Outline of the Lecture

(1) Introduction to univariate time series analysis.
(2) Stationarity.
(3) Characterizing time dependence: ACF and PACF.
(4) Modelling time dependence: the ARMA(p,q) model
(5) Examples:
   • AR(1).
   • AR(2).
   • MA(1).
(6) Lag operators, lag polynomials and invertibility.
(7) Model selection.
(8) Estimation.
(9) Forecasting.
Univariate Time Series Analysis

- We consider a single time series, \( y_1, y_2, \ldots, y_T \).
  We want to construct simple models for \( y_t \) as a function of the past: \( E[y_t | \text{history}] \).

- Univariate models are useful for:
  1. Analyzing the dynamic properties of a time series.
     What is the dynamic adjustment after a shock?
     Do shocks have transitory or permanent effects (presence of unit roots)?
  2. Forecasting.
     A model for \( E[y_t | x_t] \) is only useful for forecasting \( y_{t+1} \) if we know (or can forecast) \( x_{t+1} \).
  3. Univariate time series analysis is a way to introduce the tools necessary for analyzing more complicated models.

Stationarity

- A time series, \( y_1, y_2, \ldots, y_t, \ldots, y_T \), is strictly stationary if the joint distributions
  \[
  (y_{t_1}, y_{t_2}, \ldots, y_{t_n}) \quad \text{and} \quad (y_{t_1+h}, y_{t_2+h}, \ldots, y_{t_n+h})
  \]
  are the same for all \( h \).

- A time series is called weakly stationary or covariance stationary if
  \[
  E[y_t] = \mu \\
  V[y_t] = E[(y_t - \mu)^2] = \gamma_0 \\
  Cov[y_t, y_{t-k}] = E[(y_t - \mu)(y_{t-k} - \mu)] = \gamma_k \quad \text{for} \quad k = 1, 2, \ldots
  \]
  Often \( \mu, \gamma_0, \) and \( \gamma_k \) are assumed finite.

- On these slides we consider only stationary processes.
  Later we consider (unit root) non-stationary processes.
The Autocorrelation Function (ACF)

- For a stationary time series we define the autocorrelation function (ACF) as

\[
\rho_k = \text{Corr}(y_t, y_{t-k}) = \frac{\gamma_k}{\gamma_0} = \frac{\text{Cov}(y_t, y_{t-k})}{\sqrt{\text{V}(y_t) \cdot \text{V}(y_{t-k})}},
\]

Note that \(-1 \leq \rho_k \leq 1\), \(\rho_0 = 1\), and \(\rho_k = \rho_{-k}\).

- Recall that the ACF can (e.g.) be estimated by OLS in the regression model

\[
y_t = c + \rho_k y_{t-k} + \text{residual}.
\]

- If \(\rho_1 = \rho_2 = \ldots = 0\), it holds that

\[
V(\hat{\rho}_k) = T^{-1},
\]

and a 95% confidence band is given by \(\pm 2/\sqrt{T}\).
The Partial Autocorrelation Function (PACF)

- An alternative measure is the partial autocorrelation function (PACF), which is the correlation conditional on the intermediate values, i.e.
  \[ \text{Corr}(y_t; y_{t-k} | y_{t-1}, \ldots, y_{t-k+1}). \]

- The PACF can be estimated as the OLS estimator \( \hat{\theta}_k \) in the regression
  \[ y_t = c + \theta_1 y_{t-1} + \ldots + \theta_k y_{t-k} + \text{residual}, \]
  where the intermediate lags are included.

- If \( \theta_1 = \theta_2 = \ldots = 0 \), it holds that
  \[ V(\hat{\theta}_k) = T^{-1}, \]
  and a 95% confidence band is given by \( \pm 2/\sqrt{T} \).

