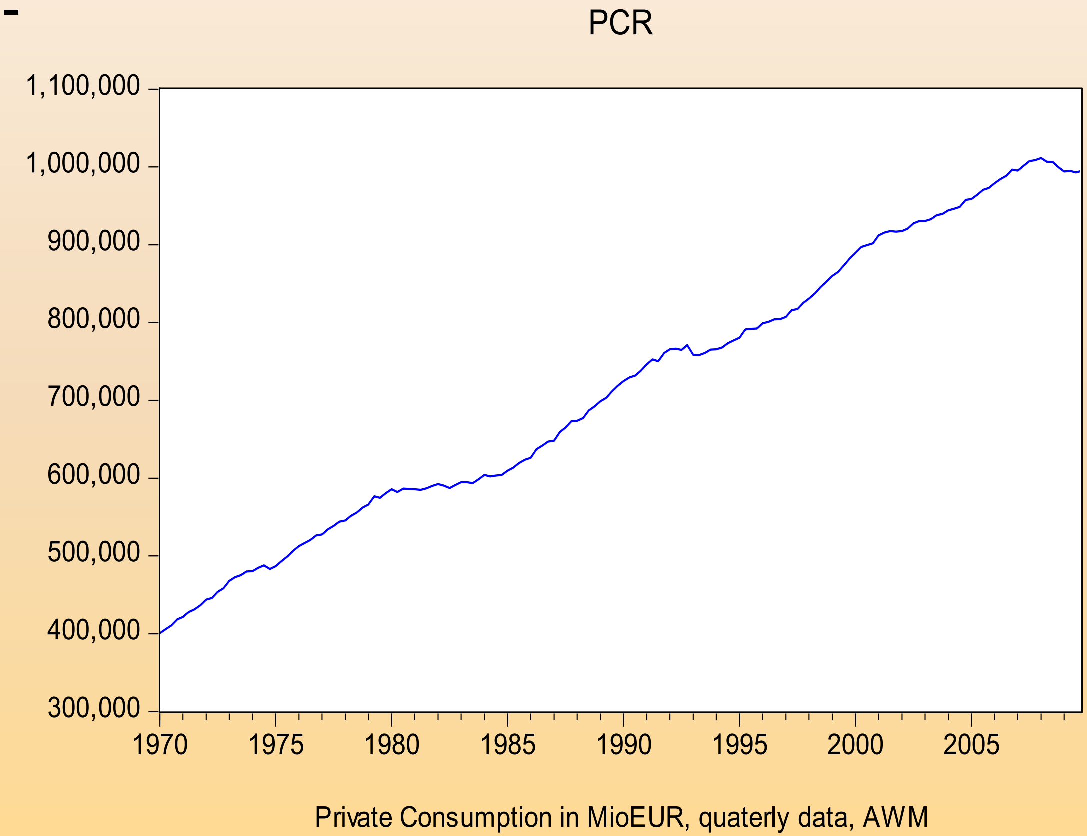
Econometrics 2 - Lecture 3

Univariate Time Series Models

Contents

Private Consumption

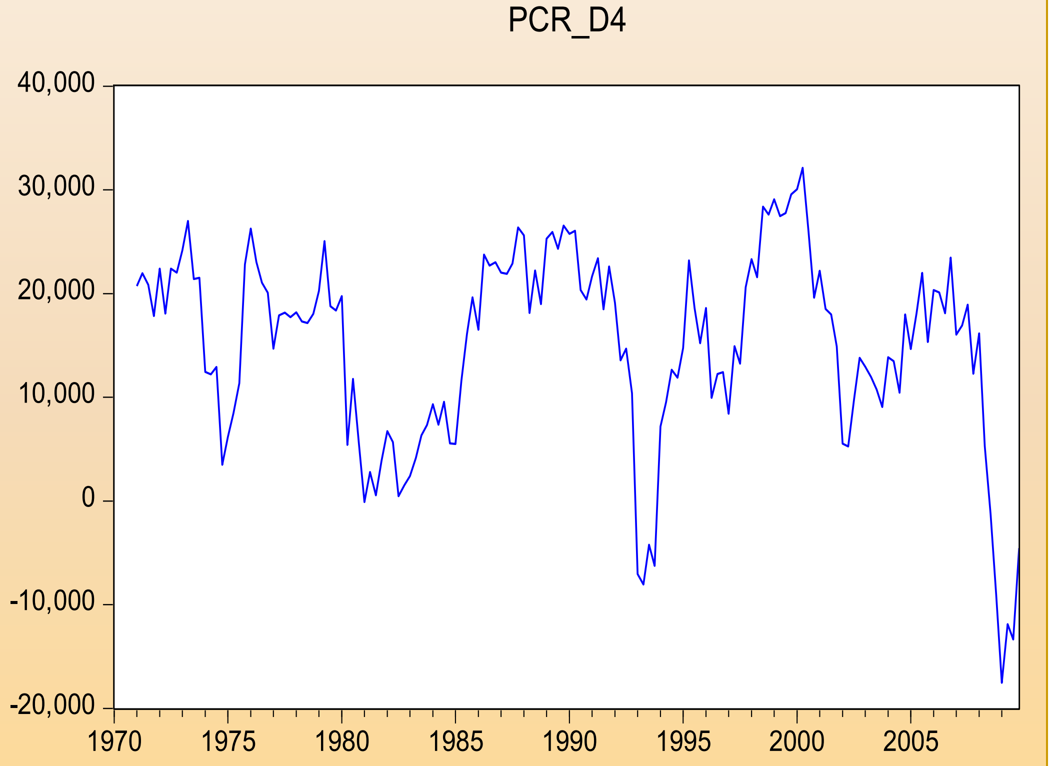
Private consumption in EURO area (16 members), seasonally adjusted, AWM database (in MioEUR)



Private Consumption, cont'd

Yearly growth of private consumption in EURO area (16 members), AWM database (in MioEUR)

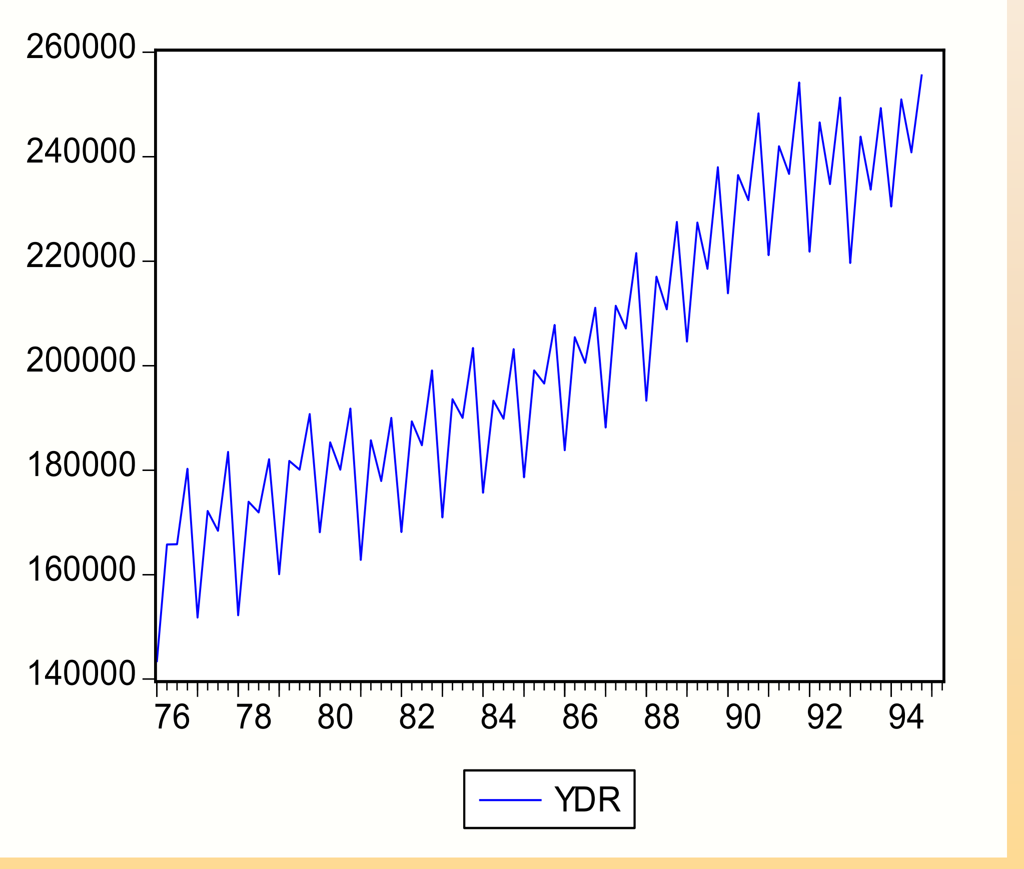
Mean growth: 15.008



Yearly growth of private consumption, AWM

Disposable Income

Disposable income
in Austria (in Mio EUR)



