Econometrics 2 – Fall 2005

Univariate Time Series Analysis; ARIMA Models

Heino Bohn Nielsen

Univariate Time Series Analysis

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

- Univariate models are useful for:
  (1) Analyzing the dynamic properties of 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 $E[y_t \mid x_t]$ 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.

Outline of the Lecture

(1) Characterizing time dependence: ACF and PACF.

(2) Modelling time dependence: the ARMA(p,q) model

(3) Examples:

- AR(1).
- AR(2).
- MA(1).

(4) Lag operators, lag polynomials and invertibility.

(5) Model selection.

(6) Estimation.

(7) Forecasting.

Characterizing Time Dependence

- For a stationary time series the autocorrelation function (ACF) is

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

An alternative measure is the partial autocorrelation function (PACF), which is the conditional correlation:.

$$\theta_k = \text{Corr}(y_t, y_{t-k} \mid y_{t-1}, ..., y_{t-k+1}).$$

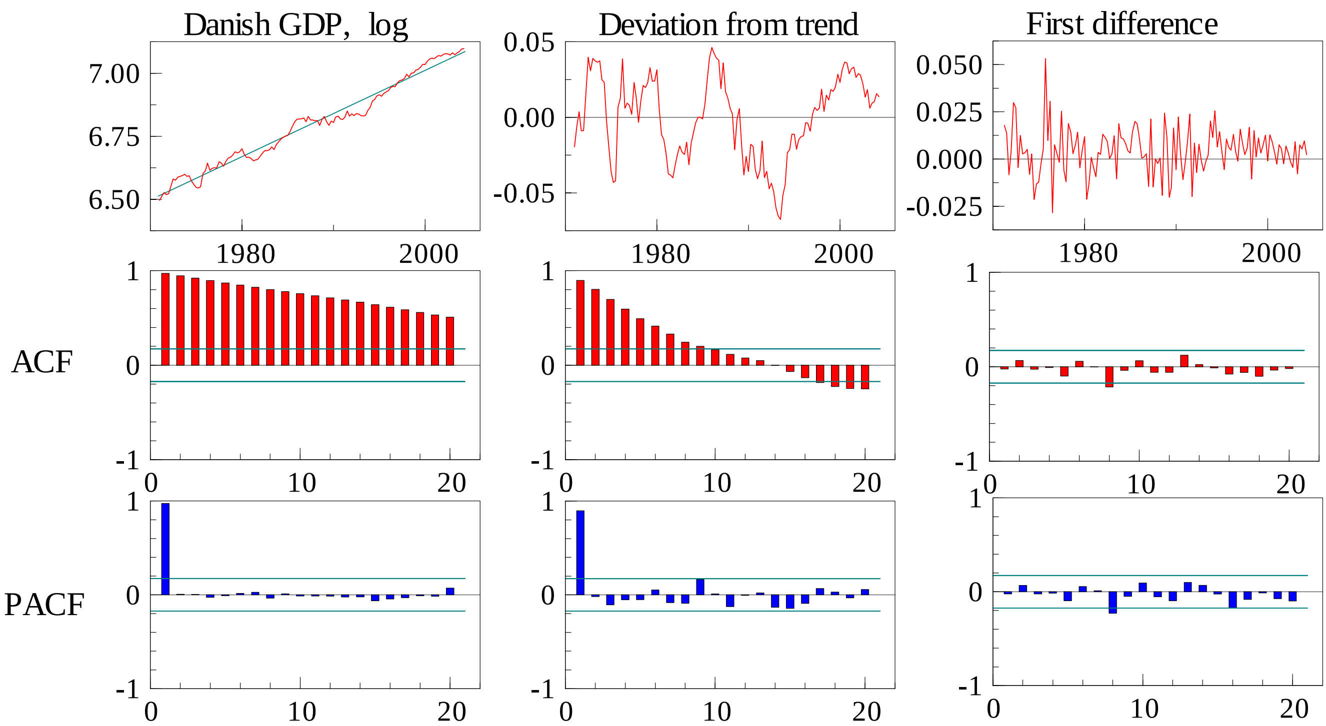
Note: ACF and PACF are bounded in $[-1; 1]$, symmetric $\rho_k = \rho_{-k}$ and $\rho_k = \theta_0 = 1$.

- Simple estimators, $\widehat{\rho}_k$ and $\widehat{\theta}_k$, can be derived from OLS regressions

$$\begin{array}{lcccccccccc} \text{ACF:} & y_t & = & c & + & & & & & \rho_k y_{t-k} & + & \text{residual} \\ \text{PACF:} & y_t & = & c & + & \theta_1 y_{t-1} & + & ... & + & \theta_k y_{t-k} & + & \text{residual} \end{array}$$

- For an IID time series it hold that $V(\widehat{\rho}_k) = V(\widehat{\theta}_k) = T^{-1}$, and a 95% confidence band is given by $\pm 2/\sqrt{T}$.

Example: Danish GDP



The ARMA(p,q) Model

- First define a white noise process, $\epsilon_t \sim IID(0, \sigma^2)$.

- The autoregressive AR(p) model is defined as

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + ... + \theta_p y_{t-p} + \epsilon_t.$$

  Systematic part of y_t is a linear function of p lagged values.
  We need p (observed) initial values: $y_{-(p-1)}, y_{-(p-2)}, ..., y_{-1}, y_0$.

- The moving average MA(q) model is defined as

$$y_t = \epsilon_t + \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2} + ... + \alpha_q \epsilon_{t-q}.$$

  y_t is a moving average of past shocks to the process.
  We need q initial values: $\epsilon_{-(p-1)} = \epsilon_{-(p-2)} = ... = \epsilon_{-1} = \epsilon_0 = 0$.

