Lecture 8: Time Series Analysis

MIT 18.S096

Dr. Kempthorne

Fall 2013
Outline

1. Time Series Analysis
   - Stationarity and Wold Representation Theorem
   - Autoregressive and Moving Average (ARMA) Models
   - Accommodating Non-Stationarity: ARIMA Models
   - Estimation of Stationary ARMA Models
   - Tests for Stationarity/Non-Stationarity
Stationarity and Wold Representation Theorem

A stochastic process \( \{..., X_{t-1}, X_t, X_{t+1}, \ldots \} \) consisting of random variables indexed by time index \( t \) is a **time series**.

The stochastic behavior of \( \{X_t\} \) is determined by specifying the probability density/mass functions (pdf’s)

\[
p(x_{t_1}, x_{t_2}, \ldots, x_{t_m})
\]

for all finite collections of time indexes

\[
\{(t_1, t_2, \ldots, t_m), \ m < \infty\}
\]
i.e., all finite-dimensional distributions of \( \{X_t\} \).

**Definition:** A time series \( \{X_t\} \) is **Strictly Stationary** if

\[
p(t_1 + \tau, t_2 + \tau, \ldots, t_m + \tau) = p(t_1, t_2, \ldots, t_m),
\]

\( \forall \tau, \forall m, \forall (t_1, t_2, \ldots, t_m) \).

(Invariance under time translation)
**Definitions of Stationarity**

**Definition:** A time series \( \{ X_t \} \) is **Covariance Stationary** if

\[
E(X_t) = \mu \\
Var(X_t) = \sigma^2_X \\
Cov(X_t, X_{t+\tau}) = \gamma(\tau)
\]

(all constant over time \( t \))

The **auto-correlation function** of \( \{ X_t \} \) is

\[
\rho(\tau) = \frac{Cov(X_t, X_{t+\tau})}{\sqrt{Var(X_t) \cdot Var(X_{t+\tau})}} = \frac{\gamma(\tau)}{\gamma(0)}
\]
Wold Representation Theorem: Any zero-mean covariance stationary time series \( \{X_t\} \) can be decomposed as \( X_t = V_t + S_t \) where

- \( \{V_t\} \) is a linearly deterministic process, i.e., a linear combination of past values of \( V_t \) with constant coefficients.
- \( S_t = \sum_{i=0}^{\infty} \psi_i \eta_{t-i} \) is an infinite moving average process of error terms, where
  - \( \psi_0 = 1, \sum_{i=0}^{\infty} \psi_i^2 < \infty \)
  - \( \{\eta_t\} \) is linearly unpredictable white noise, i.e.,
    \[ E(\eta_t) = 0, \ E(\eta_t^2) = \sigma^2, \ E(\eta_t \eta_s) = 0 \ \forall t, \ \forall s \neq t, \]
  and \( \{\eta_t\} \) is uncorrelated with \( \{V_t\} : \)
    \[ E(\eta_t V_s) = 0, \ \forall t, s \]
Intuitive Application of the Wold Representation Theorem

Suppose we want to specify a covariance stationary time series 
\( \{X_t\} \) to model actual data from a real time series 
\( \{x_t, t = 0, 1, \ldots, T\} \)

Consider the following strategy:

- Initialize a parameter \( p \), the number of past observations in 
  the linearly deterministic term of the Wold Decomposition of 
  \( \{X_t\} \)
- Estimate the linear projection of \( X_t \) on \( (X_{t-1}, X_{t-2}, \ldots, X_{t-p}) \)
  - Consider an estimation sample of size \( n \) with endpoint \( t_0 \leq T \).
  - Let \( \{j = -(p-1), \ldots, 0, 1, 2, \ldots n\} \) index the subseries of 
    \( \{t = 0, 1, \ldots, T\} \) corresponding to the estimation sample and 
    define \( \{y_j : y_j = x_{t_0-n+j}\} \), (with \( t_0 \geq n + p \))
  - Define the vector \( \mathbf{Y} \ (T \times 1) \) and matrix \( \mathbf{Z} \ (T \times [p+1]) \) as:
Estimate the linear projection of $X_t$ on $(X_{t-1}, X_{t-2}, \ldots, X_{t-p})$

$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$

$Z = \begin{bmatrix} 1 & y_0 & y_1 & \cdots & y_{(p-1)} \\ 1 & y_1 & y_0 & \cdots & y_{(p-2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & y_{n-1} & y_{n-2} & \cdots & y_{n-p} \end{bmatrix}$

