Week 2: Unit roots tests
Handling strongly dependent time series, Spurious regression, Cointegration and error correction model (ECM)

Advanced Econometrics 4EK608

Vysoká škola ekonomická v Praze

Outline

Czech terminology

Testy jednotkových kořenů,
silně závislé časové řady,
zdánlivá regrese, kointegrace,
model korekce chyby,
MA proces, AR proces,
náhodná procházka s driftem,
integrovaná časová řada, řád integrace,
korelogram, ADF test,
stochatický trend, deterministický trend,
trendově stacionární časová řada,

Examples of stochastic processes

Weakly dependent time series

- Moving average process of order one MA(1)
  $x_t = e_t + \alpha_1 e_{t-1}$, where e_t is *i.i.d.* time series.
  Observations with higher time distance than 1 are uncorrelated. This process is stationary.

- Autoregressive process of order 1: AR(1)
  $y_t = \rho_1 y_{t-1} + e_t \Rightarrow Corr(y_t, y_{t+h}) = \rho_1^h$

- If stability condition $|\rho| < 1$ holds, the process is weakly dependent because correlation converges to zero with growing h. Also, this process is stationary for $y_0 = 0$.

Examples of stochastic processes

Random walk:

$$
\begin{aligned}
y_t &= y_{t-1} + e_t \\
y_t &= y_{t-2} + e_{t-1} + e_t \\
&\dots \\
y_t &= e_t + e_{t-1} + \cdots + e_1 + y_0
\end{aligned}
$$

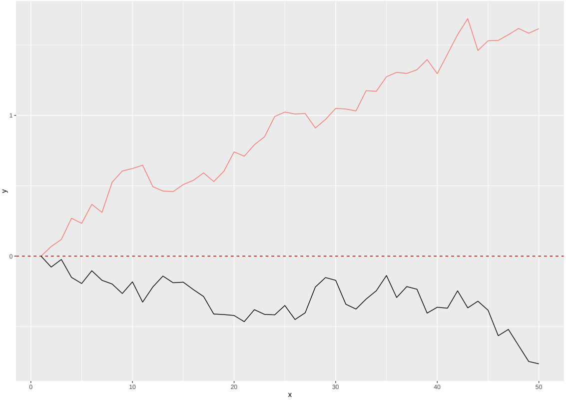
Shocks have permanent effects, the series is not covariance stationary and is strongly dependent.

$$
\begin{aligned}
E(y_t) &= E(y_0) \\
Var(y_t) &= \sigma_e^2 t \\
Corr(y_t, y_{t+h}) &= \sqrt{t/(t+h)}
\end{aligned}
$$

Correlation decreases very slowly and speed depends on t.

Examples of stochastic processes

- Two realizations of a random walk



Examples of stochastic processes

- Random walk with a drift

$$y_t = \alpha_0 + y_{t-1} + e_t \Rightarrow y_t = \alpha_0 t + e_t + e_{t-1} + \cdots + e_1 + y_0$$

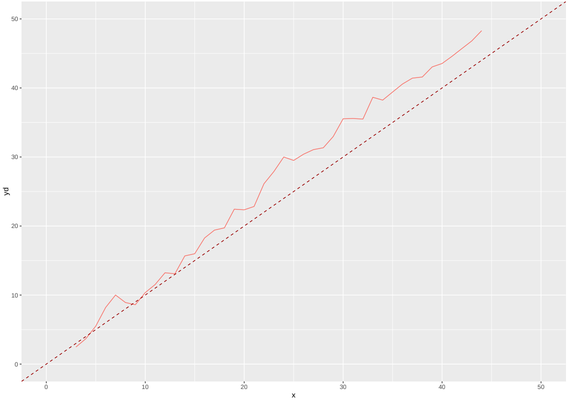
A linear trend with random walk around the trend.
It is neither covariance stationary nor weakly dependent.

$$\begin{aligned} E(y_t) &= \alpha_0 t + E(y_0) \\ Var(y_t) &= \sigma_e^2 t \\ Corr(y_t, y_{t+h}) &= \sqrt{t/(t+h)} \end{aligned}$$

Correlation decreases very slowly and decline speed depends on t.

Examples of stochastic processes

- Realization of random walk with a drift



- Different realizations of trending TS may produce similar time series.