Example: Danish GDP

![Graphs showing Danish GDP, deviation from trend, and first difference over time, along with ACF and PACF plots.](image-url)
The ARMA(p,q) Model

- We consider two simple models for $y_t$:
  - The autoregressive AR(p) model and the moving average MA(q) model.
- First define a white noise process, $\epsilon_t \sim i.i.d.(0, \sigma^2)$.
- The AR(p) model is defined as
  $$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \ldots + \theta_p y_{t-p} + \epsilon_t.$$  
  Systematic part of $y_t$ is a linear function of $p$ lagged values.
- The MA(q) model is defined as
  $$y_t = \epsilon_t + \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2} + \ldots + \alpha_q \epsilon_{t-q}.$$  
  $y_t$ is a moving average of past shocks to the process.
- They can be combined into the ARMA(p,q) model
  $$y_t = \theta_1 y_{t-1} + \ldots + \theta_p y_{t-p} + \epsilon_t + \alpha_1 \epsilon_{t-1} + \ldots + \alpha_q \epsilon_{t-q}.$$  

Dynamic Properties of an AR(1) Model

- Consider the AR(1) model
  $$Y_t = \delta + \theta Y_{t-1} + \epsilon_t.$$  
  Assume for a moment that the process is stationary. As we will see later, this requires $|\theta| < 1$.
- First we want to find the expectation.
  Stationarity implies that $E[Y_t] = E[Y_{t-1}] = \mu$. We find
  $$E[Y_t] = E[\delta + \theta Y_{t-1} + \epsilon_t] = \delta + \theta E[Y_{t-1}] + E[\epsilon_t]$$  
  $$\begin{align*}
  (1 - \theta) \mu &= \delta \\
  \mu &= \frac{\delta}{1 - \theta}.
  \end{align*}$$  
  Note the following:
  (1) The effect of the constant term, $\delta$, depends on the autoregressive parameter, $\theta$.
  (2) $\mu$ is not defined if $\theta = 1$. This is excluded for a stationary process.
Next we want to calculate the variance and the autocovariances. It is convenient to define the deviation from mean, \( y_t = Y_t - \mu \), so that

\[
Y_t = \delta + \theta Y_{t-1} + \epsilon_t \\
Y_t = (1 - \theta) \mu + \theta Y_{t-1} + \epsilon_t \\
Y_t - \mu = \theta (Y_{t-1} - \mu) + \epsilon_t \\
y_t = \theta y_{t-1} + \epsilon_t.
\]

We note that \( \gamma_0 = V[Y_t] = V[y_t] \). We find:

\[
V[y_t] = E[y_t^2] \\
= E[(\theta y_{t-1} + \epsilon_t)^2] \\
= E[\theta^2 y_{t-1}^2 + \epsilon_t^2 + 2\theta y_{t-1}\epsilon_t] \\
= \theta^2 E[y_{t-1}^2] + E[\epsilon_t^2] + 2\theta E[y_{t-1}\epsilon_t] \\
= \theta^2 V[y_{t-1}] + V[\epsilon_t] + 0.
\]

Using stationarity, \( \gamma_0 = V[y_t] = V[y_{t-1}] \), we get

\[
\gamma_0(1 - \theta^2) = \sigma^2 \quad \text{or} \quad \gamma_0 = \frac{\sigma^2}{1 - \theta^2}.
\]

The covariances are given by

\[
\gamma_1 = \text{Cov}[y_t, y_{t-1}] = E[y_t y_{t-1}] \\
= E[(\theta y_{t-1} + \epsilon_t) y_{t-1}] \\
= \theta E[y_{t-1}^2] + \theta \epsilon_t y_{t-1} \\
= \theta \sigma^2 \frac{\gamma_0}{1 - \theta^2} = \theta \gamma_0
\]

\[
\gamma_2 = \text{Cov}[y_t, y_{t-2}] = E[y_t y_{t-2}] \\
= E[(\theta y_{t-1} + \epsilon_t) y_{t-2}] \\
= E[(\theta (\theta y_{t-2} + \epsilon_{t-1}) + \epsilon_t) y_{t-2}] \\
= \theta^2 E[y_{t-2}^2] + \theta y_{t-2} \epsilon_{t-1} + y_{t-2} \epsilon_t \\
= \theta^2 E[y_{t-2}^2] + \theta E[y_{t-2} \epsilon_{t-1}] + E[y_{t-2} \epsilon_t] = \theta^2 \frac{\sigma^2}{1 - \theta^2} = \theta^2 \gamma_0
\]

\[
\gamma_k = \text{Cov}[y_t, y_{t-k}] = \theta^k \gamma_0
\]

The ACF is given by

\[
\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\theta^k \gamma_0}{\gamma_0} = \theta^k.
\]

The PACF is simply the autoregressive coefficients: \( \theta_1, 0, 0, ... \)
Examples of Stationary AR(1) Models

