# Univariate Time Series Models

Gianluca Cubadda

Università di Roma "Tor Vergata"

# Program

- Stationary time series analysis: Basic concepts. Stationarity, autocorrelation, partial autocorrelation. Linear stationary processes. ARMA models. Forecasting.

- Nonstationary time series analysis: ARIMA models. Seasonality, The Box-Jenkins approach.

- Unit roots in macroeconomic time series: Deterministic trends vs. random walks. Unit-roots tests.

- The analysis of financial time series: Volatility and conditional heteroscedasticity. GARCH and IGARCH models.

# 1 Univariate time series analysis: Basic concepts

We consider a univariate time series, $y_t, t = 1, \ldots, T$.

The information set is the series itself and its position in time.

We now review some basic concepts in time series analysis, along with simple and essential tools for descriptive analysis.

The main descriptive tool is the plot of the series, by which we represent the pair of values $(t, y_t)$ on a Cartesian plane.

The graph can immediately reveal the presence of important features, such as trend and seasonality, structural breaks and outliers, and so forth.

The series may be a transformation of the original measurements: logarithms; changes, log-differences, etc.

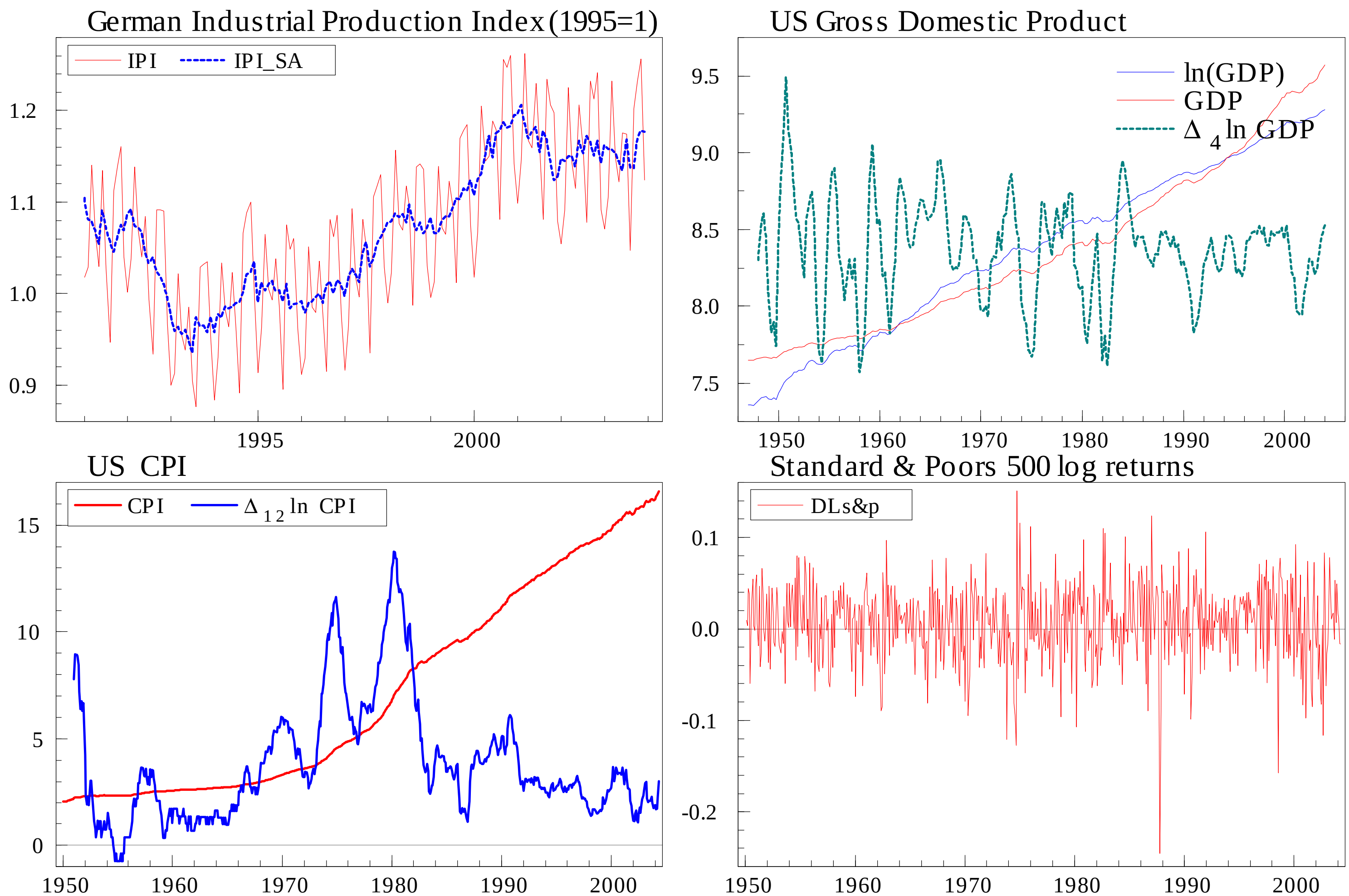


Figure 1: Plots of various time series

## 2 Stationary stochastic processes

Stochastic process: a collection of random variables $\{y_t(\omega), \omega \in \Omega,\ t \in Z\}$ defined on a probability space $(\Omega, F, P)$, where the integer number $t$ is a time-index, $\Omega$ is the sample space, $F$ is a sigma algebra defined on $\Omega$ and $P$ is a probability measure on $\Omega$. A time series is a realization of the stochastic process for a given $\omega \in \Omega$ and $t = 0, 1, 2, ..., T$.

Stationarity: $y_t$ is weakly stationary if $\forall t, k \in Z$:

$$
\begin{aligned}
\mathsf{E}(y_t) &= \mu < \infty \\
\mathsf{E}(y_t - \mu)^2 &= \gamma(0) < \infty \\
\mathsf{E}(y_t - \mu)(y_{t-k} - \mu) &= \gamma(k)
\end{aligned}
$$

$y_t$ is strictly stationary if $\forall t, k, h \in Z$:

$$
(y_t, y_{t+1}, ...y_{t+h}) \overset{d}{=} (y_{t+k}, y_{t+1+k}, ...y_{t+h+k})
$$

Strict stationarity implies weak stationarity whereas the *viceversa* is in general not true. The exception are Gaussian processes, i.e., if the distribution of $(y_t, y_{t+1}, ...y_{t+h})$ is a multivariate Gaussian for $\forall t, h \in Z$.

