Time Series

Course Lecturer: Loukia Meligkotsidou
Department of Mathematics, University of Athens

MSc in Statistics and Operations Research

Stationary ARMA Models and Box-Jenkins methodology

- Introduction
- The time series models:
  - Moving Average models [MA]
  - Autoregressive models [AR]
  - Autoregressive and Moving Average models [ARMA]
- Properties of time series models
- Estimation of time series models
- Forecasting time series models

Introduction

- Explain the movements of the time series by its own characteristics.
- Emphasis is on the analysis of probabilistic or stochastic properties of the time series.
- Construct models that explain the series y_t by using past values y_{t-i} and past stochastic error terms ε_{t-j}...
- ...instead of (or apart from) explaining y_t by using explanatory variables x_1, x_2, ..., x_k.

Introduction: Time Series models

Use lagged data points or/and lagged errors

Autoregressive models [AR(p)]

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

Moving Average models [MA(q)]

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

Autoregressive Moving Average models [ARMA(p,q)]

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

Introduction: Aim of the analysis

- Try to study the characteristics of the time series (identification step)...
- in order to build - construct and estimate an appropriate model (estimation step)...
- which can be used to explain what has generated the observed time-series data (diagnostic checking step)...
- and can be also used for predictions (forecasting step).

Introduction: Box-Jenkins methodology

- Identification step
  Use autocorrelation and partial autocorrelation of the series to find appropriate values of p and q.
- Estimation step
  Estimate the model's parameters using Maximum Likelihood or Least Squares method.
- Diagnostic checking step
  Examine if the chosen (estimated) model fits the data reasonably well - test if the residuals of the estimated model are uncorrelated, homoscedastic and normal, i.e. white noise.
- Forecasting step
  Compute forecasts based on the fitted model.

White Noise

The basic building block for all processes and models considered in time series analysis is the White Noise process

A process ε_t that satisfies the following conditions is called a White Noise process:

- $E(\varepsilon_t) = 0$, zero mean.
- $V(\varepsilon_t) = E[\varepsilon_t - E(\varepsilon_t)]^2 = E(\varepsilon_t^2) - [E(\varepsilon_t)]^2 = E(\varepsilon_t^2) = \sigma^2$, constant variance.
- $\gamma_{t-s} = Cov(\varepsilon_t, \varepsilon_s) = E[(\varepsilon_t - E(\varepsilon_t))(\varepsilon_s - E(\varepsilon_s))] = E(\varepsilon_t \varepsilon_s) = 0$ for $t \neq s$, uncorrelated elements across time.

Furthermore, if the ε_t follow the Normal distribution, i.e. $\varepsilon_t \sim N(0, \sigma^2)$ the process ε_t is called Gaussian White Noise process.

Example 1: $y_t = \mu + \varepsilon_t$

Consider a process y_t which is the sum of a constant μ plus a Gaussian White Noise process ε_t i.e. $y_t = \mu + \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma^2)$

$$E(y_t) = E(\mu + \varepsilon_t) = E(\mu) + E(\varepsilon_t) = \mu$$

$$\gamma_0 = V(y_t) = V(\mu + \varepsilon_t) = V(\mu) + V(\varepsilon_t) = \sigma^2 \text{ or}$$

$$\gamma_0 = V(y_t) = E[y_t - E(y_t)]^2 = E[y_t - \mu]^2 = E(\varepsilon_t)^2 = \sigma^2$$

$$\gamma_j = Cov(y_t, y_{t-j}) = E[(y_t - E(y_t))(y_{t-j} - E(y_{t-j}))] =$$

$$= E[(y_t - \mu)(y_{t-j} - \mu)] = E(\varepsilon_t \varepsilon_{t-j}) = 0$$

y_t moves around μ with variance σ^2 (uncorrelated $y_t's$)

Example 2: $y_t = \beta t + \varepsilon_t$

Consider a process y_t which is a time trend, βt, plus a Gaussian White Noise process ε_t i.e. $y_t = \beta t + \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma^2)$

$$E(y_t) = E(\beta t + \varepsilon_t) = E(\beta t) + E(\varepsilon_t) = \beta t$$

$$\gamma_0 = V(y_t) = V(\beta t + \varepsilon_t) = V(\beta t) + V(\varepsilon_t) = \sigma^2$$ or

$$\gamma_0 = V(y_t) = E[y_t - E(y_t)]^2 = E[y_t - \beta t]^2 = E(\varepsilon_t)^2 = \sigma^2$$

$$\gamma_j = Cov(y_t, y_{t-j}) = E[(y_t - E(y_t))(y_{t-j} - E(y_{t-j}))] =$$

$$= E[(y_t - \beta t)(y_{t-j} - \beta(t-j))] = E(\varepsilon_t \varepsilon_{t-j}) = 0$$

y_t move around trend βt with variance σ^2 (uncorrelated $y_t's$)

Moving Average process: MA(1) - Mean - Variance

Let y_t follow a Moving Average of order one, MA(1), model:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t \ , \ \varepsilon_t \sim N(0, \sigma^2)$$

$$E(y_t) = E(\mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t) = E(\mu) + E(\theta_1 \varepsilon_{t-1}) + E(\varepsilon_t)$$
$$= E(\mu) + \theta_1 E(\varepsilon_{t-1}) + E(\varepsilon_t) = \mu$$

$$V(y_t) = E[y_t - E(y_t)]^2 = E[y_t - \mu]^2 = E(\theta_1 \varepsilon_{t-1} + \varepsilon_t)^2 =$$
$$= E(\theta_1^2 \varepsilon_{t-1}^2 + \varepsilon_t^2 + 2\theta_1 \varepsilon_{t-1} \varepsilon_t) =$$
$$= E(\theta_1^2 \varepsilon_{t-1}^2) + E(\varepsilon_t^2) + E(2\theta_1 \varepsilon_{t-1} \varepsilon_t) =$$
$$= \theta_1^2 E(\varepsilon_{t-1}^2) + E(\varepsilon_t^2) + 2\theta_1 E(\varepsilon_{t-1} \varepsilon_t) =$$
$$= \theta_1^2 \sigma^2 + \sigma^2 = (1 + \theta_1^2)\sigma^2$$

Moving Average process: MA(1)
Autocovariance at lag 1, γ_1 - Autocorrelation at lag 1, ρ_1

$$\gamma_1 = Cov(y_t, y_{t-1}) = E[(y_t - E(y_t))(y_{t-1} - E(y_{t-1}))] =$$

$$= E[(y_t - \mu)(y_{t-1} - \mu)] = E[(\theta_1 \varepsilon_{t-1} + \varepsilon_t)(\theta_1 \varepsilon_{t-2} + \varepsilon_{t-1})] =$$

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

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

$$= \theta_1 \sigma^2$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta_1 \sigma^2}{(1+\theta_1^2)\sigma^2} = \frac{\theta_1}{1+\theta_1^2} \neq 0$$

Moving Average process: MA(1)
Autocovariance at lag 2, γ_2 - Autocorrelation at lag 2, ρ_2

$$\gamma_2 = Cov(y_t, y_{t-2}) = E[(y_t - E(y_t))(y_{t-2} - E(y_{t-2}))] =$$

$$= E[(y_t - \mu)(y_{t-2} - \mu)] = E[(\theta_1 \varepsilon_{t-1} + \varepsilon_t)(\theta_1 \varepsilon_{t-3} + \varepsilon_{t-2})] =$$

$$= E(\theta_1^2 \varepsilon_{t-1}\varepsilon_{t-3} + \theta_1 \varepsilon_{t-1}\varepsilon_{t-2} + \theta_1 \varepsilon_t \varepsilon_{t-3} + \varepsilon_t \varepsilon_{t-2}) =$$

$$= \theta_1^2 E(\varepsilon_{t-1}\varepsilon_{t-3}) + \theta_1 E(\varepsilon_{t-1}\varepsilon_{t-2}) + \theta_1 E(\varepsilon_t \varepsilon_{t-3}) + E(\varepsilon_t \varepsilon_{t-2}) = 0$$

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{0}{(1+\theta_1^2)\sigma^2} = 0$$

Moving Average process: MA(1)
Autocovariance at lag $k \geq 2$ - Autocorrelation at lag $k \geq$

$$\gamma_k = Cov(y_t, y_{t-k}) = E[(y_t - E(y_t))(y_{t-k} - E(y_{t-k}))] =$$

$$= E[(y_t - \mu)(y_{t-k} - \mu)] = E[(\theta_1 \varepsilon_{t-1} + \varepsilon_t)(\theta_1 \varepsilon_{t-k-1} + \varepsilon_{t-k})] =$$

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

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

$$= 0$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{0}{(1+\theta_1^2)\sigma^2} = 0$$

Moving Average process: MA(1) Useful comments

- The mean, the variance and the autocovariances-autocorrelations are constant over time.
- The MA(1) is weakly stationary process, without imposing restrictions on the model parameters.
- In the MA(1) process the autocovariance-autocorrelation at lag one is different from zero, and all the other autocorrelations are zero.

