Time Series Analysis

Discussion Section 02

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Organizational Issues

Please log in to **ISIS** (password: Zeit1718) and **download** the following files:

- GNP.dta
- acf_exercise.do

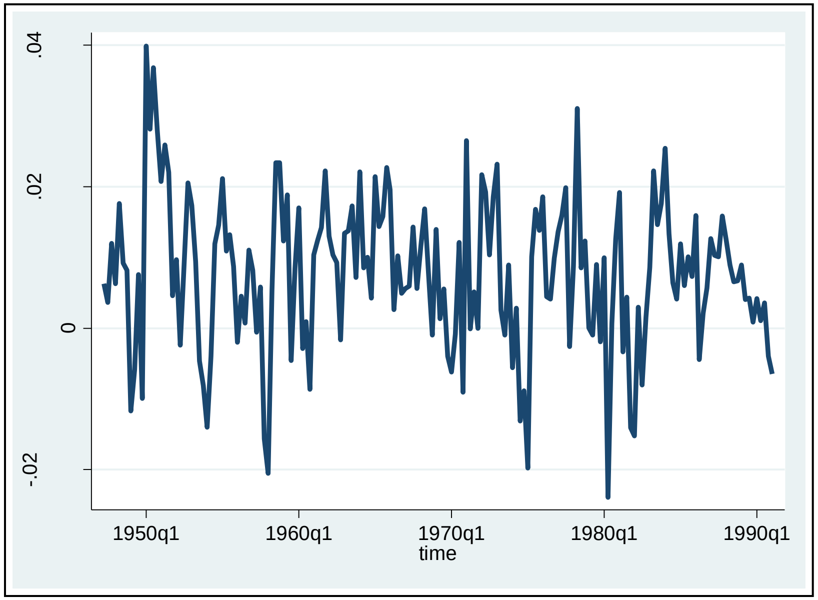
Stationary Stochastic Processes

- Introduction
- **Identification**
  - Autocorrelation Function
  - Moving Average and Autoregressive Models
  - Partial Autocorrelation Function
  - ARMA Models
- Estimation
- Diagnostic Checking
- Forecasting

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.1:

- What is meant by weak stationarity and what is the difference to strict stationarity?
- Does this series look stationary?
- Does it seem like there is serial dependence?



Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.2:

Create a scatter plot to identify if there is serial dependency for lag *k* = 1,2,3 in the **GNP** series. Add the means and the regression line.

Recall:

- Time-series operators: When a command allows a time-series *varlist*, you may include time-series operators.

Operator	Meaning
L.	lag x_{t-1}
L2.	2-period lag x_{t-2}
...	...

- `[twoway] scatter` *varlist* `[if] [in] [weight] [,options]` where *varlist* is `y_1 [y_2 [...]] x`
- **Twoway linear prediction plots**: `twoway lfit` calculates the prediction for *yvar* based on a linear regression of *yvar* on *xvar* and plots the resulting line (Sytax: `twoway lfit` *yvar xvar*)
- Use the **||-separator notation** to put on top of the scatter plot the prediction from a linear regression (`Syntax: [twoway] scatter ... [, scatter_options] || lfit ...`)

Exercise 2.2:

```
. describe

  obs:           176
 vars:             1                          21 Nov 2011 10:58
 size:           704
-------------------------------------------------------------------------------------------------------------------
              storage  display    value
variable name   type   format     label      variable label
-------------------------------------------------------------------------------------------------------------------
GNP             float  %9.0g                 Quarterly growth rate of U.S. real gross national product (1947q2 to
                                               1991q1)
-------------------------------------------------------------------------------------------------------------------
Sorted by:
```

Note:
We have **quarterly data** beginning in the **2nd** quarter of **1947** and ending in the **1st** quarter of **1991**.
However, there is no time variable in our data set.

=> create an appropriate time variable

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.3:

- What is the autocorrelation function and what is its purpose?
- Consider the formula of the Bravais-Pearson correlation coefficient and compare it to the formula for the autocorrelation coefficient at lag *k*. Explain under which assumptions both formulas yield the same result.

Bravais-Pearson correlation coefficient:

$$r = \frac{E\left[(x-\mu_x)(y-\mu_y)\right]}{\sqrt{E\left[(x-\mu_x)^2\right] E\left[(y-\mu_y)^2\right]}}$$

Autocorrelation coefficient at lag *k*:

$$\rho_k = \frac{E\left[(y_t-\mu_y)(y_{t+k}-\mu_y)\right]}{E\left[(y_t-\mu_y)^2\right]}$$

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.4:

- Calculate the values of the autocorrelation function for the **GNP** series.

  Note: `corrgram` *`varname`* tabulate autocorrelations

- Plot the autocorrelation function for the **GNP** series.

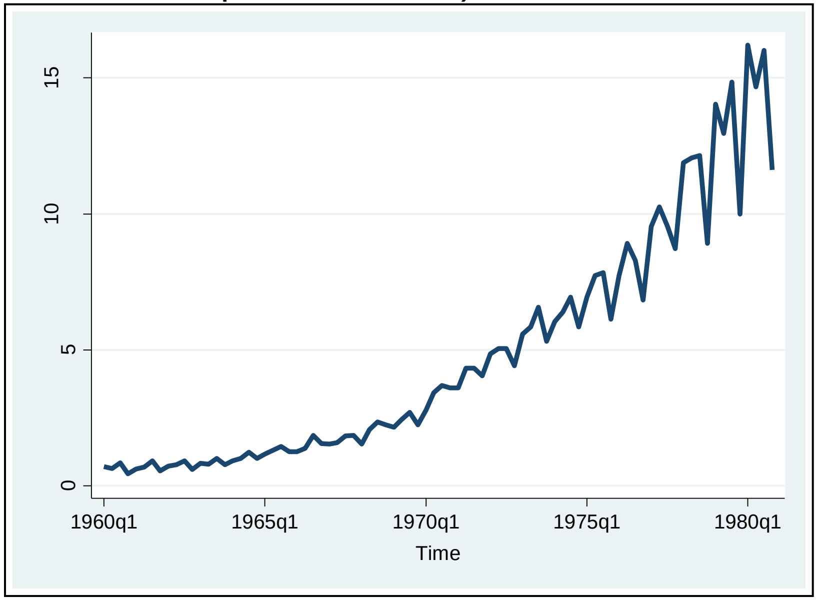
  Note: `ac` *`varname`* graph of autocorrelations

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.5-1:

What should the ACF for this series look like?

