Empirical Finance: Methods & Applications

A Statistical Evaluation of Asset Returns

Pasquale Della Corte
Imperial College London

Week 2

Introduction

The majority of financial economics graduates do not pursue a PhD in Economics/Finance.

"Many go on to … government jobs, and others to the private sector. In many of these positions, it is quite common for our graduates to be exposed to economic data and analysis, including formal econometric (e.g., regression) analysis. Many of these applications are time series in nature. What tools can we give these students to help them succeed?"

Bruce Hansen (2017). *Time-Series Econometrics for the 21st Century*.

Time and Data Collection

Whenever we collect data, time often plays an important role

- This is true for many economic and financial data,
- Asset returns, inflation rate, economic growth, and many others.

Time-series analysis studies how economic and financial data evolve over time

- We can differentiate between deterministic and stochastic patterns,
- We can attempt to forecast future data based on historical data.

What is a Time Series?

Let's Get Started

Stay focused – don't let permanent shocks take control

Stochastic Process

A stochastic process is a collection of random variables denoted as

$$(Y_1, Y_2, \ldots, Y_T) \longrightarrow \text{or simply as } \{Y_t\}$$

A simple example is an *independent and identically distributed* or *or iid* process as

$$Y_t \overset{iid}{\sim} \mathcal{D} \longrightarrow$$

D denotes a given distribution like a Normal or t-student

Another example is the *random walk* as

$$Y_t = Y_{t-1} + \varepsilon_t \longrightarrow$$

where ε_t is an iid process

A Time Series

A time series is a single realization (*non-experimental data*) of a stochastic process as

Only one history observed $\longleftarrow$ $(y_1, y_2, \ldots, y_T)$ $\longrightarrow$ or simply as $\{y_t\}$

Time series observations are close in calendar time, and thus dependent over time

Dependence on all past information $\longleftarrow$ $y_t \not\perp \mathcal{F}_{t-1}$ $\longrightarrow$ Observations can't be treated as independent

Standard cross-sectional regression theory does not apply;
Time ordering is not a detail as it fundamentally changes inference.

One Process: Two Realizations

The real world and the upside down can be viewed as two realizations of the same stochastic process

Dependence: Ripple Effect

A shock today creates effects that persist into the future.

Stochastic Process: Simulation

Simulate a random walk consisting of 1,000 observations ($t = 1, 2, \ldots, 1000$) as

$$y_{i,t} = y_{i,t-1} + \varepsilon_{i,t}$$

where

$$\varepsilon_{i,t} \sim N(0, 1)$$

and

$$y_{i,0} = 0.$$

By repeating this exercise ($i = 1, 2, \ldots, 10$), we have multiple realizations of a stochastic process.

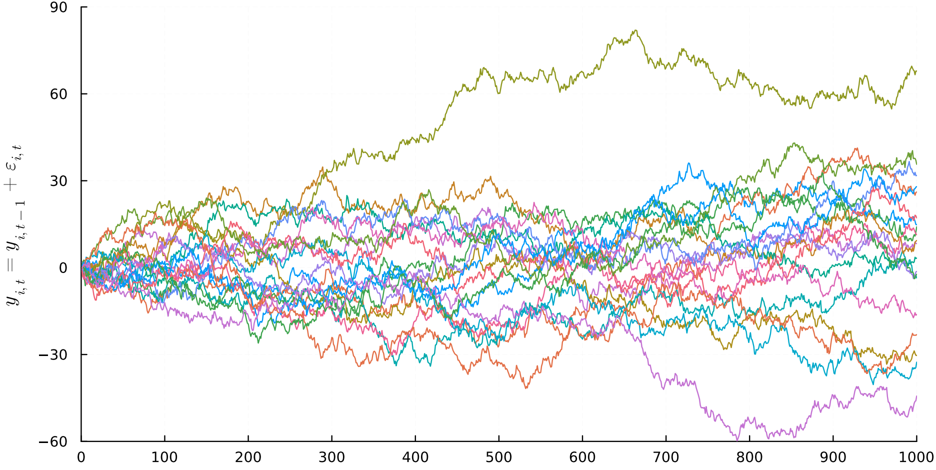
Stochastic Process: Simulation

```
# Set the seed for reproducibility
Random.seed!(9876543)

# Control variables
nstep = 1000   # Number of observations for each series
nsim  = 20     # Number of simulations

# Set the parameters for the shocks
mu    = 0      # Mean of the Normal distribution
sigma = 1      # Standard deviation of the Normal distribution

# Create a matrix to store the random walk values for each simulation
ymat  = zeros(nstep, nsim)

# Set the starting value
y0    = 0


```

Stochastic Process: Simulation

```
# Simulate the random walks
for sim in 1:nsim
y = zeros(nstep)

for i in 1:nstep
shock = rand(Normal(mu, sigma))  # Generate a random shock
if i == 1  y[i] = y0 + shock
else       y[i] = y[i-1] + shock
end
end

ymat[:, sim] = y
end

# Plot the simulated paths
gr()
plot(1:nstep, ymat,  lw=2, ylabel="Value")
```

Stochastic Process: Simulation



Author's simulations based on a Julia script using $y_{i,t} = y_{i,t-1} + \varepsilon_{i,t}$ with $\varepsilon_{i,t} \sim N(0,1)$.

Properties of a Time Series

There are two key concepts to keep in mind:

- Stationarity and Ergodicity

Stationarity

A time series is treated as a random vector with a joint distribution

- Can we infer this joint distribution? Yes, but we need a measure of regularity,
- If the underlying process changes frequently, inference becomes impossible.

Stationarity is a measure of *regularity*

- It helps estimate the unknown parameters of a joint distribution,
- Which one? Think of *means*, *variances*, and *covariances*.

There are two forms of stationarity

- Weak (or covariance) stationarity $\longrightarrow$ important for linear models,
- Strong (or strict) stationarity $\longrightarrow$ important for nonlinear models.

$\longrightarrow$ Related but not nested

Stationarity: A Counter Example

Consider the asteroid that wiped out the dinosaurs: The impact permanently changes the environment, so the past no longer describes the future.

This is not just a big shock, it is a game changer!

Weak Stationarity

A stochastic process $\{Y_t\}$ is weakly stationary if

$$\begin{array}{lcll} E(Y_t) & = & \mu & \text{for all } t \\ Var(Y_t) & = & \sigma^2 < \infty & \text{for all } t \\ Cov(Y_t, Y_{t-j}) & = & \gamma_j & \text{for all } t \text{ and } j. \end{array}$$

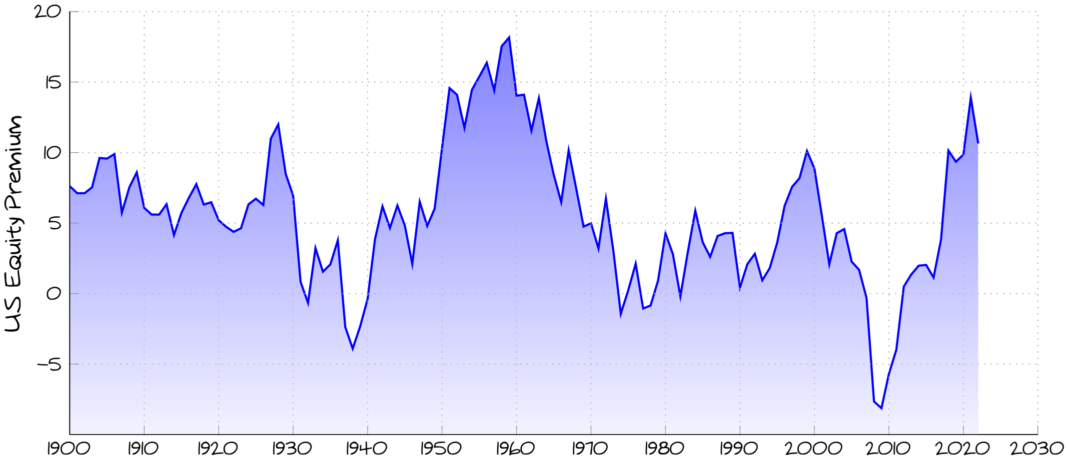
Weak Stationarity requires that the

- The unconditional mean $E(Y_t)$ is finite and constant (it exists!),
- The unconditional variance $Var(Y_t)$ is also finite and constant,
- The covariance $Cov(Y_t, Y_{t-j})$ only depends on the time gap j.

