ISA 444: Business Forecasting

13 - Autocorrelation and Seasonality

Fadel M. Megahed

Associate Professor
Department of Information Systems and Analytics
Farmer School of Business
Miami University
Email: fmegahed@miamioh.edu
Office Hours: Click here to schedule an appointment

Fall 2020

Outline

Quick Refresher on What we Covered in Previous Chapter 04

Main Learning Outcomes Discussed in Chapter 04

☑ **Explain when to use an additive vs. multiplicative model for a time series.**

☑ **Use classic decomposition methods to detrend and deseasonalize a time series.**

☑ **Use Holt-Winters method to forecast a time series with a seasonal component.**

☑ **Evaluate the application of different smoothing methods applied to a time series, and determine the best performing method.**

A Note on the Previous Assignment

I had a typo in the previous assignment that led to some confusion. Therefore, in class, we will go through constructing both Figures 4.2 and 4.3 using R (based on the Walmart_2.xlsx Dataset).

Learning Objectives for Today's Class

Main Learning Outcomes

- Define the population mean, and variance of a random variable.
- Define the population covariance, and correlation between two random variables.
- Define the population autocovariance and autocorrelation of a random variable.
- Use sample estimates of the population mean, variance, covariance, and correlation.
- Explain the properties of the large sample distribution of the sample ACF.
- Use the large sample distribution of the sample ACF to identify significant autocorrelation in a time series.
- Determine if a sample ACF plot "cuts off" or "dies down".

Outline

Definition and Notation

A random variable, Y , is the outcome of a random experiment. The random nature of Y can occur through a variety of mechanisms including sampling, natural variation, etc . In time series, we write Y_t to represent the random variable at time t , where $t = 1, 2, 3, \ldots$.

Specific observed values of a random variable are written as lower case letters, y_t.

Example to demonstrate the notation

```
pacman::p_load(tidyquant)
aapl = tq_get('AAPL', from = "2020-09-28", to    = "2020-10-10") %>%
  pull(adjusted)
```

Y_2 represents the adjusted **but not observed** closing price for the $AAPL stock on Sept. 29, 2020. When we observe a value for this we have, $y_2 = 114.09$.

Basic Population Parameter Functions

Mean Function:

$$\mu_{Y\,t} = \mu_t = E(Y_t). \tag{1}$$

Variance Function:

$$\sigma_t^2 = E[(Y_t - \mu_t)^2]. \tag{2}$$

Covariance Function: The covariance of two random variables,Y and Z is given by

$$E[(Y - \mu_Y)(Z - \mu_Z)]. \tag{3}$$

The covariance measures the *linear dependence* between two random variables.

The Correlation Coefficient between two random variables, Y and Z is given by $\rho = \frac{E[(Y-\mu_Y)(Z-\mu_Z)]}{\sigma_Y \sigma_z}$. It measures the scaled linear dependence between two random variables, and is in the interval $[-1, 1]$.

Outline

Autocovariance Function

In time series applications, often, our best predictor of a future observation is the past values of the series. Thus, we measure the linear dependence of the series over time using the autocovariance (autocorrelation) functions. For the random variable Y observed at two different times, Y_s and Y_t , the autocovariance function is defined as:

$$\gamma(s,t) = cov(Y_s, Y_t) = E[(Y_s - \mu_s)(Y_t - \mu_t)]. \tag{4}$$

Notes:

- $\gamma(s,t) = \gamma(t,s)$.
- If $\gamma(s,t) = 0$, then Y_s and Y_t are **NOT linearly related**.
- $\gamma(t,t) = \sigma_t^2$.

Autocorrelation Function

In applications, we generally use the Autocorrelation Function (ACF):

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}} = \frac{\gamma(s,t)}{\sqrt{\sigma_s^2\sigma_t^2}} \tag{5}$$

Notes:

- The ACF is in the interval $[-1, 1]$.
- The ACF measures the linear predictability of the series at time t using only information from time Y_s.

Non-graded Breakout Room Class Activity

Consider a white noise, centered moving average model, where w_t is distributed *iid* $N(0,1)$ and $Y_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1})$. **Use the Equations in the previous slides and the expected value properties for both the mean and variance to compute:**

- Population Mean: $E(Y_t) =$
- Population Variance: $\sigma^2(Y_t) =$
- Population Autocovariance between times t and $t+1$: $\gamma(t+1, 1) =$
- Population Autocorrelation between times t and $t+1$: $\rho(t+1, 1) =$
- Population Autocorrelation between times t and $t+2$: $\rho(t+2, 1) =$
- Population Autocorrelation between times t and $t+3$: $\rho(t+3, 1) =$

Outline

Definitions [1]

Sample mean:

$$\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t$$

Sample variance:

$$\hat{\sigma}_y^2 = \frac{1}{n-1}\sum_{t=1}^{n}(y_t - \bar{y})^2$$

Standard error of the mean:

$$\hat{\sigma}_{\bar{y}}^2 = \frac{1}{n-1}\sum_{t=1}^{n}(y_t - \bar{y})^2$$

Definitions [2]

Lag k Sample Autocorrelation:

$$r_k = \frac{\sum_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}$$

Comments on the Sample ACF

- The sample ACF is very useful in helping us to determine the degree of autocorrelation in our time series.
- However, the sample ACF is subject to random sampling variability. Like the sample mean, the sample ACF has a sampling distribution.