Quarterly earnings per share for the U.S. company Johnson & Johnson (first quarter of 1960 to last quarter of 1980)



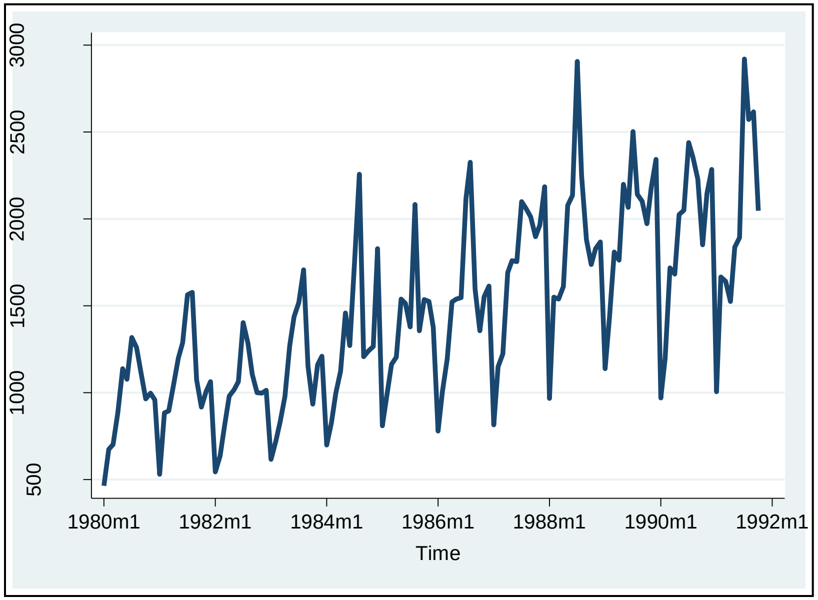
Shumway, Stoffer (2000) “Time series analysis and its applications”

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.5-2:

What should the ACF for this series look like?

Monthly Australian red wine sales
(Jan. 1990 to Oct. 1991)



Shumway, Stoffer (2000) “Time series analysis and its applications”

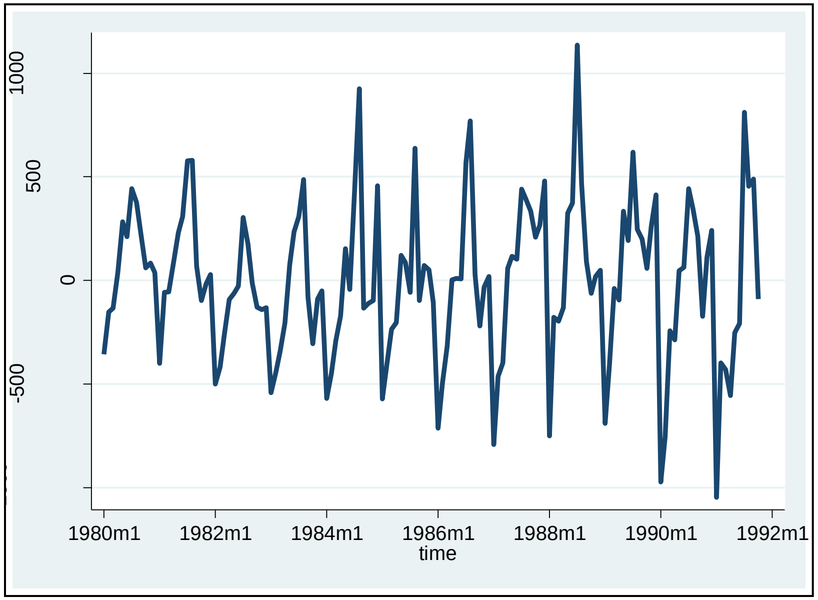
Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.5-3:

What should the ACF for this series look like?

Monthly Australian red wine sales (Jan. 1990 to Oct. 1991)

Detrended

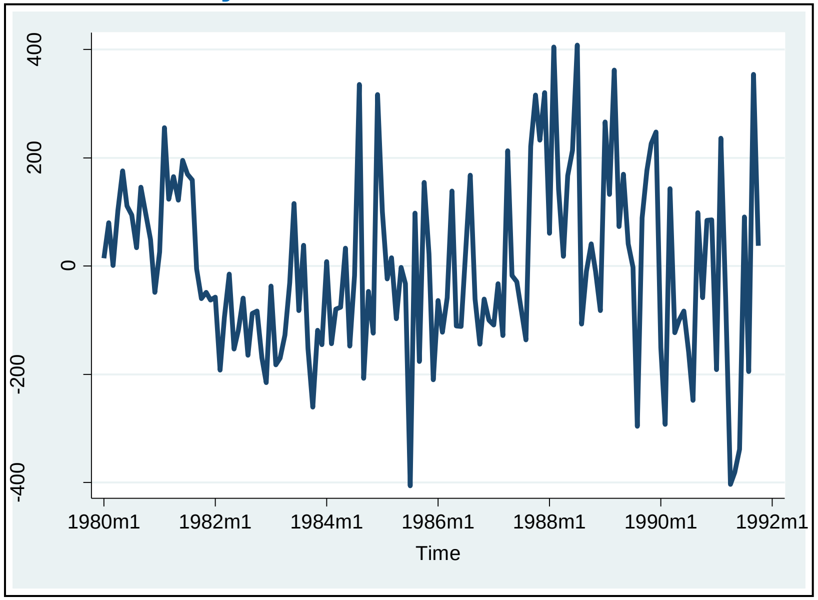


Shumway, Stoffer (2000) "Time series analysis and its applications"

Exercise 2.5-4:

What should the ACF for this series look like?

Monthly Australian red wine sales
(Jan. 1990 to Oct. 1991)
Seasonal adjusted and detrended



Shumway, Stoffer (2000) “Time series analysis and its applications”

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.6:

Write down the following models in our general notation:

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

Calculate the values of the theoretical autocorrelation function for an MA(1) process with $\mu = 0$ for lags $k = 1, 2$.

Calculate the values of the theoretical autocorrelation function for an AR(1) process with $\delta = 0$ for lags $k = 1, 2$.

Hint: $$\rho_k := \frac{\gamma_k}{\gamma_0} = \frac{Cov(y_t, y_{t-k})}{Var(y_t)} = \frac{E[(y_t - \mu)(y_{t-k} - \mu)]}{E[(y_t - \mu)^2]}$$

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.7:

Simulate the following processes with 200 observations and $\varepsilon_t \sim \text{i.i.d.}\, N(0, \sigma_\varepsilon^2 = 1)$:

- White noise
- MA(1) with θ_1 = -0.8
- MA(2) with θ_1 = -0.6 and θ_2 = 0.3

Plot in each case the ACF of the series with the help of the do-file (acf_exercise.do) and compare to its theoretical ACF you would expect.