<u>Autocovariance function</u>, $\gamma(k)$, is symmetric: $\gamma(k) = \gamma(-k)$.
The <u>partial autocovariance function</u> at lag $k$ is the covariance between $y_t$ and $y_{t-k}$ having removed the effects of $w_t = (y_{t-1}, \cdots, y_{t-k+1})$, i.e.

$$g(k) = \mathsf{E}\left\{[y_t - \mathsf{E}(y_t|w_t)][y_{t-k} - \mathsf{E}(y_{t-k}|w_t)]\right\}$$

<u>Autocorrelation function</u> (ACF):

$$\rho(k) = \gamma(k)/\gamma(0)$$

i) $\rho(0) = 1$; ii) $|\rho(k)| < 1$; iii) $\rho(k) = \rho(-k)$.
The <u>partial autocorrelation function</u> (PACF):

$$r(k) = g(k)/\left\{\mathsf{E}[y_t - \mathsf{E}(y_t|w_t)]^2 \mathsf{E}[y_{t-k} - \mathsf{E}(y_{t-k}|w_t)]^2\right\}^{1/2}$$

White noise (WN): $\varepsilon_t \sim \text{WN}(\sigma^2)$,

$$\begin{aligned} \text{E}(\varepsilon_t) &= 0, \quad \forall t, \\ \text{E}(\varepsilon_t^2) &= \sigma^2 < \infty \quad \forall t, \\ \text{E}(\varepsilon_t \varepsilon_{t-k}) &= 0, \qquad \forall t, \forall k \neq 0. \end{aligned}$$

Lag operator: $L^k y_t = y_{t-k}$, $L$ is an algebraic operator.

Wold theorem: (almost) any weakly stationary stochastic process can be represented as a linear process, i.e.

$$y_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots = \mu + \psi(L)\varepsilon_t,$$

where $\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$, with $\psi_0 = 1$ and $\sum_{j=0}^{\infty} |\psi_j| < \infty$.
Notice that

$$\text{E}(y_t) = \mu, \;\; \gamma(0) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 < \infty, \;\; \gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}.$$

## 2.1 Estimation

- sample mean $\hat{\mu} = \bar{y} = T^{-1} \sum_{t=1}^{T} y_t$

- sample variance: $\hat{\gamma}(0) = T^{-1} \sum_{t=1}^{T} (y_t - \bar{y})^2$

- sample autocovariance: $\hat{\gamma}(k) = T^{-1} \sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})$

- The ACF is estimated by $\hat{\rho}(k) = \hat{\gamma}(k)/\hat{\gamma}(0)$; the barplot $(k, \hat{\rho}(k))$ is the *correlogram*. If $y_t \sim \text{WN}(\sigma^2)$, then $T^{1/2}\hat{\rho}(k) \xrightarrow{d} \text{N}(0, 1)$.

# 3 Genesis and Properties of Autoregressive - Moving Average (ARMA) processes

A problem arises with linear stationary process: an infinite number of coefficients $\left\{\psi_j, j > 0\right\}$ need to be estimated.

Since stationarity implies $\lim\limits_{j\to\infty} \psi_j = 0$, we could approximate $\psi(L)$ by its "truncated" version $\widetilde{\psi}(L)$ such that

$$\widetilde{\psi}_j = \begin{cases} \psi_j, & j \leq m \\ 0, & j > m \end{cases}$$

where $m \to \infty$ and $m/T \to 0$ as $T \to \infty$.

However, the "best" approximation of a $\infty-$order polynomial is obtained by a rational polynomial, i.e.

$$\psi(L) \simeq \frac{\theta(L)}{\phi(L)},$$

where

$$\phi(L) = 1 - \textstyle\sum_{j=1}^{p} \phi_j L^j, \;\; p < \infty$$
$$\theta(L) = 1 + \textstyle\sum_{j=1}^{q} \theta_j L^j, \;\; q < \infty$$

Autoregressive-Moving average (ARMA) processes: A linear stationary process such that $\psi(L) = \theta(L)/\phi(L)$, which can be rewritten as

$$\phi(L) y_t = \theta(L) \varepsilon_t,$$
$$y_t = \textstyle\sum_{j=1}^{p} \phi_j y_{t-j} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

It is denoted as $y_t \sim$ARMA$(p, q)$, where $p$ is the AR order and $q$ is the MA order.

A numerical example: Consider a first-order polynomial $\phi(L) = (1 - \phi L)$ such that $|\phi| < 1$. From the relation

$$1 - (\phi L)^{n+1} = (1 - \phi L)[1 + \phi L + (\phi L)^2 + \cdots + (\phi L)^n],$$

we obtain that

$$\frac{1}{1 - \phi L} = \lim_{n \to \infty} \frac{[1 + \phi L + (\phi L)^2 + \cdots + (\phi L)^n]}{1 - (\phi L)^{n+1}} = \sum_{j=0}^{\infty} (\phi L)^j$$

Assume now that $\theta(L) = 1$ and $\phi(L) = (1 - \phi L)$ with $|\phi| < 1$. Hence, we have

$$y_t = \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \cdots = \frac{\varepsilon_t}{1 - \phi L} = \varepsilon_t + \phi y_{t-1}.$$

## 3.1 Autoregressive (AR) processes

AR(1) processes: The autoregressive process of order 1, AR(1), is generated by the equation

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

The process is stationary if $|\phi| < 1$. Indeed, by recursive substitution we obtain the Wold representation:

$$y_t = m/(1-\phi) + \varepsilon_t + \phi\varepsilon_{t-1} + \cdots + \phi^n \varepsilon_{t-n} + \cdots$$

Hence, the condition $\sum_{j=0}^{\infty} |\psi_j| = \sum_{j=0}^{\infty} |\phi^j| < \infty$ is satisfied iff $|\phi| < 1$ or, equivalently, iff the root of the equation $(1 - \phi L) = 0$, is greater then 1 in modulus.

Under stationarity, the process is uniquely characterized by its moments.

$$\begin{array}{rcl}\mathsf{E}(y_t) & = & \mu = m/(1-\phi) \\ \mathsf{Var}(y_t) & = & \gamma(0) = \mathsf{E}[y_t(y_t-\mu)] = \mathsf{E}[(m+\phi y_{t-1}+\varepsilon_t)(y_t-\mu)] \\ & = & \phi\gamma(1)+\sigma^2\end{array}$$

since $\mathsf{E}[(y_t-\mu)\varepsilon_t] = \mathsf{E}[(\varepsilon_t+\phi\varepsilon_{t-1}+\cdots)\varepsilon_t] = \sigma^2$.

$$\begin{array}{rcl}\gamma(1) & = & \mathsf{E}[y_t(y_{t-1}-\mu)] = \mathsf{E}[(m+\phi y_{t-1}+\varepsilon_t)(y_{t-1}-\mu)] \\ & = & \phi\gamma(0)\end{array}$$

since $\mathsf{E}[(y_{t-1}-\mu)\varepsilon_t] = \mathsf{E}[(\varepsilon_{t-1}+\phi\varepsilon_{t-2}+\cdots)\varepsilon_t] = 0$.