- They can be combined into the ARMA(p,q) model

$$y_t = \theta_1 y_{t-1} + ... + \theta_p y_{t-p} + \epsilon_t + \alpha_1 \epsilon_{t-1} + ... + \alpha_q \epsilon_{t-q}.$$

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

$$\begin{aligned} E[Y_t] &= E[\delta + \theta Y_{t-1} + \epsilon_t] \\ E[Y_t] &= \delta + \theta E[Y_{t-1}] + E[\epsilon_t] \\ (1-\theta)\,\mu &= \delta \\ \mu &= \frac{\delta}{1-\theta}. \end{aligned}$$

Note the following:
(1) The effect of the constant term, δ, depends on the autoregressive parameter, θ.
(2) μ 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

$$\begin{aligned} Y_t &= \delta + \theta Y_{t-1} + \epsilon_t \\ Y_t &= (1-\theta)\,\mu + \theta Y_{t-1} + \epsilon_t \\ Y_t - \mu &= \theta\,(Y_{t-1} - \mu) + \epsilon_t \\ y_t &= \theta y_{t-1} + \epsilon_t. \end{aligned}$$

- We note that $\gamma_0 = V[Y_t] = V[y_t]$. We find:

$$\begin{aligned} V[y_t] &= E[y_t^2] \\ &= E[(\theta y_{t-1} + \epsilon_t)^2] \\ &= E[\theta^2 y_{t-1}^2 + \epsilon_t^2 + 2\theta y_{t-1}\epsilon_t] \\ &= \theta^2 E[y_{t-1}^2] + E[\epsilon_t^2] + 2\theta E[y_{t-1}\epsilon_t] \\ &= \theta^2 V[y_{t-1}] + \sigma^2 + 0. \end{aligned}$$

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, $Cov[y_t, y_{t-k}] = E[y_t y_{t-k}]$, are given by

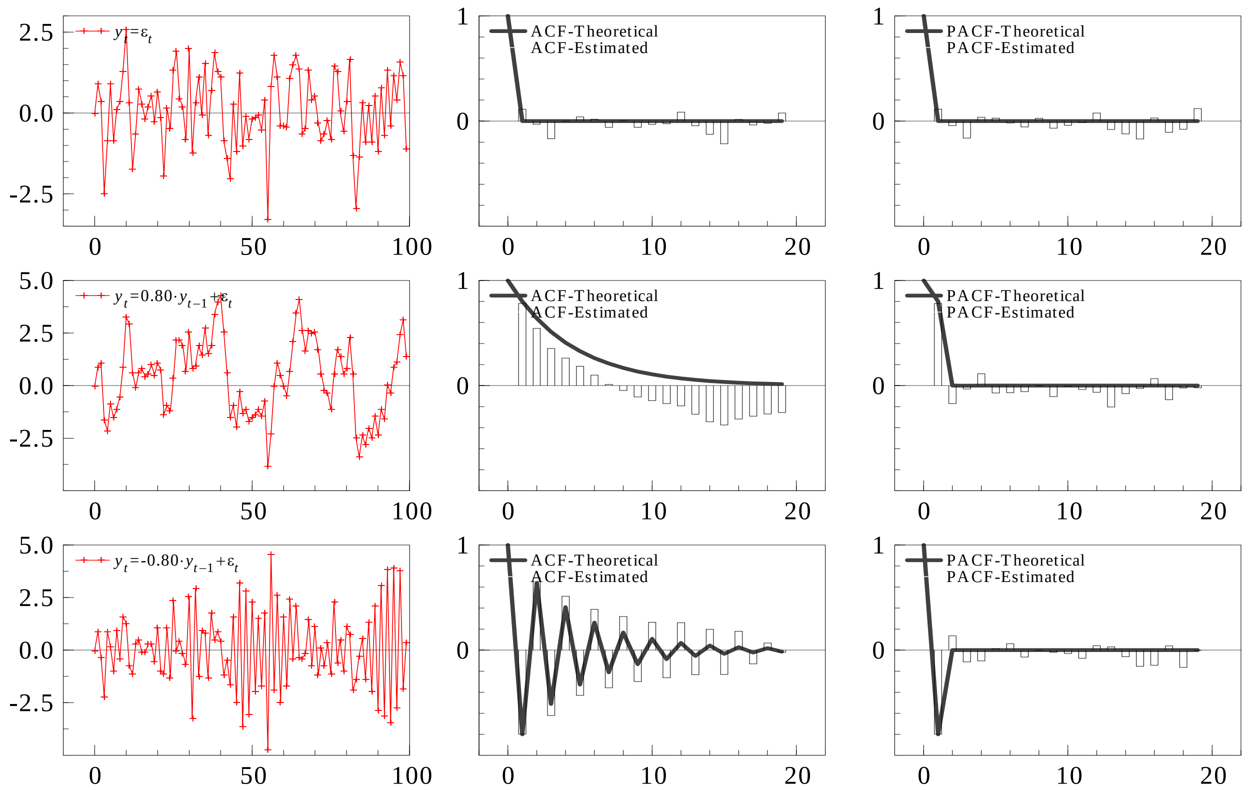
$$
\begin{aligned}
\gamma_1 &= E[y_t y_{t-1}] = E[(\theta y_{t-1} + \epsilon_t) y_{t-1}] = \theta E[y_{t-1}^2] + E[y_{t-1}\epsilon_t] &= \theta\gamma_0 = \theta\frac{\sigma^2}{1-\theta^2} \\
\gamma_2 &= E[y_t y_{t-2}] = E[(\theta y_{t-1} + \epsilon_t)\, y_{t-2}] = \theta E[y_{t-1} y_{t-2}] + E[\epsilon_t y_{t-2}] &= \theta\gamma_1 = \theta^2\frac{\sigma^2}{1-\theta^2} \\
&\vdots \\
\gamma_k &= E[y_t y_{t-k}] &= \theta^k \gamma_0
\end{aligned}
$$