Moving Average process: MA(2) - Mean - Variance

Let y_t follow a Moving Average of order two, MA(2), model:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t \;,\; \varepsilon_t \sim N(0, \sigma^2)$$

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

$$V(y_t) = E[y_t - E(y_t)]^2 = E[y_t - \mu]^2 =$$
$$E(\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t)^2 = \ldots = (1 + \theta_1^2 + \theta_2^2)\sigma^2$$

Moving Average process: MA(2)
Autocovariance at lag 1, γ_1 - Autocorrelation at lag 1, ρ_1

$$\gamma_1 = Cov(y_t, y_{t-1}) = E[(y_t - E(y_t))(y_{t-1} - E(y_{t-1}))] =$$

$$= E[(y_t - \mu)(y_{t-1} - \mu)] =$$

$$= E[(\theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \varepsilon_t)(\theta_1\varepsilon_{t-2} + \theta_2\varepsilon_{t-3} + \varepsilon_{t-1})] =$$

$$= \theta_1\sigma^2 + \theta_1\theta_2\sigma^2 = (\theta_1 + \theta_1\theta_2)\sigma^2$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(\theta_1+\theta_1\theta_2)\sigma^2}{(1+\theta_1^2+\theta_2^2)\sigma^2} = \frac{\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2} \neq 0$$

Moving Average process: MA(2)
Autocovariance at lag 2, γ_2 - Autocorrelation at lag 2, ρ_2

$$\gamma_2 = Cov(y_t, y_{t-2}) = E[(y_t - E(y_t))(y_{t-2} - E(y_{t-2}))] =$$

$$= E[(y_t - \mu)(y_{t-2} - \mu)] =$$

$$= E[(\theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \varepsilon_t)(\theta_1\varepsilon_{t-3} + \theta_2\varepsilon_{t-4} + \varepsilon_{t-2})] =$$

$$= \theta_2\sigma^2$$

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\theta_2\sigma^2}{(1+\theta_1^2+\theta_2^2)\sigma^2} = \frac{\theta_2}{1+\theta_1^2+\theta_2^2} \neq 0$$

Moving Average process: MA(2)
Autocovariance at lag k, γ_k - Autocorrelation at lag k, ρ_k

$$\gamma_k = Cov(y_t, y_{t-k}) = E[(y_t - E(y_t))(y_{t-k} - E(y_{t-k}))] =$$

$$= E[(y_t - \mu)(y_{t-k} - \mu)] =$$

$$= E[(\theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \varepsilon_t)(\theta_1\varepsilon_{t-k-1} + \theta_2\varepsilon_{t-k-2} + \varepsilon_{t-k})] =$$

$$= 0$$

Thus, $\gamma_k = 0$ for $k \geq 3$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{0}{(1+\theta_1^2+\theta_2^2)\sigma^2} = 0$$

Thus, $\rho_k = 0$ for $k \geq 3$

Moving Average process: MA(q) - Mean - Variance

Let y_t follow a Moving Average of order q, MA(q), model:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \ , \ \varepsilon_t \sim N(0, \sigma^2)$$

$$E(y_t) = E(\mu + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t) = \mu$$

$$V(y_t) = E[y_t - E(y_t)]^2 = E[y_t - \mu]^2 =$$
$$E(\theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t)^2 = (1 + \theta_1^2 + \ldots + \theta_q^2)\sigma^2$$

Moving Average process: MA(q)
Autocovariance at lag k, γ_k - Autocorrelation at lag k, ρ_k

$$\gamma_k = Cov(y_t, y_{t-k}) = E[(y_t - E(y_t))(y_{t-k} - E(y_{t-k}))] =$$

$$= E[(y_t - \mu)(y_{t-k} - \mu)] =$$

$$= (\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q)\sigma^2$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{(\theta_k+\theta_1\theta_{k+1}+\ldots+\theta_{q-k}\theta_q)\sigma^2}{(1+\theta_1^2+\ldots+\theta_q^2)\sigma^2} = \frac{\theta_k+\theta_1\theta_{k+1}+\ldots+\theta_{q-k}\theta_q}{1+\theta_1^2+\ldots+\theta_q^2} \neq 0$$

Thus, for $k = 1, 2, \ldots, q$ the autocovariances-autocorrelations are different from zero, while for $k > q$ these metrics are zero.

Moving Average process: MA(q) Useful comments

- The mean, the variance and the autocovariances-autocorrelations are constant over time.
- The MA(q) is weakly stationary process, without imposing restrictions on the model parameters.
- In the MA(q) process the autocovariance-autocorrelation at lag $1, \ldots, q$ are different from zero, and all the other autocorrelations are zero.

Autoregressive process: AR(1) - MA representation
AR(1) as a linear process

Let y_t follow an Autoregressive model of order one, AR(1):

$$y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t$$

This model can be written as a MA(∞) model:

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

$$= \delta + \phi_1\delta + \phi_1^2 y_{t-2} + \phi_1\varepsilon_{t-1} + \varepsilon_t$$

$$= \delta + \phi_1\delta + \phi_1^2(\delta + \phi_1 y_{t-3} + \varepsilon_{t-2}) + \phi_1\varepsilon_{t-1} + \varepsilon_t$$

$$= \delta + \phi_1\delta + \phi_1^2\delta + \phi_1^3 y_{t-3} + \phi_1^2\varepsilon_{t-2} + \phi_1\varepsilon_{t-1} + \varepsilon_t$$

... go backwards m periods

$$= \delta(1 + \phi_1 + \phi_1^2 + \ldots + \phi_1^m) + \sum_{i=0}^{m} \phi_1^i \varepsilon_{t-i} + \phi_1^{m+1} y_{t-m-1}$$

Autoregressive process: AR(1) - MA representation
AR(1) as a linear process

$$y_t = \delta \sum_{i=0}^{m} \phi_1^i + \sum_{i=0}^{m} \phi_1^i \varepsilon_{t-i} + \phi_1^{m+1} y_{t-m-1}$$

if $|\phi_1| < 1$, then $\phi_1^{m+1} \to 0$ as $m \to \infty$

if $|\phi_1| < 1$, then $1 + \phi_1 + \phi_1^2 + \ldots + \phi_1^m = \frac{1}{1-\phi_1}$ as $m \to \infty$

Thus, the AR(1) model can be written as a MA(∞) model:

$$y_t = \frac{\delta}{1-\phi_1} + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}$$

Note that $|\phi_1| < 1$ implies $\sum \left|\phi_1^i\right| < \infty$

Autoregressive process: AR(1) - Mean

Let y_t follow an Autoregressive model of order one, AR(1):

$y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t$ or

$y_t = \frac{\delta}{1-\phi_1} + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}$, if $|\phi_1| < 1$

The mean of the AR(1) process is given by:

$E(y_t) = E(\frac{\delta}{1-\phi_1} + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}) =$

$= E(\frac{\delta}{1-\phi_1}) + E(\sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}) =$

$= \frac{\delta}{1-\phi_1} + \sum_{i=0}^{\infty} \phi_1^i E(\varepsilon_{t-i}) =$

$= \frac{\delta}{1-\phi_1}$

Autoregressive process: AR(1) - Variance

Let y_t follow an Autoregressive model of order one, AR(1):

$y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t$ or

$y_t = \frac{\delta}{1-\phi_1} + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}$, if $|\phi_1| < 1$

The variance of the AR(1) process is given by:

$$V(y_t) = V(\frac{\delta}{1-\phi_1} + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}) =$$

$$= V(\frac{\delta}{1-\phi_1}) + V(\sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}) =$$

$$= \sum_{i=0}^{\infty} \phi_1^{2i} V(\varepsilon_{t-i}) =$$

$$= (1 + \phi_1^2 + \phi_1^4 + \ldots)\sigma^2 =$$

$$= \frac{\sigma^2}{1-\phi_1^2}$$

Autoregressive process: AR(1) - Mean

Let y_t follow an Autoregressive model of order one, AR(1):

$y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t$, $|\phi_1| < 1$ (stationary process)

$E(y_t) = E(y_{t-1}) = \ldots = \mu$

Then the mean of the AR(1) process is calculated by:

$$E(y_t) = E(\delta + \phi_1 y_{t-1} + \varepsilon_t) = E(\delta) + E(\phi_1 y_{t-1}) + E(\varepsilon_t)$$

$$\Rightarrow \mu = \delta + \phi_1 \mu \Rightarrow \mu(1 - \phi_1) = \delta \Rightarrow \mu = \frac{\delta}{1-\phi_1} = E(y_t)$$

Note that for $\phi_1 = 1$ the mean is not defined

Autoregressive process: AR(1) - Variance

Let y_t follow an Autoregressive model of order one, AR(1):

$y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t$, $|\phi_1| < 1$ (stationary process)