Weak Stationarity only applies to unconditional moments

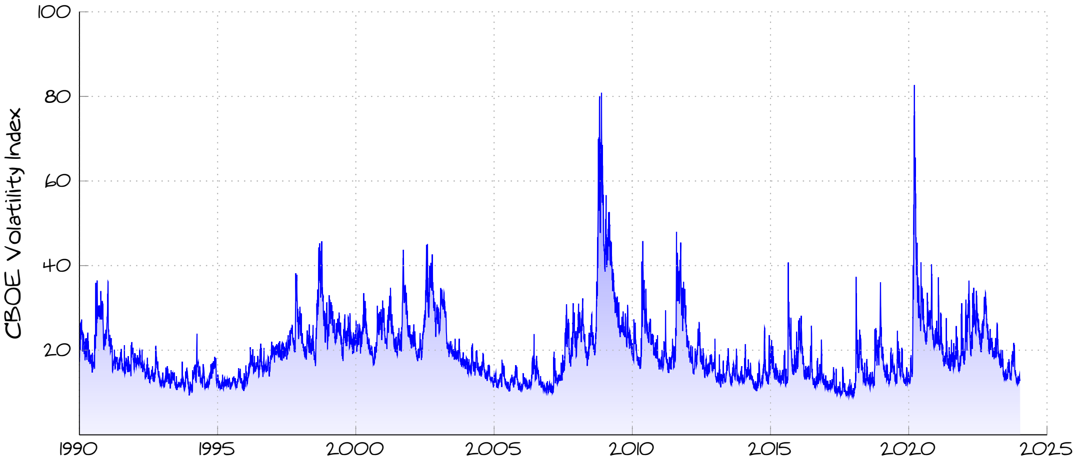
- Conditional means or variances may still vary over time.

US Equity Risk Premium



Data source: Global Financial Data.

VIX Index



Data source: Yahoo Finance.

Autocovariance and Autocorrelation

The autocorrelation of $\{Y_t\}$ is defined as

j-th order autocovariance

$$\rho_j = \frac{Cov(Y_t, Y_{t-j})}{Cov(Y_t, Y_t)} = \frac{\gamma_j}{\gamma_0}$$

j-th order autocorrelation

Consistently estimated via its sample counterpart as

$$\widehat{\rho}_j = \frac{\sum_{t=j+1}^{T}(y_t - \widehat{\mu})(y_{t-j} - \widehat{\mu})}{\sum_{t=1}^{T}(y_t - \widehat{\mu})^2} = \frac{\widehat{\gamma}_j}{\widehat{\gamma}_0}$$

Plotting ρ_j against j gives the autocorrelation function (ACF), summarizing the linear dependence in y_t.

An Example

Consider the following process

$$y = \alpha + \beta t + \varepsilon_t \quad \text{with} \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, 1).$$

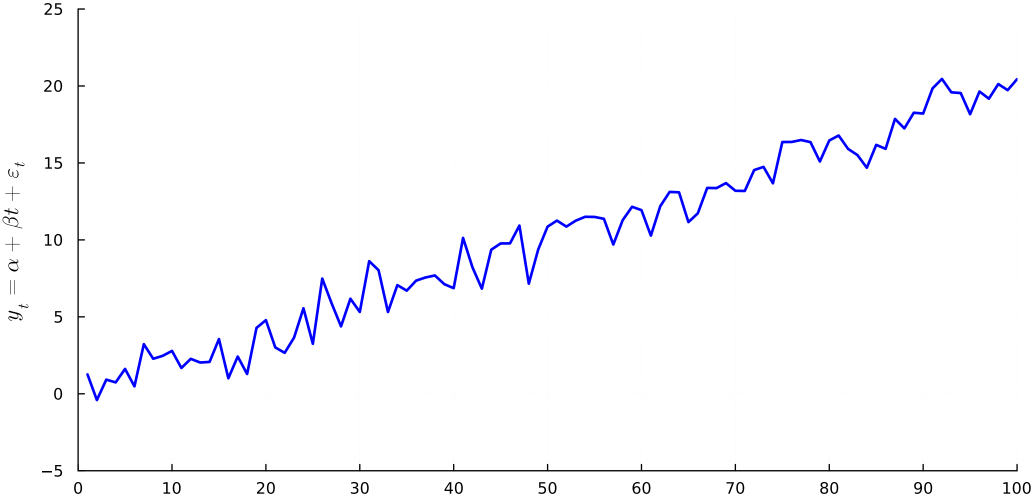
This process is not weakly stationary as

$$t \longrightarrow \text{linear trend}$$

$$\mu = \alpha + \beta t \longrightarrow \text{depends on time}$$

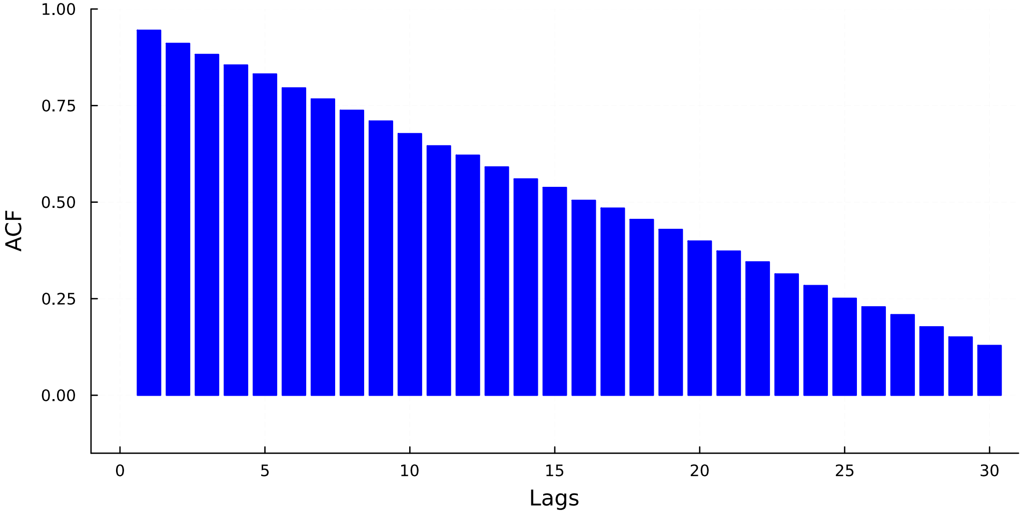
This is a trend-stationary (TS) process, i.e., stationary after we remove the trend component.

Simulated TS Process



Author's simulations based on a Julia script using $\alpha = 0$ and $\beta = 0.2$.

ACF of a TS Process



Author's simulations based on a Julia script using *alpha* $= 0$ and $\beta = 0.2$.

Strong Stationarity

What does strong stationary mean?

- The distribution of a time-series is the same through time for different sub-samples,
- We assume no distribution, only require the same probability distribution.

A stochastic process $\{Y_t\}$ is strictly stationary if

- If the joint distribution of $\{Y_t, \ldots, Y_{t+h}\}$ equals that of $\{Y_{t+\tau}, \ldots, Y_{t+\tau+h}\}$,
- The joint distribution depends on h and not on t (*time invariant*).
- Strong stationarity does not require finite moments.

Stationarity

A **strictly stationary** process with a finite second moment

$\downarrow$

weakly stationary

A **weakly stationary** process with a time invariant
joint distribution of the standardized residuals

$\downarrow$

strictly stationary

Consider the following examples

- A process with time-varying kurtosis is weakly stationary but not strictly stationary,
- A sample from $t(0, 1, \nu)$ with $\nu = 2$ (Cauchy) is strictly stationary but not weakly stationary,
- A normally distributed sample is both strictly and weakly stationary.

Ergodicity

What is the intuition behind ergodicity?

- The dynamics of a time-series is contaminated by shocks (e.g., news, events, etc),
- A shock is transitory, dissipates over time, and won't affect future values.

An example?

- The Great Depression, for example, had a major impact on the economic growth, employment rate, and stock market between 1929 and 1939 in the United States.
- Its effect has vanished by now on the same US economic and financial quantities.

What does ergodicity mean?

- Serial dependence vanishes asymptotically (generalization of the *Law of Large Numbers*),
- Sample moments converge in probability to population moments as errors vanish.

