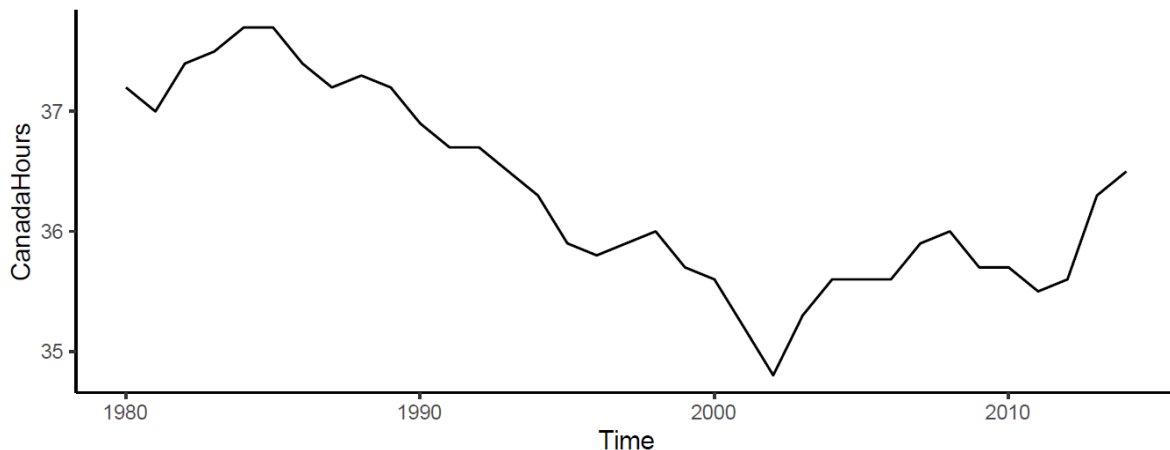


ADS 506 Module 4 Exercises: Chapter 7

This assignment is due on Day 7 of the learning week. The assignment for this module is a mixture of programming and written work. Complete this entire assignment in R Markdown. You will need to include the question and number that you are answering within your submitted assignment. **Once completed, you will knit your deliverable to a Word/PDF file.**

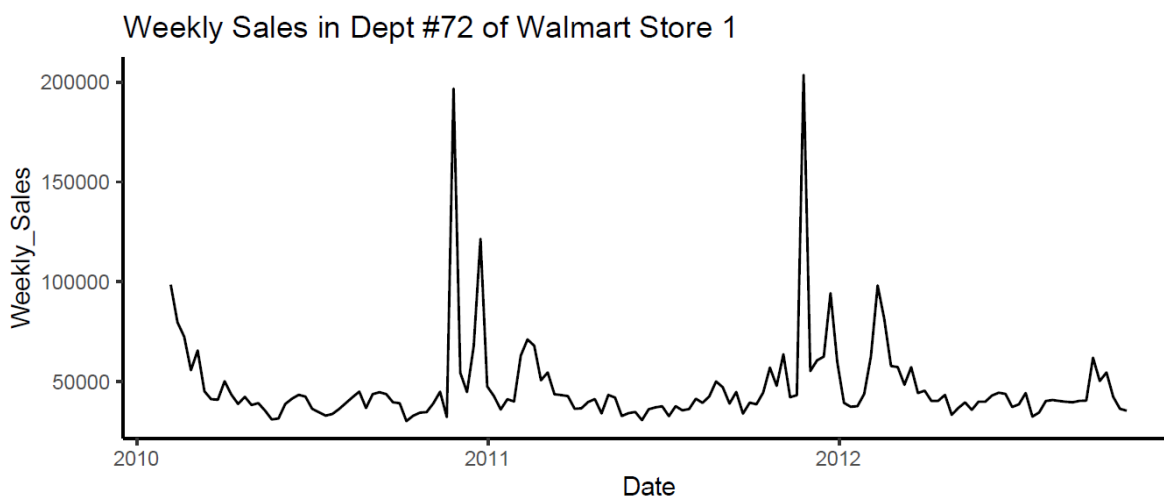
Chapter 7: Regression Models: Autocorrelation & External Info (Pages 170-178): #1, 2, & 6

1. *Analysis of Canadian Manufacturing Workers Work-Hours:* The time series plot in Figure 7.7 describes the average annual number of weekly hours spent by Canadian manufacturing workers. The data is available in CanadianWorkHours.csv.