Replacing $\gamma(1)$ in the expression for $\gamma(0)$, we obtain:

$$\gamma(0) = \frac{\sigma^2}{1-\phi^2}$$

Moreover, $\gamma(k) = \phi\gamma(k-1)$ for $k \geq 1$, so that $\gamma(k) = \phi^k\gamma(0)$.

- The autocorrelation function (ACF) is thus

$$\rho(k) = \phi^k$$

- The partial autocorrelation function (PACF) is easily obtained as

$$r(k) = \begin{cases} \rho(k), & k \leq 1 \\ 0, & k > 1 \end{cases}$$

since $y_t - \mathrm{E}[y_t|(y_{t-1}, \cdots, y_{t-k+1})] = \varepsilon_t$ for $k > 1$.

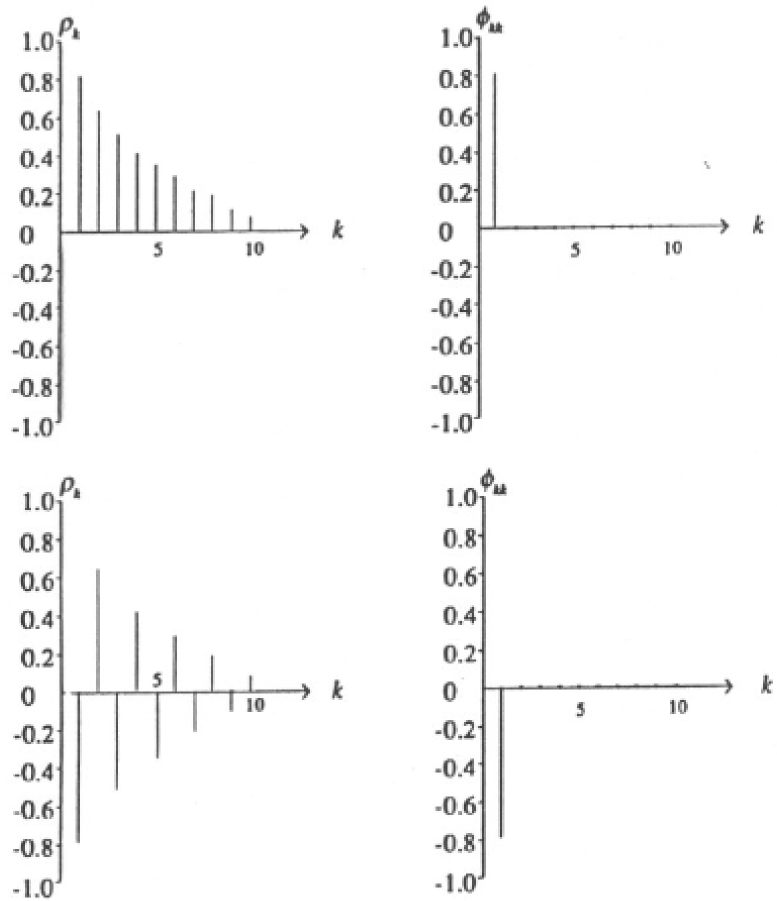


Fig. 3.1 ACF and PACF of the AR(1) process: $(1 - \phi B)\dot{Z}_t = a_t$.

Figure 2: ACF and PACF of an AR(1) process

AR(2) processes: The AR(2) process is generated by the equation

$$y_t = m + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

It can be shown that $y_t$ is stationary if the roots of $1 - \phi_1 L - \phi_2 L^2 = 0$ are greater than 1 in modulus (lie outside the unit circle). This implies the following constraints on the parameter space $(\phi_1, \phi_2)$:
i) $|\phi_2| < 1$ and ii) $|\phi_1| < 1 - \phi_2$.

The stationarity region of the AR parameters lies inside the triangle with vertices (-2,-1),(2,-1),(0,1). A pair of complex conjugate roots arises for $\phi_1^2 + 4\phi_2 < 0$.

