Univariate Time Series Models

Gianluca Cubadda
Università di Roma "Tor Vergata"

Program

- Stationary time series: Basic concepts. Stationarity, Total and partial autocorrelation, Ergodicity, Linear stationary processes, ARMA models, Outliers, Forecasting.

- Nonstationary time series: ARIMA models, The Beveridge-Nelson Trend-Cycle decomposition, Seasonality,

- Statistical inference: Estimation, Identification, Diagnostic checking.

- Unit roots in economic and financial time series: Deterministic trends vs. random walks, Unit-roots tests, Impulse response function and measures of persistence.

List of references

- Brockwell and Davis (2002) Introduction to Time Series and Forecasting, second edition, Springer-Verlag, New York.

- Hamilton (1994), Time Series Analysis, Princeton University Press.

- Wei (2006) Time Series Analysis: Univariate and Multivariate Methods, second edition, Addison-Wesley.

1 Univariate time series analysis: Basic concepts

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

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.

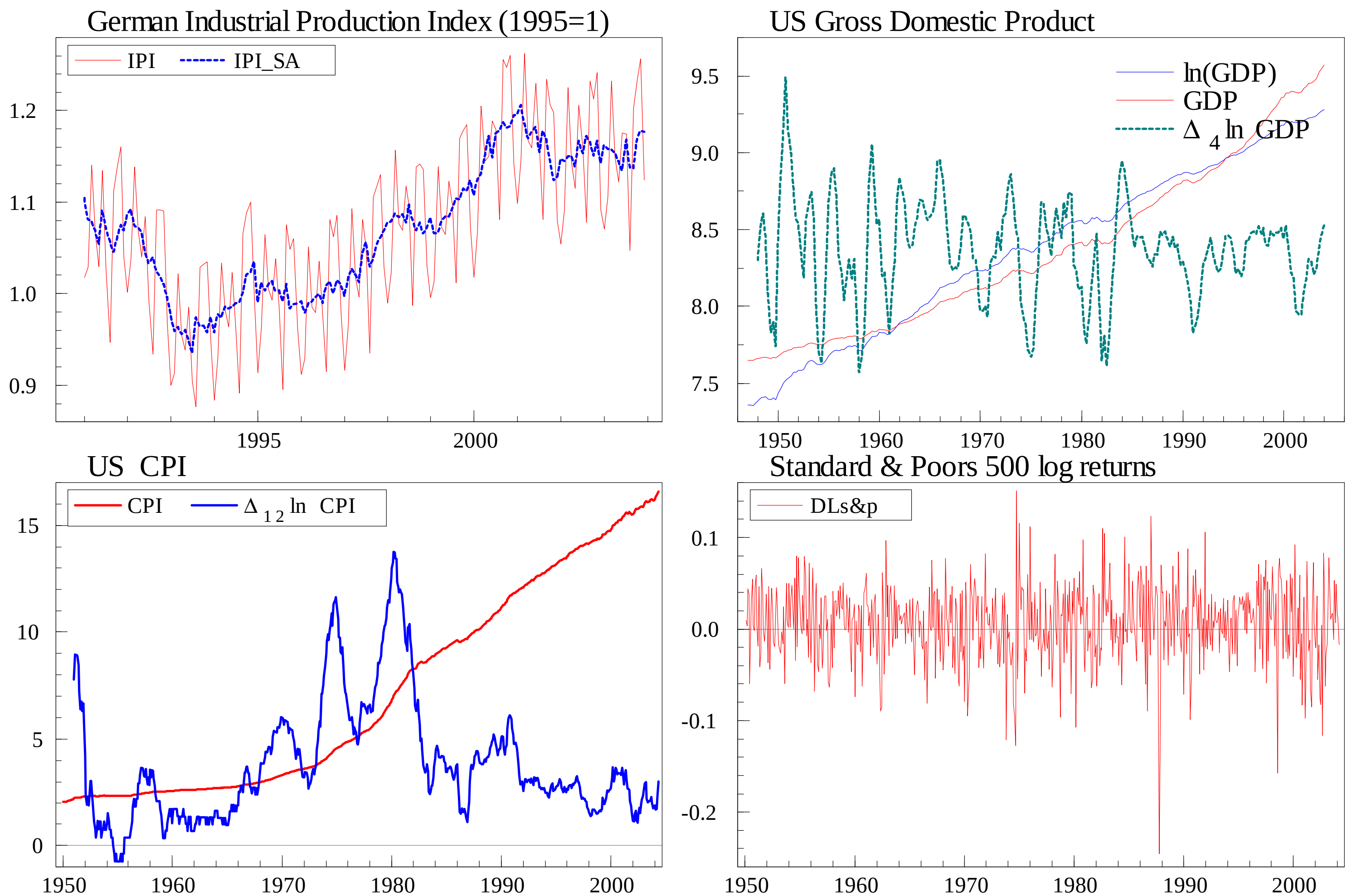


Figure 1: Plots of various time series

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 (Ω, F, P), where the integer number t is a time-index, Ω is the sample space, F is a sigma algebra defined on Ω and P is a probability measure on Ω. 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 $\forall t, k \in Z$:

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

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

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

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 $\forall t, h \in Z$.

Autocovariance function, $\gamma(k)$, is symmetric: $\gamma(k) = \gamma(-k)$.
The partial autocovariance function 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.

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

Autocorrelation function (ACF):

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

i) $\rho(0) = 1$; ii) $|\rho(k)| < 1$; iii) $\rho(k) = \rho(-k)$.
The partial autocorrelation function (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): $\varepsilon_t \sim \mathsf{WN}(\sigma^2)$,

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

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 $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ (square summability).
For a stationary linear process we have:

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

2.1 Estimation

- sample mean $\hat{\mu} = \bar{y} = T^{-1} \sum\limits_{t=1}^{T} y_t$
- 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 $y_t \sim \mathrm{WN}(\sigma^2)$, then $T^{1/2}\hat{\rho}(k) \xrightarrow{d} \mathrm{N}(0, 1)$.

2.2 Ergodicity

Ergodicity: A stationary stochastic process is ergodic for the first and second moments when the sample mean and autocovariance function are mean-square consistent. Notice that stationarity itself does not imply ergodicity.

A sufficient condition for ergodicity for the first two moments of a linear process is that

$$\sum_{j=0}^{\infty}|\psi_j| < \infty$$

(absolute summability). This implies square summability.

The above condition is equivalent to require that the autocovariance function is absolutely summable:

$$\sum_{k=0}^{\infty}|\gamma(k)| < \infty$$

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 $\sum_{j=0}^{\infty}|\psi_j| < \infty$ implies that $\lim_{j\to\infty}\psi_j = 0$, 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

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

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.

Inversion of a first-order polynomial: Consider $\phi(L) = (1 - \phi L)$ 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) process: 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 $1 - \phi L$, is greater then 1 in modulus.

The expected value can be easily computed from either the AR or the Wold representation:

$$\mathsf{E}(y_t) \;=\; \mu = m/(1-\phi)$$

The demeaned AR(1) process: In view of the above equation, we see that any stationary AR(1) process can be rewritten as a zero-mean AR(1) process:

$$(y_t - \mu) = \phi(y_{t-1} - \mu) + \varepsilon_t$$

From the demeaned AR(1) process we easily get

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

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

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

since $\mathrm{E}[(y_{t-1}-\mu)\varepsilon_t] = \mathrm{E}[(\varepsilon_{t-1} + \phi\varepsilon_{t-2} + \phi^2\varepsilon_{t-3} + \cdots)\varepsilon_t] = 0$.

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 $k > 1$.