Time Series

Time-ordered sequence of observations of a random variable

Examples:

- Annual values of private consumption
- Changes in expenditure on private consumption
- Quarterly values of personal disposable income
- Monthly values of imports

Notation:

- Random variable Y
- Sequence of observations $Y_1, Y_2, \ldots, Y_T$
- Deviations from the mean: $y_t = Y_t - E\{Y_t\} = Y_t - \mu$

Components of a Time Series

Components or characteristics of a time series are

- Trend
- Seasonality
- Irregular fluctuations

Time series model: represents the characteristics as well as possible interactions

Purpose of modeling

- Description of the time series
- Forecasting the future

Example: $Y_t = \beta t + \Sigma_i \gamma_i D_{it} + \varepsilon_t$

with D_{it} = 1 if t corresponds to i-th quarter, D_{it} = 0 otherwise

for describing the development of the disposable income

Contents

Stochastic Process

Time series: realization of a stochastic process

Stochastic process is a sequence of random variables Y_t, e.g.,

$\{Y_t, t = 1, ..., n\}$

$\{Y_t, t = -\infty, ..., \infty\}$

Joint distribution of the $Y_1, ... , Y_n$:

$p(y_1, ..., y_n)$

Of special interest

- Evolution of the expectation $\mu_t = E\{Y_t\}$ over time
- Dependence structure over time

Example: Extrapolation of a time series as a tool for forecasting

White Noise Process

White noise process x_t, $t = -\infty, ..., \infty$

- $E\{x_t\} = 0$
- $V\{x_t\} = \sigma^2$
- $Cov\{x_t, x_{t-s}\} = 0$ for all (positive or negative) integers s

i.e., a mean zero, serially uncorrelated, homoskedastic process

AR(1)-Process

States the dependence structure between consecutive observations as

$$Y_t = \delta + \theta Y_{t-1} + \varepsilon_t, \quad |\theta| < 1$$

with ε_t: white noise, i.e., $V\{\varepsilon_t\} = \sigma^2$ (see next slide)

- Autoregressive process of order 1

From $Y_t = \delta + \theta Y_{t-1} + \varepsilon_t = \delta + \theta\delta + \theta^2\delta + \ldots + \varepsilon_t + \theta\varepsilon_{t-1} + \theta^2\varepsilon_{t-2} + \ldots$ follows

$$E\{Y_t\} = \mu = \delta(1-\theta)^{-1}$$

- $|\theta| < 1$ needed for convergence! Invertibility condition

In deviations from μ, $y_t = Y_t - \mu$:

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

AR(1)-Process, cont'd

Autocovariances $\gamma_k = \text{Cov}\{Y_t, Y_{t-k}\}$

- k=0: $\gamma_0 = V\{Y_t\} = \theta^2 V\{Y_{t-1}\} + V\{\varepsilon_t\} = \ldots = \Sigma_i \theta^{2i} \sigma^2 = \sigma^2(1-\theta^2)^{-1}$
- k=1: $\gamma_1 = \text{Cov}\{Y_t, Y_{t-1}\} = E\{(\theta y_{t-1}+\varepsilon_t) y_{t-1}\} = \theta V\{y_{t-1}\} = \theta\sigma^2(1-\theta^2)^{-1}$
- In general:

  $$\gamma_k = \text{Cov}\{Y_t, Y_{t-k}\} = \theta^k \sigma^2 (1-\theta^2)^{-1},\ k = 0,\ 1, \ldots$$

  depends upon k, not upon t!

MA(1)-Process

States the dependence structure between consecutive observations as

$$Y_t = \mu + \varepsilon_t + \alpha\varepsilon_{t-1}$$

with ε_t: white noise, $V\{\varepsilon_t\} = \sigma^2$

Moving average process of order 1

$$E\{Y_t\} = \mu$$

Autocovariances $\gamma_k = Cov\{Y_t, Y_{t-k}\}$

- k=0: $\gamma_0 = V\{Y_t\} = \sigma^2(1+\alpha^2)$
- k=1: $\gamma_1 = Cov\{Y_t, Y_{t-1}\} = \alpha\sigma^2$
- $\gamma_k = 0$ for k = 2, 3, …
- Depends upon k, not upon t!

AR-Representation of MA-Process

The AR(1) can be represented as MA-process of infinite order

$$y_t = \theta y_{t-1} + \varepsilon_t = \Sigma^{\infty}_{i=0}\ \theta^i\ \varepsilon_{t-i}$$

given that $|\theta| < 1$

Similarly, the AR representation of the MA(1) process

$$y_t = \alpha y_{t-1} - \alpha^2 y_{t-2} + \ldots\ \varepsilon_t = \Sigma^{\infty}_{i=0}\ (-1)^i\ \alpha^{i+1} y_{t-i-1} + \varepsilon_t$$

given that $|\alpha| < 1$

Contents

Stationary Processes

Refers to the joint distribution of Y_t's, in particular to second moments

A process is called strictly stationary if its stochastic properties are unaffected by a change of the time origin

- The joint probability distribution at any set of times is not affected by an arbitrary shift along the time axis

Covariance function:

$$\gamma_{t,k} = \mathrm{Cov}\{Y_t, Y_{t+k}\},\ k = 0, \ 1, \ldots$$

Properties:

$$\gamma_{t,k} = \gamma_{t,-k}$$

Weak stationary process:

$$E\{Y_t\} = \mu \text{ for all } t$$

$$\mathrm{Cov}\{Y_t, Y_{t+k}\} = \gamma_k,\ k = 0, \ 1, \ldots \text{ for all } t \text{ and all } k$$

Also called covariance stationary process

AC and PAC Function

Autocorrelation function (AC function, ACF)

Independent of the scale of Y

- For a stationary process:

  $$\rho_k = \mathrm{Corr}\{Y_t, Y_{t-k}\} = \gamma_k/\gamma_0,\ k = 0,\ \ 1, \ldots$$

- Properties:
  - $|\rho_k| \leq 1$
  - $\rho_k = \rho_{-k}$
  - $\rho_0 = 1$
- Correlogram: graphical presentation of the AC function

Partial autocorrelation function (PAC function, PACF):

$$\theta_{kk} = \mathrm{Corr}\{Y_t,\ Y_{t-k}|Y_{t-1},\ldots,Y_{t-k+1}\},\ k = 0,\ \ 1, \ldots$$

- θ_{kk} is obtained from $Y_t = \theta_{k0} + \theta_{k1} Y_{t-1} + \ldots + \theta_{kk} Y_{t-k}$
- Partial correlogram: graphical representation of the PAC function

AC and PAC Function: Examples

Examples for the AC and PAC functions

- White noise

  $\rho_0 = \theta_{00} = 1$

  $\rho_k = \theta_{kk} = 0$, if $k \neq 0$
- AR(1) process, $Y_t = \delta + \theta Y_{t-1} + \varepsilon_t$

