MSDS 596 Regression & Time Series

Lecture 10 ARMA and ARIMA Models

Department of Statistics
Rutgers University

Nov 5, 2020

Schedule

Week	Date	Topic
1	9/3	Intro to linear regression (JF1,2)
2	9/10	Estimation (JF2)
3	9/17	Inference I (JF3)
4	9/24	Inference II (JF3)
5	10/1	Inference and prediction (JF4,5)
6	10/8	Explanation; model diagnostics (JF6-8)
7	10/15	Transformation and model selection (JF9-10)
8	10/22	Shrinkage methods (JF11)
9	10/29	Time series exploratory analysis (CC2,3)
10	11/5	Linear time series: ARIMA models (CC4-5)
11	11/12	Model specification and estimation (CC6,7)
12	11/19	Diagnostics and forecasting (CC8,9)
	11/26	(no class)
13	12/3	Seasonal models (CC10)
14	week of 12/7	Project & final evaluation

Models for Stationary Time Series

$\sim [0, \sigma^2]$

Let $\{y_t\}$ be an observed time series and $\{e_t\}$ an unobserved white noise. $\{y_t\}$ is said to be a **general linear process** if it can be represented as

$$y_t = \psi_0 e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots$$

$\psi_0 = 1$

- In order for the above expression to make sense, we require that

$$\sum_{i=1}^{\infty} \psi_i^2 < \infty.$$

$Var(Y_t) = \sigma^2 \cdot \left(1 + \sum_{i=1}^{\infty} \psi_i^2\right)$

- Suppose $\psi_0 = 1$ throughout.
- Autoregressive (**AR**), moving average (**MA**) and **ARMA** processes.

Example: a general linear process

$\{y_t\}$ is a **general linear process** if it can be represented as

$$y_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots$$

Suppose $\psi_i = \phi^i$ for some $|\phi| < 1$. That is,

$$y_t = e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \dots$$

- $E(Y_t) = 0$;
- $Var(Y_t) = \sigma^2/(1-\phi^2)$, where $Var(e_t) = \sigma^2$;
    - *Geometric series sum: $1 + a + a^2 + \cdots = 1/(1-a)$ for $|a| < 1$;
- $Cov(Y_t, Y_{t-k}) = \phi^k \sigma^2/(1-\phi^2)$, and $Corr(Y_t, Y_{t-k}) = \phi^k$;
- This is the GLP representation of an *AR(1)* process, and is **stationary**.

For GLPs, can show that

$$E(Y_t) = 0, \qquad \gamma_k = Cov(Y_t, Y_{t-k}) = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}.$$

Moving Average Processes

A **moving average (MA) process** is a general linear process with a finite number of nonzero ψ weights. In particular,

- The time series $\{Y_t\}$ is an MA process of order q, denoted MA(q), if for every t

  $$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q},$$

  where $\{e_t\}$ is $\mathrm{WN}(0, \sigma^2)$.
- MA processes are always stationary.

MA(1) Process

$$Y_t = e_t - \theta e_{t-1}.$$

$Y_{t-2} = e_{t-2} - \theta e_{t-3}$.

- $E(Y_t) = 0$;
- $Var(Y_t) = (1 + \theta^2)\sigma^2 = \gamma_0$;
- $Cov(Y_t, Y_{t-1}) = -\theta\sigma^2 = \gamma_1$ and
  $Cov(Y_t, Y_{t-k}) = 0 = \gamma_k$ for $k \geq 2$;
- Therefore, $\rho_1 = -\theta/(1 + \theta^2)$, and $\rho_k = 0$ for $k \geq 2$.
  (Same derivation as for the two-point moving average.)
- Notice that
  - MA(1) process has no correlation beyond lag 1;
  - ρ_1 decreases as θ increases;
  - Both θ and its reciprocal $1/\theta$ result in the same ρ: invertibility of MA processes (more on this later).
- Examples (see R markdown).

MA(2) Process

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}.$$

- $E(Y_t) = 0$;
- $Var(Y_t) = (1 + \theta_1^2 + \theta_2^2)\sigma^2 = \gamma_0$;
- $Cov(Y_t, Y_{t-1}) = (-\theta_1 + \theta_1\theta_2)\sigma^2 = \gamma_1$,
  $Cov(Y_t, Y_{t-2}) = -\theta_2\sigma^2 = \gamma_2$, and
  $Cov(Y_t, Y_{t-k}) = 0 = \gamma_k$ for $k \geq 3$;
- $\rho_k = \begin{cases} (-\theta_1 + \theta_1\theta_2) / \left(1 + \theta_1^2 + \theta_2^2\right) & k = 1 \\ -\theta_2 / \left(1 + \theta_1^2 + \theta_2^2\right) & k = 2 \\ 0 & k \geq 3 \end{cases}.$
- The ACF cuts off after lag 2.
- Examples (see R markdown).

MA(q) Process

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}.$$

- $E(Y_t) = 0$;
- $\gamma_0 = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma^2$;
- $\rho_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1+\theta_1^2+\cdots+\theta_q^2} & k = 1, 2, \ldots, q \\ 0 & k \geq q+1 \end{cases}.$
- Again, the ACF cuts off after lag q.

Autoregressive Processes

The time series $\{Y_t\}$ is an **autoregressive (AR) process** of order p, denoted AR(p), if for every t

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t,$$

where $\{e_t\}$ is WN$(0, \sigma^2)$.

