

Lab 8 Background

Gabriel

3/22/2018

Goals and learning Objectives

The goal of this lab is to investigate the predictability of a few financial objects over time. We begin with an introduction to the **random walk** process and the **autoregressive** model (please see the file **Lab 8 Background** for this information). In this lab, you will estimate the autoregressive statistical parameter ϕ for each year of the SP500 weekly closing prices. You will then perform the same task using closing prices from the Dow Jones Industrial Average. To accomplish these tasks, utilize the **Split/Apply/Combine** strategy using functions from the **plyr** or **apply** family.

The Random Walk and the AR(1) Model

Consider an experiment involving a fair coin where the researcher flips the coin 10 times and records the outcome of each trial. When the face displays a **Heads**, she records **1** and when the face displays a **Tails**, she records **-1**. The following **R** code illustrates this experiment.

```
set.seed(0)
trials <- 10
coins <- ifelse(rbinom(n=trials,size=1,prob=.5)==1,1,-1)
coins
```

```
## [1] 1 -1 -1 1 1 -1 1 1 1 1
```

Construct a random variable called Y_t , which represents the sum of coin flips over the n -trials, i.e.,

$$Y_t = Y_{t-1} + Z_t, \quad t = 1, 2, \dots, n,$$

where

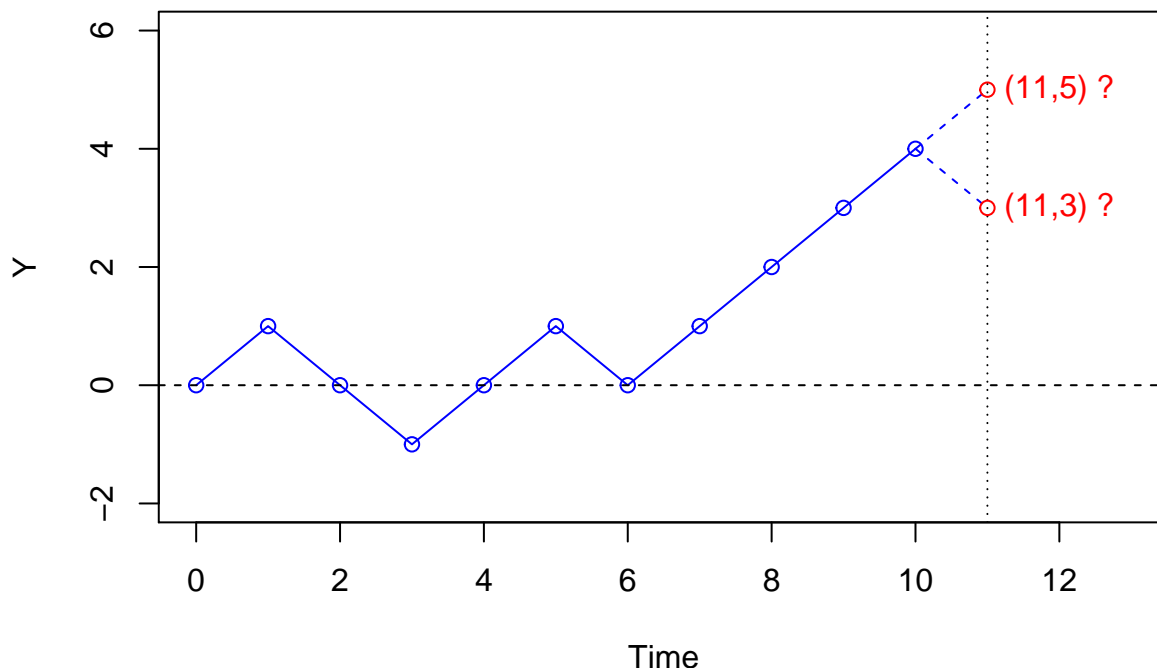
$$Z_t = \begin{cases} 1 & \text{if heads on trial } t \\ -1 & \text{if tails on trial } t \end{cases}$$

and $Y_1 = 0$. The above random variable corresponding to the n^{th} trial, Y_n , is called a **random walk**. The random walk is a time series of random variables whose updated value (Y_{t+1}) is governed by the coin flip (Z_t), at time t . For the random sequence of **coins**, Y_{11} is computed by