Building Blocks of a Time Series

White Noise

A process $\{\varepsilon_t\}$ is a white noise (WN) if

$$
\begin{array}{lcll}
E(\varepsilon_t) & = & 0 & \textit{for} \;\; t = 1, 2, \ldots \\
V(\varepsilon_t) & = & \sigma^2 < \infty & \textit{for} \;\; t = 1, 2, \ldots \\
Cov(\varepsilon_t, \varepsilon_\tau) & = & 0 & \textit{for} \;\; t \neq \tau \longrightarrow
\end{array}
$$

uncorrelated but not independent

A process $\{\varepsilon_t\}$ is an *iid* white noise or an independent white noise (IWN) if we add

$$\varepsilon_t \perp \varepsilon_\tau \quad \text{for} \quad t \neq \tau \longrightarrow$$

are independent

A process $\{\varepsilon_t\}$ is a Gaussian white noise (GWN) if we add that

$$\varepsilon_t \sim \mathcal{N}(0, \sigma^2) \longrightarrow$$

shocks are iid by construction

An Example

Daily exchange rate returns y_t, for instance, are well described by

$$\begin{aligned} y_t &= \sqrt{h_t}\varepsilon_t, \quad \varepsilon_t \sim IWN(0,1) \\ h_t &= \omega + \alpha y_{t-1}^2, \quad \omega > 0, \alpha > 0 \end{aligned}$$

- Returns are uncorrelated as $E(y_t y_{t-\tau}) = 0$,
- Returns are not independent as $E(y_t^2 | y_{t-1}) = \omega + \alpha y_{t-1}^2$,
- Returns are white noise but not independent white noise.

Absence of correlation does not imply independence

Example 1

Consider the following process

$$y_t = \varepsilon_t \quad \text{with} \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1).$$

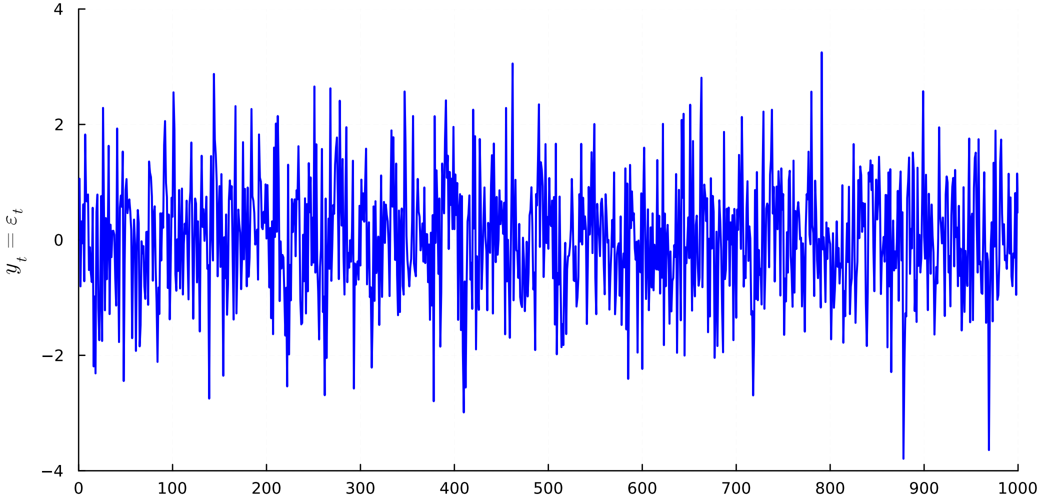
This process is weakly stationary as

$$\mu = 0 \longrightarrow \text{constant}$$

$$\gamma_0 = 1 \longrightarrow \text{constant}$$

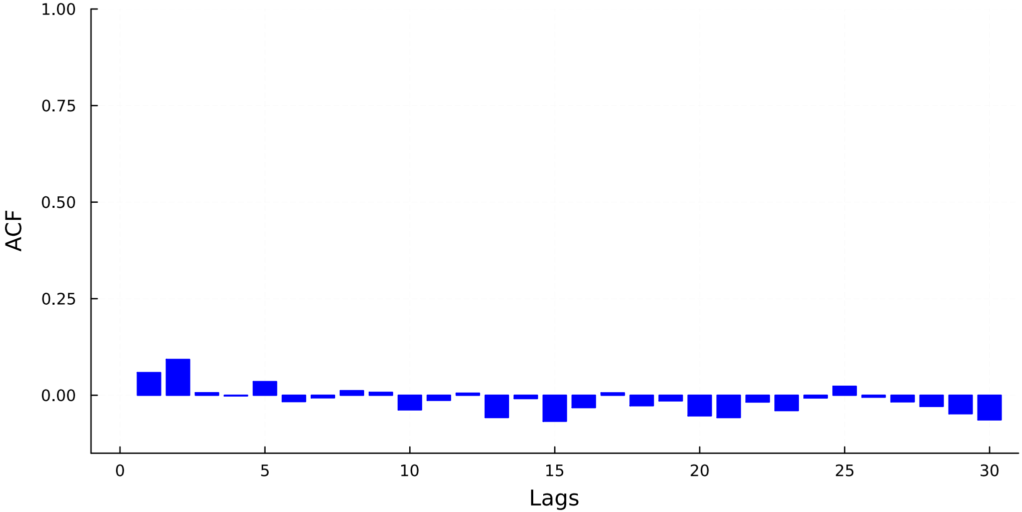
$$\gamma_j = 0 \longrightarrow \text{for } j \geq 1$$

Simulated iid Process



Author's simulations based on a Julia script.

ACF of an iid Process



Author's simulations based on a Julia script.

ARMA Models

ARMA processes, central to time-series analysis, consist of

- **Autoregressive** (AR) processes,
- **Moving Average** (MA) processes.

MA Process

Moving Average Process: MA(1)

The first-order moving average $MA(1)$ process is described as

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} \quad \text{with} \quad \varepsilon_t \sim WN(0, \sigma^2)$$

- μ and θ are parameters, and Y_t depends on the current and previous shock.

Unconditional mean

$$E(Y_t) = E(\mu) + E(\varepsilon_t) + \theta E(\varepsilon_{t-1}) = \mu$$

Unconditional variance

$$Var(Y_t) = Var(\varepsilon_t) + \theta^2 Var(\varepsilon_{t-1}) + 2\theta Cov(\varepsilon_t, \varepsilon_{t-1}) = \sigma^2(1 + \theta^2)$$

Moving Average Process: MA(1)

First-order autocovariance

$$\begin{aligned} Cov(Y_t, Y_{t-1}) &= E[(Y_t - \mu)(Y_{t-1} - \mu)] \\ &= E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})] \\ &= E(\varepsilon_t\varepsilon_{t-1} + \theta\varepsilon_t\varepsilon_{t-2} + \theta\varepsilon_{t-1}^2 + \theta^2\varepsilon_{t-1}\varepsilon_{t-2}) = \theta\sigma^2 \end{aligned}$$

First-order autocorrelation

$$Cor(Y_t, Y_{t-1}) = \frac{Cov(Y_t, Y_{t-1})}{\sqrt{V(Y_t)}\sqrt{V(Y_{t-1})}} = \frac{\theta\sigma^2}{(1+\theta^2)\sigma^2} = \frac{\theta}{(1+\theta^2)}$$

Moving Average Process: MA(1)

Higher-order autocovariance

$$
\begin{aligned}
Cov(Y_t, Y_{t-j}) &= E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-j} + \theta\varepsilon_{t-j-1})] \\
&= E(\varepsilon_t\varepsilon_{t-j} + \theta\varepsilon_t\varepsilon_{t-j-1} + \theta\varepsilon_{t-1}\varepsilon_{t-j} + \theta^2\varepsilon_{t-1}\varepsilon_{t-j-1}) = 0 \quad \text{for} \quad j > 1.
\end{aligned}
$$

Higher-order autocorrelation

$$
Cor(Y_t, Y_{t-j}) = \frac{Cov(Y_t, Y_{t-j})}{\sqrt{V(Y_t)}\sqrt{V(Y_{t-j})}} = \frac{0}{(1+\theta^2)\sigma^2} = 0 \quad \text{for} \quad j > 1.
$$

Moving Average Process: MA(1)

Consider the following process

$$y_t = 0.5 + \varepsilon_t + 0.8\varepsilon_{t-1} \quad \text{with} \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, 0.04).$$