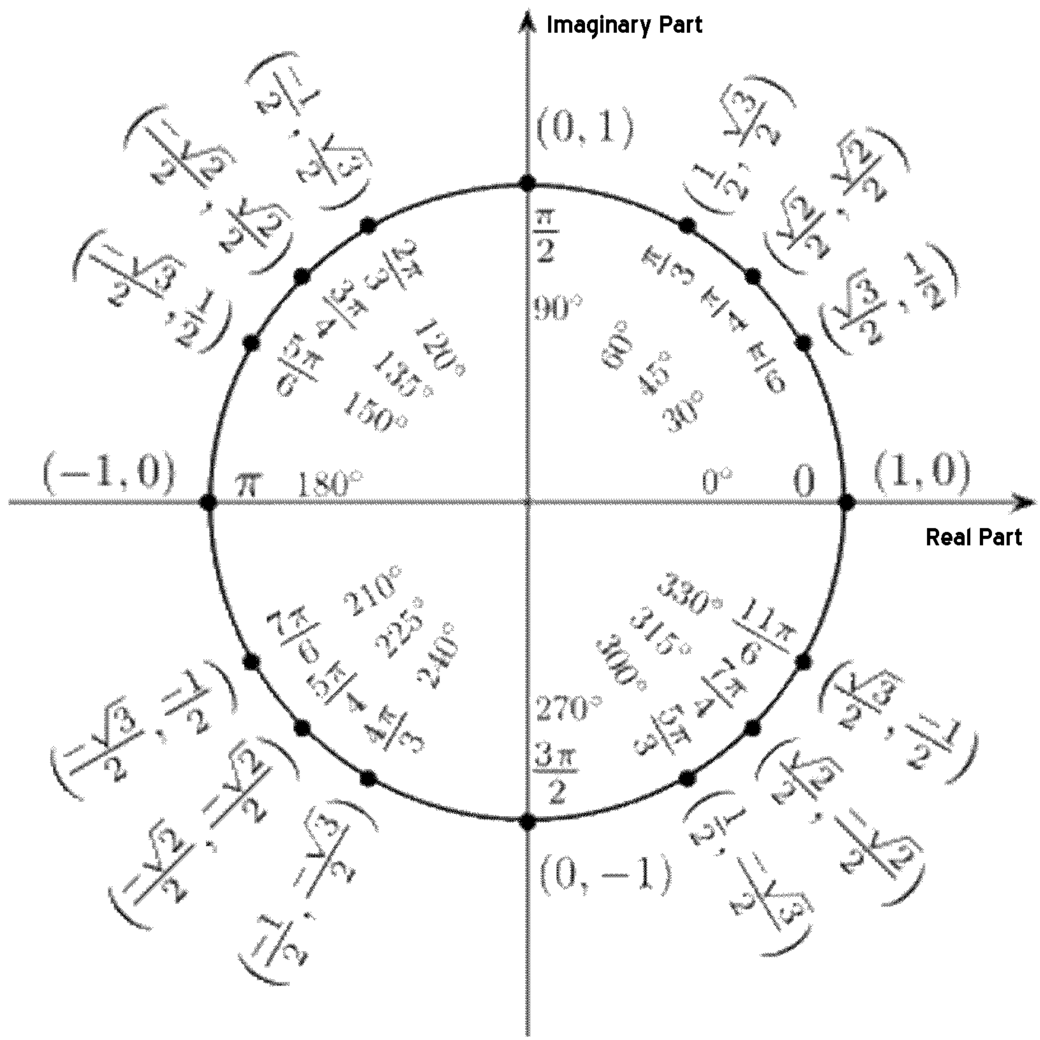


Figure 3: The complex unit circle

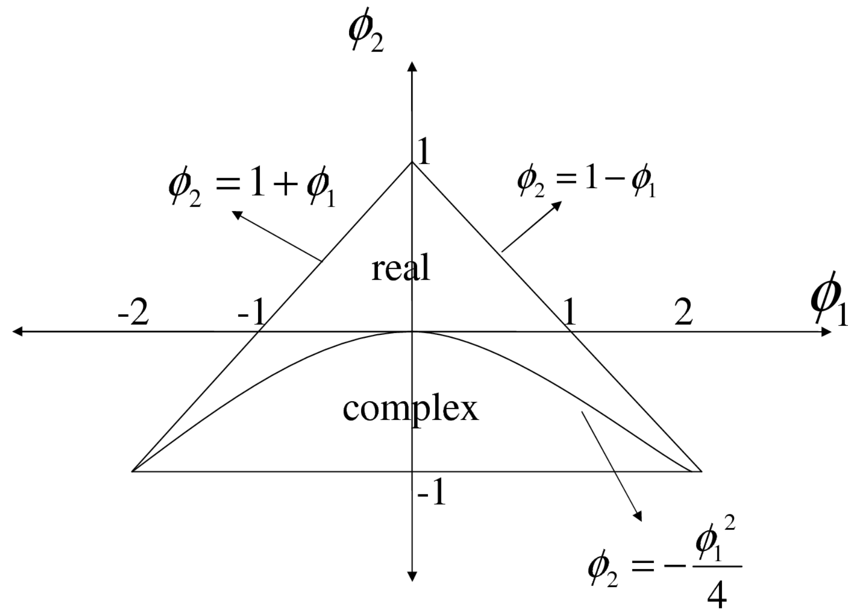


Figure 4: Stationarity region of an AR(2) process

Under stationarity, the process $y_t$ can be uniquely characterized by its moments:

- Expected value: $\mathsf{E}(y_t) = \mu = m/(1 - \phi_1 - \phi_2)$.

- Autocovariance function: it is given recursively by

$$\gamma(k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2), \quad k = 2, 3, \ldots$$

  with starting values

$$\gamma(0) = \frac{(1-\phi_2)\sigma^2}{(1+\phi_2)([(1-\phi_2)^2 - \phi_1^2]}, \quad \gamma(1) = \phi_1 \gamma(0)/(1-\phi_2).$$

The expression for $\gamma(k)$ can be derived as follows:

$$
\begin{array}{rcl}
\gamma(0) & = & \mathsf{E}[(m+\phi_1 y_{t-1}+\phi_2 y_{t-2}+\varepsilon_t)(y_t-\mu)] \\
& = & \phi_1\gamma(1)+\phi_2\gamma(2)+\sigma^2 \\
\gamma(1) & = & \mathsf{E}[(m+\phi_1 y_{t-1}+\phi_2 y_{t-2}+\varepsilon_t)(y_{t-1}-\mu)] \\
& = & \phi_1\gamma(0)+\phi_2\gamma(1) \\
\gamma(2) & = & \mathsf{E}[(m+\phi_1 y_{t-1}+\phi_2 y_{t-2}+\varepsilon_t)(y_{t-2}-\mu)] \\
& = & \phi_1\gamma(1)+\phi_2\gamma(0) \\
\ldots & \ldots & \ldots \\
\gamma(k) & = & \mathsf{E}[(m+\phi_1 y_{t-1}+\phi_2 y_{t-2}+\varepsilon_t)(y_{t-k}-\mu)] \\
& = & \phi_1\gamma(k-1)+\phi_2\gamma(k-2)
\end{array}
$$

Compute $\gamma(1)$ from the second equation, and substitute in the equation for $\gamma(2)$, then replace for $\gamma(1)$ and $\gamma(2)$ in the first expression to get $\gamma(0)$.

- ACF:

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2), \quad k = 2, 3, \ldots$$

with starting values

$$\rho(0) = 1, \quad \rho(1) = \phi_1/(1-\phi_2)$$

It is such that $\rho(k) \to 0$ as $k \to \infty$. If the roots of the AR polynomial are complex the ACF describes a damped cosine wave.

- PACF: It has a cut-off (i.e. it's equal to zero) after $k = 2$ since

$$y_t - \mathsf{E}[y_t | (y_{t-1}, \cdots, y_{t-k+1})] = \varepsilon_{t,} \quad k > 2.$$

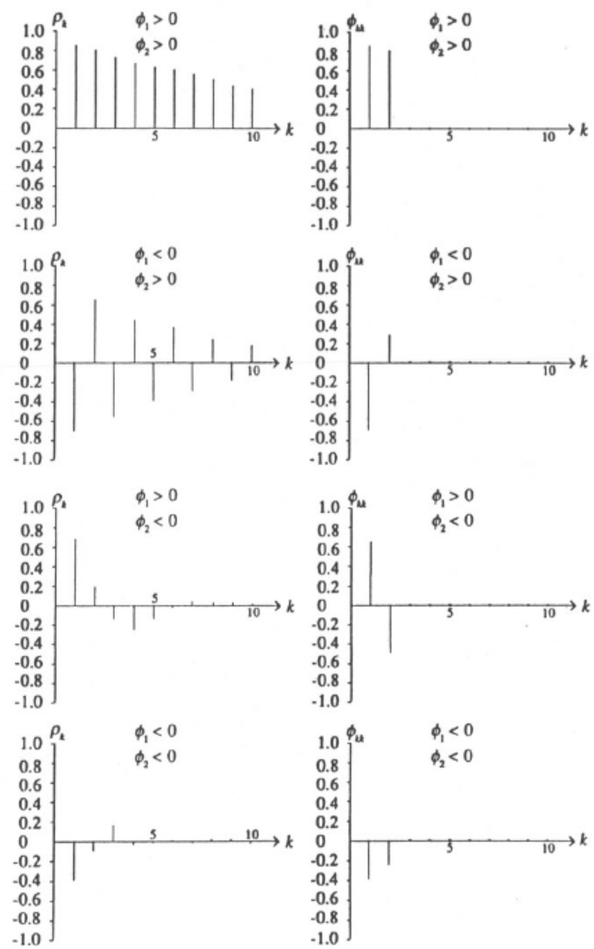


Fig. 3.7 ACF and PACF of AR(2) process: $(1-\phi_1 B-\phi_2 B^2)\dot{Z}_t = a_t$.