Apply OLS to specify the projection:

$\hat{y} = Z(Z^T Z)^{-1} Z y$

$= \hat{P}(Y_t \mid Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})$

$= \hat{y}(p)$

Compute the projection residual

$\hat{\epsilon}^{(p)} = y - \hat{y}(p)$
Apply time series methods to the time series of residuals \( \{\hat{\epsilon}_j^{(p)}\} \) to specify a moving average model:

\[
\epsilon_t^{(p)} = \sum_{i=0}^{\infty} \psi_j \eta_{t-i}
\]

yielding \( \{\hat{\psi}_j\} \) and \( \{\hat{\eta}_t\} \), estimates of parameters and innovations.

Conduct a case analysis diagnosing consistency with model assumptions:

- Evaluate orthogonality of \( \hat{\epsilon}^{(p)} \) to \( Y_{t-s} \), \( s > p \).
  - If evidence of correlation, increase \( p \) and start again.
- Evaluate the consistency of \( \{\hat{\eta}_t\} \) with the white noise assumptions of the theorem.
  - If evidence otherwise, consider revisions to the overall model:
    - Changing the specification of the moving average model.
    - Adding additional ‘deterministic’ variables to the projection model.
Note:

- Theoretically, 
  \[
  \lim_{p \to \infty} \hat{y}^{(p)} = \hat{y} = P(Y_t \mid Y_{t-1}, Y_{t-2}, \ldots)
  \]
  but if \( p \to \infty \) is required, then \( n \to \infty \) while \( p/n \to 0 \).

- Useful models of covariance stationary time series have
  - Modest finite values of \( p \) and/or include
  - Moving average models depending on a parsimonious number of parameters.
Lag Operator $L()$

**Definition** The lag operator $L()$ shifts a time series back by one time increment. For a time series $\{X_t\}$:

$$L(X_t) = X_{t-1}.$$  

Applying the operator recursively we define:

- $L^0(X_t) = X_t$
- $L^1(X_t) = X_{t-1}$
- $L^2(X_t) = L(L(X_t)) = X_{t-2}$

...  

- $L^n(X_t) = L(L^{n-1}(X_t)) = X_{t-n}$

Inverses of these operators are well defined as:

$$L^{-n}(X_t) = X_{t+n}, \text{ for } n = 1, 2, \ldots$$
Wold Representation with Lag Operators

The Wold Representation for a covariance stationary time series \( \{X_t\} \) can be expressed as
\[
X_t = \sum_{i=0}^{\infty} \psi_i \eta_{t-i} + V_t = \sum_{i=0}^{\infty} \psi_i L^i(\eta_t) + V_t = \psi(L)\eta_t + V_t
\]

where \( \psi(L) = \sum_{i=0}^{\infty} \psi_i L^i \).

**Definition** The Impulse Response Function of the covariance stationary process \( \{X_t\} \) is
\[
IR(j) = \frac{\partial X_t}{\partial \eta_{t-j}} = \psi_j.
\]

The long-run cumulative response of \( \{X_t\} \) is
\[
\sum_{i=0}^{\infty} IR(j) = \sum_{i=0}^{\infty} \psi_i = \psi(L) \text{ with } L = 1.
\]
Suppose that the operator $\psi(L)$ is invertible, i.e.,

$$
\psi^{-1}(L) = \sum_{i=0}^{\infty} \psi_i^* L^i
$$

such that

$$
\psi^{-1}(L) \psi(L) = I = L^0.
$$

Then, assuming $V_t = 0$ (i.e., $X_t$ has been adjusted to $X_t^* = X_t - V_t$), we have the following equivalent expressions of the time series model for \{X_t\}

$$
X_t = \psi(L) \eta_t
$$

$$
\psi^{-1}(L) X_t = \eta_t
$$

**Definition** When $\psi^{-1}(L)$ exists, the time series \{X_t\} is **Invertible** and has an auto-regressive representation:

$$
X_t = (\sum_{i=0}^{\infty} \psi_i^* X_{t-i}) + \eta_t
$$
Outline

1. **Time Series Analysis**
   - Stationarity and Wold Representation Theorem
   - Autoregressive and Moving Average (ARMA) Models
   - Accommodating Non-Stationarity: ARIMA Models
   - Estimation of Stationary ARMA Models
   - Tests for Stationarity/Non-Stationarity
ARMA\((p,q)\) Models

