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 Eeconometrics for the 21st Century*.

Time and Data Collection

Whenever we collect data, time often plays an important role

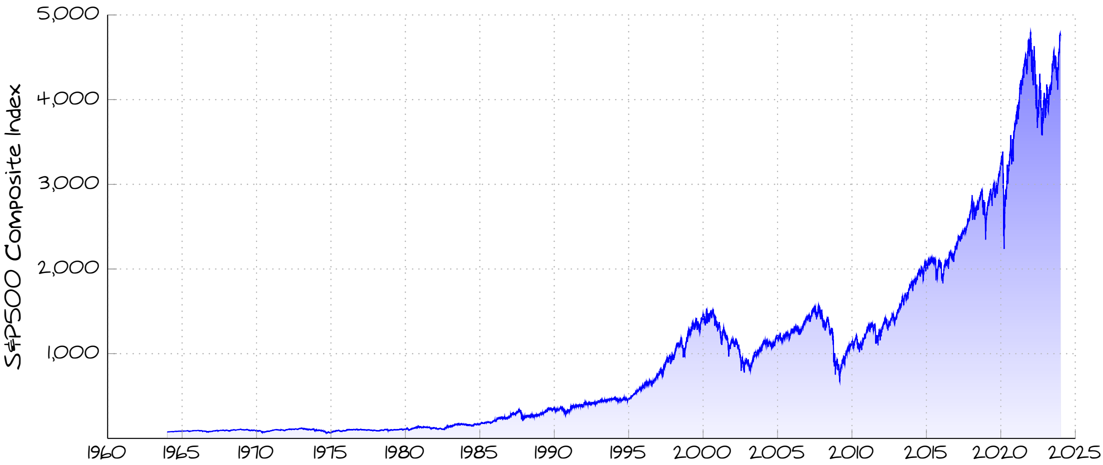
- This is true for many economic and financial data,
- Asset returns, inflation rate, economic growth, and many others.

.

With time-series analysis, we explore how most economic and financial data evolve over time

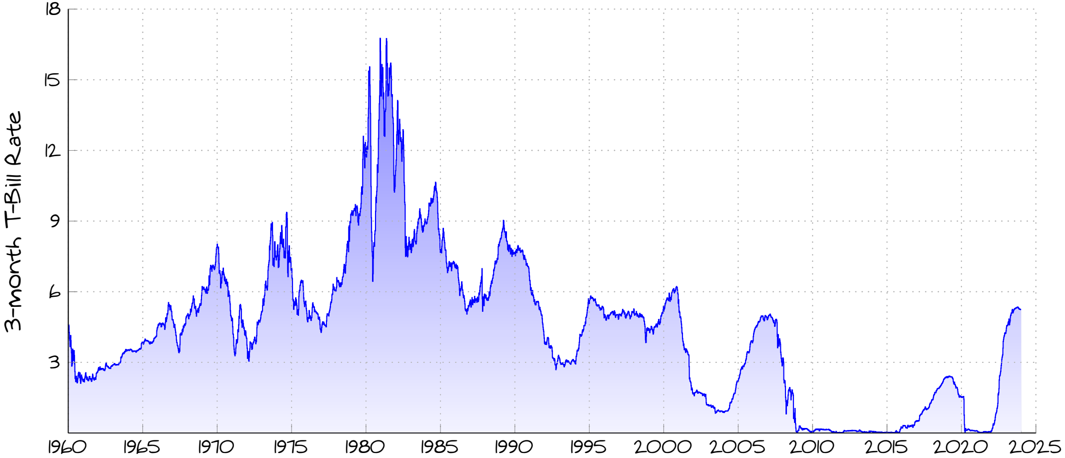
- We can differentiate between deterministic and stochastic patterns,
- We can attempt to forecast future data based on historical data.

S&P500 Index



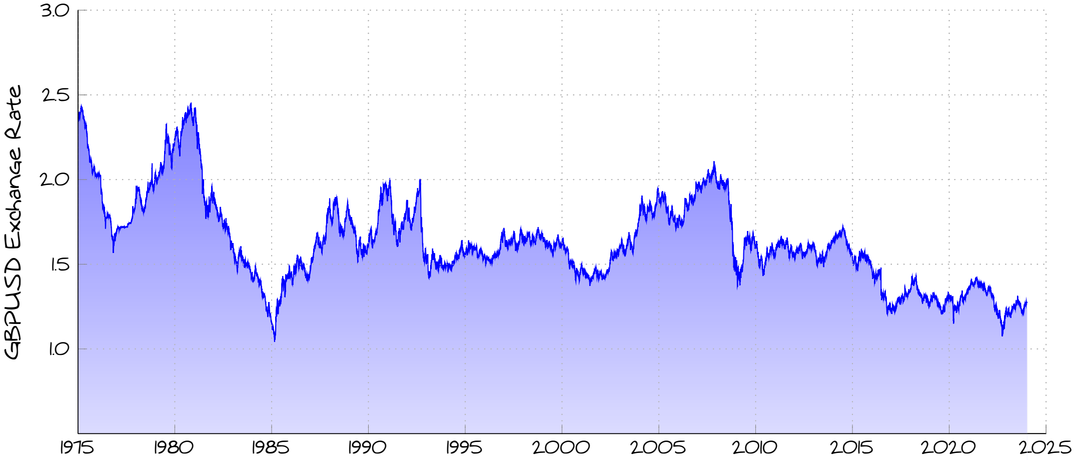
Data source: Datastream.

3-month Treasury Bill Rate



Data source: Datastream.

GBPUSD Exchange Rate



Data source: Datastream.

What is a Time Series?

A 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 as N 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 realization of a stochastic process as

$$(y_1, y_2, \ldots, y_T) \longrightarrow \text{or simply as } \{y_t\}$$

Economic and financial data are observed at different frequencies (e.g., daily, monthly, etc.),

- Collecting a time-series of returns is equivalent to observing a single realization of a stochastic process (*non-experimental data*),
- The observations of a time series, however, are close in calender time and thus dependent over time (*serially correlated*),
- Because of this dependence structure, we cannot rely on the distributional theory used for cross-sectional regressions as observations cannot be divided into independent groups.

An Example

Simulate a random walk consisting of 1000 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.