$V(y_t) = V(y_{t-1}) = \ldots = v$

Then the variance of the AR(1) process is calculated by:

$$V(y_t) = V(\delta + \phi_1 y_{t-1} + \varepsilon_t) = V(\delta) + V(\phi_1 y_{t-1}) + V(\varepsilon_t)$$

$$\Rightarrow v = \phi_1^2 v + \sigma^2 \Rightarrow v(1 - \phi_1^2) = \sigma^2 \Rightarrow v = \frac{\sigma^2}{1-\phi_1^2} = V(y_t)$$

Note that the variance is well defined if $|\phi_1| < 1$

$$V(y_t) > 0 \Rightarrow \frac{\sigma^2}{1-\phi_1^2} > 0 \Rightarrow 1 - \phi_1^2 > 0$$

$$\Rightarrow \phi_1^2 < 1 \Rightarrow -1 < \phi_1 < 1 \Rightarrow |\phi_1| < 1$$

Autoregressive process: AR(1) - Variance

Let y_t follow an Autoregressive model of order one with mean zero, i.e. $\delta = 0$:

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

Then the variance of the AR(1) process is calculated by:

$$\gamma_0 = Cov(y_t, y_t) = E[(y_t - E(y_t))(y_t - E(y_t))] = E(y_t y_t) =$$

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

$$= \phi_1^2 E(y_{t-1}^2) + E(\varepsilon_t^2) + 2\phi_1 E(y_{t-1}\varepsilon_t) \Rightarrow$$

$$\Rightarrow \gamma_0 = \phi_1^2 \gamma_0 + \sigma^2 \Rightarrow \gamma_0(1 - \phi_1^2) = \sigma^2 \Rightarrow \gamma_0 = \frac{\sigma^2}{1-\phi_1^2}$$

Autoregressive process: AR(1) Autocovariance at lag 1, γ_1 - Autocorrelation at lag 1, ρ_1

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

The autocovariance and autocorrelation at lag 1 are calculated by:

$\gamma_1 = Cov(y_t, y_{t-1}) = E[(y_t - E(y_t))(y_{t-1} - E(y_{t-1}))] =$

$= E(y_t y_{t-1}) = E[(\phi_1 y_{t-1} + \varepsilon_t)y_{t-1}] =$

$= E(\phi_1 y_{t-1}^2 + \varepsilon_t y_{t-1}) = \phi_1 \gamma_0$

$\Rightarrow \gamma_1 = \phi_1 \frac{\sigma^2}{1-\phi_1^2}$

$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\phi_1 \gamma_0}{\gamma_0} = \phi_1$

Autoregressive process: AR(1)
Autocovariance at lag 2, γ_2 - Autocorrelation at lag 2, ρ_2

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

The autocovariance and autocorrelation at lag 2 are calculated by:

$$\gamma_2 = Cov(y_t, y_{t-2}) = E[(y_t - E(y_t))(y_{t-2} - E(y_{t-2}))] =$$

$$= E(y_t y_{t-2}) = E[(\phi_1 y_{t-1} + \varepsilon_t) y_{t-2}] =$$

$$= E[y_{t-2}(\phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t)] =$$

$$= E[y_{t-2}(\phi_1^2 y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t)] =$$

$$= E(\phi_1^2 y_{t-2}^2 + \phi_1 \varepsilon_{t-1} y_{t-2} + \varepsilon_t y_{t-2}) \Rightarrow$$

$$\Rightarrow \gamma_2 = \phi_1^2 \gamma_0 \Rightarrow \gamma_2 = \phi_1^2 \frac{\sigma^2}{1-\phi_1^2}$$

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\phi_1^2 \gamma_0}{\gamma_0} = \phi_1^2$$

Autoregressive process: AR(1)
Autocovariance at lag k, γ_k - Autocorrelation at lag k, ρ_k

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

In this way, the autocovariance at lag k is:

$$\gamma_k = Cov(y_t, y_{t-k}) = E[(y_t - E(y_t))(y_{t-k} - E(y_{t-k}))] \Rightarrow \ldots \Rightarrow$$

$$\Rightarrow \gamma_k = \phi_1^k \gamma_0 \Rightarrow \gamma_k = \phi_1^k \frac{\sigma^2}{1-\phi_1^2}$$

And the autocorrelation at lag k is given by:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\phi_1^k \gamma_0}{\gamma_0} = \phi_1^k$$

Note: In the AR(1) process all the autocorrelations are different from zero, i.e. the process has infinite memory.

Autoregressive process: AR(2) - Autocovariances, γ_k

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

The variance is computed by:

$$\gamma_0 = E(y_t y_t) = E[y_t(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t)] =$$

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

$$\Rightarrow \gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2$$

The autocovariance at lag 1 is computed by:

$$\gamma_1 = E(y_t y_{t-1}) = E[y_{t-1}(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t)] =$$

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

$$\Rightarrow \gamma_1 = \phi_1 \gamma_0 + \phi_2 \gamma_1$$

Autoregressive process: AR(2) - Autocovariances, γ_k

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

The autocovariance at lag 2 is computed by:

$$\gamma_2 = E(y_t y_{t-2}) = E[y_{t-2}(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t)] =$$

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

$$\Rightarrow \gamma_2 = \phi_1 \gamma_1 + \phi_2 \gamma_0$$

Generally, the autocovariance at lag k, $k \geq 2$ is computed by:

$$\gamma_k = E(y_t y_{t-k}) = E[y_{t-k}(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t)] =$$

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

$$\Rightarrow \gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}$$

Autoregressive process: AR(2) - Autocovariances, γ_k

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

Solving the following equations

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma^2$$
$$\gamma_1 = \phi_1\gamma_0 + \phi_2\gamma_1$$
$$\gamma_2 = \phi_1\gamma_1 + \phi_2\gamma_0$$

with respect to γ_0, γ_1 and γ_2 we obtain:

$$\gamma_0 = \frac{(1-\phi_2)\sigma^2}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]}$$

$$\gamma_1 = \frac{\phi_1}{1-\phi_2}\gamma_0$$

and for $k \geq 2$

$$\Rightarrow \gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2}$$

Autoregressive process: AR(2) - Autocorrelations, ρ_k

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

The autocorrelations in the AR(2) process are given by:

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\frac{\phi_1\gamma_0}{1-\phi_2}}{\gamma_0} = \frac{\phi_1}{1-\phi_2}$$

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\phi_1\gamma_1+\phi_2\gamma_0}{\gamma_0} = \phi_2 + \phi_1\rho_1 = \phi_2 + \frac{\phi_1^2}{1-\phi_2}$$

and for $k \geq 2$

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$$

Useful comments: In the AR(2) process all the autocorrelations are different from zero. Similar calculations provide the autocovariances - autocorrelations for the general AR(p) process.

The Partial Autocorrelation Function, α_k

$$AR(1): y_t = \phi_1 y_{t-1} + \varepsilon_t$$
$$AR(2): y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$
$$AR(p): y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

First step: compute the autocovariances, γ_k

$$y_t y_{t-k} = y_{t-k}(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t) \Rightarrow$$

$$E(y_t y_{t-k}) =$$
$$E(\phi_1 y_{t-1} y_{t-k} + \phi_2 y_{t-2} y_{t-k} + \ldots + \phi_p y_{t-p} y_{t-k} + \varepsilon_t y_{t-k}) \Rightarrow$$

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \ldots + \phi_p \gamma_p + \sigma^2$$
$$\gamma_1 = \phi_1 \gamma_0 + \phi_2 \gamma_1 + \ldots + \phi_p \gamma_{p-1}$$
$$\ldots$$
$$\gamma_p = \phi_1 \gamma_{p-1} + \phi_2 \gamma_{p-2} + \ldots + \phi_p \gamma_0$$
$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \ldots + \phi_p \gamma_{k-p}, \text{ for } k > p$$

The Partial Autocorrelation Function, α_k

Second step: compute the autocorrelations, ρ_k

$$\rho_1 = \phi_1 + \phi_2\rho_1 + \ldots + \phi_p\rho_{p-1}$$
$$\ldots$$
$$\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \ldots + \phi_p$$
$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\rho_{k-p}, \text{ for } k > p$$

These equations are called Yule-Walker equations

Third step: To obtain the partial autocorrelations solve the Yule-Walker equations with respect to $\phi_1, \phi_2, \ldots$ iteratively for different values of p, i.e. for $p = 1, p = 2, \ldots$

The Partial Autocorrelation Function, α_k

$p = 1$: Solve the Yule-Walker equations with respect to ϕ_1
Equation: $\rho_1 = \phi_1$
The solution with respect to ϕ_1 is: $\phi_1 = \rho_1$
Thus, the partial autocorrelation at lag 1 is: $\alpha_1 = \hat{\phi}_1 = \hat{\rho}_1$.

