3. Stationarity, white noise, and some basic time series models

Edward Ionides

1/9/2018

Contents

Objectives

1. Define different concepts for stationarity.
2. Define white noise.
3. Use white noise to construct some basic time series models.

Definition: Weak stationarity and strict stationarity

- A time series model which is both mean stationary and covariance stationary is called **weakly stationary**.

  Typically we call "weakly stationary" just "stationary", and say "strictly stationary" otherwise.

- A time series model for which all joint distributions are invariant to shifts in time is called **strictly stationary**.
  - Formally, this means that for any collection of times $(t_1, t_2, \ldots, t_K)$, the joint distribution of observations at these times should be the same as the joint distribution at $(t_1+\tau, t_2+\tau, \ldots, t_K+\tau)$ for any τ.
  - For equally spaced observations, this becomes: for any collection of timepoints $n_1, \ldots, n_K$, and for any lag h, the joint density function of $(Y_{n_1}, Y_{n_2}, \ldots, Y_{n_K})$ is the same as the joint density function of $(Y_{n_1+h}, Y_{n_2+h}, \ldots, Y_{n_K+h})$.
  - In our general notation for densities, this strict stationarity requirement can be written as

$$f_{Y_{n_1},Y_{n_2},\ldots,Y_{n_K}}(y_1, y_2, \ldots, y_K) \quad (1)$$
$$= f_{Y_{n_1+h},Y_{n_2+h},\ldots,Y_{n_K+h}}(y_1, y_2, \ldots, y_K). \quad (2)$$

- Strict stationarity implies weak stationarity (check this). Note that we only defined weak stationarity for equally spaced observations.

Question: How could we assess whether a weak stationary model is appropriate for a time series dataset?

Suppose observations are equally spaced scross time. Can calculate averages in chunks (i.e. time windows) for mean-stationary. Can calculate sample ACF in time windows. Perhaps the sample ACF decays quickly to zero, that would support **weakly stationary**, since a trend typically shows up as a **long range** pattern in the sample ACF.

Question: How could we assess whether a strictly stationary model is appropriate for a time series dataset?

We essentially can't.

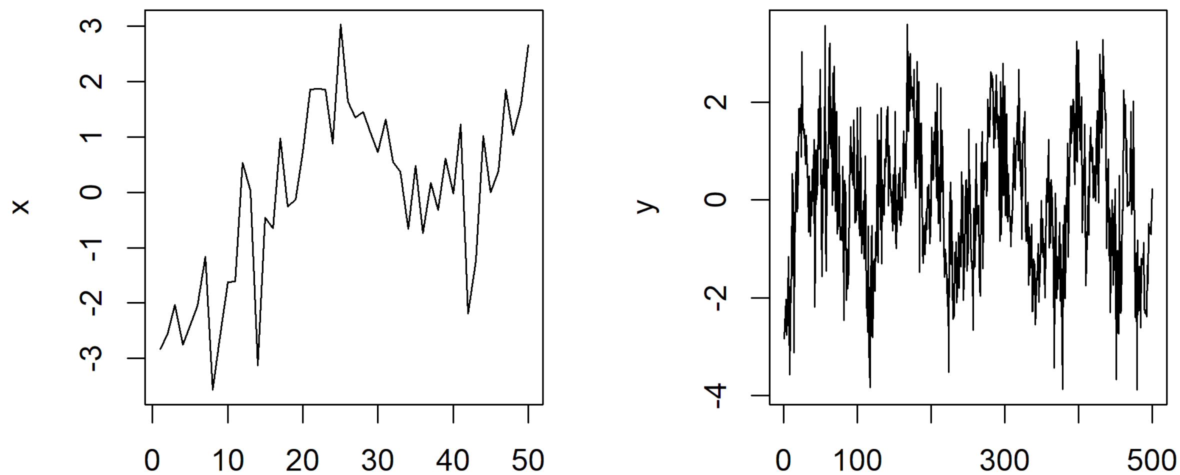
Question: Is it usual for time series to be well modeled as stationary (either weakly or strictly)?

ANSWER: It sometimes happens. However, systems often change over time, and that may be one of the things we are interested in.

Question: If data do not often show stationary behavior, why do many fundamental models have stationarity?

Stationary models are building blocks for non-stationary models.

Question: Is a stationary model appropriate for either (or both) the time series below? Explain.



Here, $X = Y[1:50]$, so we really need the entire data set to be able to assess stationarity.

