1. Signal Plus Noise Model with linear trend. 2. Signal Plus Noise Model with Cos trend. 3. AR(1) with a positive real root 4. AR(2) with 2 positive real roots 5. AR(10) with all positive real roots 6. AR(1) with a negative real root | 1. AR(2) with 2 negative real roots 2. AR(10) with all negative real roots 3. AR(2) with complex conjugate roots 4. ARIMA(pdq) with p = 0, q = 0, d = 1 and s = 12. 5. ARIMA(pdq) with p = 2, q = 0, d = 1 and s = 0 |

|  |  |
| --- | --- |
| **Forecasts** | List all possible models that could generate the forecasts to the left. Just put the letter(s) from the model(s) above below. There may be more than one answer. |
| Example: | Answer(s): F,G,H |
|  | Answer(s):K |
|  | Answer(s):C,D |
|  | Answer(s):A |
|  | Answer(s):J |
|  | Answer(s):ARIMA(0,1,0) |
|  | Answer(s):I |

According to the asynchronous materials aruma model (0,1,0) would predict the last point for all points so would fit fig. 5 but then J would have fit but it has s=12 which changes it. K has a p =2 which would go up and down.

1. (3 pts) Estimate the period(s) that is/are associated with this spectral density. It just needs to be close… not exact.



1. (3pts) What are the minimum and maximum number of peaks an AR(7) could have in the true Spectral density?

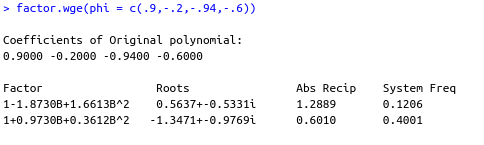
Maximum would be 3 AR(2) with conjugate roots and one AR(1). So it would be 3 + 1(at 0 frequency) = 4. Minimum would be 1 at 0 frequency.

For questions 13-15, consider the model below.

1. (4pts) Describe the behavior of realizations generated from the model. Specifically comment on any wandering, oscillating and or periodic behavior and the dominance of those behaviors in generated realizations.

The AR part of the model is not stationary and neither has a (1-B)^d or

nor (1-B^s) factors in it. Couldn’t generate any plot.



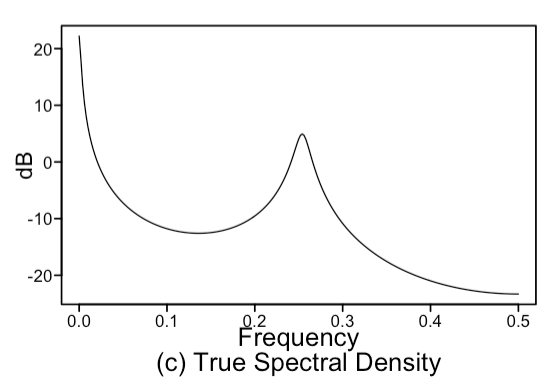
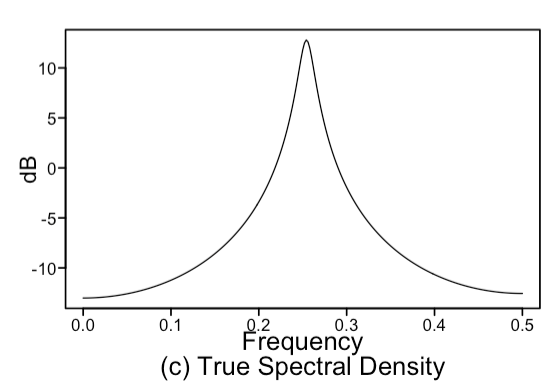
1. (4pts) Write the model in the last question in GLP form (just the first 4 terms.).



1. (4pts) Is the model stationary or non-stationary … invertible or non-invertible? Why?

The model is non-stationary as well as non-invertible because there was a root inside the unit circle for both the AR part and the MA part.

