Mile stone II

Multivariate time series

- Spurios regression and Cointegration
- Modeling cross- and autocorrelation
- Stationary VAR-Models
- Non-stationary VAR-Models

Describing uncertainty

- Econometric models often are based on the ,,first" and ,,second" moments of the conditional distribution $p(Y|X)$ by specify the expectation and the variance of Y under the assumption that X is known, i.e. $\mathrm{E}(Y|X)$ and $\mathrm{Var}(Y|X)$

- Alternatively, specify the first and second moments of the joint distribution $p(X, Y)$ of the random vector $(X, Y)'$ through the expectation vector and the covariance matrix:

$$\begin{pmatrix} \mathrm{E}(X) \\ \mathrm{E}(Y) \end{pmatrix}, \qquad \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X, Y) \\ \mathrm{Cov}(X, Y) & \mathrm{Var}(Y) \end{pmatrix}.$$

II.1 Spurios regression and Cointegration

Consider a regression model where both the response Y_t and the predictor X_t are time series:

$$Y_t = \beta_0 + \beta_1 X_t + u_t. \tag{1}$$

If X_t and Y_t are stationary time series, then two cases have to be distinguished concerning u_t:

- Econometrics I: u_t is a white noise process, i.e. $\mathrm{E}(u_t) = 0$, $\mathrm{Var}(u_t) = \sigma^2$, $\mathrm{Corr}(u_t, u_s) = 0$ for all $s \neq t$ $\Rightarrow$ OLS estimation is BLUE

Spurios regression and Cointegration

- Econometrics II: u_t is a stationary process, $\mathrm{E}(u_t) = 0$, $\mathrm{Var}(u_t) = \sigma^2$, $\mathrm{Corr}(u_t, u_s) = \rho(t - s)$ for all $s \neq t$ (correlation depends on the lag $h = t - s$ between t and s) $\Rightarrow$ OLS is consistent, but inefficient; include ARMA terms

What happens, if the time series X_t and Y_t are non-stationary?

- u_t might be a non-stationary process $\Rightarrow$ spurios regression may occur

- u_t is stationary $\Rightarrow$ the time series are cointegrated

Stationarity versus non-stationarity

Examples from Econometrics II: assume that $u_t \sim \mathrm{Normal}\,(0, 2)$ is a white noise process

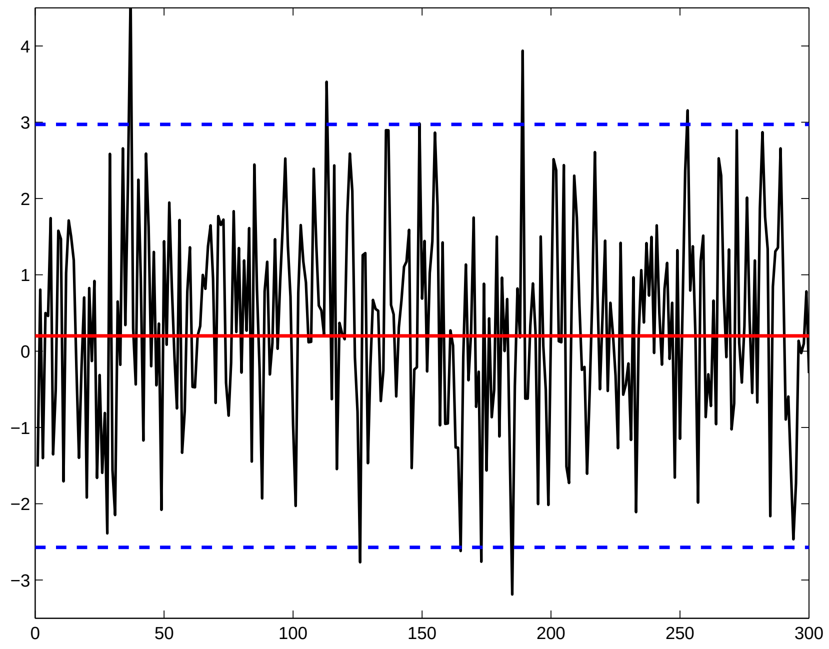
- Random process: $Y_t = 0.2 + u_t$ - stationary; no autocorrelation; $\mathrm{E}(Y_t) = \mathrm{E}(Y_t|Y_{t-1}) = 0.2$, $\mathrm{Var}(Y_t) = \mathrm{Var}(Y_t|Y_{t-1}) = 2$,