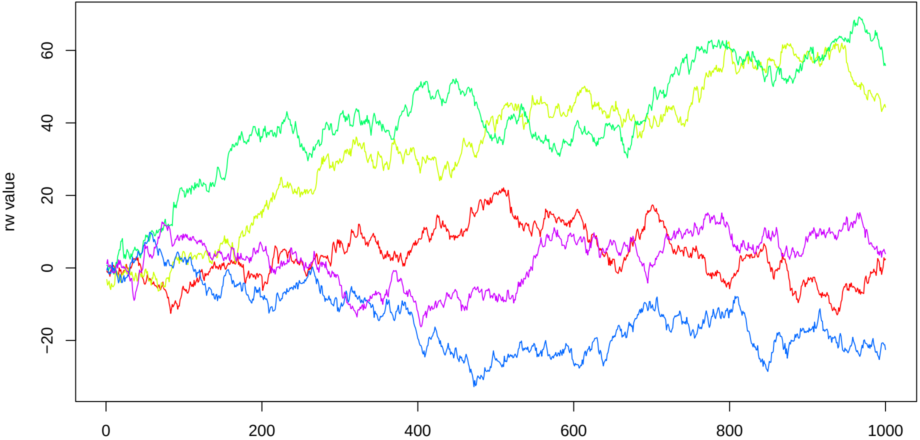
An Example

```
# Set the seed for reproducibility
set.seed(9876543)

# Control variables
nstep       = 1000    # Number of observations for each series
nsim        = 5       # 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       = matrix(0, nrow = nstep, ncol = nsim)

# Set the starting value
y0         = 0

```

An Example

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

        for (i in 1:nstep) {
          shock = rnorm(1, mean = mu, sd = sigma)

          if (i == 1) {
            y[i] = y0 + shock
          } else {
            y[i] = y[i-1] + shock
          }
        }