$p = 2$: Solve the Yule-Walker equations with respect to ϕ_2
Equations:
$\rho_1 = \phi_1 + \phi_2\rho_1$
$\rho_2 = \phi_1\rho_1 + \phi_2$
The solution with respect to ϕ_2 is: $\phi_2 = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}$
Thus, the partial autocorrelation at lag 2 is: $\alpha_2 = \hat{\phi}_2 = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1-\hat{\rho}_1{}^2}$
and so on...
to obtain general formulas which provide the partial autocorrelations.

The Partial Autocorrelation Function : AR(1) model

Recall that in the AR(1) process, the autocorrelations are given by:
$\rho_1 = \phi_1, \rho_2 = \phi_1^2, \rho_3 = \phi_1^3, \ldots$
Applying the general formulas of partial autocorrelations, using the AR(1) autocorrelation estimates, we obtain:

$\alpha_1 = \rho_1 = \phi_1$, different from zero

$\alpha_2 = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} = \frac{\phi_1^2 - \phi_1^2}{1-\phi_1^2} = 0$

$\alpha_3 = 0$

Useful comments: In the AR(1) model, the partial autocorrelation at lag 1 is different from zero.
Generalization: In the AR(p) model, the partial autocorrelation at lag 1, 2, . . . , p are different from zero, while the remaining partial autocorrelations are zero.

Autoregressive Moving Average: ARMA(1,1) - Mean

Let y_t follow an ARMA(1,1) model:

$y_t = \delta + \phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$, $|\phi_1| < 1$ (stationary process)

$E(y_t) = E(y_{t-1}) = \ldots = \mu$

Then the mean of the ARMA(1,1) process is calculated by:

$$E(y_t) = E(\delta + \phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t) =$$

$$= E(\delta) + \phi_1 E(y_{t-1}) + \theta_1 E(\varepsilon_{t-1}) + E(\varepsilon_t) =$$

$$\Rightarrow \mu = \delta + \phi_1 \mu \Rightarrow \mu(1 - \phi_1) = \delta \Rightarrow \mu = \frac{\delta}{1-\phi_1} = E(y_t)$$

Note that for $\phi_1 = 1$ the mean is not defined.

Autoregressive Moving Average: ARMA(1,1) - Variance

Let y_t follow an ARMA(1,1) model with mean zero:

$y_t = \phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$, $|\phi_1| < 1$ (stationary process)

Then the variance of the ARMA(1,1) process is calculated by:

$$\gamma_0 = V(y_t) = E[(y_t - E(y_t))^2] = E(y_t^2) =$$

$$= E(\phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t)^2 =$$

$$= E(\phi_1^2 y_{t-1}^2 + \theta_1^2 \varepsilon_{t-1}^2 + \varepsilon_t^2 + 2\phi_1\theta_1 y_{t-1}\varepsilon_{t-1} + 2\phi_1 y_{t-1}\varepsilon_t + 2\theta_1 \varepsilon_t \varepsilon_{t-1})$$

$$= \phi_1^2 \gamma_0 + \theta_1^2 \sigma^2 + \sigma^2 + 2\phi_1\theta_1\sigma^2$$

$$\Rightarrow \gamma_0(1 - \phi_1^2) = \sigma^2(1 + \theta_1^2 + 2\theta_1\phi_1)$$

$$\Rightarrow \gamma_0 = \frac{(1+\theta_1^2+2\theta_1\phi_1)\sigma^2}{1-\phi_1^2}$$

Note that the variance is well defined if $|\phi_1| < 1$

Autoregressive Moving Average: ARMA(1,1) - γ_1 - ρ_1

Let y_t follow an ARMA(1,1) model with mean zero:

$y_t = \phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad , \; |\phi_1| < 1$ (stationary process)

The autocovariance and the autocorrelation at lag 1 are given by:

$$\gamma_1 = Cov(y_t, y_{t-1}) = E[(y_t - E(y_t))(y_{t-1} - E(y_{t-1}))] =$$

$$= E(y_t y_{t-1}) = E[(\phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t) y_{t-1}] =$$

$$= E(\phi_1 y_{t-1}^2 + \theta_1 \varepsilon_{t-1} y_{t-1} + \varepsilon_t y_{t-1}) =$$

$$= \phi_1 \gamma_0 + \theta_1 \sigma^2 = \phi_1 \frac{(1+\theta_1^2+2\theta_1\phi_1)\sigma^2}{1-\phi_1^2} + \theta_1 \sigma^2 \Rightarrow$$

$$\Rightarrow \gamma_1 = \frac{(\phi_1+\theta_1)(1+\phi_1\theta_1)}{1-\phi_1^2} \sigma^2$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(\phi_1+\theta_1)(1+\phi_1\theta_1)}{1+\theta_1^2+2\theta_1\phi_1}$$

Autoregressive Moving Average: ARMA(1,1) - γ_2 - ρ_2

Let y_t follow an ARMA(1,1) model with mean zero:

$y_t = \phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$, $|\phi_1| < 1$ (stationary process)

The autocovariance γ_2 of the ARMA(1,1) process is calculated by:

$$\gamma_2 = Cov(y_t, y_{t-2}) = E[(y_t - E(y_t))(y_{t-2} - E(y_{t-2}))] =$$

$$= E(y_t y_{t-2}) = E[(\phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t) y_{t-2}] =$$

$$= E(\phi_1 y_{t-1} y_{t-2} + \theta_1 \varepsilon_{t-1} y_{t-2} + \varepsilon_t y_{t-2}) =$$

$$\Rightarrow \gamma_2 = \phi_1 \gamma_1$$

Then the autocorrelation at lag 1 is calculated by:

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\phi_1 \gamma_1}{\gamma_0} = \phi_1 \rho_1$$

Autoregressive Moving Average: ARMA(1,1) - γ_k - ρ_k

Let y_t follow an ARMA(1,1) model with mean zero:

$y_t = \phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$, $|\phi_1| < 1$ (stationary process)

The autocovariance, γ_k, and the autocorrelation, ρ_k, $k \geq 2$ of the ARMA(1,1) process are given by:

$\gamma_k = \phi_1 \gamma_{k-1}$

$\rho_k = \phi_1 \rho_{k-1}$

Backward Operators

The backward operators are defined as follows:

$$B\varepsilon_t = \varepsilon_{t-1}, B^2\varepsilon_t = \varepsilon_{t-2}, \ldots B^k\varepsilon_t = \varepsilon_{t-k}$$

The AR(p) process can be written:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t \Rightarrow$$

$$\Rightarrow y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \ldots - \phi_p y_{t-p} = \varepsilon_t \Rightarrow$$

$$\Rightarrow (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p) y_t = \varepsilon_t \Rightarrow$$

$$\Rightarrow \Phi(B) y_t = \varepsilon_t$$

The polynomial $\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p$ is called the characteristic polynomial of the AR(p) model.

Backward Operators

The MA(q) process can be written:

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

$$\Rightarrow y_t = (1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q)\varepsilon_t \Rightarrow$$

$$\Rightarrow y_t = \Theta(B)\varepsilon_t$$

Backward Operators

The ARMA(p) process can be written:

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

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

$$\Rightarrow (1 - \phi_1 B - \ldots - \phi_p B^p) y_t = (1 + \theta_1 B + \ldots + \theta_q B^q) \varepsilon_t \Rightarrow$$

$$\Rightarrow \Phi(B) y_t = \Theta(B) \varepsilon_t$$

Stationarity

Example: AR(1) process: $y_t = \phi_1 y_{t-1} + \varepsilon_t \Rightarrow (1 - \phi_1 B) y_t = \varepsilon_t$
That is, the characteristic polynomial of the AR(1) model is:
$\Phi(z) = 1 - \phi_1 z$

$\Phi(z) = 0 \Rightarrow 1 - \phi_1 z = 0 \Rightarrow \phi_1 z = 1 \Rightarrow z = \frac{1}{\phi_1}$

Stationary solution if : $|z| > 1 \Rightarrow \left|\frac{1}{\phi_1}\right| > 1 \Rightarrow |\phi_1| < 1$

An AR(p) process is stationary if the roots of the characteristic polynomial $\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p$ lie outside the unit cirlce.
Therefore, an AR(p) process is stationary if we impose some restrictions on the autoregressive coefficients.

Backward operator

Note that manipulating operators like $\Phi(z)$ is like manipulating polynomials:

$$\frac{1}{1-\phi z} = 1 + \phi z + \phi^2 z^2 + \phi^3 z^3 + \dots$$

provided that $|\phi| < 1$ and $|z| < 1$.
Remember:

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots$$

provided that $|x| < 1$.

Invertibility

Generally speaking, y_t is invertible if $\Phi(B)y_t = \varepsilon_t$.

The Autoregressive processes, AR(p), are invertible without imposing any restrictions on the model parameters.

The Moving average processes are invertible if we impose restrictions on the model parameters.

