Time series. Stationarity, spurious regression and cointegration.

Jakub Mućk
SGH Warsaw School of Economics

Introduction

- Time series y_t a series of observations indexed in time order, where $t = 1, 2, \ldots, T$.
- Lag operator L:
$$L(y_t) = y_{t-1}. \tag{1}$$
- Difference operator/first difference Δ:
$$\Delta(y_t) = (1 - L)\, y_t = y_t - y_{t-1}. \tag{2}$$
- Growth rates measure the percentage changes of y_t within a specific period:
$$g = \frac{y_t - y_{t-1}}{y_{t-1}}. \tag{3}$$
- Logarithmic growth rates:
$$\Delta \ln y_t = \ln y_t - \ln y_{t-1} = \ln \frac{y_t}{y_{t-1}} = \ln \frac{y_{t-1} \times (1+g)}{y_{t-1}} \approx g. \tag{4}$$

Stationarity

- **Weak stationarity (wide-sense stationarity)** is satisfied when series have:
  1. constant mean:
  $$\mathbb{E}(y_t) = \mu \tag{5}$$
  2. constant variance :
  $$\text{var}(y_t) = \sigma^2 \tag{6}$$
  3. covariance doesn't depend on time t:
  $$\text{cov}(y_t, y_{t+s}) = \text{cov}(y_t, y_{t-s}) = \gamma_s \tag{7}$$
- Stationary time series have the property of **mean reversion**
- Nonstationary time series series has a **unit root**
- Intuition: stationary series fluctuate around sample mean.

- The order of integration is the minimum number of differences required to obtain stationary series.
- Stationary series are integrated of order 0, i.e. $y_t \sim I(0)$.
- If series is non-stationary but its first difference is stationary then this series is integrated of order 1, i.e., $y_t \sim I(1)$.
- In general,

$$y_t \sim I(d) \iff \Delta^d y_t \sim I(0). \tag{8}$$

- **Difference-stationary**:

$$y_t \sim I(1) \iff \Delta y_t \sim I(0). \tag{9}$$

- **Trend-stationary**:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t \iff \varepsilon_t \sim I(0). \tag{10}$$

- The choice between difference and trend stationarity is sometimes arbitrary. But it has implication for y_t.

Stationary processes

- White noise.
- Autoregressive process.

Non-stationary processes

- Random walk.

- **White noise:**

$$y_t = \varepsilon_t, \tag{11}$$

where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

- **AR(1) process:**

$$y_t = \rho y_{t-1} + \varepsilon_t, \tag{12}$$

where $|\rho| < 1$ and $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

- Selected properties of y_t when it follows AR(1) process:

$$\mathbb{E}(y_t) = 0, \tag{13}$$

$$Var(y_t) = \frac{\sigma^2}{1-\rho^2}. \tag{14}$$

- Both white noise and AR(1) processes are stationary.

- Let μ denotes constant term. Then AR(1) with constant:

$$y_t = \mu + \rho y_{t-1} + \varepsilon_t \tag{15}$$

where $|\rho| < 1$ and $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

- It's useful to rewrite (15) as follows:

$$(y_t - \mu) = \rho\left(y_{t-1} - \mu\right) + \varepsilon_t \tag{16}$$

- The long-run expected value:

$$\mathbb{E}(y_t) = \mu/(1-\rho) \tag{17}$$

- Finally, we AR(1) model might be extended by linear trend. Then

$$(y_t - \mu - \delta t) = \rho\left(y_{t-1} - \mu - \delta(t-1)\right) + \varepsilon_t \tag{18}$$

- **Random walk** process is an example of **nonstationary process**.

$$y_t = y_{t-1} + \varepsilon_t \tag{19}$$

where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma^2\right)$

- The sample means depends on the time span.
- Using recursive substitution we can show that random walk process is **wandering**

$$\begin{aligned} y_1 &= y_0 + \varepsilon_1 \\ y_2 &= y_1 + \varepsilon_2 = y_0 + \varepsilon_1 + \varepsilon_2 = y_0 + \sum_{k=1}^{2} \varepsilon_k \\ &\dots \\ y_t &= y_0 + \sum_{k=1}^{t} \varepsilon_k \end{aligned} \tag{20}$$

- Cumulative sum of shocks/innovations $(\sum_{k=1}^{t} \varepsilon_k)$ is often said to be **stochastic trend**.