    ymat[,sim] = y
  }

```

An Example



Author's simulations based on an R script using $Y_{i,t} = Y_{i,t-1} + \varepsilon$.

What are the Properties of a Time Series?

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 stochastic process changes frequently, we have no hope.

Stationarity is a probabilistic measure of *regularity* (or constancy)

- 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,
- Weak and strong stationarity are related but not nested.

Weak Stationarity

What does weak stationary mean?

- The dynamics of a time-series is contaminated by shocks (e.g., news, events, etc),
- A shock, however, is transitory and dissipate over time (e.g., lose its energy),
- A shock, in other words, at some point will no longer 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.

Weak Stationarity

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

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

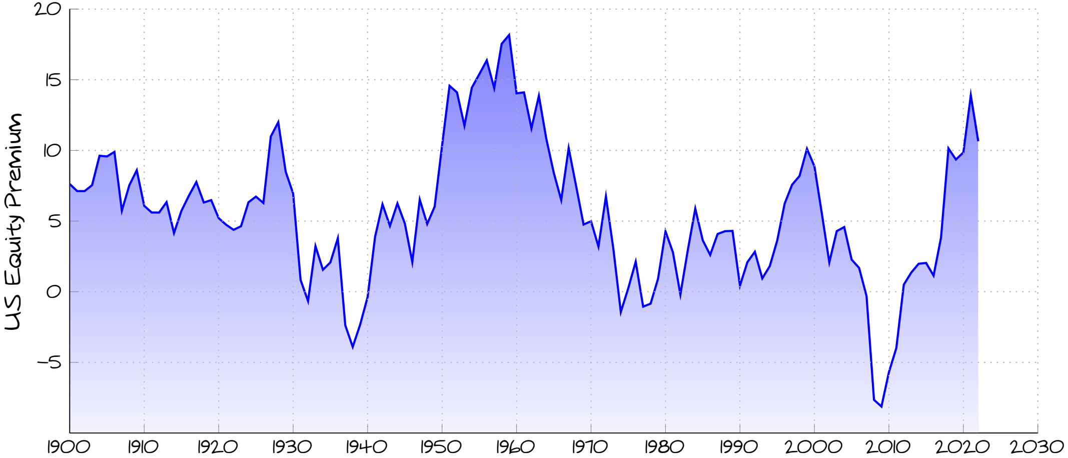
Weak Stationarity requires that the

- 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

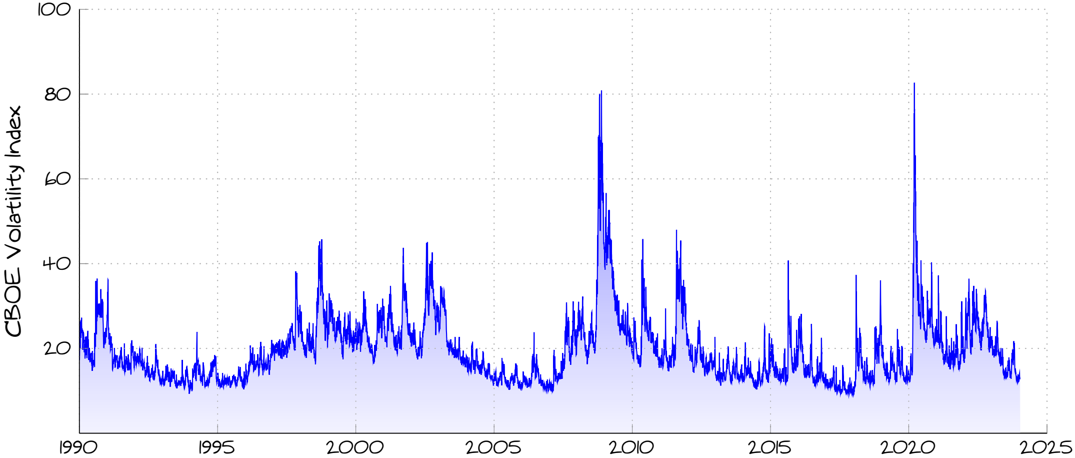
- A process may have predictable varying conditional mean/variance.

US Equity Risk Premium



Data source: Global Financial Data.

CBOE Volatility 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 yields the autocorrelation function (ACF)

- A summary of the linear dependence of $\{y_t\}$.

Weak Stationarity: Example 1

Consider the following process

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

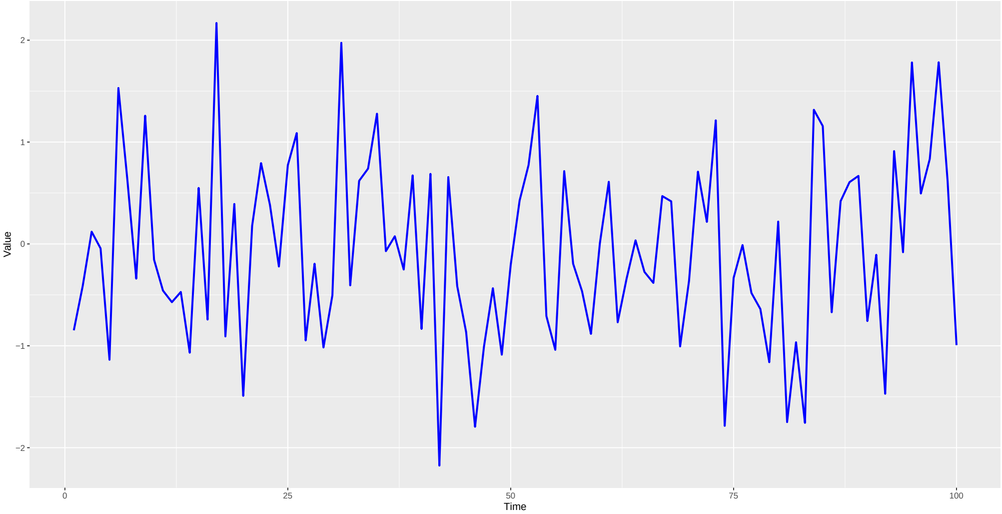
This process is weakly stationary as

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

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

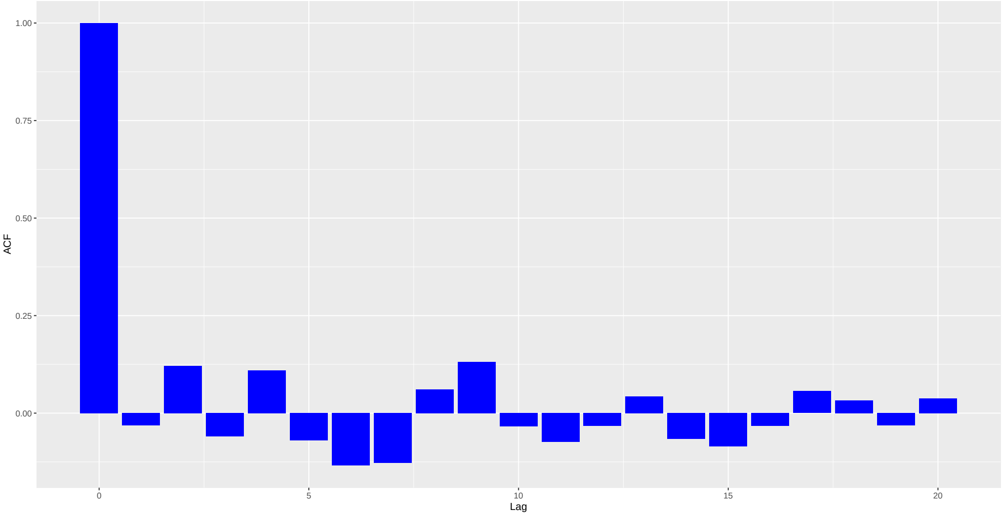
$$\gamma_j = 0 \longrightarrow \text{for } j \geq 1$$

Weak Stationarity: Simulated iid Process



Author's simulations based on an R script.

Weak Stationarity: ACF of an iid Process



Author's simulations based on an R script.

Weak Stationarity: Example 2

Consider the following process

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

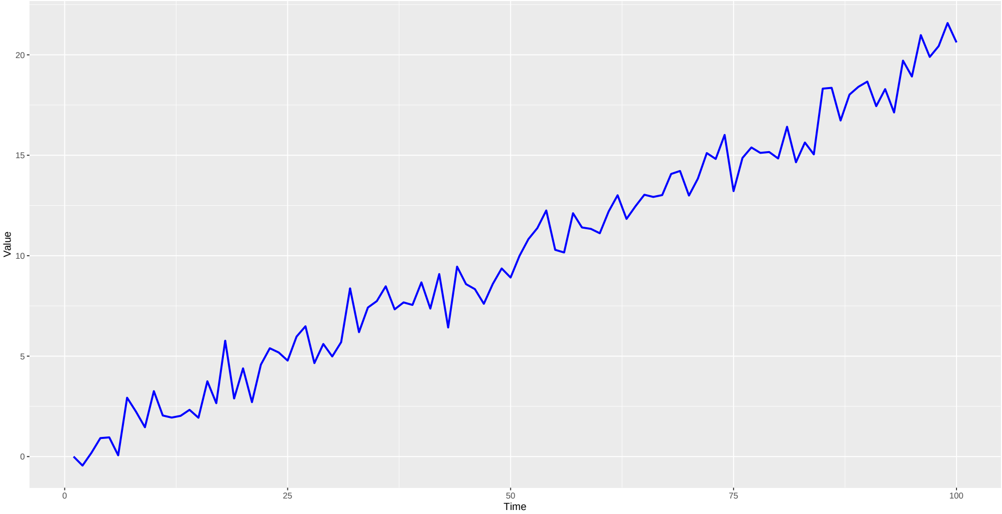
This process is not weakly stationary as

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

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

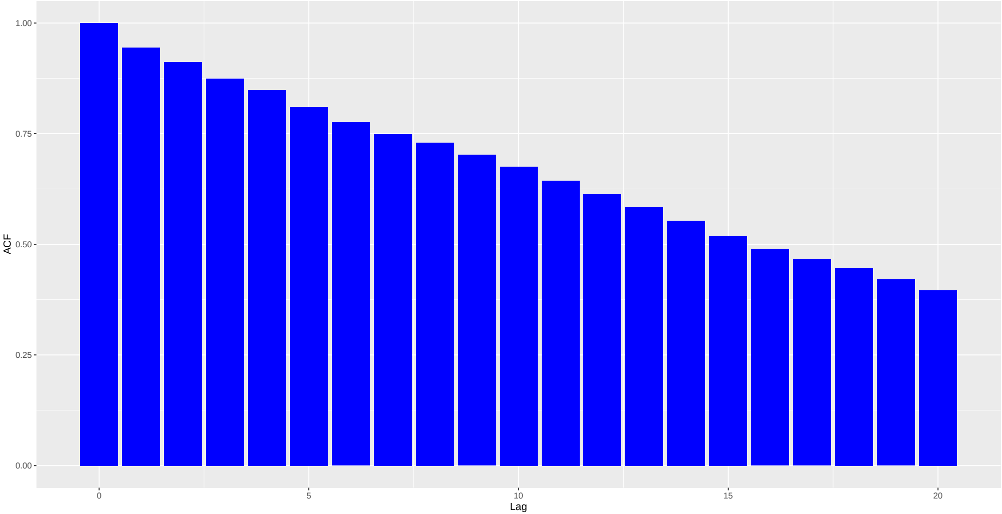
This is a trend-stationary (TS) process, i.e., stationary after we remove the trend component.

Weak Stationarity: Simulated TS Process



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

Weak Stationarity: ACF of a TS Process



Author's simulations based on an R 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 (e.g., for different sub-samples),
- We assume no distribution but only require that the probability distribution is the same.

A stochastic process $\{Y_t\}$ is strictly stationary if the joint distribution of $\{Y_t, \ldots, Y_{t+h}\}$ is the same as $\{Y_{t+\tau}, \ldots, Y_{t+\tau+h}\}$. Using the cumulative distribution, we have

$$F(y_t, \ldots, y_{t+h}) = F(y_{t+\tau}, \ldots, y_{t+\tau+h}) \quad \text{for all } \tau$$

- The joint distribution depends on h and not on t (*time invariant*).