Consider the MA(q) process:

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

$$\Rightarrow y_t = (1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q)\varepsilon_t \Rightarrow$$

$$\Rightarrow y_t = \Theta(B)\varepsilon_t \Rightarrow \Theta^{-1}(B)y_t = \varepsilon_t$$

So y_t is invertible if $\Theta^{-1}(B)$ converges. It converges if the roots of the polynomial $\Theta(B) = 1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q = 0$ lie outside the unit cirlce.

Invertibility - Example MA(1) process

Consider the MA(1) process:

$$y_t = \theta_1 \varepsilon_{t-1} + \varepsilon_t \Rightarrow$$

$$\Rightarrow y_t = (1 + \theta_1 B)\varepsilon_t \Rightarrow$$

If $|\theta_1| < 1$, we can write:

$$\varepsilon_t = (1 + \theta_1 B)^{-1} y_t \Leftrightarrow$$

$$\Leftrightarrow \varepsilon_t = (1 - \theta_1 B + \theta_1^2 B^2 - \theta_1^3 B^3 + \ldots) y_t \Leftrightarrow$$

$$\Leftrightarrow \varepsilon_t = \sum_{i=0}^{\infty} (-\theta_1)^i y_{t-i}$$

and y_t is invertible.

Estimation of ARMA models

Estimation of ARMA(p,q) models can be done by using

- Least Squares method
  - Minimize the sum of squared residuals of the model under consideration
  - The idea is based on the minimization of a function: no need for distributional assumptions for the error process
- Maximum Likelihood method
  - Maximize the likelihood (log-likelihood) function
  - Use a distribution for the error process

Least Squares Method: AR(p) model

Suppose that y_t is a stationary process, and we want to estimate an AR(p) model:

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

parameter vector: $\theta = (\delta, \phi_1, \phi_2 \ldots, \phi_p)$
Rewrite the model as follows:

$$\varepsilon_t = y_t - \delta - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \ldots - \phi_p y_{t-p}$$

and then minimize the sum of squared errors with respect to the model parameters, i.e

$$\min_\theta \sum_{t=1}^{T} \varepsilon_t^2$$

Least Squares Method: ARMA(p,q) model

Suppose that y_t is a stationary process with mean zero, and we want to estimate an ARMA(p,q) model:

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

parameter vector: $\theta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$
We can write the model as follows:

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

$$\Rightarrow (1 - \phi_1 B - \ldots - \phi_p B^p) y_t = (1 + \theta_1 B + \ldots + \theta_q B^q) \varepsilon_t \Rightarrow$$

$$\Rightarrow \Phi(B) y_t = \Theta(B) \varepsilon_t \Rightarrow \varepsilon_t = \Theta^{-1}(B) \Phi(B) y_t$$

and then minimize the sum of squared errors with respect to the model parameters, i.e

$$\min_\theta \sum_{t=1}^T \varepsilon_t^2$$

Least Squares Method: Useful comments

- For autoregressive models (AR) least squares estimation is straightforward, since the derivatives of the function of the sum of squared residuals with respect to the model parameters are obtained easily, and the corresponding system of equations to be solved is linear.
- For moving average (MA) and autoregressive moving average (ARMA) models, non-linear least squares estimation procedures/routines must be used, due to the MA part of the model (the roots of the corresponding polynomial lie outside the unit circle).

Maximum Likelihood method

- The approach is based on calculating the likelihood i.e. the joint probability density $f_{Y_1,Y_2,\dots,Y_T}(y_1, y_2, \dots, y_T; \theta)$, which might be viewed as the probability density of having observed this particular sample.
- The maximum likelihood estimate (MLE) of θ is the value for which this sample is most likely to have been observed.
- That is, it is the value of θ that maximizes $f_{Y_1,Y_2,\dots,Y_T}(y_1, y_2, \dots, y_T; \theta)$

- step 1: compute the likelihood (or log-likelihood)
- step 2: maximize the likelihood function

Maximum Likelihood method - Compute the likelihood

The likelihood is the joint probability density $f(y_1, y_2, \ldots, y_T; \theta)$ which can be computed as follows:

$$L(\theta; y) = f(y_1, y_2, \ldots, y_T|\theta) =$$

$$= f(y_T|y_1, \ldots, y_{t-1}, \theta) f(y_1, \ldots, y_{t-1}, \theta)$$

$$= f(y_T|y_1, \ldots, y_{t-1}, \theta) f(y_{T-1}|y_1, \ldots, y_{t-2}, \theta) f(y_1, \ldots, y_{t-2}, \theta)$$

$$\ldots$$

$$= f(y_T|y_1, \ldots, y_{t-1}, \theta) f(y_{T-1}|y_1, \ldots, y_{t-2}, \theta) \ldots f(y_2|y_1, \theta) f(y_1|\theta)$$

$$= f(y_1|\theta) \prod_{t=2}^{T} f(y_t|y_{t-1}, \theta)$$

Maximum Likelihood method - Maximize the likelihood

We maximize the likelihood function $L(\theta; y)$ with respect to θ

$$L(\theta; y) = f(y_1|\theta) \prod_{t=2}^{T} f(y_t|y_{t-1}, \theta)$$

or the logarithm of the likelihood function $log[L(\theta; y)]$

$$log[L(\theta; y)] = log[f(y_1|\theta)] + \sum_{t=2}^{T} log[f(y_t|y_{t-1}, \theta)]$$

Maximum Likelihood method: Conditional-Exact likelihood

In order to calculate the likelihood of ARMA models, one will have to decide how to treat the initial values (known as initial conditions) of the $y's$ and the $\varepsilon's$, i.e. the initial values of y and ε that appear in the likelihood function.

There are two approaches:

- Conditional likelihood: treat the initial values as given, i.e. compute the likelihood conditional on the initial values (simplifies the computation of the likelihood function).
- Exact likelihood: treat the initial values as unknown, i.e. consider them to be random variables, which usually follow a Normal distribution with mean and variance based on the unconditional mean and variance of the y_t process.

AR(1) model: Exact likelihood

Suppose we have observed a sample $y_1, y_2, \ldots, y_T$ of size T, and we want to estimate an AR(1) model:

$$y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2).$$

Parameter vector: $\theta = (\delta, \phi_1, \sigma^2)$.

- Consider the probability distribution of y_1, $f(y_1|\theta)$.
  It is a random variable with mean $\frac{\delta}{1-\phi_1}$ and variance $\frac{\sigma^2}{1-\phi_1^2}$
  i.e. $y_1 \sim N(\frac{\delta}{1-\phi_1}, \frac{\sigma^2}{1-\phi_1^2})$.
  Thus, the density of the first observation is given by:
  $f(y_1|\theta) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2/(1-\phi_1^2)}} \exp\left[\frac{-[y_1-(\delta/(1-\phi_1))]^2}{2\sigma^2/(1-\phi_1^2)}\right]$.

AR(1) model: Exact likelihood

- Next consider the distribution of $y_2|y_1$, $f(y_2|y_1,\theta)$

  $y_2 = \delta + \phi_1 y_1 + \varepsilon_2, \varepsilon_2 \sim N(0,\sigma^2)$

  Thus, $y_2|y_1 \sim N(\delta + \phi_1 y_1, \sigma^2)$

  Therefore, the density of the $y_2|y_1$ is given by:

  $f(y_2|y_1,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-[y_2-(\delta+\phi_1 y_1)]^2}{2\sigma^2}\right]$

AR(1) model: Exact likelihood

- In general, the distribution of $y_t|y_{t-1}$, $f(y_t|y_{t-1},\theta)$ can be calculated as follows:

  $y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$

  Thus, $y_t|y_{t-1} \sim N(\delta + \phi_1 y_{t-1}, \sigma^2)$

  Therefore, the density of the $y_t|y_{t-1}$ is given by:

  $f(y_t|y_{t-1},\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-[y_t-(\delta+\phi_1 y_{t-1})]^2}{2\sigma^2}\right]$

AR(1) model: Exact likelihood

The likelihood of the complete sample can be calculated by:

$$L(\theta; y) = f(y_1, y_2, \ldots, y_T|\theta) = f(y_1|\theta) \prod_{t=2}^{T} f(y_t|y_{t-1}, \theta) =$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2/(1-\phi_1^2)}} \exp\left[\frac{-[y_1-(\delta/(1-\phi_1))]^2}{2\sigma^2/(1-\phi_1^2)}\right] \cdot$$

$$\cdot \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-[y_t-(\delta+\phi_1 y_{t-1})]^2}{2\sigma^2}\right]$$

AR(1) model: Exact likelihood

The log-likelihood of the complete sample can be calculated by:

$$log[L(\theta; y)] = log[f(y_1|\theta)] + \sum_{t=2}^{T} log[f(y_t|y_{t-1}, \theta)] =$$

$$= -\frac{1}{2} log(2\pi) - \frac{1}{2} log(\sigma^2/(1-\phi_1^2)) - \frac{[y_1-(\delta/(1-\phi_1))]^2}{2\sigma^2/(1-\phi_1^2)}$$

$$-[(T-1)/2]log(2\pi) - [(T-1)/2]log(\sigma^2) - \sum_{t=2}^{T} \frac{[y_t-\delta-\phi_1 y_{t-1})]^2}{2\sigma^2}$$