$y_t = 0.8y_{t-1} + \varepsilon_t$

$y_t = -0.8y_{t-1} + \varepsilon_t$

$y_t = 0.5y_{t-1} + \varepsilon_t$

$y_t = 0.95y_{t-1} + \varepsilon_t$

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

$y_t = 1.05y_{t-1} + \varepsilon_t$
Dynamic Properties of an AR(2) Model

- Consider the AR(2) model given by
  \[ Y_t = \delta + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \epsilon_t. \]

- Again we find the mean under stationarity:
  \[
  E[Y_t] = \delta + \theta_1 E[Y_{t-1}] + \theta_2 E[Y_{t-2}] + E[\epsilon_t]
  \quad \Rightarrow \quad
  E[Y_t] = \frac{\delta}{1 - \theta_1 - \theta_2} = \mu.
  \]

- We then define the process \( y_t = Y_t - \mu \) for which it holds that
  \[ y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \epsilon_t. \]

- Multiplying both sides with \( y_t \) and taking expectations yields
  \[
  E[y_t^2] = \theta_1 E[y_{t-1}y_t] + \theta_2 E[y_{t-2}y_t] + E[\epsilon_t y_t]
  \gamma_0 = \theta_1 \gamma_1 + \theta_2 \gamma_2 + \sigma^2
  \]
  Multiplying instead with \( y_{t-1} \) yields
  \[
  E[y_t y_{t-1}] = \theta_1 E[y_{t-1}y_{t-1}] + \theta_2 E[y_{t-2}y_{t-1}] + E[\epsilon_t y_{t-1}]
  \gamma_1 = \theta_1 \gamma_0 + \theta_2 \gamma_1
  \]
  Multiplying instead with \( y_{t-2} \) yields
  \[
  E[y_t y_{t-2}] = \theta_1 E[y_{t-1}y_{t-2}] + \theta_2 E[y_{t-2}y_{t-2}] + E[\epsilon_t y_{t-2}]
  \gamma_2 = \theta_1 \gamma_1 + \theta_2 \gamma_0
  \]
  Multiplying instead with \( y_{t-3} \) yields
  \[
  E[y_t y_{t-3}] = \theta_1 E[y_{t-1}y_{t-3}] + \theta_2 E[y_{t-2}y_{t-3}] + E[\epsilon_t y_{t-3}]
  \gamma_3 = \theta_1 \gamma_2 + \theta_2 \gamma_1
  \]

- These are the so-called Yule-Walker equations.
• To find the variance we can solve the substitute $\gamma_1$ and $\gamma_2$ into the equation for $\gamma_0$. This is, however, a bit tedious.

• We can find the autocorrelations, $\rho_k = \gamma_k / \gamma_0$, as

$$
\begin{align*}
\rho_1 &= \theta_1 + \theta_2 \rho_1 \\
\rho_2 &= \theta_1 \rho_1 + \theta_2 \\
\rho_k &= \theta_1 \rho_{k-1} + \theta_2 \rho_{k-2}, \quad k \geq 3
\end{align*}
$$

or alternatively that

$$
\begin{align*}
\rho_1 &= \frac{\theta_1}{1 - \theta_2} \\
\rho_2 &= \frac{\theta_1^2}{1 - \theta_2} + \theta_2 \\
\rho_k &= \theta_1 \rho_{k-1} + \theta_2 \rho_{k-2}, \quad k \geq 3.
\end{align*}
$$
Dynamic Properties of a MA(1) Model

• Consider the MA(1) model

\[ Y_t = \mu + \epsilon_t + \alpha \epsilon_{t-1}. \]

• The mean is given by

\[ E[Y_t] = E[\mu + \epsilon_t + \alpha \epsilon_{t-1}] = \mu \]

which is here identical to the constant term.

• Next we find the variance:

\[
V[Y_t] = E[(Y_t - \mu)^2] \\
= E[(\epsilon_t + \alpha \epsilon_{t-1})^2] \\
= E[\epsilon_t^2] + E[\alpha^2 \epsilon_{t-1}^2] + E[2\alpha \epsilon_t \epsilon_{t-1}] \\
= (1 + \alpha^2) \sigma^2.
\]