- A *iid* sample from a Cauchy distribution is strictly stationary but not weakly stationary as its variance is infinite. Strong stationarity does not require finite variance.

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 drawn $t(0, 1, \nu)$ with $\nu = 2$ is strictly stationary but not weakly stationary,
- A normally distributed sample is both strictly and weakly stationary.

Ergodicity

What does Ergodicity mean?

- Ergodicity is a generalization of the *Law of Large Numbers*,
- Ergodicity implies that serial dependence vanishes asymptotically.

Ergodic Theorem

- If a stationary process $\{Y_t\}$ is ergodic and its k^{th} moment μ_k is finite, then averages will converge to their expectations

$$T^{-1}\sum_{t=1}^{T} Y_t^k \xrightarrow{p} \mu_k$$

- Sample moments converge in probability to population moments as errors vanish

$$\widehat{\mu} \xrightarrow{p} \mu \qquad \widehat{\gamma}_j \xrightarrow{p} \gamma_j \qquad \widehat{\rho}_j \xrightarrow{p} \rho_j$$

Does Stationarity imply Ergodicity?

Consider the following process

$$Y_t = \alpha + \varepsilon_t \quad \text{with} \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \quad \text{and} \quad \alpha \sim \mathcal{N}(0, \sigma^2)$$

- α and ε_t are independent, and α is drawn only once.

The process is weakly stationary as

$$\begin{aligned} \mu &= E(\alpha) + E(\varepsilon_t) &&= 0 \\ \gamma_0 &= V(\alpha) + V(\varepsilon_t) &&= s^2 + \sigma^2 \\ \gamma_j &= E[(\alpha + \varepsilon_t)(\alpha + \varepsilon_{t-j})] &&= s^2 \end{aligned}$$

The process is not ergodic as it converges in probability to α not 0

$$T^{-1} \sum_{t=1}^{T} Y_t = T^{-1} \sum_{t=1}^{T} (\alpha + \varepsilon_t) = \alpha + T^{-1} \sum_{t=1}^{T} \varepsilon_t \to \alpha \neq 0$$

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

White Noise

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

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

- 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.

ARMA Models

Autoregressive moving average (ARMA) processes, central to time-series analysis, consist of

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

What is a Moving-Average Process?

Moving Average Process: MA(1)

Take a **first-order moving average** or $MA(1)$ process

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

- 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

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

Higher-order autocovariance

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

Moving Average Process: MA(1)

First-order autocorrelation

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

Higher-order autocorrelation

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

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

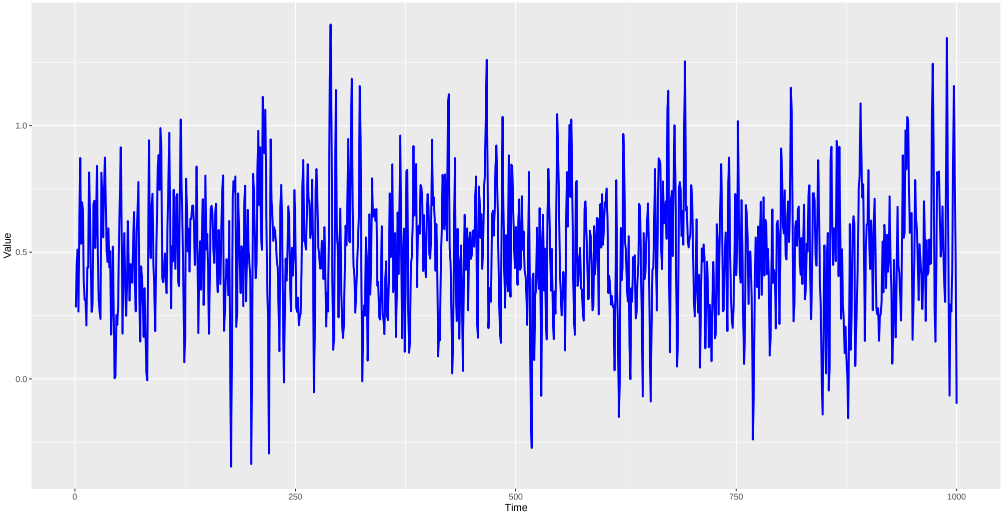
$$\mu = 0.5 \longrightarrow \text{constant}$$

$$\gamma_0 = 0.66 \longrightarrow \sigma^2(1+\theta^2)$$

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

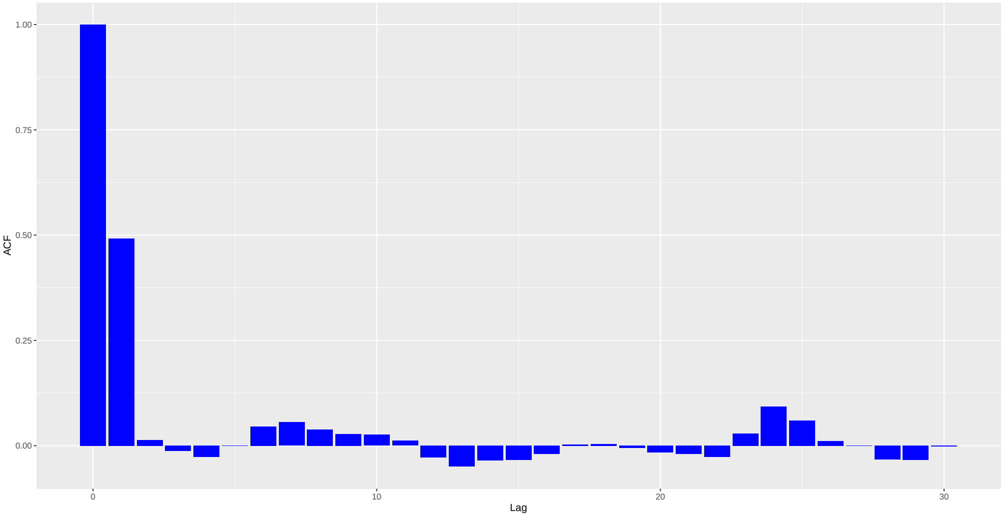
$$\rho_j = 0 \longrightarrow \text{for } j > 1$$

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