- Assume $E(Y_t) = 0$ (because if not, can replace Y_t with $Y_t - \mu$);

AR(1)

$$Y_t = \phi Y_{t-1} + e_t \qquad \leftarrow \textit{the "defining recursion"}$$

- $Var(Y_t) = \phi^2 Var(Y_{t-1}) + \sigma^2$, where $Var(Y_t) = Var(Y_{t-1})$ if stationary;
- Thus, $\gamma_0 = Var(Y_t) = \sigma^2/(1-\phi^2)$, if $|\phi| < 1$;
- $Cov(Y_t, Y_{t-k}) = \underbrace{E(Y_t Y_{t-k})}_{=\gamma_k} = \phi \underbrace{E(Y_{t-1}Y_{t-k})}_{\gamma_{k-1}} + E(e_t Y_{t-k})$;
- Thus for $k = 1, 2, \ldots,$

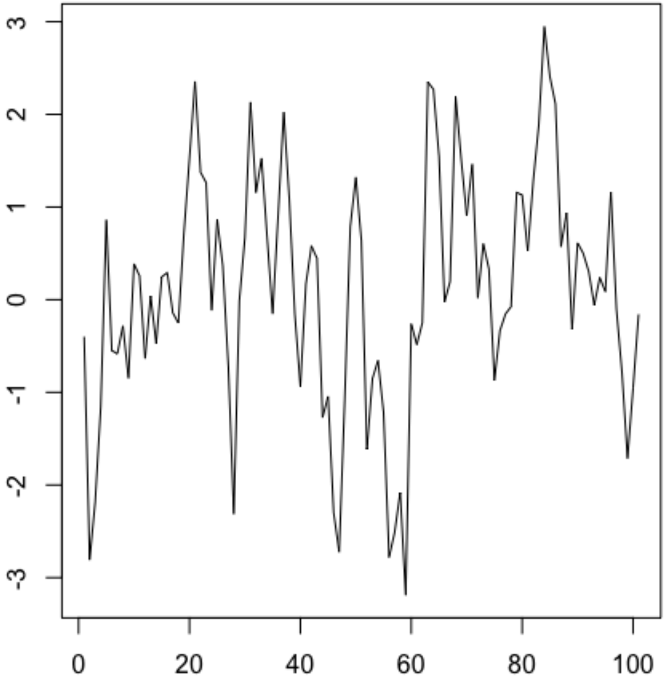
$$\gamma_k = \phi\gamma_{k-1} = \phi(\phi\gamma_{k-2}) = \cdots = \phi^k\gamma_0.$$

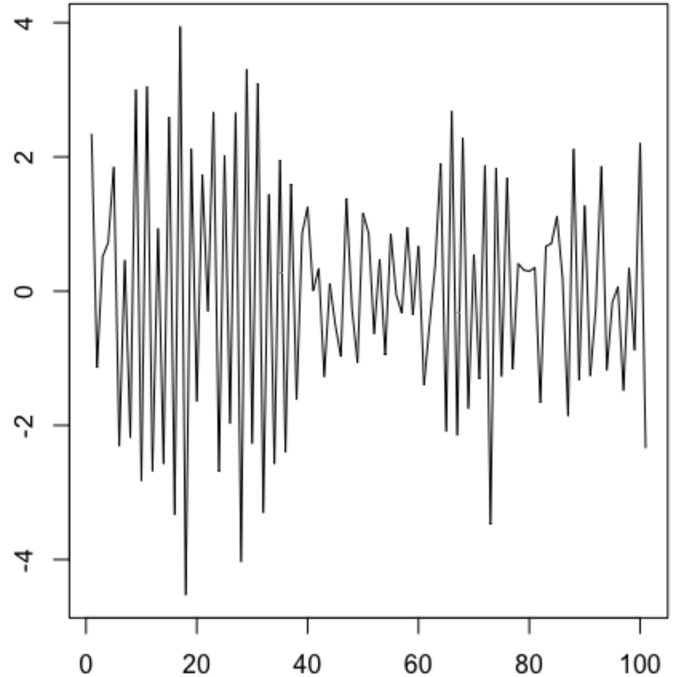
- Similarly, $\rho_k = \gamma_k/\gamma_0 = \phi^k$. That is, the ACF of AR(1) exhibits exponential decay (up to sign) according to the damping factor ϕ.
- These equations called the Yule-Walker equations for AR(1).
- Examples (some below; more in R markdown).

AR(1)

```
T=10000
x1=arima.sim(model=list(ar=c(0.8)),T)
x2=arima.sim(model=list(ar=c(-0.8)),T)
plot.ts(x1[100:200]); plot.ts(x2[100:200])
```