• The covariances are given by

\[
\gamma_1 = Cov[Y_t, Y_{t-1}] \\
= E[(Y_t - \mu)(Y_{t-1} - \mu)] \\
= E[(\epsilon_t + \alpha \epsilon_{t-1})(\epsilon_{t-1} + \alpha \epsilon_{t-2})] \\
= E[\epsilon_t \epsilon_{t-1} + \alpha \epsilon_t \epsilon_{t-2} + \alpha \epsilon_{t-1}^2 + \alpha^2 \epsilon_{t-1} \epsilon_{t-2}] \\
= \alpha \sigma^2.
\]

\[
\gamma_2 = Cov[Y_t, Y_{t-2}] \\
= E[(Y_t - \mu)(Y_{t-2} - \mu)] \\
= E[(\epsilon_t + \alpha \epsilon_{t-1})(\epsilon_{t-2} + \alpha \epsilon_{t-3})] \\
= E[\epsilon_t \epsilon_{t-2} + \alpha \epsilon_t \epsilon_{t-3} + \alpha \epsilon_{t-1} \epsilon_{t-2} + \alpha^2 \epsilon_{t-1} \epsilon_{t-3}] \\
= 0.
\]

\[
\gamma_k = 0 \quad \text{for} \quad k = 3, 4, \ldots
\]

• The ACF is given by

\[
\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\alpha \sigma^2}{(1 + \alpha^2) \sigma^2} = \frac{\alpha}{(1 + \alpha^2)}
\]

\[
\rho_k = 0, \quad k \geq 2.
\]
The Lag– and Difference Operators

• Now we introduce an important tool called the lag-operator, $L$. It has the property that

$$L \cdot y_t = y_{t-1},$$

and, for example,

$$L^2 y_t = L(Ly_t) = Ly_{t-1} = y_{t-2}.$$ 

• Also define the first difference operator, $\Delta = 1 - L$, such that

$$\Delta y_t = (1 - L) y_t = y_t - Ly_t = y_t - y_{t-1}.$$ 

• The operators $L$ and $\Delta$ are not functions, but can be used in calculations.
Lag Polynomials

- Consider as an example the AR(2) model

\[ y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \epsilon_t. \]

That can be written as

\[ y_t - \theta_1 y_{t-1} - \theta_2 y_{t-2} = \epsilon_t \]
\[ y_t - \theta_1 L y_t - \theta_2 L^2 y_t = \epsilon_t \]
\[ (1 - \theta_1 L - \theta_2 L^2)y_t = \epsilon_t \]
\[ \theta(L)y_t = \epsilon_t, \]

where

\[ \theta(L) = 1 - \theta_1 L - \theta_2 L^2 \]

is a polynomial in \( L \), denoted a lag-polynomial.

- Standard rules for calculating with polynomials also hold for polynomials in \( L \).

Characteristic Equations and Roots

- For a model

\[ y_t - \theta_1 y_{t-1} - \theta_2 y_{t-2} = \epsilon_t \]
\[ \theta(L)y_t = \epsilon_t, \]

we define the characteristic equation as

\[ \theta(z) = 1 - \theta_1 z - \theta_2 z^2 = 0. \]

The solutions, \( z_1 \) and \( z_2 \), are denoted characteristic roots.

- An AR(p) has \( p \) roots.
  Some of them may be complex values, \( h \pm v \cdot i \), where \( i = \sqrt{-1} \).

- Recall, that the roots can be used for factorizing the polynomial

\[ \theta(z) = 1 - \theta_1 z - \theta_2 z^2 = (1 - \phi_1 z)(1 - \phi_2 z), \]

where \( \phi_1 = z_1^{-1} \) and \( \phi_2 = z_2^{-1} \) are the inverse roots.
Invertibility of Polynomials

- Define the inverse of a polynomial, $\theta^{-1}(L)$ of $\theta(L)$, so that
  \[ \theta^{-1}(L)\theta(L) = 1. \]

- Consider the AR(1) case, $\theta(L) = 1 - \theta L$, and look at the product
  \[
  (1 - \theta L) \left( 1 + \theta L + \theta^2 L^2 + \theta^3 L^3 + \ldots + \theta^k L^k \right)
  = (1 - \theta L) + (\theta L - \theta^2 L^2) + (\theta^2 L^2 - \theta^3 L^3) + (\theta^3 L^3 - \theta^4 L^5) + \ldots
  = 1 - \theta^{k+1} L^{k+1}.
  \]
  If $|\theta| < 1$, it holds that $\theta^{k+1} L^{k+1} \to 0$ as $k \to \infty$ implying that
  \[
  \theta^{-1}(L) = (1 - \theta L)^{-1} = \frac{1}{1 - \theta L} = 1 + \theta L + \theta^2 L^2 + \theta^3 L^3 + \ldots = \sum_{i=0}^{\infty} \theta^i L^i.
  \]

- If $\theta(L)$ is a finite polynomial, the inverse polynomial, $\theta^{-1}(L)$, is infinite.