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

This process is weakly stationary as

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

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

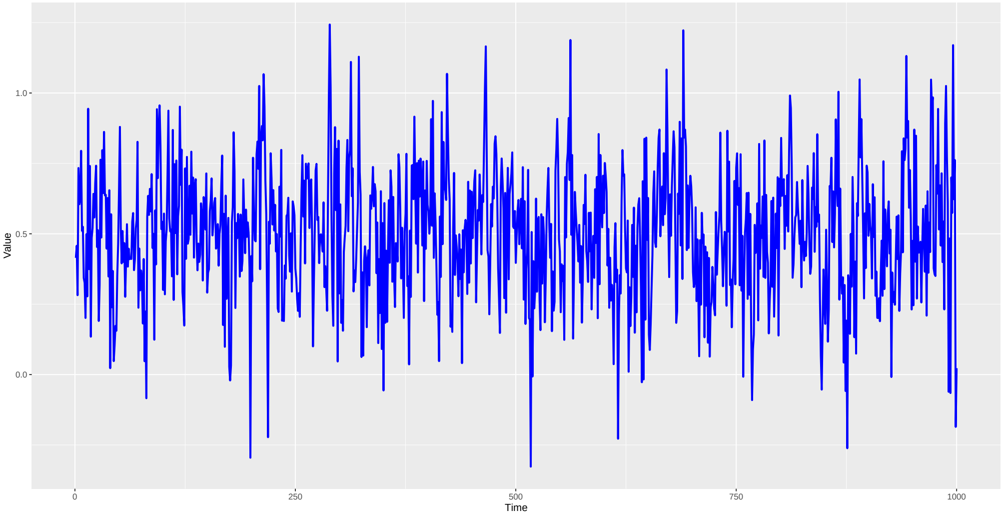
$$\gamma_1 = 0.018 \longrightarrow \sigma^2(\theta_1 + \theta_1\theta_2)$$

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

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

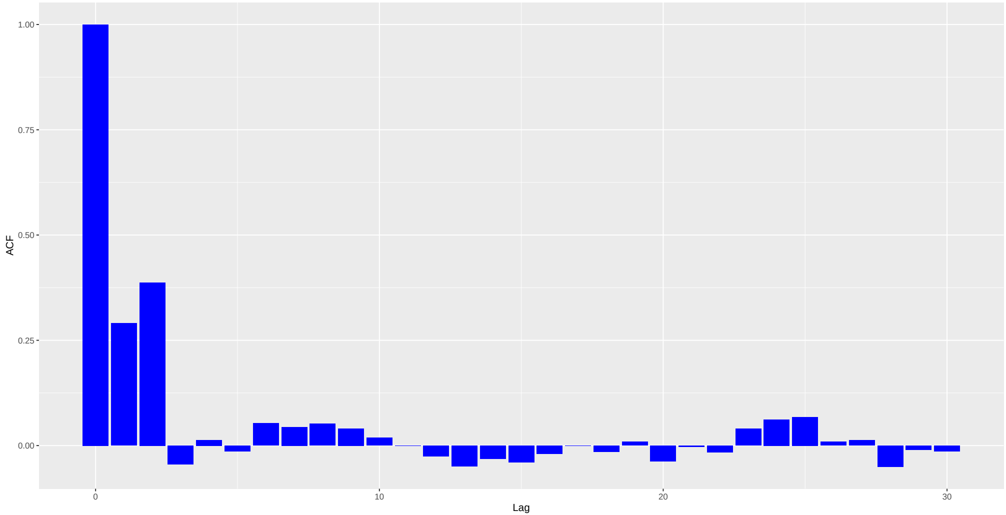
$$\rho_j = 0.37 \longrightarrow \gamma_2/\gamma_0$$

Moving Average Process: Simulated MA(2) Process



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

Moving Average Process: ACF of MA(2) Process



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

What is an Autoregressive Process?

Autoregressive Process: AR(1)

First-order autoregressive process or AR(1) process

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

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

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

the AR(1) process has an $MA(\infty)$ representation when $|\phi| < 1$.

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

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

Higher-order autocovariance

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

Autoregressive Process: AR(1)

First-order autocorrelation

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

Higher-order autocorrelation

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

Autoregressive Process: AR(1)

Consider the following process

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

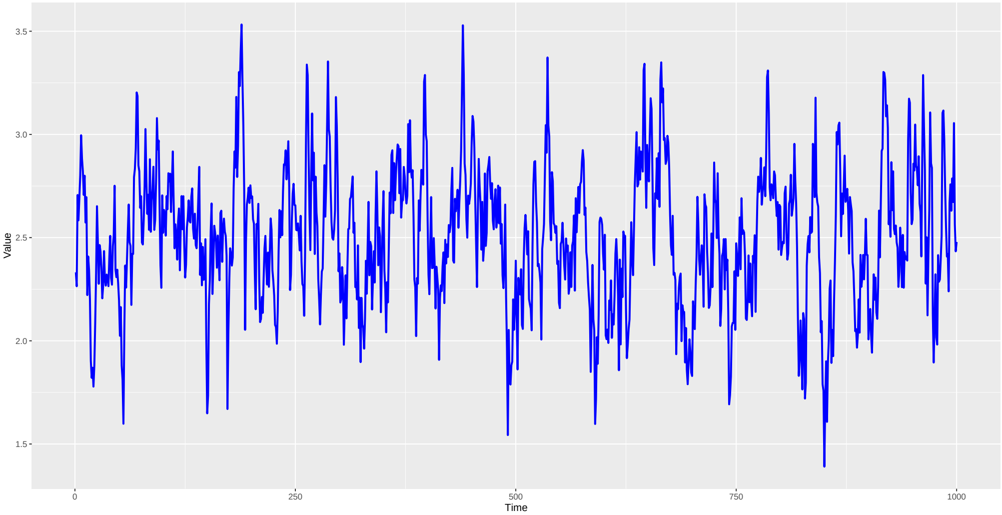
This process is weakly stationary as

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

$$\gamma_0 = 0.054 \longrightarrow \sigma^2(1-\phi^2)$$

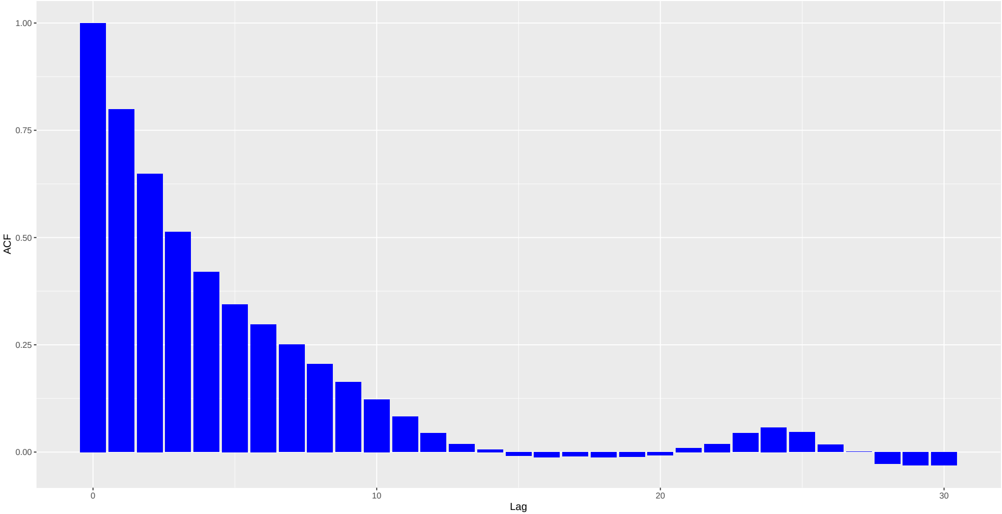
$$\rho_j = 0.80^j \longrightarrow \phi^j$$

Autoregressive Process: Simulated AR(1) Process



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

Autoregressive Process: ACF of AR(1) Process



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

Autoregressive Process: Special Case

If $c = 0$ and $\phi = 1$, we have a naïve random walk

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

which can be rewritten by back-substitution as

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

The naïve random walk is non-stationary as the variance grows over time

$$Var(Y_t) = \sum_{s=1}^{t} Var(\varepsilon_s) = t\sigma^2$$

and the impact of a single shock is permanent and never dissipates.

Autoregressive Process: Special Case

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

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

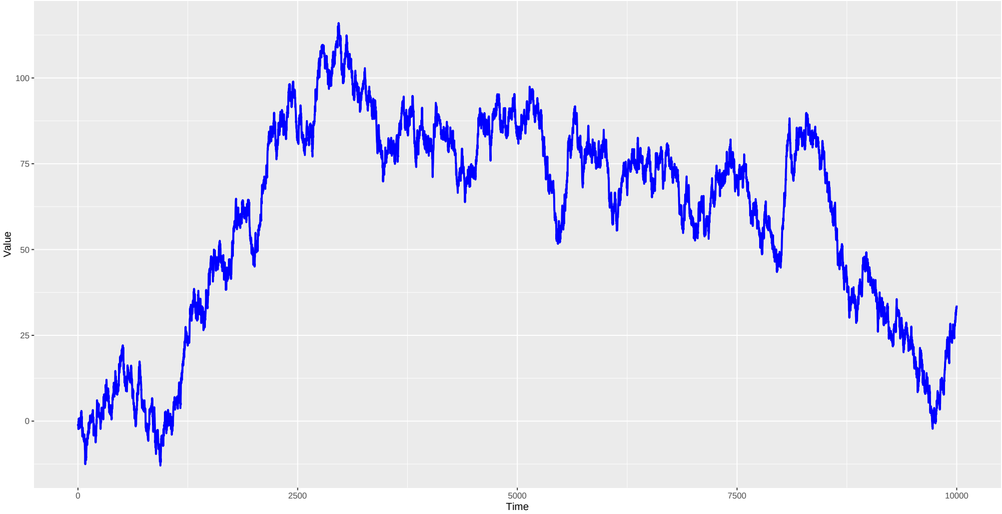
which can be rewritten by back-substitution as

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

The random walk with drift is non-stationary as both mean and variance grow over time

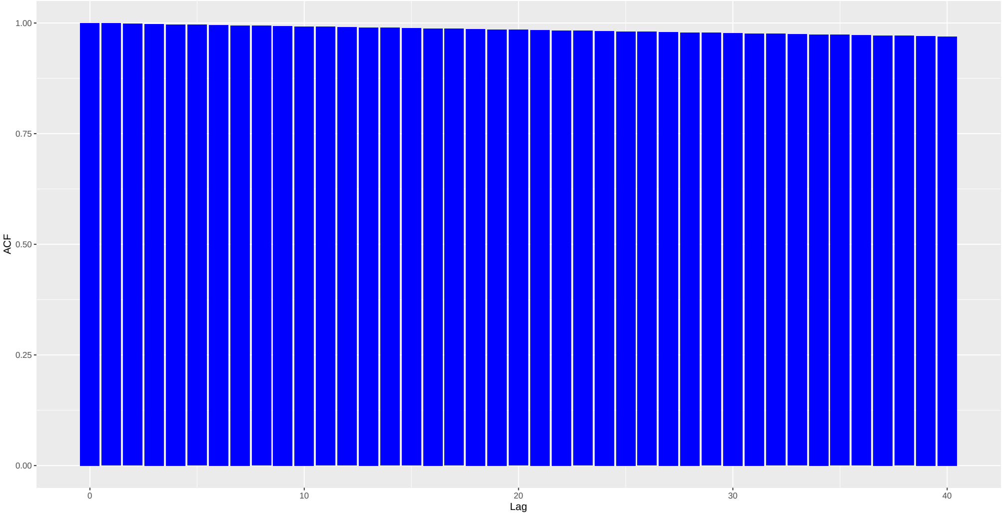
$$\begin{aligned} E(Y_t) &= Y_0 + tc \\ Var(Y_t) &= \sum_{s=1}^{t} Var(\varepsilon_s) = t\sigma^2 \end{aligned}$$

Simulated Naïve Random Walk