$$\begin{aligned} Y_{10} &= Y_9 + Z_{10} \\ &= (Y_8 + Z_9) + Z_{10} \\ &= (Y_7 + Z_8) + (Z_9 + Z_{10}) \\ &\vdots \\ &= Y_0 + (Z_1 + Z_2 + Z_3 + Z_4 + Z_5 + Z_6 + Z_7 + Z_8 + Z_9 + Z_{10}) \\ &= 0 + 1 - 1 - 1 + 1 + 1 - 1 + 1 + 1 + 1 + 1 = 4 \end{aligned}$$

Below is a plot of the time series Y_1, Y_2, \dots, Y_{10} . This graph is constructed by plotting the cumulative sum of coin flips as a function of time.

Time Series Plot: Random Walk (Coin Flips)



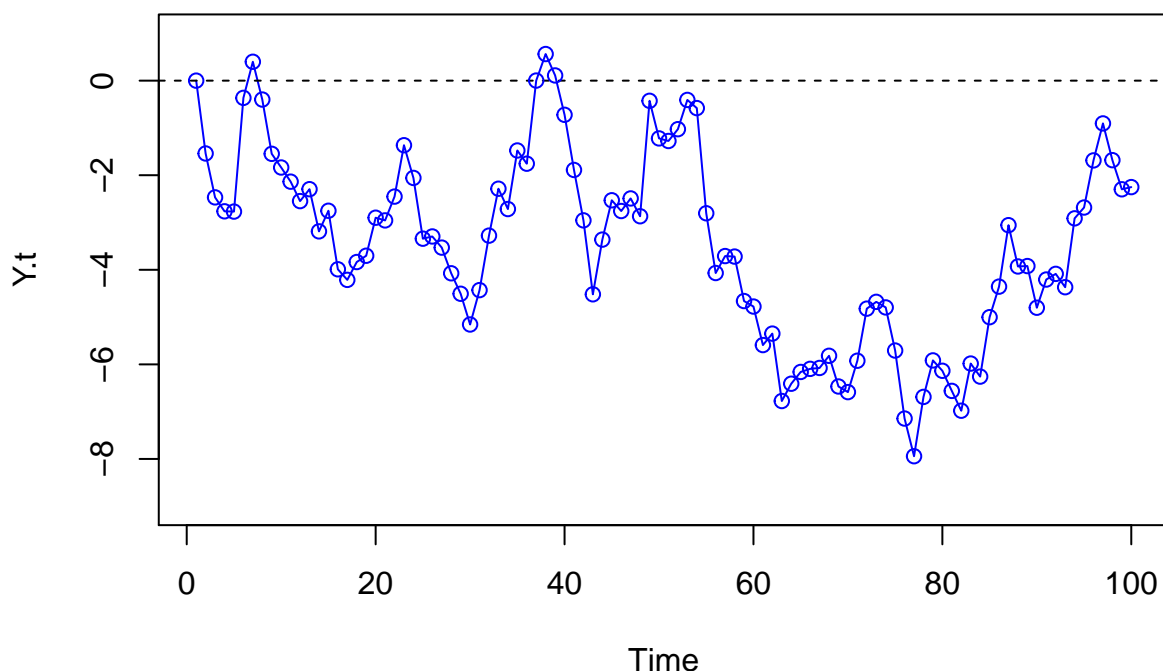
Further suppose that the researcher would like to predict the $(n+1)^{th}$ trial of the random walk, i.e., is Y_{11} going to take on the value 3 or 5? Is this forecast of Y_{11} possible? Obviously forecasting Y_{11} is not feasible because you cannot predict the future outcome of a coin flip.

Below is a more general expression of the random walk model that assigns a continuous error structure on Z_t . This model is given by

$$Y_t = Y_{t-1} + Z_t, \quad Z_t \stackrel{iid}{\sim} N(0, \sigma^2). \quad (1)$$

The error structure can further be generalized to any continuous (or discrete) distribution but the normal distribution is a natural choice. The plot displayed below shows a realization of the random model with $\sigma^2 = 1$ and time steps $t = 1, 2, \dots, 100$, i.e., sample size $n = 100$.