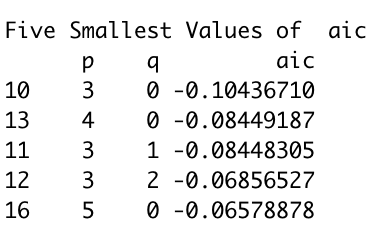
1. (3pts) Assume the data has passed through a filter. The spectral densities are displayed below both before and after the filter is applied to the data. Which is filter was applied to the data?
   1. Moving Average Filter
   2. 1st Difference Filter

BEFORE FILTER: AFTER FILTER: 

1. (4pts) Name one reason why we should consider only invertible models?

2 Non-invertible models can have same autocorrelation which is undesirable makes for a model multiplicity.

1. (3 pts) True / False Consider the output from a call to aic5.wge() below:



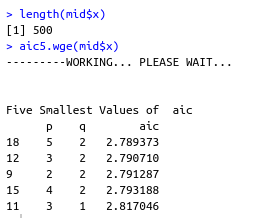
True or False: The ARMA(3,0) model must be more useful in forecasting than the ARMA(3,2) model.

False : Aic is just one of many ways of measuring performance.

Questions 19 – 20 Consider the data on the Wall in the file, *MidtermSummer2020InClass.csv*.

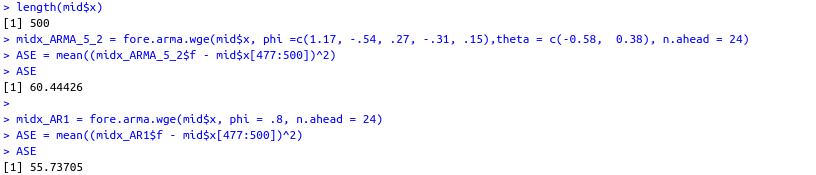
1. (3 pts) First let’s assume the data come from a stationary process. What is the most favored correlation structure (identify p and q) with respect to the AIC? Assume you only want to consider ARMA(p,q) models with p < 15 and q < 5. All you have to do here is identify p and q.

The best ARMA Model with lowest AIC measure is ARMA(5,2)



1. Next we would like to compare the usefulness of a stationary and non-stationary model in terms of ASE rather than the AIC.
   1. (*6 pts) Find* and report the ASE (and show your code) for a horizon of 24 from the following two models:

There weren’t Any model displayed for A. Will use ARMA(5,2) from above instead.

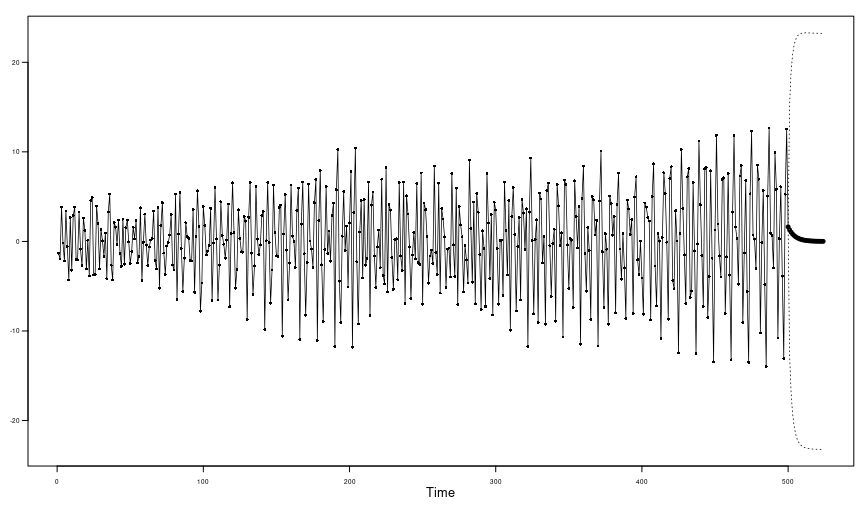


Model B has the lower ASE which means its better when it comes to ASE.

* 1. (1 pt) Given your answer in part a, which model would be most favorable for forecasting the next 24 observations with respect to the ASE?