White noise

A time series model $\epsilon_{1:N}$ which is weakly stationary with

$$\begin{aligned} \mathbb{E}[\epsilon_n] &= 0 && (3)\\ \mathrm{Cov}(\epsilon_m, \epsilon_n) &= \begin{cases} \sigma^2, & \text{if } m = n \\ 0, & \text{if } m \neq n \end{cases}, && (4) \end{aligned}$$

is said to be **white noise** with variance σ^2.

- The "noise" is because there's no pattern, just random variation. If you listened to a realization of white noise as an audio file, you would hear a static sound.
- The "white" is because all freqencies are equally represented. This will become clear when we do frequency domain analysis of time series.
- Signal processing—sending and receiving signals on noisy channels—was a motivation for early time series analysis.

Example: Gaussian white noise

In time series analysis, a sequence of independent identically distributed (IID) Normal random variables with mean zero and variance σ^2 is known as **Gaussian white noise**. We write this model as

$$\epsilon_{1:N} \sim \text{IID } N[0, \sigma^2].$$

Example: Binary white noise

Let $\epsilon_{1:N}$ be IID with

$$\epsilon_n = \begin{cases} 1, & \text{with probability } 1/2 \\ -1, & \text{with probability } 1/2 \end{cases}. \tag{5}$$

We can check that $\mathbb{E}[\epsilon_n] = 0$, $\text{Var}(\epsilon_n) = 1$ and $\text{Cov}(\epsilon_m, \epsilon_n) = 0$ for $m \neq n$. Therefore, $\epsilon_{1:N}$ is white noise.

Similarly, for any $p \in (0, 1)$, we could have

$$\epsilon_n = \begin{cases} (1-p)/p, & \text{with probability } p \\ -1, & \text{with probability } 1-p \end{cases}. \tag{6}$$

Example: Sinusoidal white noise

Let $\epsilon_n = \sin(2\pi n U)$, with a single draw $U \sim \text{Uniform}[0, 1]$ determining the time series model for all $n \in 1 : N$.

A. Show that $\epsilon_{1:N}$ is weakly stationary, and is white noise!

ANS. We have $\mathbb{E}(\epsilon_n) = \int_0^1 \sin(2\pi n u)\, du = \left[-\frac{1}{2\pi n}\cos(2\pi n u)\right]_0^1 = 0$. Then, for $m \neq n$,

$$\begin{aligned} \text{Cov}(\epsilon_m, \epsilon_n) = \mathbb{E}(\epsilon_m \epsilon_n) &= \int_0^1 \sin(2\pi m u)\sin(2\pi n u)\, du \\ &= \frac{1}{2}\int_0^1 \cos(2\pi(m-n)u) - \cos(2\pi(m+n)u)\, du \\ &= 0. \end{aligned}$$

Now, for $m = n$,

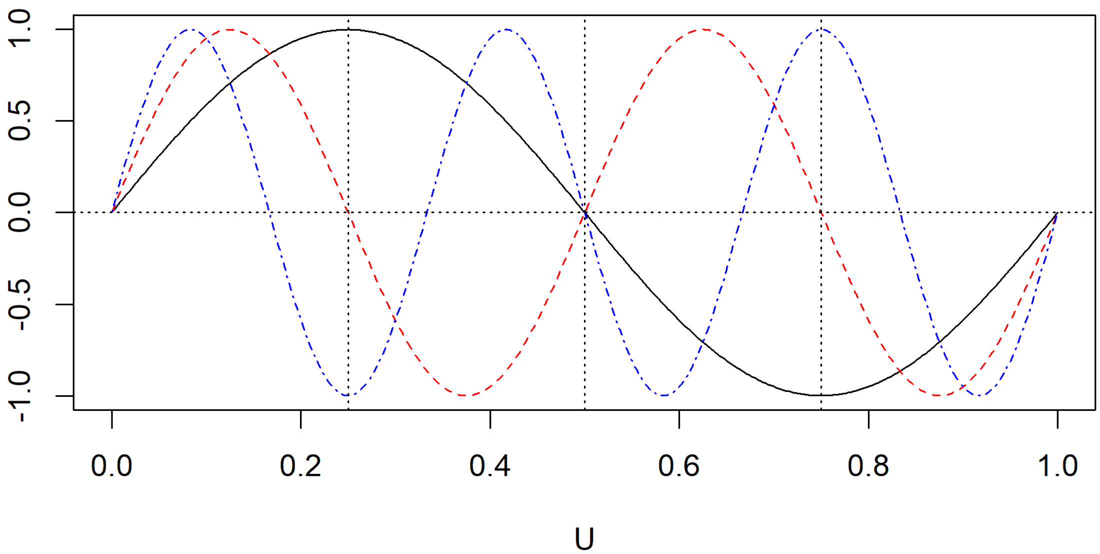
$$\text{Cov}(\epsilon_m, \epsilon_n) = \int_0^1 \sin^2(2\pi n u)\, du = \frac{1}{2}\int_0^1 \left[1 - \cos(4\pi n u)\right] du = \frac{1}{2}.$$