AR(1) model: Conditional likelihood

If we consider that the initial value is known/given i.e. is the first observation in the sample, then the conditional likelihood is:

$$L(\theta; y) = f(y_2, \ldots, y_T | y_1, \theta) =$$

$$= f(y_T | y_1, \ldots, y_{t-1}, \theta) f(y_{T-1} | y_1, \ldots, y_{t-2}, \theta) \ldots f(y_2 | y_1, \theta)$$

$$= \prod_{t=2}^{T} f(y_t | y_{t-1}, \theta)$$

$$= \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_t - \delta - \phi_1 y_{t-1})^2}{2\sigma^2}\right]$$

$$= (2\pi\sigma^2)^{-(T-1)/2} \exp\left[-\frac{1}{2} \sum_{t=2}^{T} \frac{(y_t - \delta - \phi_1 y_{t-1})^2}{\sigma^2}\right]$$

AR(1) model: Conditional likelihood

In the conditional likelihood approach, the conditional log-likelihood for the AR(1) model is given by:

$$log[L(\theta; y)] = \sum_{t=2}^{T} log[f(y_t|y_{t-1}, \theta)] =$$

$$= -[(T-1)/2]log(2\pi) - [(T-1)/2]log(\sigma^2) - \sum_{t=2}^{T} \frac{[y_t-\delta-\phi_1 y_{t-1}]^2}{2\sigma^2}$$

Note that, maximization of the conditional log-likelihood with respect to δ and ϕ_1 is equivalent to minimization of

$$\sum_{t=2}^{T}(y_t - \delta - \phi_1 y_{t-1})^2$$

which is achieved by an ordinary least squares (OLS) regression of y_t on a constant and its own lagged values.

MA(1) model: Conditional likelihood

Suppose we have observed a sample $y_1, y_2, \ldots, y_T$ of size T, and we want to estimate an MA(1) model:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

Parameter vector: $\theta = (\mu, \theta_1, \sigma^2)$.

To calculate the conditional likelihood for the MA(1) model, we condition on the initial values of the $\varepsilon' s$.

- Based on the MA(1) model, $y_1 = \mu + \theta_1 \varepsilon_0 + \varepsilon_1$
  Conditional on $\varepsilon_0 = 0$, $y_1 = \mu + \varepsilon_1$ or $\varepsilon_1 = y_1 - \mu$ and $(y_1 | \varepsilon_0 = 0) \sim N(\mu, \sigma^2)$
  The conditional density of the first observation is given by:
  $f(y_1 | \varepsilon_0 = 0, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_1-\mu)^2}{2\sigma^2}\right]$

MA(1) model: Conditional likelihood

- Based on the MA(1) model, $y_2 = \mu + \theta_1 \varepsilon_1 + \varepsilon_2$
  $\varepsilon_2 = y_2 - \mu - \theta_1 \varepsilon_1 = y_2 - \mu - \theta_1 (y_1 - \mu)$
  Thus, the conditional distribution of y_2 given $y_1, \varepsilon_0 = 0, \theta$, is given by:
  $(y_2 | y_1, \varepsilon_0 = 0, \theta) \sim N(\mu + \theta_1 \varepsilon_1, \sigma^2)$
  The conditional density of the second observation is given by:
  $f(y_2 | y_1, \varepsilon_0 = 0, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_2 - \mu - \theta_1 \varepsilon_1)^2}{2\sigma^2}\right]$

MA(1) model: Conditional likelihood

- In general, the distribution of $y_t|y_1,\ldots,y_{t-1},\varepsilon_0=0,\theta$, $f(y_t|y_1,\ldots,y_{t-1},\varepsilon_0=0,\theta)$ can be calculated as follows:

  $$y_t=\mu+\theta_1\varepsilon_{t-1}+\varepsilon_t \Rightarrow \varepsilon_t=y_t-\mu-\theta_1\varepsilon_{t-1}$$

  Thus, $(y_t|y_1,\ldots,y_{t-1},\varepsilon_0=0,\theta)\sim N(\mu+\theta_1\varepsilon_{t-1},\sigma^2)$

  Therefore, the density of the $y_t|y_1,\ldots,y_{t-1},\varepsilon_0=0,\theta$ is given by:

  $$f(y_t|y_1,\ldots,y_{t-1},\varepsilon_0=0,\theta)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[\frac{-[y_t-\mu-\theta_1\varepsilon_{t-1}]^2}{2\sigma^2}\right]$$

MA(1) model: Conditional likelihood

The conditional likelihood for the MA(1) model is given by:

$$L(\theta; y) = f(y_1, y_2, \ldots, y_T | \varepsilon_0 = 0, \theta) =$$

$$= f(y_1 | \varepsilon_0 = 0, \theta) \prod_{t=2}^{T} f(y_t | \varepsilon_0 = 0, y_1, \ldots, y_{t-1}, \theta) =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_1 - \mu)^2}{2\sigma^2}\right] \cdot \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_t - \mu - \theta_1 \varepsilon_{t-1})^2}{2\sigma^2}\right]$$

$$= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-\varepsilon_t^2}{2\sigma^2}\right]$$

$$= (2\pi\sigma^2)^{-T/2} \exp\left[-\sum_{t=1}^{T} \frac{\varepsilon_t^2}{2\sigma^2}\right]$$

MA(1) model: Conditional likelihood

The conditional log-likelihood for the MA(1) model is given by:

$$log[L(\theta; y)] = \sum_{t=1}^{T} log[f(y_t|y_{t-1}, \varepsilon_0 = 0, \theta)] =$$

$$= -\frac{T}{2} log(2\pi) - \frac{T}{2} log(\sigma^2) - \sum_{t=1}^{T} \frac{\varepsilon_t^2}{2\sigma^2}$$

Maximum likelihood Method: Useful comments

- Two approaches: exact and conditional likelihood.
- AR(p) models - conditional likelihood: maximization is straightforward, since the derivatives of the log-likelihood function with respect to the model parameters are obtained easily, and the corresponding system of equations to be solved is linear (similar to least squares method).
- AR(p) models - exact likelihood: maximization requires iterative or numerical procedures.
- MA(q), ARMA(p,q) models - conditional likelihood: maximization requires iterative or numerical procedures (due to the MA part of the model, i.e. the $\varepsilon's$ are constructed iteratively creating nonlinearities - the roots of the corresponding polynomial lie outside the unit circle).

Model fit

Different measures can be used to assess the model fit.

- Akaike's information criterion (AIC):

$$AIC_m = -2\log(L_m) + 2np_m \text{ or}$$
$$AIC_m = -2\log(L_m)/T + 2np_m/T$$

- Schwartz's information criterion or Bayesian information criterion (BIC):

$$BIC_m = -2\log(L_m) + np_m\log(T) \text{ or}$$
$$BIC_m = -2\log(L_m)/T + np_m\log(T)/T$$

where L_m is the value of the likelihood for the fitted model m, np_m is the number of model parameters, T is sample size (number of observations).

- Small values of AIC, BIC indicate better fit.

Diagnostic checking

After estimating an identified model, the residuals must be (resemble) a white noise process, i.e. must be:

- Uncorrelated
- Homoskedastic
- Normally distributed

In ARMA(p,q) models, the residuals are estimated through the following formula:

$$\hat{\varepsilon}_t = \hat{\Theta}^{-1}(B)\hat{\Phi}(B)y_t$$

Test for Autocorrelation of Residuals

Different tests can be used to test for autocorrelation of residuals:

- Bartlett test
- Box-Pierce and Ljung-Box test
- Autocorrelation and partial autocorrelation plots
- Durbin-Watson test (autocorrelation test at lag 1)
- Breusch-Godfrey test

Test for Autocorrelation of Residuals

Estimate the autocorrelation of residuals at lag k:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{T-k} \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2}$$

Bartlett's test (for a particular lag k):
$H_0 : \rho_k = 0$
$H_1 : \rho_k \neq 0$

If the residuals are random (white noise), then the sampling distribution of $\hat{\rho}_k$ is approximately normal, i.e. $\hat{\rho}_k \sim N(0, \frac{1}{T})$ test statistic: $Z = \frac{\hat{\rho}_k - 0}{\sqrt{1/T}} \sim N(0, 1)$

Reject H_0, at level of significance α, if the observed value of the test statistic $Z < -Z_{1-\alpha/2}$ or $Z > Z_{1-\alpha/2}$

Test for Autocorrelation of Residuals

$H_0: \rho_1 = \rho_2 = \ldots = \rho_m = 0$, for a fixed m
$H_1: \rho_i \neq 0$, for at least one $i \leq m$

Box-Pierce test statistic: $Q = T\sum_{k=1}^{m} \hat{\rho}_k^2 \sim \chi^2_{m-np}$