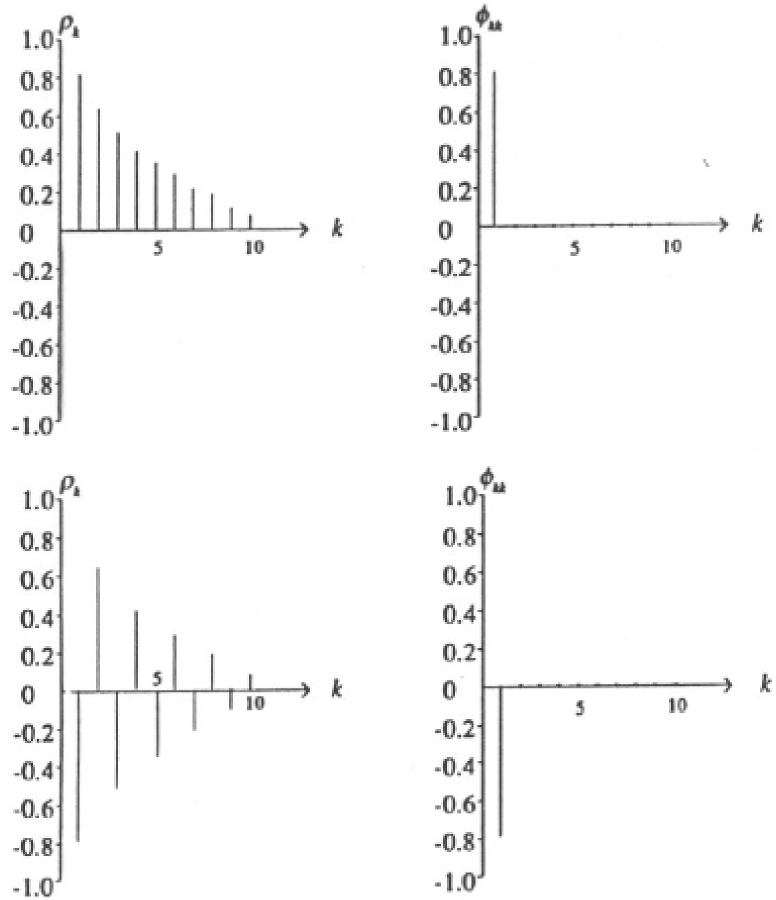


Fig. 3.1 ACF and PACF of the AR(1) process: $(1-\phi B)\dot{Z}_t = a_t$.

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

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

or equivalently

$$\Delta y_t = m + \varepsilon_t \tag{1}$$

It is a non-stationary AR(1) process since $\phi = 1$. By recursive substitutions we get

$$y_t = y_0 + mt + \sum_{j=0}^{t-1} \varepsilon_{t-j}$$

The above result can also be obtained by applying the cumulating operator

$$S(L) = (1 + L + L^2 + \cdots + L^{t-1})$$

on both sides of (1) and noticing

$$S(L)(1 - L) = 1 - L^t$$

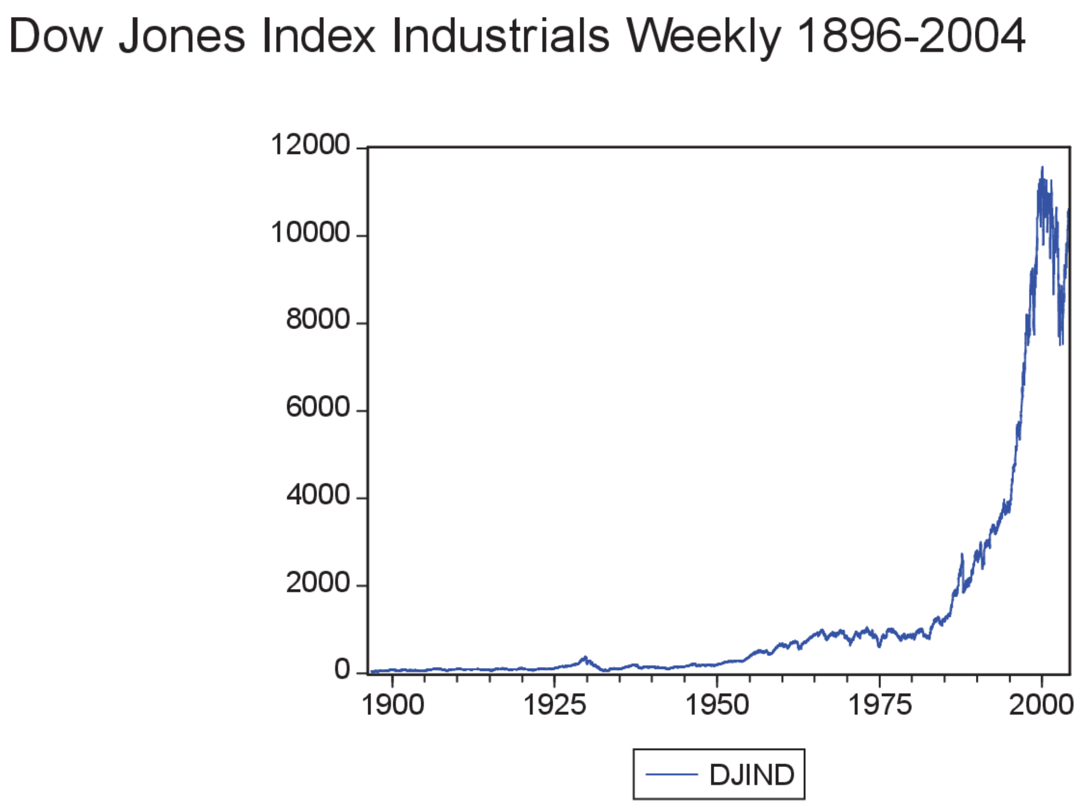
Treating the initial condition y_0 as fixed, it's easy to compute the moments:

$$\begin{aligned} \mathsf{E}(y_t) &= y_0 + mt \\ \mathsf{Var}(y_t) &= t\sigma^2 \\ \mathsf{Cov}(y_t, y_{t-k}) &= (t-k)\sigma^2 \;\; 0 \le k < t \end{aligned}$$

which vary over time.

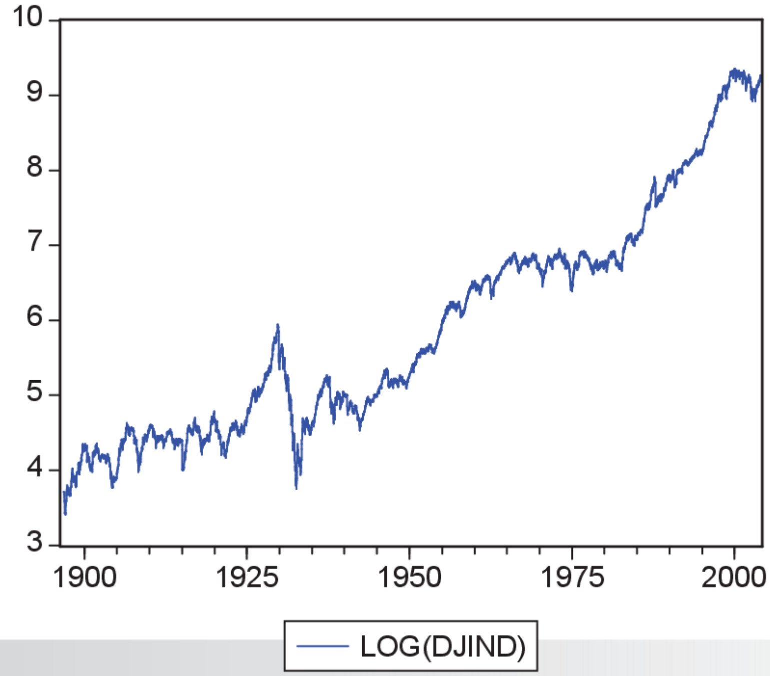
The term m is called the <u>drift</u> of the RW.