**Definition:** The times series \( \{X_t\} \) follows the ARMA\((p,q)\) Model with auto-regressive order \( p \) and moving-average order \( q \) if

\[
X_t = \mu + \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-1} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) + \eta_t + \theta_1\eta_{t-1} + \theta_2\eta_{t-2} + \cdots + \theta_q\eta_{t-q}
\]

where \( \{\eta_t\} \) is \( WN(0,\sigma^2) \), “White Noise” with

\[
E(\eta_t) = 0, \quad \forall t
\]

\[
E(\eta_t^2) = \sigma^2 < \infty, \quad \forall t, \quad \text{and} \quad E(\eta_t\eta_s) = 0, \quad \forall t \neq s
\]

With lag operators

\[
\phi(L) = (1 - \phi_1L - \phi_2L^2 - \cdots - \phi_pL^p)
\]

\[
\theta(L) = (1 + \theta_1L + \theta_2L^2 + \cdots + \theta_qL^q)
\]

we can write

\[
\phi(L) \cdot (X_t - \mu) = \theta(L)\eta_t
\]

and the Wold decomposition is

\[
X_t = \mu + \psi(L)\eta_t, \quad \text{where} \quad \psi(L) = [\phi(L)]^{-1}\theta(L)
\]
AR(p) Models

Order-\(p\) Auto-Regression Model: AR(p)

\[ \phi(L) \cdot (X_t - \mu) = \eta_t \text{ where} \]
\[ \{\eta_t\} \text{ is } WN(0, \sigma^2) \text{ and} \]
\[ \phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots + \phi_p L^p \]

Properties:

- Linear combination of \(\{X_t, X_{t-1}, \ldots, X_{t-p}\}\) is \(WN(0, \sigma^2)\).
- \(X_t\) follows a linear regression model on explanatory variables \((X_{t-1}, X_{t-2}, \ldots, X_{t-p})\), i.e

\[ X_t = c + \sum_{j=1}^{p} \phi_j X_{t-j} + \eta_t \]

where \(c = \mu \cdot \phi(1)\), (replacing \(L\) by 1 in \(\phi(L)\)).
AR(p) Models

Stationarity Conditions
Consider $\phi(z)$ replacing $L$ with a complex variable $z$.

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

Let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be the $p$ roots of $\phi(z) = 0$.

$$\phi(L) = (1 - \frac{1}{\lambda_1} L) \cdot (1 - \frac{1}{\lambda_2} L) \cdots (1 - \frac{1}{\lambda_p} L)$$

Claim: $\{X_t\}$ is covariance stationary if and only if all the roots of $\phi(z) = 0$ (the “characteristic equation”) lie outside the unit circle $\{z : |z| \leq 1\}$, i.e., $|\lambda_j| > 1$, $j = 1, 2, \ldots, p$

- For complex number $\lambda$: $|\lambda| > 1$,

$$ (1 - \frac{1}{\lambda} L)^{-1} = 1 + \left(\frac{1}{\lambda}\right)L + \left(\frac{1}{\lambda}\right)^2 L^2 + \left(\frac{1}{\lambda}\right)^3 L^3 + \cdots = \sum_{i=0}^{\infty} \left(\frac{1}{\lambda}\right)^i L^i $$

* $\phi^{-1}(L) = \prod_{j=1}^{p} \left[\left(1 - \frac{1}{\lambda_j} L\right)^{-1}\right]$
Suppose \( \{X_t\} \) follows the \( AR(1) \) process, i.e.,
\[
X_t - \mu = \phi(X_{t-1} - \mu) + \eta_t, \quad t = 1, 2, \ldots
\]
where \( \eta_t \sim WN(0, \sigma^2) \).