Ljung-Box test statistic: $LB = T(T+2)\sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{T-k} \sim \chi^2_{m-np}$
np: is the number of ARMA parameters, i.e. $np = p + q$, that have been estimated in the model under consideration

Reject H_0, at level of significance α, if the observed value of the test statistic $Q > \chi^2_{m-np,1-\alpha}$ ($LB > \chi^2_{m-np,1-\alpha}$)

Test for Heteroskedasticity of Residuals

Different tests (directly or indirectly) can be used to test for heteroscedasticity of residuals:

- Autocorrelation test of squared residuals
- Autocorrelation and partial autocorrelation plots of squared residuals
- Goldfeld-Quandt test
- Breusch-Pagan test
- White test
- ARCH LM test of heteroscedasticity

Normality Test for residuals

Different tests can be used to test for normality of residuals:

- Jarque-Bera test
- Kolmogorov-Smirnov test
- Shapiro-Wilk test
- QQ-plot

Jarque-Bera Normality Test of Residuals

$H_0 : \varepsilon_t$ follow a Normal distribution
$H_1 : \varepsilon_t$ are not Normal

Jarque-Bera test statistic: $JB = \frac{T-np}{6}[S^2 + \frac{(K-3)^2}{4}] \sim \chi^2_2$

where S denotes the skewness, K denotes the kurtosis, T is the sample size, and np is the number of parameters in the model under consideration

Reject H_0, at level of significance α, if the observed value of the test statistic $JB > \chi^2_{2,1-\alpha}$

Forecasting: the Loss Function

- Suppose we are interested in forecasting the value of y_{t+1} based on a set of observations $y_t, y_{t-1}, \ldots, y_{t-m+1}$ i.e. based on the m most recent values of the series
- Let $\hat{y}_{t+1|t}$ denote the forecast of y_{t+1}. The usefulness of this forecast can be specified with respect to a certain loss function
- The most commonly used loss function is the quadratic loss function. The best prediction $\hat{y}_{t+1|t}$ according to the quadratic loss function is that which minimizes the Mean Squared Error:

$$MSE(\hat{y}_{t+1|t}) = E(y_{t+1} - \hat{y}_{t+1|t})^2$$

Forecasting: Conditional Expectation

Theorem: The minimum Mean Squared Error predictor $\hat{y}_{t+1|t}$ of y_{t+1} is given by the conditional expectation:

$$\hat{y}_{t+1|t} = E(y_{t+1}|y_t, y_{t-1}, \ldots, y_{t-m+1}) = E(y_{t+1}|\tilde{y}_t)$$

Proof

Let $\hat{y}_{t+1|t} = g(y_t, y_{t-1}, \ldots, y_{t-m+1}) = g(\tilde{y}_t)$, where g is any function of the most recent values of the series

$$E[y_{t+1} - \hat{y}_{t+1|t}]^2 = E[y_{t+1} - g(\tilde{y}_t)]^2 =$$

$$= E[y_{t+1} - E(y_{t+1}|\tilde{y}_t) + E(y_{t+1}|\tilde{y}_t) - g(\tilde{y}_t)]^2 =$$

$$= E[y_{t+1} - E(y_{t+1}|\tilde{y}_t)]^2 + E[E(y_{t+1}|\tilde{y}_t) - g(\tilde{y}_t)]^2 +$$

$$+2E[[y_{t+1} - E(y_{t+1}|\tilde{y}_t)][E(y_{t+1}|\tilde{y}_t) - g(\tilde{y}_t)]] +$$

$$= E[y_{t+1} - E(y_{t+1}|\tilde{y}_t)]^2 + E[E(y_{t+1}|\tilde{y}_t) - g(\tilde{y}_t)]^2$$

Forecasting: Conditional Expectation

Denote $\eta_{t+1} = [y_{t+1} - E(y_{t+1}|\tilde{y}_t)][E(y_{t+1}|\tilde{y}_t) - g(\tilde{y}_t)]$

Law of iterated expectations: $E(\eta_{t+1}) = E_{\tilde{y}_t}(E[\eta_{t+1}|\tilde{y}_t])$

$E[\eta_{t+1}|\tilde{y}_t] = E[[y_{t+1} - E(y_{t+1}|\tilde{y}_t)][E(y_{t+1}|\tilde{y}_t) - g(\tilde{y}_t)]|\tilde{y}_t] =$

$= [E(y_{t+1}|\tilde{y}_t) - g(\tilde{y}_t)] \cdot E[[y_{t+1} - E(y_{t+1}|\tilde{y}_t)]|\tilde{y}_t] =$

$= [E(y_{t+1}|\tilde{y}_t) - g(\tilde{y}_t)] \cdot 0 = 0$

Thus, $E(\eta_{t+1}) = E_{\tilde{y}_t}(E[\eta_{t+1}|\tilde{y}_t]) = 0$

Forecasting: Conditional Expectation

Therefore, $E[y_{t+1} - \hat{y}_{t+1|t}]^2 = E[y_{t+1} - g(\tilde{y}_t)]^2 =$

$$= E[y_{t+1} - E(y_{t+1}|\tilde{y}_t)]^2 + E[E(y_{t+1}|\tilde{y}_t) - g(\tilde{y}_t)]^2$$

The function $g(\tilde{y}_t)$ that makes the Mean Squared Error as small as possible is the function that makes the second term zero, i.e. $E(y_{t+1}|\tilde{y}_t) = g(\tilde{y}_t)$

That is, the forecast $g(\tilde{y}_t)$ that minimize the mean squared error is the conditional expectation $E(y_{t+1}|\tilde{y}_t)$

The Mean Squared Error of this optimal forecast is:

$$E[y_{t+1} - \hat{y}_{t+1|t}]^2 = E[y_{t+1} - g(\tilde{y}_t)]^2 = E[y_{t+1} - E(y_{t+1}|\tilde{y}_t)]^2$$

Forecasting a MA(1) process

Consider the MA(1) model: $y_t = \mu + \theta_1\varepsilon_{t-1} + \varepsilon_t$

The one-step ahead forecast, $\hat{y}_{t+1|t}$, is computed as follows:

$$\hat{y}_{t+1|t} = E(y_{t+1}|\Phi_t) = E[\mu + \theta_1\varepsilon_t + \varepsilon_{t+1}|\Phi_t] =$$

$$= E(\mu|\Phi_t) + E(\theta_1\varepsilon_t|\Phi_t) + E(\varepsilon_{t+1}|\Phi_t) =$$

$$= \mu + \theta_1\varepsilon_t$$

The Mean Squared Error of $\hat{y}_{t+1|t}$ is:

$$MSE(\hat{y}_{t+1|t}) = E[y_{t+1} - \hat{y}_{t+1|t}]^2 =$$

$$= E[\mu + \theta_1\varepsilon_t + \varepsilon_{t+1} - \mu - \theta_1\varepsilon_t]^2$$

$$= E(\varepsilon_{t+1})^2 = \sigma^2$$

Forecasting a MA(1) process

The forecast error is given by:

$\hat{\varepsilon}_{t+1} = y_{t+1} - \hat{y}_{t+1|t} = \varepsilon_{t+1}$ [see previews slide]

The variance of the forecast error is given by:

$V(\hat{\varepsilon}_{t+1}) = V(y_{t+1} - \hat{y}_{t+1|t}) = V(\varepsilon_{t+1}) = \sigma^2$

- the variance of the forecast error is equal to the mean squared error of the forecast
- the standard error of the one-step ahead forecast in the MA(1) model is given by the square root of its variance or by the square root of the mean squared error of the forecast
- it is useful to evaluate the accuracy of the forecasts as well to construct confidence intervals

Forecasting a MA(1) process

Consider the MA(1) model: $y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t$

The two-step ahead forecast, $\hat{y}_{t+2|t}$, is computed as follows:

$$\hat{y}_{t+2|t} = E(y_{t+2}|\Phi_t) = E[\mu + \theta_1 \varepsilon_{t+1} + \varepsilon_{t+2}|\Phi_t] =$$

$$= E(\mu|\Phi_t) + E(\theta_1 \varepsilon_{t+1}|\Phi_t) + E(\varepsilon_{t+2}|\Phi_t) =$$

$$= \mu$$

The Mean Squared Error of $\hat{y}_{t+2|t}$ is:

$$MSE(\hat{y}_{t+2|t}) = E[y_{t+2} - \hat{y}_{t+2|t}]^2 =$$

$$= E[\mu + \theta_1 \varepsilon_{t+1} + \varepsilon_{t+2} - \mu]^2$$

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

Forecasting a MA(1) process

The forecast error is given by:

$\hat{\varepsilon}_{t+2} = y_{t+2} - \hat{y}_{t+2|t} = \theta_1 \varepsilon_{t+1} + \varepsilon_{t+2}$ [see previews slide]

The variance of the forecast error is given by:

$$V(\hat{\varepsilon}_{t+2}) = V(y_{t+2} - \hat{y}_{t+2|t}) = V(\theta_1 \varepsilon_{t+1} + \varepsilon_{t+2}) = (1 + \theta_1)\sigma^2$$