Examples of stochastic processes

$$y_t = \boxed{1 \cdot y_{t-1}} + \boxed{u_t} = y_{t-1} + u_t$$

- Unit root process: $y_t = y_{t-1} + u_t$;
  u_t is a weakly dependent series.
- Random walk is a special case of the unit root process where: $u_t \sim Distr(0, \sigma_u^2), iid$

We need to distinguish strongly and weakly dependent TS:

- Economic reasons:
  In strongly dependent series, shocks or policy changes have long or permanent effects; in weakly dependent series, their effect is only temporary.
- Statistical reasons:
  Analysis with strongly dependent series must be handled in specific ways.

Integrated series

Terminology - Order of integration

- Weakly dependent TS are integrated of order zero: $I(0)$.
- If we have to difference a TS once to get a weakly dependent TS, then it is integrated of order 1: $I(1)$.
- Example of a $I(1)$ process:

$$\begin{aligned} y_t = y_{t-1} + e_t \quad &\Rightarrow \Delta y_t = y_t - y_{t-1} = e_t \\ \log y_t = \log y_{t-1} + e_t &\Rightarrow \Delta \log y_t = e_t \end{aligned}$$

- A time series is integrated of order d: $I(d)$, if it becomes a weakly dependent TS after being differenced d times.

Unit roots tests

Unit root tests help to decide if a time series is $I(0)$ or not

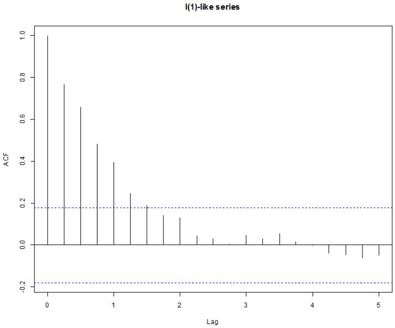
- Use either some informal procedure or a unit root test
- Informal procedures
  - Analyze autocorrelation of the first order

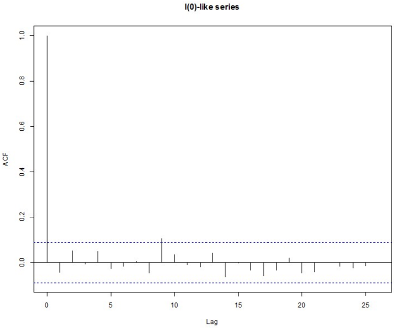
$$\hat{\rho}_1 = C\hat{o}rr(y_t, y_{t-1})$$

  - If $\hat{\rho}_1$ approaches 1, it indicates that the series can have unit root. Alternatively, it could have a deterministic trend.
- We can analyze sample autocorrelations using a correlogram

Unit root tests

Correlogram: $\rho_h = \frac{cov(y_t, y_{t-h})}{\sigma_{y_t} \cdot \sigma_{y_{t-h}}}$





$I(1)$-like series $I(0)$-like series

Unit root tests

Dickey-Fuller (DF) test – motivation

Unit root test in an $ar(1)$ process:

$y_t = \alpha + \rho y_{t-1} + e_t$

$H_0 : \rho = 1, \quad H_1 : \rho < 1$

- Under H_0, y_t has a unit root.
  - For $\rho = 1 \ \wedge \ \alpha = 0 \rightarrow y_t$ is a random walk.
  - For $\rho = 1 \ \wedge \ \alpha \neq 0 \rightarrow y_t$ is a randomw walk with a drift and $E(y_t)$ is a linear function of t .
- Under H_1, y_t is a weakly dependent $ar(1)$ process.

Unit root tests

Dickey-Fuller (DF) test – motivation

Unit root test in an $ar(1)$ process:

$y_t = \alpha + \rho y_{t-1} + e_t$

$H_0 : \rho = 1, \quad H_1 : \rho < 1$

For DF tests, $H_1 : \rho < 1$ is a common simplification to the full space of alternatives to $H_0 : \rho = 1$.

- For $|\rho| < 1$, y_t is weakly dependent (as $plim\, \rho^h = 0$) However, if unit root is likely to be present, the probability of $\rho < 0$ is negligible.
- We usually ignore the possibility of $\rho > 1$, as it would lead to explosive behavior in y_t.
  $\ldots |\rho| > 1$ would allow for explosive oscillations in y_t.

Dickey Fuller (DF) test

- Basic equation for unit root test in an $ar(1)$ process:
  $y_t = \alpha + \rho y_{t-1} + e_t$
- For DF test, we apply a suitable transformation to y_t:
  we subtract y_{t-1} from both sides of the equation:

  $\Delta y_t = \alpha + (\rho - 1) y_{t-1} + e_t;$ apply substitution: $\theta = (\rho - 1)$

  i.e.