  $\rho_k = \theta^k$, $k = 0, \quad 1, \ldots$

  $\theta_{00} = 1$, $\theta_{11} = \theta$, $\theta_{kk} = 0$ for $k > 1$
- MA(1) process, $Y_t = \mu + \varepsilon_t + \alpha\varepsilon_{t-1}$

  $\rho_0 = 1$, $\rho_1 = -\alpha/(1 + \alpha^2)$, $\rho_k = 0$ for $k > 1$

  PAC function: damped exponential if $\alpha > 0$, otherwise alternating and damped exponential

AC and PAC Function: Estimates

- Estimator for the AC function ρ_k:

$$r_k = \frac{\sum_t (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_t (y_t - \bar{y})^2}$$

- Estimator for the PAC function θ_{kk}: coefficient of Y_{t-k} in the regression of Y_t on $Y_{t-1}, \ldots, Y_{t-k}$

AR(1) Processes, Verbeek, Fig. 8.1

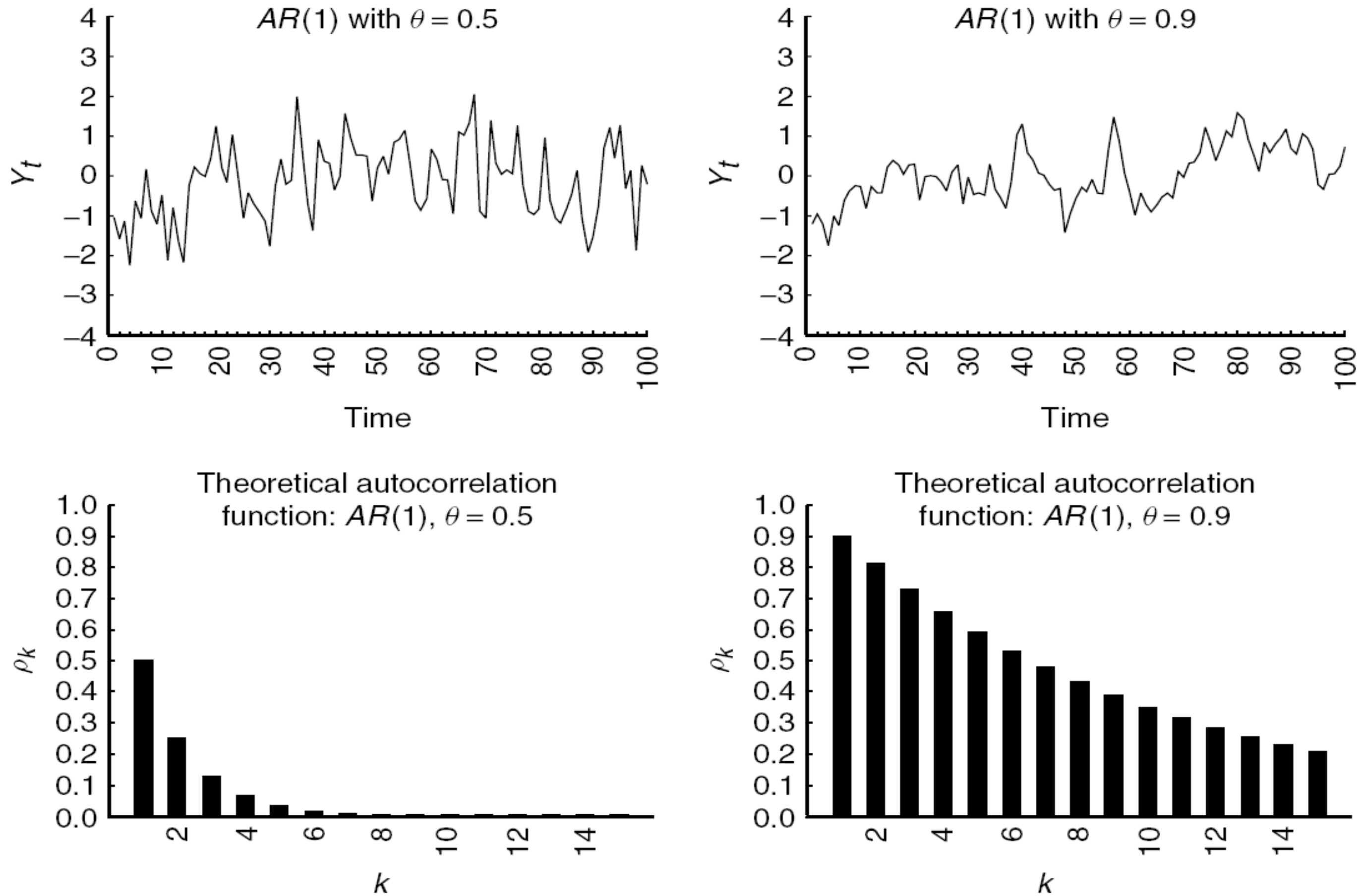


Figure 8.1 First-order autoregressive processes: data series and autocorrelation functions