Time Series Plot: Random Walk



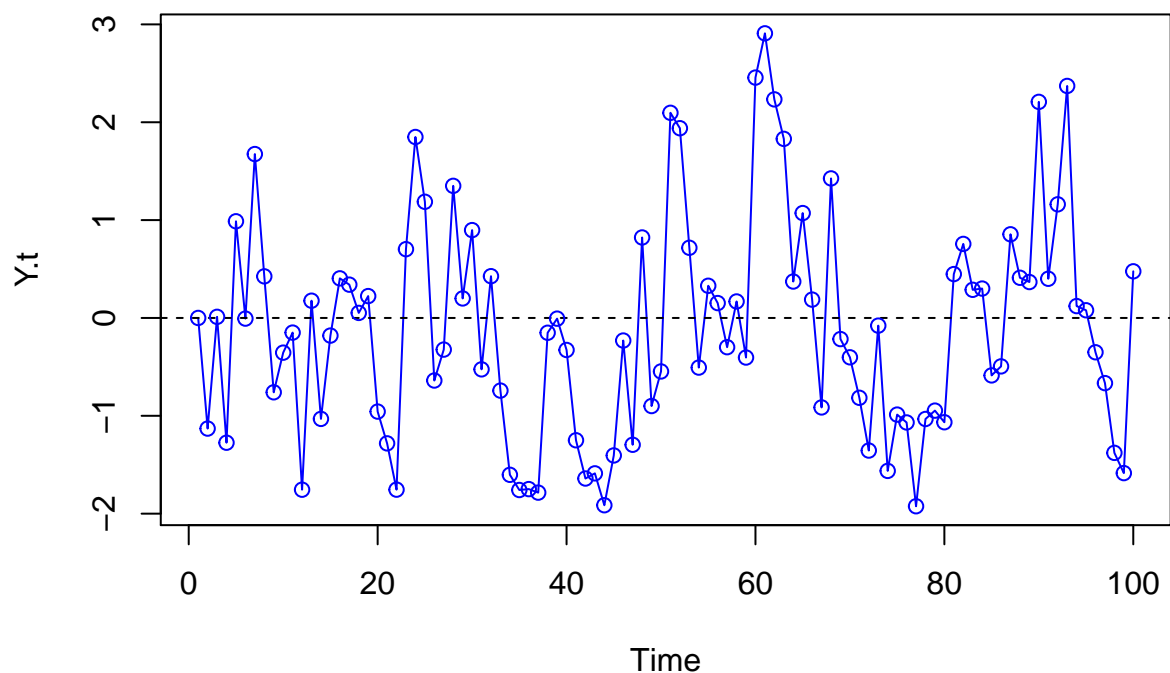
The Zero Mean Autoregressive Model

Recall when asking the question “*is the random walk time series predictable?*”; we already know the answer is no. Now consider the question “*when is a time series predictable?*” To answer this question, we should include past information to successfully predict the future case Y_{n+1} . The basic or foundational time-series model is the mean zero autoregressive lag-1 process, given by the formula

$$Y_t = \phi Y_{t-1} + Z_t, \quad Z_t \stackrel{iid}{\sim} N(0, \sigma^2). \quad (2)$$