  $\Delta y_t = \alpha + \theta y_{t-1} + e_t;$ now:

  $$\begin{aligned} H_0 &: \rho = 1 \Leftrightarrow \boxed{H_0 : \theta = 0} \\ H_1 &: \rho < 1 \Leftrightarrow \boxed{H_1 : \theta < 0} \end{aligned}$$

- We use a t-ratio for testing $H_0 : \theta = 0$. However:
  Under H_0, t-ratios don't have a t-distribution, but follow a DF-distribution. (-negative- critical values of the DF distribution are much farther from zero)
- Critical values for the DF distribution are available from statistical tables and implemented in most relevant SW packages.

DF test & ADF test

Unit root time series can manifest various levels of complexity. Hence, DF test is usually performed using the following three specifications:

$$\Delta y_t = \boxed{\theta y_{t-1}} + e_t \quad \text{random walk}$$
$$\Delta y_t = \alpha + \boxed{\theta y_{t-1}} + e_t \quad \text{random walk with a drift}$$
$$\Delta y_t = \alpha + \boxed{\theta y_{t-1}} + \delta t + e_t \quad \text{random walk with a drift and trend}$$

DF test is the same $(H_0 : \theta = 0)$ for all specifications /critical values diffFer/

Augmented Dickey-Fuller (ADF) test is a common generalization of DF test (example: Augmentation of the DF test for the 2^{nd} specification)

$$\Delta y_t = \alpha + \boxed{\theta y_{t-1}} + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p} + e_t$$

- When estimating θ, we control for possible $ar(p)$ behavior in Δy_t.
- ADF test has the same null hypothesis as a DF test $\rightarrow H_0 : \theta = 0$.

Unit root tests in R: package {urca}

Description of the options for the `ur.df()` function:

1. `type "none"`
   $\Delta y_t = \theta y_{t-1} + e_t$
   `tau1`: we test for $H_0 : \theta = 0$ (unit root)

2. `type "drift"`
   $\Delta y_t = \alpha + \theta y_{t-1} + e_t$
   `tau2`: $H_0 : \theta = 0$ (unit root)
   `phi1`: $H_0 : \theta = \alpha = 0$ (unit root and no drift)

3. `type "trend"`
   $\Delta y_t = \alpha + \theta y_{t-1} + \delta t + e_t$
   `tau3`: $H_0 : \theta = 0$ (unit root)
   `phi2`: $H_0 : \theta = \alpha = \delta = 0$ (unit root, no drift, no trend)
   `phi3`: $H_0 : \theta = \delta = 0$ (unit root and no trend)

Unit root tests

- ADF test for TS with trend

$$\Delta y_t = \alpha + \theta y_{t-1} + \delta t + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p} + e_t$$

  Under the alternative hypothesis of no unit root, the process is trend-stationary.

- The critical values in the ADF distribution with time trend are even more negative as compared to random walk and random walk with a drift.

- When using DF/ADF specification 1 or 2 (R-W, R-W with drift) to test for unit root in a clearly trending TS, the test would not have sufficient power (we would not reject H_0 for trending weakly dependent TS).

Unit roots and trend-stationary series

- $\Delta y_t = \alpha + \theta y_{t-1} + \delta t + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p} + e_t$
- Terminology:
  - Stochastic trend: $\theta = 0$
    Also called **difference-stationary process**: y_t can be turned into $I(0)$ series by differencing. Terminology emphasizes stationarity after differencing y_t instead of weak dependence in differenced TS.
  - Deterministic trend: $\delta \neq 0, \quad \theta < 0$
    Also called **trend-stationary process**: has a linear trend, not a unit root. y_t is weakly dependent - $I(0)$ - around its trend. We can use such series in LRMs, if trend is also used as regressor.
- DF/ADF tests are not precise tools. Distinguishing between stochastic and deterministic trend is not easy (sample size!).

Handling trend-stationary time series

- Trend-stationary TS fulfill TS.1' assumption (look at Week1 presentation).

  We can use them in regressions if we have time trend among regressors.

Handling strongly dependent time series

- Strongly dependent time series do not fulfill TS.1' assumption (look at Week1 presentation). We cannot use them in regressions directly.
- Sometimes, taking logarithms helps.
- Sometimes, we can transform such series into weakly dependent time series.
- Differencing is popular, but it has drawbacks.