MA(1) Processes, Verbeek, Fig. 8.2

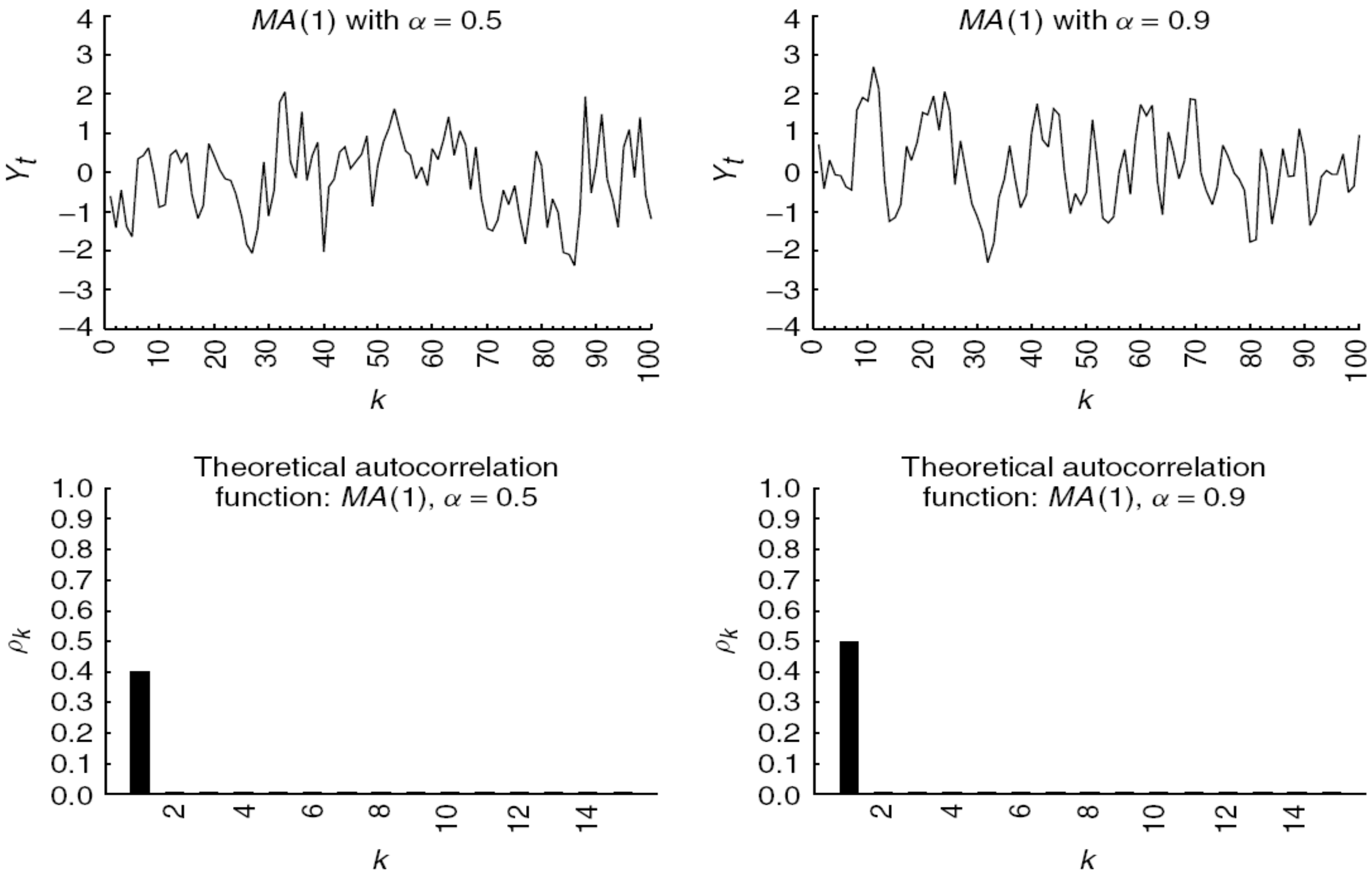


Figure 8.2 First-order moving average processes: data series and autocorrelation functions

Contents

The ARMA(p,q) Process

Generalization of the AR and MA processes: ARMA(p,q) process

$$y_t = \theta_1 y_{t-1} + \ldots + \theta_p y_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \ldots + \alpha_q \varepsilon_{t-q}$$

with white noise ε_t

Lag (or shift) operator L ($Ly_t = y_{t-1}$, $L^0 y_t = Iy_t = y_t$, $L^p y_t = y_{t-p}$)

ARMA(p,q) process in operator notation

$$\theta(L)y_t = \alpha(L)\varepsilon_t$$

with operator polynomials $\theta(L)$ and $\alpha(L)$

$$\theta(L) = I - \theta_1 L - \ldots - \theta_p L^p$$

$$\alpha(L) = I + \alpha_1 L + \ldots + \alpha_q L^q$$

Lag Operator

Lag (or shift) operator L

- $Ly_t = y_{t-1}$, $L^0 y_t = Iy_t = y_t$, $L^p y_t = y_{t-p}$
- Algebra of polynomials in L like algebra of variables

Examples:

- $(I - \phi_1 L)(I - \phi_2 L) = I - (\phi_1 + \phi_2)L + \phi_1\phi_2 L^2$
- $(I - \theta L)^{-1} = \Sigma^{\infty}_{i=0}\theta^i L^i$
- MA(∞) representation of the AR(1) process

  $$y_t = (I - \theta L)^{-1}\varepsilon_t$$

  the infinite sum defined only (e.g., finite variance) $|\theta| < 1$
- MA(∞) representation of the ARMA(p,q) process

  $$y_t = [\theta\,(L)]^{-1}\alpha(L)\varepsilon_t$$

  similarly the AR(∞) representations; invertibility condition: restrictions on parameters

Invertibility of Lag Polynomials

Invertibility condition for $I - \theta L$: $|\theta| < 1$

Invertibility condition for $I - \theta_1 L - \theta_2 L^2$:

- $\theta(L) = I - \theta_1 L - \theta_2 L^2 = (I - \phi_1 L)(I - \phi_2 L)$ with $\phi_1 + \phi_2 = \theta_1$ and $-\phi_1\phi_2 = \theta_2$
- Invertibility conditions: both $(I - \phi_1 L)$ and $(I - \phi_2 L)$ invertible; $|\phi_1| < 1$, $|\phi_2| < 1$
- Characteristic equation: $\theta(z) = (1 - \phi_1 z)(1 - \phi_2 z) = 0$
- Characteristic roots: solutions z_1, z_2 from $(1 - \phi_1 z)(1 - \phi_2 z) = 0$
- Invertibility conditions: $|z_1| > 1$, $|z_2| > 1$

Can be generalized to lag polynomials of higher order

Unit root: a characteristic root of value 1

- Polynomial $\theta(z)$ evaluated at $z = 1$: $\theta(1) = 0$, if $\Sigma_i \theta_i = 1$
- Simple check, no need to solve characteristic equation

Contents

Types of Trend

Trend: The expected value of a process Y_t increases or decreases with time

- Deterministic trend: a function $f(t)$ of the time, describing the evolution of $E\{Y_t\}$ over time

  $$Y_t = f(t) + \varepsilon_t,\ \varepsilon_t\text{: white noise}$$

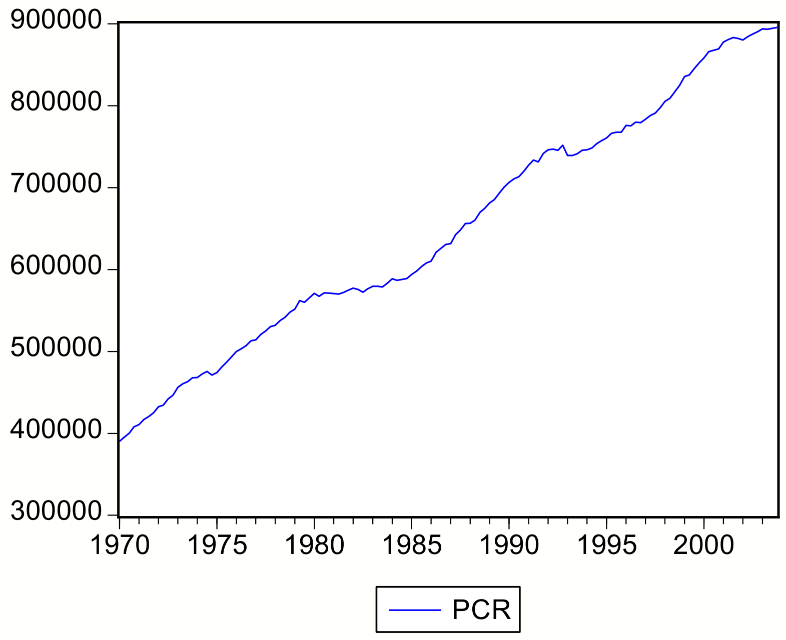
  Example: $Y_t = \alpha + \beta t + \varepsilon_t$ describes a linear trend of Y; an increasing trend corresponds to $\beta > 0$
- Stochastic trend: $Y_t = \delta + Y_{t-1} + \varepsilon_t$ or

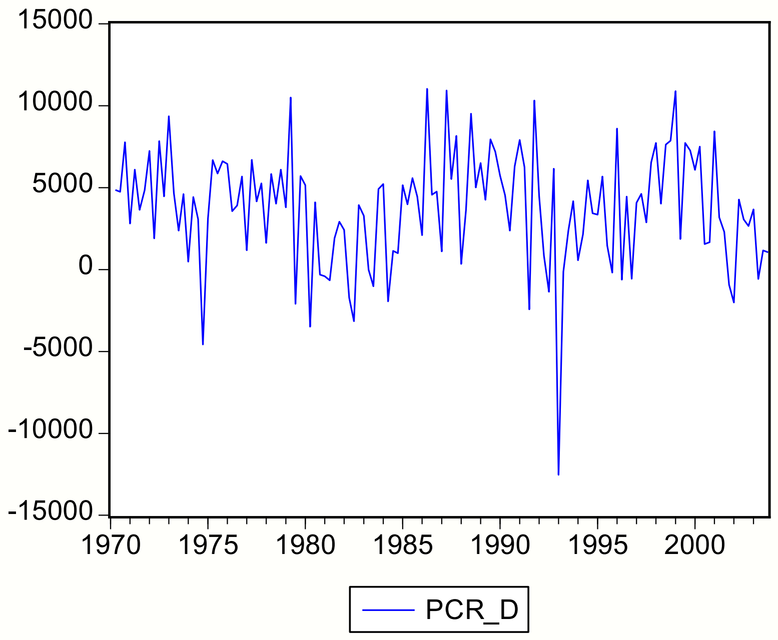
  $$\Delta Y_t = Y_t - Y_{t-1} = \delta + \varepsilon_t,\ \varepsilon_t\text{: white noise}$$