Notice:

ACF MA(1)

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \dfrac{-\theta_1}{1+\theta_1^2} & k = 1 \\ 0 & k > 1 \end{cases}$$

ACF MA(2)

$$\rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2} \qquad \rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}$$

$$\rho_k = 0 \quad \text{for } k > 2$$

Exercise 2.8:

Simulate the following processes with 2000 observations and $\varepsilon_t \sim$ i.i.d. $N\left(0, \sigma_\varepsilon^2 = 1\right)$:

- AR(1) with $\varphi_1 = 0.9$
- AR(2) with $\varphi_1 = 0.9$ und $\varphi_2 = -0.7$

Plot in each case the ACF of the series with the help of the do-file (acf_exercise.do) for the last 200 observations and compare to its theoretical ACF you would expect.

Notice:

ACF AR(1)

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \varphi_1^k$$

ACF AR(2)

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\varphi_1}{1-\varphi_2}, \quad \rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\varphi_1^2}{1-\varphi_2} + \varphi_2$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2}$$

Stochastic first-order difference equation

Date	Equation
1	$y_1 = \varphi_1 \cdot y_0 + (\delta + \varepsilon_1)$
2	$y_2 = \varphi_1 \cdot y_1 + (\delta + \varepsilon_2)$
	$= \varphi_1 \cdot [\varphi_1 \cdot y_0 + (\delta + \varepsilon_1)] + (\delta + \varepsilon_2)$
	$= \varphi_1^2 \cdot y_0 + (\varphi_1 + 1)\delta + \varphi_1 \varepsilon_1 + \varepsilon_2$
3	$y_3 = \varphi_1 \cdot y_2 + (\delta + \varepsilon_3)$
	$= \varphi_1^3 \cdot y_0 + (\varphi_1^2 + \varphi_1 + 1)\delta + \varphi_1^2 \varepsilon_1 + \varphi_1 \varepsilon_2 + \varepsilon_3$
	$\vdots$
t	$y_t = \varphi_1^t \cdot y_0 + (\varphi_1^{t-1} + \varphi_1^{t-2} + \ldots + \varphi_1 + 1) \cdot \delta + \sum_{j=1}^{t} \varphi_1^{t-j} \cdot \varepsilon_j$

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

AR(1) as First-order difference equation

$$y_t = \varphi_1^t \cdot y_0 + \left(\varphi_1^{t-1} + \varphi_1^{t-2} + \ldots + \varphi_1 + 1\right) \cdot \delta + \sum_{j=1}^{t} \varphi_1^{t-j} \cdot \varepsilon_j$$

$$E[y_t] = \varphi_1^t \cdot y_0 + \left(\varphi_1^{t-1} + \varphi_1^{t-2} + \ldots + \varphi_1 + 1\right) \cdot \delta$$

$$E[y_{t+s}] = \varphi_1^{t+s} \cdot y_0 + \left(\varphi_1^{t+s-1} + \varphi_1^{t+s-2} + \ldots + \varphi_1^{t-1} + \ldots + \varphi_1 + 1\right) \cdot \delta$$

$$E[y_t] \neq E[y_{t+s}]$$

To get stationarity, we need to impose conditions.

AR(1) as First-order difference equation:

$$y_t = \varphi_1^t \cdot y_0 + \left(\varphi_1^{t-1} + \varphi_1^{t-2} + \ldots + \varphi_1 + 1\right) \cdot \delta + \sum_{j=1}^{t} \varphi_1^{t-j} \cdot \varepsilon_j$$

If $t \rightarrow \infty$ and $|\varphi_1|<1$

$$\lim y_t = \delta \big/ (1-\varphi_1) + \sum_{j=0}^{\infty} \varphi_1^j \cdot \varepsilon_{t-j} \qquad \Rightarrow E[y_t] = \delta \big/ (1-\varphi_1)$$

$t \rightarrow \infty$ means: the process has started a long time ago "**stochastic initial conditions**" ➔ That is why we only use the last 200 observations from our simulated 2000 data points!

$|\varphi_1|<1$ means: dependence can't be too strong

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.9:

- What is the **partial** autocorrelation function (PACF) and what is its purpose?
- Which regression should you run to estimate the partial autocorrelation coefficient for y_t at lag $k = 3$?
- What should the PACFs for the simulated processes from **Exercise 5.7** and **Exercise 5.8** look like?

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.10:

- Summarize the stylized shapes for the autocorrelation and partial autocorrelation functions for the following stochastic processes: white noise, AR(1), AR(2), MA(1) and MA(2).
- Which model in general is more applicable to model strong serial dependence?

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Organizational Issues

Please log in to **ISIS2** (password: Zeit1718) and **download** the following file:

- inventories.dta

Univariate Box-Jenkins models for stationary time series

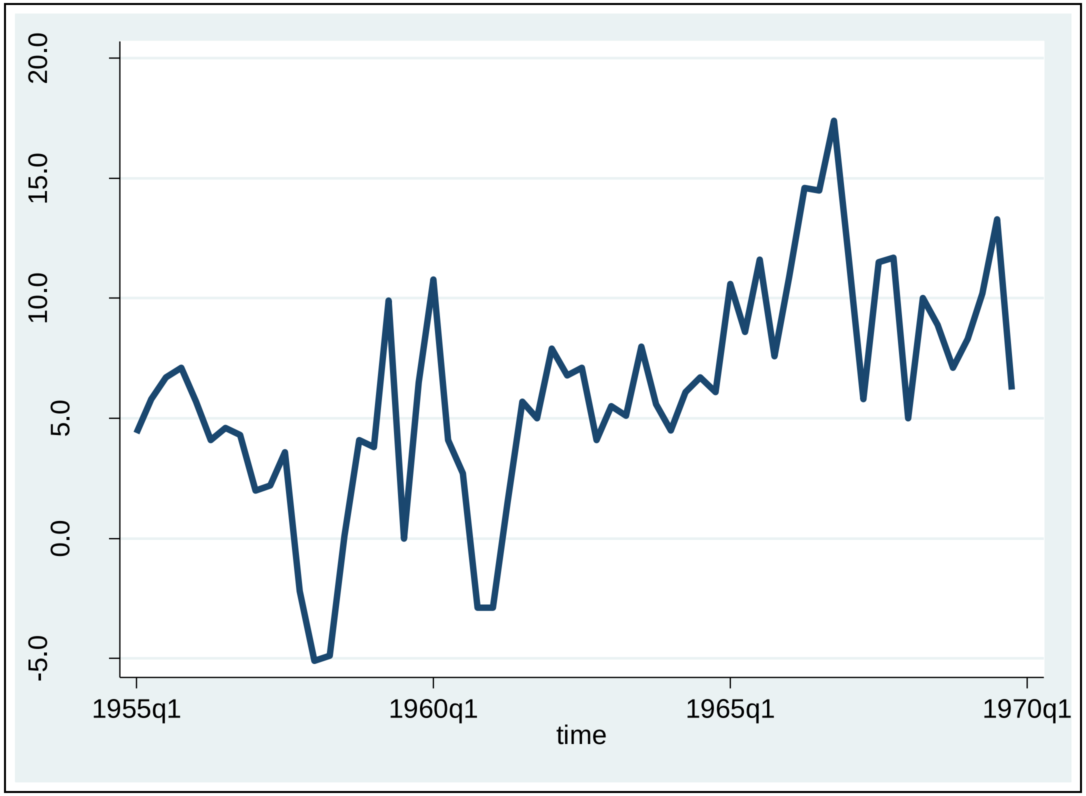
General Procedure:

1. Identification
2. Estimation
   - Solution of the Yule-Walker equations (AR processes)
   - Least Squares Estimation (AR processes)
   - Maximum Likelihood Estimation
   - Conditional Maximum Likelihood Estimation
3. Diagnostic Checking
4. Forecasting

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Business Inventories – Original Series

Y: “Quarterly change in business inventories”



- 60 observations from 1955q1 trough 1969q4
- the data have been seasonally adjusted

Pankratz (1983) “Forecasting with univariate Box-Jenkins models”

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.11: Identification

- What is the first step in the Box-Jenkins methodology?

  Examine the **original series**. Is the series stationary?

  - The observations seem to fluctuate around a constant mean.
  - The variance seems to be constant over time.

- What is the next step in the Box-Jenkins methodology?

  Plot the **ACF** and **PACF** and describe their pattern. Use about one-forth of the number of observations. Identify the most appropriate time series process for this data.

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.12: Estimation

Estimate the parameter of the

AR(1): $y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$

- Solution of the Yule-Walker equations
- Least Squares Estimation
- Maximum Likelihood Estimation
- Conditional Maximum Likelihood Estimation

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Estimation – Solution of the Yule-Walker equations

Yule-Walker equations:

$$\rho_1 = \varphi_1\rho_0 + \varphi_2\rho_1 + \ldots + \varphi_p\rho_{p-1}$$
$$\vdots$$
$$\rho_p = \varphi_1\rho_{p-1} + \varphi_2\rho_{p-2} + \ldots + \varphi_p$$

If ρ_1, ρ_2, ..., ρ_p are known, the equation can be solved for the autoregressive parameters φ_1, φ_2, ..., φ_p.

For an **AR(1)** process it reduces to:

$$\rho_1 = \varphi_1\rho_0$$

Estimation – Least Squares Estimation

We can estimate δ, φ_1, φ_2, ..., φ_p by ordinary least squares (these estimates minimize the sum of squared residuals):

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \delta + \varepsilon_t$$

Under the following **assumption** OLS provides consistent estimators:

$$E(y_{t-j}\varepsilon_t) = 0 \quad \text{for } j = 1,2,3,\ldots,p$$

For an **AR(1)** process it reduces to:

$$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$$

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Estimation – Maximum Likelihood Estimation

Calculate the probability density:

$$f_{Y_T, Y_{T-1}, \ldots, Y_1}(y_T, y_{T-1}, \ldots, y_1; \boldsymbol{\theta})$$

“probability of having observed the particular sample y_1, y_2, ..., y_T”

The maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ is the value of $\boldsymbol{\theta}$ that maximizes this probability.

Assumption: $\varepsilon_t \sim \text{i.i.d.}\, N\left(0, \sigma_\varepsilon^2\right)$

The joint density is a product of **conditional densities:**

$$f_{Y_T, Y_{T-1}, \ldots, Y_1}(y_T, y_{T-1}, \ldots, y_1 | \boldsymbol{\theta}) = f_{Y_1}(y_1; \boldsymbol{\theta}) \cdot \prod_{t=2}^{T} f_{Y_t | Y_{t-1}, Y_{t-2}, \ldots, Y_1}(y_t | y_{t-1}, y_{t-2}, \ldots, y_1; \boldsymbol{\theta})$$

For an **AR(1)** process each factor (t = 2, ..., T) reduces to:

$$f_{Y_t | Y_{t-1}, Y_{t-2}, \ldots, Y_1}(y_t | y_{t-1}, y_{t-2}, \ldots, y_1; \boldsymbol{\theta}) = f_{Y_t | Y_{t-1}}(y_t | y_{t-1}; \boldsymbol{\theta})$$

Estimation – Maximum Likelihood Estimation

Likelihood function for an **AR(1)** process:

$$f_{Y_T,Y_{T-1},\ldots,Y_1}\left(y_T,y_{T-1},\ldots,y_1;\boldsymbol{\theta}\right)=f_{Y_1}\left(y_1;\boldsymbol{\theta}\right)\cdot\prod_{t=2}^{T}f_{Y_t|Y_{t-1}}\left(y_t\mid y_{t-1};\boldsymbol{\theta}\right)$$

Log likelihood function for an **AR(1)** process:

$$L(\boldsymbol{\theta})=\log\left[f_{Y_1}\left(y_1;\boldsymbol{\theta}\right)\right]+\sum_{t=2}^{T}\log\left[f_{Y_t|Y_{t1}}\left(y_t\mid y_{t-1};\boldsymbol{\theta}\right)\right]$$

$$L(\boldsymbol{\theta})=\log\left[\frac{1}{\sqrt{2\pi}\sqrt{\sigma_\varepsilon^2/\left(1-\varphi_1^2\right)}}\exp\left[\frac{-\left\{y_1-\left[\delta/1-\varphi_1\right]\right\}^2}{2\sigma_\varepsilon^2/\left(1-\varphi_1^2\right)}\right]\right]$$

$$+\sum_{t=2}^{T}\log\left[\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\exp\left[\frac{-\left\{y_{t+1}-\delta-\varphi_1 y_t\right\}^2}{2\sigma_\varepsilon^2}\right]\right]$$

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Estimation – Maximum Likelihood Estimation

