

EAS509 Exam 2

Submit your answers as a single pdf attach all R code. Failure to do so will result in grade reduction.

The exam must be done individually, with no discussion or help with others. Breaking this rule will result in an automatic 0 grade.

Part A (30 points) - each question worth 1 points

Some questions have multiple answers

1. Which simple forecasting method says the forecast is equal to the mean of the historical data?

- a. Average Method
- b. Naïve Method
- c. Seasonal Naïve Method
- d. Drift Method

Answer:

2. Which simple forecasting method says the forecast is equal to the last observed value?

- a. Average Method
- b. Naïve Method
- c. Seasonal Naïve Method
- d. Drift Method

Answer:

3. Which simple forecasting method is equivalent to extrapolating a line draw between the first and last observations?

- a. Average Method
- b. Naïve Method
- c. Seasonal Naïve Method
- d. Drift Method

Answer:

4. Which of the following is an assumption made about forecasting residuals during point forecast?

- a. Residuals are normally distributed
- b. Residuals are uncorrelated
- c. Residuals have constant variance

- d. None of the above

Answer:

5. Which of the following is an assumption made about forecasting residuals during interval forecasting?
(multiple answers)
- a. Residuals have mean zero
 - b. Residuals are normally distributed
 - c. Residuals have constant variance
 - d. None of the above

Answer: all should present for full score

6. What is the consequence of forecasting residuals that are not uncorrelated?
- a. Prediction intervals are difficult to calculate
 - b. Information is left in the residuals that should be used
 - c. Forecasts are biased
 - d. None of the above

Answer:

7. What is the consequence of forecasting residuals that don't have mean zero?
- a. Prediction intervals are difficult to calculate
 - b. Information is left in the residuals that should be used
 - c. Forecasts are biased
 - d. None of the above

Answer:

8. Which measure of forecast accuracy is scale independent?
- a. MAE
 - b. MSE
 - c. RMSE
 - d. MAPE

Answer:

9. Calculation of forecasts is based on what?
- a. Test set
 - b. Training set
 - c. Both
 - d. Neither

Answer:

10. Forecast accuracy is based on what?

- a. Test set
- b. Training set
- c. Both
- d. Neeither

Answer:

11. A series that is influenced by seasonal factors is known as what?

- a. Trend
- b. Seasonal
- c. Cyclical
- d. White Noise

Answer:

12. Data that exhibits rises and falls that are not of a fixed period is known as what?

- a. Trend
- b. Seasonal
- c. Cyclical
- d. White Noise

Answer: either or all is ok for full credit

13. Data that is uncorrelated over time is known as what?

- a. Trend
- b. Seasonal
- c. Cyclical
- d. White Noise

Answer:

14. Which of the following time series decomposition models is appropriate when the magnitude of the seasonal fluctuations are not proportional to the level?

- a. Additive
- b. Multiplicative
- c. Both
- d. Neither

Answer:

15. Which of the following time series decomposition models is appropriate when the magnitude of the seasonal fluctuations are proportional to the level?

- a. Additive

- b. Multiplicative
- c. Both
- d. Neither

Answer:

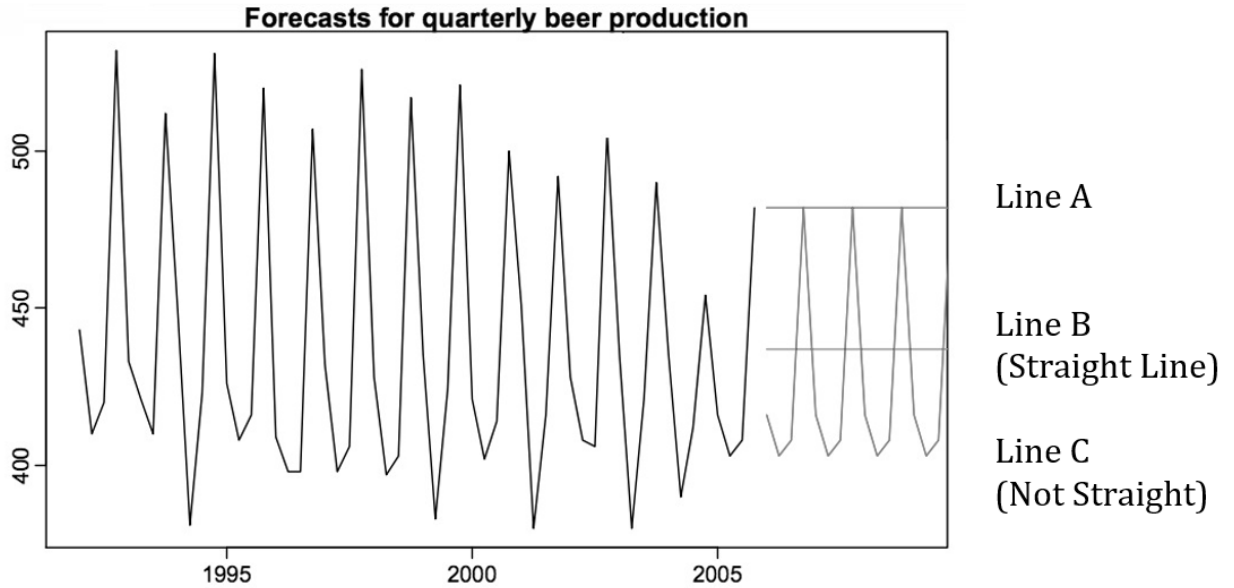


Exhibit 1

16. Refer to Exhibit 1. Line A is which simple forecasting method?

- a. Average Method
- b. Naïve Method
- c. Seasonal Naïve Method
- d. Drift

Answer:

17. Refer to Exhibit 1. Line B is which simple forecasting method?

- a. Average Method
- b. Naïve Method
- c. Seasonal Naïve Method
- d. Drift Method

Answer:

18. Refer to Exhibit 1. Line C is which simple forecasting method?

- a. Average Method
- b. Naïve Method
- c. Seasonal Naïve Method
- d. Drift Method

Answer:

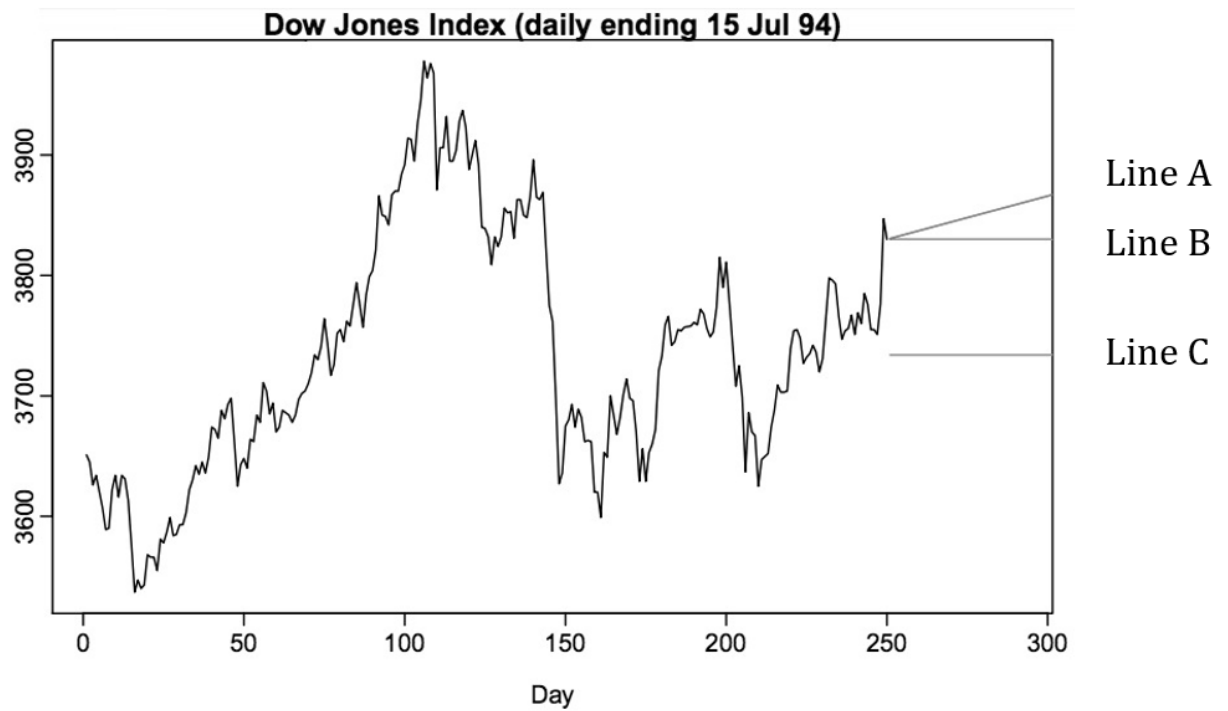


Exhibit 2

19. Refer to Exhibit 2. Line A is which simple forecasting method?

- a. Average Method
- b. Naïve Method
- c. Seasonal Naïve Method
- d. Drift Method

Answer:

20. Refer to Exhibit 2. Line B is which simple forecasting method?

- a. Average Method
- b. Naïve Method
- c. Seasonal Naïve Method
- d. Drift Method

Answer:

21. Refer to Exhibit 2. Line C is which simple forecasting method?

- a. Average Method
- b. Naïve Method
- c. Seasonal Naïve Method
- d. Drift Method

Answer:

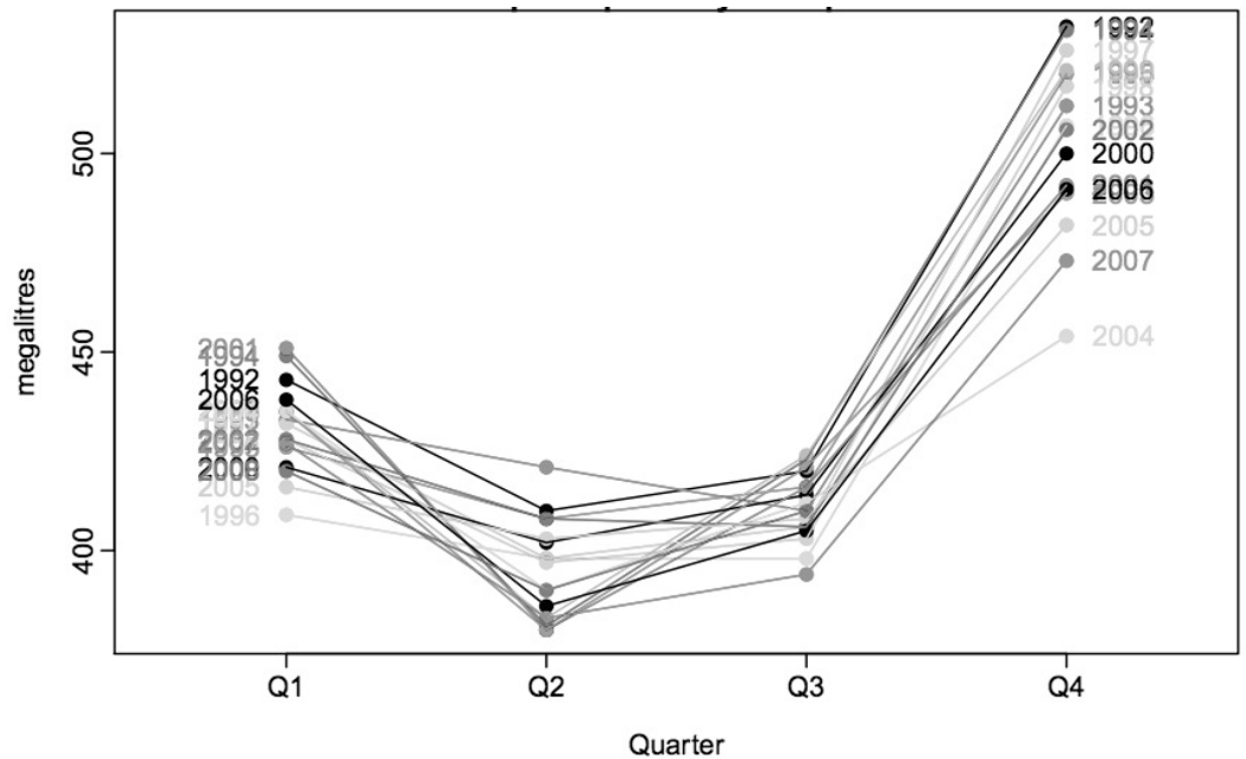


Exhibit 3

22. Refer to Exhibit 3. The peaks are in which quarter?

- a. Quarter 1
- b. Quarter 2
- c. Quarter 3
- d. Quarter 4

Answer:

23. Refer to Exhibit 3. The trough are in which quarter?

- a. Quarter 1
- b. Quarter 2
- c. Quarter 3
- d. Quarter 4

Answer: there are few in Q3 but largely it is Q2

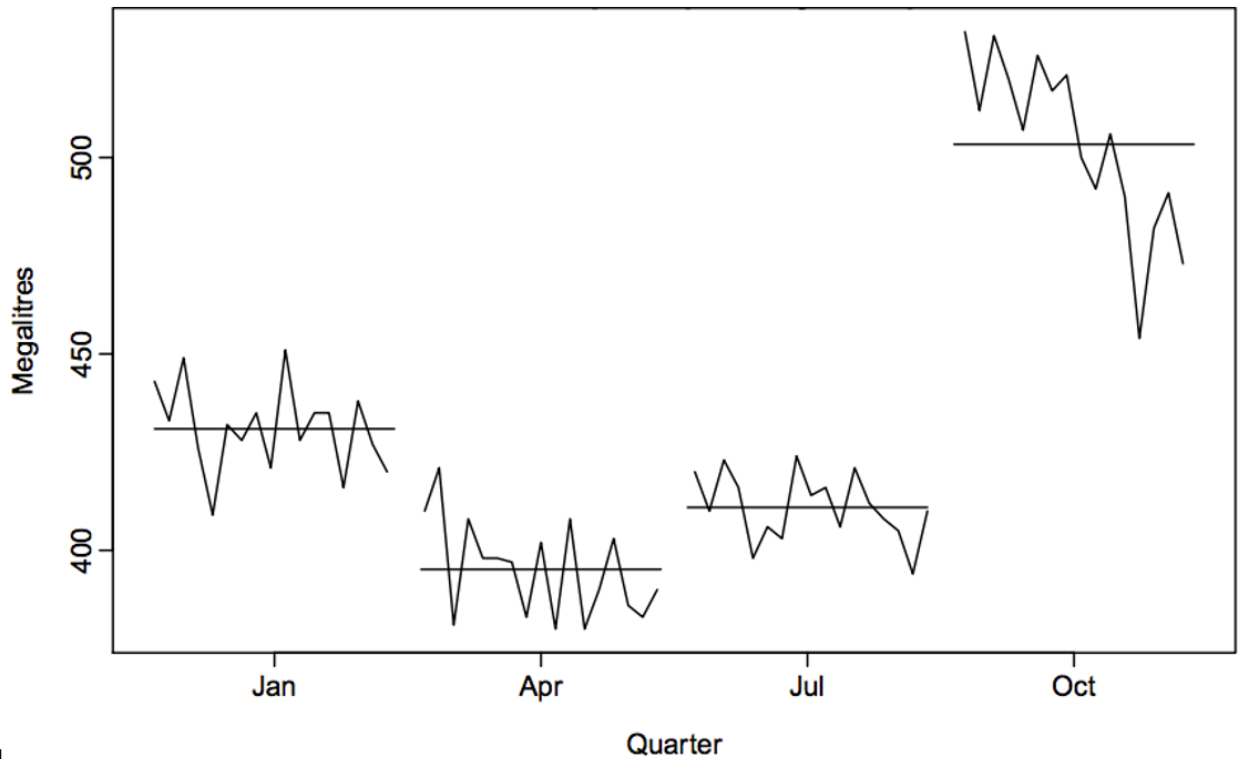


Exhibit 4

24. Refer to Exhibit 4. The peaks are in which quarter?

- a. Quarter 1
- b. Quarter 2
- c. Quarter 3
- d. Quarter 4

Answer:

25. Refer to Exhibit 4. The trough are in which quarter?

- a. Quarter 1
- b. Quarter 2
- c. Quarter 3
- d. Quarter 4

Answer:

26. Refer to Exhibit 4. In which quarter is there a decline in the seasonal affect?

- a. Quarter 1
- b. Quarter 2
- c. Quarter 3
- d. Quarter 4

Answer:

Figure 5

Year 1				Year 2			
Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
10	6	8	12	11	7	9	13

27. Refer to Figure 5. Using the average method, what is the forecast of Quarter 2 of Year 3? (Don't use a calculator.)

- a. 7
- b. 9.5
- c. 13.85
- d. 13

Answer:

28. Refer to Figure 5. Using the naïve method, what is the forecast of Quarter 2 of Year 3? (Don't use a calculator.)

- a. 7
- b. 9.5
- c. 13.85
- d. 13

Answer:

29. Refer to Figure 5. Using the seasonal naïve method, what is the forecast of Quarter 2 of Year 3? (Don't use a calculator.)

- a. 7
- b. 9.5
- c. 13.85
- d. 13

Answer:

30. Refer to Figure 5. Using the drift method, what is the forecast of Quarter 2 of Year 3? (Don't use a calculator.)

- a. 7
- b. 9.5
- c. 13.85
- d. 13

Answer:

Part B (30 points)

Choose a series from `us_employment.csv`, the total employment in leisure and hospitality industry in the United States (see, title column).

- Produce an STL decomposition of the data and describe the trend and seasonality. (4 points)
- Do the data need transforming? If so, find a suitable transformation.(4 points)
- Are the data stationary? If not, find an appropriate differencing which yields stationary data.(4 points)
- Identify a couple of ARIMA models that might be useful in describing the time series. Which of your models is the best according to their AICc values?(5 points)
- Estimate the parameters of your best model and do diagnostic testing on the residuals. Do the residuals resemble white noise? If not, try to find another ARIMA model which fits better.(5 points)
- Forecast the next 3 years of data. Get the latest figures from <https://fred.stlouisfed.org/categories/11> to check the accuracy of your forecasts. (5 points)
- Eventually, the prediction intervals are so wide that the forecasts are not particularly useful. How many years of forecasts do you think are sufficiently accurate to be usable? (3 points)

Part C (40 points)

Consider following transactions: (8 points)

- Eggs, Bread, Milk, Bananas, Onion, Yogurt
- Dill, Eggs, Bread, Bananas, Onion, Yogurt
- Apple, Eggs, Bread, Milk
- Corn, Bread, Milk, Teddy Bear, Yogurt
- Corn, Eggs, Ice Cream, Bread, Onion

a) Calculate by hand support, confidence and lift for following rules (without usage of apriory library, show your work)

- $\{\text{Bananas}\} \rightarrow \{\text{Yogurt}\}$ (2 points)

`N=`
`N_bananas =`
`N_yogurt =`
`N_bananas_yogurt =`

`support =`
`confidence =`

`support_yogurt =`

`lift =`

- $\{\text{Corn, Bread}\} \rightarrow \{\text{Onion}\}$ (3 points)

- {Bread}->{Milk, Yogurt} (3 points)

Part D (32 points)

Online_Retail2.csv contains transaction from online store in long format (i.e. single item per line and lines with same InvoiceNo is single transaction).

- a) Read data and convert it to transactions (hint: transactions function and format argument). (4 points)
- b) Run summary on transactions. How many transactions are there? How many unique items? (4 points)
- c) Inspect (with inspect) first three transactions. What items are in basket with transaction id 536366? (4 points)
- d) Visualize top 10 frequent items. What is the most frequent? (4 points)
- e) We want to look at rule which would have at least 100 transactions. What support is corresponding to that? (4 points)
- f) Calculate rules with a rule. Use previously calculated support, confidence of 0.9 and maxlen of 4 (we are looking into the rules with up to 4 items). (4 points)
- g) List top 10 by confidence. What is the sense of confidence (explain on example of the top rule)? (4 points)
- h) List top 10 by lift. What is the sense of lift (explain on example of the top rule)? (4 points)