B. Show that $\epsilon_{1:N}$ is NOT strictly stationary.

ANS. We have $\mathbb{P}(\epsilon_1 > 1-\delta,\, \epsilon_2 \in [-\delta, \delta]) > 0$ but $\mathbb{P}(\epsilon_2 > 1-\delta,\, \epsilon_3 \in [-\delta, \delta]) = 0$. Thus, the model is not invariant to time lags.

- These are exercises in working with sines and cosines.

- As well as providing a concrete example of a weakly stationary time series that is not strictly stationary, practice working with sines and cosines will come in handy later when we work in the frequency domain.
- As a hint for B, consider the following plot of $\epsilon_{1:3}$ as a function of U. ϵ_1 is shown as the black solid line; ϵ_2 is the red dashed line; ϵ_3 is the blue dot-dash line.



- Now we're going to use white noise as a building block for various other time series models.

Reminder: why do we need time series models?

- All statistical tests (i.e., whenever we use data to answer a question) rely on having a model for the data. The model is sometimes called the **assumptions** for the test.
- If our model is wrong, then any conclusions drawn from it may be wrong. Our error could be small and insignificant, or disastrous.
- Time series data collected close in time are often more similar than a model with IID variation would predict. We need models that have this property, and we must work out how to test interesting hypotheses for these models.

Autoregressive models

The AR(p) autoregressive model

- The order p autoregressive model, abbreviated to AR(p), is

$$[M1] \qquad Y_n = \phi_1 Y_{n-1} + \phi_2 Y_{n-2} + \dots + \phi_p Y_{n-p} + \epsilon_n,$$

  where $\{\epsilon_n\}$ is a white noise process.

- Often, we consider the **Gaussian AR(p)** model, where $\{\epsilon_n\}$ is a Gaussian white noise process.
- M1 is a **stochastic difference equation**. It is a difference equation (also known as a recurrence relation) since each time point is specified recursively in terms of previous time points. Stochastic just means random.
- To complete the model, we need to **initialize** the solution to the stochastic difference equation. Supposing we want to specify a distribution for $Y_{1:N}$, we have some choices in how to set up the **initial values**.
  1. We can specify $Y_{1:p}$ explicitly, to get the recursion started.
  2. We can specify $Y_{1-p:0}$ explicitly.
  3. For either of these choices, we can define these initial values either to be additional parameters in the model (i.e., not random) or to be specified random variables.
  4. If we want our model is strictly stationary, we must initialize so that $Y_{1:p}$ have the proper joint distribution for this stationary model.
- Let's investigate a particular Gaussian AR(1) process, as an exercise.

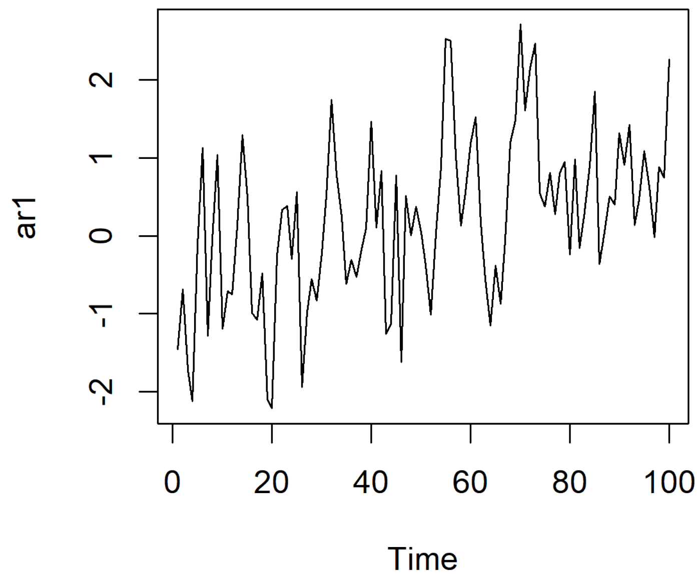
$$[M2] \qquad Y_n = 0.6 Y_{n-1} + \epsilon_n,$$

  where $\epsilon_n \sim \mathrm{IID} N[0,1]$. We will initialize with $Y_1 \sim N[0, 1.56^2]$.