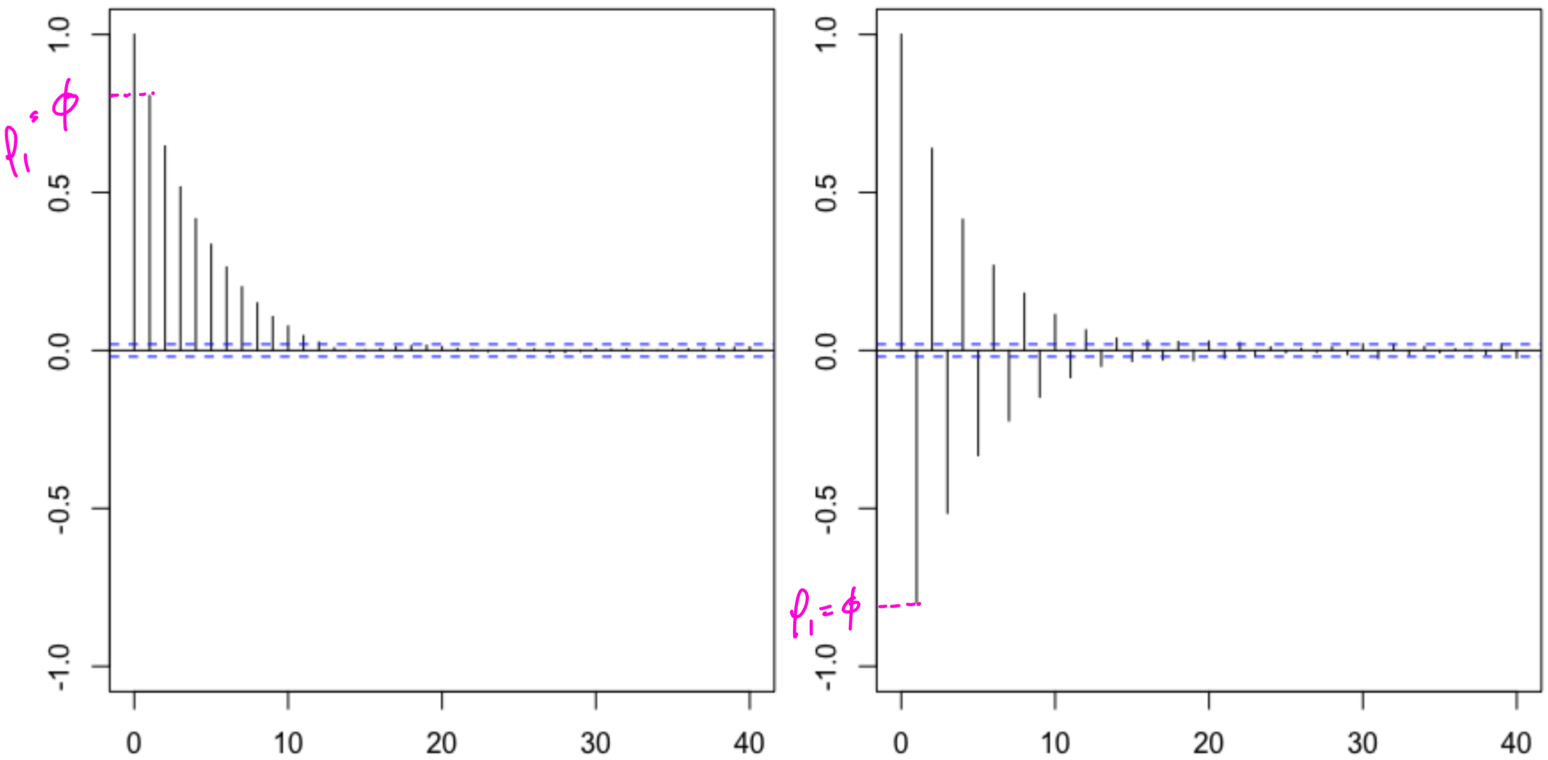
AR(1): $\phi_1 = 0.8$ *vs.* $\phi_1 = -0.8$





AR(1)

AR(1): $\phi_1 = 0.8$ *vs.* $\phi_1 = -0.8$



AR(1): GLP representation

- Write the defining recursion of AR(1) as

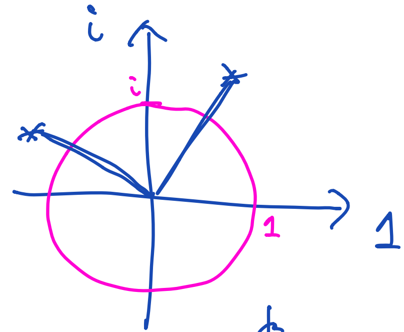
$$
\begin{aligned}
Y_t &= \phi Y_{t-1} + e_t = \phi\left(\phi Y_{t-2} + e_{t-1}\right) + e_t = \cdots \\
&= e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \cdots + \phi^{k-1} e_{t-k+1} + \phi^k Y_{t-k} \\
&\vdots \\
&= e_t + \underbrace{\phi}_{\psi_1} e_{t-1} + \underbrace{\phi^2}_{\psi_2} e_{t-2} + \underbrace{\phi^3}_{\psi_3} e_{t-3} \ldots
\end{aligned}
$$

  which is the GLP representation as we've seen before.
- Since we require $\sum_{i=1}^{\infty} \psi_i^2 < \infty$ for a GLP,

$$|\phi| < 1 \qquad \Longleftrightarrow \qquad \text{AR(1) is stationary.}$$

- In general, an AR process is said to be causal if it possesses a GLP representation.

AR(2)



$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t.$$

Stationarity?

- Introduce the **characteristic polynomial** of AR(2):

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2,$$

and the **characteristic equation**: $\phi(z) = 0$. and |roots| > 1

solve

- The characteristic equation has two **roots**: $\frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}$, which may be real or complex.
- An AR(2) process is stationary if and only if both roots are **larger than 1 in modulus**.
  - Equivalently, $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$, $|\phi_2| < 1$.

polyroot( )

AR(2)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t.$$

The ACF of AR(2):

- Consider $E(Y_t Y_{t-k}) = E(Y_{t-k}(\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t))$.
- Yule-Walker equations for AR(2):