Dow Jones: Nonstationary series



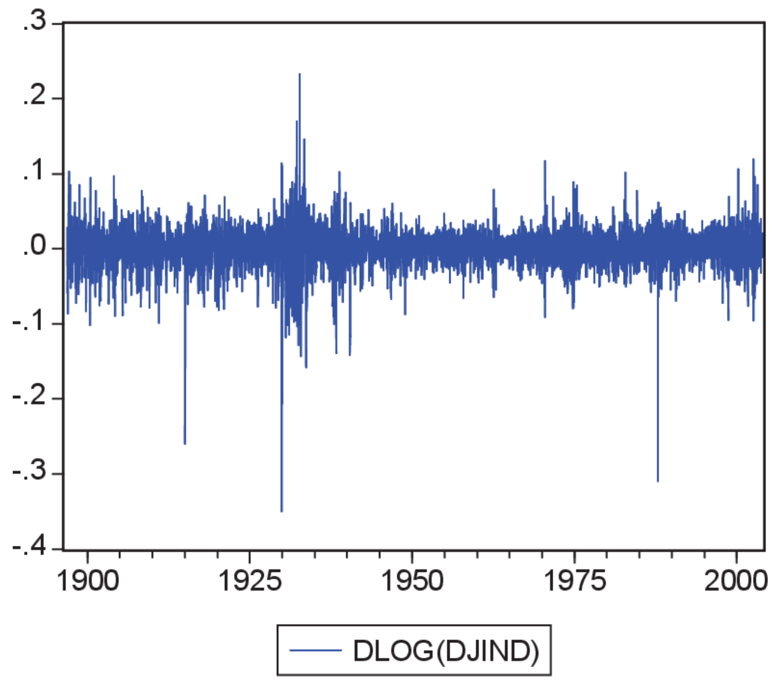
Log Dow Jones: Nonstationary series

Log transformed Dow Jones Index Industrials Weekly 1896-2004. *Exercise (2)*: approximate mean weekly growth.



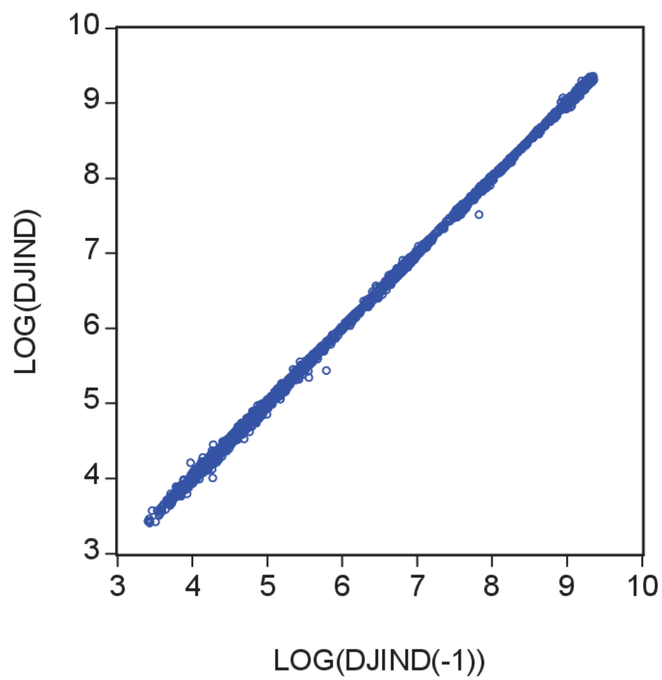
Returns (DLog) DJI: Nonstationary series?

Returns Dow Jones Index Industrials Weekly 1896-2004



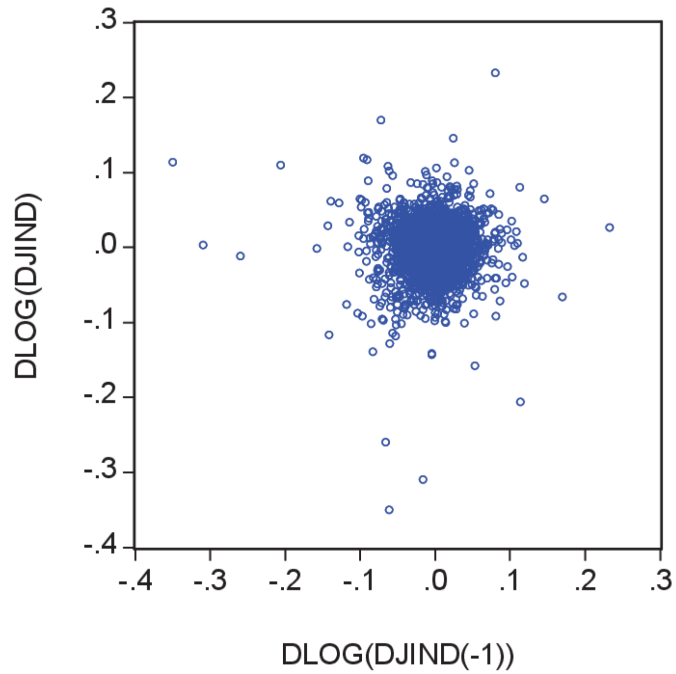
Scatter plot log Dow Jones vs. (-1)

Log Dow Jones Index Industrials Weekly 1896-2004



Scatter plot dlog Dow Jones vs (-1)

Returns (dlog) Dow Jones Index Industrials Weekly 1896-2004



SACF/PACF Log Dow Jones

Log Dow Jones Index Industrials Weekly 1896-2004

Correlogram of LOG(DJIND)

Date: 02/05/04 Time: 16:08
Sample: 5/27/1896 1/27/2004
Included observations: 5596

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.999	0.999	5590.1	0.000
		2	0.998	-0.006	11172.	0.000
		3	0.998	-0.009	16746.	0.000
		4	0.997	-0.004	22312.	0.000
		5	0.996	-0.004	27870.	0.000
		6	0.995	-0.004	33420.	0.000
		7	0.994	-0.001	38961.	0.000
		8	0.993	-0.005	44494.	0.000
		9	0.993	-0.005	50018.	0.000
		10	0.992	-0.004	55534.	0.000
		11	0.991	0.002	61042.	0.000
		12	0.990	-0.002	66541.	0.000
		13	0.989	-0.001	72031.	0.000
		14	0.988	0.001	77514.	0.000
		15	0.988	0.009	82988.	0.000
		16	0.987	0.011	88453.	0.000
		17	0.986	0.002	93911.	0.000
		18	0.985	0.002	99361.	0.000
		19	0.984	-0.002	104803	0.000
		20	0.983	0.002	110237	0.000

SACF/PACF dLog Dow Jones

Returns (dlog) Dow Jones Index Industrials Weekly 1896-2004

Correlogram of DLOG(DJIND)

Date: 02/05/04 Time: 16:09
Sample: 5/27/1896 1/27/2004
Included observations: 5595

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.021	0.021	2.5084	0.113
		2	0.044	0.044	13.432	0.001
		3	0.022	0.020	16.163	0.001
		4	-0.009	-0.012	16.596	0.002
		5	-0.008	-0.009	16.938	0.005
		6	0.006	0.007	17.154	0.009
		7	0.010	0.011	17.733	0.013
		8	-0.019	-0.020	19.785	0.011
		9	0.027	0.026	23.818	0.005
		10	0.002	0.002	23.842	0.008
		11	-0.015	-0.016	25.109	0.009
		12	0.004	0.003	25.209	0.014
		13	-0.008	-0.007	25.591	0.019
		14	-0.024	-0.023	28.872	0.011
		15	-0.008	-0.007	29.239	0.015
		16	-0.025	-0.023	32.651	0.008
		17	0.044	0.048	43.461	0.000
		18	-0.011	-0.011	44.080	0.001
		19	-0.021	-0.025	46.634	0.000
		20	0.019	0.020	48.709	0.000

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 $1 - \phi_1 L - \phi_2 L^2$ are greater than 1 in modulus (lie outside the unit circle).

The above condition is equivalent to: i) $|\phi_2| < 1$ and ii) $|\phi_1| < 1 - \phi_2$. The stationarity region of the AR(2) parameters lies inside the triangle with vertices (-2,-1),(2,-1),(0,1). A pair of complex conjugate roots arises for $\phi_1^2 + 4\phi_2 < 0$.