This process is weakly stationary as

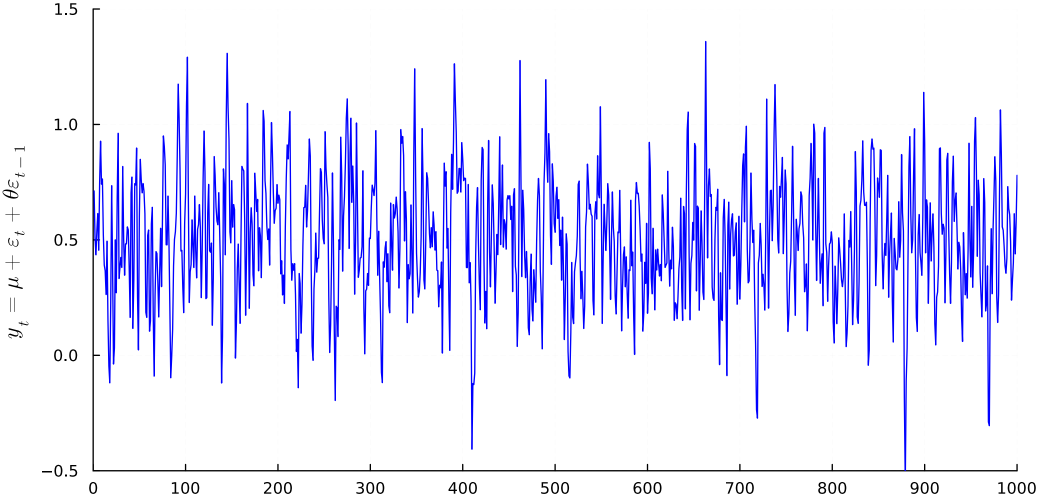
$$\mu = 0.5 \longrightarrow \text{constant}$$

$$\gamma_0 = 0.066 \longrightarrow \sigma^2(1+\theta^2)$$

$$\rho_1 = 0.49 \longrightarrow \theta/(1+\theta^2)$$

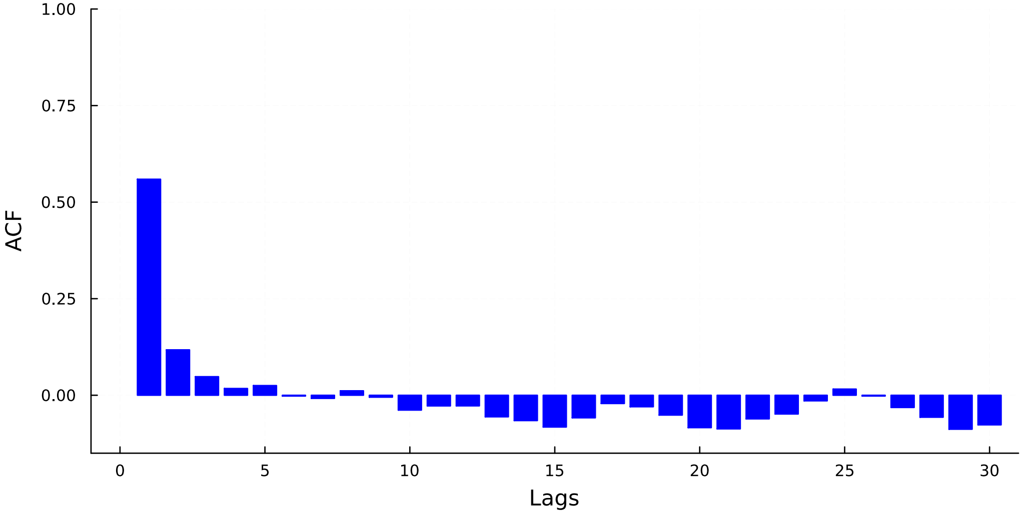
$$\rho_j = 0 \longrightarrow \text{for } j > 1$$

Moving Average Process: Simulated MA(1) Process



Author's simulations based on an R script using 1000 observations.

Moving Average Process: ACF of MA(1) Process



Author's simulations based on an R script using 1000 observations.

Moving Average Process: MA(q)

Take a **q-th order moving average** or $MA(q)$ process

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \quad \text{with} \quad \varepsilon_t \sim WN(0, \sigma^2)$$

Properties

$$\begin{aligned} E(Y_t) &= \mu \\ Var(Y_t) &= \sigma^2 (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2) \end{aligned}$$

Note that I could have also written $Var(Y_t) = \sigma^2 \sum_{i=0}^{q} \theta_i^2$ by setting $\theta_0 = 1$.

Moving Average Process: MA(q)

Properties (*cont'd*)

$$Cov(Y_t, Y_{t-j}) = \begin{cases} \sigma^2 \sum_{i=0}^{q-j} \theta_i \theta_{i+j} & \text{for } j \leq q \\ 0 & \text{for } j > q \end{cases}$$

$$Cor(Y_t, Y_{t-j}) = \begin{cases} \sum_{i=0}^{q-j} \theta_i \theta_{i+j} / \sum_{i=0}^{q} \theta_i^2 & \text{for } j \leq q \\ 0 & \text{for } j > q \end{cases}$$

Autocovariances and autocorrelations are non-zero up to q, and then become zero.

Moving Average Process: MA(2)

Consider the following process

$$y_t = 0.5 + \varepsilon_t + 0.3\varepsilon_{t-1} + 0.5\varepsilon_{t-2} \quad \text{with} \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, 0.04).$$

This process is weakly stationary as

$$\mu = 0.5 \longrightarrow \text{constant}$$

$$\gamma_0 = 0.054 \longrightarrow \sigma^2(1 + \theta_1^2 + \theta_2^2)$$

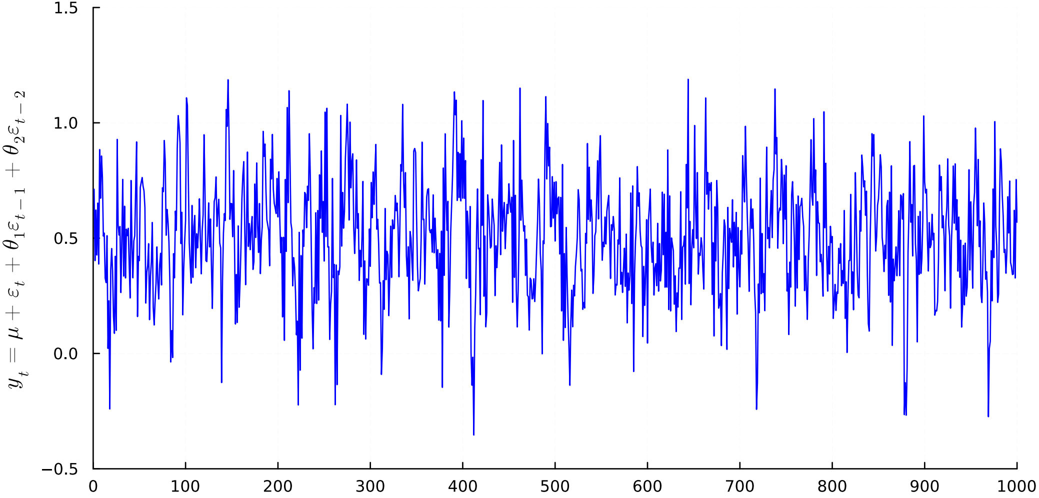
$$\gamma_1 = 0.018 \longrightarrow \sigma^2(\theta_1 + \theta_1\theta_2)$$

$$\gamma_2 = 0.020 \longrightarrow \sigma^2\theta_2$$

$$\rho_1 = 0.33 \longrightarrow \gamma_1/\gamma_0$$

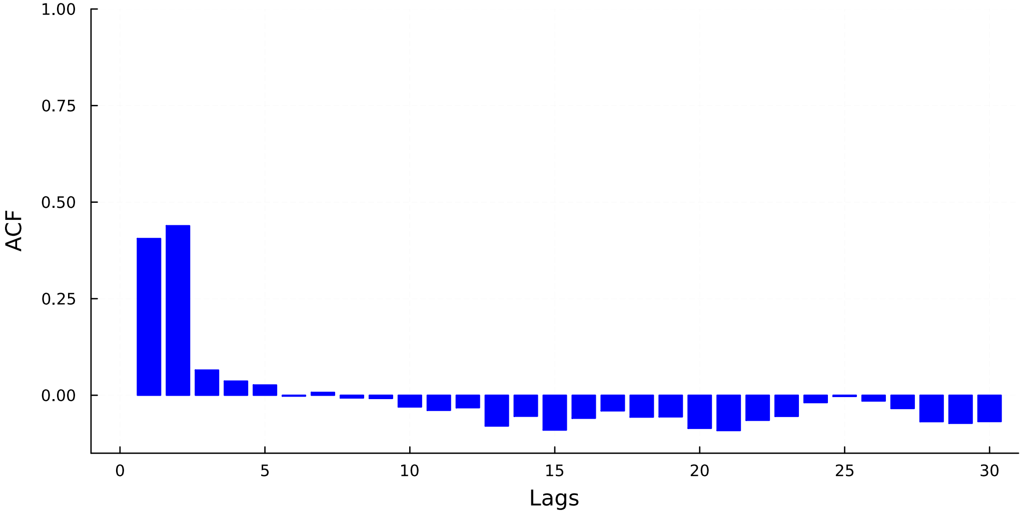
$$\rho_j = 0.37 \longrightarrow \gamma_2/\gamma_0$$

Moving Average Process: Simulated MA(2) Process



Author's simulations based on a Julia script using 1000 observations.

Moving Average Process: ACF of MA(2) Process



Author's simulations based on a Julia script using 1000 observations.