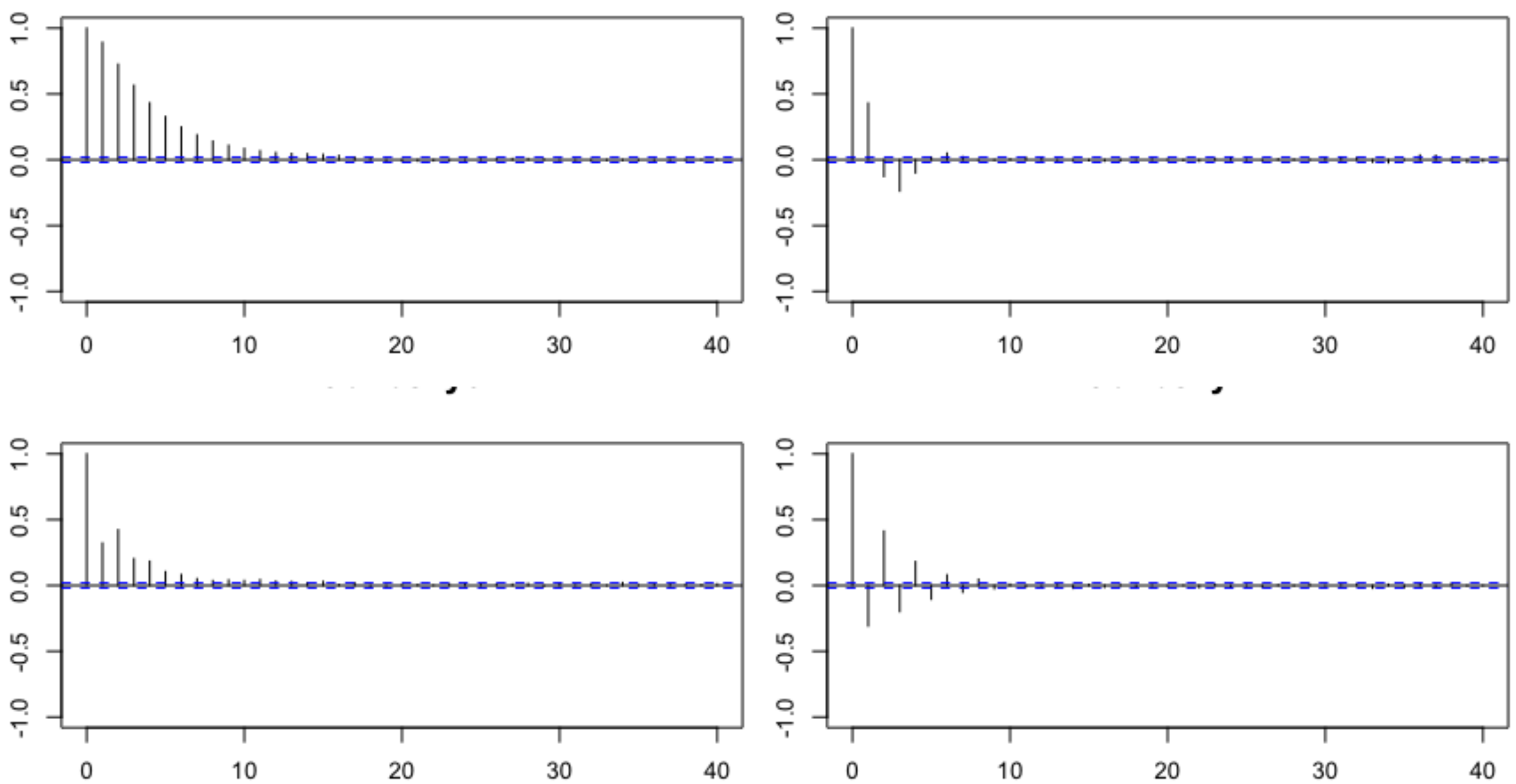
$$\gamma_k = \begin{cases} \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma^2 & \text{for } k = 0, \\ \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} & \text{for } k > 0, \end{cases}$$

  and the ACF satisfies $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$ for $k > 0$.
- Use $k = 1$, $\rho_0 = 1$, and $\rho_{-1} = \rho_1$ to start the recursion:

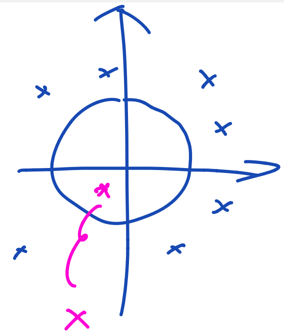
$$\begin{aligned} \rho_1 &= \phi_1/(1-\phi_2) \\ \rho_2 &= \phi_1\rho_1 + \phi_2\rho_0 = \left\{\phi_2(1-\phi_2) + \phi_1^2\right\}/(1-\phi_2) \\ &\vdots \end{aligned}$$

Autoregressive process: AR(2)

AR(2), clockwise from top left:
$(\phi_1, \phi_2) = (1.2, -.35)\,;\ (.6, -.4)\,;\ (-.2, -.35)\,;\ (-.2, .35)\,.$



AR(p)



$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t,$$

Stationarity:

- The characteristic polynomial for an AR(p) process is

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p,$$

  and the characteristic equation: $\phi(z) = 0$.

- The AR(p) process is stationary if and only if all roots to the characteristic equation are larger than 1 in modulus.
- In particular, stationarity implies that

$$\phi_1 + \phi_2 + \cdots + \phi_p < 1, \quad \& \quad |\phi_p| < 1,$$

  although these alone aren't enough.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t,$$

The ACF of AR(p):

- Again consider $E(Y_t Y_{t-k})$, noticing that $E(e_t Y_t) = \sigma^2$.

  $= f(\underbrace{X_{t-1}, \cdots X_{t-p}}_{\perp e_t}, e_t)$

- Yule-Walker equations for AR(p):

$$\begin{cases} \gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_k \gamma_{k-p} + \sigma^2 & \text{when } k = 0, \\ \gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_k \gamma_{k-p} & \text{when } k > 0. \end{cases}$$

- Solve the first $p+1$ equations corresponding to $k = 0, 1, \ldots, p$ to get $\gamma_0, \gamma_1, \ldots, \gamma_p$, and then compute the γ_k recursively for $k \geq p+1$.

Invertibility

Take the MA(1) process, $Y_t = e_t - \theta e_{t-1}$. $\Rightarrow Y_{t-1} = e_{t-1} - \theta e_{t-2}$

- The ACF is

$$\rho_1 = -\theta/(1+\theta^2), \& \quad \rho_k = 0, \; k \geq 2.$$

  Both θ and its reciprocal $1/\theta$ result in the same ρ, creating non-identifiability of the model.

- Can the MA process be expressed as an AR process?

$$\begin{aligned} e_t &= Y_t + \theta\left(Y_{t-1} + \theta e_{t-2}\right) = Y_t + \theta Y_{t-1} + \theta^2 e_{t-2} = \cdots \\ &= Y_t + \theta Y_{t-1} + \theta^2 Y_{t-2} + \theta^3 Y_{t-3} + \cdots \end{aligned}$$

  Or alternatively,

$$Y_t = \left(-\theta Y_{t-1} - \theta^2 Y_{t-2} - \theta^3 Y_{t-3} \cdots\right) + e_t$$

- If $|\theta| < 1$, the MA(1) process has been "inverted" into an AR(∞) process, and we say that the MA(1) process is invertible.

