Predictive Systems

Outline: (PART 1)

Chapter 3 : Linear Stochatsic Models

Linear Stationary Models

- AR
- MA
- ARMA

Linear NonStationary Models

- ARIMA
- SARIMA

Motivation

Stochastic Models assume that:

$$\boldsymbol{z_t = Deterministic\ component + Stochastic\ component}$$

$$cov(\varepsilon_t, \varepsilon_{t+1}) \neq 0 \Longrightarrow \varepsilon_{t+1} = f(\varepsilon_0, .., \varepsilon_t)$$

Such that the observations of the stochastic component are correlated.

Stochastic Process

A stochastic process $Z = (z_t)_{t \in \mathbb{N}}$ is a collection of random variables indexed by a set of time T and valued within a state space S.

In our case, we deal with:

- Discrete time stochastic process.
- Real valued stochastic process.

The observations $(z_1, z_2, \dots, z_n)$ are often called a ***realization*** of a stochastic process Z.

Generally, this stochastic process can be described by n-dimensional probability distribution (joint probability) $P(z_1, z_2, \dots, z_n)$.

Stochastic Process : Moments

The nth moment of Z:

$$E(Z^n) = \begin{cases} \sum_{i=0}^{\infty} z_i^n p(Z = z_i) \\ \int_{-\infty}^{+\infty} z^n f(z) dz \end{cases}$$

The nth central moment of Z: Moment about the mean.

$$E\left((Z-\mu)^n\right) = \begin{cases} \sum_{i=0}^{\infty} (z_i - \mu)^n p(Z = z_i) \\ \int_{-\infty}^{+\infty} (z-\mu)^n f(z) dz \end{cases}$$

Stochastic Process : Moments

In the discrete case:

- The first moment of Z: it is the mean of Z.

$$E(Z) = \sum_{i=0}^{\infty} z_i p(Z = z_i)$$

- The second central moment of Z: It is the variance of Z.

$$V(Z) = E\left((Z - \mu)^2\right) = \sum_{i=0}^{\infty} (z_i - \mu)^2 p(Z = z_i)$$

Stochastic Process : Assumptions

To ensure that the selected moments are able to capture the properties of the stochastic process, some assumptions must be made:

- **The stochastic process is linear.**
- **The stochastic process is ergodic.**
- **The stochastic process if stationary.**

Definition: Ergodicity

A random process is said to be ergodic if all the moments of a realization approach their stochastic process if the size of the realization becomes infinite.

Stochastic Process : Assumptions

Definition: strictly stationary stochastic process (strict sense stationarity)

The stochastic process $Z = (z_t)_{t \in \mathbb{N}}$ is said to be strictly stationary, if :

1) The marginal distributions of all the random variables are identical:
   - The mean is constant over time: $\forall t \in \mathbb{N} \quad E(z_t) = \mu$
   - The variance is constant over time: $\forall t \in \mathbb{N} \quad V(z_t) = \sigma_z^2$
2) The finite-dimensional distributions of any set of variables depends only on the lags between them: the dependence between variables depends only on their lags.

These two conditions can be summarized by:

$$P\left(z_{t_1}, z_{t_2}, \dots, z_{t_m}\right) = P\left(z_{t_{1+k}}, z_{t_{2+k}}, \dots, z_{t_{m+k}}\right)$$

Stochastic Process : Assumptions

Definition: weakly stationary stochastic process (weak sense stationarity)

The stochastic process $Z = (z_t)_{t \in \mathbb{N}}$ is said to be weakly stationary if:

The mean is constant over time: $\forall t \in \mathbb{N}$

$$E(z_1) = E(z_2) = \cdots = E(z_n) = \mu$$

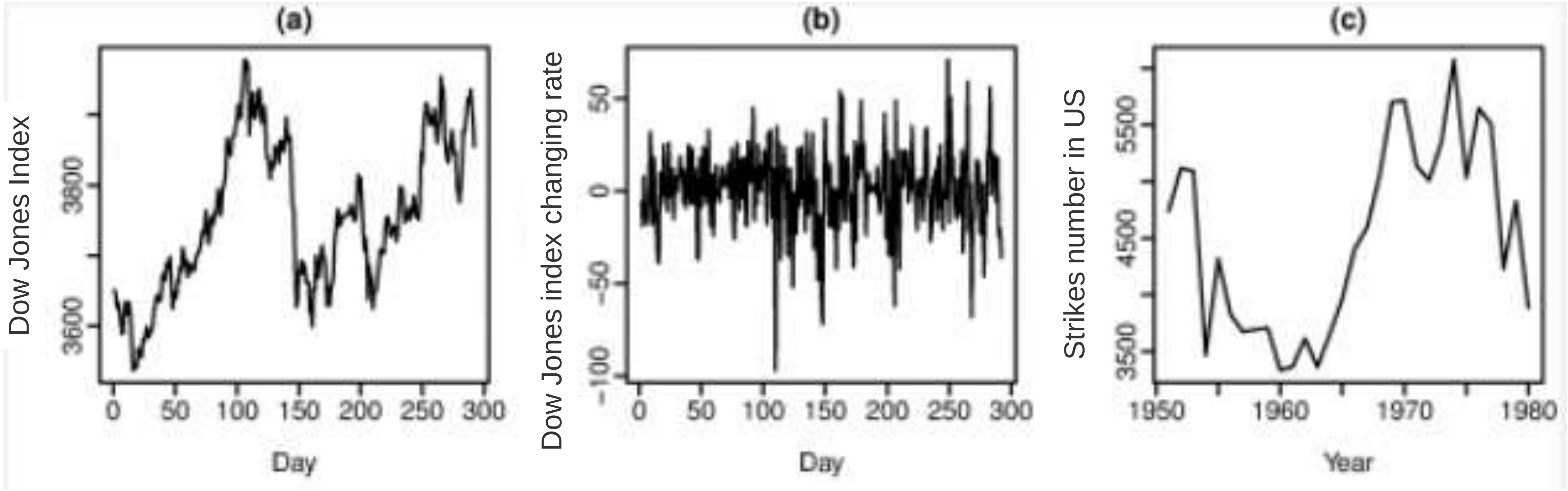
The variance is constant over time: $\forall t \in \mathbb{N}$

$$V(z_1) = V(z_2) = \cdots = V(z_n) = \sigma_z^2$$

The covariance between two variables depends only on their separation (lag/shift) k: $\forall t, k \in \mathbb{N}$