AR Processes

Autoregressive Process: AR(1)

The first-order autoregressive process AR(1) process is defined as

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t \quad \text{with} \quad \varepsilon_t \sim WN(0, \sigma^2)$$

Solve the first-order difference equation by backward substitution

$$\begin{aligned}
Y_t &= c + \phi Y_{t-1} + \varepsilon_t \\
&= c + \phi\left(c + \phi Y_{t-2} + \varepsilon_{t-1}\right) + \varepsilon_t \\
&= c + \phi\left(c + \phi\left(c + \phi Y_{t-3} + \varepsilon_{t-2}\right) + \varepsilon_{t-1}\right) + \varepsilon_t \\
&= \vdots \\
&= c \sum_{i=0}^{t-1} \phi^i + \sum_{i=0}^{t-1} \phi^i \varepsilon_{t-i} + \phi^t Y_0.
\end{aligned}$$

Autoregressive Process: AR(1)

If $|\phi| < 1$, the process is stationary since

$$\lim_{t\to\infty} \phi^t Y_0 \to 0, \tag{1}$$

Recall the property of an absolutely convergent geometric series

$$\sum_{i=0}^{\infty} \phi^i = (1-\phi)^{-1} \tag{2}$$

Using (1) and (2), we show that

AR(1) has $MA(\infty)$ representation

$$Y_t = c\sum_{i=0}^{t-1} \phi^i + \sum_{i=0}^{t-1} \phi^i \varepsilon_{t-i} + \phi^t Y_0 = \frac{c}{1-\phi} + \boxed{\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}}$$

Autoregressive Process: AR(1)

Unconditional mean

$$
\begin{aligned}
E(Y_t) &= E\left(\frac{c}{1-\phi} + \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}\right) \\
&= \frac{c}{1-\phi} + \sum_{i=0}^{\infty} \phi^i E(\varepsilon_{t-i}) = \frac{c}{1-\phi}
\end{aligned}
$$

They are all equal to zero

Unconditional variance

$$
\begin{aligned}
Var(Y_t) &= Var\left(\frac{c}{1-\phi} + \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}\right) \\
&= Var(\varepsilon_{t-i}) \textstyle\sum_{i=0}^{\infty} \phi^{2i} + \displaystyle\sum_{i=0}^{\infty} \sum_{\substack{j=0 \\ i \neq j}}^{\infty} \phi^{i+j} Cov(\varepsilon_{t-i}\varepsilon_{t-j}) = \frac{\sigma^2}{1-\phi^2}
\end{aligned}
$$

They are all equal to zero

convergent series
$1/(1-\phi^2)$

Autoregressive Process: AR(1)

First-order autocovariance

$$\begin{aligned} Cov(Y_t, Y_{t-1}) &= Cov(\mu + \phi Y_{t-1} + \varepsilon_t, Y_{t-1}) \\ &= \phi Var(Y_{t-1}) = \frac{\phi\sigma^2}{1-\phi^2} \end{aligned}$$

First-order autocorrelation

$$Cor(Y_t, Y_{t-1}) = \frac{Cov(Y_t, Y_{t-1})}{\sqrt{V(Y_t)}\sqrt{V(Y_{t-1})}} = \phi$$

Autoregressive Process: AR(1)

Higher-order autocovariance

$$
\begin{aligned}
Cov(Y_t, Y_{t-j}) &= Cov(\mu + \phi Y_{t-1} + \varepsilon_t, Y_{t-j}) \\
&= \phi Cov(Y_{t-1}, Y_{t-j}) \\
&= \phi^2 Cov(Y_{t-2}, Y_{t-j}) \\
&= \vdots \\
&= \phi^j Cov(Y_{t-j}, Y_{t-j}) = \frac{\phi^j \sigma^2}{1-\phi^2}
\end{aligned}
$$

Higher-order autocorrelation

$$
Cor(Y_t, Y_{t-j}) = \phi^j
$$

Autoregressive Process: AR(1)

Consider the following process

$$y_t = 0.5 + 0.8y_{t-1} + \varepsilon_t \quad \text{with} \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, 0.04).$$

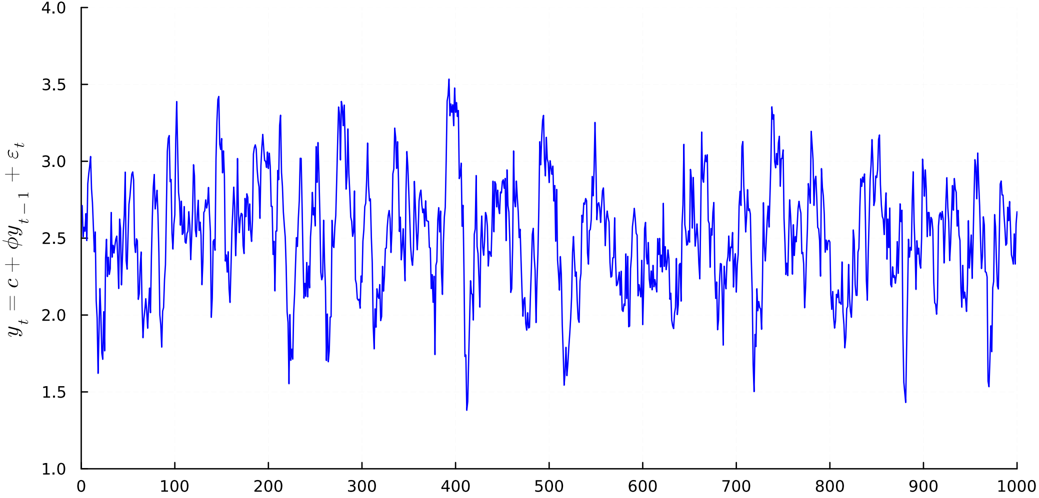
This process is weakly stationary as

$$\mu = 2.5 \longrightarrow c/(1-\phi)$$

$$\gamma_0 = 0.11 \longrightarrow \sigma^2/(1-\phi^2)$$

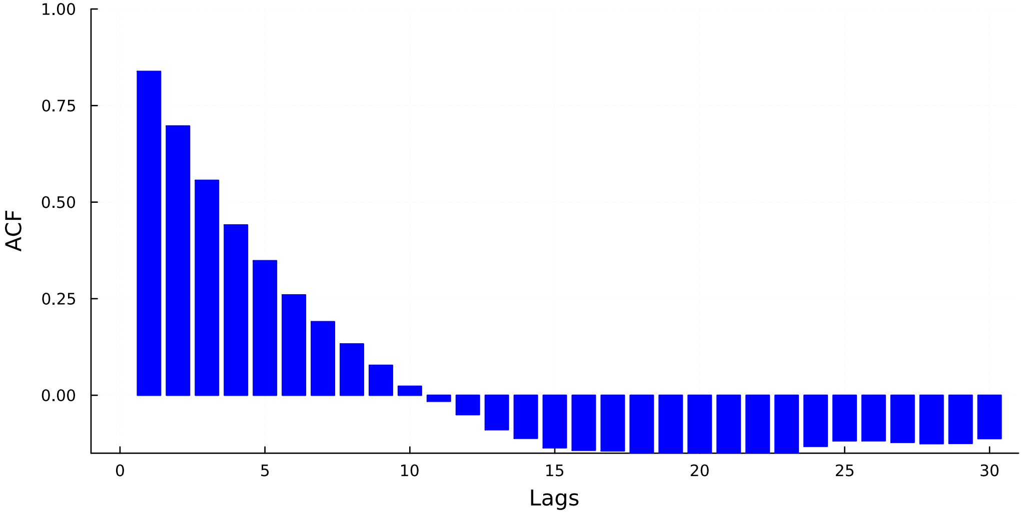
$$\rho_j = 0.80^j \longrightarrow \phi^j$$

Simulated AR(1) Process



Author's simulations based on a Julia script using 1000 observations.

ACF of AR(1) Process



Author's simulations based on a Julia script using 1000 observations.