Invertibility

For MA(q) process,

- The characteristic polynomial for an MA(q) process is

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_q z^q,$$

  and the characteristic equation: $\theta(z) = 0$.
- The MA(q) process $\{Y_t\}$ is said to be invertible if and only if all roots to the characteristic equation are larger than 1 in modulus.
- For a given ACF, there is a unique set of parameter values that yield an invertible MA process.
- For this reason, we require MA processes to be invertible.

(Mixed) Autoregressive Moving Average Processes

The time series $\{Y_t\}$ is said to be an **(mixed) autoregressive moving average process** of order (p, q), denoted ARMA(p, q), if for every t

$$\begin{aligned} Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t \\ &\quad -\theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}, \end{aligned}$$

where $\{e_t\}$ is a $\mathrm{WN}(0, \sigma^2)$.

ARMA(1, 1)

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1}.$$

- Notice $E(e_t Y_t) = \sigma^2$, and
  $E(e_{t-1} Y_t) = E(e_{t-1}(\phi Y_{t-1} + e_t - \theta e_{t-1})) = (\phi - \theta)\sigma^2$.
- Calculate $E(Y_t Y_{t-k})$ to get the Yule-Walker equations

$$\begin{aligned} \gamma_0 &= \phi\gamma_1 + \{1 - \theta(\phi - \theta)\}\sigma^2 \\ \gamma_1 &= \phi\gamma_0 - \theta\sigma^2 \\ \gamma_k &= \phi\gamma_{k-1} \qquad \text{for } k \geq 2 \end{aligned}$$

- We obtain

$$\begin{aligned} \gamma_0 &= \left(1 - 2\phi\theta + \theta^2\right)\sigma^2 / \left(1 - \phi^2\right) \\ \frac{\gamma_k}{\gamma_0} = \rho_k &= (1 - \theta\phi)(\phi - \theta)\phi^k / \left(1 - \phi^2\right) \qquad \text{for } k \geq 1. \end{aligned}$$

Notice that starting from lag 2, the ACF exhibits exponential decay w/ damping factor ϕ.

ARMA(1, 1)

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1}.$$

- The GLP representation of ARMA(1, 1) is

$$Y_t = e_t + (\phi - \theta) \sum_{i=1}^{\infty} \phi^{i-1} e_{t-i},$$

$\sim \psi_k$'s

that is, $\psi_i = (\phi - \theta)\,\phi^{i-1}$ for $i \geq 1$.

- The ARMA(1, 1) process is stationary if $|\phi| < 1$, i.e. the AR(1) component is stationary.

ARMA(p, q)

$$
\begin{aligned}
Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t \\
&\quad -\theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q},
\end{aligned}
$$

- The characteristic polynomials for an ARMA(p,q) process are defined as

$$
\begin{aligned}
\phi(z) &= 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p, \\
\theta(z) &= 1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_q z^q.
\end{aligned}
$$

- The ARMA(p, q) process is stationary if its AR(p) component is stationary, i.e. if all roots to $\phi(z) = 0$ are larger than 1 in modulus.
- The ARMA(p, q) process is invertible if its MA(q) component is invertible, i.e. if all roots to $\theta(z) = 0$ are larger than 1 in modulus.
- We require ARMA processes to be stationary and invertible.

ARMA(p, q)

$$\begin{aligned} Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t \\ &\quad -\theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}, \end{aligned}$$

- The GLP representation of ARMA(p, q) is

$$\begin{aligned} \psi_0 &= 1 \\ \psi_1 &= -\theta_1 + \phi_1 \\ \psi_2 &= -\theta_2 + \phi_2 + \phi_1 \psi_1 \\ &\vdots \\ \psi_k &= -\theta_k + \phi_p \psi_{k-p} + \phi_{p-1} \psi_{k-p+1} + \cdots + \phi_1 \psi_{k-1} \end{aligned}$$

  assuming $\psi_k = 0$ for all $k < 0$, and $\theta_k = 0$ for all $k > q$.
- The ACF of ARMA(p, q) satisfies for all $k > q$,

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}.$$

  For $k \leq q$, ρ_k also involves the θ terms.

AR(p) estimation: Yule-Walker equations

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \cdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \cdots \\ \gamma_p \end{pmatrix},$$

and

$$\sigma^2 = \gamma_0 - \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_p \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \cdots \\ \gamma_p \end{pmatrix}.$$

The estimates of ϕ and σ^2 can be obtained by plugging in the sample autocovariances $\hat{\gamma}_0, \hat{\gamma}_1, \ldots \hat{\gamma}_p$ and solving the equations.

Estimation for ARMA processes

- Estimation using Yule-Walker equations is a **method of moments (MoM)** approach to parameter estimation. Note that we do not use the MoM approach to estimate MA processes, or the MA component of ARMA processes, because of instability of the resulting estimates (parameters are highly nonlinear functions of the data).
- **Least squares** (conditional and unconditional) estimation;
- **Maximum likelihood** estimation.