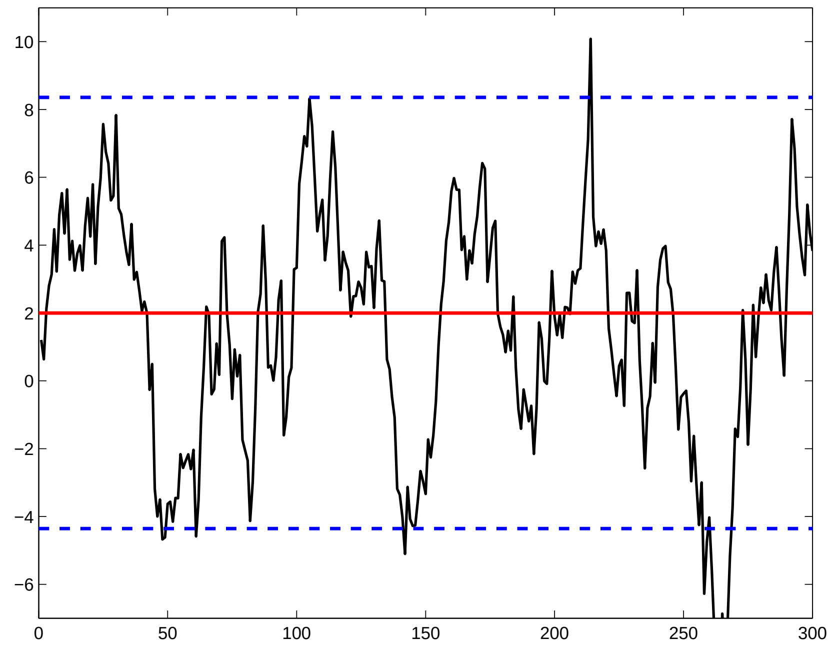
- AR(1)-process: $Y_t = 0.9Y_{t-1} + 0.2 + u_t$ - stationary ($\varphi = 0.9$ satisfies the stationarity condition $|\varphi| < 1$); autocorrelation: $\mathrm{Corr}(u_t, u_s) = \varphi^{t-s}$; $\mathrm{E}(Y_t|Y_{t-1}) = 0.9Y_{t-1} + 0.2 \neq \mathrm{E}(Y_t) = 2$; $\mathrm{Var}(Y_t|Y_{t-1}) = 2 \neq \mathrm{Var}(Y_t) = 10.53$.

- Random walk with drift: $Y_t = Y_{t-1} + 0.2 + u_t$ - non-stationary; $\mathrm{E}(Y_t|Y_0) = \mathrm{E}(Y_0) + 0.2t$; $\mathrm{Var}(Y_t|Y_0) = 2t$.

Stationarity versus non-stationarity

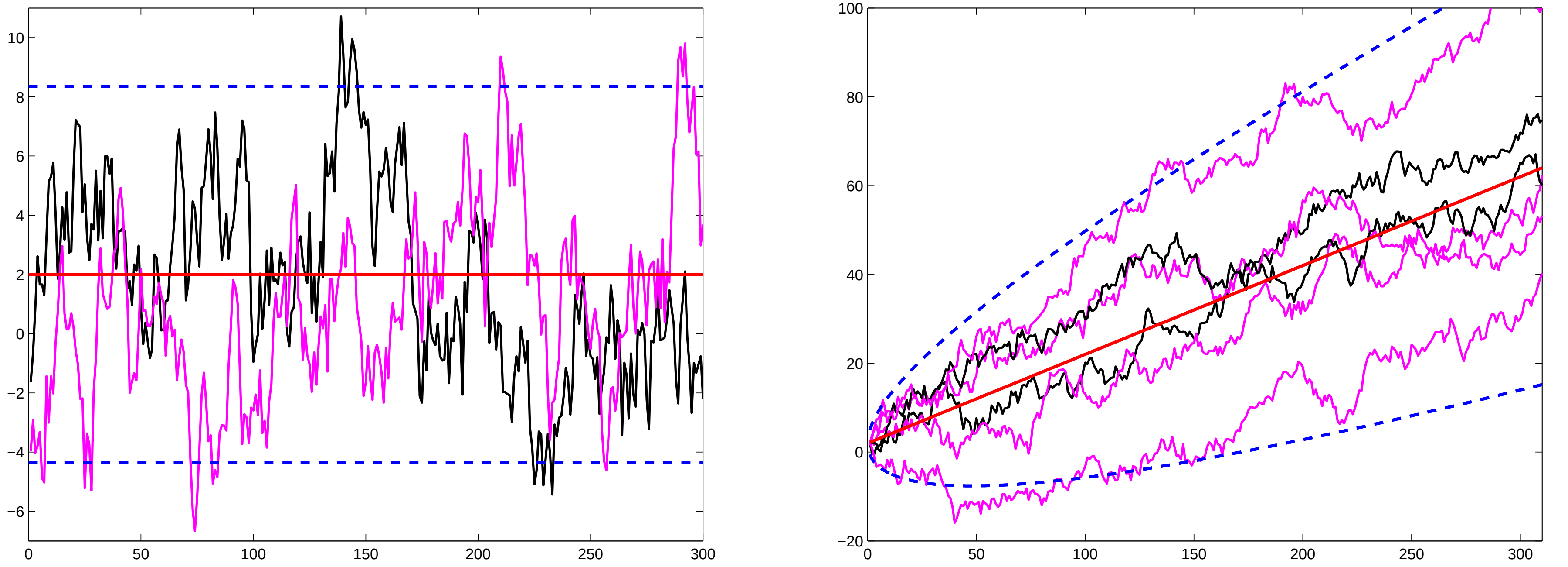
A random process versus a stationary AR process:





Stationarity versus non-stationarity

Stationary AR processes versus random walk processes:



First differences

Recall from Econometrics II:

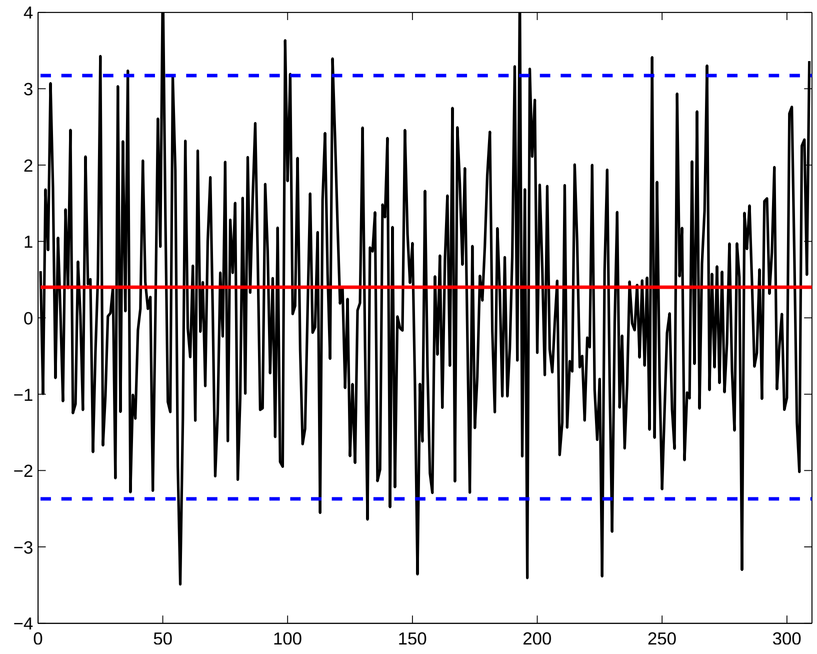
- A random walk with (or without drift) is an AR(1) process, where the AR(1)-coefficient is equal to 1 („unit root")
- Use the (augmented) Dickey-Fuller test to test the null hypothesis, that a time series Y_t is non-stationary
- The process of **first differences**, i.e.

$$\nabla Y_t = Y_t - Y_{t-1},$$

  is often stationary; Y_t is integrated of order 1 (I(1)-process).

First differences

A random walk process and the corresponding process of first differences



Spurios regression

If the response Y_t and the predictor X_t in a regression model are non-stationary time series, then spurios regression might be present.

- Spurios regression means that Y_t and X_t are independent, but the regression coefficient β_1 in regression model (2) is highly significant (large t-value, small p-value).

- With increasing number of observations T, the t-value might even converge to ∞, although the null hypothesis $\beta_1 = 0$ is true. The risk of rejecting a true null hypothesis $\beta_1 = 0$ may be considerably larger (up to 100%) than the assumed significance level of, say, 5%.

Spurios regression

- What is the reason? If the response Y_t and the predictor X_t are non-stationary time series, then the error term u_t in regression model (2) **might be** a non-stationary process.

- If the error term u_t in regression model (2) is a non-stationary process, then econometric inference for the regression parameters β_0 and β_1 may be misleading.

- However, the error term u_t in regression model (2) **could be** a stationary process, even if the response Y_t and the predictor X_t are non-stationary time series. In this case, the time series X_t and Y_t are called **cointegrated**.