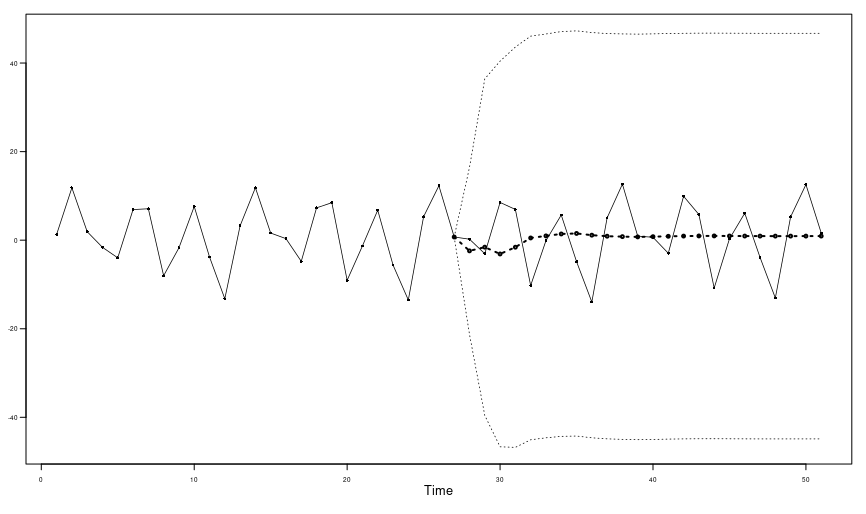
Model B

* 1. Use the model you chose in part b to forecast the next 24 observations from the end of the dataset.

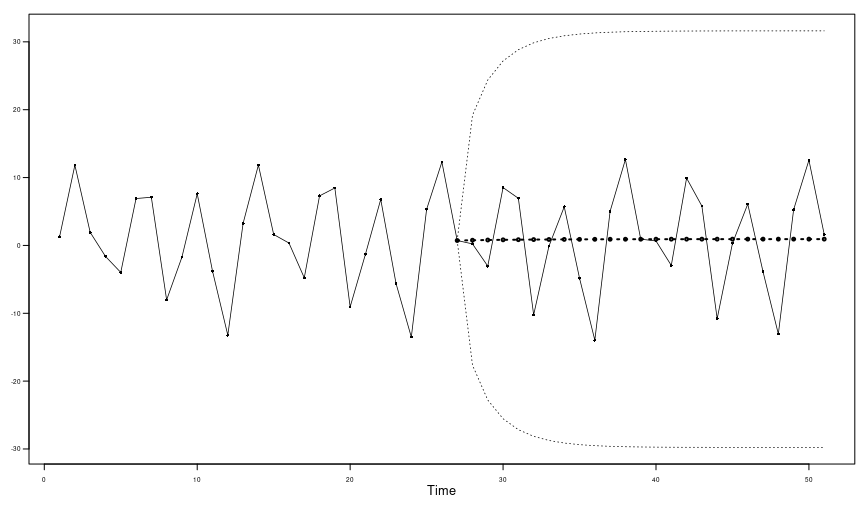


BONUS: Plotting the entire dataset with the forecasts makes the forecasts tough to see and assess. Come up with a good visualization that you would show your client that would clearly show them the 24 forecasts superimposed on the last 24 observations (visualizing the ASE.) Note these are the forecasts used to find the ASE **not** the forecasts from the end of the realization.

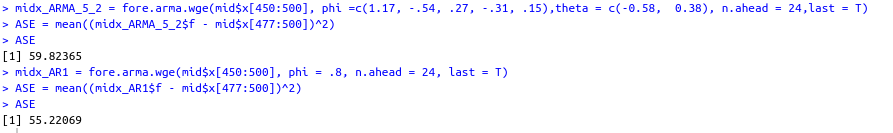
ARMA(5,2)



AR(1)



Both were plotted from 450 to 500 points.



APPENDIX