- $\mathbb{E}(y_t)$ depends on initial value:

$$\mathbb{E}(y_t) = \mathbb{E}(y_0 + \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_t) = y_0 \tag{21}$$

- But the variance depends on time and cannot be limited:

$$\text{var}(y_t) = \text{var}(\varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_t) = t\sigma^2 \tag{22}$$

Hence, the condition on constant variance is not satisfied. Therefore, the random walk process is nonstationary.

Random Walk with drift and trend

- We can include constant term into DGP of basic random walk process. It's known as **random walk with drift**

$$y_t = \mu + y_{t-1} + \varepsilon_t \tag{23}$$

where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

- The mean and variance:

$$\begin{aligned} \mathbb{E}(y_t) &= t\mu + y_0 + \mathbb{E}(\varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_t) = t\mu + y_0, \\ \text{var}(y_t) &= \text{var}(\varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_t) = t\sigma^2. \end{aligned} \tag{24}$$

- Finally, we might include deterministic trend

$$y_t = \mu + \beta t + y_{t-1} + \varepsilon_t, \tag{25}$$

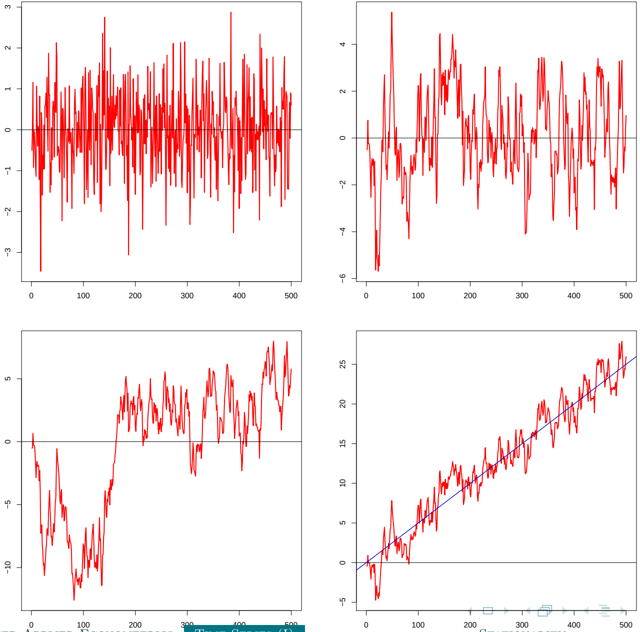
where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

- The addition of deterministic trend strengthens the trend behavior:

$$\mathbb{E}(y_t) = t\mu + \left(\frac{t^2 + t}{2}\right)\beta + y_0 t. \tag{26}$$

Examples (simulated data)

SGH



Dickey-Fuller test

Dickey–Fuller test (basic) I

- The general assumption: y_t is generated by AR(1):

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{27}$$

- The general idea is to test whether ρ is equal or **significantly less than one**.
- **The null is that there is unit root** (y_t is nonstationary).
- Estimating ρ in equation (27) and calculating t-statistics might lead to completely meaningless result (spurious regression). Therefore, y_t should be differenced:

$$\Delta y_t = (\rho - 1)\, y_{t-1} + \varepsilon_t = \gamma y_{t-1} + \varepsilon_t \tag{28}$$

where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ and $\gamma = (\rho - 1)$.

- The null (y_t is not stationary) and alternative (y_t is stationary)

$$\begin{array}{llcll} \mathcal{H}_0 : & \rho = 1 & \iff & \mathcal{H}_0 : & \gamma = 0 \\ \mathcal{H}_1 : & \rho < 1 & \iff & \mathcal{H}_1 : & \gamma < 0 \end{array} \tag{29}$$

- To check stationary we estimate equation (27) and calculate t-statistic. Here, **we cannot use t distribution** because the **calculated t statistic is not t distributed**. Therefore we use τ **distribution** which equals t-statistics. Critical values for τ statistic are computed from numerical distributions.
- The null is rejected when τ **is below its critical value**.
- One might include in regression (27) deterministic components: **constant** or **time trend**. The null and alternative will be the same.