A Special Case

If $c \neq 0$ and $\phi = 1$, we have the random walk with drift

$$Y_t = c + Y_{t-1} + \varepsilon_t \longrightarrow$$

By setting $c = 0$, we have the naïve random walk

By back-substitution, we obtain

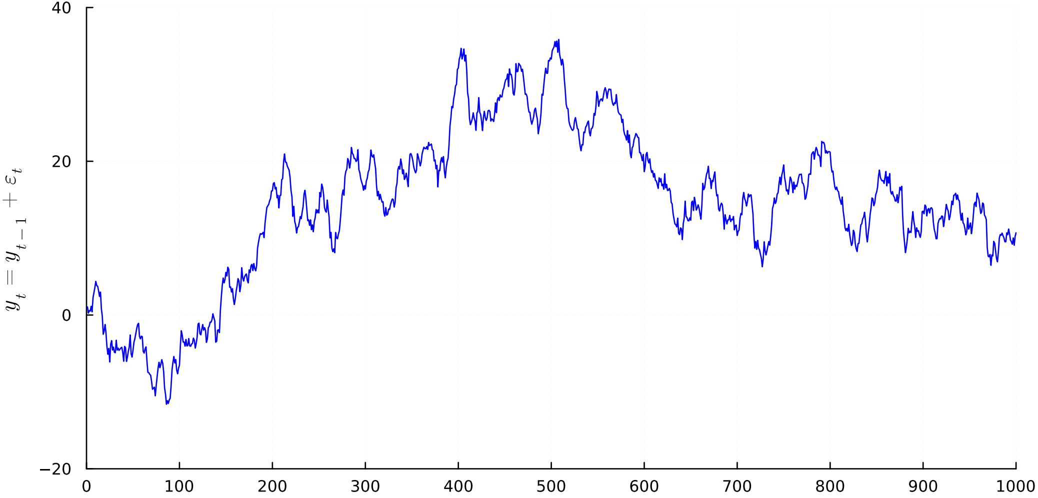
$$Y_t = Y_0 + tc + \sum_{s=1}^{t} \varepsilon_s,$$

This process is non-stationary as both mean and variance grow over time

$$\begin{aligned} E(Y_t) &= Y_0 + tc \\ Var(Y_t) &= \sum_{s=1}^{t} Var(\varepsilon_s) = t\sigma^2 \end{aligned}$$

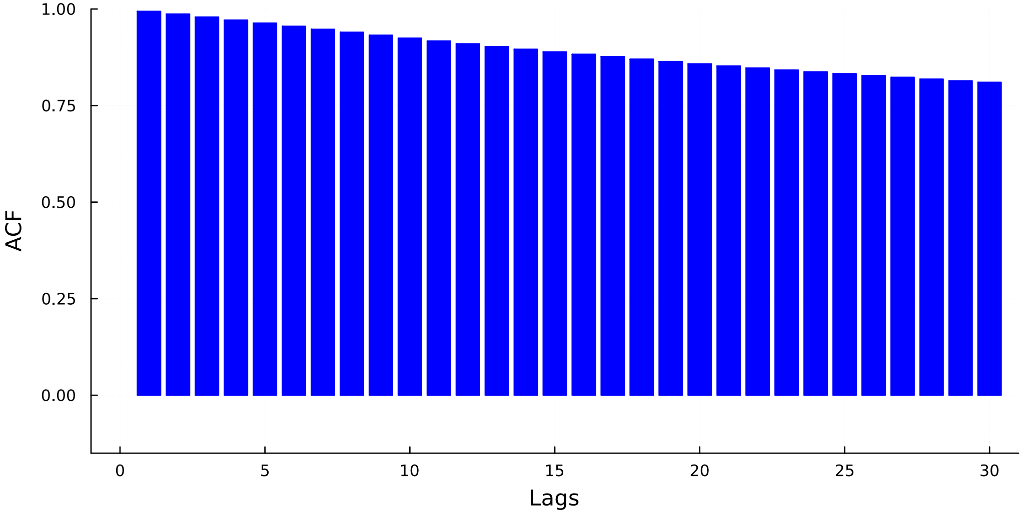
The impact of a single shock is permanent and never dissipates

Simulated Naïve Random Walk



Author's simulations based on a Julia script using 1000 observations.

ACF of a Naïve Random Walk



Author's simulations based on a Julia script using 1000 observations.

Other Cases

If $\phi > 1$, the process is explosive

- It grows exponentially over time, not a good description of most economic time series
- Typical exceptions are *hyperinflation episodes* or *asset price bubbles*.

If $\phi < -1$, the process is explosively oscillating

- Shocks alternate in sign and grow in magnitude.
- Such dynamics are rarely observed in economic time series.

Autoregressive Process: AR(p)

Take a p-**th order autoregressive** or $AR(p)$ process

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t \quad \text{with} \quad \varepsilon_t \sim WN(0, \sigma^2)$$

What are the key properties of this model?

- Mean?
- Variance?
- Autocovariance and Autocorrelation?

We must identify the stationarity conditions

Autoregressive Process: AR(p)

Rewrite the $AR(p)$ process as a $VAR(1)$ process

$$\underbrace{\begin{bmatrix} Y_t - c \\ Y_{t-1} - c \\ Y_{t-2} - c \\ \vdots \\ Y_{t-p+1} - c \end{bmatrix}}_{\boldsymbol{Y}_t} = \underbrace{\begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}}_{\Phi} \underbrace{\begin{bmatrix} Y_{t-1} - c \\ Y_{t-2} - c \\ Y_{t-3} - c \\ \vdots \\ Y_{t-p} - c \end{bmatrix}}_{\boldsymbol{Y}_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\boldsymbol{\varepsilon}_t}$$

This is a system of p-equations where

- The first equation is the $AR(p)$ process,
- Other equations are just identities.

Autoregressive Process: AR(p)

We have the following specification

$$\boldsymbol{Y}_t = \Phi \boldsymbol{Y}_{t-1} + \varepsilon_t$$

By recursive substitutions, we obtain

$$\begin{aligned}
\boldsymbol{Y}_t &= \Phi \left(\Phi \boldsymbol{Y}_{t-2} + \varepsilon_{t-1} \right) + \varepsilon_t \\
&= \Phi^2 \left(\Phi \boldsymbol{Y}_{t-3} + \varepsilon_{t-2} \right) + \Phi \varepsilon_{t-1} + \varepsilon_t \\
&= \vdots \\
&= \sum_{i=0}^{t-1} \Phi^i \varepsilon_{t-i} + \Phi^t \boldsymbol{Y}_0
\end{aligned}$$

The system is stationary if

$$\lim_{t \to \infty} \Phi^t \to 0 \longrightarrow$$

Since Φ is matrix,
what does it mean?

Autoregressive Process: AR(p)

Consider the eigenvalue decomposition

$$\Phi^i = Q\Lambda^i Q^{-1}$$

where

$$\Lambda^i = \begin{bmatrix} \lambda_1^i & 0 & \dots & 0 \\ 0 & \lambda_2^i & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p^i \end{bmatrix} \longrightarrow$$ diagonal matrix of eigenvalues

and

$$Q \longrightarrow$$ matrix of eigenvectors

Autoregressive Process: AR(p)

The system is stationary if

$$\lim_{t\to\infty} \Phi^t \to 0 \quad \Longleftrightarrow \quad \lim_{t\to\infty} \Lambda^t \to 0$$

which requires

$$|\lambda_i| < 1 \quad \text{for all} \quad i. \longrightarrow$$

The eigenvalues lie inside
the unit circle

The eigenvalues λ_i can be viewed as the roots of the characteristic polynomial

$$\lambda^p - \phi_1\lambda^{p-1} - \phi_2\lambda^{p-2} - \cdots - \phi_{p-1}\lambda - \phi_p = 0.$$

Each root describes how a shock propagates over time:

- $|\lambda_i| < 1$: shocks decay $\Rightarrow$ stable dynamics,
- $|\lambda_i| = 1$: shocks persist $\Rightarrow$ unit root behavior,
- $|\lambda_i| > 1$: shocks amplify $\Rightarrow$ explosive dynamics.

Example 1: AR(2) Process

Consider the following $AR(2)$ process

$$y_t = 0.5 + 0.6 y_{t-1} + 0.2 y_{t-2} + \varepsilon_t.$$

Rewrite the $AR(2)$ process as a $VAR(1)$ process

$$\underbrace{\begin{bmatrix} y_t - 0.5 \\ y_{t-1} - 0.5 \end{bmatrix}}_{\boldsymbol{y}_t} = \underbrace{\begin{bmatrix} 0.6 & 0.2 \\ 1 & 0 \end{bmatrix}}_{\Phi} \underbrace{\begin{bmatrix} y_{t-1} - c \\ y_{t-2} - c \end{bmatrix}}_{\boldsymbol{y}_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}}_{\boldsymbol{\varepsilon}_t}$$

Note: we only care about the matrix Φ in practice.

