

28Oct2023_Assignment 1.1_GR

Gabi Rivera

2023-10-28

Chapter 1: Approaching Forecasting (Page 23) #1-5

Impact of September 11 on Air Travel in the United States: The Research and Innovative Technology Administration's Bureau of Transportation Statistics (BTS) conducted a study to evaluate the impact of the September 11, 2001, terrorist attack on U.S. transportation. The study report and the data can be found at www.bts.gov/publications/estimated_impacts_of_9_11_on_us_travel. The goal of the study was stated as follows:

The purpose of this study is to provide a greater understanding of the passenger travel behavior patterns of persons making long distance trips before and after September 11.

The report analyzes monthly passenger movement data between January 1990 and April 2004. Data on three monthly time series are given in the file Sept11Travel.xls for this period:

- (1) actual airline revenue passenger miles (Air),
- (2) rail passenger miles (Rail), and
- (3) vehicle miles traveled (Auto).

In order to assess the impact of September 11, BTS took the following approach: Using data before September 11, it forecasted future data (under the assumption of no terrorist attack). Then, BTS compared the forecasted series with the actual data to assess the impact of the event.

#1 Is the goal of this study descriptive or predictive?

The goal of the study is descriptive to see the impact of Sep 11 terrorist attack on passengers long distance travel behavior.

#2 What is the forecast horizon to consider in this task? Are next-month forecasts sufficient?

The forecast horizon is from Sep 2001 to the last record which is Apr 2004. But it also possible that only half of the that range of time is sufficient to view the effects of the terrorist attack on long distance travel behavior. Although, forecasting next-month is too abrupt to see if there is a decreasing trend and also a pick up after sometime.

#3 What level of automation does this forecasting task require? Consider the four questions related to automation.

Automation is not necessary in this case because the forecasting is a one time event. Historic data will be used to forecast air RPM, Rail PM, and VMT spanning from Sep 2001 to Apr 2004 with the assumption that there had been no terrorist attack. The result is said to be compared against the actual data from that range of time to assess the effect of the event. This hints that the intent of this task is for one time descriptive report.

#4 What is the meaning of $t = 1, 2, 3$ in the Air series? Which time period does $t = 1$ refer to?

$t = 1, 2, 3$ in the air series denotes to the time period per month. So for $t = 1$, this refers to the airline RPM during entire month of Jan 1990.

#5. What are the values for y_1, y_2 , and y_3 in the Air series?

y1 = 35,153,577 RPM y2 = 454,115,779 PM y3 = 163.28 VMT

Chapter 2: Time Series Data (Page 43) #3

Shipments of Household Appliances: The file ApplianceShipments.xls contains the series of quarterly shipments (in millions of USD) of U.S. household appliances between 1985-1989.(6)

#(a) Create a well-formatted time plot of the data.