- To avoid the danger of autocorrelation of error test we might extend test regression by autoregression part of higher order:

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{P} \alpha_i \Delta y_{t-i} + \varepsilon_t \tag{30}$$

- The null and the alternative are the same as in the basic version.

$$\begin{aligned} \mathcal{H}_0 : \quad & \gamma = 0 \\ \mathcal{H}_1 : \quad & \gamma < 0 \end{aligned} \tag{31}$$

- In practice, including autoregressive part in DF regression is very often approach.

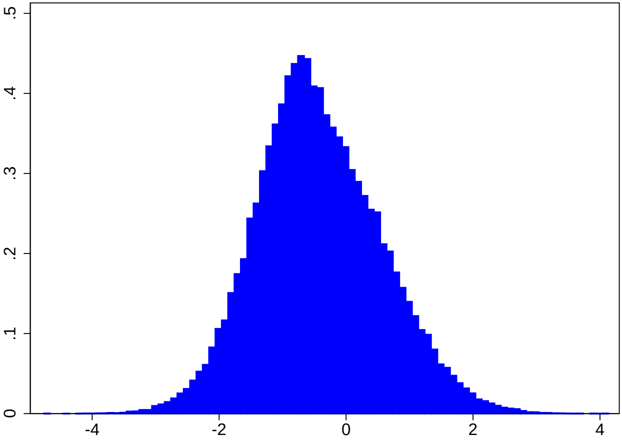
Distribution of the Dickey-Fuller statistic

- The distribution of the Dickey-Fuller statistic is approximated numerically.
- Under the null, the series y_t has a unit root, i.e., $\rho = 1$ and

$$y_t = y_{t-1} + \varepsilon_t \tag{32}$$

  where $\varepsilon_t \in \mathcal{N}(0, \sigma^2)$.
- Assumptions:
  - $T = 1000$;
  - $\sigma = 0.1$;
  - Number of replications: 100000.

Simulations results – distribution of the Dickey-Fuller statistic



qunatile	0.001	0.01	0.5	0.1	0.5
value	-3.304	-2.569	-1.938	-1.614	-0.500

Critical values for the (augmented) Dickey-Fuller test

SGH

- Critical values for the (augmented) Dickey-Fuller test are different from the t-student distribution.
- There could be slight differences in critical values between software since they are calculated numerically.

Regression	1%	5%	10%
$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$	-2.56	-1.94	-1.62
$\Delta y_t = \mu + \gamma y_{t-1} + \varepsilon_t$	-3.43	-2.86	-2.57
$\Delta y_t = \mu + \delta t + \gamma y_{t-1} + \varepsilon_t$	-3.96	-3.41	-3.13
t-student statistic	-2.33	-1.65	-1.28

Note: above critical values are taken from Davidson and MacKinnon (1993)

Empirical example: unemployment rate U_t

The Dickey-Fuller regression:

$$\Delta U_t = \underset{(0.185)}{0.087} - \underset{(0.0144)}{0.0311}\, U_{t-1}, \qquad (33)$$

Test statistic: $-0.031/0.014 \approx -2.16$
Critical value (10% significance level):
-2.570
$\Longrightarrow$ null cannot be rejected $\mathcal{H}_0$

The Dickey-Fuller regression:

$$\Delta U_t = \underset{(0.185)}{0.087} - \underset{(0.0144)}{0.0311}\, U_{t-1}, \qquad (33)$$

Test statistic: $-0.031/0.014 \approx -2.16$
Critical value (10% significance level): -2.570
$\Longrightarrow$ null cannot be rejected $\mathcal{H}_0$

[But:] serially correlated residuals
Correlation between residuals and their lags ≈ 0.65

Empirical example: unemployment rate U_t

The Dickey-Fuller regression:

$$\Delta U_t = \underset{(0.185)}{0.087} - \underset{(0.0144)}{0.0311}\, U_{t-1}, \qquad (33)$$