- The characteristic equation for the \( AR(1) \) model is
  \[
  (1 - \phi z) = 0
  \]
  with root \( \lambda = \frac{1}{\phi} \).
- The \( AR(1) \) model is covariance stationary if (and only if)
  \[
  |\phi| < 1 \quad (\text{equivalently } |\lambda| > 1)
  \]
- The first and second moments of \( \{X_t\} \) are
  \[
  E(X_t) = \mu
  \]
  \[
  Var(X_t) = \sigma^2_X = \sigma^2 / (1 - \phi) \quad (= \gamma(0))
  \]
  \[
  Cov(X_t, X_{t-1}) = \phi \cdot \sigma^2_X
  \]
  \[
  Cov(X_t, X_{t-j}) = \phi^j \cdot \sigma^2_X \quad (= \gamma(j))
  \]
  \[
  Corr(X_t, X_{t-j}) = \phi^j = \rho(j) \quad (= \gamma(j)/\gamma(0))
  \]
AR(1) Model

- For $\phi : |\phi| < 1$, the Wold decomposition of the $AR(1)$ model is:
  \[ X_t = \mu + \sum_{j=0}^{\infty} \phi^j \eta_{t-j} \]
  - For $\phi : 0 < \phi < 1$, the $AR(1)$ process exhibits exponential mean-reversion to $\mu$.
  - For $\phi : 0 > \phi > -1$, the $AR(1)$ process exhibits oscillating exponential mean-reversion to $\mu$.
- For $\phi = 1$, the Wold decomposition does not exist and the process is the simple random walk (non-stationary!).
- For $\phi > 1$, the $AR(1)$ process is explosive.

**Examples of $AR(1)$ Models** (mean reverting with $0 < \phi < 1$)
- Interest rates (Ornstein Uhlenbeck Process; Vasicek Model)
- Interest rate spreads
- Real exchange rates
- Valuation ratios (dividend-to-price, earnings-to-price)
Second Order Moments of $AR(p)$ Processes

From the specification of the $AR(p)$ model:

$$(X_t - \mu) = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-1} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) + \eta_t$$

we can write the Yule-Walker Equations ($j = 0, 1, \ldots$)

$$E[(X_t - \mu)(X_{t-j} - \mu)] = \phi_1 E[(X_{t-1} - \mu)(X_{t-j} - \mu)] + \phi_2 E[(X_{t-1} - \mu)(X_{t-j} - \mu)] + \cdots + \phi_p E[(X_{t-p} - \mu)(X_{t-j} - \mu)] + E[\eta_t(X_{t-j} - \mu)]$$

$$\gamma(j) = \phi_1 \gamma(j-1) + \phi_2 \gamma(j-2) + \cdots + \phi_p \gamma(j-p) + \delta_{0,j} \sigma^2$$

Equations $j = 1, 2, \ldots p$ yield a system of $p$ linear equations in $\phi_j$: 

\[\begin{align*}
\gamma(0) &= \phi_1 \gamma(-1) + \phi_2 \gamma(-2) + \cdots + \phi_p \gamma(-p) + \delta_{0,0} \sigma^2 \\
\gamma(1) &= \phi_1 \gamma(0) + \phi_2 \gamma(-1) + \cdots + \phi_p \gamma(-p+1) + \delta_{0,1} \sigma^2 \\
&\vdots \\
\gamma(p) &= \phi_1 \gamma(p-1) + \phi_2 \gamma(p-2) + \cdots + \phi_p \gamma(0) + \delta_{0,p} \sigma^2
\end{align*}\]
Yule-Walker Equations

\[
\begin{pmatrix}
\gamma(1) \\
\gamma(2) \\
\vdots \\
\gamma(p)
\end{pmatrix}
= 
\begin{bmatrix}
\gamma(0) & \gamma(-1) & \gamma(-2) & \cdots & \gamma(-(p - 1)) \\
\gamma(1) & \gamma(0) & \gamma(-1) & \cdots & \gamma(-(p - 2)) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\gamma(p - 1) & \gamma(p - 2) & \gamma(p - 3) & \cdots & \gamma(0)
\end{bmatrix}
\begin{pmatrix}
\phi_1 \\
\phi_2 \\
\vdots \\
\phi_p
\end{pmatrix}
\]

- Given estimates \(\hat{\gamma}(j), j = 0, \ldots, p\) (and \(\hat{\mu}\)) the solution of these equations are the Yule-Walker estimates of the \(\phi_j\); using the property \(\gamma(-j) = \gamma(+j), \forall j\)

- Using these in equation 0

\[
\gamma(0) = \phi_1 \gamma(-1) + \phi_2 \gamma(-2) + \cdots + \phi_p \gamma(-p) + \delta_{0,0} \sigma^2
\]

provides an estimate of \(\sigma^2\)

\[
\hat{\sigma}^2 = \hat{\gamma}(0) - \sum_{j=1}^{p} \phi_j \hat{\gamma}(j)
\]