Figure 5: ACF and PACF of an AR(2) process

AR($p$) processes: The AR($p$) process is generated by the equation

$$y_t = m + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \text{WN}(\sigma^2)$$

$$\phi(L) y_t = m + \varepsilon_t, \quad \phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p.$$

- $y_t$ is stationary if the $p$ roots of $\phi(L)$ are outside the unit circle.

- $\text{E}(y_t) = \mu = m/\phi(1)$, where $\phi(1) = 1 - \phi_1 - \cdots - \phi_p$.

- The Autocovariance Function is

$$\begin{aligned}\gamma(k) &= \phi_1 \gamma(k-1) + \cdots + \phi_p \gamma(k-p), \quad \text{for } k > 0 \\ \gamma(k) &= \phi_1 \gamma(k-1) + \cdots + \phi_p \gamma(k-p) + \sigma^2, \quad \text{for } k = 0\end{aligned}$$

- ACF is given by the Yule-Walker system of equations:

$$\rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2) + \cdots + \phi_p\rho(k-p), \quad k = 1, 2, \ldots, p$$

- PACF: It has a cut-off after $k = p$ since

$$y_t - \mathsf{E}[y_t|(y_{t-1}, \cdots, y_{t-k+1})] = \varepsilon_t, \quad k > p.$$

## 3.2 Moving Average (MA) processes

In the Wold representation set $\psi_j = \theta_j, j \leq q$ and $\psi_j = 0, j > q$. This gives the MA($q$) process

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

where $\varepsilon_t \sim \mathrm{WN}(\sigma^2)$.

<u>Stationarity</u>: Since the condition $\sum_j |\psi_j| < \infty$ holds, the MA($q$) process is always stationary.

MA(1) processes: The MA(1) process is generated by the equation

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} = \mu + (1+\theta L)\varepsilon_t$$

The moments are obtained as follows

$$\begin{aligned}
\mathsf{E}(y_t) &= \mu + \mathsf{E}(\varepsilon_t) + \theta\mathsf{E}(\varepsilon_{t-1}) = \mu \\
\gamma(0) &= \mathsf{E}(y_t - \mu)^2 = \mathsf{E}(\varepsilon_t + \theta\varepsilon_{t-1})^2 \\
&= \mathsf{E}(\varepsilon_t^2) + 2\theta\mathsf{E}(\varepsilon_t\varepsilon_{t-1}) + \theta^2\mathsf{E}(\varepsilon_{t-1}^2) = \sigma^2(1+\theta^2) \\
\gamma(1) &= \mathsf{E}[(y_t - \mu)(y_{t-1} - \mu)] \\
&= \mathsf{E}[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})] = \theta\sigma^2 \\
\gamma(k) &= \mathsf{E}[(y_t - \mu)(y_{t-k} - \mu)] \\
&= \mathsf{E}[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-k} + \theta\varepsilon_{t-k-1})] = 0, \quad k > 1
\end{aligned}$$

- ACF has a cutoff at $k = 1$:

$$\begin{array}{rcll} \rho(0) & = & 1 & \\ \rho(1) & = & \frac{\theta}{1+\theta^2} & \\ \rho(k) & = & 0, & \qquad k > 1 \end{array}$$

<u>Invertibility</u>: $y_t \sim$MA(1) is invertible if $|\theta| < 1$. Consider the process

$$\tilde{y}_t = \mu + \varepsilon_t + \tilde{\theta}\varepsilon_{t-1}$$

with $\tilde{\theta} = 1/\theta$ and $\varepsilon_t \sim$ WN($\tilde{\sigma}^2$).
The process $\tilde{y}_t$ has the same moments $\mu$, $\gamma(0)$ and $\gamma(1)$, as

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$$

with $\sigma^2 = \tilde{\theta}^2\tilde{\sigma}^2$. Hence, $\rho(1) = \theta^{-1}/(1+\theta^{-2}) = \theta/(1+\theta^2)$ in both cases.

The two processes have identical properties and cannot be discriminated from a time series This problem is known as identifiability and is remedied upon by constraining $\theta$ in the interval (-1,+1).

The term invertibility stems from the possibility of rewriting the process as an infinite autoregression, AR($\infty$), with coefficients $\pi_j$ that are convergent:

$$y_t + \pi_1 y_{t-1} + \pi_2 y_{t-2} + \cdots + \pi_k y_{t-k} + \cdots = m + \varepsilon_t, \quad \sum_{j=1}^{\infty} |\pi_j| < \infty$$

The sequence of weights $\pi_j = (-\theta)^j$ converges if and only if $|\theta| < 1$.

- PACF: Since an invertible MA(1) process can be rewritten as AR($\infty$), its PACF has no cutoff but it decays exponentially.