Outline

Properties [1]

- A common heuristic is that at least 50 observations are needed to give a reliable estimate of the population ACF, and that the sample ACF should be computed up to lag $K = \frac{n}{4}$, where n is the length of the series available for training.
- Under general conditions, for large n, and $k = 1, 2, \ldots$, the ACF follows an approximate normal distribution with zero mean and standard deviation given by $\frac{1}{\sqrt{n}}$.
- This result can be used to give us a cutoff to determine if there is a statistically significant amount of autocorrelation for a given lag in a series.
- `R` uses a cutoff of $\pm 1.96\frac{1}{\sqrt{n}}$ to determine statistical significance of the sample ACF.
  - That is if the sample ACF is **within** $\pm 1.96\frac{1}{\sqrt{n}}$, it is considered **NOT** significant.
  - If the sample ACF is greater than $+1.96\frac{1}{\sqrt{n}}$, then there is significant positive autocorrelation at a particular lag.

Properties [2]

- If the sample ACF is less than $-1.96\frac{1}{\sqrt{n}}$, then there is significant negative autocorrelation at a particular lag.

Example 1: White Noise

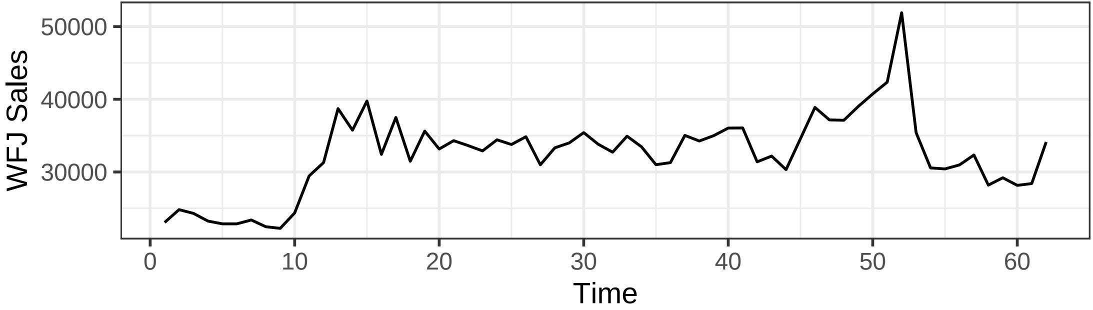
In this live coding session, we will generate the following time-series:

- White Noise
- Centered Moving Average of the White Noise Data

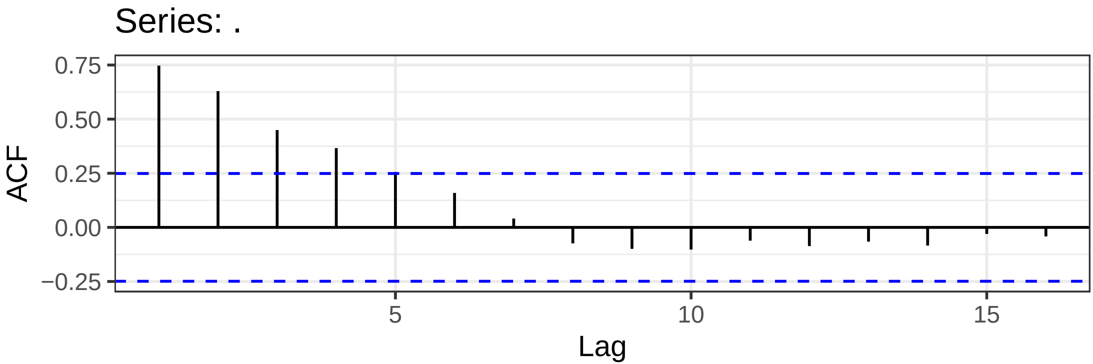
We will plot both time-series as well as their corresponding ACFs. We will also commit on the obtained results.

Example 2: The WFJ Sales Data [1]

In this live-demo example, we will use R to plot the ACF for the WFJ Sales Data. Note that this corresponds to Figure 6.2 in your textbook; however `R` uses constant significance limits.



Example 2: The WFJ Sales Data [2]



Example 2: The WFJ Sales Data [3]

```
## $coefficients
##                  Estimate   Std. Error  t value     Pr(>|t|)
## (Intercept) 8337.7020663 2.682195e+03 3.108537 2.892362e-03
## Lag1           0.7486362 8.150525e-02 9.185128 5.621917e-13

## $adj.r.squared
## [1] 0.5814924
```

Outline

Summary of Main Points

Main Learning Outcomes

- Define the population mean, and variance of a random variable.
- Define the population covariance, and correlation between two random variables.
- Define the population autocovariance and autocorrelation of a random variable.
- Use sample estimates of the population mean, variance, covariance, and correlation.
- Explain the properties of the large sample distribution of the sample ACF.
- Use the large sample distribution of the sample ACF to identify significant autocorrelation in a time series.
- Determine if a sample ACF plot "cuts off" or "dies down".

Things to Do for Next Class

- Thoroughly read Chapter 6.1 of our textbook.
- Go through the slides, examples and make sure you have a good understanding of what we have covered.

ISA 444: Business Forecasting

13 - Autocorrelation and Seasonality

Fadel M. Megahed

Associate Professor
Department of Information Systems and Analytics
Farmer School of Business
Miami University
Email: fmegahed@miamioh.edu
Office Hours: Click here to schedule an appointment

Fall 2020