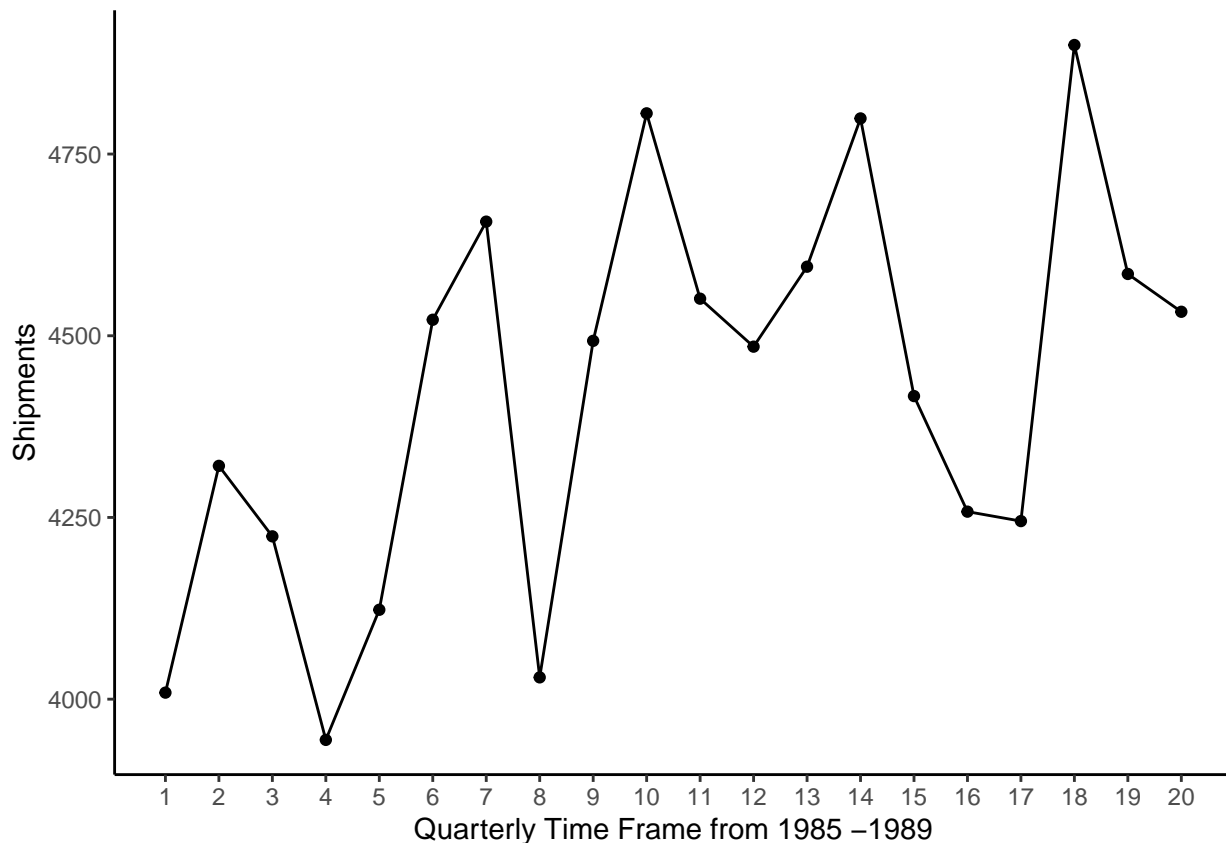
#Import Dataset:

```
df <- read_excel('ApplianceShipments.xlsx')
head(df)
```

```
## # A tibble: 6 x 4
##   Quarter Shipments   Qtr     t
##   <chr>      <dbl> <dbl> <dbl>
## 1 Q1-1985      4009     1     1
## 2 Q2-1985      4321     2     2
## 3 Q3-1985      4224     3     3
## 4 Q4-1985      3944     4     4
## 5 Q1-1986      4123     1     5
## 6 Q2-1986      4522     2     6
```

#Plot Months

```
ggplot(df, aes(t, Shipments)) +
  xlab('Quarterly Time Frame from 1985 -1989') + ylab('Shipments') +
  geom_point() +
  scale_x_continuous(n.breaks=20) +
  geom_line() +
  theme_classic()
```



#(b) Which of the four components (level, trend, seasonality, noise) seem to be present in this series?

There seems to have an upward trend with a cyclical pattern having curve lines going up and down every full quarter.

Performance Evaluation (Pages 67-68) #2 and 3

Forecasting Shampoo Sales: The file ShampooSales.xls contains data on the monthly sales of a certain shampoo over a three year period.

If the goal is forecasting sales in future months, which of the following steps should be taken? (choose one or more)

- partition the data into training and validation periods
- examine time plots of the series and of model forecasts only for the training period
- look at MAPE and RMSE values for the training period
- look at MAPE and RMSE values for the validation period
- compute naive forecasts

The following steps were chosen:

- partition the data into training and validation periods - an important preliminary step for forecasting methods. Forecast methods can be trained and later on assess for predictive performance using the validation results.
- look at MAPE and RMSE values for the validation period - measuring the predictive accuracy of the forecasting model during validation is essential to understand how well the performance is. Using validation data is preferable because the records are not used to select predictors or model parameters.
- compute naive forecasts - showing the most recent information is more likely one of the key forecasts that can be looked at as the baseline data for comparative assessment.

Performance Evaluation (Pages 67-68) #2 and 3

Performance on Training and Validation Data: Two different models were fit to the same time series. The first 100 time periods were used for the training period and the last 12 periods were treated as a validation period. Assume that both models make sense practically and fit the data reasonably well. Below are the RMSE values for each of the models:

	Training Period	Validation Period
Model A	543	690
Model B	669	675

Figure 1: A caption

- (a) Which model appears more useful for retrospectively describing the different components of this time series? Why?

For the training period, it seems that model A is performing better with lower values indicating better accuracy. It's the opposite with validation period, but the scores are only slightly different from each other with Model B having relatively better accuracy. But for overall, it might be more useful to use Model A with the lowest and presumably better accuracy to describe the components of the time series.

- (b) Which model appears to be more useful for forecasting purposes? Why?

For forecasting, I would look at the validation period instead of the training period. Model B with the lowest value and higher accuracy appears to be the more useful model for forecasting in this case. Using Model B, the reliability of the forecast from comparing the actual against the predictions is accurate.