Example: Regression involving random walk processes

Assume that $Y_t = Y_{t-1} + u_t^Y$ and $X_t = X_{t-1} + u_t^X$ follow independent random walk processes, i.e. u_t^Y and u_t^X are independent white noise processes. Rewrite regression model (2):

$$
\begin{aligned}
u_{t-1} &= Y_{t-1} - \beta_1 X_{t-1} - \beta_0, \\
u_t &= Y_t - \beta_1 X_t - \beta_0 = Y_{t-1} + u_t^Y - \beta_1 X_{t-1} - \beta_1 u_t^X - \beta_0 \\
\Rightarrow u_t &= u_{t-1} + u_t^Z,
\end{aligned}
$$

where $u_t^Z = u_t^Y - \beta_1 u_t^X$ is a white noise process (superposition of the white noise processes u_t^X and u_t^Y). **The error term u_t follows a random walk.**

How to deal with spurios regression?

- Check stationarity/non-stationarity for the response and all random predictors in a regression model.

- Check stationarity of the OLS residuals.

- Check for extremely large t-values in combination with a Durbin Watson statistic d close to 0 (remember $d \approx 2(1 - r_1)$, where r_1 is the autocorrelation of the residuals at lag 1).

If spurios regression seems to be present, then consider first differences instead of levels for all non-stationary variables.

Example: Regression involving random walk processes

Assume that $Y_t = Y_{t-1} + u_t^Y$ and $X_t = X_{t-1} + u_t^X$ follow independent random walk processes. Then

$$Y_t = \beta_0 + \beta_1 X_t + u_t, \qquad Y_{t-1} = \beta_0 + \beta_1 X_{t-1} + u_{t-1},$$

where u_t follows a random walk. Hence,

$$\begin{aligned} Y_t - Y_{t-1} &= \beta_1 (X_t - X_{t-1}) + (u_t - u_{t-1}), \\ \nabla Y_t &= \beta_1 \nabla X_t + \nabla u_t. \end{aligned} \tag{2}$$

Considering the first differences ∇X_t and ∇Y_t instead of the levels X_t and Y_t for the non-stationary variables leads to the regression model (2) with stationary error distribution $\nabla u_t = u_t - u_{t-1}$.

EVIEWS Exercise : Case Study `nasdaq2`

Monthly equity prices X_t and Y_t of two US firm listed at the NASDAQ (January 1992 to December 2011)

- Unit root hypothesis (including intercept) not rejected for $\log Y_t$ (p-value: 0.58) and $\log X_t$ (p-value: 0.39)
- OLS regression in the levels: significantly negative regression coefficient $\Rightarrow \mathrm{E}(\log Y_t | X_t) = -0.21 \cdot \log X_t - 0.74$
- Durbin-Watson statistics very small (≈ 0.1), ACF of the residuals: AR(1) coefficient close to 1 - spurios regression?

EVIEWS Exercise : Case Study `nasdaq2`

- Unit root hypothesis (no intercept) for the residual series not rejected (p-value: 0.14) $\Rightarrow$ spurios regression!

- Unit root hypothesis (including intercept) rejected for $\nabla \log Y_t$ (p-value: 0.0) and $\nabla \log X_t$ (p-value: 0.0): difference processes are stationary

- OLS regression in the differences: regression coefficient of X_t not significant (p-value: 0.967) $\Rightarrow \mathrm{E}(\nabla \log Y_t | X_t) = \mathrm{E}(\nabla \log Y_t)$

- F-test for the null hypothesis $\beta_0 = \beta_1 = 0$ not rejected (p-value: 0.967) $\Rightarrow \mathrm{E}(\nabla \log Y_t) = 0$; $\log Y_t$ is a random walk

Cointegration

If the response Y_t and the predictor X_t in regression model (2) are non-stationary time series, but the error term u_t is a stationary process, then the time series X_t and Y_t are called **cointegrated**.

A linear combination of X_t and Y_t is a stationary process (i.e. a I(0)-process):

$$u_t = Y_t - \beta_1 X_t - \beta_0 = \begin{pmatrix} -\beta_1 \\ 1 \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} - \beta_0$$

If X_t and Y_t are I(1)-processes, then a linear combination of X_t and Y_t is a I(0)-process, i.e. has a lower order of integration. The processes move together

Cointegration

Two random walk processes versus two cointegrated processes (the black process is the same in both figures)

Testing for Cointegration

If β_0 and β_1 were known, then we could compute $u_t = Y_t - \beta_1 X_t - \beta_0$ and perform a unit root test on u_t.

Engle-Granger two-step method:

- Run OLS regression with the levels Y_t and X_t and determine the OLS residuals $\hat{u}_t$.

- If the unit root hypothesis (no intercept) for the residual series $\hat{u}_t$ is rejected, then X_t and Y_t are co-integrated.