$$Cov(z_t, z_{t-k}) = \gamma_k$$

Stochastic Process : Assumptions

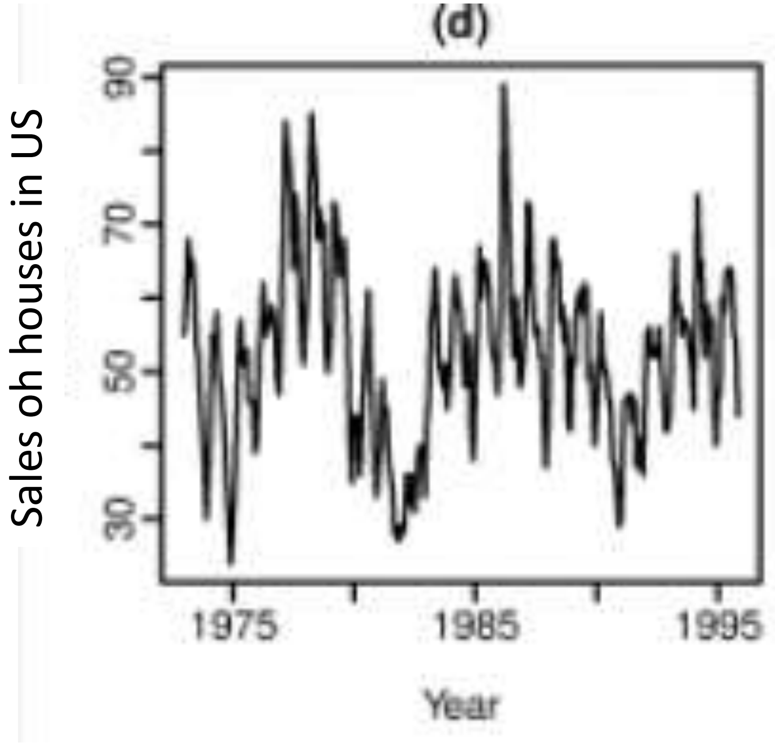


Nonstationary

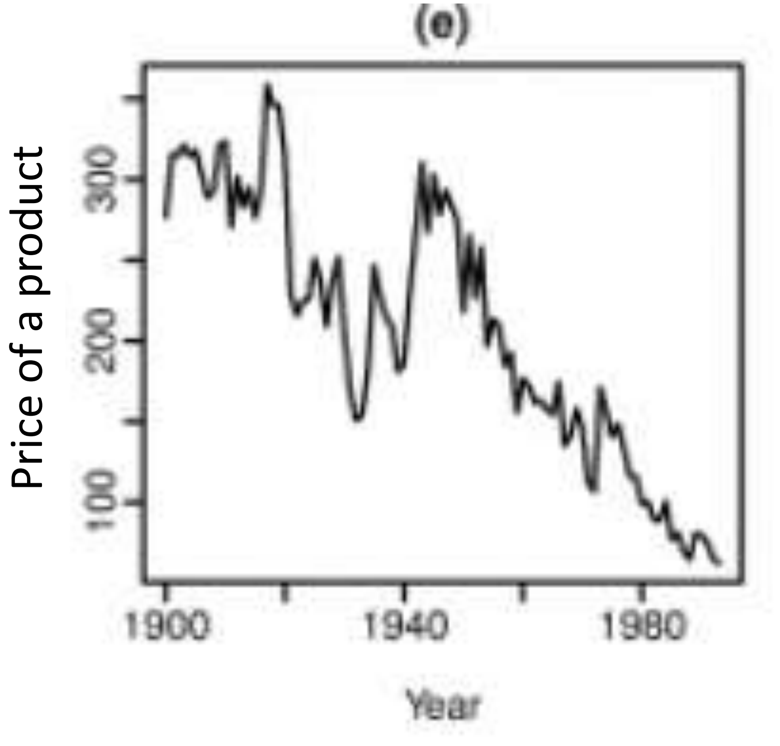
Stationary

Nonstationary

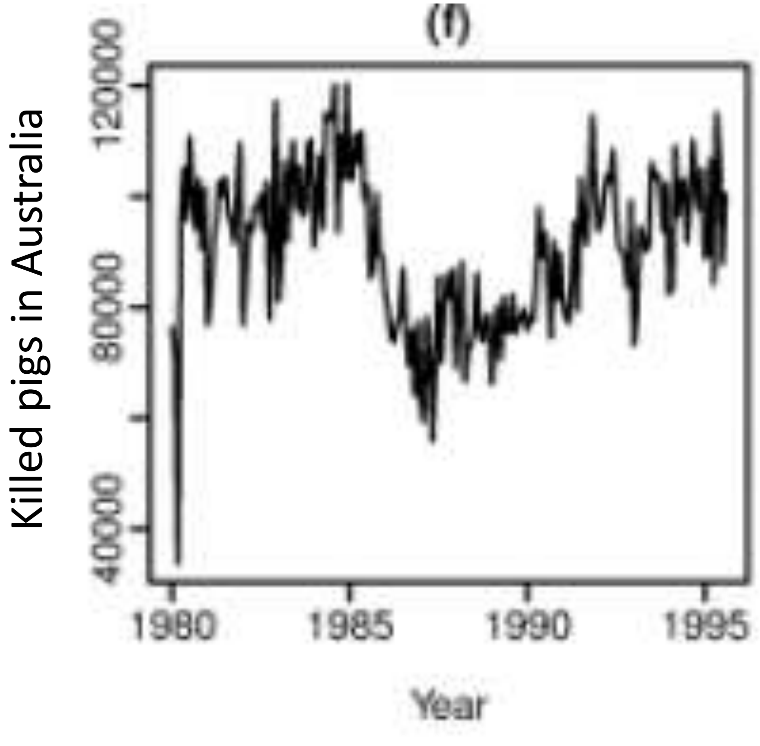
Stochastic Process : Assumptions



Nonstationary

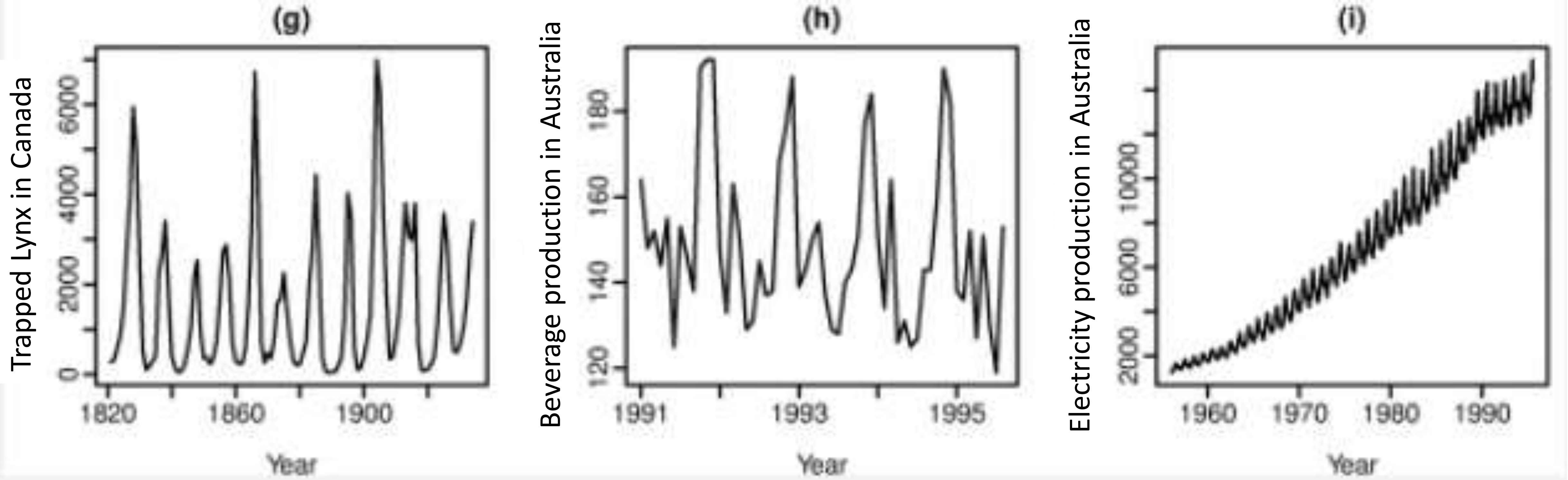


Nonstationary



Nonstationary

Stochastic Process : Assumptions



Nonstationary

Nonstationary

Nonstationary

Réf.

Autocorrelation Function (ACF)

The lag k auto-covariance is defined as: $\forall t \in \{k+1, \dots\}$

$$\gamma_k = Cov(z_t, z_{t-k}) = E((z_t - \mu)(z_{t-k} - \mu))$$

For $k = 0$:

$$\gamma_0 = Cov(z_t, z_t) = E\left((z_t - \mu)^2\right) = V(z_t) = \sigma_z^2$$

The lag k autocorrelation is defined as: $\forall t \in \{k+1, \dots\}$

$$\rho_k = \frac{Cov(z_t, z_{t-k})}{\sqrt{V(z_t)V(z_{t-k})}} = \frac{\gamma_k}{\sigma_z^2} = \frac{\gamma_k}{\gamma_0}$$

Notice:

We have that $\rho_k = \rho_{-k}$ that's why only the positive half of ACF is usually given

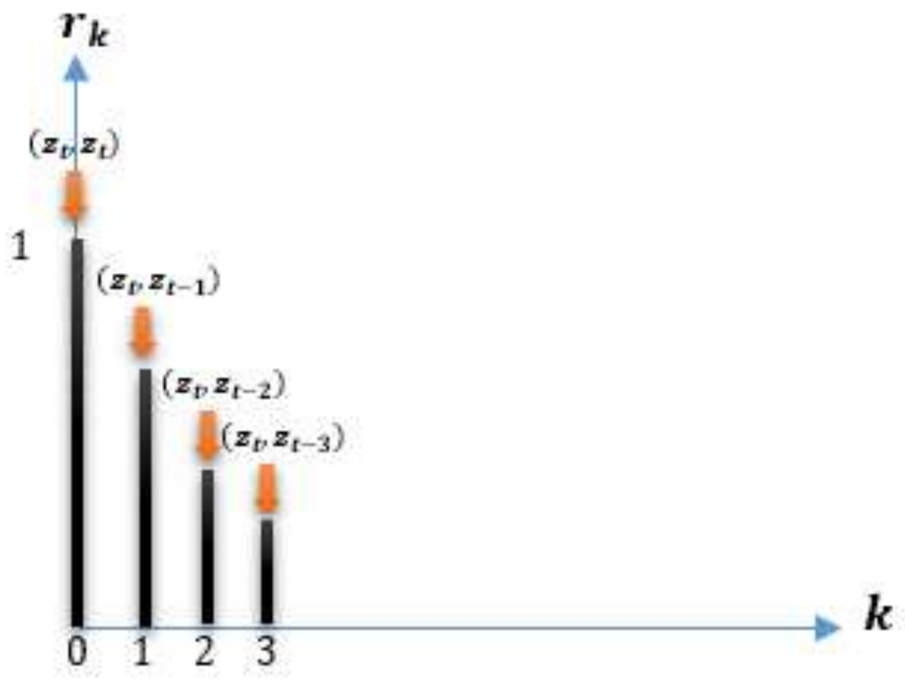
Sample Autocorrelation Function (SACF)

Sample autocorrelation function is an autocorrelation applied to a realization $(z_1, z_2, \dots, z_n)$ of the stochastic process Z, and since it depends on the lag k, we call it autocorrelation function:

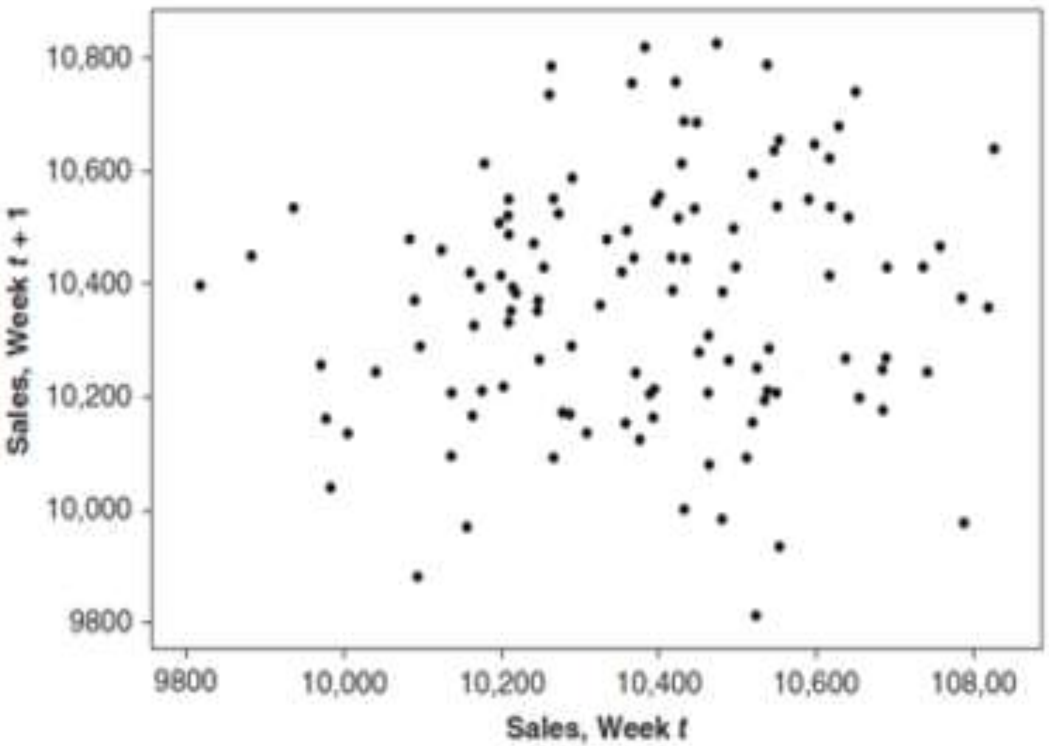
$\forall k \in \{0,1,\dots\}$

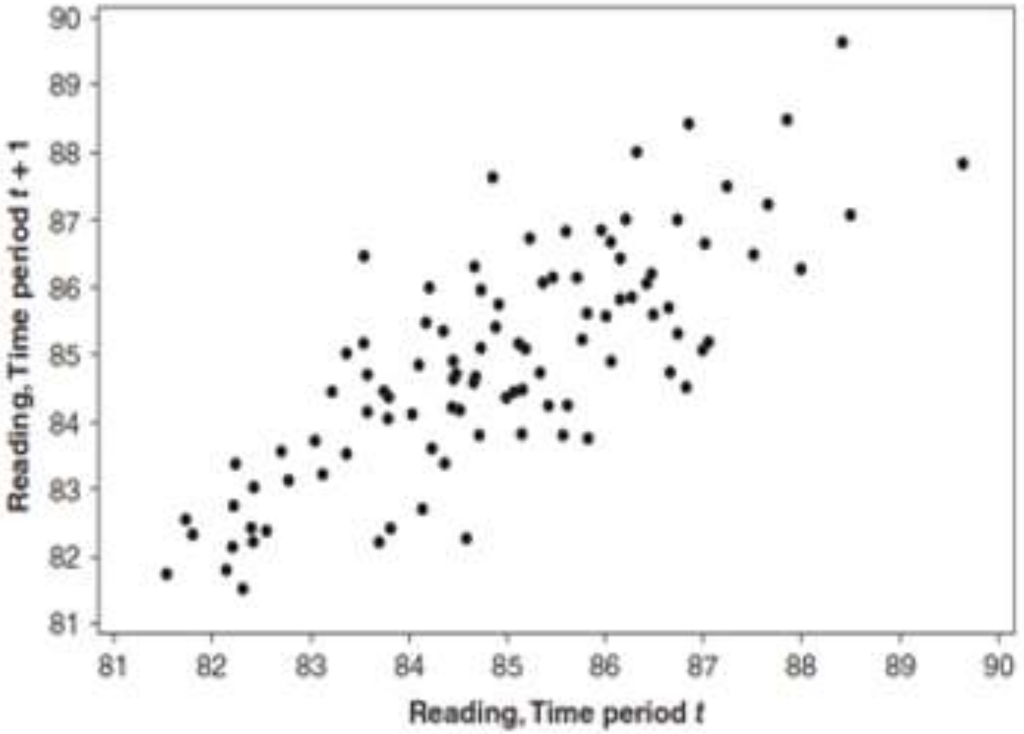
$$r_k = \frac{\sum_{t=k+1}^{n}(z_t - \bar{z})(z_{t-k} - \bar{z})}{\sum_{t=1}^{n}(z_t - \bar{z})^2}$$

Such that $-1 < r_k < 1$.



Sample Autocorrelation Function (SACF)





Differentiation

There exist three main types of differentiation:

- Trend differentiation of order d: it is used to remove the trend.
  - Trend differentiation of order 1.
  - Trend differentiation of order 2.
- Seasonal differentiation of order D: it is used to remove the seasonality.
- Seasonal-Trend differentiation of order (D, d) it is used to remove the seasonality then the trend.

Differentiation : Backshift Notation

One application of the operator B to z_t shifts the data back one observation.

$$Bz_t = z_{t-1}$$

And, one application of the operator B^2 to z_t shifts the data back two observations.

$$B^2 z_t = z_{t-2}$$

In general, the application of the backshift operator of order d to z_t shifts the data back d observations:

$$B^d z_t = z_{t-d}$$

Trend differentiation of order d

The trend differentiation of order d is an operation that allows to remove the trend component :
$\forall t \in \{d+1, \dots, n\}$

$$z_t^{(d)} = z_t - d z_{t-1} - \frac{d(d-1)}{2!} z_{t-2} - \frac{d(d-1)(d-2)}{3!} z_{t-3} - \cdots$$

This operation is represented by $(1-B)^d$ operator, which is computed by the following equation:

$$z_t^{(d)} = (1-B)^d z_t$$

Z	z_1	z_2	$\cdots$	z_d	z_{d+1}	z_{d+2}	$\cdots$	z_n
$Z^{(d)}$					$z_{d+1}^{(d)}$	$z_{d+2}^{(d)}$	$\cdots$	$z_n^{(d)}$

Trend differentiation of order $d = 1$

For $d = 1$, this operation is represented by $(1 - B)$ operator:

$\forall t \in \{2, \dots, n\}$

$$z_t^{(1)} = z_t - z_{t-1} = z_t - Bz_t = (1 - B)z_t$$

This yields another time series named $Z^{(1)} = (z_t^{(1)})_t$ of size $n - 1$, since it is impossible to compute $z_1^{(1)}$ of the first observation.

Z	z_1	z_2	z_3	$\cdots$	z_{n-1}	z_n
$Z^{(1)}$		$z_2^{(1)}$	$z_3^{(1)}$	$\cdots$	$z_{n-1}^{(1)}$	$z_n^{(1)}$

Trend differentiation of order $d = 2$

Consider the trend differentiation of order 1: $\forall t \in \{1, \dots, n\}$

$$z_t^{(1)} = z_t - z_{t-1}$$

So, the second trend differentiation of order 1 on $z_t^{(1)}$ is: $\forall t \in \{2, \dots, n\}$

$$z_t^{(2)} = z_t^{(1)} - z_{t-1}^{(1)} = (z_t - z_{t-1}) - (z_{t-1} - z_{t-2})$$
$$= z_t - 2z_{t-1} + z_{t-2} = z_t - 2Bz_t + B^2 z_t = (1 - 2B + B^2)z_t = (1 - B)^2 z_t$$

This yields another time series named $Z^{(2)} = (z_t^{(2)})_t$ of size $n - 2$, since it is impossible to compute $z_1^{(2)}$ and $z_2^{(2)}$ of the first observation.

Z	z_1	z_2	z_3	z_4	$\cdots$	z_{n-1}	z_n
$Z^{(1)}$		$z_2^{(1)}$	$z_3^{(1)}$	$z_4^{(1)}$	$\cdots$	$z_{n-1}^{(1)}$	$z_n^{(1)}$
$Z^{(2)}$			$z_3^{(2)}$	$z_4^{(2)}$	$\cdots$	$z_{n-1}^{(2)}$	$z_n^{(2)}$

Seasonal differentiation of order D

The seasonal differentiation of order D is the difference between each observation and the same observation from the previous period of length D.
$\forall t \in \{D+1, \ldots, n\}$

$$z_t^{(D)} = z_t - z_{t-D} = z_t - B^D z_t = (1 - B^D) z_t$$

Where D is the length of the seasonal period.