Forecasting a MA(1) process

Consider the MA(1) model: $y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t$

The s-step ahead forecast, $\hat{y}_{t+s|t}$, is computed as follows:

$$\hat{y}_{t+s|t} = E(y_{t+s}|\Phi_t) = E[\mu + \theta_1 \varepsilon_{t+s-1} + \varepsilon_{t+s}|\Phi_t] = \mu$$

The Mean Squared Error of $\hat{y}_{t+s|t}$ is:

$$MSE(\hat{y}_{t+s|t}) = E[y_{t+s} - \hat{y}_{t+s|t}]^2 = E[\mu + \theta_1 \varepsilon_{t+s-1} + \varepsilon_{t+s} - \mu]^2$$

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

Notice that

$$E(\varepsilon_{t+1}|\Phi_t) = E(\varepsilon_{t+2}|\Phi_t) = \ldots = E(\varepsilon_{t+s}|\Phi_t) = 0$$

$$E(\varepsilon_1|\Phi_t) = \varepsilon_1, E(\varepsilon_2|\Phi_t) = \varepsilon_2, \ldots, E(\varepsilon_t|\Phi_t) = \varepsilon_t$$

Forecasting a MA(1) process

The forecast error is given by:

$\hat{\varepsilon}_{t+s} = y_{t+s} - \hat{y}_{t+s|t} = \theta_1 \varepsilon_{t+s-1} + \varepsilon_{t+s}$ [see previews slide]

The variance of the forecast error is given by:

$$V(\hat{\varepsilon}_{t+s}) = V(y_{t+s} - \hat{y}_{t+s|t}) = V(\theta_1 \varepsilon_{t+s-1} + \varepsilon_{t+s}) = (1 + \theta_1)\sigma^2$$

Forecasting an AR(1) process

Consider the AR(1) model: $y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t$

The one-step ahead forecast, $\hat{y}_{t+1|t}$, is computed as follows:

$$\hat{y}_{t+1|t} = E(y_{t+1}|\Phi_t) = E[\delta + \phi_1 y_t + \varepsilon_{t+1}|\Phi_t] =$$

$$= E(\delta|\Phi_t) + E(\phi_1 y_t|\Phi_t) + E(\varepsilon_{t+1}|\Phi_t) =$$

$$= \delta + \phi_1 y_t$$

The Mean Squared Error of $\hat{y}_{t+1|t}$ is:

$$MSE(\hat{y}_{t+1|t}) = E[y_{t+1} - \hat{y}_{t+1|t}]^2 =$$

$$= E[\delta + \phi_1 y_t + \varepsilon_{t+1} - \delta - \phi_1 y_t]^2$$

$$= E(\varepsilon_{t+1})^2 = \sigma^2$$

Forecasting an AR(1) process

The forecast error is given by:

$\hat{\varepsilon}_{t+1} = y_{t+1} - \hat{y}_{t+1|t} = \varepsilon_{t+1}$ [see previews slide]

The variance of the forecast error is given by:

$V(\hat{\varepsilon}_{t+1}) = V(y_{t+1} - \hat{y}_{t+1|t}) = V(\varepsilon_{t+1}) = \sigma^2$

Forecasting an AR(1) process

Consider the AR(1) model: $y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t$

The two-step ahead forecast, $\hat{y}_{t+2|t}$, is computed as follows:

$$\hat{y}_{t+2|t} = E(y_{t+2}|\Phi_t) = E[\delta + \phi_1 y_{t+1} + \varepsilon_{t+2}|\Phi_t] =$$

$$= E(\delta|\Phi_t) + E(\phi_1 y_{t+1}|\Phi_t) + E(\varepsilon_{t+1}|\Phi_t) =$$

$$= \delta + \phi_1(\delta + \phi_1 y_t) = \delta + \phi_1 \delta + \phi_1^2 y_t$$

$$= \delta(1 + \phi_1) + \phi_1^2 y_t$$

or $\hat{y}_{t+2|t} = \delta + \phi_1 \hat{y}_{t+1|t}$

Forecasting an AR(1) process

Consider the AR(1) model: $y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t$

The Mean Squared Error of $\hat{y}_{t+2|t}$ is:

$$MSE(\hat{y}_{t+2|t}) = E[y_{t+2} - \hat{y}_{t+2|t}]^2 =$$

$$= E[\delta + \phi_1 y_{t+1} + \varepsilon_{t+2} - (\delta + \phi_1 \delta + \phi_1^2 y_t)]^2$$

$$= E[\delta + \phi_1(\delta + \phi_1 y_t + \varepsilon_{t+1}) + \varepsilon_{t+2} - \delta - \phi_1 \delta - \phi_1^2 y_t]^2$$

$$= E[\delta + \phi_1 \delta + \phi_1^2 y_t + \phi_1 \varepsilon_{t+1} + \varepsilon_{t+2} - \delta - \phi_1 \delta - \phi_1^2 y_t]^2$$

$$= E[\phi_1 \varepsilon_{t+1} + \varepsilon_{t+2}]^2 = (1 + \phi_1^2)\sigma^2$$

Forecasting an AR(1) process

The forecast error is given by:

$\hat{\varepsilon}_{t+2} = y_{t+2} - \hat{y}_{t+2|t} = \phi_1 \varepsilon_{t+1} + \varepsilon_{t+2}$ [see previews slide]

The variance of the forecast error is given by:

$V(\hat{\varepsilon}_{t+2}) = V(y_{t+2} - \hat{y}_{t+2|t}) = V(\phi_1 \varepsilon_{t+1} + \varepsilon_{t+2}) = (1 + \phi_1^2)\sigma^2$

Forecasting an AR(1) process

Consider the AR(1) model: $y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t$

The s-step ahead forecast, $\hat{y}_{t+s|t}$, is computed as follows:

$$\hat{y}_{t+s|t} = E(y_{t+s}|\Phi_t) = E[\delta + \phi_1 y_{t+s-1} + \varepsilon_{t+s}|\Phi_t] =$$

$$= E(\delta|\Phi_t) + E(\phi_1 y_{t+s-1}|\Phi_t) + E(\varepsilon_{t+s}|\Phi_t) = \delta + \phi_1 E(y_{t+s-1}|\Phi_t) =$$

$$= \delta + \phi_1(\delta + \phi_1 E(y_{t+s-2}|\Phi_t)) = \delta + \phi_1\delta + \phi_1^2 E(y_{t+s-2}|\Phi_t) = \ldots =$$

$$= \delta(1 + \phi_1 + \phi_1^2 + \ldots + \phi_1^{s-1}) + \phi_1^s y_t$$

or $\hat{y}_{t+s|t} = \delta + \phi_1 \hat{y}_{t+s-1|t}$

Forecasting an AR(1) process

Consider the AR(1) model: $y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t$

The Mean Squared Error of $\hat{y}_{t+s|t}$ is:

$$MSE(\hat{y}_{t+s|t}) = E[y_{t+s} - \hat{y}_{t+s|t}]^2 = \ldots = \sigma^2(1 + \phi_1^2 + \ldots + \phi_1^{2(s-1)})$$

The forecast error is given by:

$$\hat{\varepsilon}_{t+s} = y_{t+s} - \hat{y}_{t+s|t} =$$

$$= \varepsilon_{t+s} + \phi_1 \varepsilon_{t+s-1} + \phi_1^2 \varepsilon_{t+s-2} + \ldots + \phi_1^{s-1} \varepsilon_{t+1}$$

The variance of the forecast error is given by:

$$V(\hat{\varepsilon}_{t+s}) = V(y_{t+s} - \hat{y}_{t+s|t}) = \sigma^2(1 + \phi_1^2 + \ldots + \phi_1^{2(s-1)})$$

Similar computations provide the forecasts and the Mean Squared Error for any MA, AR, ARMA model

Confidence Intervals of the Forecasts

Consider a time series model [AR, MA, ARMA]. Assuming Normal errors, i.e. the process $\varepsilon_t \sim N(0, \sigma^2)$, the forecast errors follow approximately (asymptotically) a normal distribution:

$$\hat{\varepsilon}_{t+s} = y_{t+s} - \hat{y}_{t+s|t} \sim N[0, V(\hat{\varepsilon}_{t+s})]$$

and a $(1-\alpha)100\%$ confidence interval for the forecast y_{t+s} is computed by

$$[\hat{y}_{t+s|t} - Z_{1-\frac{\alpha}{2}}\sqrt{V(\hat{\varepsilon}_{t+s})}, \hat{y}_{t+s|t} + Z_{1-\frac{\alpha}{2}}\sqrt{V(\hat{\varepsilon}_{t+s})}$$

where α is the level of significance and $Z_{1-\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}$ percentile of the standard normal distribution

Application to financial and economic series

- Time series modeling and forecasting financial return series
- Example 1: ARMA modeling and forecasting Johnson & Johnson quarterly data