  - describes an irregular or random fluctuation of the differences ΔY_t around the expected value δ
  - AR(1) – or AR(p) – process with unit root
  - "random walk with trend"

Example: Private Consumption

Private consumption, AWM database; level values (PCR) and first differences (PCR_D)





Mean of PCD_D: 3740

Trends: Random Walk and AR Process

Random walk: $Y_t = Y_{t-1} + \varepsilon_t$; random walk with trend: $Y_t = 0.1 + Y_{t-1} + \varepsilon_t$; AR(1) process: $Y_t = 0.2 + 0.7 Y_{t-1} + \varepsilon_t$; ε_t simulated from $N(0,1)$

Random Walk with Trend

The random walk with trend $Y_t = \delta + Y_{t-1} + \varepsilon_t$ can be written as

$$Y_t = Y_0 + \delta t + \Sigma_{i \le t}\, \varepsilon_i$$

δ: trend parameter

Components of the process

- Deterministic growth path $Y_0 + \delta t$
- Cumulative errors $\Sigma_{i \le t}\, \varepsilon_i$

Properties:

- Expectation $Y_0 + \delta t$ is not a fixed value!
- $V\{Y_t\} = \sigma^2 t$ becomes arbitrarily large!
- $\mathrm{Corr}\{Y_t, Y_{t-k}\} = \sqrt{(1-k/t)}$
- Non-stationarity

Random Walk with Trend, cont’d

From

$$Corr(Y_t, Y_{t-k}) = \sqrt{\frac{t-k}{t}}$$

follows

- For fixed k, Y_t and Y_{t-k} are the stronger correlated, the larger t
- With increasing k, correlation tends to zero, but the slower the larger t (long memory property)

Comparison of random walk with the AR(1) process $Y_t = \delta + \theta Y_{t-1} + \varepsilon_t$

- AR(1) process: ε_{t-i} has the lesser weight, the larger i
- AR(1) process similar to random walk when θ is close to one

Non-Stationarity: Consequences

AR(1) process $Y_t = \theta Y_{t-1} + \varepsilon_t$

- OLS Estimator for θ:

$$\hat{\theta} = \frac{\sum_t y_t y_{t-1}}{\sum_t y_t^2}$$

- For |θ| < 1: the estimator is
  - Consistent
  - Asymptotically normally distributed
- For θ = 1 (unit root)
  - θ is underestimated
  - Estimator not normally distributed
  - Spurious regression problem

Spurious Regression

Random walk without trend: $Y_t = Y_{t-1} + \varepsilon_t$, ε_t: white noise

- Realization of Y_t: is a non-stationary process, stochastic trend?
- $V\{Y_t\}$: a multiple of t

Specified model: $Y_t = \alpha + \beta t + \varepsilon_t$

- Deterministic trend
- Constant variance
- Misspecified model!

Consequences for OLS estimator for β

- t- and F-statistics: wrong critical limits, rejection probability too large
- R^2 indicates explanatory potential although Y_t random walk without trend
- Granger & Newbold, 1974

Contents

How to Model Trends?

Specification of

- Deterministic trend, e.g., $Y_t = \alpha + \beta t + \varepsilon_t$: risk of wrong decisions
- Stochastic trend: analysis of differences ΔY_t if a random walk, i.e., a unit root, is suspected

Consequences of spurious regression are more serious

Consequences of modeling differences ΔY_t:

- Autocorrelated errors
- Consistent estimators
- Asymptotically normally distributed estimators
- HAC correction of standard errors

Elimination of a Trend

In order to cope with non-stationarity

- Trend-stationary process: the process can be transformed in a stationary process by subtracting the deterministic trend
- Difference-stationary process, or integrated process: stationary process can be derived by differencing

Integrated process: stochastic process Y is called

- integrated of order one if the first differences yield a stationary process: $Y \sim I(1)$
- integrated of order d, if the d-fold differences yield a stationary process: $Y \sim I(d)$

Trend-Elimination: Examples

Random walk $Y_t = \delta + Y_{t-1} + \varepsilon_t$ with white noise ε_t

$$\Delta Y_t = Y_t - Y_{t-1} = \delta + \varepsilon_t$$

- ΔY_t is a stationary process
- A random walk is a difference-stationary or $I(1)$ process

Linear trend $Y_t = \alpha + \beta t + \varepsilon_t$

- Subtracting the trend component $\alpha + \beta t$ provides a stationary process
- Y_t is a trend-stationary process

Integrated Stochastic Processes

Random walk $Y_t = \delta + Y_{t-1} + \varepsilon_t$ with white noise ε_t is a difference-stationary or $I(1)$ process

Many economic time series show stochastic trends

From the AWM Database

	Variable	***d***
YER	GDP, real	1
PCR	Consumption, real	1-2
PYR	Household's Disposable Income, real	1-2
PCD	Consumption Deflator	2

ARIMA(p,d,q) process: d-th differences follow an ARMA(p,q) process

Contents

Unit Root Tests

AR(1) process $Y_t = \delta + \theta Y_{t-1} + \varepsilon_t$ with white noise ε_t

- Dickey-Fuller or DF test (Dickey & Fuller, 1979)

  Test of H_0: $\theta = 1$ against H_1: $\theta < 1$
- KPSS test (Kwiatkowski, Phillips, Schmidt & Shin, 1992)

  Test of H_0: $\theta < 1$ against H_1: $\theta = 1$
- Augmented Dickey-Fuller or ADF test

  extension of DF test
- Various modifications like Phillips-Perron test, Dickey-Fuller GLS test, etc.

Dickey-Fuller‘s Unit Root Test

AR(1) process $Y_t = \delta + \theta Y_{t-1} + \varepsilon_t$ with white noise ε_t

OLS Estimator for θ:

$$\hat{\theta} = \frac{\sum_t y_t y_{t-1}}{\sum_t y_t^2}$$

Distribution of *DF*

$$DF = \frac{\hat{\theta} - 1}{s_{\hat{\theta}}}$$

- If $|\theta| < 1$: approximately $t(T-1)$
- If $\theta = 1$: Dickey & Fuller critical values

DF test for testing H_0: $\theta = 1$ against H_1: $\theta < 1$

- $\theta = 1$: characteristic polynomial has unit root

Dickey-Fuller Critical Values

Monte Carlo estimates of critical values for

- DF_0: Dickey-Fuller test without intercept
- DF: Dickey-Fuller test with intercept
- DF_τ: Dickey-Fuller test with time trend

T		p = 0.01	p = 0.05	p = 0.10
25	DF_0	-2.66	-1.95	-1.60
	DF	-3.75	-3.00	-2.63
	DF_τ	-4.38	-3.60	-3.24
100	DF_0	-2.60	-1.95	-1.61
	DF	-3.51	-2.89	-2.58
	DF_τ	-4.04	-3.45	-3.15
N(0,1)		-2.33	-1.65	-1.28

Unit Root Test: The Practice

AR(1) process $Y_t = \delta + \theta Y_{t-1} + \varepsilon_t$ with white noise ε_t

can be written with $\pi = \theta - 1$ as

$$\Delta Y_t = \delta + \pi Y_{t-1} + \varepsilon_t$$

DF tests H_0: $\pi = 0$ against H_1: $\pi < 0$

test statistic for testing $\pi = \theta - 1 = 0$ identical with *DF* statistic

$$DF = \frac{\hat{\theta} - 1}{se(\hat{\theta})} = \frac{\hat{\pi}}{se(\hat{\pi})}$$