Simulating an autoregressive model

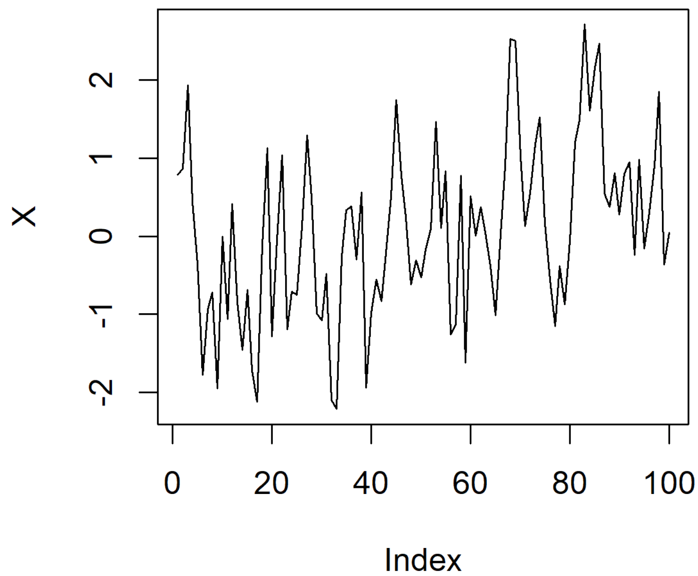
- Looking at simulated sample paths is a good way to get some intuition about a random process model.
- We will do this for the AR(1) model M2
- One approach is to use the `arima.sim` function in R.

```r
set.seed(123456789)
ar1 <- arima.sim(list(ar=0.6),n=100,sd=1)
plot(ar1,type="l")
```



- Does your intuition tell you that these data are evidence for a model with a linear trend?
- The eye looks for patterns in data, and often finds them even when there is no strong statistical evidence.
- That is why we need statistical tests!
- It is easy to see patterns even in white noise. Dependent models produce spurious patterns even more often.
- Play with simulating different models with different seeds to train your intuition.
- A direct approach to simulating model M2 is to write out the model equation explicitly.

```
set.seed(123456789)
N <- 100
X <- numeric(N)
X[1] <- rnorm(1,sd=1.56)
for(n in 2:N) X[n] <- 0.6 * X[n-1] + rnorm(1)
plot(X,type="l")
```



- This looks very similar to the `arima.sim` simulation, except for a difference near the start. Can you explain this? Hint: How does `arima.sim` initialize the simulation?

What are the advantages and disadvantages of `arima.sim` over the direct simulation method?

Exercise: Compute the autcovariance function for model M2.

Moving average models

The MA(q) moving average model

- The order q moving average model, abbreviated to MA(q), is

$$[M3] \qquad Y_n = \epsilon_n + \theta_1 \epsilon_{n-1} + \dots + \theta_q \epsilon_{n-q},$$

where $\{\epsilon_n\}$ is a white noise process.

Moving average models are always stationary.

- To fully specify $Y_{1:N}$ we must specify the joint distribution of $\epsilon_{1-q:N}$.
- Often, we consider the **Gaussian MA(q)** model, where ϵ_n is a Gaussian white noise process.
- Let's investigate a particular Gaussian MA(2) process, as an exercise.

$$[M4] \qquad Y_n = \epsilon_n + 1.5\epsilon_{n-1} + \epsilon_{n-2},$$

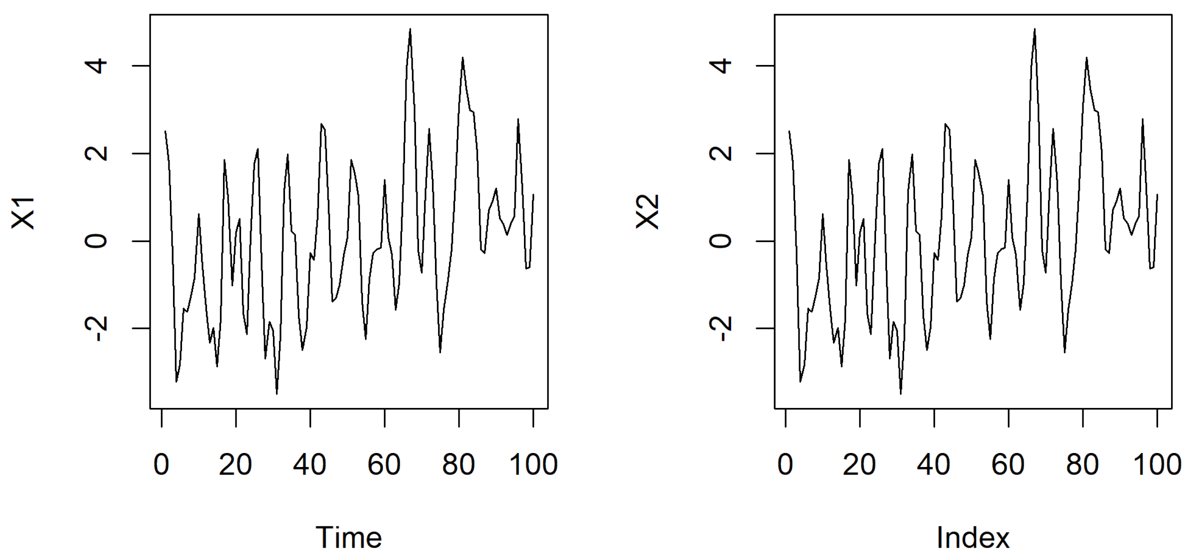
where $\epsilon_n \sim \mathrm{IID} N[0,1]$.