Log likelihood function for an **AR(1)** process:

$$L(\boldsymbol{\theta}) = \log\left[\frac{1}{\sqrt{2\pi}\sqrt{\sigma_\varepsilon^2 / (1-\varphi_1^2)}}\exp\left[\frac{-\{y_1 - [\delta/1-\varphi_1]\}^2}{2\sigma_\varepsilon^2/(1-\varphi_1^2)}\right]\right] + \sum_{t=2}^{T}\log\left[\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\exp\left[\frac{-\{y_{t+1}-\delta-\varphi_1 y_t\}^2}{2\sigma_\varepsilon^2}\right]\right]$$

$$L(\boldsymbol{\theta}) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log[\sigma_\varepsilon^2/(1-\varphi_1^2)] - \frac{\{y_1-[\delta/1-\varphi_1]\}^2}{2\sigma_\varepsilon^2/(1-\varphi_1^2)} - \left[\frac{(T-1)}{2}\right]\cdot\log(2\pi) - \left[\frac{(T-1)}{2}\right]\cdot\log(\sigma_\varepsilon^2) - \sum_{t=2}^{T}\left[\frac{\{y_t-\delta-\varphi_1 y_{t-1}\}^2}{2\sigma_\varepsilon^2}\right]$$

In principle, the maximization requires differentiating and setting the result equal to zero. In praxis, it requires iterative or numerical procedures.

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Estimation – Maximum Likelihood Estimation

`arima` *`varname`*`, ar(`*`numlist`*`) ma(`*`numlist`*`)` estimates an AR(p) MA(q) model using the maximum likelihood method.

Example:

`arima` *`varname`*`, ar(1) ma(1/3)` estimates an ARMA(1,3) model, because *`numlist 1/3`* denotes three numbers: 1, 2, 3

Notation in Stata:

$$y_t = \sum_{i=1}^{p} \varphi_i y_{t-i} \boxed{+} \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Estimation – Conditional ML Estimation

Treat the value of y_1 as deterministic and maximize the likelihood conditioned of the first observation:

$$f_{Y_T, Y_{T-1}, \ldots, Y_2 | Y_1}\left(y_T, y_{T-1}, \ldots, y_2 | y_1; \boldsymbol{\theta}\right)$$

For an **AR(1)** process:

$$L\left(y_T, y_{T-1}, \ldots, y_2 | y_1; \boldsymbol{\theta}\right) = -\left[\frac{(T-1)}{2}\right] \cdot \log(2\pi) - \left[\frac{(T-1)}{2}\right] \cdot \log\left(\sigma_\varepsilon^2\right) - \sum_{t=2}^{T}\left[\frac{\{y_t - \delta - \varphi_1 y_{t-1}\}^2}{2\sigma_\varepsilon^2}\right]$$

Maximization with respect to δ and φ_1 is equivalent to minimize:

$$\sum_{t=2}^{T}\{y_t - \delta - \varphi_1 y_{t-1}\}^2$$

This can be achieved by an OLS regression of y_t on a constant and it's own lagged values.

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Estimation – Conditional ML Estimation

For an **AR(1)** process:

$$L(y_T, y_{T-1}, \ldots, y_2 | y_1; \boldsymbol{\theta}) = -\left[\frac{(T-1)}{2}\right] \cdot \log(2\pi) - \left[\frac{(T-1)}{2}\right] \cdot \log(\sigma_\varepsilon^2) - \sum_{t=2}^{T} \left[\frac{\{y_t - \delta - \varphi_1 y_{t-1}\}^2}{2\sigma_\varepsilon^2}\right]$$

The conditional MLE of the variance is found by differentiating the log likelihood with respect to σ_ε^2:

$$\hat{\sigma}_\varepsilon^2 = \sum_{t=2}^{T} \left[\frac{\{y_t - \hat{\delta} - \hat{\varphi}_1 y_{t-1}\}^2}{T-1}\right]$$

The conditional MLE is the average squared residuals from the OLS regression.

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise (inventories.dta)

Exercise 2.13: Diagnostic Checking

Do diagnostic checking for the ML estimation of the

AR(1): $y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$

- Is the series stationary?
- Are the estimated coefficients significant?
- Is the AR(1) model appropriate?

Diagnostic Checking – Residual ACF and PACF

How should the residuals “behave”?

Diagnostic Checking – Joint Hypothesis Test

H_0: All autocorrelation coefficients are zero

Box and Pierce

$$Q = T\sum_{k=1}^{K}\hat{\rho}_k^2 \quad \sim\chi^2 \quad \text{with } K-p-q \text{ degrees of freedom}$$

Box and Ljung (refined test)

$$Q = T(T+2)\sum_{k=1}^{K}\frac{1}{T-k}\hat{\rho}_k^2 \quad \sim\chi^2 \quad \text{with } K-p-q \text{ degrees of freedom}$$

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.14: Forecasting

- Estimate the AR(1) model by OLS with only the first 56 observations.
- What is the optimal forecast? In which sense is it optimal? What “assumptions” are part of the information set?
- Calculate forecasts for 1969q1 to 1969q4.
- Compare these forecasts to the actual values and compute percentage forecast errors.
- What is the forecast for December 2010?

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Organizational Issues

Please log in to **ISIS** (password: Zeit1718) and **download** the following file:

- coal_production.dta

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Univariate Box-Jenkins models for stationary time series

General Procedure:

1. Identification
2. Estimation
   - Solution of the Yule-Walker equations
   - Least Squares Estimation
   - Maximum Likelihood Estimation
3. Diagnostic Checking
   - Significance of estimated coefficients
   - Test of individual values of residual ACF and PACF
   - Joint test that residuals are white noise (Box and Ljung)
   - Comparison of different candidate models (AIC and BIC)
4. Forecasting

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Some helpful formulas

Yule-Walker equations:

$$\begin{aligned}\rho_1 &= \varphi_1\rho_0 + \varphi_2\rho_1 + \ldots + \varphi_p\rho_{p-1} \\ &\vdots \\ \rho_p &= \varphi_1\rho_{p-1} + \varphi_2\rho_{p-2} + \ldots + \varphi_p\end{aligned}$$

Box and Ljung (refined test):

$$Q = T(T+2)\sum_{k=1}^{K}\frac{1}{T-k}\hat{\rho}_k^2 \quad \sim\chi^2 \quad \text{with } K-p-q \text{ degrees of freedom}$$

Akaike's Information Criterion (AIC):

$$AIC = \log\hat{\sigma}^2 + 2\frac{p+q}{T}$$

Bayesian Information Criterion (BIC):

$$BIC = \log\hat{\sigma}^2 + \frac{p+q}{T}\log T$$

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Some helpful Stata commands

- **`generate`** ***`newvar = exp`***
  creates a new variable, the values of the variable are specified by =*exp*. It allows you to create a new variable that is an algebraic expression of other variables.
- **`_n`**
  contains the number of the current observation; it can be used with mathematical operators.
- **`format`** ***`timevar %fmt`***
  allows you to specify the display format for variables.