Z	z_1	z_2	$\cdots$	z_D	z_{D+1}	$\cdots$	z_{n-1}	z_n
$Z^{(D)}$					$z_{D+1}^{(D)}$	$\cdots$	$z_{n-1}^{(D)}$	$z_n^{(D)}$

Seasonal-Trend differentiation of order (D, d)

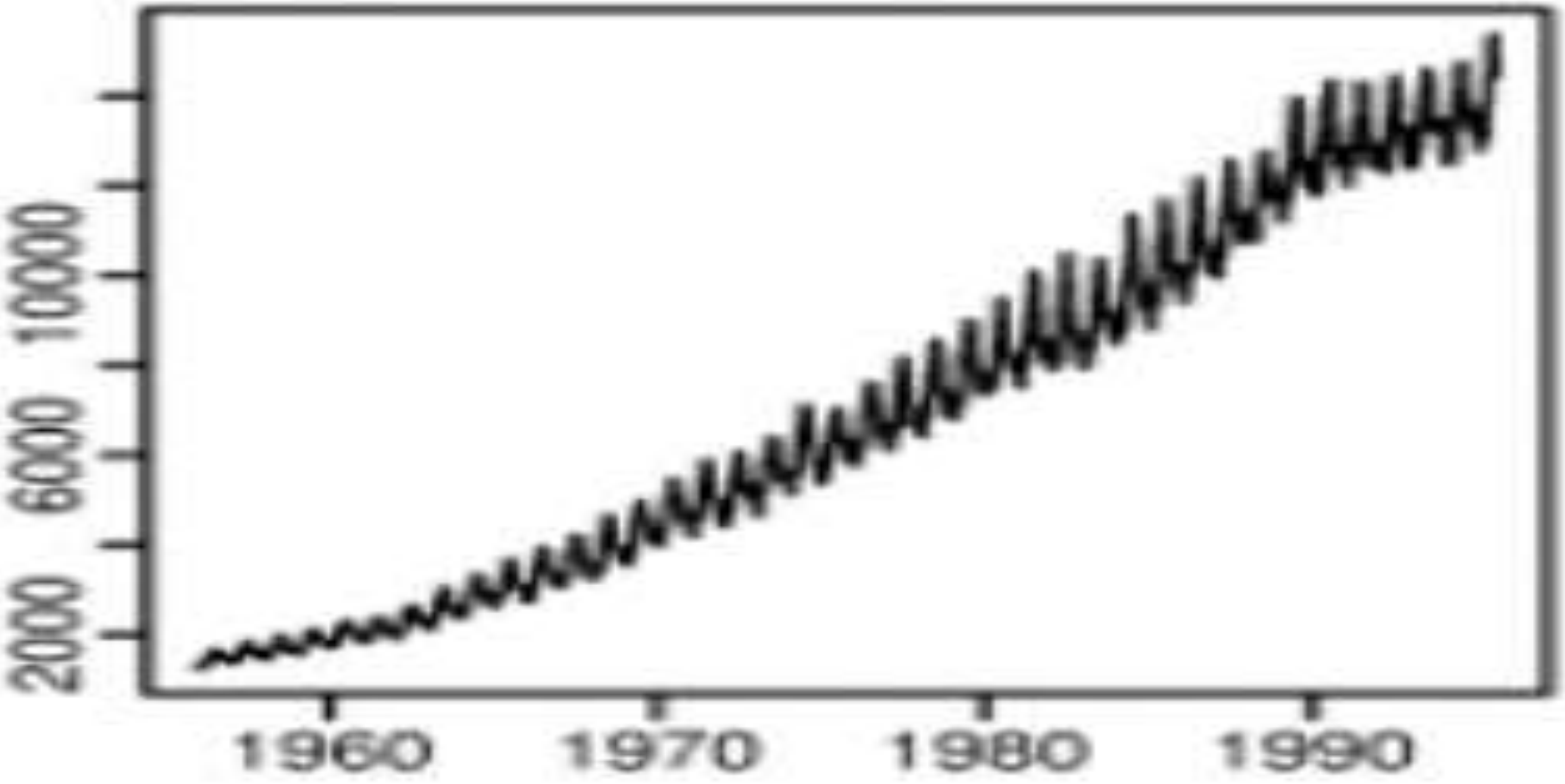
In fact, we can have series with trend and seasonality. In that case, it is suggested to take the seasonal differentiation, first, then the trend differentiation. $\forall t \in \{D+2, \dots, n\}$

$$z_t^{(D)} = z_t - z_{t-D}$$

$$z_t^{(D,1)} = z_t^{(D)} - z_{t-1}^{(D)} = (z_t - z_{t-D}) - (z_{t-1} - z_{t-1-D})$$
$$= z_t - z_{t-1} - z_{t-D} - z_{t-D-1} = (1 - B - B^D - B^{D+1})z_t$$
$$= (1 - B)(1 - B^D)z_t$$

Z	z_1	z_2	$\cdots$	z_D	z_{D+1}	z_{D+2}	$\cdots$	z_{n-1}	z_n
$Z^{(D)}$					$z_{D+1}^{(D)}$	$z_{D+2}^{(D)}$	$\cdots$	$z_{n-1}^{(D)}$	$z_n^{(D)}$
$Z^{(D,1)}$						$z_{D+2}^{(D,1)}$	$\cdots$	$z_{n-1}^{(D,1)}$	$z_n^{(D,1)}$

Distributional Power Transformations



Distributional Power Transformations

Box-Cox transformation (1964): power transformation.

$$f^{BC}(z_t, \lambda) = \begin{cases} \dfrac{z_t^{\lambda} - 1}{\lambda} & if \;\; \lambda \neq 0 \\ \log(z_t) & if \;\; \lambda = 0 \end{cases}$$

- $\lambda = 0.5$ implies square root transformation.
- $\lambda = 0.5$ implies the inverse of square root transformation.
- $\lambda = -1$ implies the inverse transformation.

Distributional Power Transformations

Bickel and Doksum transformation (1981): signed power transformation.

$$f^{SP}(z_t, \lambda) = \frac{sign(z_t).|z_t^{\lambda}|-1}{\lambda} \qquad \text{for } \lambda > 0$$

Burbidge et al. (1988): Inverse hyperbolic transformation.

$$f^{IHS}(z_t, \lambda) = \frac{\sinh^{-1}(\lambda z_t)}{\lambda} \qquad \text{for } \lambda > 0$$

Wold's Decomposition

Theorem:

Every weakly stationary, purely nondeterministic, stochastic process $z_t - \mu$ can be written as a linear combination (or linear filter) of a sequence of uncorrelated random variables:

$$z_t - \mu = \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \cdots = \sum_{i=0}^{\infty} \Psi_i \, \varepsilon_{t-i}$$

Such that $\Psi_0 = 1$ and $(\Psi_1, \Psi_2, \dots)$ are the model's parameters.

Wold's Decomposition

White noise (innovations)

The observations of ε_t are uncorrelated random variables called innovations, drawn from a fixed distribution:

$$E(\varepsilon_t) = 0$$

$$V(\varepsilon_t) = E(\varepsilon_t^2) = \sigma^2 < \infty$$

$$Cov(\varepsilon_t, \varepsilon_{t-k}) = E[\varepsilon_t . \varepsilon_{t-k}] = 0 \text{ for all } k \neq 0$$

Such a sequence is called a white noise process, the innovations are occasionally denoted as $\varepsilon_t \sim WN(0, \sigma^2)$.

Wold's Decomposition

This model leads to compute the autocorrelation of z_t:

Given that: $E\left[\varepsilon_{t-i}.\varepsilon_{t-j}\right] = 0$ for any $i \neq j$.

$$\gamma_0 = V(z_t) = E\left((z_t - \mu)^2\right) = E\left((\varepsilon_t + \Psi_1\varepsilon_{t-1} + \Psi_2\varepsilon_{t-2} + \cdots)^2\right)$$

$$= E(\varepsilon_t^2) + \theta_1^2 E(\varepsilon_{t-1}^2) + \theta_2^2 E(\varepsilon_{t-2}^2) + \cdots$$

$$= \sigma^2 + \Psi_1^2\sigma^2 + \Psi_2^2\sigma^2 + \cdots$$

Then:

$$\boldsymbol{\gamma_0 = \sigma^2 \sum_{i=0}^{\infty} \Psi_i^2}$$

Wold's Decomposition

$$\gamma_k = E((z_t - \mu)(z_{t-k} - \mu))$$

$$= E((\varepsilon_t + \Psi_1\varepsilon_{t-1} + \Psi_2\varepsilon_{t-2} + \cdots + \Psi_k\varepsilon_{t-k} + \cdots)(\varepsilon_{t-k} + \Psi_1\varepsilon_{t-k-1} + \cdots))$$

$$= \sigma^2(1.\Psi_1 + \Psi_1\Psi_{k+1} + \Psi_2\Psi_{k+2} + \cdots)$$

Then:

$$\boldsymbol{\gamma_k = \sigma^2 \sum_{i=0}^{\infty} \Psi_i \Psi_{i+k}}$$

Wold's Decomposition

This implies:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{i=0}^{\infty} \Psi_i \Psi_{i+k}}{\sum_{i=0}^{\infty} \theta_i^2}$$

Notice:

In order to say that the linear filter converges, the number of parameters must be infinite, i.e. the weights must be assumed to be absolutely summable:

$$\sum_{i=0}^{\infty} |\Psi_i| < \infty$$

This condition on the parameters is equivalent to assuming that the process is stationary.