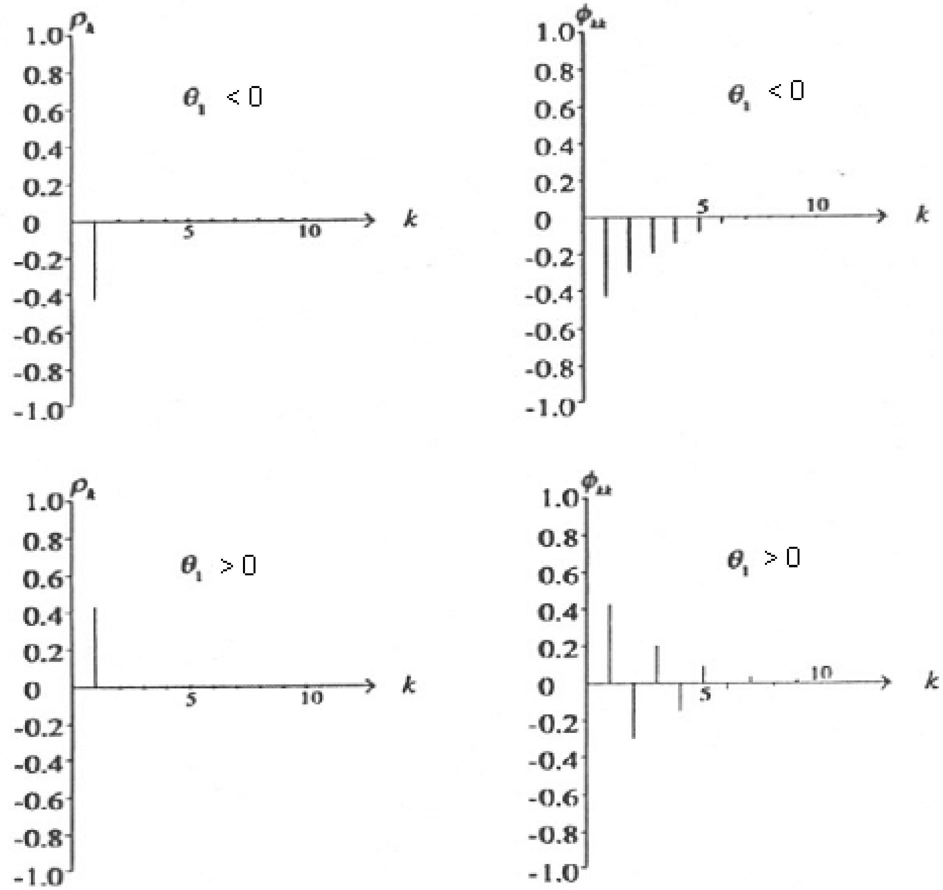


**Fig. 3.10** ACF and PACF of MA(1) processes

Figure 6: ACF and PACF of an MA(1) process

MA($q$) processes: The MA($q$) process is generated by the equation

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

is invertible if the roots of $\theta(L) = 0$ are outside the unit circle.
The moments are obtained as follows

$$\begin{aligned}
\mathsf{E}(y_t) &= \mu \\
\gamma(0) &= \mathsf{E}(y_t - \mu)^2 = \mathsf{E}(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q})^2 \\
&= \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2) \\
\gamma(k) &= \mathsf{E}[(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q})(\varepsilon_{t-k} + \theta_1 \varepsilon_{t-k-1} + \cdots + \theta_q \varepsilon_{t-k-q})] \\
&= \sigma^2(\theta_k + \theta_{k+1}\theta_1 + \cdots + \theta_{q-k}\theta_q) \\
\gamma(k) &= 0, \quad k > q
\end{aligned}$$

- ACF has a cutoff at $k = q$.

- PACF has no cutoff, it is similar has the ACF of an AR($q$) process.

## 3.3 ARMA processes

ARMA($p, q$) processes: The ARMA($p, q$) process is generated by the equation

$$y_t = m + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

$$\phi(L) y_t = m + \theta(L) \varepsilon_t$$

Stationarity: $y_t$ is stationary if the roots of the AR polynomial $\phi(L)$ lie outside the unit circle.

Invertibility: $y_t$ is invertible if the roots of the MA polynomial $\theta(L)$ lie outside the unit circle. Invertibility implies that $y_t$ can be written as an AR($\infty$) process with declining coefficients.

ARMA(1,1) processes: The ARMA(1,1) process is generated by the equation

$$y_t = m + \phi y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$$

The moments are obtained as follows

$$\begin{aligned}
\mathsf{E}(y_t) &= m + \phi\mathsf{E}(y_{t-1}) = m + \phi\mu = m/(1-\phi) \\
\gamma(0) &= \mathsf{E}[y_t(y_t-\mu)] = \mathsf{E}[(m+\phi y_{t-1})(y_t-\mu)] + \mathsf{E}[\varepsilon_t(y_t-\mu)] \\
&+ \mathsf{E}\left\{\theta\varepsilon_{t-1}[\phi(y_{t-1}-\mu) + \varepsilon_t + \theta\varepsilon_{t-1})\right\} \\
&= \phi\gamma(1) + \sigma^2(1+\theta\phi+\theta^2) \\
\gamma(1) &= \mathsf{E}[y_t(y_{t-1}-\mu)] = \mathsf{E}[(m+\phi y_{t-1})(y_{t-1}-\mu)] + \mathsf{E}[\varepsilon_t(y_{t-1}-\mu)] \\
&+ \mathsf{E}\left\{\theta\varepsilon_{t-1}[\phi(y_{t-2}-\mu) + \varepsilon_{t-1} + \theta\varepsilon_{t-2})\right\} \\
&= \phi\gamma(0) + \theta\sigma^2 \\
\gamma(k) &= \mathsf{E}[y_t(y_{t-k}-\mu)] = \phi\gamma(k-1), \quad k > 1
\end{aligned}$$

- Both ACF and PACF have no cutoff!