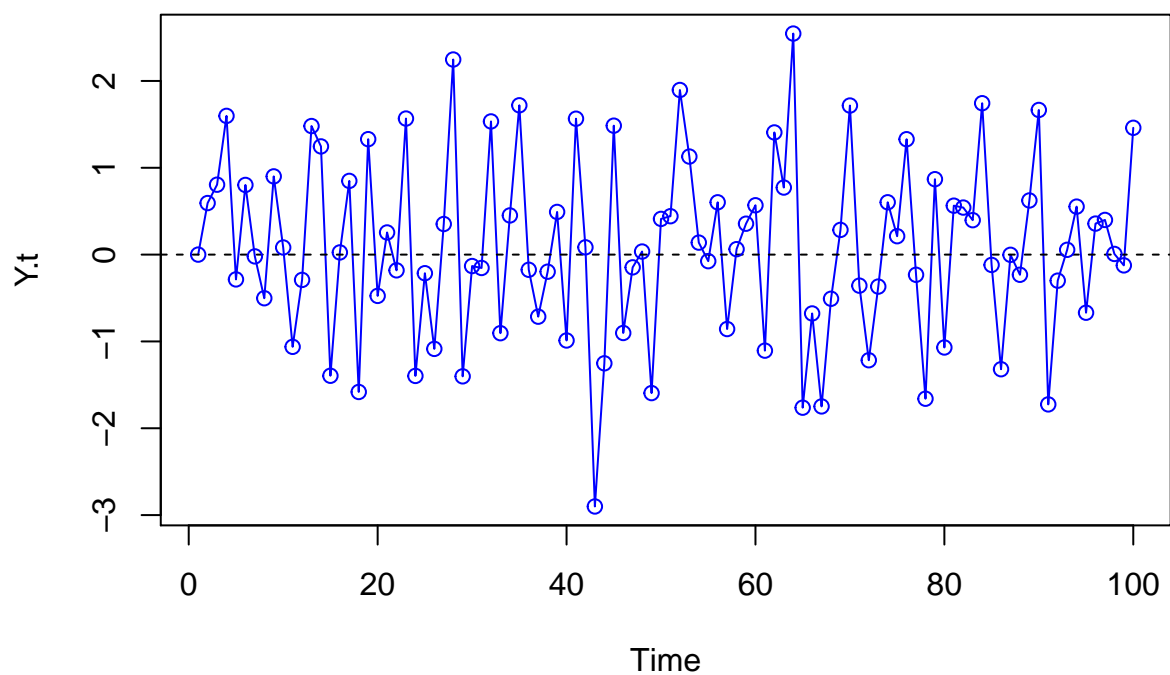
In the above model, the case Y_t depends on the previous case Y_{t-1} , plus some random error Z_t . The random error is assumed normal with zero mean and variance σ^2 , but this error structure is often assumed white noise. The parameter ϕ can take on any real number, but specific ranges of ϕ guarantee better behavior of the process. This autoregressive model only depends on 1 lag and hence is called the AR(1) model. To get a feel for what an AR(1) process looks like, consider the following plot of $n = 100$ AR(1) realization using $\phi = .50$ and $\sigma^2 = 1$.