- When all the estimates \(\hat{\gamma}(j)\) and \(\hat{\mu}\) are unbiased, then the Yule-Walker estimates apply the **Method of Moments** Principle of Estimation.
MA(q) Models

Order-q Moving-Average Model: MA(q)

\((X_t - \mu) = \theta(L)\eta_t, \text{ where}\)

\(\{\eta_t\} \text{ is } WN(0, \sigma^2) \text{ and}\)

\(\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q\)

Properties:

- The process \(\{X_t\}\) is invertible if all the roots of \(\theta(z) = 0\) are outside the complex unit circle.
- The moments of \(X_t\) are:
  \(E(X_t) = \mu\)
  \(Var(X_t) = \gamma(0) = \sigma^2 \cdot (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\)
  \(Cov(X_t, X_{t+j}) = \begin{cases} 
    0, & j > q \\
    \sigma^2 \cdot (\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \cdots + \theta_q\theta_{q-j}), & 1 < j \leq q
  \end{cases}\)
Outline

1 Time Series Analysis

- Stationarity and Wold Representation Theorem
- Autoregressive and Moving Average (ARMA) Models
- Accommodating Non-Stationarity: ARIMA Models
- Estimation of Stationary ARMA Models
- Tests for Stationarity/Non-Stationarity
Many economic time series exhibit non-stationary behavior consistent with random walks. Box and Jenkins advocate removal of non-stationary trending behavior using

**Differencing Operators:**

\[ \Delta = 1 - L \]
\[ \Delta^2 = (1 - L)^2 = 1 - 2L + L^2 \]
\[ \Delta^k = (1 - L)^k = \sum_{j=0}^{k} \binom{k}{j} (-L)^j, \text{ (integral } k > 0) \]

- If the process \( \{X_t\} \) has a linear trend in time, then the process \( \{\Delta X_t\} \) has no trend.
- If the process \( \{X_t\} \) has a quadratic trend in time, then the second-differenced process \( \{\Delta^2 X_t\} \) has no trend.
**Examples of Non-Stationary Processes**

**Linear Trend Reversion Model:** Suppose the model for the time series \( \{X_t\} \) is:
\[
X_t = TD_t + \eta_t, \text{ where}
\]
- \( TD_t = a + bt \), a deterministic (linear) trend
- \( \eta_t \sim AR(1) \), i.e.,
  \[
  \eta_t = \phi \eta_{t-1} + \xi_t, \text{ where } |\phi| < 1 \text{ and }
  \{\xi_t\} \text{ is } WN(0, \sigma^2).
  \]

The moments of \( \{X_t\} \) are:
\[
E(X_t) = E(TD_t) + E(\eta_t) = a + bt
\]
\[
Var(X_t) = Var(\eta_t) = \sigma^2/(1 - \phi).
\]

The differenced process \( \{\Delta X_t\} \) can be expressed as
\[
\Delta X_t = b + \Delta \eta_t
\]
\[
= b + (\eta_t - \eta_{t-1})
\]
\[
= b + (1 - L)\eta_t
\]
Non-Stationary Trend Processes

Pure Integrated Process I(1) for \( \{X_t\} \):
\[
X_t = X_{t-1} + \eta_t, \text{ where } \eta_t \text{ is } WN(0, \sigma^2).
\]
Equivalently:
\[
\Delta X_t = (1 - L)X_t + \eta_t, \text{ where } \{\eta_t\} \text{ is } WN(0, \sigma^2).
\]
Given \( X_0 \), we can write \( X_t = X_0 + TS_t \) where
\[
TS_t = \sum_{j=0}^{t} \eta_j
\]
The process \( \{ TS_t \} \) is a **Stochastic Trend** process with
\[
TS_t = TS_{t-1} + \eta_t, \text{ where } \{\eta_t\} \text{ is } WN(0, \sigma^2).
\]
Note:
- The Stochastic Trend process is not perfectly predictable.
- The process \( \{ X_t \} \) is a **Simple Random Walk** with white-noise steps. It is non-stationary because given \( X_0 \):
  - \( Var(X_t) = t\sigma^2 \)
  - \( Cov(X_t, X_{t-j}) = (t - j)\sigma^2 \) for \( 0 < j < t \).
  - \( Corr = (X_t, X_{t-j}) = \sqrt{t-j}/\sqrt{t} = \sqrt{1-j/t} \)
ARIMA(p,d,q) Models

**Definition:** The time series \( \{X_t\} \) follows an ARIMA\((p, d, q)\) model (“Integrated ARMA”) if \( \{\Delta^d X_t\} \) is stationary (and non-stationary for lower-order differencing) and follows an ARMA\((p, q)\) model.