ARMA Models in AR and MA form

- Using lag polynomials we can rewrite the stationary ARMA(p,q) model as
  \[
  y_t - \theta_1 y_{t-1} - \ldots - \theta_p y_{t-p} = \epsilon_t + \alpha_1 \epsilon_{t-1} + \ldots + \alpha_q \epsilon_{t-q}
  \]
  \[
  \theta(L)y_t = \alpha(L)\epsilon_t. \quad \text{(*)}
  \]
  where $\theta(L)$ and $\alpha(L)$ are finite polynomials.

- If $\theta(L)$ is invertible, (*) can be written as the infinite MA($\infty$) model
  \[
  y_t = \theta^{-1}(L)\alpha(L)\epsilon_t
  \]
  \[
  y_t = \epsilon_t + \gamma_1 \epsilon_{t-1} + \gamma_2 \epsilon_{t-2} + \ldots
  \]
  This is called the MA representation.

- If $\alpha(L)$ is invertible, (*) can be written as an infinite AR($\infty$) model
  \[
  \alpha^{-1}(L)\theta(L)y_t = \epsilon_t
  \]
  \[
  y_t - \gamma_1 y_{t-1} - \gamma_2 y_{t-2} - \ldots = \epsilon_t.
  \]
  This is called the AR representation.
Invertibility and Stationarity

• A finite order MA process is stationary by construction.
  – It is a linear combination of stationary white noise terms.
  – Invertibility is sometimes convenient for estimation and prediction.

• An infinite MA process is stationary if the coefficients, $\alpha_i$, converge to zero.
  – We require that $\sum_{i=1}^{\infty} \alpha_i^2 < \infty$.

• An AR process is stationary if $\theta(L)$ is invertible.
  – This is important for interpretation and inference.
  – In the case of a root at unity standard results no longer hold.
    We return to unit roots later.

• Consider again the AR(2) model
  $$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 = (1 - \phi_1 L) (1 - \phi_2 L).$$
  The polynomial is invertible if the factors $(1 - \phi_i L)$ are invertible, i.e. if
  $$|\phi_1| < 1 \quad \text{and} \quad |\phi_2| < 1.$$

• In general a polynomial, $\theta(L)$, is invertible if the characteristic roots, $z_1, \ldots, z_p$, are larger than one in absolute value.
  In complex cases, this corresponds to the roots being outside the complex unit circle.
  (Modulus larger than one).
ARMA Models and Common Roots

- Consider the stationary ARMA(p,q) model
  \[ y_t - \theta_1 y_{t-1} - \ldots - \theta_p y_{t-p} = \epsilon_t + \alpha_1 \epsilon_{t-1} + \ldots + \alpha_q \epsilon_{t-q} \]
  \[ \theta(L) y_t = \alpha(L) \epsilon_t \]
  \[ (1 - \phi_1 L) (1 - \phi_2 L) \cdots (1 - \phi_p L) y_t = (1 - \xi_1 L) (1 - \xi_2 L) \cdots (1 - \xi_q L) \epsilon_t. \]

- If \( \phi_i = \xi_j \) for some \( i, j \), they are denoted common roots or canceling roots. The ARMA(p,q) model is equivalent to a ARMA(p-1,q-1) model.

- As an example, consider
  \[ y_t - y_{t-1} + 0.25 y_{t-2} = \epsilon_t - 0.5 \epsilon_{t-1} \]
  \[ (1 - L + 0.25 L^2) y_t = (1 - 0.5 L) \epsilon_t \]
  \[ (1 - 0.5 L) (1 - 0.5 L) y_t = (1 - 0.5 L) \epsilon_t \]
  \[ (1 - 0.5 L) y_t = \epsilon_t. \]

Unit Roots and ARIMA Models

- A root at one is denoted a unit root, and has important consequences for the analysis. We consider tests for unit roots and unit root econometrics later.