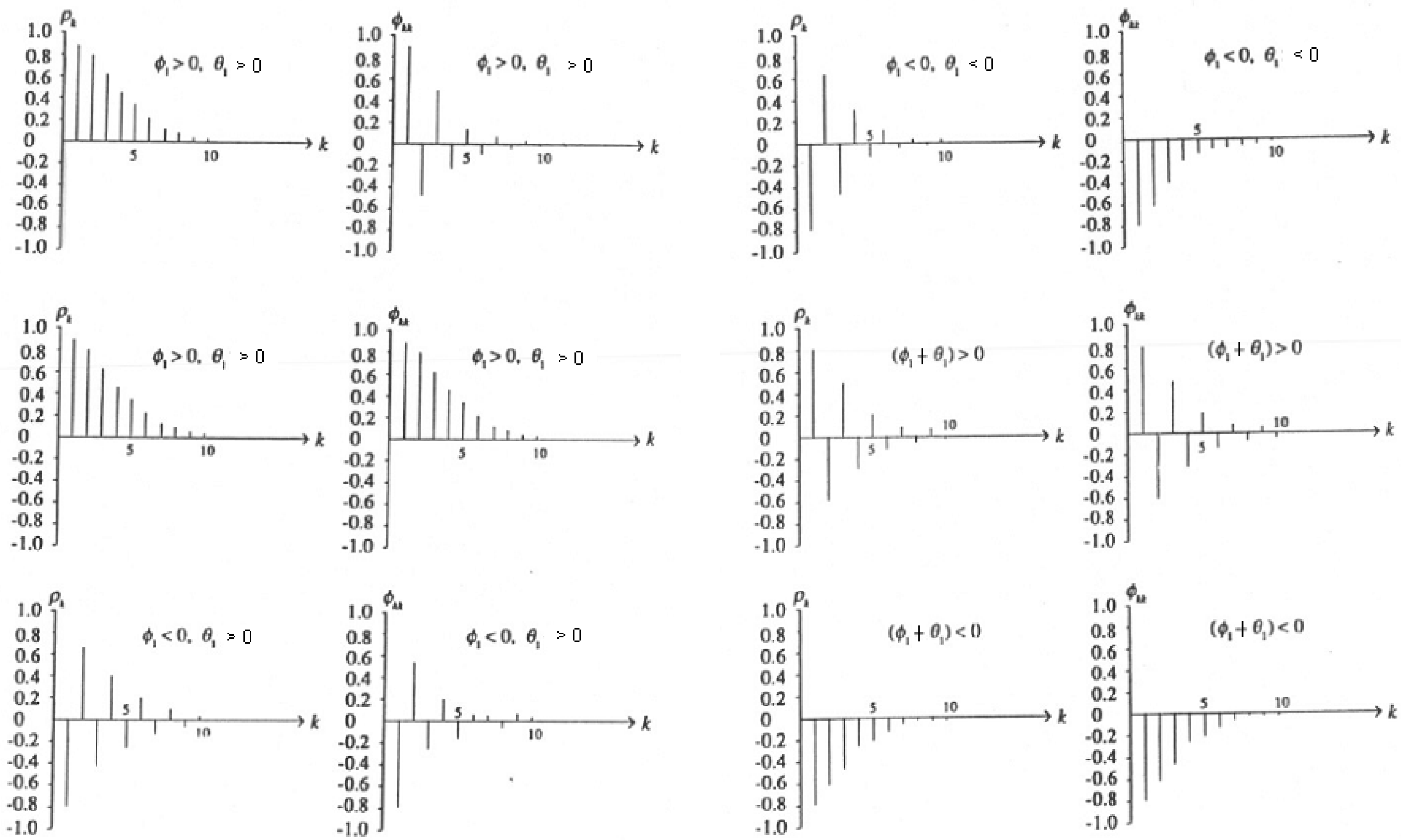


Fig. 3.14 ACF and PACF of ARMA(1,1) model

Fig. 3.14 *(continued)*

Figure 7: ACF and PACF of an ARMA(1,1) process.

## 3.4 Forecasting from ARMA Models

Let $y_t$ ~ARMA$(p, q)$ and $I_t = \{y_t, y_{t-1}, \cdots\}$. The best linear unbiased predictor of $y_{t+h}$ is given by:

$$y_t(h) = \mathsf{E}(y_{t+h}|I_t), \quad h = 1, 2, ...,$$

where $\mathsf{E}(y_{t+h}|I_t)$ is the expected value of $y_{t+h}$ conditional to $I_t$, which is the called the natural filtration of the process $y_t$.

From the expression

$$y_{t+h} = m + \phi_1 y_{t+h-1} + \cdots + \phi_p y_{t+h-p} + \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \cdots + \theta_q \varepsilon_{t+h-q}$$

we get

$$\begin{aligned} y_t(h) &= \quad m + \phi_1 \mathsf{E}(y_{t+h-1}|I_t) + \cdots + \phi_p \mathsf{E}(y_{t+h-p}|I_t) \\ &+ \quad \mathsf{E}(\varepsilon_{t+h}|I_t) + \theta_1 \mathsf{E}(\varepsilon_{t+h-1}|I_t) + \cdots + \theta_q \mathsf{E}(\varepsilon_{t+h-q}|I_t) \end{aligned}$$

It is then possible to recursively compute the optimal $h-$step ahead predictor $y_t(h)$ given that

$$\mathsf{E}(y_{t+h-i}|I_t) = \begin{cases} y_{t+h-i}, & i \geq h \\ y_t(h-i), & i < h \end{cases}$$

$$\mathsf{E}(\varepsilon_{t+h-i}|I_t) = \begin{cases} \varepsilon_{t+h-i}, & i \geq h \\ 0, & i < h \end{cases}$$

<u>Example</u>: Let assume that $y_t \sim$ARMA(1,1), we get

$$y_t(h) = m + \phi\mathsf{E}(y_{t+h-1}|I_t) + \mathsf{E}(\varepsilon_{t+h}|I_t) + \theta\mathsf{E}(\varepsilon_{t+h-1}|I_t),$$

which implies

$$\begin{aligned} y_t(1) &= m + \phi y_t + \theta\varepsilon_t \\ y_t(h) &= m + \phi y_t(h-1) = m(1+\phi+\cdots+\phi^{h-2}) + \phi^{h-1} y_t(1) \\ &= m(1+\phi+\cdots+\phi^{h-1}) + \phi^h y_t + \phi^{h-1}\theta\varepsilon_t, \quad h > 1 \end{aligned}$$

Since any ARMA$(p,q)$ admits the Wold representation $y_t = \mu + \psi(L)\varepsilon_t$, where $\psi(L) = \theta(L)/\phi(L)$, we can rewrite $h-$step ahead predictor as

$$y_t(h) - \mu = \mathsf{E}(\textstyle\sum_{j=0}^{h-1}\psi_j\varepsilon_{t+h-j} + \sum_{j=h}^{\infty}\psi_j\varepsilon_{t+h-j}|I_t) = \sum_{j=h}^{\infty}\psi_j\varepsilon_{t+h-j}$$

Hence, the $h-$step ahead prediction error is

$$\varepsilon_t(h) = y_{t+h} - y_t(h) = \textstyle\sum_{j=0}^{h-1}\psi_j\varepsilon_{t+h-j}$$

Since $\varepsilon_t(h) \sim$MA$(h-1)$, we have that

$$\begin{array}{rcl} \mathsf{E}[\varepsilon_t(h)] & = & 0, \\ \sigma^2(h) & \equiv & \mathsf{Var}[\varepsilon_t(h)] \;=\; \sigma^2\textstyle\sum_{j=0}^{h-1}\psi_j^2, \end{array}$$

We note that $\sigma^2(h)$ is a non-decreasing function of $h$ such that

$$\lim_{h\to\infty}\sigma^2(h) = \gamma(0)$$

When $\varepsilon_t$ is a Gaussian white-noise, it follows that

$$\varepsilon_t(h)/\sigma(h) \sim N(0,1)$$

since $\varepsilon_t(h)$ is a linear combination of i.i.d. $N(0,\sigma^2)$ random variables. Hence, the $100(1-\alpha)\%$ confidence interval for $y_{t+h}$ is

$$y_t(h) - z_{\alpha/2}\sigma(h) < y_{t+h} < y_t(h) + z_{\alpha/2}\sigma(h)$$

Remark: when the model parameters are estimated, the above formula underestimates the true sample variability.