Handling strongly dependent time series

Example

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t \qquad Y_t, X_t \sim I(1) \qquad (1)$$

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + \varepsilon_{t-1} \qquad \varepsilon_t \sim i.i.d. \qquad (2)$$

$$\Delta Y_t = \beta_2 \Delta X_t + v_t \qquad v_t = \varepsilon_t - \varepsilon_{t-1} \qquad (3)$$

1. If we work with logarithms, it has an additional advantage: $\log Y_t - \log Y_{t-1} = \log\left(\frac{Y_t}{Y_{t-1}}\right) \doteq \frac{Y_t - Y_{t-1}}{Y_{t-1}}$
   i.e.: nice interpretation as rate of growth of Y_t
2. Three problems
   1. v_t is no more i.i.d.
   2. We loose information linked with the levels of variables, short term relation are stressed
   3. Estimates often generate bad long-term predictions: $\Delta \hat{Y}_t = \hat{\beta}_2 \Delta X_t$; ... what if $\beta_1 \neq 0$?

Handling strongly dependent time series

Some properties of integrated processes

1. The sum of stationary and non-stationary series must be non-stationary.
2. Consider a process $y_t = \alpha + \beta x_t$:
   - If x_t is stationary then y_t will be stationary.
   - If x_t is non-stationary then y_t will be non-stationary.
3. If two time series are integrated of different orders, then any linear combination of the series will be integrated at the higher of the two orders of integration.
4. Sometimes it turns out a linear combination of two $I(d)$ series is integrated of order less then d.

Spurious regression or cointegration

- **Spurious regression** Regressing one $I(1)$-series on another $I(1)$-series may lead to extremely high t-statistics even if the series are completely independent. Similarly, the R^2 of such regressions tend to be very high.
  Regression analysis involving time series that have a unit root may generate completely misleading inferences.

- **Cointegration** Fortunately, regressions with $I(1)$-variables are not always spurious: If there is a stable relationship between time series that, individually, display unit root behavior, these time series are called "cointegrated".

Spurious regression or cointegration

General definition of cointegration

Two $I(1)$-time series y_t, x_t are said to be cointegrated if there exists a stable relationship between them, where:

$$y_t = \alpha + \beta x_t + e_t, \qquad e_t \sim I(0)$$

Cointegration (CI) test if CI parameters are known

For residuals of the known CI relationship:

$$e_t := y_t - \alpha - \beta x_t,$$

test whether the residuals have a unit root. If the unit root H_0 is rejected, y_t, x_t are cointegrated.

Spurious regression or cointegration

- **Testing for CI if the parameters are unknown**
  If the potential relationship is unknown, it can be estimated by OLS. After that, we test whether the regression residuals have a unit root. If the unit root is rejected, this means that y_t, x_t are cointegrated. Due to the pre-estimation of parameters, critical values are different than in the case of known parameters. (Software handles this automatically.)
- **The CI relationship may include a time trend**
  If the two series have differential time trends (drifts in this case), the deviation between them may still be $I(0)$ but with a linear time trend. In this case one should include a time trend in the CI-regression. Also, we have to use different critical values when testing residuals. (Software handles this automatically.)

Cointegration tests based on regression residuals

Engle-Granger test estimates a p-lag ADF equation:

$$\Delta\hat{u}_t = \theta\,\hat{u}_{t-1} + \sum_{j=1}^{p} \Delta\hat{u}_{t-j} + e_t$$

- Esentially, this is an ADF test on $\hat{u}_t$ $[\theta = (\rho - 1)]$
- Specific critical values apply (farther from 0 than t or DF).

Phillips-Ouliaris test estimates a DF equation:

$$\Delta\hat{u}_t = \theta\,\hat{u}_{t-1} + e_t$$

- The t-ratio is based on robust standard errors, different estimator exist for the robust standard errors.

In both cases (EG and PO), H_0 of unit root in $\hat{u}$ i.e. “no-cointegration” is tested.

Error correction model (ECM)

- It can be shown that when variables are cointegrated, i.e. when there exists a long-term relationship among them, their short-term dynamics are related as in a so-called error correction model (ECM).