Time Series Plot: AR(1) Phi=0.50

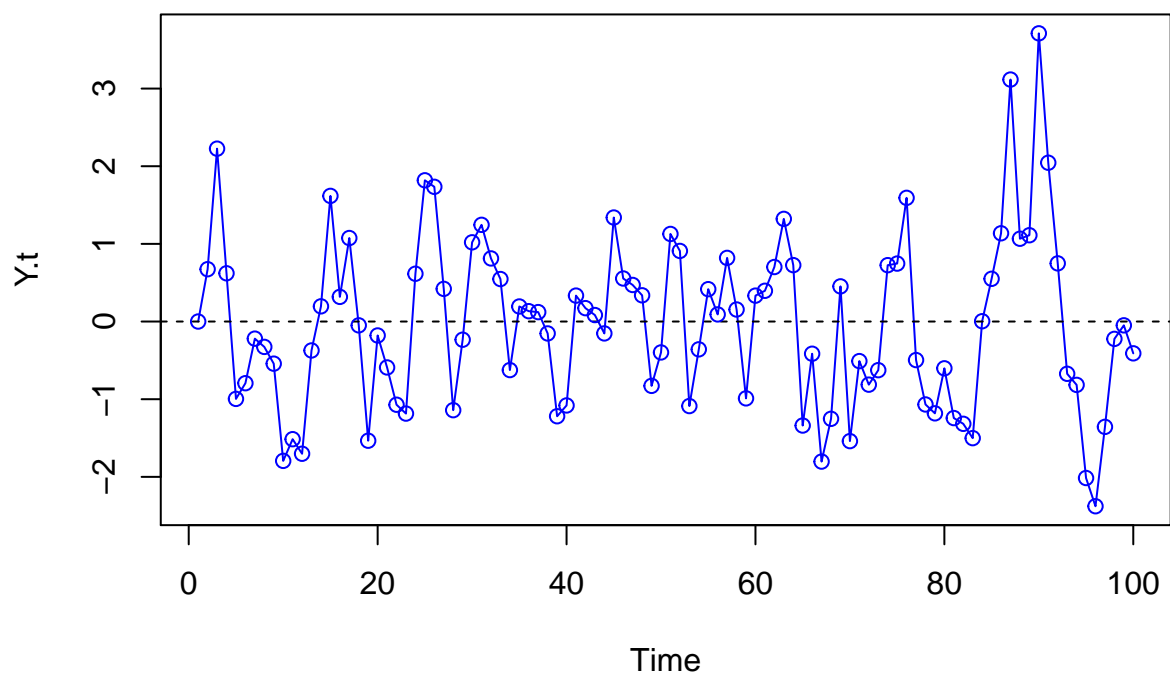


To gain more insight on how the AR(1) process changes with different values of ϕ , consider the following 5 realizations, each of size $n = 100$ and autoregressive parameter ranging from $\phi = .05, .40, .80, .99, 1.05$. Notice that as ϕ ranges from 0 to 1, the AR(1) realizations go from a time series of **white noise** to the **random walk** process. When $\phi > 1$, the process is explosive.