- Consider an ARMA(p,q) model
  \[ \theta(L) y_t = \alpha(L) \epsilon_t. \]
  If there is a unit root in the AR polynomial, we can factorize into
  \[ \theta(L) = (1 - L) (1 - \phi_2 L) \cdots (1 - \phi_p L) = (1 - L) \theta^*(L), \]
  and we can write the model as
  \[ \theta^*(L) (1 - L) y_t = \alpha(L) \epsilon_t \]
  \[ \theta^*(L) \Delta y_t = \alpha(L) \epsilon_t. \]

- An ARMA(p,q) model for \( \Delta^d y_t \) is denoted an ARIMA(p,d,q) model for \( y_t \).
Example: Danish Real House Prices

- Consider the Danish real house prices in logs, $p_t$. An AR(2) model yields
  \[
  p_t = 1.551\, p_{t-1} - 0.5734\, p_{t-2} + 0.003426
  \]
  \[(20.7) \quad (-7.56) \quad (1.30)\]
  
  The lag polynomial is given by
  \[
  \theta(L) = 1 - 1.551 \cdot L + 0.5734 \cdot L^2,
  \]
  with inverse roots given by 0.9422 and 0.6086.

- One root is close to unity and we estimate an ARIMA(2,1,0) model for $p_t$:
  \[
  \Delta p_t = 1.323\, \Delta p_{t-1} - 0.4853\, \Delta p_{t-2} + 0.0009959
  \]
  \[(16.6) \quad (-6.12) \quad (0.333)\]
  
  The lag polynomial is given by
  \[
  \theta(L) = 1 - 1.323 \cdot L + 0.4853 \cdot L^2,
  \]
  with complex (inverse) roots given by
  \[
  0.66140 \pm 0.21879 \cdot i, \quad \text{where} \quad i = \sqrt{-1}.
  \]

ARIMA($p,d,q$) Model Selection

- Find a transformation of the process that is stationary, e.g. $\Delta^d Y_t$.

- Recall, that for the stationary AR($p$) model
  - The ACF is infinite but convergent.
  - The PACF is zero for lags larger than $p$.

- For the MA($q$) model
  - The ACF is zero for lags larger than $q$.
  - The PACF is infinite but convergent.

- The ACF and PACF contains information $p$ and $q$.
  Can be used to select relevant models.
• If alternative models are nested, they can be tested.

• Model selection can be based on information criteria

\[
IC = \log \hat{\sigma}^2 + \text{penalty}(T, \#\text{parameters})
\]

Measures the likelihood + penalty

A penalty for the number of parameters

The information criteria should be minimized!

• Three important criteria

\[
AIC = \log \hat{\sigma}^2 + \frac{2 \cdot k}{T}
\]

\[
HQ = \log \hat{\sigma}^2 + \frac{2 \cdot k \cdot \log(\log(T))}{T}
\]

\[
BIC = \log \hat{\sigma}^2 + \frac{k \cdot \log(T)}{T},
\]

where \( k \) is the number of estimated parameters, e.g. \( k = p + q \).

Example: Consumption-Income Ratio

(A) Consumption and income, logs.

(B) Consumption-Income ratio, logs.

(C) ACF for series in (B)

(D) PACF for series in (B)
<table>
<thead>
<tr>
<th>Model</th>
<th>T</th>
<th>p</th>
<th>log-lik</th>
<th>SC</th>
<th>HQ</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMA(2,2)</td>
<td>130</td>
<td>5</td>
<td>300.82151</td>
<td>-4.4408</td>
<td>-4.5063</td>
<td>-4.5511</td>
</tr>
<tr>
<td>ARMA(2,1)</td>
<td>130</td>
<td>4</td>
<td>300.39537</td>
<td>-4.4717</td>
<td>-4.5241</td>
<td>-4.5599</td>
</tr>
<tr>
<td>ARMA(2,0)</td>
<td>130</td>
<td>3</td>
<td>300.38908</td>
<td>-4.5090</td>
<td>-4.5483</td>
<td>-4.5752</td>
</tr>
<tr>
<td>ARMA(1,2)</td>
<td>130</td>
<td>4</td>
<td>300.42756</td>
<td>-4.4722</td>
<td>-4.5246</td>
<td>-4.5604</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>130</td>
<td>3</td>
<td>299.99333</td>
<td>-4.5030</td>
<td>-4.5422</td>
<td>-4.5691</td>
</tr>
<tr>
<td>ARMA(1,0)</td>
<td>130</td>
<td>2</td>
<td>296.17449</td>
<td>-4.4816</td>
<td>-4.5078</td>
<td>-4.5258</td>
</tr>
<tr>
<td>ARMA(0,0)</td>
<td>130</td>
<td>1</td>
<td>249.82604</td>
<td>-3.8060</td>
<td>-3.8191</td>
<td>-3.8281</td>
</tr>
</tbody>
</table>