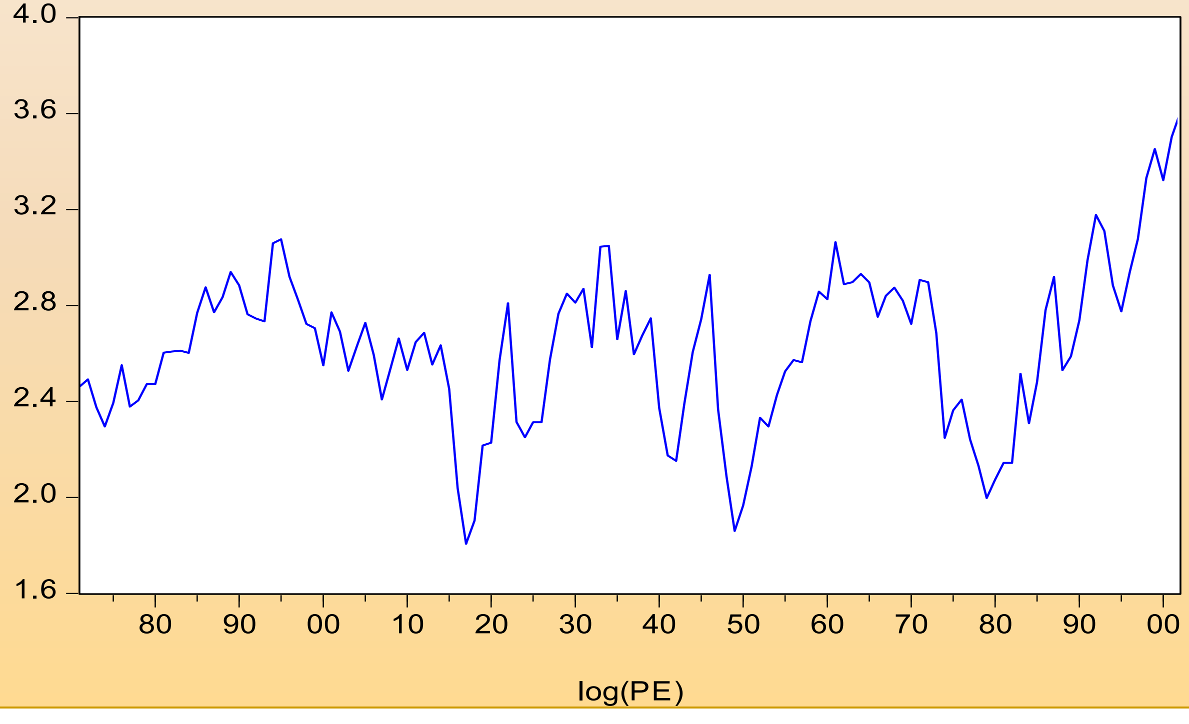
Two steps:

1. Regression of ΔY_t on Y_{t-1}: OLS-estimator for $\pi = \theta - 1$
2. Test of H_0: $\pi = 0$ against H_1: $\pi < 0$ based on *DF*; critical values of Dickey & Fuller

Example: Price/Earnings Ratio

Verbeek's data set PE: annual time series data on composite stock price and earnings indices of the S&P500, 1871-2002

- PE: price/earnings ratio
  - Mean 14.6
  - Min 6.1
  - Max 36.7
  - St.Dev. 5.1
- Log(PE)
  - Mean 2.63
  - Min 1.81
  - Max 3.60
  - St.Dev. 0.33



Price/Earnings Ratio, cont'd

Fitting an AR(1) process to the log PE ratio data gives:

$$\Delta Y_t = 0.335 - 0.125 Y_{t-1}$$

with *t*-statistic -2.569 (Y_{t-1}) and *p*-value 0.1021

- *p*-value of the DF statistic (-2.569): 0.102
  - 1% critical value: -3.48
  - 5% critical value: -2.88
  - 10% critical value: -2.58
- H_0: θ = 1 (non-stationarity) cannot be rejected for the log PE ratio

Unit root test for first differences: DF statistic -7.31, *p*-value 0.000 (1% critical value: -3.48)

- log PE ratio is $I(1)$

However: for sample 1871-1990: DF statistic -3.65, *p*-value 0.006

Unit Root Test: Extensions

DF test so far for a model with intercept: $\Delta Y_t = \delta + \pi Y_{t-1} + \varepsilon_t$

Tests for alternative or extended models

- DF test for model without intercept: $\Delta Y_t = \pi Y_{t-1} + \varepsilon_t$
- DF test for model with intercept and trend: $\Delta Y_t = \delta + \gamma t + \pi Y_{t-1} + \varepsilon_t$

DF tests in all cases H_0: $\pi = 0$ against H_1: $\pi < 0$

Test statistic in all cases

$$DF =$$

Critical values depend on cases; cf. Table on slide 42

KPSS Test

A process $Y_t = \delta + \varepsilon_t$ with white noise ε_t

- Test of H_0: no unit root (Y_t is stationary), against H_1: $Y_t \sim I(1)$
- Under H_0:
  - Average $\dot{y}$ is a consistent estimate of δ
  - Long-run variance of ε_t is a well-defined number
- KPSS (Kwiatkowski, Phillips, Schmidt, Shin) test statistic

$$KPSS = \sum$$

  with $S_t^2 = \Sigma_i^t e_i$ and the variance estimate s^2 of the residuals $e_i = Y_t - \dot{y}$
- Bandwidth or lag truncation parameter m for estimating s^2

$$s^2 = \sum \hat{}$$

- Critical values from Monte Carlo simulations

ADF Test

Extended model according to an AR(p) process:

$$\Delta Y_t = \delta + \pi Y_{t-1} + \beta_1 \Delta Y_{t-1} + \ldots + \beta_p \Delta Y_{t-p+1} + \varepsilon_t$$

Example: AR(2) process $Y_t = \delta + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \varepsilon_t$ can be written as

$$\Delta Y_t = \delta + (\theta_1 + \theta_2 - 1) Y_{t-1} - \theta_2 \Delta Y_{t-1} + \varepsilon_t$$

the characteristic equation $(1 - \phi_1 L)(1 - \phi_2 L) = 0$ has roots $\theta_1 = \phi_1 + \phi_2$ and $\theta_2 = - \phi_1 \phi_2$

a unit root implies $\phi_1 = \theta_1 + \theta_2 = 1$:

Augmented DF (ADF) test

- Test of H_0: $\pi = 0$ against H_1: $\pi < 0$
- Needs its own critical values
- Extensions (intercept, trend) similar to the DF-test
- Phillips-Perron test: alternative method; uses HAC-corrected standard errors

Price/Earnings Ratio, cont'd

Extended model according to an AR(2) process gives:

$$\Delta Y_t = 0.366 - 0.136 Y_{t-1} + 0.152 \Delta Y_{t-1} - 0.093 \Delta Y_{t-2}$$

with t-statistics -2.487 (Y_{t-1}), 1.667 (ΔY_{t-1}) and -1.007 (ΔY_{t-2}) and p-values 0.119, 0.098 and 0.316

- p-value of the DF statistic 0.121
  - 1% critical value: -3.48
  - 5% critical value: -2.88
  - 10% critical value: -2.58
- Non-stationarity cannot be rejected for the log PE ratio

Unit root test for first differences: DF statistic -7.31, p-value 0.000 (1% critical value: -3.48)

- log PE ratio is $I(1)$

However: for sample 1871-1990: DF statistic -3.52, p-value 0.009

Unit Root Tests in GRETL

For marked variable:

- Variable > Unit root tests > Augmented Dickey-Fuller test

  Performs the
  - DL test (choose zero for “lag order for ADL test”) or the
  - ADL test
  - with or without constant, trend, squared trend
- Variable > Unit root tests > ADF-GLS test

  Performs the
  - DL test (choose zero for “lag order for ADL test”) or the
  - ADL test
  - with or without a trend, which are estimated by GLS
- Variable > Unit root tests > KPSS test