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

ACF of a Naïve Random Walk



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

Autoregressive Process: AR(1)

If $c \neq 0$ and $\phi > 1$, we have an explosive AR process

- The process displays exponential growth and high sensitivity to initial conditions and do not seem to be good descriptions for most economic time series.

If $c \neq 0$ and $\phi < 1$, we have an explosive oscillating AR process

- The process displays explosive oscillating growth and does not appear to be empirically relevant for economic applications.

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?

To address them, we must identify the stationarity conditions

- To derive the stationarity conditions, we will rewrite the $AR(p)$ process as a first-order vector autoregressive or $VAR(1)$ process.

Autoregressive Process: AR(p)

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

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

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

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

and

$$Q \longrightarrow \text{matrix of eigenvectors}$$

Autoregressive Process: AR(p)

The system is thus stationary if

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

which requires

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

The eigenvalues λ_i can also be seen as the roots to

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

AR(2) Process: Example 1

Consider the following $AR(2)$ process

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

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.

AR(2) Process: Example I

Take the eigenvalue decomposition of Φ as

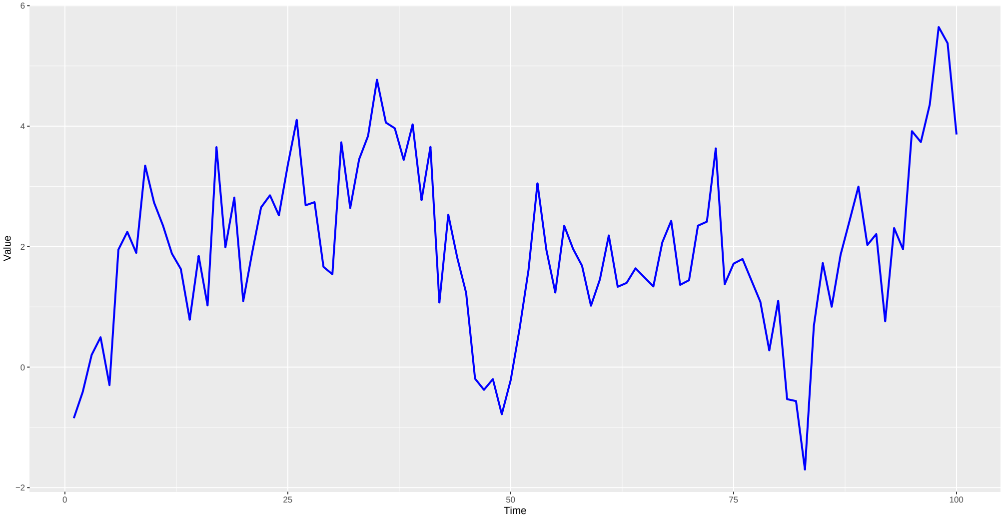
$$\underbrace{\begin{bmatrix} 0.6 & 0.2 \\ 1 & 0 \end{bmatrix}}_{\Phi} = \underbrace{\begin{bmatrix} 0.97 & 0.23 \\ -0.77 & 0.64 \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} 0.84 & 0 \\ 0 & -0.24 \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} 0.97 & 0.23 \\ -0.77 & 0.64 \end{bmatrix}^{-1}}_{Q^{-1}}$$

We thus have a stationary process since

$$\lambda_1 = 0.84 \longrightarrow |\lambda_1| < 1$$

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

AR(2) Process: Example 1



Author's simulations based on an R script using $c = 0.5$, $\phi_1 = 0.6$ and $\phi_2 = 0.2$.

AR(2) Process: Example 2

Consider the following $AR(2)$ process

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

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.

AR(2) Process: Example I

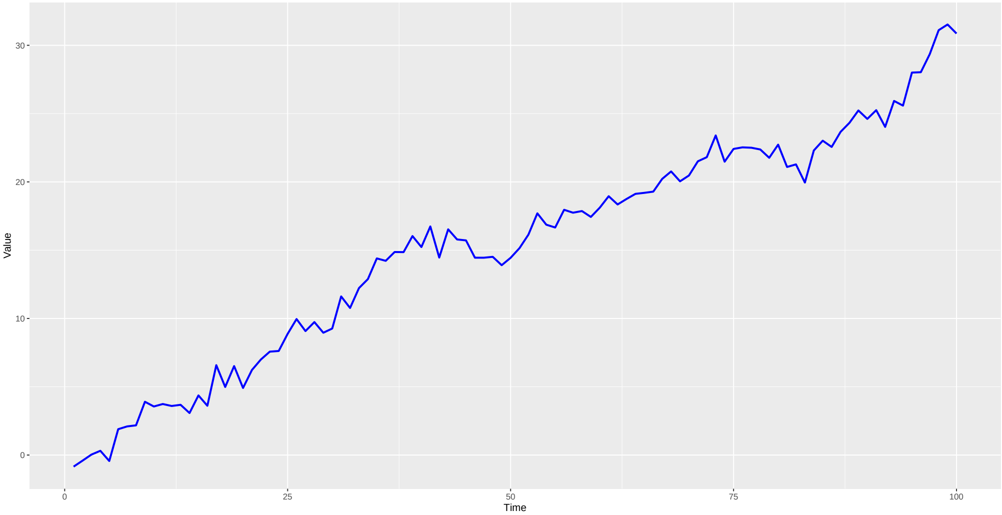
Take the eigenvalue decomposition of Φ as

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

We thus have a non-stationary process since

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

AR(2) Process: Example 2



Author's simulations based on an R script using $c = 0.5$, $\phi_1 = 0.6$ and $\phi_2 = 0.4$.

Autoregressive Process: AR(p)

Rules to check 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.

The stationarity conditions can also be derived using the lag operator.

Autoregressive Process: AR(p)

Unconditional mean