Nonstationary time series

- Example: a time series $\{Y_t\}$ is called a random walk if it satisfies:

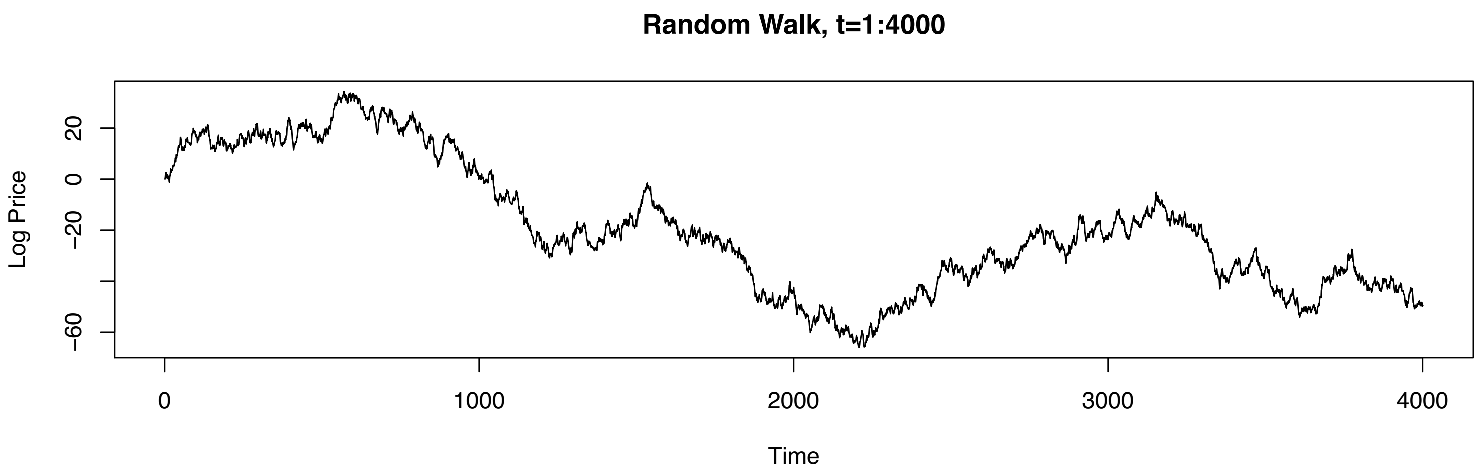
$$Y_t = Y_{t-1} + e_t,$$

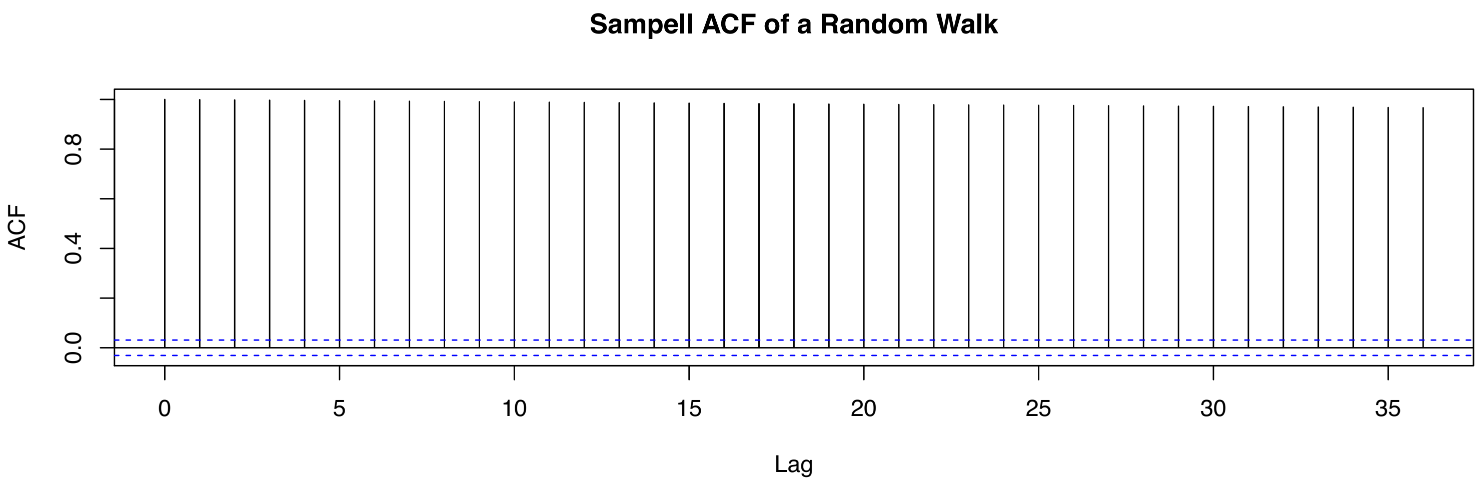
  where e_t are i.i.d. with mean zero and variance σ^2.
- It is an AR(1) model with coefficient $\phi_1 = 1$ and $\phi_0 = 0$.
  - Nonstationary: the variance diverges as t increases.
  - Strong memory: the sample ACF approaches 1 for any finite lag.
  - Unpredictable: l-step ahead forecast is $\hat{r}_h(l) = Y_h$, and $\mathrm{Var}(e_h(l)) = l\sigma^2$.
- A random walk with drift takes the form $Y_t = \mu + Y_{t-1} + e_t$.
  - Same properties as a random walk;
  - In addition, it has a time trend with slope μ:

$$Y_t = \mu t + Y_0 + e_1 + e_2 + \cdots + e_t.$$

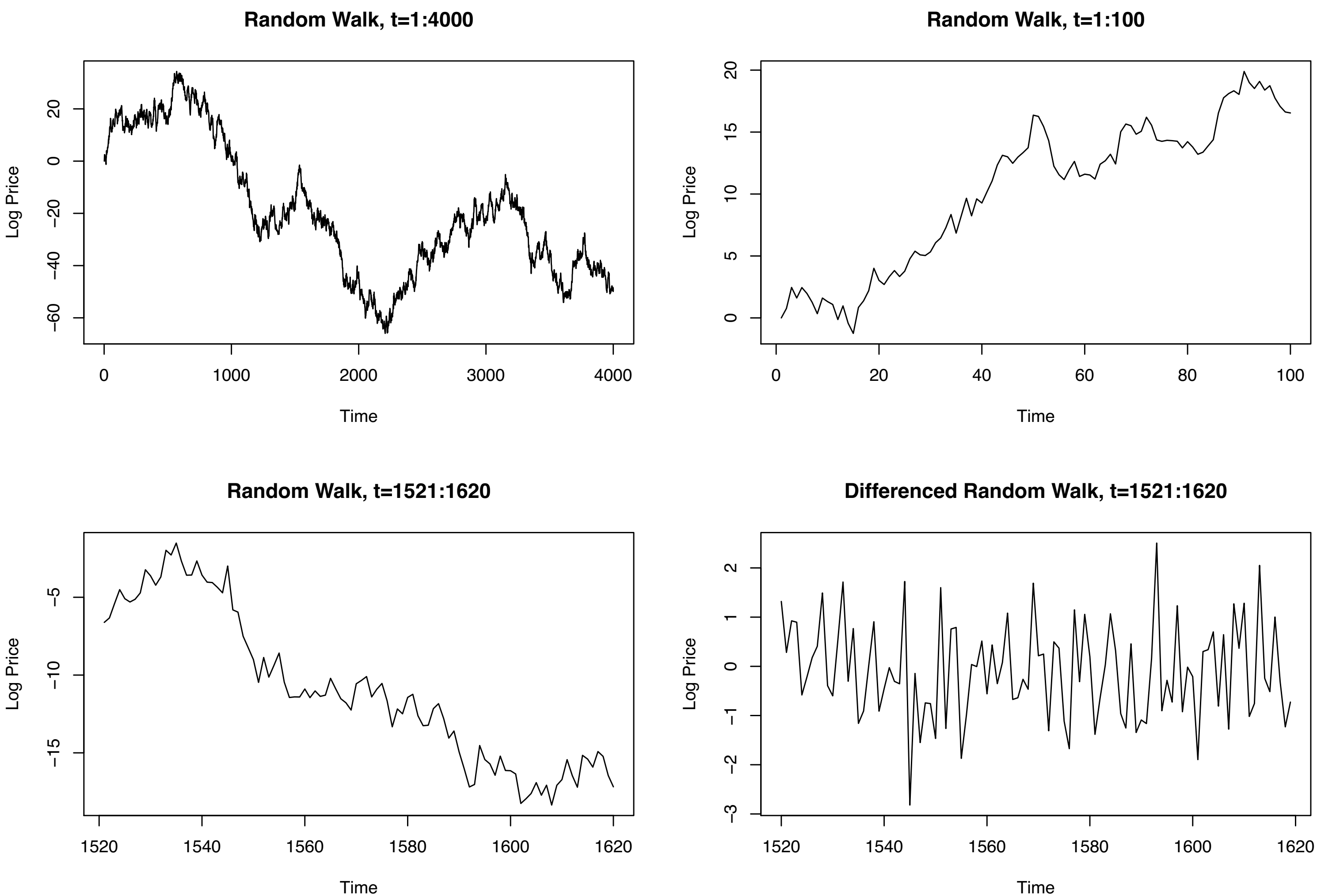
- Differencing: $\nabla Y_t := Y_t - Y_{t-1}$ leads to a white noise (with nonzero mean, if there is a drift).