Autoregressive Process of order 1

$$z_t - \mu = \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \cdots = \sum_{i=0}^{\infty} \Psi_i\, \varepsilon_{t-i}$$

Taking $\mu = 0$ and $\Psi_{\mathrm{i}} = \varphi^i$ the linear filter becomes:

$$z_t = \varepsilon_t + \varphi \varepsilon_{t-1} + \varphi^2 \varepsilon_{t-2} + \varphi^3 \varepsilon_{t-3} + \cdots = \sum_{i=0}^{\infty} \varphi^i\, \varepsilon_{t-i}$$

$$= \varepsilon_t + \varphi(\varepsilon_{t-1} + \varphi \varepsilon_{t-2} + \varphi^2 \varepsilon_{t-3} + \cdots)$$

$$= \varphi z_{t-1} + \varepsilon_t$$

This is known as first order autoregressive model and can be computed using the backshift operator as:

$$(1 - \varphi B) z_t = \varepsilon_t$$

Autoregressive Process of order 1

Stationarity Conditions of AR(1):

This model can converge only if $|\varphi| < 1$ which refers to the stationarity condition of this process.

ACF of AR(1) Process

$$z_t - \varphi z_{t-1} = \varepsilon_t$$

Let's multiply the two sides by z_{t-k} such that $k > 0$:

$$z_t z_{t-k} - \varphi z_{t-1} z_{t-k} = \varepsilon_t z_{t-k}$$

Autoregressive Process of order 1

Let's take the expectation:

$$E(z_t z_{t-k}) - \varphi E(z_{t-1} z_{t-k}) = E(\varepsilon_t z_{t-k})$$

$$\gamma_k - \varphi \gamma_{k-1} = E(\varepsilon_t z_{t-k})$$

We have:

$$z_t = \varepsilon_t + \varphi \varepsilon_{t-1} + \varphi^2 \varepsilon_{t-2} + \varphi^3 \varepsilon_{t-3} + \cdots$$

$$z_{t-k} = \varepsilon_{t-k} + \varphi \varepsilon_{t-1-k} + \varphi^2 \varepsilon_{t-2-k} + \cdots$$

Autoregressive Process of order 1

$$\varepsilon_t z_{t-k} = \varepsilon_t \varepsilon_{t-k} + \varphi \varepsilon_t \varepsilon_{t-1-k} + \varphi^2 \varepsilon_t \varepsilon_{t-2-k} + \cdots$$

$$= \sum_{i=0}^{\infty} \varphi^i \varepsilon_t \varepsilon_{t-i-k}$$

$$E(\varepsilon_t z_{t-k}) = E\left(\sum_{i=0}^{\infty} \varphi^i \varepsilon_t \varepsilon_{t-i-k}\right) = 0$$

This is true if $k + i > 0$.

Autoregressive Process of order 1

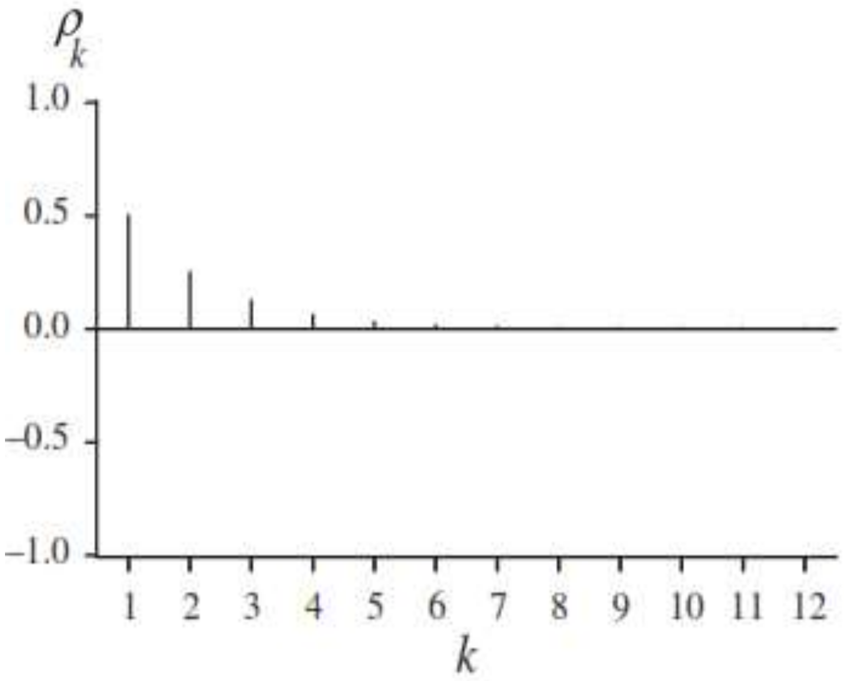
Thus: $$\gamma_k = \varphi\gamma_{k-1} \text{ for all } k > 0$$

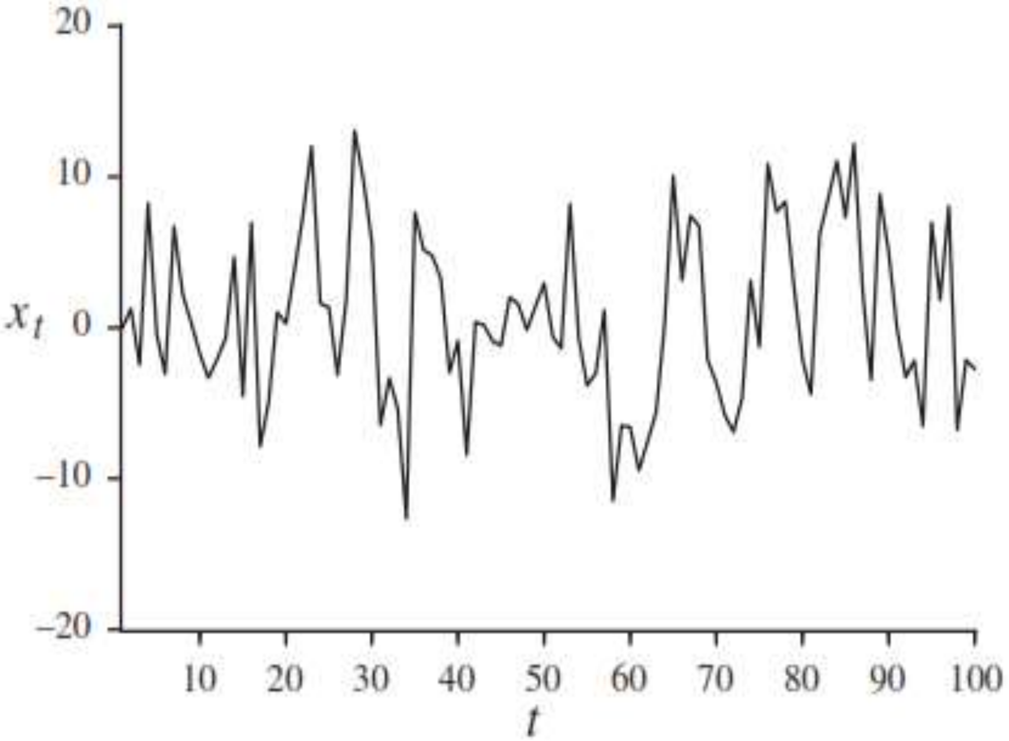
Consequently: $$\gamma_k = \varphi^k\gamma_0$$

$$\boldsymbol{\rho_k = \varphi^k}$$

- If $\varphi > 0$ the AFC will decay exponentially to zero.
- If $\varphi < 0$ the AFC will decay oscillatory to zero.
- While φ is close to ± 1 (nonstationary boundaries) both decays will be slow.