- The ACF is given by

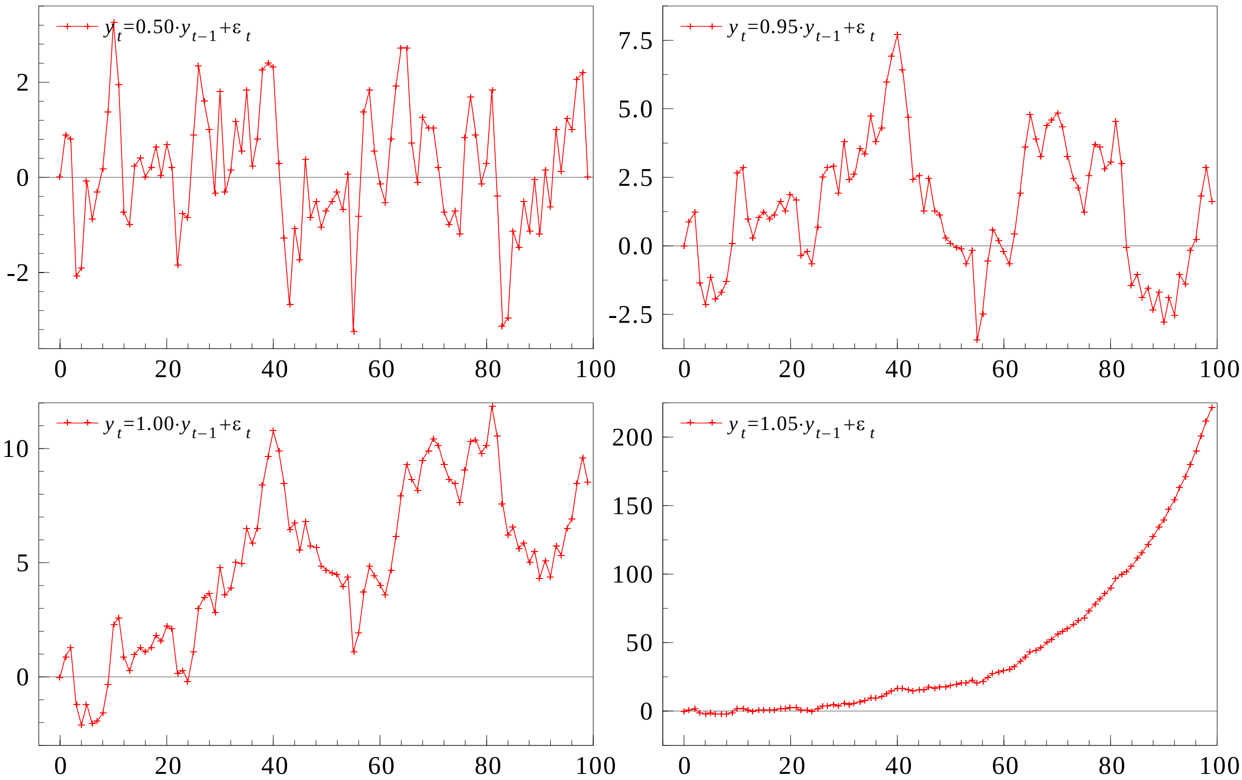
$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\theta^k \gamma_0}{\gamma_0} = \theta^k.$$

- The PACF is simply the autoregressive coefficients: $\theta_1, 0, 0, \ldots$

Examples of Stationary AR(1) Models



Examples of AR(1) Models



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:

$$\begin{aligned} E[Y_t] &= \delta + \theta_1 E[Y_{t-1}] + \theta_2 E[Y_{t-2}] + E[\epsilon_t] \\ E[Y_t] &= \frac{\delta}{1-\theta_1-\theta_2} = \mu. \end{aligned}$$

- 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

$$\begin{aligned} E\left[y_t^2\right] &= \theta_1 E\left[y_{t-1}y_t\right] + \theta_2 E\left[y_{t-2}y_t\right] + E\left[\epsilon_t y_t\right] \\ \gamma_0 &= \theta_1\gamma_1 + \theta_2\gamma_2 + \sigma^2 \end{aligned}$$

Multiplying instead with y_{t-1} yields

$$\begin{aligned} E\left[y_t y_{t-1}\right] &= \theta_1 E\left[y_{t-1}y_{t-1}\right] + \theta_2 E\left[y_{t-2}y_{t-1}\right] + E\left[\epsilon_t y_{t-1}\right] \\ \gamma_1 &= \theta_1\gamma_0 + \theta_2\gamma_1 \end{aligned}$$

Multiplying instead with y_{t-2} yields

$$\begin{aligned} E\left[y_t y_{t-2}\right] &= \theta_1 E\left[y_{t-1}y_{t-2}\right] + \theta_2 E\left[y_{t-2}y_{t-2}\right] + E\left[\epsilon_t y_{t-2}\right] \\ \gamma_2 &= \theta_1\gamma_1 + \theta_2\gamma_0 \end{aligned}$$

Multiplying instead with y_{t-3} yields

$$\begin{aligned} E\left[y_t y_{t-3}\right] &= \theta_1 E\left[y_{t-1}y_{t-3}\right] + \theta_2 E\left[y_{t-2}y_{t-3}\right] + E\left[\epsilon_t y_{t-3}\right] \\ \gamma_3 &= \theta_1\gamma_2 + \theta_2\gamma_1 \end{aligned}$$

- These are the so-called Yule-Walker equations.

- To find the variance we can substitute γ_1 and γ_2 into the equation for γ_0. This is, however, a bit tedious.