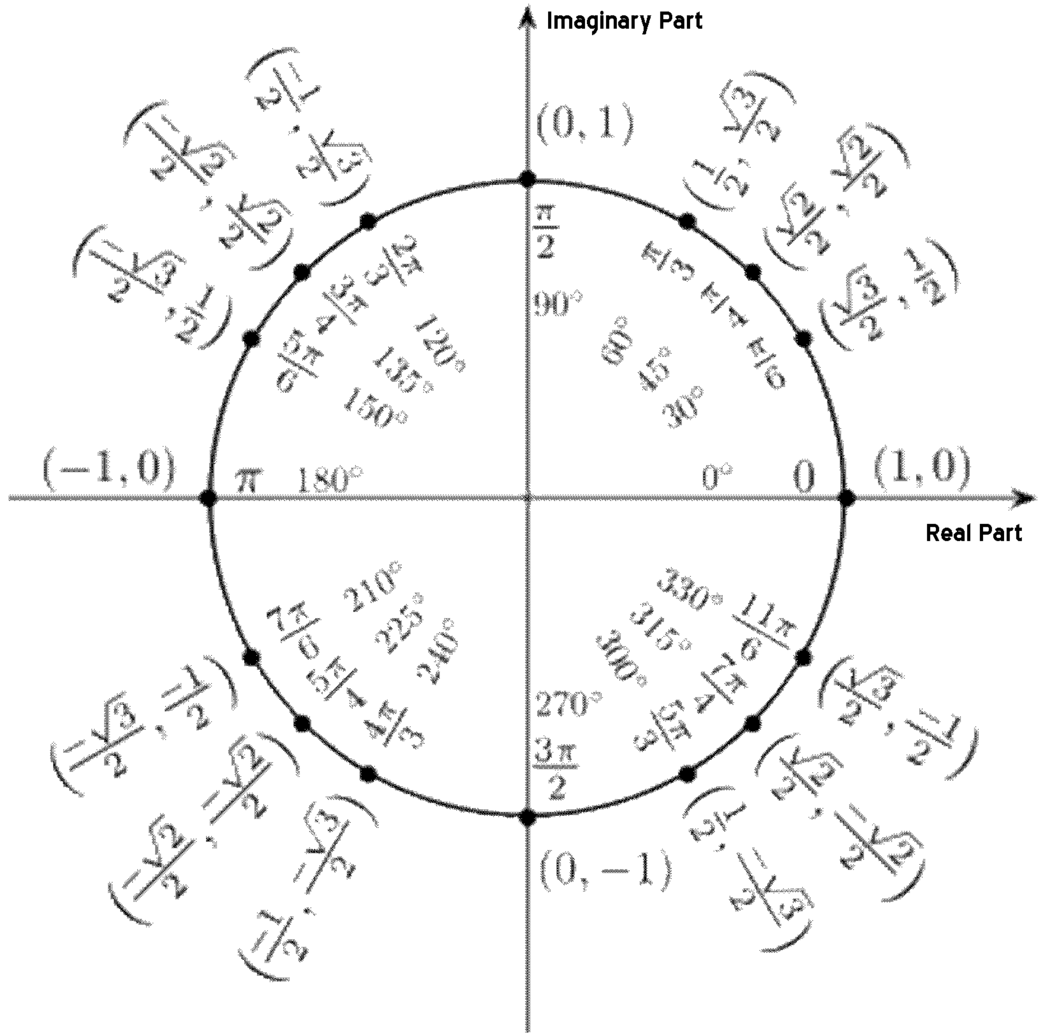


Figure 2: The complex unit circle

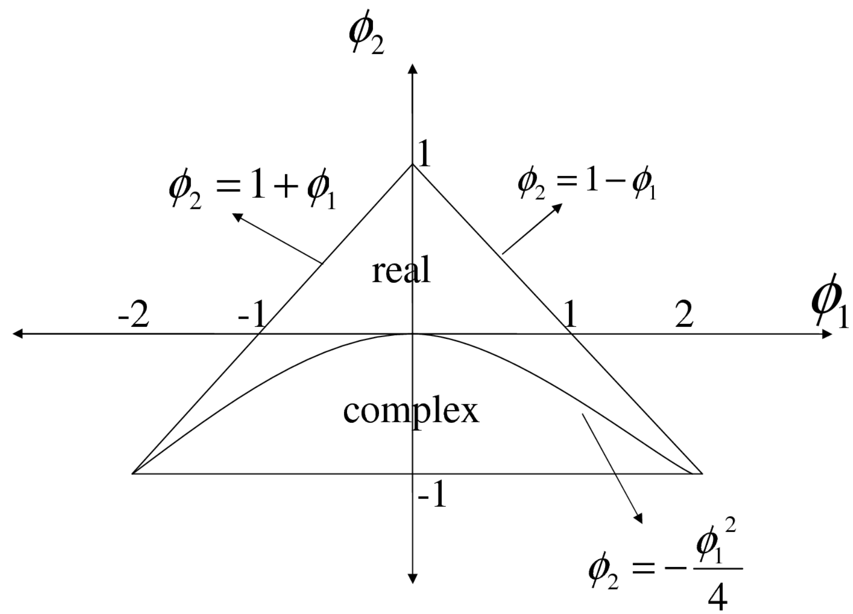


Figure 3: Stationarity region of an AR(2) process

Under stationarity, we can easily compute the first two moments of the AR(2) process:

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

- The demeaned AR(2) process:

$$(y_t - \mu) = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \varepsilon_t$$

- 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 from the demeaned AR(2) equation as follows:

$$
\begin{aligned}
\gamma(0) &= \mathsf{E}[\phi_1(y_{t-1}-\mu)y_t + \phi_2(y_{t-2}-\mu)y_t + \varepsilon_t y_t] \\
&= \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma^2 \\
\gamma(1) &= \mathsf{E}[\phi_1(y_{t-1}-\mu)y_{t-1} + \phi_2(y_{t-2}-\mu)y_{t-1} + \varepsilon_t y_{t-1}] \\
&= \phi_1\gamma(0) + \phi_2\gamma(1) \\
\gamma(k) &= \mathsf{E}[\phi_1(y_{t-1}-\mu)y_{t-k} + \phi_2(y_{t-2}-\mu)y_{t-k} + \varepsilon_t y_{t-k}] \\
&= \phi_1\gamma(k-1) + \phi_2\gamma(k-2), \quad k = 2, 3, \ldots
\end{aligned}
$$

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

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

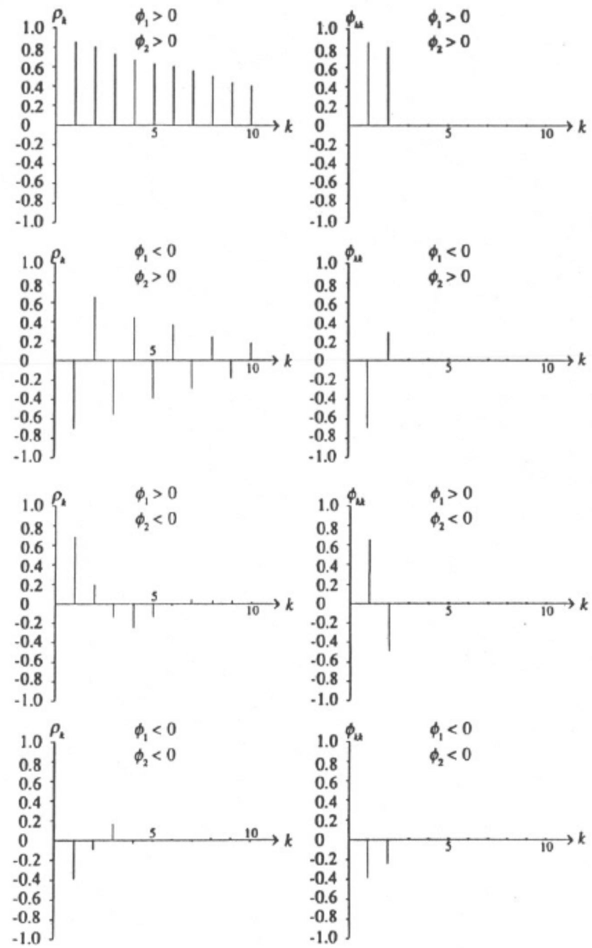


Fig. 3.7 ACF and PACF of AR(2) process: $(1-\phi_1 B-\phi_2 B^2)\dot{Z}_t = a_t$.