Simulating a moving average model

- Let's simulate M4 using both the methods we tried for the autoregressive model

```
N <- 100
set.seed(123456789)
X1 <- arima.sim(list(ma=c(1.5,1)),n=N,sd=1)
set.seed(123456789)
epsilon <- rnorm(N+2)
X2 <- numeric(N)
for(n in 1:N) X2[n] <- epsilon[n+2]+1.5*epsilon[n+1]+epsilon[n]
oldpars <- par(mfrow=c(1,2))
plot(X1,type="l")
plot(X2,type="l")
```

Argument 'ma=' as uppposed to 'ar=' when we simulate the AR(1) model.




```
par(oldpars)
```

- `X1` and `X2` look identical. We can check this

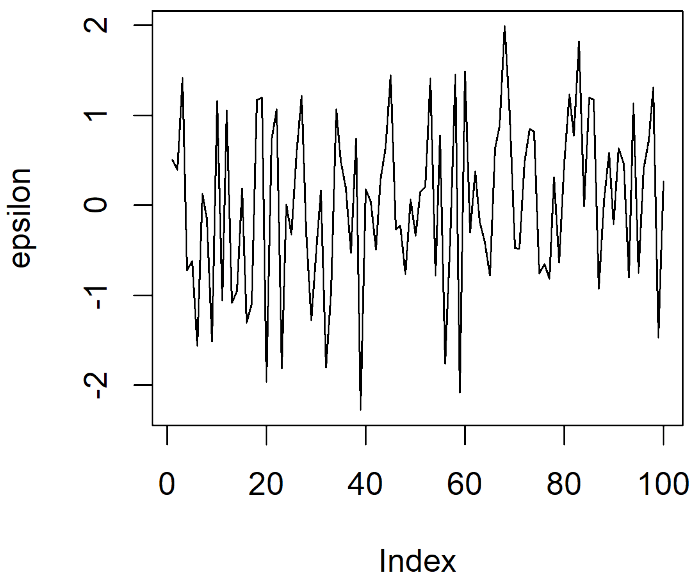
```
all(X1==X2)
```

```
## [1] TRUE
```

- Perhaps you agree that the spurious evidence for a trend that we saw for the AR(1) model is still somewhat present for the MA(2) simulation.

- We could be curious about what the underlying white noise process looks like

```
N <- 100
set.seed(123456789)
epsilon <- rnorm(N)
plot(epsilon,type="l")
```



- To me, the trend-like behavior is not visually apparent in the white noise that "drives" the AR and MA models.

A random walk model

- The **random walk** model is

$$[M5] \qquad Y_n = Y_{n-1} + \epsilon_n$$

, where $\{\epsilon_n\}$ is white noise. Unless otherwise specified, we usually initialize with $Y_0 = 0$.

Mean stationary but not covariance stationary. The variance is increasing with time - thus the covariance function is not only a function of the time "lag" (check first problem of HW2)

- If $\{\epsilon_n\}$ is Gaussian white noise, then we have a Gaussian random walk.

- The random walk model is a special case of AR(1) with $\phi_1 = 1$.

- The stochastic difference equation in M5 has an exact solution,

$$Y_n = \sum_{k=1}^{n} \epsilon_k.$$

Exact solution as long as we initialize $Y_0 = 0$.