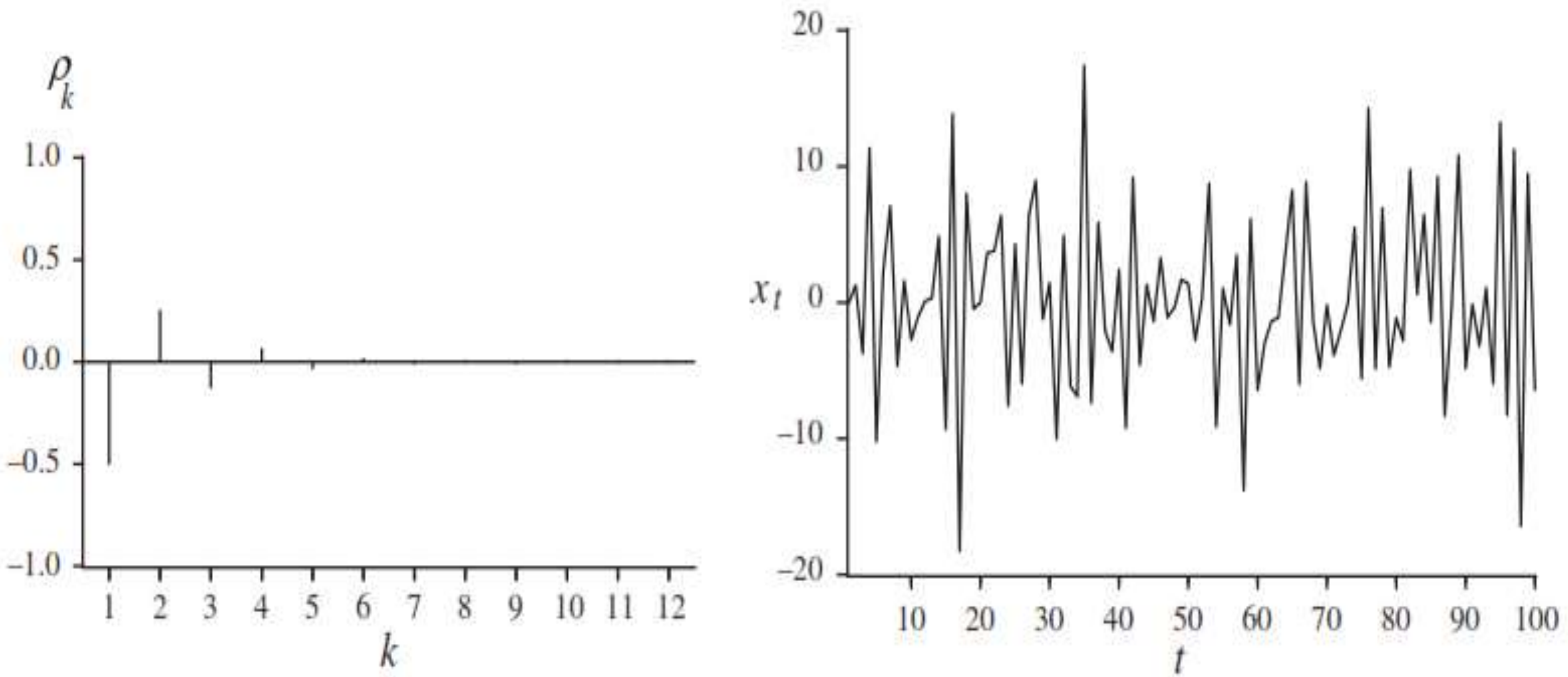
Autoregressive Process of order 1





$AR(1)$ process of $\varphi = 0.5$, $\varepsilon_t \sim NIDW(0,25)$ and $z_1 = 0$

Autoregressive Process of order 1



$AR(1)$ process of $\varphi = -0.5$, $\varepsilon_t \sim NIDW(0{,}25)$ and $z_1 = 0$

Moving Average Process of order 1

$$z_t - \mu = \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \cdots = \sum_{i=0}^{\infty} \Psi_i \, \varepsilon_{t-i}$$

Now let's consider $\mu = 0$ and $\Psi_1 = -\theta_1$ and $\Psi_{\mathrm{i}} = 0$ for any $i \geq 2$. The linear filter becomes:

$$z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Or

$$z_t = (1 - \theta_1 B) \, \varepsilon_t$$

This is called $MA(1)$ process.

Moving Average Process of order 1

ACF of MA(1) Process

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

$$\gamma_1 = -\sigma^2\theta_1$$

$$\gamma_k = 0 \text{ for } k > 1$$

Hence, the ACF is described by:

$$\boldsymbol{\rho_1} = -\frac{\boldsymbol{\theta_1}}{\mathbf{1}+\boldsymbol{\theta_1^2}} \text{ and } \rho_k = 0 \;\text{ for } k > 1$$

Moving Average Process of order 1

The equation of ρ_1 can be written as the quadratic equation:

$$\rho_1\theta_1^2 + \theta_1 + \rho_1 = 0$$

The solution of this equation is:

$$\theta_1 = \frac{-1 \pm \sqrt{1 - 4\rho_1^2}}{2\rho_1}$$

Since θ_1 must be real, this requires that:

$$1 - 4\rho_1^2 > 0$$

Which implies that:

$$-0.5 < \rho_1 < 0.5$$

$$|\rho_1| < 0.5$$

Moving Average Process of order 1

Notice: Stationarity Conditions of MA(1)

Since MA process consists of a finite number of parameters, all MA processes are stationary.

Note that:

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

$$\varepsilon_t = z_t - \theta\varepsilon_{t-1}$$

$$\varepsilon_t = z_t - \theta(z_{t-1} - \theta\varepsilon_{t-2})$$

$$\varepsilon_t = z_t - \theta z_{t-1} + \theta^2\varepsilon_{t-2}$$

$$\varepsilon_t = z_t - \theta z_{t-1} + \theta^2(z_{t-2} - \theta\varepsilon_{t-3})$$

$$\varepsilon_t = z_t - \theta z_{t-1} + \theta^2 z_{t-2} - \theta^3\varepsilon_{t-3}$$

$$\vdots$$

Moving Average Process of order 1

$$\varepsilon_t = \sum_{i=0}^{\infty} (-\theta)^i \, z_{t-i} = AR(\infty)$$

We require $|\theta| < 1$ so that the most recent observations have higher weights than the most distant observations. Hence, the invertibility constraint.

Invertibility Conditions of MA(1):

To insure the converging autoregressive representation, the restriction $|\theta| < 1$ must be imposed.

Autoregressive Process of order p

The general $AR(p)$ model:

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \cdots + \varphi_p z_{t-p} + \varepsilon_t$$

Using the backshift operator, we get:

$$z_t - \varphi_1 B z_t - \varphi_2 B^2 z_t - \cdots - \varphi_p B^p z_t = \varepsilon_t$$

$$(1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p) z_t = \varepsilon_t$$

$$\varphi(B) z_t = \varepsilon_t$$

Such that: $$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$$

Autoregressive Process of order p

So:

$$z_t = \varphi^{-1}(B)\varepsilon_t$$

Let's take:

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

We have:

$$\varphi(B)\psi(B) = 1$$

$$\left(1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p\right).\left(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots\right) = 1$$

Autoregressive Process of order p

$$\psi_0 + (\psi_1 - \varphi_1\psi_0)B + (\psi_2 - \varphi_1\psi_1 - \varphi_2\psi_0)B^2 + \cdots$$
$$+ \left(\psi_j - \varphi_1\psi_{j-1} - \varphi_2\psi_{j-2} - \cdots - \varphi_p\psi_{j-p}\right)B^j + \cdots = 1$$

So:

$$\begin{cases} \psi_0 = 1 \\ \psi_j = 0 \quad for\; j < 0 \\ \psi_j - \varphi_1\psi_{j-1} - \varphi_2\psi_{j-2} - \cdots - \varphi_p\psi_{j-p} = 0 \; for\; all\; j = 1,2 \ldots \end{cases}$$

Its polynomial characteristic equation is: (by substituting $\psi_{j-p} = \frac{1}{m^p}$)

$$m^p - \varphi_1 m^{p-1} - \varphi_2 m^{p-2} - \cdots - \varphi_p = 0$$

Autoregressive Process of order p

Stationarity Conditions of AR(p):

This model can converge only if the roots of the polynomial characteristic equation are such that:

$$|\boldsymbol{m_i}| < \mathbf{1} \text{ for } i = 1, \dots, p$$

We can easily show that:

$$\gamma_k = cov(z_t, z_{t-k})$$

$$= cov(\varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \cdots + \varphi_p z_{t-p} + \varepsilon_t, z_{t-k})$$

$$= \sum_{i=1}^{p} \varphi_i \, cov(z_{t-i}, z_{t-k}) + cov(\varepsilon_t, z_{t-k})$$

Autoregressive Process of order p

$$\gamma_k = \sum_{i=1}^{p} \varphi_i \gamma_{k-i} + \begin{cases} \sigma^2 & if\ k = 0 \\ 0 & if\ k > 0 \end{cases}$$

The following equation is called the **Yule-Walker** equations:

for $k = 1,2 \ldots$

$$\boldsymbol{\gamma_k = \sum_{i=1}^{p} \varphi_i \gamma_{k-i}}$$

So we have:

$$\gamma_0 = \sum_{i=1}^{p} \varphi_i \gamma_i + \sigma^2$$

Autoregressive Process of order p

Thus, it can be noticed that the ACF of the AR(p) model satisfies the Yule-Walker equations:

for $k = 1,2 \ldots$

$$\boldsymbol{\rho_k = \sum_{i=1}^{p} \varphi_i \rho_{k-i}}$$

So, the behavior of the ACF is determined by the p^{th} order difference equations:

for $k = 1,2 \ldots$

$$\boldsymbol{\varphi(B)\rho_k = 0}$$

Autoregressive Process of order p

This implies that the ACF of an AR(p) model can be found through the p roots of the associated polynomial characteristic equation:

$$m^p - \varphi_1 m^{p-1} - \varphi_2 m^{p-2} - \cdots - \varphi_p = 0$$

For example, if the roots are all distinct and real, we have:

for $k = 1,2, \ldots$

$$\rho_k = c_1 m_1^k + c_2 m_2^k + \cdots + c_p m_p^k$$

Such that $c_1, c_2, \ldots, c_p$ are constants.