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

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

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

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

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

- The demeaned process is $\phi(L)(y_t - \mu) = \varepsilon_t$

- 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 $\varepsilon_t \sim \text{WN}(\sigma^2)$.

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

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 $k = 1$:

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

<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 μ, $\gamma(0)$ 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 the first two moments. This problem is known as non-identifiability and is remedied upon by constraining θ in the interval (-1,+1).

The term invertibility stems from the possibility of rewriting the process as an infinite autoregression, AR(∞), with coefficients π_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(∞), its PACF has no cutoff but it decays exponentially.

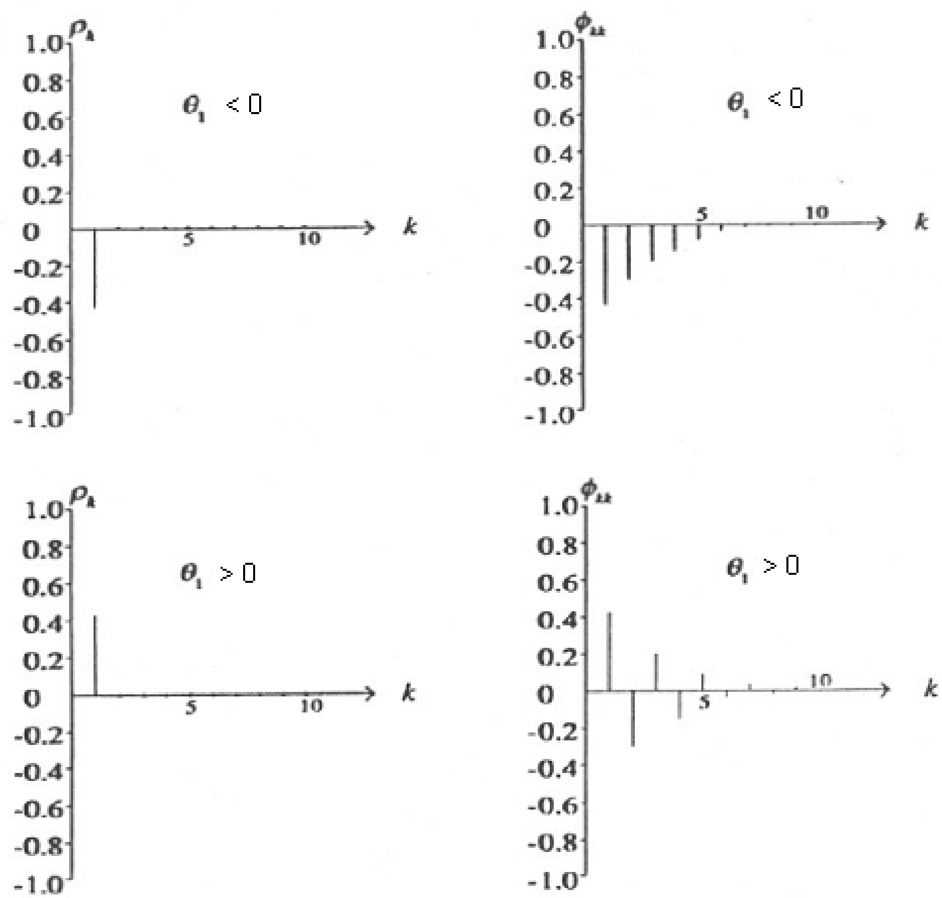


Fig. 3.10 ACF and PACF of MA(1) processes

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 as 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

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

where

$$\begin{aligned}\phi(L) &= 1 - \phi_1 L - \cdots - \phi_p L^p, \\ \theta(L) &= 1 + \theta_1 L + \cdots + \theta_q L^q\end{aligned}$$

Remark: the polynomials $\phi(L)$ and $\theta(L)$ should have no common roots, otherwise a reduction of the orders p, q would occur after cancelling the common roots.

Stationarity: y_t is stationary if the roots of the AR polynomial $\phi(L)$ lie outside the unit circle.
Stationarity implies that y_t can be written as an MA(∞) process with declining coefficients:

$$y_t = \phi(1)^{-1}m + \phi(L)^{-1}\theta(L)\varepsilon_t = \mu + \psi(L)\varepsilon_t$$

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(∞) process with declining coefficients:

$$\theta(L)^{-1}\phi(L)(y_t - \mu) = \pi(L)(y_t - \mu) = \varepsilon_t$$

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 expected value is:

$$\mathsf{E}(y_t) \;\;=\;\; m + \phi \mathsf{E}(y_{t-1}) = m + \phi\mu = m/(1-\phi)$$

- The demeaned process is:

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

- The autocovariance function is obtained as follows

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

- Both ACF and PACF have no cutoff!

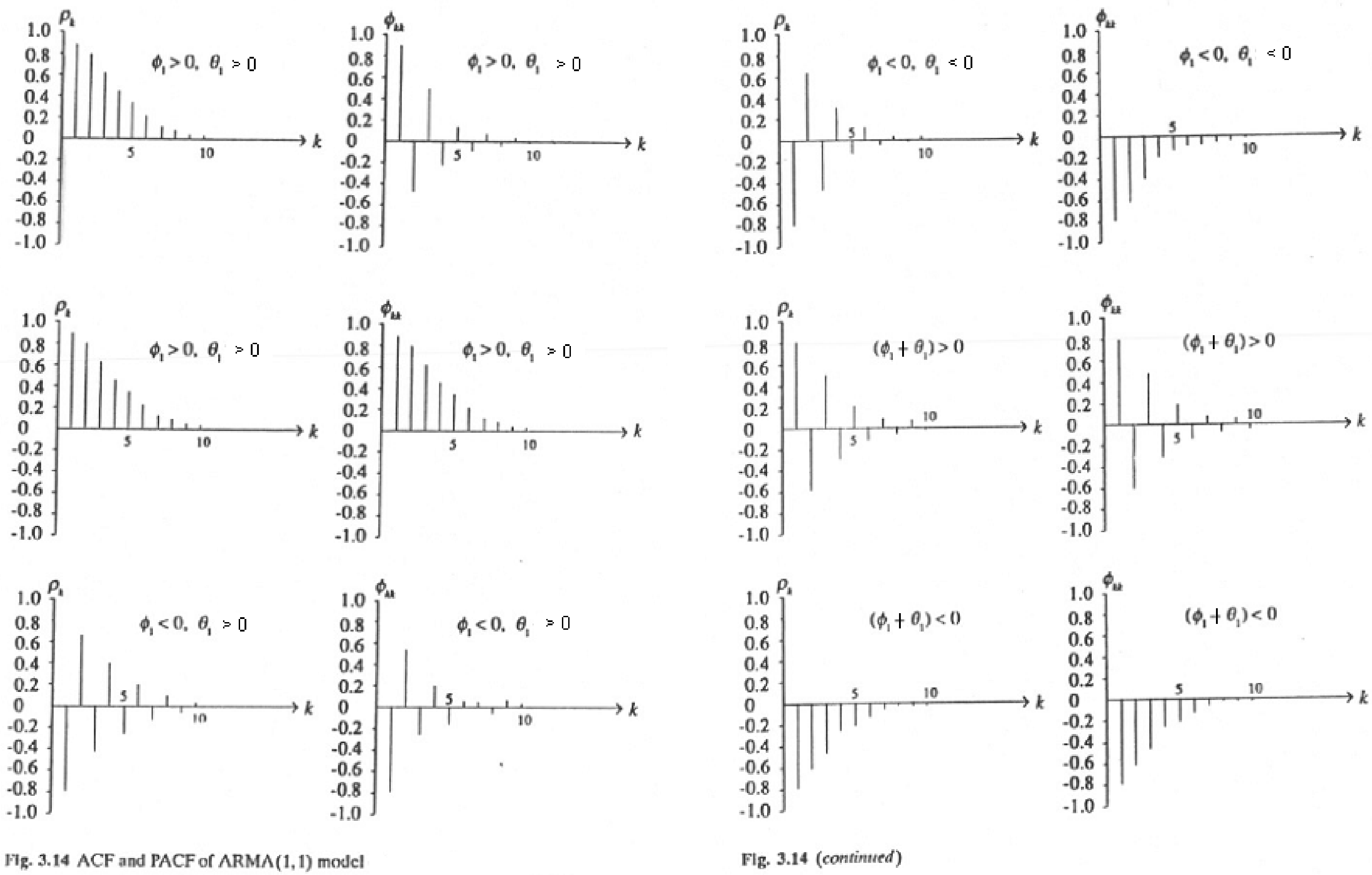


Figure 7: ACF and PACF of an ARMA(1,1) process.