- We can also call $Y_{0:N}$ an **integrated white noise process**. We think of summation as a discrete version of integration.
- If data $y^*_{1:N}$ are modeled as a random walk, the value of Y_0 is usually an unknown. Rather than introducing an unknown parameter to our model, we may initialize our model at time t_1 with $Y_1 = y^*_1$.
- The **first difference** time series $z^*_{2:N}$ is defined by

$$z^*_n = \Delta y^*_n = y^*_n - y^*_{n-1}$$

- From a time series of length N, we only get $N-1$ first differences.
- A random walk model for $y^*_{1:N}$ is essentially equivalent to a white noise model for $z^*_{2:N} = \Delta y^*_{2:N}$, apart from the issue of initialization.
- The **random walk with drift** model is given by the difference equation

$$[M6] \qquad Y_n = Y_{n-1} + \mu + \epsilon_n,$$

driven by a white noise process $\{\epsilon_n\}$. This has solution

$$Y_n = Y_0 + n\mu + \sum_{k=1}^{n} \epsilon_k.$$

Here, μ is the "drift" (here, μ isn't actually the expected value of an arbitrary Y).

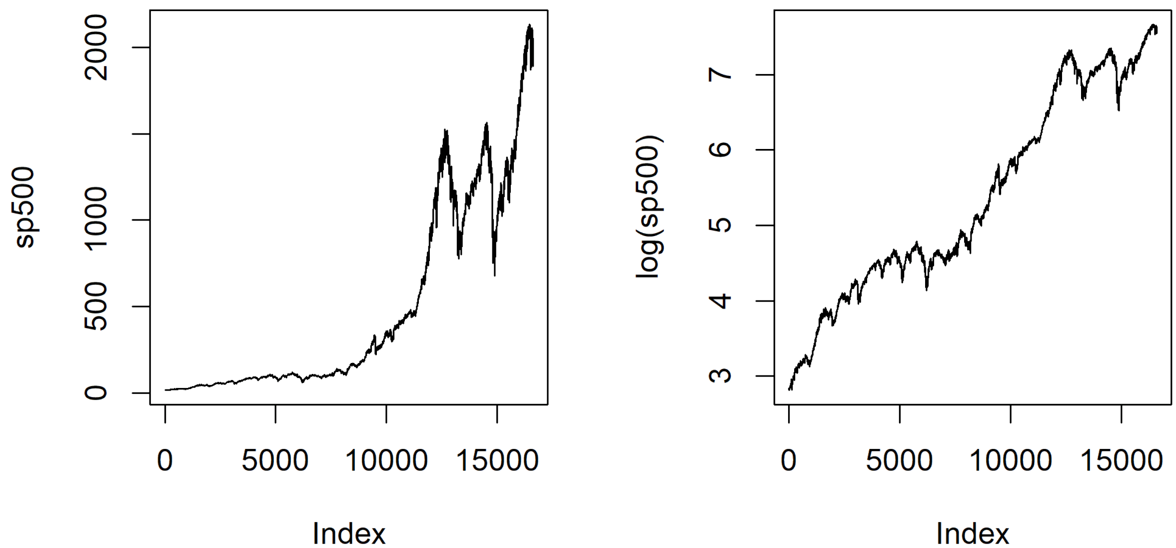
- As for the random walk without drift, we must define Y_0 to initialize the model and complete the model specification. Unless otherwise specified, we usually initialize with $Y_0 = 0$.

Exercise: compute the mean and covariance functions for the random walk model with and without drift.

An example of a random walk: Modeling financial markets

- The theory of efficient financial markets suggests that the logarithm of a stock market index (or, for that matter, the value of an individual stock or other investment) might behave like a random walk with drift
- Let's test that out on daily S&P 500 data, downloaded from yahoo.com.