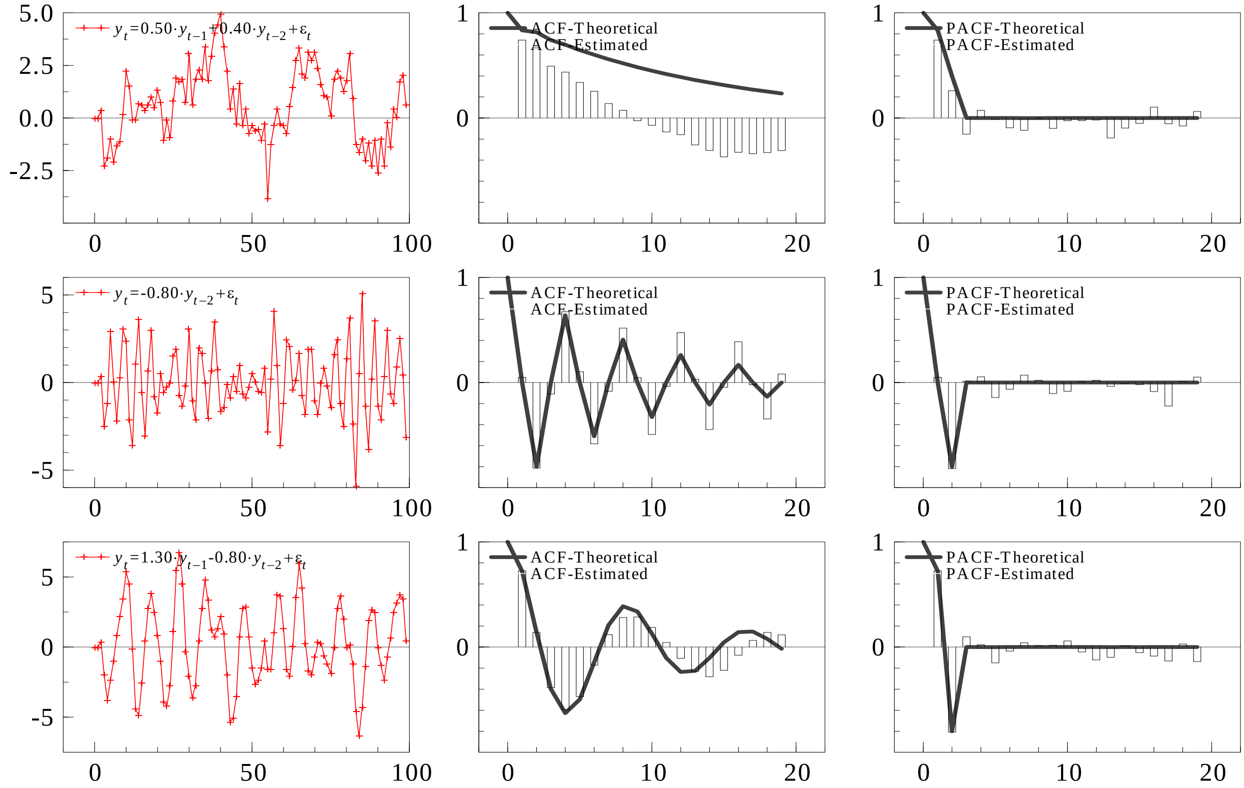
- We can find the autocorrelations, $\rho_k = \gamma_k/\gamma_0$, as

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

or alternatively that

$$\begin{aligned} \rho_1 &= \frac{\theta_1}{1-\theta_2} \\ \rho_2 &= \frac{\theta_1^2}{1-\theta_2} + \theta_2 \\ \rho_k &= \theta_1\rho_{k-1} + \theta_2\rho_{k-2}, \quad k \geq 3. \end{aligned}$$

Examples of AR(2) Models



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.

- Define the deviation from mean: $y_t = Y_t - \mu$.

- Next we find the variance:

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

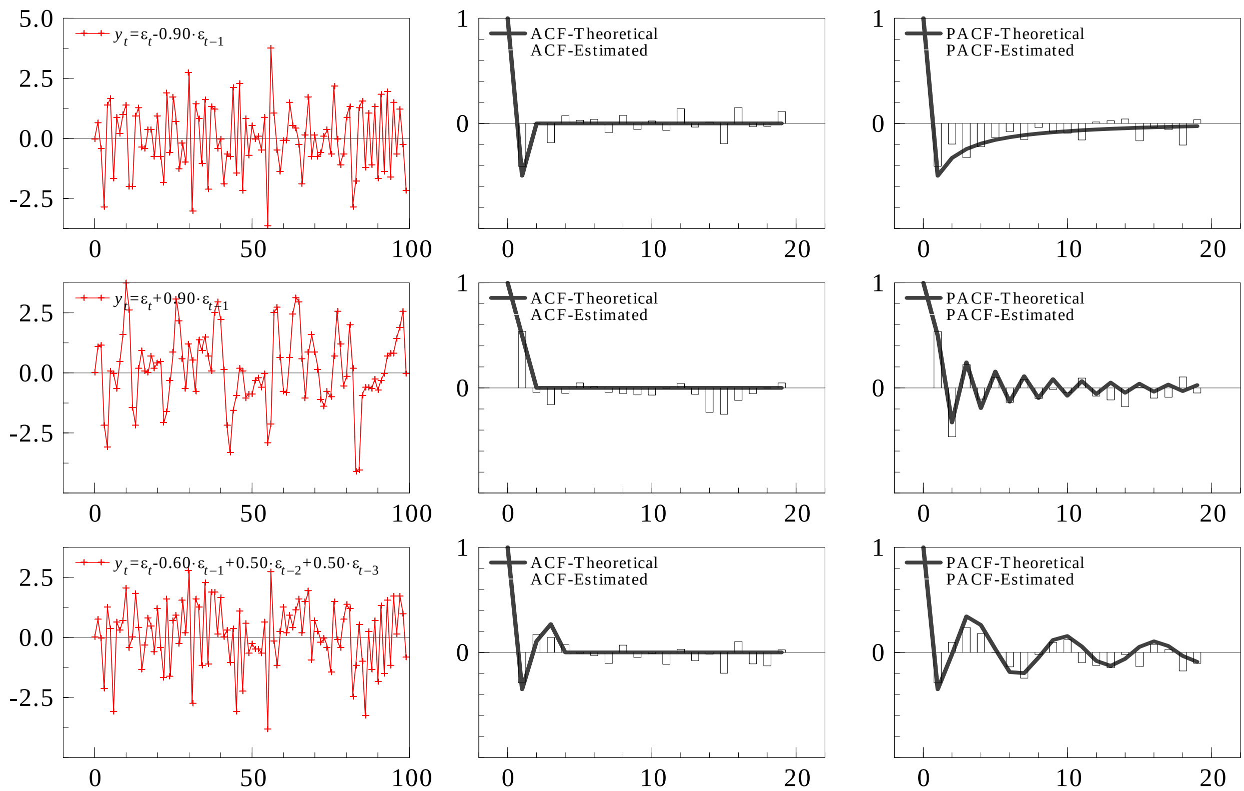
- The covariances, $Cov[y_t, y_{t-k}] = E[y_t y_{t-k}]$, are given by