3.4 Outliers in ARMA models

- Many economic and financial time series display outlying observations, which are due to specific events (errors in the data, strikes, changes in regulations, natural disasters, etc.).

- The presence of these outliers may induce misspecification of the ARMA model as well as biases in parameter estimation.

- Outliers can be modelled as additional deterministic components to the basic ARMA set-up. Their treatment leads to ARMA models with exogenous variables (ARMAX) that are impulse dummies:

$$I_t(\tau) = \begin{cases} 1 & \text{if } \tau = t \\ 0 & \text{if } \tau \neq t \end{cases}$$

  where τ is the time in which the event generating the outlier occurs.

- Additive Outlier (AO):

$$y_t = \mu + \frac{\theta(L)}{\phi(L)}\varepsilon_t + \omega I_t(\tau), \quad \omega \in \mathbb{R}$$

  An AO shows up at time τ only with magnitude ω, subsequent observations are unaffected.

- Innovational Outlier (IO):

$$y_t = \mu + \frac{\theta(L)}{\phi(L)}[\omega I_t(\tau) + \varepsilon_t],$$

  An IO is characterized by having an impact of magnitude $\psi_h \omega$ at time $\tau + h$ for $h = 0, 1, 2, ...$

- Transient Change (TC):

$$y_t = \mu + \frac{\theta(L)}{\phi(L)}\varepsilon_t + \frac{\omega}{1-\delta L} I_t(\tau), \quad |\delta| < 1$$

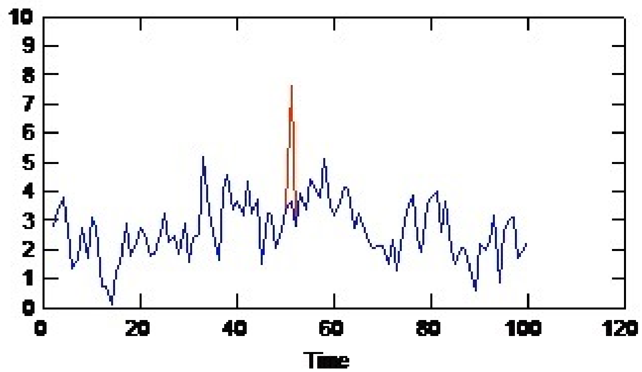
  A TC is characterized by having an impact of magnitude $\delta^h \omega$ at time $\tau + h$. Asymptotically, $\mathrm{E}(y_{\tau+h})$ returns to μ.

- Level Shift (LS):

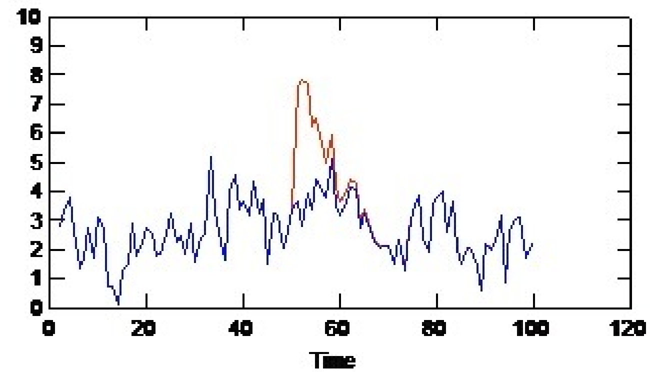
$$y_t = \mu + \frac{\theta(L)}{\phi(L)}\varepsilon_t + \frac{\omega}{1-L} I_t(\tau),$$

  A LS implies that $\mathrm{E}(y_{\tau+h})$ move permanently to $(\mu + \omega)$. In contrast to a TC, a LS affects all the subsequent observations forever.

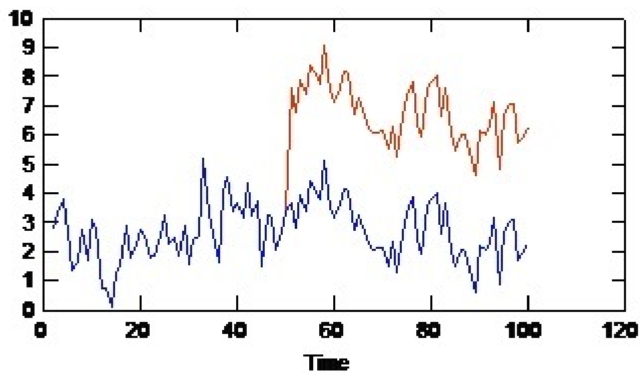
Additive Outlier



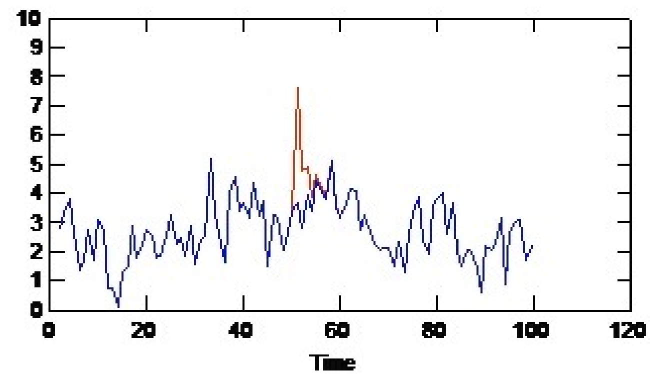
Innovational Outlier



Level Shift Outlier



Transient Change Outlier



3.5 Forecasting from ARMA Models

Let $y_t \sim$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

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

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

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

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

Example: 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 ε_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.

3.6 Forecasting the random walk

In order to compute the $h-$step ahead prediction of the RW (with drift), it is convenient to rely on the expression:

$$y_{t+h} = mh + y_t + \varepsilon_{t+h} + \varepsilon_{t+h-1} + \cdots + \varepsilon_{t+1},$$

which is easily obtained by recursive substitutions.
From the above equation we get:

$$\begin{aligned} y_t(h) &= mh + y_t \\ \varepsilon_t(h) &= \varepsilon_{t+h} + \varepsilon_{t+h-1} + \cdots + \varepsilon_{t+1} \\ \sigma^2(h) &= h\sigma^2 \end{aligned}$$

<u>Remark</u>: $\lim_{h\to\infty} \sigma^2(h) = \infty \Rightarrow$ it is less and less likely to find y_{t+h} close to $y_t(h)$ as the forecasting horizion h increases.