Error Correction Models - motivation

Autoregressive distributed lag models

- Autoregressive distributed lag model with one regressor

$$\mathrm{ADL}(p,q):\; y_t = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{j=0}^{q} \gamma_j x_{t-j} + u_t, \quad u_t \sim iid(0,\sigma^2)$$

- There are many useful modifications/simplifications to the ADL(p,q) process. For example:

$$\mathrm{ADL}(1,1):\; y_t = \beta_0 + \beta_1 y_{t-1} + \gamma_0 x_t + \gamma_1 x_{t-1} + u_t\,. \qquad (4)$$

Additional ADL$(1,1)$ restriction: $\beta_1 = 1$ and $\gamma_1 = -\gamma_0$

gives a model in 1^{st} diffs.: $\Delta y_t = \beta_0 + \gamma_0 \Delta x_t + u_t\,.$

Error Correction Models - motivation

For ADL(1,1) model (4), suppose there is an equilibrium value x° and in the absence of shocks, $x_t \to x^\circ$ as $t \to \infty$. Then, assuming absence of u_t errors, y_t converges to steady state: y°.

Hence, the ADL(1,1) model (4) can be re-written as:

$$y^\circ = \beta_0 + \beta_1 y^\circ + (\gamma_0 + \gamma_1)x^\circ$$

Solving this for y° as a function of x°, we get

$$y^\circ = \frac{\beta_0}{1-\beta_1} + \frac{\gamma_0+\gamma_1}{1-\beta_1}x^\circ = \frac{\beta_0}{1-\beta_1} + \lambda x^\circ$$

where $\lambda \equiv \frac{\gamma_0+\gamma_1}{1-\beta_1}$ and $|\beta_1| < 1$ is assumed.

Error Correction Models - motivation

$$y^{\circ} = \frac{\beta_0}{1-\beta_1} + \lambda x^{\circ}$$

$$\lambda \equiv \frac{\gamma_0 + \gamma_1}{1-\beta_1}$$

- λ is the long-run derivative of y° with respect to x°.
- λ is an elasticity if both y° and x° are in logs.
- $\hat{\lambda}$ can be computed directly from the estimated parameters of the ADL(1,1) model (4).

Error Correction Models - motivation

The ADL(1,1) equation (4) - repeated here for convenience:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \gamma_0 x_t + \gamma_1 x_{t-1} + u_t,$$

can be equivalently rewritten as follows:

$$\Delta y_t = \beta_0 + (\beta_1 - 1)(y_{t-1} - \lambda x_{t-1}) + \gamma_0 \Delta x_t + u_t. \qquad (5)$$

Again, $\lambda \equiv \frac{\gamma_0 + \gamma_1}{1 - \beta_1}$ and $|\beta_1| < 1$ is assumed.

Equation (5) is an error-correction model (ECM).

Error Correction Models

ECM: $\Delta y_t = \beta_0 + (\beta_1 - 1)(y_{t-1} - \lambda x_{t-1}) + \gamma_0 \Delta x_t + u_t.$

- $(y_{t-1} - \lambda x_{t-1})$ measures the extent to which the long run equilibrium between y_t and x_t is not satisfied (at $t-1$).
- Consequently, $(\beta_1 - 1)$ can be interpreted as the proportion of the disequilibrium $(y_{t-1} - \lambda x_{t-1})$ that is reflected in the movement of y_t, i.e. in Δy_t.
- $(\beta_1 - 1)(y_{t-1} - \lambda x_{t-1})$ is the **error-correction term**.
- Many ADL(p, q) specifications can be re-written as ECMs.
- **ECMs can be used with non stationary TS** (Week 3).
- ECMs $(\beta_1 - 1)$ is essentially the same as θ from Partial adjustment model (see Week 4).

Box 1: Partial Adjustment Models
(Will be put into context in the 4th week)

$$Y_t^* = \alpha + \beta X_t + u_t \tag{6}$$

Y_t^*: optimal value or target value or long-run equilibrium value

$$Y_t - Y_{t-1} = \theta(Y_t^* - Y_{t-1}) \tag{7}$$

$0 < \theta < 1$

$\theta \sim$ coefficient of adjustment

$Y_t = \theta Y_t^* + (1-\theta)Y_{t-1}$

As $\theta \to 1$, the speed of contemporaneous adjustment of Y_t towards Y_t^* grows.

Substituting Y_t^* from (6) to (7) yields

$$Y_t = \alpha\theta + \beta\theta X_t + (1-\theta)Y_{t-1} + \theta u_t \tag{8}$$

We estimate (8) and then calculate parameters in (6) and (7).

Error Correction Models

Some more complicated ECMs:

1) We can use higher order lags, e.g. ADL(2,2):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \gamma_0 x_t + \gamma_1 x_{t-1} + \gamma_2 x_{t-2} + u_t,$$

to establish ECMs. It is again possible to rearrange and re-parametrize ADL(2,2) to get an ECM. More than one re-parameterization is possible.

2) More than two variables can enter into an equilibrium relationship.