$$\begin{aligned}
\gamma_1 &= E[y_t y_{t-1}] \\
&= E[(\epsilon_t + \alpha\epsilon_{t-1})(\epsilon_{t-1} + \alpha\epsilon_{t-2})] \\
&= E[\epsilon_t\epsilon_{t-1} + \alpha\epsilon_t\epsilon_{t-2} + \alpha\epsilon_{t-1}^2 + \alpha^2\epsilon_{t-1}\epsilon_{t-2}] = \alpha\sigma^2 \\
\gamma_2 &= E[y_t y_{t-2}] \\
&= E[(\epsilon_t + \alpha\epsilon_{t-1})(\epsilon_{t-2} + \alpha\epsilon_{t-3})] \\
&= E[\epsilon_t\epsilon_{t-2} + \alpha\epsilon_t\epsilon_{t-3} + \alpha\epsilon_{t-1}\epsilon_{t-2} + \alpha^2\epsilon_{t-1}\epsilon_{t-3}] = 0 \\
&\vdots \\
\gamma_k &= E[y_t y_{t-k}] = 0
\end{aligned}$$

- The ACF is given by

$$\begin{aligned}
\rho_1 &= \frac{\gamma_1}{\gamma_0} = \frac{\alpha\sigma^2}{(1+\alpha^2)\,\sigma^2} = \frac{\alpha}{(1+\alpha^2)} \\
\rho_k &= 0, \quad k \geq 2.
\end{aligned}$$

Examples of MA Models



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

- Also define the first difference operator, $\Delta = 1 - L$, such that
  $$\Delta y_t = (1 - L)\, y_t = y_t - L y_t = y_t - y_{t-1}.$$

- The operators L and Δ 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
  $$\begin{aligned} y_t - \theta_1 y_{t-1} - \theta_2 y_{t-2} &= \epsilon_t \\ y_t - \theta_1 L y_t - \theta_2 L^2 y_t &= \epsilon_t \\ (1 - \theta_1 L - \theta_2 L^2) y_t &= \epsilon_t \\ \theta(L) y_t &= \epsilon_t, \end{aligned}$$
  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

$$\begin{aligned} y_t - \theta_1 y_{t-1} - \theta_2 y_{t-2} &= \epsilon_t \\ \theta(L) y_t &= \epsilon_t, \end{aligned}$$

  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

$$\begin{aligned} & (1 - \theta L)\left(1 + \theta L + \theta^2 L^2 + \theta^3 L^3 + \dots + \theta^k L^k\right) \\ = \; & (1 - \theta L) + \left(\theta L - \theta^2 L^2\right) + \left(\theta^2 L^2 - \theta^3 L^3\right) + \left(\theta^3 L^3 - \theta^4 L^4\right) + \dots \\ = \; & 1 - \theta^{k+1} L^{k+1}. \end{aligned}$$

  If $|\theta| < 1$, it holds that $\theta^{k+1} L^{k+1} \to 0$ as $k \to \infty$ implying that

$$\theta^{-1}(L) = (1 - \theta L)^{-1} = \frac{1}{1 - \theta L} = 1 + \theta L + \theta^2 L^2 + \theta^3 L^3 + \dots = \sum_{i=0}^{\infty} \theta^i L^i.$$

- 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

$$\begin{aligned} y_t - \theta_1 y_{t-1} - ... - \theta_p y_{t-p} &= \epsilon_t + \alpha_1 \epsilon_{t-1} + ... + \alpha_q \epsilon_{t-q} \qquad (*) \\ \theta(L) y_t &= \alpha(L)\epsilon_t. \end{aligned}$$

  where $\theta(L)$ and $\alpha(L)$ are finite polynomials.

- If $\theta(L)$ is invertible, $(*)$ can be written as the infinite MA(∞) model

$$\begin{aligned} y_t &= \theta^{-1}(L)\alpha(L)\epsilon_t \\ y_t &= \epsilon_t + \gamma_1 \epsilon_{t-1} + \gamma_2 \epsilon_{t-2} + ... \end{aligned}$$

  This is called the MA representation.

- If $\alpha(L)$ is invertible, $(*)$ can be written as an infinite AR(∞) model

$$\begin{aligned} \alpha^{-1}(L)\theta(L) y_t &= \epsilon_t \\ y_t - \gamma_1 y_{t-1} - \gamma_2 y_{t-2} - ... &= \epsilon_t. \end{aligned}$$

  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, α_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.

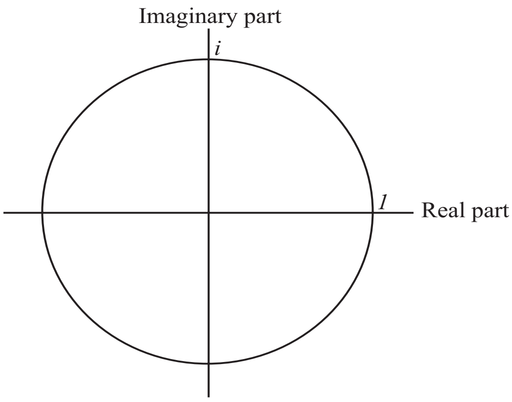
- Consider again the AR(2) model

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 = (1 - \phi_1 L)(1 - \phi_2 L).$$

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, $z_1, ..., z_p$, 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).