- a. If we computed the autocorrelation of this series, would the lag-1 autocorrelation exhibit negative, positive, or no autocorrelation? How can you see this from the plot?
 - b. Compute the autocorrelation and produce an ACF plot. Verify your answer to the previous question.
2. *Forecasting Walmart Stock:* Figure 7.10 shows a time plot of Wal-Mart daily closing prices between February 2001 and February 2002. The data is available at finance.yahoo.com and in WalmartStock.csv. The ACF plots of these daily closing prices and its lag-1 differenced series are in Figure 7.11. Table 7.4 shows the output from fitting an AR(1) model to the series of closing prices and to the series of differences. Use all the information to answer the following questions.

- a. Create a time plot of the differenced series.
 - b. Which of the following is/are relevant for testing whether this stock is a random walk?
 - The autocorrelations of the closing price series.
 - The AR(1) slope coefficient for the closing price series.
 - The AR(1) constant coefficient for the closing price series.
 - The autocorrelations of the differenced series.
 - The AR(1) slope coefficient for the differenced series.
 - The AR(1) constant coefficient for the differenced series.
 - c. Recreate the AR(1) model output for the Close price series shown in the left panel of Table 7.4. Does the AR model indicate that this is a random walk? Explain how you reached your conclusion.
 - d. What are the implications of finding that a time series is a random walk? Choose the correct statement(s) below:
 - It is impossible to obtain useful forecasts of the series.
 - The series is random.
 - The changes in the series from one period to the other are random.
6. *Forecasting Weekly Sales at Walmart:* The data in WalmartStore1Dept72.csv is a subset from a larger datasets on weekly department-wise sales at 45 Walmart stores, which were released by Walmart as part of a hiring contest hosted on kaggle.com. The file includes data on a single department at one specific store.



The fields include:

- Date - the week

- Weekly_Sales - sales for the given department in the given store
- IsHoliday - whether the week is a special holiday week
- Temperature - average temperature in the region
- Fuel_Price - cost of fuel in the region
- Markdown1-5 - anonymized data related to promotional markdowns that Walmart is running. Markdown data is only available after Nov 2011, and is not available for all stores all the time.
- CPI - the consumer price index
- Unemployment - the unemployment rate

Figure 7.15 shows a time plot of weekly sales in this department. We are interested in creating a forecasting model for weekly sales for the next 26 weeks.

- a. Recreate the time plot of the weekly sales data. Which systematic patterns appear in this series?
- b. Create time plots of the other numerical series (Temperature, Fuel_Price, CPI, and Unemployment). Also create scatter plots of the sales series against each of these four series (each point in the scatter plot will be a week). From the charts, which of the four series would potentially be useful as external predictors in a regression model for forecasting sales?

The following questions are not in your textbook. You will need to also complete these programming questions in R Markdown.

- c. Fit an ARIMA model with 1 lag and external predictors for Weekly_Sales that treats Nov 4, 2011 to Oct 26, 2012 as the training period, and the next 26 weeks as the validation period. Compute the RMSE for the training period.
- d. Create a mean forecasts for the validation period. Create a time plot of these forecasts and a plot of the forecast errors series. Compute the RMSE for the training period.
- e. Compare the performance of the ARIMA model to the mean forecasts. Which one performs better?
- f. Plot the ARIMA model forecasted values. Use WalmartStore1Dept72_validation.csv for your regression model data.