# 4 Nonstationary processes

Integrated processes: $y_t$ is said to be integrated of order $d$, $y_t \sim I(d)$, if

$$\Delta^d y_t = \mu + \psi(L)\varepsilon_t, \quad \textstyle\sum_{j=0}^{\infty}|\psi_j| < \infty$$

where $\Delta = 1 - L \Rightarrow \Delta^2 = (1-L)^2 = 1 - 2L + L^2$.

Random Walk (RW): the RW is generated by the equation

$$y_t = \mu + y_{t-1} + \varepsilon_t = y_0 + \mu t + \textstyle\sum_{j=0}^{t-1}\varepsilon_{t-j}$$

Note: RW$\sim I(1)$ since $\Delta y_t = \mu + \varepsilon_t$, $\mu$ is the "drift" of the RW.

Integrated Random Walk (IRW): the IRW is generated by the equation

$$y_t = 2y_{t-1} - y_{t-2} + \varepsilon_t = y_0 + \Delta y_0 t + \textstyle\sum_{i=0}^{t-1}\sum_{j=i}^{t-1}\varepsilon_{t-j}$$

Note: IRW$\sim I(2)$ since $\Delta^2 y_t = \varepsilon_t$.

## 4.1 The Box-Jenkins Approach

1. Identification of the orders $p, d, q, P, D, Q$. The integration orders are determined first. The MA and AR orders are determined from the analysis of the correlogram.

2. Estimation of the parameters (maximum likelihood)

3. Diagnostic checking and goodness of fit

   - Significance tests for the parameters

   - Normality tests on residuals $e_t = \frac{\widehat{\theta}(L)}{\widehat{\phi}(L)} y_t$

- Autocorrelation tests on residuals. Ljung-Box test statistic

$$Q(m) = T(T+2)\sum_{k=1}^{m}(T-k)^{-1}\hat{\rho}_e^2(k)$$

Under $H_0$, $Q(m) \sim \chi^2$ with $m-(p+q)$ degrees of freedom.

- Goodness of fit. Coefficient of determination.

- Selection criteria. Choose the ARMA$(p^*, q^*)$ such that either

$$\min\left\{AIC(p,q) = \ln\hat{\sigma}^2 + 2\frac{p+q}{T}\right\},$$

or

$$\min\left\{BIC(p,q) = \ln\hat{\sigma}^2 + \ln T\frac{p+q}{T}\right\}.$$

# 5 Unit-roots in macroeconomic time series

Consider the AR(1) process $y_t = \phi y_{t-1} + \varepsilon_t$. The test of $H_0 : \phi = 1$ versus $H_1 : \phi < 1$ is known as a unit root test. We may reparametrize the model as

$$\Delta y_t = \rho y_{t-1} + \varepsilon_t, \quad \rho = (\phi - 1),$$

and formulate the unit root test as $H_0 : \rho = 0$ versus $H_1 : \rho \in (-2, 0)$.

Dickey-Fuller test: The test statistic is

$$\hat{\tau} = \frac{\widehat{\rho}}{s} \left( \sum_{t=2}^{T} y_{t-1}^2 \right)^{\frac{1}{2}},$$

where $s^2$ is the residual variance and $\widehat{\rho}$ is the OLS estimate of the regression coefficient. Under $H_0$, $\hat{\tau}$ has no longer a limit $N(0, 1)$ distribution. Its critical values were tabulated by Fuller (1977).

The distribution of the test statistic is not invariant to the deterministic kernel.

If a constant term is included in the model, we have $y_t = \alpha + u_t$, where $u_t = \phi u_{t-1} + \varepsilon_t$. We may reparametrize the model as

$$\Delta y_t = -\rho\alpha + \rho y_{t-1} + \varepsilon_t, \quad \rho = (\phi - 1),$$

and the $t$ statistic for the null $H_0 : \rho = 0$ is denoted $\hat{\tau}_\mu$. Critical values are in Fuller (1977).

If the model is extended to allow for a linear trend, we have $y_t = \alpha + \beta t + u_t$. We may reparametrize the model as

$$\Delta y_t = (\phi\beta - \rho\alpha) - \rho\beta t + \rho y_{t-1} + \varepsilon_t, \quad \rho = (\phi - 1),$$

and the $t$ statistic for the null $H_0 : \rho = 0$ is denoted $\hat{\tau}_\tau$. Critical values are in Fuller (1977).

## 5.1 Deterministic Trends vs. Random Walks.

Nelson and Plosser (1982) contrast 2 candidate DGPs:

- Trend-stationary (TS) processes: $y_t = \alpha^* + \beta^* t + I(0)$

- Difference-stationary (DS) processes: $y_t \sim I(1) +$ drift.

Both are nested into the process

$$\phi(L) y_t = \alpha^* + \beta^* t + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2) \tag{1}$$

$y_t \sim$ DS if $\phi(1) = 0$ (and $\beta^* = 0$).

Rewriting the AR polynomial as

$$\phi(L) = \phi(1)L + \Delta\phi^\dagger(L),$$

with

$$\phi^\dagger(L) = 1 - \phi_1^\dagger L - \ldots - \phi_{p-1}^\dagger L^{p-1}, \;\; \phi_j^\dagger = -\sum_{i=j+1}^{p} \phi_i$$

we can rewrite (1):

$$\Delta y_t = \alpha^* + \beta^* t + \rho y_{t-1} + \sum_{j=1}^{p-1} \phi_j^\dagger \Delta y_{t-j} + \varepsilon_t, \tag{2}$$

where $\rho = -\phi(1) = \sum_{j=1}^{p} \phi_j - 1$.

The test of $H_0 : \rho = 0$ is known as the Augmented Dickey-Fuller (ADF) test. The test statistic has the same limit distribution as $\hat{\tau}_\tau$.

Notice that under the null $y_t \sim I(1)$, that is $y_t \in$ DS. The alternative is $\rho < 0$ and implies that $y_t \in$ TS.

In their highly influential paper, N&P applied the ADF test to a subset of annual US macroeconomic time series and concluded that most series were consistent with the DS hypothesis.

As we shall see, a DS process is driven by a stochastic trend, or permanent component. Once it is proven that there exists a permanent component, the next step is to determine its importance.

Another stylized fact is that the ACF of $\Delta \ln y_t$ displays a non-negative value at lag 1 and is zero elsewhere. This would imply that most of the fluctuations are permanent (i.e. the transitory component makes little contribution) and real disturbances are a more important source of output fluctuations than monetary disturbances.