**Issues:**

- Determining the order of differencing required to remove time trends (deterministic or stochastic).
- Estimating the unknown parameters of an ARIMA\((p, d, q)\) model.
- Model Selection: choosing among alternative models with different \((p, d, q)\) specifications.
Outline

1. Time Series Analysis
   - Stationarity and Wold Representation Theorem
   - Autoregressive and Moving Average (ARMA) Models
   - Accommodating Non-Stationarity: ARIMA Models
   - Estimation of Stationary ARMA Models
   - Tests for Stationarity/Non-Stationarity
Maximum-Likelihood Estimation

- Assume that \( \{\eta_t\} \) are i.i.d. \( N(0, \sigma^2) \) r.v.'s.
- Express the \( ARMA(p, q) \) model in state-space form.
- Apply the prediction-error decomposition of the log-likelihood function.
- Apply either or both of
  - Limited Information Maximum-Likelihood (LIML) Method
    - Condition on the first \( p \) values of \( \{X_t\} \)
    - Assume that the first \( q \) values of \( \{\eta_t\} \) are zero.
  - Full Information Maximum-Likelihood (FIML) Method
    - Use the stationary distribution of the first \( p \) values to specify the exact likelihood.
Model Selection

Statistical model selection criteria are used to select the orders \((p, q)\) of an ARMA process:

- Fit all \(ARMA(p, q)\) models with \(0 \leq p \leq p_{\text{max}}\) and \(0 \leq q \leq q_{\text{max}}\), for chosen values of maximal orders.
- Let \(\tilde{\sigma}^2(p, q)\) be the MLE of \(\sigma^2 = \text{Var}(\eta_t)\), the variance of ARMA innovations under Gaussian/Normal assumption.
- Choose \((p, q)\) to minimize one of:
  - Akaike Information Criterion
    \[
    AIC(p, q) = \log(\tilde{\sigma}^2(p, q)) + 2\frac{p+q}{n}
    \]
  - Bayes Information Criterion
    \[
    BIC(p, q) = \log(\tilde{\sigma}^2(p, q)) + \log(n)\frac{p+q}{n}
    \]
  - Hannan-Quinn Criterion
    \[
    HQ(p, q) = \log(\tilde{\sigma}^2(p, q)) + 2\log(\log(n))\frac{p+q}{n}
    \]
Outline

1. Time Series Analysis
   - Stationarity and Wold Representation Theorem
   - Autoregressive and Moving Average (ARMA) Models
   - Accommodating Non-Stationarity: ARIMA Models
   - Estimation of Stationary ARMA Models
   - Tests for Stationarity/Non-Stationarity
Dickey-Fuller (DF) Test: Suppose \( \{ X_t \} \) follows the AR(1) model

\[
X_t = \phi X_{t-1} + \eta_t, \text{ with } \{ \eta_t \} \text{ a WN}(0, \sigma^2).
\]

Consider testing the following hypotheses:

- \( H_0: \phi = 1 \) (unit root, non-stationarity)
- \( H_1: |\phi| < 1 \) (stationarity)

(“Autoregressive Unit Root Test”)

- Fit the AR(1) model by least squares and define the test statistic:

\[
t_{\phi=1} = \frac{\hat{\phi} - 1}{se(\hat{\phi})}
\]

where \( \hat{\phi} \) is the least-squares estimate of \( \phi \) and \( se(\hat{\phi}) \) is the least-squares estimate of the standard error of \( \hat{\phi} \).

- If \( |\phi| < 1 \), then \( \sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} N(0, (1 - \phi^2)) \).

- If \( \phi = 1 \), then \( \hat{\phi} \) is super-consistent with rate \( (1/T) \), \( \sqrt{T}t_{\phi=1} \) has DF distribution.
References on Tests for Stationarity/Non-Stationarity*

Unit Root Tests ($H_0$: Nonstationarity)
- Dickey and Fuller (1979): Dickey-Fuller (DF) Test
- Said and Dickey (1984): Augmented Dickey-Fuller (ADF) Test
- Phillips and Perron (1988) Unit root (PP) tests

Stationarity Tests ($H_0$: stationarity)

* Optional reading