Format (`fmt`)	Description	Coding
%td	daily	0 = 01jan1960, 1 = 02jan1960
%tw	weekly	0 = 1960w1, 1 = 1960w2
%tm	monthly	0 = 1960m1, 1 = 1960m2
%tq	quarterly	0 = 1960q1, 1 = 1960q2
%th	halfyearly	0 = 1960h1, 1 = 1960h2
%ty	yearly	1960 = 1960, 1961 = 1961

- **`tsset`** ***`timevar`***
  declares the data to be a time series and designates that *timevar* represents time.
- **`tsline`** ***`varname`***
  draws line plots for time-series data.

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Some helpful Stata commands

- **ac *varname*, lags(#)**
  produces a graph of # autocorrelations with pointwise confidence intervals based on Bartlett's formula.
- **pac *varname*, lags(#)**
  produces a graph of # partial autocorrelations with confidence intervals calculated using a standard error of 1/sqrt(n).
- **corrgram *varname*, lags(#)**
  produces a table of # autocorrelations, partial autocorrelations, and Portmanteau (Q) statistics.
- **summarize *varlist***
  calculates and displays a variety of univariate summary statistics.
- **display**
  can be used as a substitute for a hand calculator.
- **regress *depvar* [*indepvars*]**
  fits a model of *depvar* on *indepvars* using linear regression.
- **L.*varname***
  refers to the lagged value of variable *varname*.

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Some helpful Stata commands

- **`arima`** ***`varname`*, `ar(`*`numlist`*`) ma(`*`numlist`*`)`**
  estimates an AR(p) MA(q) model using the maximum likelihood method.
- ***`numlist`***
  is a list of numbers; example: *1/3* three numbers: 1, 2, 3.
- **`predict`** ***`newvar`*, `residuals`**
  predicts the residuals from the last estimation.
- **`chi2(`*`n`*`,`*`x`*`)`**
  returns the cumulative chi-squared distribution with n degrees of freedom for n > 0.
- **`log(`*`x`*`)`**
  returns the natural logarithm of x if x>0. This is a synonym for ln(x).

Exercise 2.15: Coal Production

Follow the **Univariate Box-Jenkins models for stationary time series** to estimate an **appropriate** model (if necessary consider different candidate models).

Percentiles of the chi-squared distribution

Percentiles of the χ^2 Distribution

	Percent									
df	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
1	0.000039	0.000157	0.000982	0.003932	0.015791	2.705544	3.841459	5.023886	6.634897	7.879439
2	0.010025	0.020101	0.050636	0.102587	0.210721	4.605170	5.991465	7.377759	9.210340	10.596635
3	0.071722	0.114832	0.215795	0.351846	0.584374	6.251388	7.814728	9.348404	11.344867	12.838156
4	0.206989	0.297109	0.484419	0.710723	1.063623	7.779440	9.487729	11.143287	13.276704	14.860259
5	0.411742	0.554298	0.831212	1.145476	1.610308	9.236357	11.070498	12.832502	15.086272	16.749602
6	0.675727	0.872090	1.237344	1.635383	2.204131	10.644641	12.591587	14.449375	16.811894	18.547584
7	0.989256	1.239042	1.689869	2.167350	2.833107	12.017037	14.067140	16.012764	18.475307	20.277740
8	1.344413	1.646497	2.179731	2.732637	3.489539	13.361566	15.507313	17.534546	20.090235	21.954955
9	1.734933	2.087901	2.700390	3.325113	4.168159	14.683657	16.918978	19.022768	21.665994	23.589351
10	2.155856	2.558212	3.246973	3.940299	4.865182	15.987179	18.307038	20.483177	23.209251	25.188180
11	2.603222	3.053484	3.815748	4.574813	5.577785	17.275009	19.675138	21.920049	24.724970	26.756849
12	3.073824	3.570569	4.403789	5.226029	6.303796	18.549348	21.026070	23.336664	26.216967	28.299519
13	3.565035	4.106915	5.008751	5.891864	7.041505	19.811929	22.362032	24.735605	27.688250	29.819471
14	4.074675	4.660425	5.628726	6.570631	7.789534	21.064144	23.684791	26.118948	29.141238	31.319350
15	4.600916	5.229349	6.262138	7.260944	8.546756	22.307130	24.995790	27.488393	30.577914	32.801321
16	5.142205	5.812213	6.907664	7.961646	9.312236	23.541829	26.296228	28.845351	31.999927	34.267187
17	5.697217	6.407760	7.564186	8.671760	10.085186	24.769035	27.587112	30.191009	33.408664	35.718466
18	6.264805	7.014911	8.230746	9.390455	10.864936	25.989423	28.869299	31.526378	34.805306	37.156451
19	6.843971	7.632730	8.906517	10.117013	11.650910	27.203571	30.143527	32.852327	36.190869	38.582257
20	7.433844	8.260398	9.590778	10.850812	12.442609	28.411981	31.410433	34.169607	37.566235	39.996846
21	8.033653	8.897198	10.282898	11.591305	13.239598	29.615089	32.670573	35.478876	38.932173	41.401065
22	8.642716	9.542492	10.982321	12.338015	14.041493	30.813282	33.924439	36.780712	40.289360	42.795655
23	9.260425	10.195716	11.688552	13.090514	14.847956	32.006900	35.172462	38.075627	41.638398	44.181275
24	9.886234	10.856362	12.401150	13.848425	15.658684	33.196244	36.415028	39.364077	42.979820	45.558512
25	10.519652	11.523975	13.119720	14.611408	16.473408	34.381587	37.652484	40.646469	44.314105	46.927890
26	11.160237	12.198147	13.843905	15.379157	17.291885	35.563171	38.885139	41.923170	45.641683	48.289882
27	11.807587	12.878504	14.573383	16.151396	18.113896	36.741217	40.113272	43.194511	46.962942	49.644915
28	12.461336	13.564710	15.307861	16.927875	18.939243	37.915923	41.337138	44.460792	48.278236	50.993376
29	13.121149	14.256455	16.047072	17.708366	19.767744	39.087470	42.556968	45.722286	49.587885	52.335618
30	13.786720	14.953457	16.790772	18.492661	20.599235	40.256024	43.772972	46.979242	50.892181	53.671962

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.16: Forecasting

- Estimate an **AR(2)** model by means of the **ML Method** using only the first **84** observations, i.e. exclude completely the last year.