EVIEWS Exercise : Case Study `exchange-rates`

Daily exchange rate X_t of the Swedish krona against the Euro and Y_t of the Norwegian krone against the Euro (January 1, 2002 to April 4, 2012)

- Unit root hypothesis (including intercept) not rejected for $\log Y_t$ (p-value: 0.08) and $\log X_t$ (p-value: 0.17)
- OLS regression in the levels: significantly positive regression coefficient $\Rightarrow \mathrm{E}(\log Y_t|X_t) = 0.62 \cdot \log X_t + 0.7$
- Unit root hypothesis (no intercept) for the residual series rejected (p-value: 0.0047) $\Rightarrow$ time series are cointegrated

Cointegration

Some open issues:

- Which variables is the left hand side, which one is the right hand side variable?
- To forecasts future values $Y_{t+1}, Y_{t+2}, \ldots$, we need to forecast future values of $X_{t+1}, X_{t+2}, \ldots$.
- How to proceed, if we have more than two time series?

Modeling cross- and autocorrelation

- Multivariate time series $\mathbf{y}_t, t = 1, \ldots, T$: simultaneous modeling of more than one time series (z.B. GDP, industrial production, inflation)

- $\mathbf{y}_t$ is a realization of a multivariate stochastic process $\mathbf{Y}_t, t = 1, \ldots, T$:

$$\mathbf{y}_t = \begin{pmatrix} y_{1t} \\ \vdots \\ y_{mt} \end{pmatrix}, \qquad \mathbf{Y}_t = \begin{pmatrix} Y_{1t} \\ \vdots \\ Y_{mt} \end{pmatrix}$$

Individual modeling of each process

$\{y_{1t}\}, t = 1, \ldots, T$: Realization of Y_{1t}

$\vdots$

$\{y_{mt}\}, t = 1, \ldots, T$: Realization of Y_{mt}

Independent individual modeling of each process Y_{jt}, e.g. AR(1):

$$Y_{jt} = \varphi_j Y_{j,t-1} + c_j + u_{jt}, \; u_{jt} \sim \mathrm{Normal}\left(0, \sigma^2_{u,j}\right),$$
$$\mathrm{Cov}(u_{jt}, u_{kt}) = 0, \quad \forall j \neq k,$$
$$\mathrm{Cov}(u_{jt}, u_{k,t-h}) = 0, \quad \forall h \neq 0.$$

Individual modeling of each process

Conditional distribution of Y_{jt}, given all past values $\mathbf{Y}_{t-1}$:

- Conditional expectation:
  $\mathrm{E}(Y_{jt}|\mathbf{Y}_{t-1}) = \varphi_j Y_{j,t-1} + c_j$ (independent of $Y_{k,t-1}$)

- Conditional variance:
  $\mathrm{Var}(Y_{jt}|\mathbf{Y}_{t-1}) = \sigma^2_{u,j}$

- Conditional covariance:
  $\mathrm{Cov}(Y_{jt}, Y_{kt}|\mathbf{Y}_{t-1}) = \mathrm{Cov}(u_{jt}, u_{kt}) = 0$ (Y_{jt} and Y_{kt} are uncorrelated)

Individual modeling of each process

Dependence structure:

- autocorrelation within each time series:
  $\text{Cov}(Y_{jt}, Y_{j,t-h}) = \sigma^2_{u,j}\varphi^h_j$

- no simultaneous cross correlation ($j \neq k$):
  $\text{Cov}(Y_{jt}, Y_{kt}) = 0$

- no cross correlation across time ($h \neq 0$):
  $\text{Cov}(Y_{jt}, Y_{k,t-h}) = 0$

Joint modeling of all processes

Multivariate time series:

- Conditional expectation $\mathrm{E}(Y_{jt}|\mathbf{Y}_{t-1})$ depends not only on $Y_{j,t-1}$, but may also depend on all other past values $Y_{k,t-1}, k \neq j$.

- Simultaneous correlation of Y_{jt} and Y_{kt} for all $k \neq j$

- Y_{jt} is allowed to be correlated with past and future values of all processes $Y_{k,t-h}, k \neq j$

EViews Exercise - Case Study Industrial Production

Industrial production - quarterly data for various European countries from 1970:1 to 2001:4; consider France (`ip-fra`), Germany (`ip-deu`), and Spain (`ip-esp`)

- Show time series plot of all time series
- Define process of relative differences through `dlog`
- Discuss autocorrelation for each time series, simultaneous correlation, and cross correlation between the different countries
- Regression modeling of one time series in terms of lagged values of the other time series