  Performs the KPSS test with or without a trend

Contents

ARMA Models: Application

Application of the ARMA(p,q) model in data analysis: Three steps

1. Model specification, i.e., choice of p, q (and d if an ARIMA model is specified)
2. Parameter estimation
3. Diagnostic checking

Estimation of ARMA Models

The estimation methods are

- OLS estimation
- ML estimation

AR models: the explanatory variables are

- Lagged values of the explained variable Y_t
- Uncorrelated with error term ε_t
- OLS estimation

MA Models: OLS Estimation

MA models:

- Minimization of sum of squared deviations is not straightforward
- E.g., for an MA(1) model, $S(\mu,\alpha) = \Sigma_t[Y_t - \mu - \alpha\Sigma_{j=0}(-\alpha)^j(Y_{t-j-1} - \mu)]^2$
  - $S(\mu,\alpha)$ is a nonlinear function of parameters
  - Needs Y_{t-j-1} for j=0,1,…, i.e., historical Y_s, $s < 0$
- Approximate solution from minimization of

$$S^*(\mu,\alpha) = \Sigma_t[Y_t - \mu - \alpha\Sigma_{j=0}^{t-2}(-\alpha)^j(Y_{t-j-1} - \mu)]^2$$

- Nonlinear minimization, grid search

ARMA models combine AR part with MA part

ML Estimation

Assumption of normally distributed ε_t

Log likelihood function, conditional on initial values

$$\log L(\alpha,\theta,\mu,\sigma^2) = -(T-1)\log(2\pi\sigma^2)/2 - (1/2)\, \Sigma_t\, \varepsilon_t^2/\sigma^2$$

ε_t are functions of the parameters

- AR(1): $\varepsilon_t = y_t - \theta_1 y_{t-1}$
- MA(1): $\varepsilon_t = \Sigma_{j=0}^{t-1}(-\alpha)^j y_{t-j}$

Initial values: y_1 for AR, $\varepsilon_0 = 0$ for MA

- Extension to exact ML estimator
- Again, estimation for AR models easier
- ARMA models combine AR part with MA part

Model Specification

Based on the

- Autocorrelation function (ACF)
- Partial Autocorrelation function (PACF)

Structure of AC and PAC functions typical for AR and MA processes

Example:

- MA(1) process: $\rho_0 = 1$, $\rho_1 = \alpha/(1-\alpha^2)$; $\rho_i = 0$, $i = 2, 3, \ldots$; $\theta_{kk} = \alpha^k$, $k = 0, 1, \ldots$
- AR(1) process: $\rho_k = \theta^k$, $k = 0, 1, \ldots$; $\theta_{00} = 1$, $\theta_{11} = \theta$, $\theta_{kk} = 0$ for $k > 1$

Empirical ACF and PACF give indications on the process underlying the time series

ARMA(p,q)-Processes

Condition for	**AR(p)** $\theta(L)Y_t = \varepsilon_t$	**MA(q)** $Y_t = \alpha(L)\varepsilon_t$	**ARMA(p,q)** $\theta(L)Y_t=\alpha(L)\varepsilon_t$
Stationarity	roots z_i of $\theta(z)=0$: $\lvert z_i\rvert > 1$	always stationary	roots z_i of $\theta(z)=0$: $\lvert z_i\rvert > 1$
Invertibility	always invertible	roots z_i of $\alpha(z)=0$: $\lvert z_i\rvert > 1$	roots z_i of $\alpha(z)=0$: $\lvert z_i\rvert > 1$
AC function	damped, infinite	$\rho_k = 0$ for $k > q$	damped, infinite
PAC function	$\theta_{kk} = 0$ for $k > p$	damped, infinite	damped, infinite

Empirical AC and PAC Function

Estimation of the AC and PAC functions

AC ρ_k:

$$r_k = \frac{\sum_t (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_t (y_t - \bar{y})^2}$$

PAC θ_{kk}: coefficient of Y_{t-k} in regression of Y_t on $Y_{t-1}, \ldots, Y_{t-k}$

MA(q) process: standard errors for r_k, $k > q$, from

$$\sqrt{T}(r_k - \rho_k) \rightarrow N(0, v_k)$$

with $v_k = 1 + 2\rho_1^2 + \ldots + 2\rho_k^2$

- test of H_0: $\rho_1 = 0$: compare $\sqrt{T} r_1$ with critical value from N(0,1), etc.

AR(p) process: test of H_0: $\rho_k = 0$ for $k > p$ based on asymptotic distribution

$$\sqrt{T}\,\theta_{kk} \rightarrow N(0,1)$$

Diagnostic Checking

ARMA(*p*,*q*): Adequacy of choices *p* and *q*

Analysis of residuals from fitted model:

- Correct specification: residuals are realizations of white noise
- Box-Ljung Portmanteau test: for a ARMA(*p*,*q*) process

  $$Q_K = T(T+2)\sum_{k=1}^{K}\frac{1}{T-k}r_k^2$$

  follows the Chi-squared distribution with *K-p-q df*

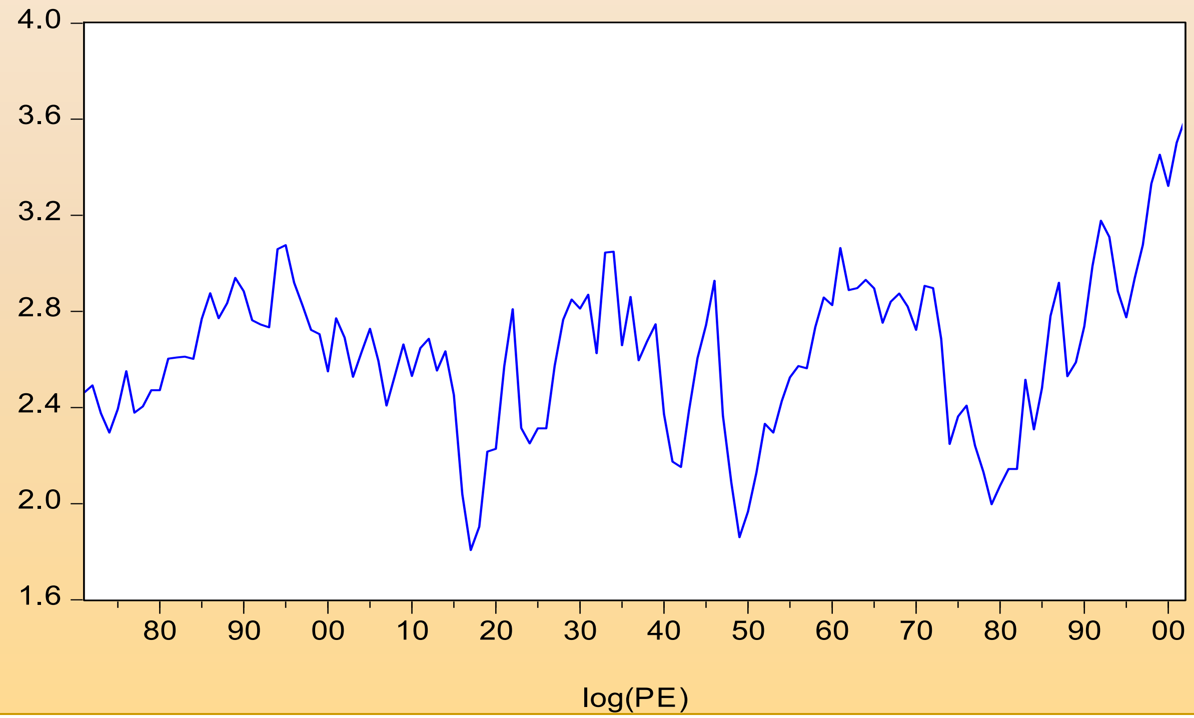
Overfitting

- Starting point: a general model
- Comparison with a model with reduced number of parameters: choose model with smallest *BIC* or *AIC*
- *AIC*: tends to result asymptotically in overparameterized models