- Calculate forecasts for all months from **1959m1** to **1961m12**.
  Hint:
  Previously increase the number of observations form 96 to 120 in order to create the additional time periods, i.e. from 1960m1 to 1961m12 (24 additional time periods). To this end you can use the following STATA-code:

```
. set obs 120
. replace time = tm(1952m1)+_n-1
. format time %tm
. tsset time
```

- Plot the original series together with the predicted values. Add to the plot the unconditional mean of the original series.
  What can you conclude?

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Dynamic forecasts in Stata

`dynamic(time_constant)` specifies how lags of y_t in the model are to be handled. If `dynamic()` is not specified, actual values are used everywhere that lagged values of y_t appear in the model to produce one-step-ahead forecasts.

`dynamic(time constant)` produces dynamic (also known as recursive) forecasts. `time constant` specifies when the forecast is to switch from one step ahead to dynamic. In dynamic forecasts, references to y_t evaluate to the prediction of y_t for all periods at or after time constant; they evaluate to the actual value of y_t for all prior periods.

For example, `dynamic(10)` would calculate predictions in which any reference to y_t with $t < 10$ evaluates to the actual value of y_t and any reference to y_t with $t \geq 10$ evaluates to the prediction of y_t. This means that one-step-ahead predictions are calculated for $t < 10$ and dynamic predictions thereafter.

Stata manual “arima postestimation”

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.17:

Consider the following true model for a time series

$y_t = 0.3 + 0.5\, y_{t-1} - 0.4\, \varepsilon_{t-1} + \varepsilon_t$

where ε_t is a zero mean error process.

What is the (unconditional) mean of the series, y_t ?

(1) 0.6

(2) 0.3

(3) 0.0

(4) 0.4

Brooks (2008) “Introductory Econometrics for Finance”, Multiple Choice Test Bank Questions

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.18:

Which of the following sets of characteristics would usually best describe an autoregressive process of order 3 (i.e. an AR(3))?

(1) A slowly decaying acf and pacf

(2) An acf and a pacf with 3 significant spikes

(3) A slowly decaying acf, and a pacf with 3 significant spikes

(4) A slowly decaying pacf and an acf with 3 significant spikes

Brooks (2008) “Introductory Econometrics for Finance”, Multiple Choice Test Bank Questions

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.19:

A process, x_t, which has a zero mean, a constant variance, and zero autocovariance for all non-zero lags is best described as

(1) A white noise process

(2) A covariance stationary process

(3) An autocorrelated process

(4) A moving average process

Brooks (2008) "Introductory Econometrics for Finance", Multiple Choice Test Bank Questions

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Lag-operator notation

Any stationary **ARMA(*p*,*q*)** model

$$y_t = \delta + \varepsilon_t + \underbrace{\varphi_1 y_{t-1} + \ldots + \varphi_p y_{t-p}}_{\text{"AR-part"}} \underbrace{- \Theta_1 \varepsilon_{t-1} + \ldots - \Theta_q \varepsilon_{t-q}}_{\text{"MA-part"}}$$

can be written in **lag-operator** notation, i.e.

$$\underbrace{(1 - \varphi_1 L + \ldots - \varphi_p L^p)}_{:= a_p(L)} \tilde{y}_t = \underbrace{(1 - \Theta_1 L + \ldots - \Theta_q L^q)}_{:= b_q(L)} \varepsilon_t \ , \ with\ \tilde{y}_t := y_t - \mu$$

Lag-operator **polynomial** of order **p** with **AR**-coefficients.

Lag-operator **polynomial** of order **q** with **MA**-coefficients.

Time series $\mathbf{y_t}$ written as **deviation form the mean**.

In short: $a_p(L)\tilde{y}_t = b_q(L)\varepsilon_t \ , with\ \tilde{y}_t := y_t - \mu$

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Lag-operator notation $a_p(L)\tilde{y}_t = b_q(L)\varepsilon_t$

How to write any stationary **ARMA(*p*,*q*)** in lag operator notation **?**

Proceed as follows:

1. Write y_t in deviations from its mean: $\tilde{y}_t := y_t - \mu$
2. Multiply the result from step 1 by the **appropriate** lag-operator polynomial $a_p(L)$, whose general form is $(1 - \varphi_1 L - \varphi_2 L^2 - \ldots - \varphi_p L^p)$.
3. Multiply the random shock, ε_t ,by the **appropriate** lag-operator polynomial $b_q(L)$, whose general form is $(1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q)$.
4. Equate the results from steps 2 and 3.

You can **check the correctness** of your result by multiplying out both sides of the equation $a_p(L)\tilde{y}_t = b_q(L)\varepsilon_t$ and comparing it to model written in our usual notation.

Pankratz (1983) “Forecasting with univariate Box-Jenkins models”, p. 99

Exercise 2.20:

Estimated ARMA(p,q) with T = 60 observations:

$\hat{\mu} = 101.26 \quad \hat{\varphi}_1 = 0.62 \quad \hat{\theta}_1 = -0.58 \quad \hat{\sigma}_\varepsilon = 16$

a) Determine the order of the estimated ARMA model.

b) Write down the estimated model in our usual notation.

c) Write down the model in lag operator notation.

The last observation in this data series is y_{60} = 96.91 and the predicted value (using the model from above) is $\hat{y}_{60} = 98.28$.

d) With forecast origin T = 60 calculate the first three forecasts from this model.

e) Construct confidence intervals around the three point forecasts.

Hint: $$\left[\tilde{y}_{T+l}/\Omega_T \;\pm\; z_{1-\frac{\alpha}{2}}\left(1+\psi_1^2+\ldots+\psi_{l-1}^2\right)^{\frac{1}{2}}\sigma_\varepsilon\right] = \left[\tilde{y}_{T+l}/\Omega_T \;\pm\; z_{1-\frac{\alpha}{2}}\left(1+\sum_{j=1}^{l-1}\psi_j^2\right)^{\frac{1}{2}}\sigma_\varepsilon\right]$$

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.21:

- Write down the following stationary ARMA(2,1) in lag operator notation:

$$y_t = 3 + 0.6y_{t-1} + 0.2y_{t-2} + \varepsilon_t - 0.4\varepsilon_{t-1}$$

- Write down the first difference operator, Δ, in lag operator notation.