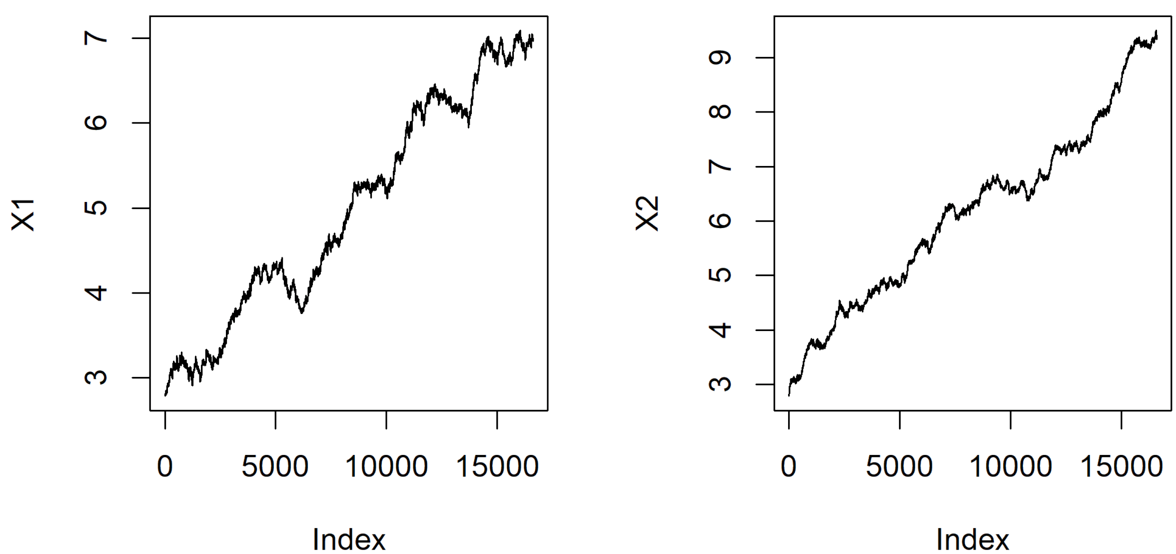
```r
dat <- read.table("sp500.csv",sep=",",header=TRUE)
N <- nrow(dat)
sp500 <- dat$Close[N:1] # data are in reverse order in sp500.csv
par(mfrow=c(1,2))
plot(sp500,type="l")
plot(log(sp500),type="l")
```



- To train our intuition, we can compare the data with simulations from a fitted model. A simple starting point is a Gaussian random walk with drift, having parameters estimated by

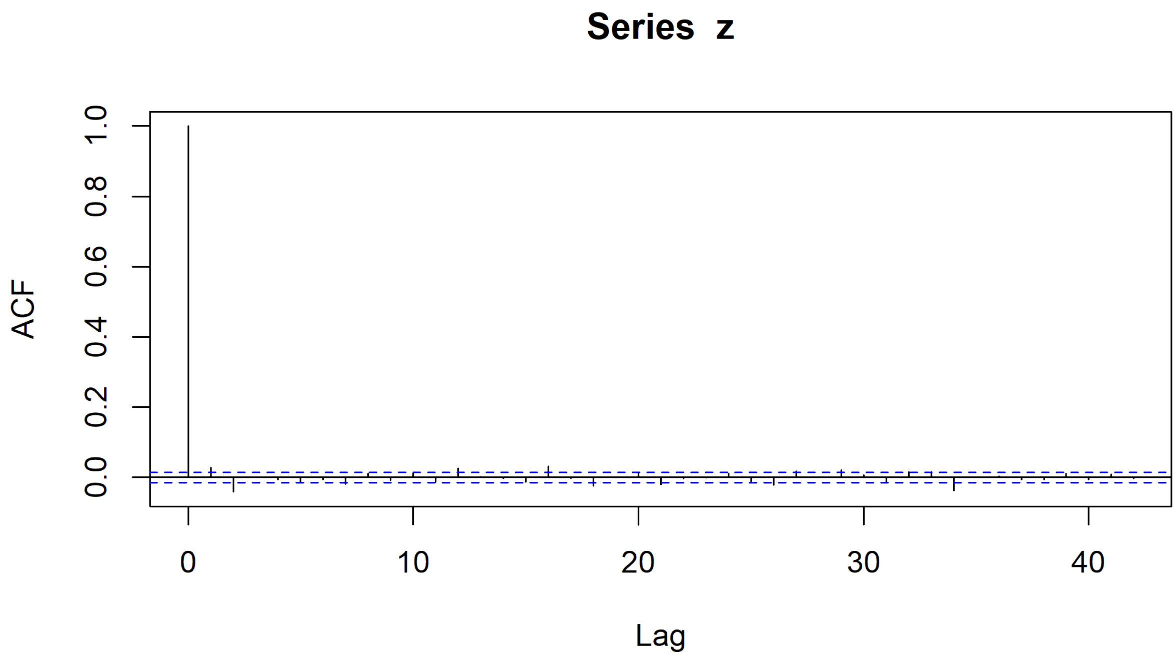
```r
mu <- mean(diff(log(sp500)))
sigma <- sd(diff(log(sp500)))
set.seed(95483123)
X1 <- log(sp500[1])+cumsum(c(0,rnorm(N-1,mean=mu,sd=sigma)))
set.seed(324324587)
X2 <- log(sp500[1])+cumsum(c(0,rnorm(N-1,mean=mu,sd=sigma)))
par(mfrow=c(1,2))
plot(X1,type="l")
plot(X2,type="l")
```

Random walk model here only initialized with our first data point (doesn't depend on any others). Here, the drift is estimated using the mean of the cumulative differences