Example 1: AR(2) Process

Take the eigenvalue decomposition of Φ as

$$\underbrace{\begin{bmatrix} 0.6 & 0.2 \\ 1 & 0 \end{bmatrix}}_{\Phi} = \underbrace{\begin{bmatrix} 0.97 & 0.23 \\ -0.77 & 0.64 \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} 0.84 & 0 \\ 0 & -0.24 \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} 0.97 & 0.23 \\ -0.77 & 0.64 \end{bmatrix}^{-1}}_{Q^{-1}}$$

We thus have a stationary process since

$$\lambda_1 = 0.84 \longrightarrow |\lambda_1| < 1$$
$$\lambda_2 = -0.24 \longrightarrow |\lambda_2| < 1$$

Example 2: AR(2) Process

Consider the following $AR(2)$ process

$$y_t = 0.5 + 0.6 y_{t-1} + 0.4 y_{t-2} + \varepsilon_t.$$

Rewrite the $AR(2)$ process as a $VAR(1)$ process

$$\underbrace{\begin{bmatrix} y_t - 0.5 \\ y_{t-1} - 0.5 \end{bmatrix}}_{\boldsymbol{Y}_t} = \underbrace{\begin{bmatrix} 0.6 & 0.4 \\ 1 & 0 \end{bmatrix}}_{\Phi} \underbrace{\begin{bmatrix} y_{t-1} - c \\ y_{t-2} - c \end{bmatrix}}_{\boldsymbol{Y}_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}}_{\boldsymbol{\varepsilon}_t}$$

Note: we only care about the matrix Φ in practice.

Example 2: AR(2) Process

Take the eigenvalue decomposition of Φ as

$$\underbrace{\begin{bmatrix} 0.6 & 0.4 \\ 1 & 0 \end{bmatrix}}_{\Phi} = \underbrace{\begin{bmatrix} 0.93 & 0.37 \\ -0.71 & 0.71 \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} 1.00 & 0 \\ 0 & -0.40 \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} 0.93 & 0.37 \\ -0.71 & 0.71 \end{bmatrix}^{-1}}_{Q^{-1}}$$

We thus have a non-stationary process since

$$\lambda_1 = 1.00 \longrightarrow |\lambda_1| = 1$$
$$\lambda_2 = -0.40 \longrightarrow |\lambda_2| < 1$$

Autoregressive Process: AR(p)

Quick rules for the stability of a p-order system

- A necessary condition for all $|\lambda_i| < 1 : \sum_{i=1}^{p} \phi_i < 1$,
- A sufficient condition for all $|\lambda_i| < 1 : \sum_{i=1}^{p} |\phi_i| < 1$,
- At least one root equals unity if $\sum_{i=1}^{p} \phi_i = 1$,
- a unit root process has one or more roots equals unity.

There exists other ways to derive stationarity conditions.

Autoregressive Process: AR(p)

Unconditional mean

$$E\left(Y_t\right) = \frac{c}{1-\sum_{i=1}^{p}\phi_i}$$

Unconditional variance

$$V\left(Y_t\right) = \frac{\sigma^2}{1-\sum_{i=1}^{p}\phi_i^2}$$

Autocovariances are computed by solving the Yule-Walker equations.

What is an ARMA Process?

ARMA Models

The **moving-average autoregressive process** or $ARMA(p,q)$ is

$$Y_t = c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

with

$$\varepsilon_t \sim WN(0, \sigma^2)$$

$ARMA(p,q)$ models can arise from the aggregation of simple time series

- High order $ARMA$ are rarely used for economic/financial data.
- $ARMA$ with p and q less than 3 are generally sufficient for most economic/financial data.

How to Select p and q?

Parsimony

Occam's razor (or **lex parsimoniae**)

- Law of parsimony attributed to William of Ockham (1285–1349),
- Having two competing modes that give the same prediction, the simpler one is always better,
- Simpler explanations, other things being equal, are generally better than more complex ones.

Complex models

- Can track the data quite well over the historical period for which parameters are estimated.
- Often perform poorly when used for out-of-sample forecasting.

Box and Jenkins (1976)

The most common approach for time-series model selection

- Transform the data to induce stationarity,
- Make an initial guess for p and q for an $ARMA(p, q)$,
- Estimate the parameters p and q for an $ARMA(p, q)$,
- Perform diagnostic analysis to confirm the model is consistent with the data.

The initial guess for p and q requires both

- Autocorrelation function (ACF),
- Partial autocorrelation function (PACF).

Selection Procedure

The autocorrellation function (ACF) is the plot of ρ_j against j

- If data follow an $MA(q)$ process, then $\rho_j = 0$ for $j > q$.
- If data follow an $AR(p)$ process, then ρ_j gradually decays toward zero.

Selection Procedure

What is the partial autocorrelation?

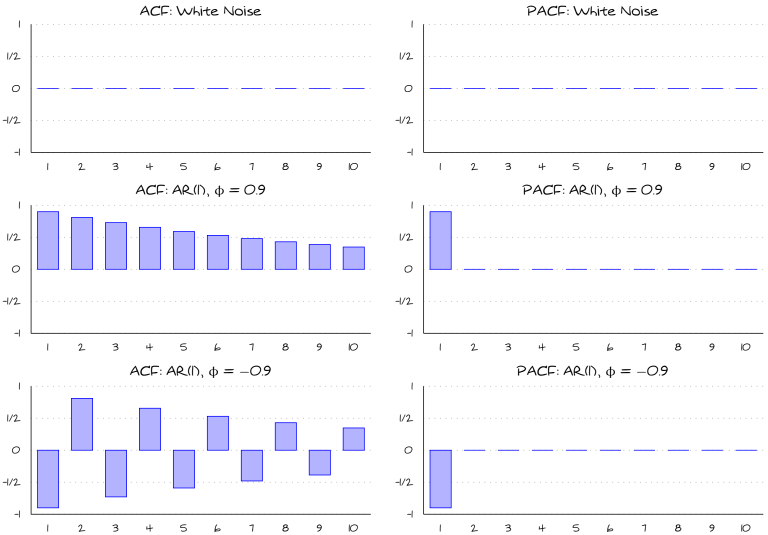
- The j-th partial autocorrelation φ_j measures the direct effect of Y_{t-j} on Y_t,
- We need to remove the effects of $Y_{t-1}, \ldots, Y_{t-j+1}$ via the regression

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} \ldots + \varphi_{p-1} Y_{t-j+1} + \varphi_j Y_{t-j} + \varepsilon_t$$

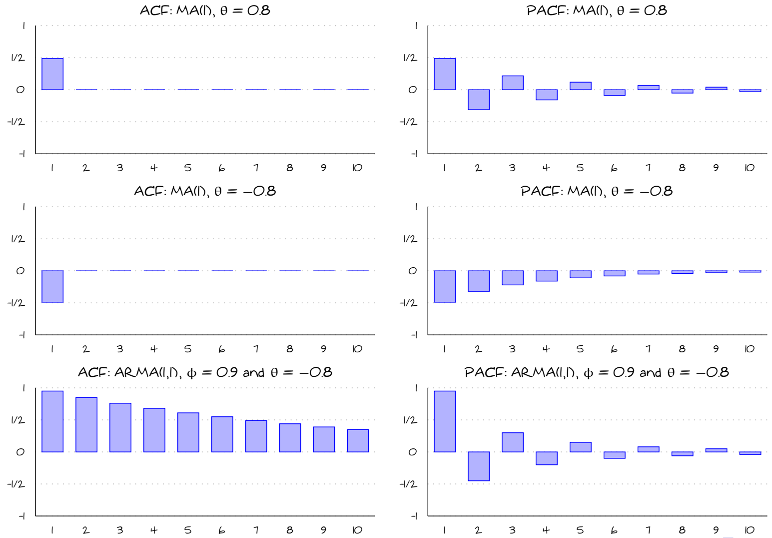
The partial autocorrelation function is the plot of φ_j against j

- If data were generated by an $AR(p)$ process, then $\varphi_j = 0$ for $j > p$,
- If data were generated by $MA(q)$ process, then φ_j will asymptotically approach zero.

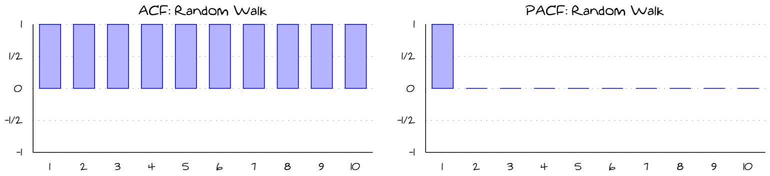
ACFs and PACFs: A Summary



ACFs and PACFs: A Summary



ACFs and PACFs: A Summary



ACFs and PACFs: A Summary

Process	ACF ρ_j	PACF φ_j
WN	zero for $j > 0$	zero for $j > 0$
AR(1)	decays towards 0	zero for $j > 1$
AR(p)	decays towards 0	zero for $j > p$
MA(1)	zero for $j > 1$	decays towards 0
MA(q)	zero for $j > q$	decays towards 0
ARMA(p,q)	decays towards 0	decays towards 0

Information Criteria for Model Selection

Akaike Information Criterion

$$AIC = -2\ln(L) + 2K$$

Schwarz Information Criterion

$$BIC = -2\ln(L) + K\ln(T)$$

- L: maximized value of the likelihood function.
- K: Number of parameters (e.g., $p + q + 1$).
- T: Number of observations.