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

where $\sum_{i=1}^{p}\phi_i < 1$.

Unconditional variance

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

Autoregressive Process: AR(p)

We often need to compute the autocovariance analytically

- Easy for a *MA* but demanding for *AR* processes,
- The Yule-Walker equations simplify the computation.

The **Yule-Walker equations** are obtained as follows

- Multiply both side of the equation by Y_{t-j} for $j = 0, 1, \ldots, p$,
- Take the expectations of both sides,
- Need $p + 1$ equations for an $AR(p)$ model.

A system of equations where the solutions provide autocovariances.

Yule-Walker: AR(p) Process

Yule-Walker equations for an $AR(p)$ process

$$
\begin{aligned}
E\left(Y_t Y_t\right) \\
E\left(Y_t Y_{t-1}\right) \\
E\left(Y_t Y_{t-2}\right) \\
\vdots \\
E\left(Y_t Y_{t-p}\right)
\end{aligned}
$$

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.

ARMA Models

Granger and Morris (1976) show

- $Y_{1,t}$ is an $ARMA(p_1, q_1)$ process and $Y_{2,t}$ is an $ARMA(p_2, q_2)$ process,
- $Y_{1,t}$ and $Y_{2,t}$ may be contemporaneously cross-correlated,
- $Y_{1,t} + Y_{2,t}$ is an $ARMA\left(p, q\right)$ process where $p = p_1 + p_2$ and $q = \max\left(p_1 + q_2, q_1 + p_2\right)$.

For example, if $Y_{1,t}$ is an $AR(1)$ process and $Y_{2,t}$ is an $AR(1)$ process, then $Y_1 + Y_2$ is an $ARMA\left(2, 1\right)$ process.

How to Select p and q?

Model Selection

The principle of parsimony – using as few parameters as possible – plays a critical role when we construct an empirical model.

When you estimate a model, more parameters are likely to generate better in-sample fit but poor out-of-sample performance.

In-sample forecasting means that the unknown parameters are estimated using the full-sample information *(look-ahead bias)*.

Out-of-sample forecasting means that the unknown parameters are estimated only using the available information at the time the forecast is produced.

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.

Finance Literature The belief is that simpler models provide more robust forecasts.

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

The **partial autocorrelation** is different from the autocorrelation

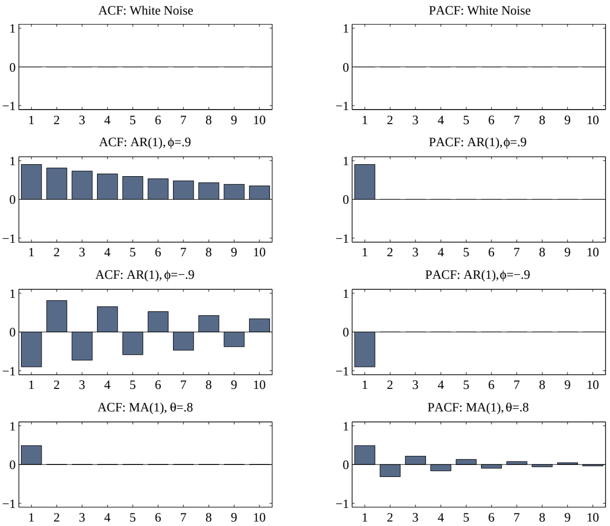
- The j^{th} partial autocorrelation relates Y_t and Y_{t-j} but remove the effects of $Y_{t-1}, \ldots, Y_{t-j+1}$,
- The j^{th} partial autocorrelation φ_j is computed by running the following regression

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

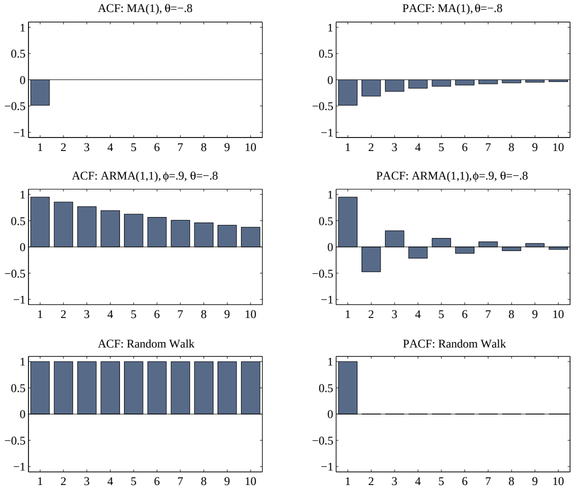
The **partial autocorrelation function** $(PACF)$ is the plot of φ_j against j

- 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 of ARMA Models



ACFs and PACFs of ARMA Models



ACFs and PACFs of ARMA Models

Process	ACF	PACF
WN	$\rho_j = 0$ for all j	$\varphi_j = 0$ for all j
$AR(1)$	$\rho_j = \phi^j$	$\begin{cases} \varphi_j \neq 0 & \textit{for} \quad j \leq 1 \\ \varphi_j = 0 & \textit{for} \quad j > 1 \end{cases}$
$AR(p)$	Decays towards 0	$\begin{cases} \varphi_j \neq 0 & \textit{for} \quad j \leq p \\ \varphi_j = 0 & \textit{for} \quad j > 1 \end{cases}$