Example: ACF does not decay

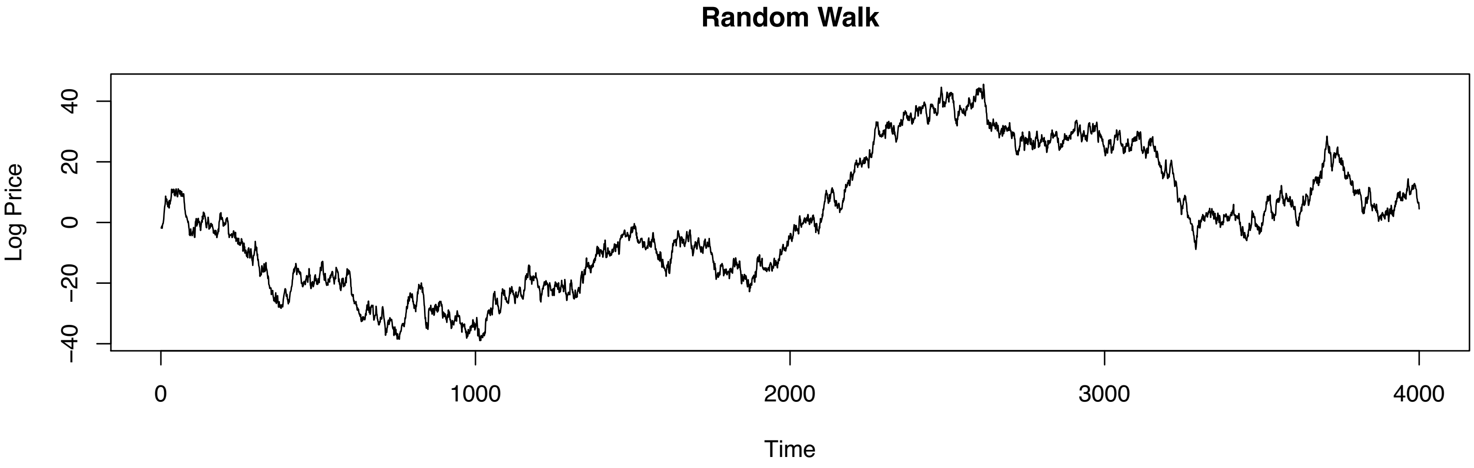


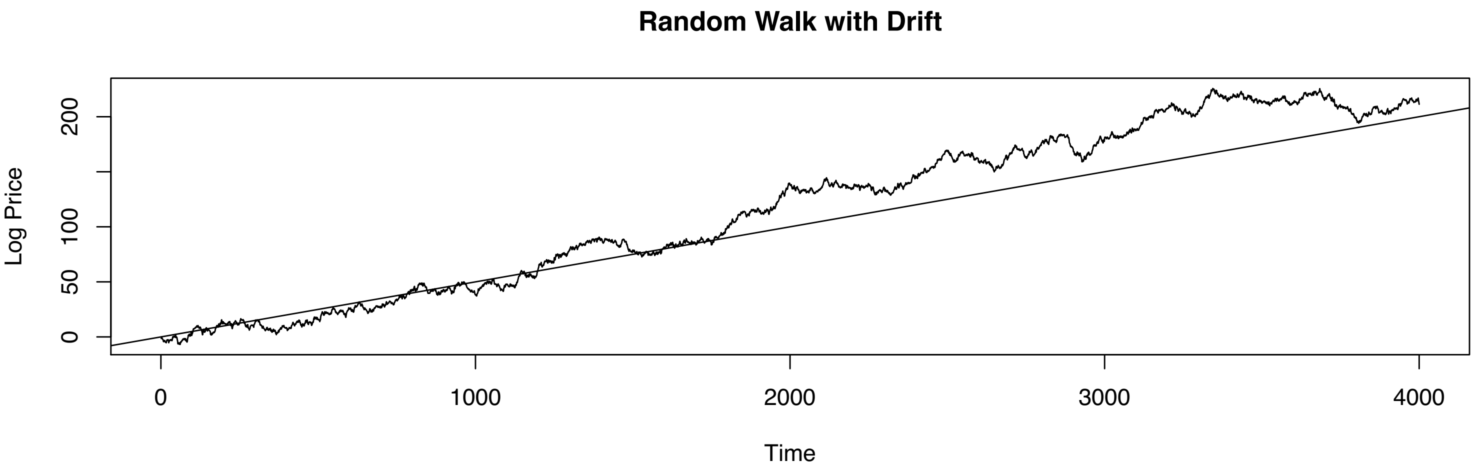


Example: Trend over a Short Time Period



Example: Random Walk with a Drift





The Backshift Operator

The **backshift operator** B operates on the time index of a series and shifts time back one time unit to form a new series. That is,

$$Y_{t-1} = BY_t.$$

- B is linear: $B(aY_t + bX_t + c) = aBY_t + bX_t + c$;
- For any positive integer m, $B^m Y_t = Y_{t-m}$.
- A general AR(p) process can be written as

$$\phi(B)Y_t = e_t$$

  where $\phi(B)$ is the AR characteristic polynomial evaluated at B;
- A general MA(q) process can be written as

$$Y_t = \theta(B)e_t$$

  where $\theta(B)$ is the MA characteristic polynomial evaluated at B;
- Combining the two, a general ARMA(p, q) process can be written as

$$\phi(B)Y_t = \theta(B)e_t$$

The Differencing Operator

The **differencing operator** $\nabla = 1 - B$. Therefore,

$$\nabla Y_t = (1 - B)Y_t = Y_t - Y_{t-1}.$$

- The **second difference**

$$\nabla^2 Y_t = (1 - B)^2 Y_t = (1 - B)(Y_t - Y_{t-1}) = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2});$$

- The d**-th difference** $\nabla^d Y_t = (1 - B)^d Y_t$;
- To be distinguished from the **seasonal difference** of period s:

$$\nabla_s Y_t = (1 - B^s) Y_t = Y_t - Y_{t-s}.$$

ARIMA Models

- A time series $\{Y_t\}$ is said to be an autoregressive integrated moving average process of order $(p, 1, q)$, denoted ARIMA(p, 1, q), if the differenced series $W_t = \nabla Y_t = Y_t - Y_{t-1}$ follows a ARMA(p, q) model.
- In general, a time series $\{Y_t\}$ is said to be an autoregressive integrated moving average process of order (p, d, q), denoted by ARIMA(p, d, q), if the d-th difference $\nabla^d Y_t$ follows a ARMA(p, q) model.

$$\begin{aligned}
\nabla Y_t &= Y_t - Y_{t-1} \\
\nabla^2 Y_t &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2} \\
\nabla^3 Y_t &= (Y_t - 2Y_{t-1} + Y_{t-2}) - (Y_{t-1} - 2Y_{t-2} + Y_{t-3})
\end{aligned}$$

ARIMA(p, 1, q): unit root nonstationarity

$W_t = \nabla Y_t$, where

$$\begin{aligned} W_t &= \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t \\ &\quad -\theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}. \end{aligned}$$

Solving for Y_t, we have

$$\begin{aligned} Y_t &= (1+\phi_1) Y_{t-1} + (\phi_2 - \phi_1) Y_{t-2} + \cdots + \left(\phi_p - \phi_{p-1}\right) Y_{t-p} \\ &\quad -\phi_p Y_{t-p-1} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \end{aligned}$$

Why isn't this an ARMA(p+1, q) model? CP of the AR component:

$$\begin{aligned} \phi(z) &= 1 - (1+\phi_1) z - (\phi_2 - \phi_1) z^2 - \cdots - \left(\phi_p - \phi_{p-1}\right) z^p + \phi_p z^{p+1} \\ &= \left(1 - \phi_1 z^2 - \cdots - \phi_p z^p\right)(1 - z) \end{aligned}$$

for which $z = 1$ is a root: nonstationarity! Therefore, we do not assume the time series starts at time $-\infty$; instead, we assume it starts at $t = -m$ for some $m < \infty$, and let $Y_t = 0$ for all $t < m$.

Examples of ARIMA Models

- Random walk (ARIMA(0, 1, 0), or the I(1) process): $e_t = \nabla Y_t$.
- IMA(1, 1) = ARIMA(0, 1, 1): $W_t = \nabla Y_t = e_t - \theta e_{t-1}$;
- ARI(1, 1) = ARIMA(1, 1, 0): $\nabla Y_t = \nabla Y_{t-1} + e_t$, for $|\phi| < 1$;
- IMA(2, 2): $W_t = \nabla^2 Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$.