Example: Price/Earnings Ratio

Data set PE: PE = price/earnings, LOGPE = log(PE)

- Log(PE)
  - Mean 2.63
  - Min 1.81
  - Max 3.60
  - Std 0.33



PE Ratio: AC and PAC Function

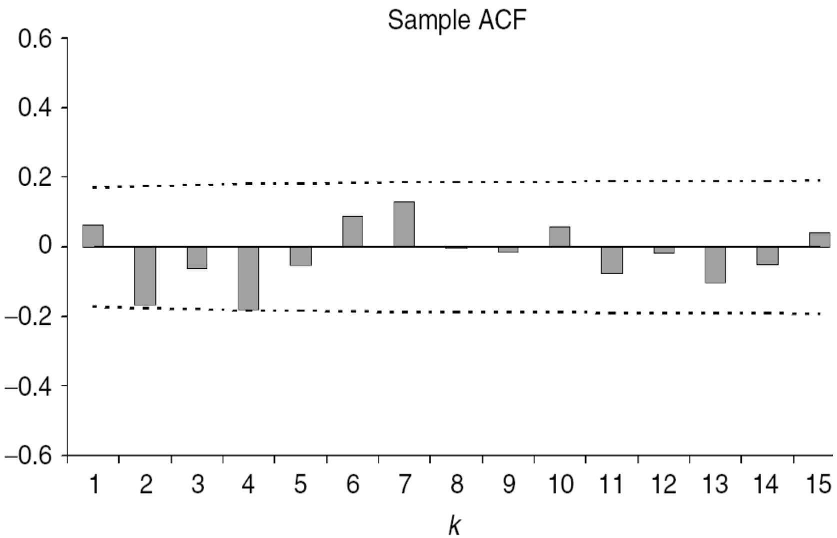


Figure 8.7 Sample autocorrelation function of log(P/E)

At level 0.05 significant values:

- ACF: $k = 4$
- PACF: $k = 2, 4$

suggests MA(4), but not very clear

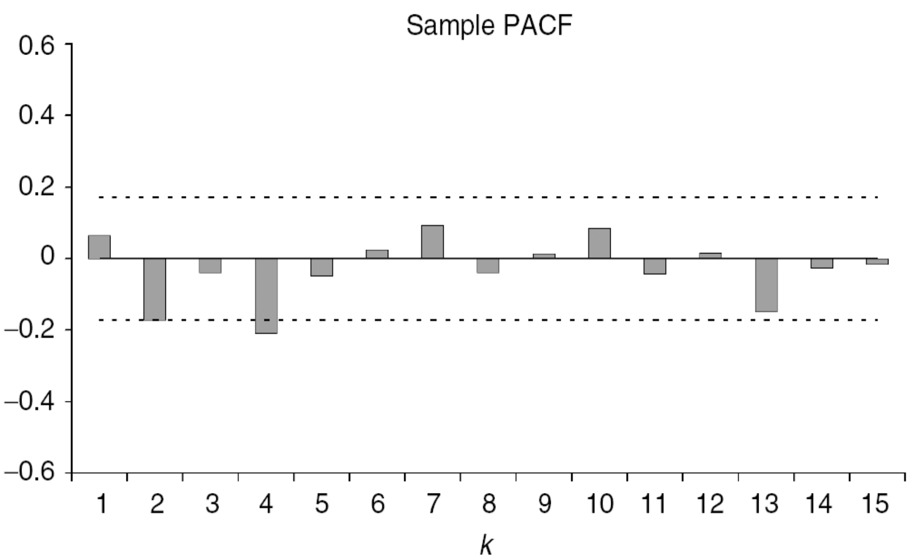


Figure 8.8 Sample partial autocorrelation function of log(P/E)

PE Ratio: MA (4) Model

MA(4) model for differences log PE_t - log PE_{t-1}

Function evaluations: 37
Evaluations of gradient: 11

Model 2: ARMA, using observations 1872-2002 (T = 131)
Estimated using Kalman filter (exact ML)
Dependent variable: d_LOGPE
Standard errors based on Hessian

	coefficient	std. error	t-ratio	p-value	
const	0,00804276	0,0104120	0,7725	0,4398	
theta_1	0,0478900	0,0864653	0,5539	0,5797	
theta_2	-0,187566	0,0913502	-2,053	0,0400	**
theta_3	-0,0400834	0,0819391	-0,4892	0,6247	
theta_4	-0,146218	0,0915800	-1,597	0,1104	

Mean dependent var	0,008716	S.D. dependent var	0,181506
Mean of innovations	-0,000308	S.D. of innovations	0,174545
Log-likelihood	42,69439	Akaike criterion	-73,38877
Schwarz criterion	-56,13759	Hannan-Quinn	-66,37884

PE Ratio: AR(4) Model

AR(4) model for differences log PE_t - log PE_{t-1}

Function evaluations: 36
Evaluations of gradient: 9

Model 3: ARMA, using observations 1872-2002 (T = 131)
Estimated using Kalman filter (exact ML)
Dependent variable: d_LOGPE
Standard errors based on Hessian

	coefficient	std. error	t-ratio	p-value	
const	0,00842210	0,0111324	0,7565	0,4493	
phi_1	0,0601061	0,0851737	0,7057	0,4804	
phi_2	-0,202907	0,0856482	-2,369	0,0178	**
phi_3	-0,0228251	0,0853236	-0,2675	0,7891	
phi_4	-0,206655	0,0850843	-2,429	0,0151	**

Mean dependent var	0,008716	S.D. dependent var	0,181506
Mean of innovations	-0,000315	S.D. of innovations	0,173633
Log-likelihood	43,35448	Akaike criterion	-74,70896
Schwarz criterion	-57,45778	Hannan-Quinn	-67,69903

PE Ratio: Various Models

Diagnostics for various competing models: Δy_t = log PE_t - log PE_{t-1}

Best fit for

- BIC: MA(2) model $\Delta y_t = 0.008 + e_t - 0.250\, e_{t-2}$
- AIC: AR(2,4) model $\Delta y_t = 0.008 - 0.202\, \Delta y_{t-2} - 0.211\, \Delta y_{t-4} + e_t$

Model	Lags	AIC	BIC	Q_{12}	*p*-value
MA(4)	1–4	-73.389	-56.138	5.03	0.957
AR(4)	1–4	-74.709	-57.458	3.74	0.988
MA	2, 4	-76.940	-65.440	5.48	0.940
AR	2, 4	**-78.057**	-66.556	4.05	0.982
MA	2	-76.072	**-67.447**	9.30	0.677
AR	2	-73.994	-65.368	12.12	0.436

Time Series Models in GRETL

Variable > Unit root tests > (a) Augmented Dickey-Fuller test, (b) ADL-GLS test, (c) KPSS test

a) DF test or ADL test with or without constant, trend and squared trend
b) DF test or ADL test with or without trend, GLS estimation for demeaning and detrending
c) KPSS (Kwiatkowski, Phillips, Schmidt, Shin) test

Model > Time Series > ARIMA

- Estimates an ARMA model, with or without exogenous regressors

Your Homework

1. Use Verbeek's data set INCOME (quarterly data for the total disposable income and for consumer expenditures for 1/1971 to 2/1985 in the UK) and answer the questions a., b., c., d., e., and f. of Exercise 8.3 of Verbeek. Confirm your finding in question c. using the KPSS test.
2. For the AR(2) model $y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \varepsilon_t$, show that (a) the model can be written as $\Delta y_t = \delta y_{t-1} - \theta_2 \Delta y_{t-1} + \varepsilon_t$ with $\delta = \theta_1 + \theta_2 - 1$, and that (b) $\theta_1 + \theta_2 = 1$ corresponds to a unit root of the characteristic equation $\theta(z) = 1 - \theta_1 z - \theta_2 z^2 = 0$.