$MA(1)$	$\begin{cases} \rho_j \neq 0 & \textit{for} \quad j \leq 1 \\ \rho_j = 0 & \textit{for} \quad j > 1 \end{cases}$	Decays towards 0
$MA(q)$	$\begin{cases} \rho_j \neq 0 & \textit{for} \quad j \leq q \\ \rho_j = 0 & \textit{for} \quad j > q \end{cases}$	Decays towards 0
$ARMA(p, q)$	Decays towards 0	Decays towards 0

Inference

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

- the first s autocorrelations are all zero,
- $H_0 : \rho_1 = \rho_2 = \ldots = \rho_s = 0$ (and homoskedasticity)

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

Information Criteria

Information criteria penalize for including additional regressors.

Akaike Information Criteria (AIC)

$$AIC = -2 \ln \left(\theta | y\right) + 2K$$

Schwartz or **Bayesian Information Criterion** (BIC)

$$BIC = -2 \ln \left(\theta | y\right) + K \ln T$$

- $(\theta | y)$ is the likelihood evaluated at the parameter estimates,
- K is the number of parameters,
- T the number of observations

Information Criteria

Under the assumption of *iid* normally distributed residuals.

Akaike Information Criteria (AIC)

$$AIC = T \ln \widehat{\sigma}^2 + 2K$$

Schwartz or **Bayesian Information Criterion** (BIC)

$$BIC = T \ln \widehat{\sigma}^2 + K \ln T$$

- $\widehat{\sigma}^2$ is the estimated variance of the regression errors,
- K is the number of parameters,
- T the number of observations

Nonstationary Time Series

Many economic and financial time series exhibit trending behavior or nonstationarity in the mean (e.g., asset prices, exchange rates and the levels of macroeconomic aggregates like real GDP).

An important task is determining the most appropriate form of the trend in the data. If the data are trending, then some form of trend removal is required.

Two common detrending procedures

- First differencing for $I(1)$ time series,
- Time-trend regression for trend stationary $I(0)$ time series.

Nonstationary Time Series

Consider the stylized trend-cycle decomposition of a time series

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

We have the following cases

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

Nonstationary Time Series

Unit root tests

- Test the null hypothesis that $\phi = 1$ (non-stationary) against the alternative hypothesis that $\phi < 1$ (trend stationary) or $\phi < 1$ & $\delta = 0$ (stationary) process,
- Called unit root tests because under the null hypothesis the autoregressive polynomial has a root equal to unity.

Unit root tests, in practice, face a number of drawbacks

- They have nonstandard and non-normal asymptotic distributions. Critical values must be calculated using simulation methods.
- The distributions are affected by the inclusion of deterministic terms (i.e., constant, time trend, dummy variables) and so different sets of critical values must be used.

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

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

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

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

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