Hint: $\Delta y_t := y_t - y_{t-1}$

- Write down the simple four-period moving average in lag operator notation.

- Multiply the first difference operator and the simple four-period moving average (both in lag operator notation). Describe the result.

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Organizational Issues

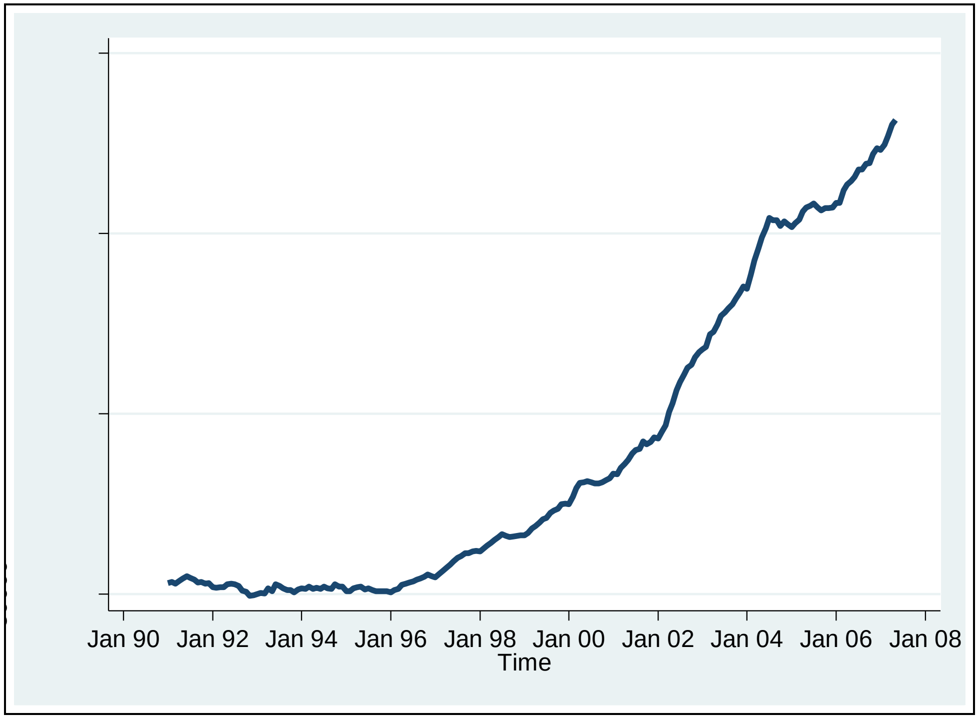
Please log in to **ISIS** (password: Zeit1718) and **download** the following file:

- ahp.dta

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Data set

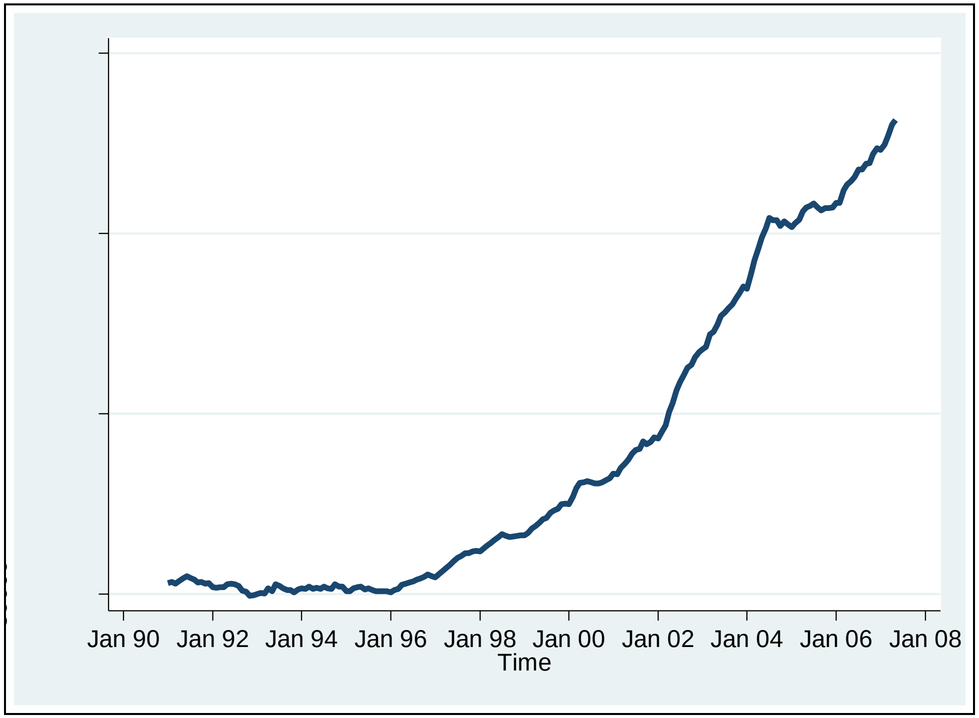
Monthly observations from January 1991 to May 2007 of UK average house prices (`ahp.dta`)




```
.generate time = tm(1991m1)+_n-1
.format time %tm
.tsset time
.tsline ahp
```

Exercise 2.22:

- Does this series look stationary?
- What would the autocorrelation function (ACF) for this series look like?



Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.23:

- Based on the PACF which model seems suitable for the `ahp` series?
- Estimate the model (maximum likelihood).
- Round the estimated coefficients to two decimal places and write down the estimated model in our usual notation and in lag operator notion (both in deviations from the mean).
- Show that the AR polynomial has a unit root.

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.24:

- Calculate the monthly percentage change (`dhp`).

$$dhp_t = 100 \cdot \frac{ahp_t - ahp_{t-1}}{ahp_{t-1}}$$

- Plot the `dhp` series.
- Is this series stationary?
- Identify candidate models and estimate them.
- Which of the candidate models is the best in terms of AIC (BIC)?

$$AIC = \log \hat{\sigma}^2 + 2\frac{p+q}{T} \qquad BIC = \log \hat{\sigma}^2 + \frac{p+q}{T}\log T$$

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 2.25:

- Estimate an AR(2) model for the `dhp` series (ML-Estimation).
- Estimate an AR(2) model for the `dhp` series (ML-Estimation) without observations from 2007 (i.e. exclude the last five observations from your analysis sample).
- What is the optimal forecast? In which sense is it optimal? What “assumptions” are part of the information set?
- Calculate forecasts for January up to May 2007.
- Compare these forecasts to the actual values and compute the forecast errors.
- What is the forecast for May 2010?

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de