Solution to the AR(1) Model

- Consider the model

$$\begin{aligned} Y_t &= \delta + \theta Y_{t-1} + \epsilon_t \\ (1 + \theta L) Y_t &= \delta + \epsilon_t. \end{aligned}$$

The solution is given as

$$\begin{aligned} Y_t &= (1 + \theta L)^{-1} (\delta + \epsilon_t) \\ &= \left(1 + \theta L + \theta^2 L^2 + \theta^3 L^3 + ...\right)(\delta + \epsilon_t) \\ &= \left(1 + \theta + \theta^2 + \theta^3 + ...\right)\delta + \epsilon_t + \theta\epsilon_{t-1} + \theta^2\epsilon_{t-2} + \theta^3\epsilon_{t-3} + ... \end{aligned}$$

- This is the MA-representation. The expectation is given by

$$E[Y_t] = \left(1 + \theta + \theta^2 + \theta^3 + ...\right)\delta \to \frac{\delta}{1-\theta}.$$

- An alternative solution method is recursive subtitution:

$$
\begin{aligned}
Y_t &= \delta + \theta Y_{t-1} + \epsilon_t \\
&= \delta + \theta(\delta + \theta Y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
&= (1+\theta)\,\delta + \epsilon_t + \theta\epsilon_{t-1} + \theta^2 Y_{t-2} \\
&= (1+\theta)\,\delta + \epsilon_t + \theta\epsilon_{t-1} + \theta^2(\delta + \theta Y_{t-3} + \epsilon_{t-2}) \\
&= \left(1+\theta+\theta^2\right)\delta + \epsilon_t + \theta\epsilon_{t-1} + \theta^2\epsilon_{t-2} + \theta^3 Y_{t-3} \\
&\vdots \\
&= \left(1+\theta+\theta^2+\theta^3+\ldots\right)\delta + \epsilon_t + \theta\epsilon_{t-1} + \theta^2\epsilon_{t-2} + \ldots + \theta^{t-1}Y_1
\end{aligned}
$$

where we see the effect of the initial observation.

- The expectation is

$$
E[Y_t] = \left(1+\theta+\theta^2+\theta^3+\ldots\right)\delta + \theta^{t-1}Y_1 \to \frac{\delta}{1-\theta}.
$$

ARMA Models and Common Roots

- Consider the stationary ARMA(p,q) model

$$
\begin{aligned}
y_t - \theta_1 y_{t-1} - \ldots - \theta_p y_{t-p} &= \epsilon_t + \alpha_1\epsilon_{t-1} + \ldots + \alpha_q\epsilon_{t-q} \\
\theta(L)y_t &= \alpha(L)\epsilon_t \\
(1-\phi_1 L)(1-\phi_2 L)\cdots\left(1-\phi_p L\right)y_t &= (1-\xi_1 L)(1-\xi_2 L)\cdots\left(1-\xi_q L\right)\epsilon_t.
\end{aligned}
$$

- 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

$$
\begin{aligned}
y_t - y_{t-1} + 0.25y_{t-2} &= \epsilon_t - 0.5\epsilon_{t-1} \\
\left(1 - L + 0.25L^2\right)y_t &= (1-0.5L)\,\epsilon_t \\
(1-0.5L)(1-0.5L)\,y_t &= (1-0.5L)\,\epsilon_t \\
(1-0.5L)\,y_t &= \epsilon_t.
\end{aligned}
$$

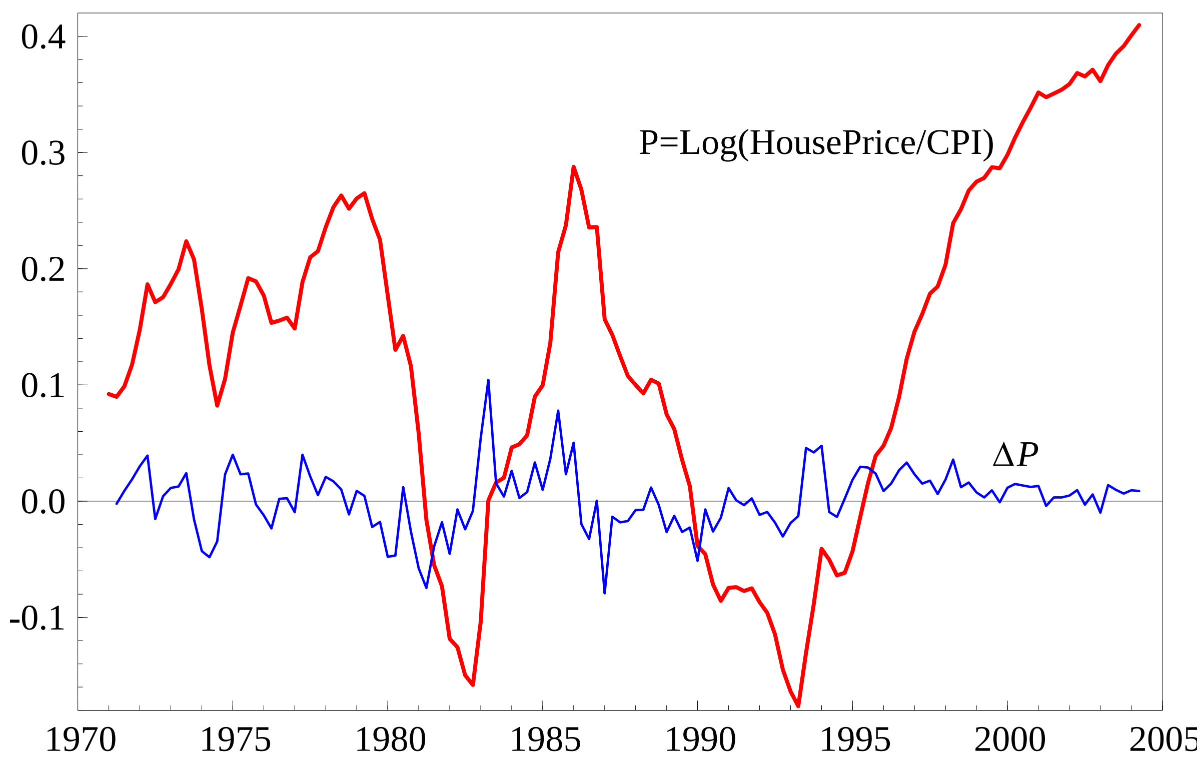
Unit Roots and ARIMA Models

- A root at one is denoted a unit root.
  We consider the consequences later, here we just remove them by first differences.