- Seems reasonable enough so far. Let's plot the sample autocorrelation function (sample ACF) of `diff(log(sp500))`.
- It is bad style to refer to quantities using computer code notation. We should set up mathematical notation in the text. Let's try again…
- Let $y^*_{1:N}$ be the time series of S&P 500 daily closing values downloaded from yahoo.com. Let $z^*_n = \Delta \log y^*_n = \log y^*_n - \log y^*_{n-1}$.
- The temporal difference of the log of the value of an investment is called the **return** on the investment. Let's plot the sample autocorrelation function of the time series of S&P 500 returns, $z^*_{2:N}$.

```
z <- diff(log(sp500))
acf(z)
```



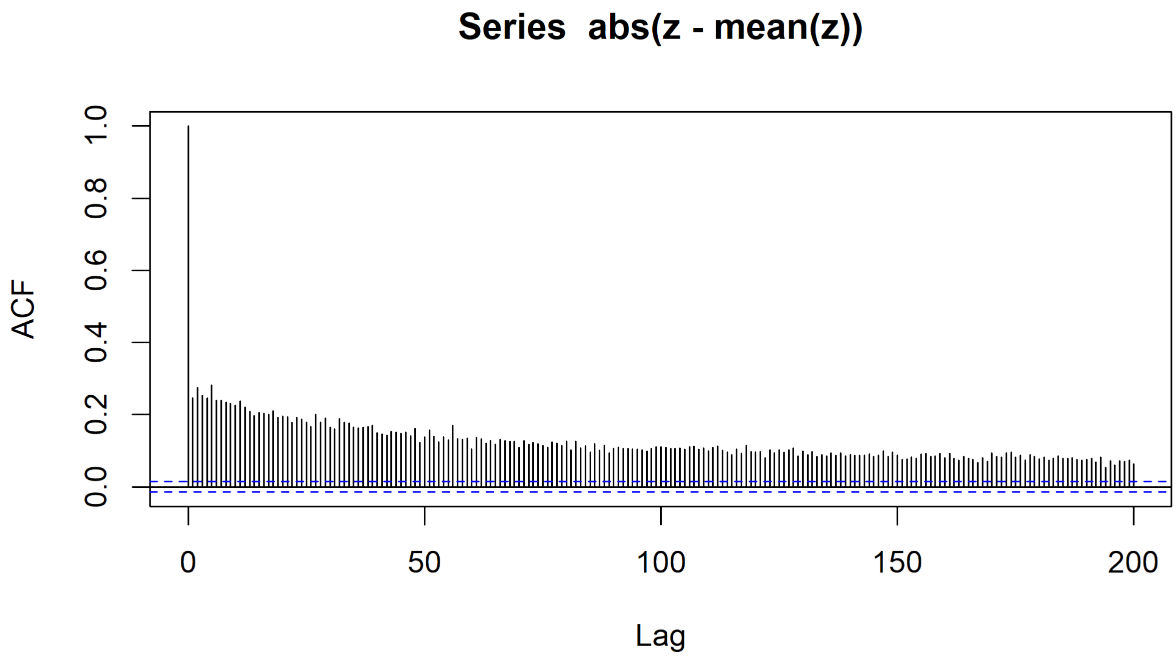
- This looks pretty close to the ACF of white noise. There is some evidence for a small nonzero autocorrelation at lags 0 and 1.
- Here, we have have a long time series ($N = 16616$). For such a long time series, statistically significant effects may be practically insignificant.

Question: Why may the length of the time series be relevant when considering practical versus statistical significance?

- It seems like the S&P 500 returns (centered, by subtracting the sample mean) may be a real-life time series well modeled by white noise.
- However, things become less clear when we look at the absolute value of the centered returns.

Question: How should we interpret the following plot? To what extent does this plot refute the white noise model for the centered returns (or, equivalently, the random walk model for the log index value)?

```
acf(abs(z-mean(z)),lag.max=200)
```



Implies that magnitudes are correlated. If a white noise model isn't iid, the correlation in absolute values is possible. But for iid white noise, the absolute values are uncorrelated

Let ξ_n be iid whitenoise. Let H_n be a stationary process independent of ξ_n. Claim: $\xi_n H_n$ is white noise with correlated absolute value.

Volatility and market inefficiencies

- Nowadays, nobody is surprised that the sample ACF of a financial return time series shows little or no evidence for autocorrelation.
- Deviations from the efficient market hypothesis, if you can find them, are of interest.
- Also, it remains a challenge to find good models for **volatility**, the conditional variance process of a financial return model.