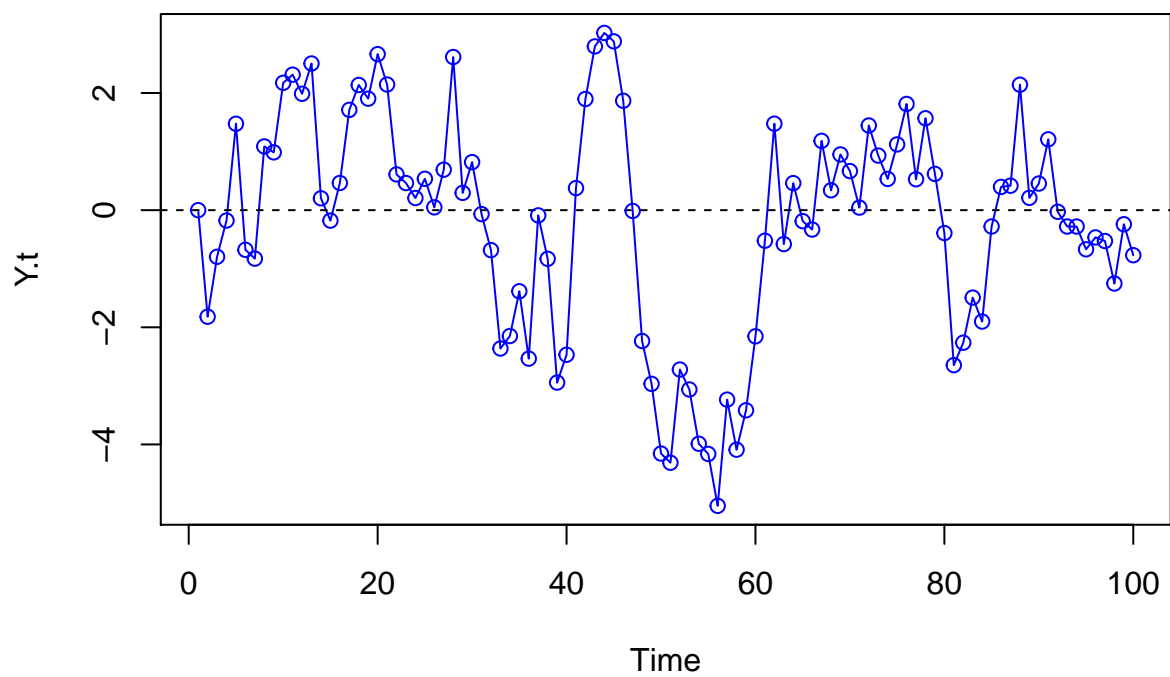
Time Series Plot: AR(1) Phi=0.05



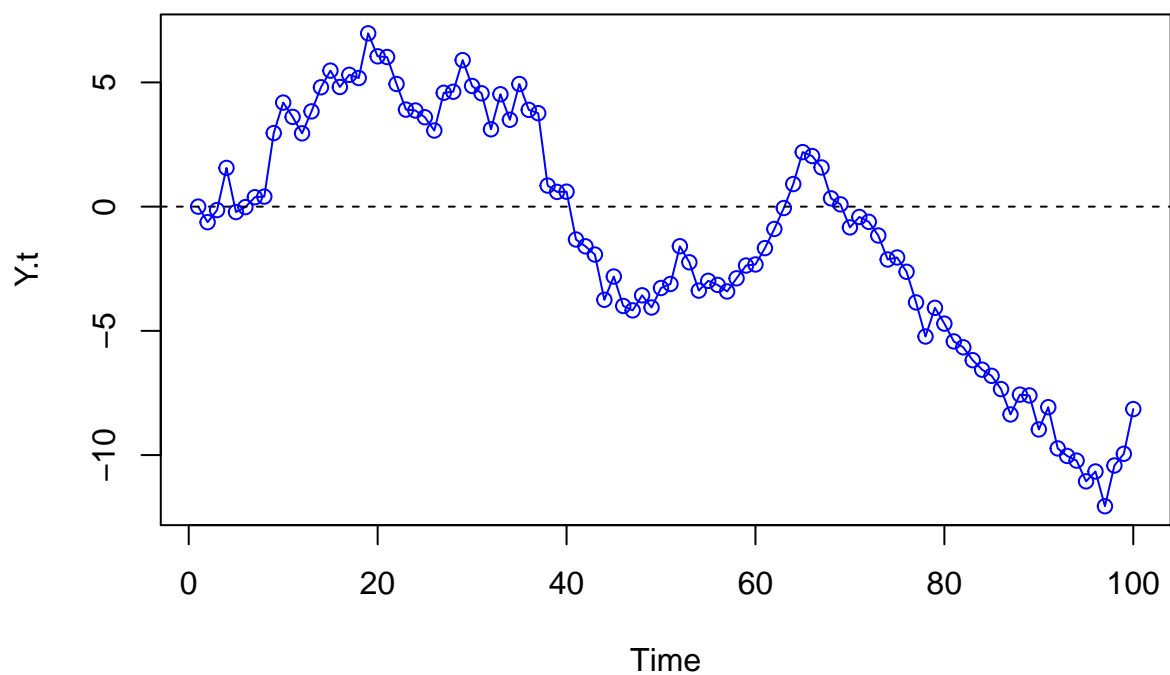
Time Series Plot: AR(1) $\Phi=0.40$



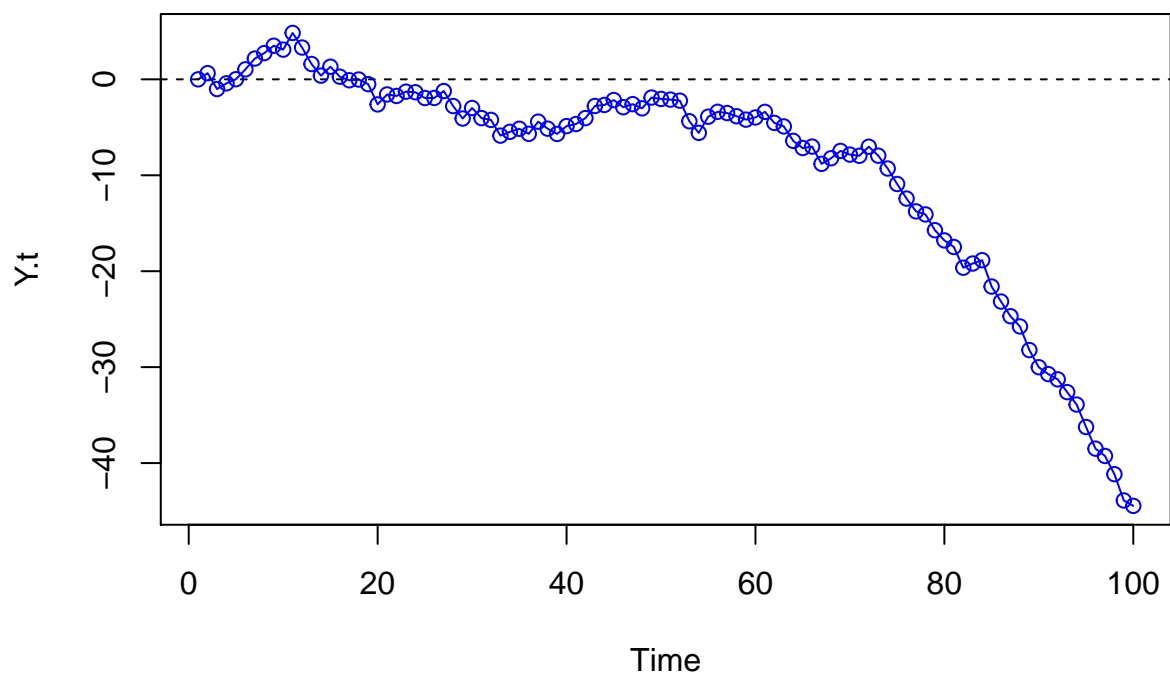
Time Series Plot: AR(1) $\Phi=0.80$



Time Series Plot: AR(1) $\Phi=0.99$



Time Series Plot: AR(1) $\Phi=1.05$



Predictability

To answer the question, “*when is a time series predictable?*”; we can now narrow the scope to the AR(1) process and base predictability on the parameter ϕ . We consider four cases:

- $\phi = 0$: This represents a sequence of white noise. To predict the future case Y_{n+1} , we would use the regular sample mean of the data. This does not incorporate any past information and hence may not be a reliable forecast.
- $|\phi| < 1$: This is the best range for the autoregressive parameter. When forecasting Y_{n+1} , the process uses past information for the time series and also exhibits an important property called stationarity. To learn more about stationarity, take a course in Time Series Analysis.
- $|\phi| = 1$: This represents the random walk process. This is not predictable and not stationary.
- $|\phi| > 1$: This represents an explosive AR(1) process. Although stationarity does exist, this series is difficult and unnatural to model and hence, we typically direct our attention to the case $|\phi| < 1$.

To summarize: The the worst case scenario is when $|\phi| = 1$, which is often refereed to as a unit root. The best case scenario is when $|\phi| < 1$, which indicates a *predictive* time series. The case when $\phi = 0$ is fine but we are not using any past information to model the future cases. The case when $|\phi| > 1$ generates inflated behavior, which is also not desirable.