4 Nonstationary processes

Integrated processes: An ARMA process is integrated of order d, denoted as $y_t \sim I(d)$ or $y_t \sim$ ARIMA(p, d, q), if

$$\phi(L)\Delta^d y_t = m + \theta(L)\varepsilon_t,$$

where $\Delta = 1 - L$, and the roots of both $\phi(L)$ and $\theta(L)$ lie outside the unit circle.
The simplest $I(1)$ process is the RW, i.e the ARIMA$(0, 1, 0)$:

$$\Delta y_t = m + \varepsilon_t$$

The simplest $I(2)$ process is the Integrated Random Walk, i.e the ARIMA$(0, 2, 0)$:

$$\Delta^2 y_t = (1 - 2L + L^2)y_t = \varepsilon_t,$$

which can be rewritten by recursive substitutions as:

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

4.1 The Beveridge-Nelson decomposition

Beveridge and Nelson (1981) proved that any ARIMA($p,1,q$) process can be decomposed as the sum of a RW (with drift) and an I(0) process. Indeed, by expanding the polynomial $\psi(L)$ on 1 we get

$$\psi(L) = \psi(1) + (1-L)\psi^*(L)$$

where $\psi^*(L) = \sum_{j=0}^{\infty} \psi_j^* L^J$ is such that $\psi_j^* = -\sum_{i>j}^{\infty} \psi_i$ and $\lim_{j\to\infty} |\psi_j^*| = 0$. Hence, we can rewrite the Wold representation as

$$\Delta y_t = \mu + \psi(1)\varepsilon_t + (1-L)\psi^*(L)\varepsilon_t, \tag{2}$$

Applying $S(L)$ to both sides of (2) and assuming $\varepsilon_t = 0$ for $t \leq 0$ we get

$$y_t = \underbrace{y_0 + \mu t + \sum_{j=0}^{t-1} \psi(1)\varepsilon_{t-j}}_{\text{RW with drift = "stochastic trend"}} + \underbrace{\psi^*(L)\varepsilon_t}_{\text{I(0) = "cycle"}}$$

4.2 Seasonality

Seasonal Differences:

$$\Delta_s = 1 - L^s = (1 - L)(1 + L + \cdots + L^{s-1})$$

where s denotes the number of seasons in an year. Usually, $s = 4$ or 12.

Seasonal Integration ($s = 4$):

$$\Delta_4 = (1 - L)(1 + L + L^2 + L^3) = (1 - L)(1 + L)(1 + L^2),$$
$$(1 + L^2) = (1 - iL)(1 + iL),$$

where $i = (-1)^{1/2}$. Hence, Δ_4 has 4 roots on the unit circle. A similar result applies to the monthly case ($s = 12$) as well.

Let us write $f_\omega(t) = a\cos(t\omega) + b\sin(t\omega)$, a function with period $P = 2\pi/\omega$, where $\omega \in [0, \pi]$. Then we have

$$\begin{array}{rcll}
(1-L)f_0(t) & = & (1-L)[a\cos(t0)] & = 0 \\
(1+L)f_\pi(t) & = & (1+L)[a\cos(t\pi)] & = 0 \\
(1+L^2)f_{\pi/2}(t) & = & (1+L^2)[a\cos(t\pi/2) + b\sin(t\pi/2)] & = 0
\end{array}$$

Hence, we define

$$y_t \sim I_\pi(1) \text{ iff } (1+L)y_t \sim I(0)$$
$$y_t \sim I_{\pi/2}(1) \text{ iff } (1+L^2)y_t \sim I(0)$$
$$y_t \sim I_0(1) \text{ iff } (1-L)y_t \sim I(0)$$

A process y_t such that

$$(1-L^4)y_t \sim I(0),$$

is then $I(1)$ at both 0 and seasonal frequencies $\pi, \pi/2$. It is denoted $y_t \sim SI(1)$. A similar definition applies to the monthly case as well.

Seasonal Processes: The most popular seasonal generalization of ARIMA models leads to the following ARIMA$(p,d,q) \times (P,D,Q)_s$ model:

$$\phi(L)\Phi(L^s)\Delta^d\Delta_s^D y_t = \mu + \theta(L)\Theta(L^s)\varepsilon_t,$$

where $\Phi(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \cdots - \Phi_P L^{Ps}$ is the seasonal AR polynomial in L^s with order P, and $\Theta(L^s) = 1 + \Theta_1 L^s + \Theta_2 L^{2s} + \cdots + \Theta_Q L^{Qs}$ is the seasonal MA polynomial with order Q.

An important special case is the Airline model ARIMA$(0,1,1) \times (0,1,1)_s$:

$$(1-L)(1-L^s)y_t = (1+\theta L)(1+\Theta L^s)\varepsilon_t,$$

with $|\theta| < 1, |\Theta| < 1$.

The autocovariance function of the Airline model is

$$
\begin{aligned}
\gamma(0) &= (1+\theta^2)(1+\Theta^2)\sigma^2 \\
\gamma(1) &= \theta(1+\Theta^2)\sigma^2 \\
\gamma(k) &= 0 \quad \text{for } k = 2, \ldots, s-2 \\
\gamma(s-1) &= \theta\Theta\sigma^2 \\
\gamma(s) &= \Theta(1+\theta^2)\sigma^2 \\
\gamma(s+1) &= \theta\Theta\sigma^2 \\
\gamma(k) &= 0 \quad \text{for } k > s+1
\end{aligned}
$$

4.3 The Box-Jenkins Approach

1. **Identification of the orders** p, d, q (and P, D, Q for seasonal models).

   - The integration order d (and D for seasonal models) is determined first. In the past, this was basically done by graphical methods. Nowadays, we rely on unit roots tests, which we will see later.

   - In the past, The MA and AR orders p, q (and P, Q for seasonal models) were determined from the analysis of the correlogram. Nowadays, we rely on information criteria, which we will see soon.

2. **Estimation of the parameters**

- Conditional maximum likelihood (CML). Given a sample of T observations, the initial values $(y_1, ...y_p)$ and $(\varepsilon_1, ...\varepsilon_q)$ are treated as fixed $\Rightarrow$ they do not enter in the likelihood function.

- Unconditional ML (UML). The initial values are treated as realizations of the stochastic process $\Rightarrow$ they enter in the likelihood function.

- In large samples, CML and UML are equivalent.

- For pure AR processes, CML is the same as OLS.

3. **Diagnostic checking**

- Significance tests for the parameters. t tests, F tests.

- Normality tests on residuals $e_t = \frac{\widehat{\phi}(L)}{\widehat{\theta}(L)} y_t$. Jarque–Bera test; under $\mathrm{H}_0 : \varepsilon_t \sim i.i.d.N(0, \sigma^2)$, the test statistic (based on the sample 3rd and 4th moments of the residuals) converges to a $\chi^2(2)$.

- Autocorrelation tests on residuals. Ljung-Box test; under $H_0 : \varepsilon_t \sim WN$, the test statistic

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

converges to a $\chi^2(m-p-q)$.

- Goodness of fit. Coefficient of determination.

Information criteria (IC): composed by (i) an inverse measure of fit; (ii) an increasing function of the number of parameters. Given a set of candidate models, the preferred model is the one with the minimum IC $\Rightarrow$ best compromise between fit and parsimony.

In practice, estimate all the ARMA models whose orders are at maximum p_{max},q_{max}, then choose the ARMA(p^*, q^*) such that one of the following is satisfied:

$$(p^*, q^*) = \arg\min \left\{ AIC(p, q) = \ln \hat{\sigma}^2 + 2\frac{p+q}{T} \right\},$$
$$(p^*, q^*) = \arg\min \left\{ HQIC(p, q) = \ln \hat{\sigma}^2 + 2 \ln \left[\ln (T) \right] \frac{p+q}{T} \right\},$$
$$(p^*, q^*) = \arg\min \left\{ BIC(p, q) = \ln \hat{\sigma}^2 + \ln (T) \frac{p+q}{T} \right\}.$$

where AIC (HIC) [BIC] stands for Akaike (Hannan-Quinn) [Bayes] IC.

Remarks:

- For $T > 15$, the penalty term of AIC (BIC) is the smallest (largest) one $\Rightarrow$ for the sample sizes typically used in economics and finance, AIC tends to choose less parsimonious models than BIC, HQIC is a middle way.