Notice:

The ACF of an $AR(p)$ model can be described by a **mixture** of **damped exponentials** or **damped oscillation** (for real roots $m_i \in \mathbb{R}$) and **damped sin wave** (sinusoid) expressions (for complex roots $m_i \in \mathbb{C}$)

Autoregressive Process of order $p = 2$

For more simplicity, let's consider the example of AR(2) process:

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \varepsilon_t$$

The backshift equation:

$$(1 - \varphi_1 B - \varphi_2 B^2) z_t = \varepsilon_t$$

$$\varphi(B)\psi(B) = 1$$

$$(1 - \varphi_1 B - \varphi_2 B^2).(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots) = 1$$

$$\psi_0 + (\psi_1 - \varphi_1 \psi_0) B + (\psi_2 - \varphi_1 \psi_1 - \varphi_2 \psi_0) B^2 + \cdots$$
$$+ \left(\psi_j - \varphi_1 \psi_{j-1} - \varphi_2 \psi_{j-2}\right) B^j + \cdots = 1$$

Autoregressive Process of order $p = 2$

$$\begin{cases} \psi_0 = 1 \\ (\psi_1 - \varphi_1\psi_0) = 0 \\ (\psi_j - \varphi_1\psi_{j-1} - \varphi_2\psi_{j-2}) = 0 \; for \; all \; j = 2,3,\ldots \end{cases}$$

Its polynomial characteristic equation is: (by substituting $\psi_{j-p} = \frac{1}{m^p}$)

$$m^2 - \varphi_1 m - \varphi_2 = 0$$

The roots of this equation are:

$$m_1 = \frac{\varphi_1 + \sqrt{\varphi_1^2 + 4\varphi_2}}{2} \quad \text{and} \quad m_2 = \frac{\varphi_1 - \sqrt{\varphi_1^2 + 4\varphi_2}}{2}$$

Autoregressive Process of order $p = 2$

Stationarity conditions of AR(2) model:

So that AR(2) be stationary the roots should satisfy:

$$|m_1| < 1 \text{ and } |m_2| < 1$$

And it can be shown that these conditions imply this set of restrictions:

If the roots are real:

$$\begin{cases} \varphi_1 + \varphi_2 < 1 \\ \varphi_2 - \varphi_1 < 1 \\ -1 < \varphi_2 < 1 \end{cases}$$

If the roots are complex:

$$\varphi_2 < 0$$

Autoregressive Process of order $p = 2$

Here, we have:

$$\gamma_k = \varphi_1 \gamma_{k-1} + \varphi_2 \gamma_{k-2} + \begin{cases} \sigma^2 \ \ if\ k = 0 \\ 0 \ \ if\ k > 0 \end{cases}$$

For $k = 1{,}2 \ldots$ the Yule-Walker equations are:

$$\gamma_k = \varphi_1 \gamma_{k-1} + \varphi_2 \gamma_{k-2}$$

Similarly:

for $k = 1{,}2 \ldots$

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2}$$

Autoregressive Process of order $p = 2$

The Yule-Walker equations can be solved recursively as:

$$\rho_0 = 1$$

$$\rho_1 = \varphi_1 \rho_0 + \varphi_2 \rho_{-1} = \varphi_1 + \varphi_2 \rho_1$$

$$\rho_1 = \frac{\varphi_1}{1 - \varphi_2}$$

$$\rho_2 = \varphi_1 \rho_1 + \varphi_2$$

$$\rho_3 = \varphi_1 \rho_2 + \varphi_2 \rho_1$$

$$\vdots$$

Autoregressive Process of order $p = 2$

A general solution of ρ_k can be obtained through the roots m_1 and m_2 of the associated polynomial characteristic equation:

$$m^2 - \varphi_1 m - \varphi_2 = 0$$

There are three cases:

- Existence of two real roots: $m_1, m_2 \in \mathbb{R}$
- Existence of two complex conjugates roots: $m_1, m_2 \in \mathbb{C}$
- Existence of one real root: $m_0 \in \mathbb{R}$

Autoregressive Process of order $p = 2$

Case 1:

If m_1 and m_2 are distinct real roots, we have:

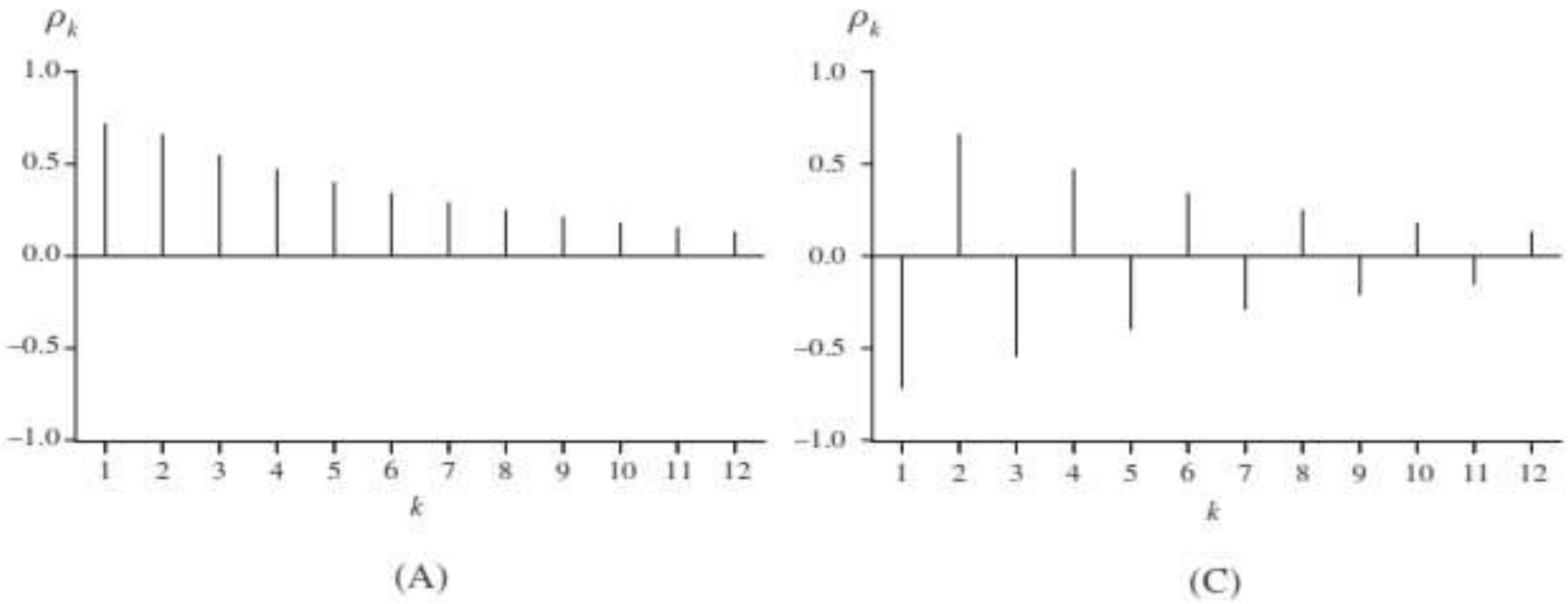
$$\rho_k = c_1 m_1^k + c_2 m_2^k \quad \text{for } k = 0,1,2, \ldots$$

Where c_1 and c_2 are constants and can be obtained, for example, from ρ_0 and ρ_1.

Notice:

Since, for stationarity we have $|m_1| < 1$ and $|m_2| < 1$, in this case the ACF is a **mixture of two damped exponentials (A)** terms or **damped oscillation (C)**, depending on the signs of the roots.

Autoregressive Process of order $p = 2$



- In **A** we have: $\varphi_1 = 0.5$ and $\varphi_2 = 0.3$.
- In **C** we have: $\varphi_1 = -0.5$ and $\varphi_2 = 0.3$.

Autoregressive Process of order $p = 2$

Case 2:

If m_1 and m_2 are complex conjugates roots in the form $a \pm ib$, we have:

for $k = 0,1,2, \ldots$

$$\rho_k = R^k(c_1 \cos(\lambda k) + i\, c_2 \sin(\lambda k))$$

Where:

$$R = |m_i| = \sqrt{a^2 + b^2}$$

And λ is determined by:

$$\cos(\lambda) = \frac{a}{R} \text{ and } \sin(\lambda) = \frac{b}{R}$$

Autoregressive Process of order $p = 2$

Hence:

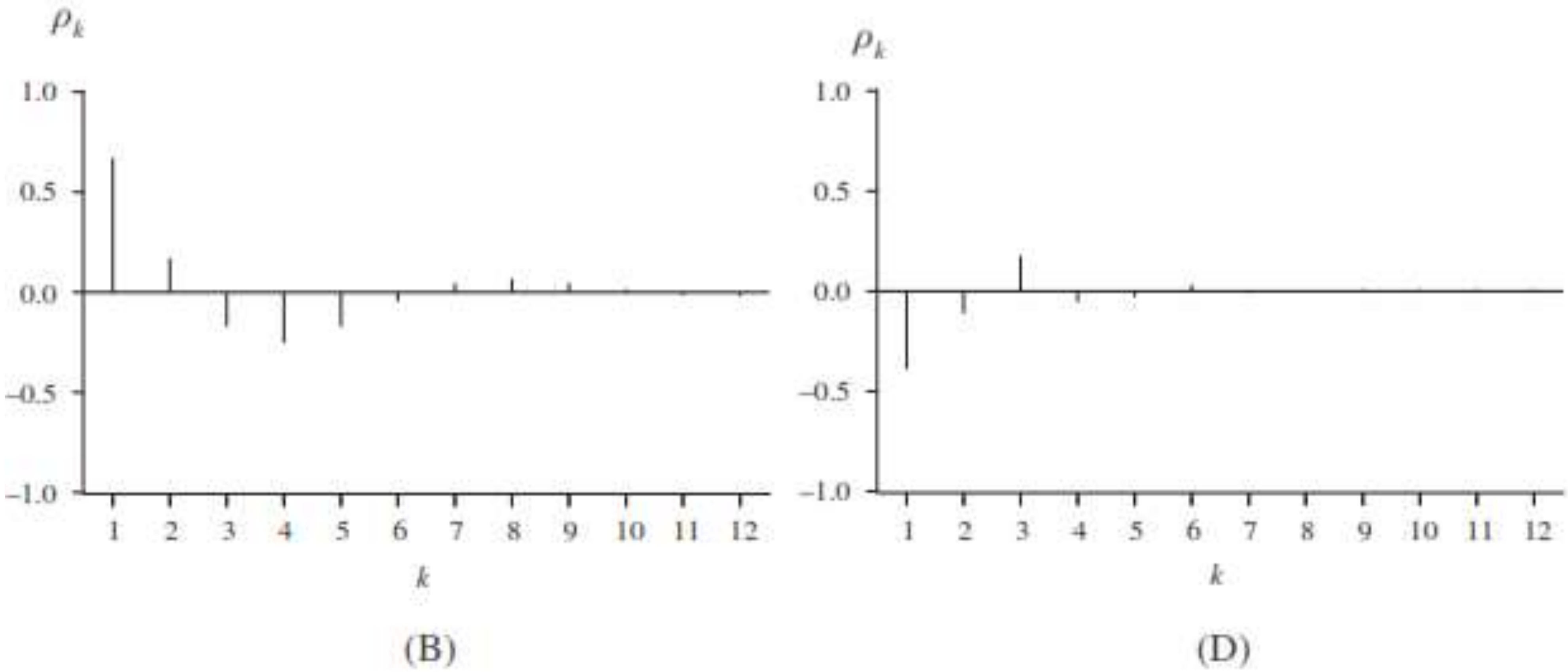
$$m_1 = a + ib = R(\cos(\lambda) + i\sin(\lambda))$$
$$m_2 = a - ib = R(\cos(\lambda) - i\sin(\lambda))$$