Interpretation: The closer this estimated ϕ is to 1 gives an indication on how predictable the time series is. If ϕ is exactly 1, then the series is a random walk and hence, is not predictable. Due to random chance, estimated ϕ values will never equal 1. If $\hat{\phi}$ is very close to 1, then the time series is hard to predict.

The Autoregressive Model and Estimation

The more common AR(1) model allows for a drift term μ . The autoregressive lag-1 process, now having two parameters, is given by the formula

$$Y_t = \mu + \phi(Y_{t-1} - \mu) + Z_t, \quad Z_t \stackrel{iid}{\sim} N(0, \sigma^2). \quad (3)$$

The above model can be expressed as

$$Y_t = \mu(1 - \phi) + \phi Y_{t-1} + Z_t, \quad Z_t \stackrel{iid}{\sim} N(0, \sigma^2).$$

To estimate the model parameters we can use common techniques, i.e., least squares.

$$Q(\mu, \phi) = \sum_{t=2}^n (Y_t - (\mu + \phi(Y_{t-1} - \mu)))^2. \quad (4)$$

The minimum value of Q is achieved at the least squares estimators $\hat{\mu}$ and $\hat{\phi}$.

One fun technique to easily estimate ϕ is to line the data set up as displayed below. Consider the following table for $n = 100$ cases.

Y_{t-1}	Y_t
x_1	x_2
x_2	x_3
x_3	x_4
x_4	x_5
x_5	x_6
\vdots	\vdots
x_{98}	x_{99}
x_{99}	x_{100}
x_{100}	

The first column of the dataframe is the first $n - 1$ cases and the second column are cases 2 through n . After constructing the two columns, you can use ordinary linear regression techniques to estimate ϕ , i.e., regress Y_t on the variable Y_{t-1} . The estimated slope $\hat{\beta}_1$ is thus the estimated AR(1) coefficient $\hat{\phi}$. Note that this new dataframe only has $n - 1$ cases.