- BIC and HQIC asymptotically choose the true orders with probability one, whereas AIC has a non-null probability of choosing an overparameterized model even in the limit.

- Notwithstanding the above, AIC is often preferred in empirical applications because the cost of underparameterization, i.e. lack of consistency, is statistically larger than the one of overparameterization i.e. lack of efficiency.

5 Unit-root tests

A stylized fact: Many (possibly log-transformed) economic and financial time series display a tendency to grow approximately linearly over time.

Nelson and Plosser (1982) contrast 2 candidate data generating processes:
Difference-stationary (DS) processes: $y_t \sim$ I(1) + drift.
Trend-stationary (TS) processes: $y_t \sim$ I(0) + linear deterministic trend
Both are nested into the unobserved components (UC) process:

$$\begin{aligned} y_t &= \alpha + \beta t + u_t, \\ \phi(L) u_t &= \varepsilon_t. \end{aligned}$$

If $\phi(L) = (1 - L)\phi^*(L)$ and all the $(p - 1)$ roots of $\phi^*(L)$ lie outside the unit circle $\Rightarrow y_t$ is DS. Notice that in this case we have $\phi(1) = 0$.
If all the roots of $\phi(L)$ lie outside the unit circle $\Rightarrow y_t$ is TS.

Dickey-Fuller (DF) test: assume that u_t is an AR(1) process. By premultiplying both sides of the first UC equation with $\phi(L)$ we get:

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

which we may reparametrize as

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

and perform a $t-$test for $H_0 : \rho = 0$, i.e. $\phi = 1$, vs. $H_1 : \rho \in (-2, 0)$, i.e. $|\phi| < 1$.

Remark 1: Under H_0, we have $\Delta y_t = \beta + \varepsilon_t \Rightarrow$ the slope β^* annihilates.

Remark 2: The limit distribution of the $t-$test statistic under H_0 is no longer a $N(0, 1)$ since the CLT does not apply to I(1) processes.

DF test with no trend: assume that the data do not display a trending behavior. Hence, $\beta = 0$ in the UC process. The DF regression becomes

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

and we still perform a $t-$test for $H_0 : \rho = 0$ vs. $H_1 : \rho \in (-2, 0)$. Notice that under H_0 we have: $\Delta y_t = \varepsilon_t \Rightarrow$ the drift is null.

DF test with no constant: assume that any deterministic term is present in the data. With $\alpha = \beta = 0$ in the UC process, the DF regression becomes

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

and we again perform a $t-$test for $H_0 : \rho = 0$ vs. $H_1 : \rho \in (-2, 0)$.

Remark: The distribution of the test statistic is not invariant to the deterministic kernel. Each of the three tests has its own critical values. They were tabulated by Fuller (1977).

Augmented Dickey-Fuller (ADF) test: assume that u_t is an AR(p) process. Rewriting the AR polynomial as

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

with

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

and premultiplying both sides of the first UC equation with $\phi(L)$ we get:

$$\Delta y_t = \underbrace{(\rho\beta + \phi^\dagger(1)\beta - \rho\alpha)}_{\alpha^*} + \underbrace{(-\rho\beta)}_{\beta^*} t + \rho y_{t-1} + \sum_{j=1}^{p-1} \phi_j^\dagger \Delta y_{t-j} + \varepsilon_t, \quad \rho = -\phi(1).$$

Finally, we perform a $t-$test for $H_0 : \rho = 0$ vs. $H_1 : \rho < 0$. Notice that under H_0 we have: $\phi^\dagger(L)\Delta y_t = \phi^\dagger(1)\beta + \varepsilon_t \Rightarrow$ the slope β^* annihilates. The limit distribution of the test statistic is the same as the DF one.

Example: DF test trending Log Dow Jones

Augmented Dickey-Fuller Unit Root Test on LOGDJIND

Null Hypothesis: LOGDJIND has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 2 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.899565	0.6546
Test critical values:	1% level	-3.959678	
	5% level	-3.410608	
	10% level	-3.127081	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOGDJIND)
Method: Least Squares
Date: 02/25/04 Time: 17:45
Sample(adjusted): 11/25/1896 1/27/2004
Included observations: 5593 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOGDJIND(-1)	-0.001452	0.000764	-1.899565	0.0575
D(LOGDJIND(-1))	0.020812	0.013367	1.556905	0.1195
D(LOGDJIND(-2))	0.044448	0.013367	3.325219	0.0009
C	0.005362	0.002716	1.974263	0.0484
@TREND(11/04/1896)	1.49E-06	7.10E-07	2.092154	0.0365

R-squared	0.003154	Mean dependent var	0.001001
Adjusted R-squared	0.002441	S.D. dependent var	0.026108
S.E. of regression	0.026076	Akaike info criterion	-4.454697
Sum squared resid	3.799650	Schwarz criterion	-4.448771
Log likelihood	12462.56	F-statistic	4.420447
Durbin-Watson stat	2.001370	Prob(F-statistic)	0.001440

(A)DF test nontrending Returns Dow Jones

Augmented Dickey-Fuller Unit Root Test on D(LOGDJIND)

Null Hypothesis: D(LOGDJIND) has a unit root
Exogenous: Constant
Lag Length: 1 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-50.07399	0.0001
Test critical values:	1% level	-3.431339	
	5% level	-2.861862	
	10% level	-2.566984	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOGDJIND,2)
Method: Least Squares
Date: 02/25/04 Time: 17:44
Sample(adjusted): 11/25/1896 1/27/2004
Included observations: 5593 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LOGDJIND(-1))	-0.936123	0.018695	-50.07399	0.0000
D(LOGDJIND(-1),2)	-0.043745	0.013361	-3.274109	0.0011
C	0.000937	0.000349	2.684011	0.0073

R-squared	0.490468	Mean dependent var	4.53E-06
Adjusted R-squared	0.490286	S.D. dependent var	0.036532
S.E. of regression	0.026082	Akaike info criterion	-4.454612
Sum squared resid	3.802691	Schwarz criterion	-4.451057
Log likelihood	12460.32	F-statistic	2690.426
Durbin-Watson stat	2.001286	Prob(F-statistic)	0.000000

6 Impulse response function and persistence measures

The impulse response function (IRF) is a standard tool in illustrating the dynamic behavior of a time series model, e.g. $y_t \sim$ARIMA$(p, 1, q)$.

The IRF measures the effect of an innovation occurring at time t, $y_t - \mathsf{E}[y_t | I_{t-1}]$, on $y_{t+h}, h = 1, \ldots, \infty$.

$$\mathsf{IRF}(h) = \frac{\partial y_{t+h}}{\partial \varepsilon_t}, \quad h = 0, 1, \ldots$$

In a linear time series model the IRF is time invariant and a function of h alone.

If $y_t \sim I(1)$,

$$\Delta y_t = \mu + \psi(L)\varepsilon_t,$$

the IRF is given by the deterministic first order difference equation:

$$\text{IRF}(h) = \text{IRF}(h-1) + \psi_h$$

with starting value IRF(0) $= 1$; thus,

$$\text{IRF}(h) = 1 + \psi_1 + \cdots + \psi_h$$

(cumulation of psi-weights)

Measures of persistence: Let $y_t \sim I(1)$, $\Delta y_t = \mu + \psi(L)\varepsilon_t$.

- Campbell and Mankiw (1987) measure

$$\lim_{h\to\infty} \text{IRF}(h) = \psi(1) = \theta(1)/\phi(1)$$

Equivalently, we can interpret $\psi(1)$ as the revision in the long run prediction of y_t due to the occurrence of a unit shock at t ($\varepsilon_t = 1$):

$$\lim_{h\to\infty} [\text{E}(y_{t+h}|I_t) - \text{E}(y_{t+h}|I_{t-1})] = \psi(1)\varepsilon_t,$$

- Cochrane (1988) normalized variance ratio

$$V = [\psi(1)\sigma]^2/\gamma_0$$

It normalizes the variance of the innovations of the stochastic trend with the unconditional variance of Δy_t.