Again, c_1 and c_2 are constants.

Notice:

In this case, the ACF has the form of a **damped sine wave (B-D)**, with damping factor R and frequency λ (the period is $2\pi/\lambda$).

Autoregressive Process of order $p = 2$



- In **B** we have: $\varphi_1 = 1$ and $\varphi_2 = -0.5$.
- In **D** we have: $\varphi_1 = -0.5$ and $\varphi_2 = -0.3$.

Autoregressive Process of order $p = 2$

Case 3:

If there is one real root m_0, we have:

for $k = 0,1,2,\ldots$

$$\rho_k = (c_1 + c_2 k) m_0^k$$

Notice:

In this case, the ACF will exhibit an exponential decay pattern.

Autoregressive Process of order $p = 2$

Exercise:

Consider the realization of the two different AR(2) processes:

$$z_t = 4 + 0.4z_{t-1} + 0.5z_{t-2} + \varepsilon_t$$

$$z_t = 4 + 0.8z_{t-1} - 0.5z_{t-2} + \varepsilon_t$$

1. Give the characteristic equation for each process.
2. Compute the roots for each process.
3. Compute the ACF for each process.
4. What is the behavior of the ACF for each process?

Partial Autocorrelation Function

Notice:

Since, all AR processes have ACFs that damp out, it is sometimes difficult to distinguish between processes of different orders. To aid with such discrimination, the partial ACF (PACF) may be used.

The correlation between two random variables is often due to both variables being correlated with a third.

This internal correlation can be viewed by the expression (Yule-Walker) of the ACF of the AR process:

$$\rho_k = \sum_{i=1}^{p} \varphi_i \rho_{k-i} \text{ for } k = 1,2 \ldots$$

Partial Autocorrelation Function

Definition: PACF

The $\boldsymbol{k^{th}}$ partial autocorrelation functions is the last coefficient $\boldsymbol{\varphi_{kk}}$ of the AR(k) process:

$$z_t = \varphi_{k1} z_{t-1} + \varphi_{k2} z_{t-2} + \cdots + \varphi_{kk} z_{t-k} + \varepsilon_t$$

It measures the additional correlation between z_t and z_{t-k} after adjustments have been made for the intervening lags.

Partial Autocorrelation Function

In general, the $\boldsymbol{\varphi_{kk}}$ can be obtained from the Yule-Walker equations that correspond to AR(k) process. For AR(k) process, the Yule-Walker equation can be expressed as: (with $p = k$ and $\varphi_i = \varphi_{ii}$)

for $j = 1,2, \ldots$

$$\rho_j = \sum_{i=1}^{k} \varphi_{ki} \rho_{j-i}$$

So:

$$\begin{gathered}
\rho_1 = \varphi_{k1} + \varphi_{k2}\rho_1 + \cdots + \varphi_{kk}\rho_{k-1} \\
\rho_2 = \varphi_{k1}\rho_1 + \varphi_{k2} + \cdots + \varphi_{kk}\rho_{k-2} \\
\vdots \\
\rho_k = \varphi_{k1}\rho_{k-1} + \varphi_{k2}\rho_{k-2} + \cdots + \varphi_{kk}
\end{gathered}$$

Partial Autocorrelation Function

Using matrix expression:

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \varphi_{k1} \\ \varphi_{k2} \\ \vdots \\ \varphi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}$$

So:

$$\varphi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_1 & 1 \end{vmatrix}}$$

Partial Autocorrelation Function

Example:

For $k = 1$:

$$\varphi_{11} = \rho_1 = \varphi$$

For $k = 2$:

$$\varphi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

Partial Autocorrelation Function

It follows from the definition of φ_{kk} that the PACFs of the AR processes follow the patterns:

AR(1): $\varphi_{11} = \rho_1 = \varphi$ and $\boldsymbol{\varphi_{kk}} = \mathbf{0}$ for $k > 1$

AR(2): $\varphi_{11} = \rho_1, \ \varphi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$ and $\boldsymbol{\varphi_{kk}} = \mathbf{0}$ for $k > 2$

$$\vdots$$

AR(p): $\varphi_{11} \neq 0, \ \varphi_{22} \neq 0, \dots, \varphi_{pp} \neq 0$ and $\boldsymbol{\varphi_{kk}} = \mathbf{0}$ for $k > p$

Hence, the partial autocorrelations for lags larger than the order of the process are zero.

Partial Autocorrelation Function

Order determination of AR(p) process

Consequently, an AR(p) process is described by:

- The ACF is infinite in extent and is dominated by a **mixture** of **damped exponentials** or **damped oscillation** (for real roots $m_i \in \mathbb{R}$) and/or **damped sin wave** (sinusoid) expressions (for complex roots $m_i \in \mathbb{C}$).
- The PACF are zero for lags larger than p.

Moving Average Process of order q

In general MA of order q is:

$$z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

Using the backshift operator, we get:

$$z_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)\varepsilon_t$$

$$z_t = \theta(B)\varepsilon_t$$

Such that: $$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

Moving Average Process of order q

So:

$$\theta^{-1}(B)z_t = \varepsilon_t$$

Let's take:

$$\theta^{-1}(B) = \pi(B)$$

We have:

$$\theta(B)\pi(B) = 1$$

$$\left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right).\left(\pi_0 + \pi_1 B + \pi_2 B^2 + \cdots \pi_q B^q\right) = 1$$

$$\pi_0 + (\pi_1 - \theta_1\pi_0)B + (\pi_2 - \theta_1\pi_1 - \theta_2\pi_0)B^2 + \cdots$$
$$+ \left(\pi_j - \theta_1\pi_{j-1} - \theta_2\pi_{j-2} - \cdots - \theta_q\pi_{j-q}\right)B^j + \cdots = 1$$

Moving Average Process of order q

So:

$$\begin{cases} \pi_0 = 1 \\ \pi_j = 0 \quad for\ j < 0 \\ \pi_j - \theta_1 \pi_{j-1} - \theta_2 \pi_{j-2} - \cdots - \theta_p \pi_{j-q} = 0\ for\ all\ j = 1,2 \ldots, q \end{cases}$$

Its polynomial characteristic equation is: (by substituting $\pi_{j-q} = \frac{1}{m^q}$)

$$\boldsymbol{m^q} - \theta_1 \boldsymbol{m^{q-1}} - \theta_2 \boldsymbol{m^{q-2}} - \cdots - \theta_q = \mathbf{0}$$

Invertibility Conditions of MA(q):

This model can converge only if the roots of the polynomial characteristic equation are such that:

$$|\boldsymbol{m_i}| < \mathbf{1} \text{ for } i = 1, \ldots, q$$

Moving Average Process of order q

The ACF of $\boldsymbol{MA(q)}$:

$$\gamma_k = cov(z_t, z_{t-k}) = E(Z_t . Z_{t-k})$$

For $k \leq q$

$$\gamma_k = \left(-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_q\theta_{q-k}\right)\sigma^2$$

$$\gamma_0 = cov(z_t, z_t) = E(Z_t^2) = \left(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2\right)\sigma^2$$

So:

$$\begin{cases} \rho_k = \dfrac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_q\theta_{q-k}}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2} \quad for\ k \leq q \\ \rho_k = 0 \quad for\ k > q \end{cases}$$

Moving Average Process of order q

Order determination of MA(q) process

Consequently, for MA(q) process:

- The PACF is infinite in extent and is dominated by a **mixture** of **damped exponentials** or **damped oscillation** (for real roots $m_i \in \mathbb{R}$) and/or **damped sin wave** (sinusoid) expressions (for complex roots $m_i \in \mathbb{C}$).

- The ACF are zero for lags larger than q.

Notice:

- The PACF pattern of MA(q) is similar to the ACF pattern of AR(p).
- The ACF pattern of MA(q) is similar to the PACF pattern of AR(p).