- 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
$$\theta(L) = (1-L)(1-\phi_2 L)\cdots\left(1-\phi_p L\right) = (1-L)\theta^*(L),$$
  and we can write the model as
$$\begin{aligned}\theta^*(L)(1-L)y_t &= \alpha(L)\epsilon_t \\ \theta^*(L)\Delta y_t &= \alpha(L)\epsilon_t.\end{aligned}$$

- 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



- Estimating an AR(2) model for 1972:1-2004:2 yields

$$p_t \;=\; \underset{(21.0)}{1.545}\, p_{t-1} \;-\; \underset{(-7.58)}{0.5646}\, p_{t-2} \;+\; \underset{(1.29)}{0.003359}$$

The lag polynomial is given by

$$\theta(L) = 1 - 1.545 \cdot L + 0.5646 \cdot L^2,$$

with inverse roots given by 0.953 and 0.592.

- One root is close to unity and we estimate an ARIMA(1,1,0) model for p_t :

$$\Delta p_t \;=\; \underset{(7.35)}{0.5442}\, \Delta p_{t-1} \;+\; \underset{(0.416)}{0.0008369}\ .$$

The second root is basically unchanged.

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

$$IC = \underbrace{\log \widehat{\sigma}^2}_{\text{Measures the likelihood}} + \underbrace{\text{penalty}(T, \#parameters)}_{\text{A penalty for the number of parameters}}$$

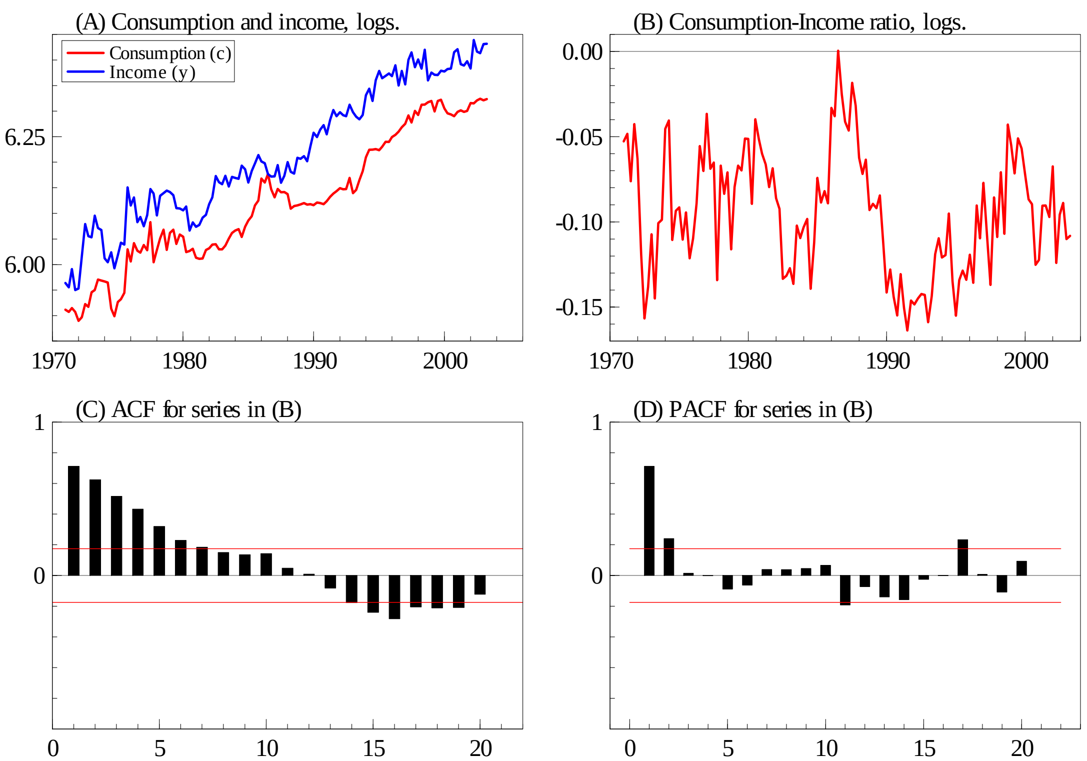
The information criteria should be minimized!

- Three important criteria

$$\begin{aligned} AIC &= \log \widehat{\sigma}^2 + \frac{2 \cdot k}{T} \\ HQ &= \log \widehat{\sigma}^2 + \frac{2 \cdot k \cdot \log(\log(T))}{T} \\ BIC &= \log \widehat{\sigma}^2 + \frac{k \cdot \log(T)}{T}, \end{aligned}$$

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

Example: Consumption-Income Ratio



```
Model        T p   log-lik      SC      HQ     AIC
ARMA(2,2) 130 5 300.82151  -4.4408  -4.5063  -4.5511
ARMA(2,1) 130 4 300.39537  -4.4717  -4.5241  -4.5599
ARMA(2,0) 130 3 300.38908  -4.5090  -4.5483  -4.5752
ARMA(1,2) 130 4 300.42756  -4.4722  -4.5246  -4.5604
ARMA(1,1) 130 3 299.99333  -4.5030  -4.5422  -4.5691
ARMA(1,0) 130 2 296.17449  -4.4816  -4.5078  -4.5258
ARMA(0,0) 130 1 249.82604  -3.8060  -3.8191  -3.8281
```



```
---- Maximum likelihood estimation of ARFIMA(1,0,1) model ----
The estimation sample is:  1971 (1) - 2003 (2)
The dependent variable is:  cy (ConsumptionData.in7)
           Coefficient   Std.Error     t-value     t-prob
AR-1          0.857361     0.05650        15.2      0.000
MA-1         -0.300821     0.09825       -3.06      0.003
Constant     -0.0934110    0.009898      -9.44      0.000
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)
           Coefficient   Std.Error    t-value      t-prob
AR-1          0.536183     0.08428       6.36       0.000
AR-2          0.250548     0.08479       2.95       0.004
Constant     -0.0935407    0.009481     -9.87       0.000
log-likelihood 300.389084
sigma 0.0239238 sigma^2 0.000572349
```

Estimation of ARMA Models

- The natural estimator is maximum likelihood. With normal errors

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

  where ϵ_t is the residual.