Decision Rule

- Information criteria trade off goodness of fit versus complexity,
- Choose the model with the lowest AIC or BIC value.

A Test for Serial Correlation

The **Ljung-Box** or **Q-statistic**

- Guess (p, q) and fit an ARMA(p, q) model,
- Compute the first s autocorrelations of the residuals,
- Test $H_0 : \rho_1 = \rho_2 = \ldots = \rho_s = 0$ (and homoskedasticity) using

Under H_0, the residulas are uncorrlated up to s

$$Q = T\,(T+2)\sum_{i=1}^{s}\frac{\rho_i^2}{T-i} \sim \chi_s^2$$

A rule of thumb is s = ln(T) with s > p+q and T = sample data

Decision Rule:

- p-value < 0.05, reject H_0, i.e., residuals are not uncorrelated,
- p-value > 0.05, do not reject H_0, i.e., residuals look like white noise.

Spurious Regressions

Spurious Regressions

When you run a regression with weakly stationary variables

- Take two stationary yet independent series, say x and y
- Regress one series on the other one,
- β will be close to zero and insignificant,
- R^2 will also be close to zero.

This is NOT the case when x and y are independent yet nonstationary

Houston, we have a problem!

Spurious Regressions

Take two nonstationary yet independent series

- Regress one series on the other one,
- β will be not converge to zero and likely be significant,
- R^2 will be very high.

Why is spurious?

- The regression statistically proves a relationship that does not exist.
- It is a statistical mirage caused by the trending nature of the data.

A regression is spurious when a random walk
is regressed onto another independent random walk.

Spurious Regressions

Any red flag?

- The regression residuals are highly persistent and contain a unit root (*nonstationary*),
- Results changes dramatically after the lagged dependent, variable is added as new regressor,
- The previously significant coefficient will become insignificant.

What is the lesson?

- Always check the stationarity of the residuals,
- The regression is spurious if the residuals are nonstationary,
- We cannot reject the null hypothesis of the unit root test.

Just because two series move together does not mean they are related!

Spurious Regressions: The Cause

What causes it?

Nonstationary series contain stochastic trends: Take the RW for example

$$y_t = y_{t-1} + \varepsilon_t$$

Rewrite it using its MA representation as

$$y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$$

What is the intuition?

- The current value y_t remembers every single shock from the past.
- The sequence of shocks causes short-lived trends (local trends).

A spurious regression happens when there are similar local trends!

An Example

```julia
# Set the seed for reproducibility
Random.seed!(9876543)

# Number of observations
nstep = 10000

# Parameters for y
mu_y    = 0      # constant
sigma_y = 1      # Standard deviation of the noise

# Parameters for x
mu_x    = 0      # constant
sigma_x = 1      # Standard deviation of the noise

# Simulate the process
ey = mu_y .+ sigma_y .* randn(nstep)
ex = mu_x .+ sigma_x .* randn(nstep)
```

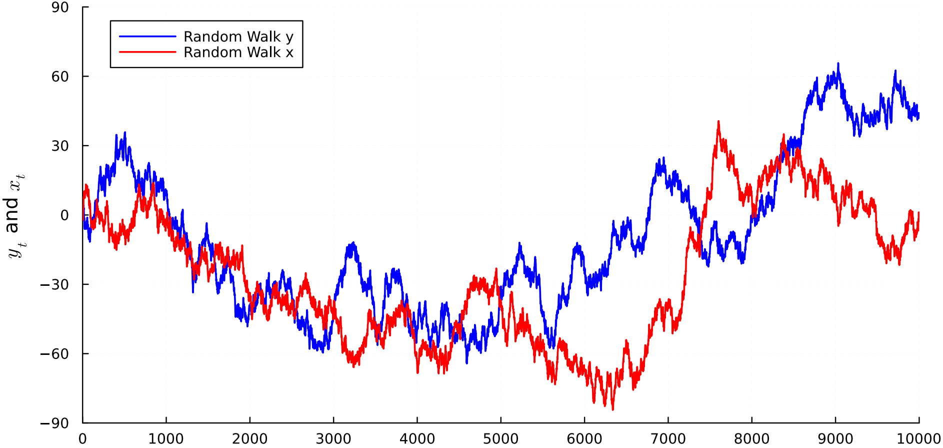
An Example

```
# Generate two random walks
rw_y = cumsum(ey)
rw_x = cumsum(ex)

# Create a DataFrame for the random walks
df = DataFrame(rw_y = rw_y, rw_x = rw_x)

# Perform a regression of rw_y on rw_x
model = lm(@formula(rw_y ~ rw_x), df)

# Print regression summary and R-squared
println(model)
println("\nR-squared: $r_squared")
```

An Example

```
# Generate two random walks
rw_y = cumsum(ey)
rw_x = cumsum(ex)

# Create a DataFrame for the random walks
df = DataFrame(rw_y = rw_y, rw_x = rw_x)

# Perform a regression of rw_y on rw_x
model = lm(@formula(rw_y ~ rw_x), df)

# Print regression summary and R-squared
println(model)
println("\nR-squared: $r_squared")

#Plot
plot(1:nstep,  rw_y, lw=2)
plot!(1:nstep, rw_x, lw=2)
ylabel!(L"$y_{t}$ and $x_{t}$")

```

An Example



Based on a Julia script using $y_t = y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N(0,1)$ and $x_t = x_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N(0,1)$

Cointegration

Cointegration

What is cointegration? y and x are cointegrated if

- y is nonstationary and x is nonstationary,
- There exists a linear combination of y and x that is stationary
- In short, the regression residuals are stationary.

When y and x are cointegrated

- We can regress y onto x
- The coefficient will converge to the true value super fast,
- The OLS estimator is superconsistent

Houston, we do not have a problem!

Cointegration

But there is a technical issue

- We can construct the t-test as usual,
- However, we cannot compare it to the critical values of t distribution or normal distribution.

Stock and Watson (1993) suggest using the dynamic OLS

$$y_t = \alpha + \beta x_t + \sum_{i=-p}^{p} \gamma \Delta x_{t-i} + \varepsilon_t$$

where we add leads and lags of changes in x_t as new regressors.

Now β is asymptotically normally distributed so that
the normal distribution can be used in hypothesis testing.

Unit Root Tests

Nonstationary Time Series

Consider the stylized trend-cycle decomposition of a time series

$$
\begin{aligned}
Y_t &= D_t + z_t \\
\text{Linear Trend} \longleftarrow \boxed{D_t} &= \kappa + \delta t \\
\text{AR(1) component} \longleftarrow \boxed{z_t} &= \phi z_{t-1} + \underbrace{\boxed{\varepsilon_t}}_{WN(0,\sigma^2)}
\end{aligned}
$$

We have the following cases

- if $|\phi| < 1$ and $\delta = 0 \longrightarrow Y_t$ is stationary,
- If $|\phi| < 1 \longrightarrow Y_t$ is trend-stationary (deterministic trend),
- If $|\phi| = 1 \longrightarrow Y_t$ is nonstationary (a stochastic trend).

Nonstationary Time Series

Unit root tests

$$H_0 : \phi = 1 \quad \text{nonstationary}$$

$$H_A : |\phi| < 1 \quad \text{trend stationary}$$

pick one H_A $\longleftarrow$

$$H_A : |\phi| < 1 \quad \& \quad \delta = 0 \quad \text{stationary}$$

Practical limitations of unit root tests

- Nonstandard distributions $\longrightarrow$ critical values are simulated,
- Critical values change with constants and trends $\longrightarrow$ choose your H_A wisely!

Case I: Constant Only

The test regression is

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t$$

which includes a constant to capture the nonzero mean under the alternative.

Test hypothesis

$$\begin{array}{lcll} H_0 & : & \phi = 1 & \Longrightarrow Y_t \sim I(1) \quad \text{without a drift} \\ H_1 & : & |\phi| < 1 & \Longrightarrow Y_t \sim I(0) \quad \text{with non-zero mean} \end{array}$$

Appropriate for non-trending economic and financial series like interest rates and exchange rates.

Case II: Constant and Time Trend

The test regression is

$$y_t = c + \delta t + \phi y_{t-1} + \varepsilon_t$$

which includes a constant and a deterministic trend.

Test hypothesis

$$\begin{aligned} H_0 &: \quad \phi = 1 \qquad \Longrightarrow y_t \sim I(1) \quad \text{with drift} \\ H_1 &: \quad |\phi| < 1 \quad \Longrightarrow y_t \sim I(0) \quad \text{with time trend} \end{aligned}$$

Appropriate for trending time series like macroeconomic aggregates like real GDP and inflation.