---- Maximum likelihood estimation of ARFIMA(1,0,1) model ----
The estimation sample is: 1971 (1) - 2003 (2)
The dependent variable is: cy (ConsumptionData.in7)

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR-1</td>
<td>0.857361</td>
<td>0.05650</td>
<td>15.2</td>
</tr>
<tr>
<td>MA-1</td>
<td>-0.300821</td>
<td>0.09825</td>
<td>-3.06</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.0934110</td>
<td>0.009898</td>
<td>-9.44</td>
</tr>
</tbody>
</table>

log-likelihood 299.993327
sigma 0.0239986 sigma^2 0.000575934

---- Maximum likelihood estimation of ARFIMA(2,0,0) model ----
The estimation sample is: 1971 (1) - 2003 (2)
The dependent variable is: cy (ConsumptionData.in7)

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR-1</td>
<td>0.536183</td>
<td>0.08428</td>
<td>6.36</td>
</tr>
<tr>
<td>AR-2</td>
<td>0.250548</td>
<td>0.08479</td>
<td>2.95</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.0935407</td>
<td>0.009481</td>
<td>-9.87</td>
</tr>
</tbody>
</table>

log-likelihood 300.389084
sigma 0.0239238 sigma^2 0.000572349
Estimation of ARMA Models

- The natural estimator is **maximum likelihood**. With normal errors

\[
\log L(\theta, \alpha, \sigma^2) = -\frac{T}{2} \log(2\pi\sigma^2) - \sum_{t=1}^{T} \frac{\epsilon_t^2}{2\sigma^2},
\]

where \( \epsilon_t \) is the residual.

- For an AR(1) model we can write the residual as

\[
\epsilon_t = Y_t - \delta - \theta Y_{t-1},
\]

and OLS coincides with ML.

- Usual to condition on the initial values. Alternatively we can postulate a distribution for the first observation, e.g.

\[
Y_1 \sim N\left(\frac{\delta}{1-\theta}, \frac{\sigma^2}{1-\theta^2}\right),
\]

where the mean and variance are chosen as implied by the model for the rest of the observations. We say that \( Y_1 \) is chosen from the invariant distribution.

- For the MA(1) model

\[
Y_t = \mu + \epsilon_t + \alpha \epsilon_{t-1},
\]

the residuals can be found recursively as a function of the parameters

\[
\begin{align*}
\epsilon_1 &= Y_1 - \mu \\
\epsilon_2 &= Y_2 - \mu - \alpha \epsilon_1 \\
\epsilon_3 &= Y_3 - \mu - \alpha \epsilon_2 \\
&\vdots
\end{align*}
\]

Here, the initial value is \( \epsilon_0 = 0 \), but that could be relaxed if required by using the invariant distribution.

- The likelihood function can be maximized wrt. \( \alpha \) and \( \mu \).
Forecasting

• Easy to forecast with ARMA models. Main drawback is that here is no economic insight.

• We want to predict $y_{T+k}$ given all information up to time $T$, i.e. given the information set

$$I_T = \{y_{-\infty}, ..., y_{T-1}, y_T\}.$$  

The optimal predictor is the conditional expectation

$$y_{T+k|T} = E[y_{T+k} | I_T].$$

• Consider the ARMA(1,1) model

$$y_t = \theta \cdot y_{t-1} + \epsilon_t + \alpha \epsilon_{t-1}, \quad t = 1, 2, ..., T.$$

• To forecast we
  – Substitute the estimated parameters for the true.
  – Use estimated residuals up to time $T$. Hereafter, the best forecast is zero.

• The optimal forecasts will be

$$y_{T+1|T} = E[\theta \cdot y_T + \epsilon_{T+1} + \alpha \epsilon_T | I_T]$$  
  $$= \hat{\theta} \cdot y_T + \hat{\alpha} \cdot \epsilon_T$$  

$$y_{T+2|T} = E[\theta \cdot y_{T+1} + \epsilon_{T+2} + \alpha \cdot \epsilon_{T+1} | I_T]$$  
  $$= \hat{\theta} \cdot y_{T+1|T}. $$