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

$$\epsilon_t = Y_t - \delta - \theta_1 \cdot Y_{t-1},$$

  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

$$\begin{aligned} \epsilon_1 &= Y_1 - \mu \\ \epsilon_2 &= Y_2 - \mu - \alpha\epsilon_1 \\ \epsilon_3 &= Y_3 - \mu - \alpha\epsilon_2 \\ &\vdots \end{aligned}$$

  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. α and μ.

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

$$\mathcal{I}_T = \{y_{-\infty}, ..., y_{T-1}, y_T\}.$$

  The optimal predictor is the conditional expectation

$$y_{T+k|T} = E[y_{T+k} \mid \mathcal{I}_T].$$

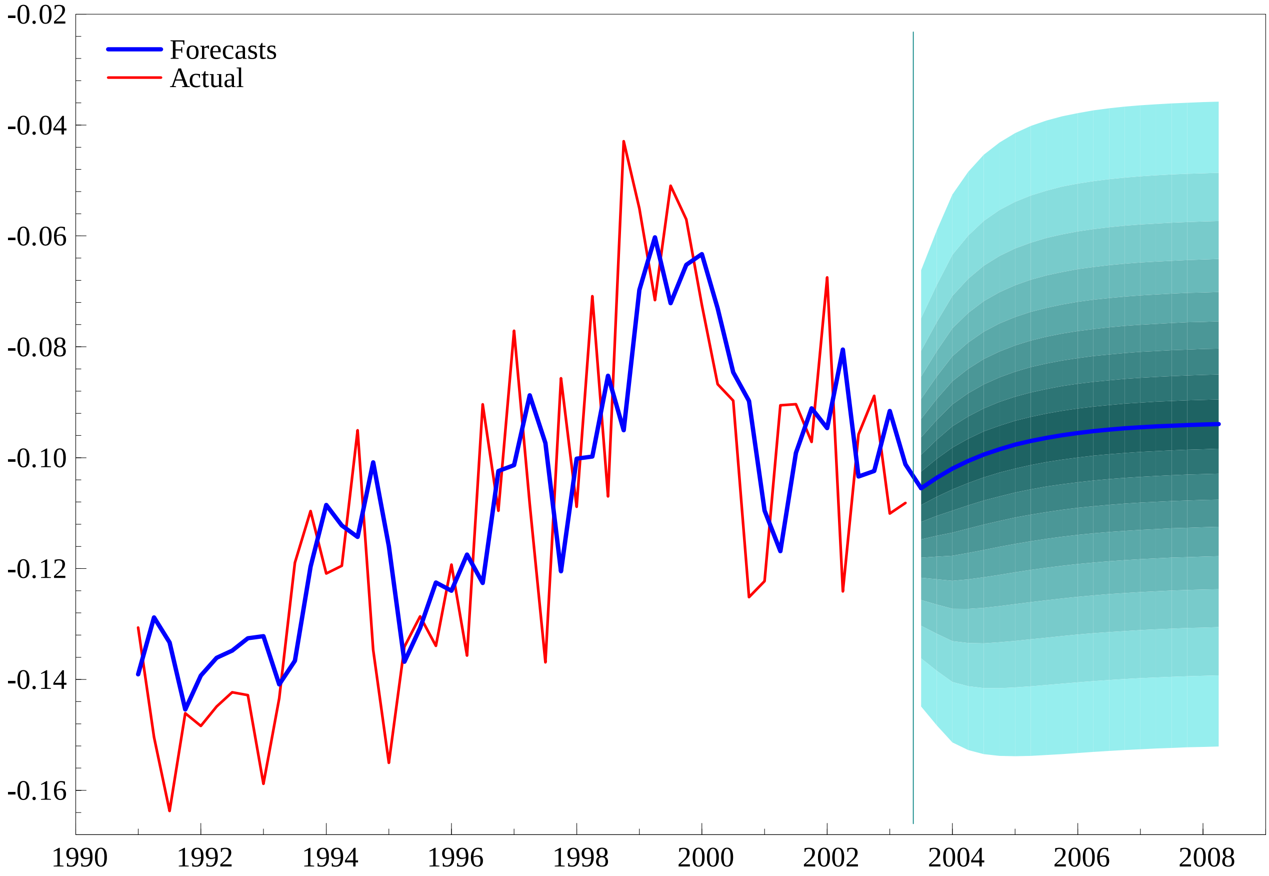
- Consider the ARMA(1,1) model

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

- 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

$$\begin{aligned}
y_{T+1|T} &= E[\theta \cdot y_T + \epsilon_{T+1} + \alpha \cdot \epsilon_T \mid \mathcal{I}_T] \\
&= \widehat{\theta} \cdot y_T + \widehat{\alpha} \cdot \widehat{\epsilon}_T \\
y_{T+2|T} &= E[\theta \cdot y_{T+1} + \epsilon_{T+2} + \alpha \cdot \epsilon_{T+1} \mid \mathcal{I}_T] \\
&= \widehat{\theta} \cdot y_{T+1|T}.
\end{aligned}$$


Forecasts
Actual
-0.02
-0.04
-0.06
-0.08
-0.10
-0.12
-0.14
-0.16
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008