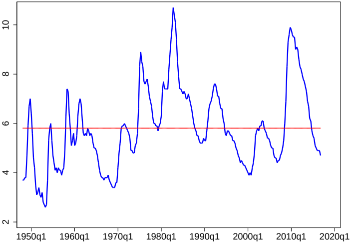
Test statistic: $-0.031/0.014 \approx -2.16$
Critical value (10% significance level): -2.570
$\Longrightarrow$ null cannot be rejected $\mathcal{H}_0$

[But:] serially correlated residuals
Correlation between residuals and their lags ≈ 0.65

The Dickey-Fuller regression:

$$\Delta U_t = \underset{(0.066)}{0.2854} - \underset{(0.011)}{0.049}\, U_{t-1} + \underset{(0.045)}{0.662}\, \Delta U_{t-1}, \qquad (34)$$

Test statistic: $-0.049/0.011 \approx -4.47$
Critical value (1% significance leveli): -3.458
$\Longrightarrow$ the $\mathcal{H}_0$ can be rejected (about unit root) at 1% significance level.



	the ADF statistic	p
$\ln GDP_t$	-2.07	1
$\Delta \ln GDP_t$	-11.19	0

where p is the number of lags in the Dickey-Fuller regression.

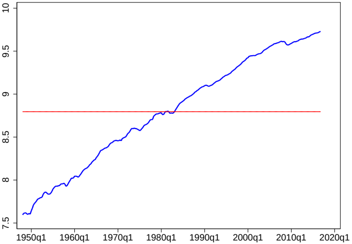
Critical values

1%	5%	10%
-3.458	-2.879	-2.570

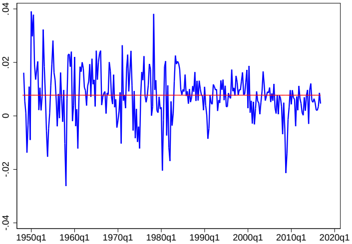
What is the order of integration of GDP_t?

Is the $\ln GDP_t$ difference-stationary?

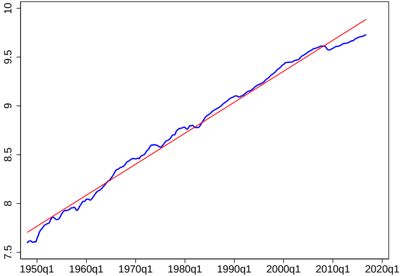
$\ln GDP_t$



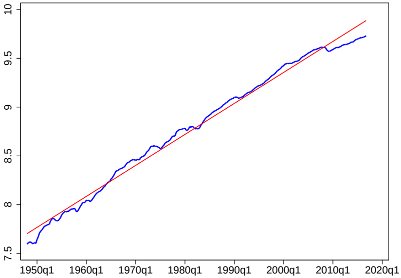
$\Delta \ln GDP_t$



Is the $\ln GDP_t$ stationarity around a (deterministic) trend?

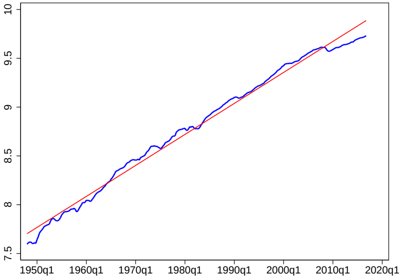


Is the $\ln GDP_t$ stationarity around a (deterministic) trend?



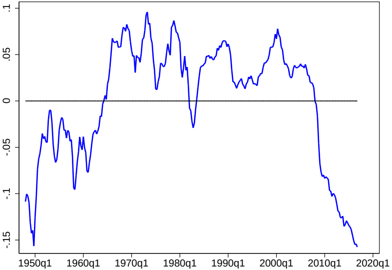
The ADF test statistic: -1.232
Critical value (10% significance level): -3.130.

Is the $\ln GDP_t$ stationarity around a (deterministic) trend?



The ADF test statistic: -1.232
Critical value (10% significance level): -3.130.

Residuals from the $\ln GDP_t$ regressed on a deterministic trend.



Spurious Regressions

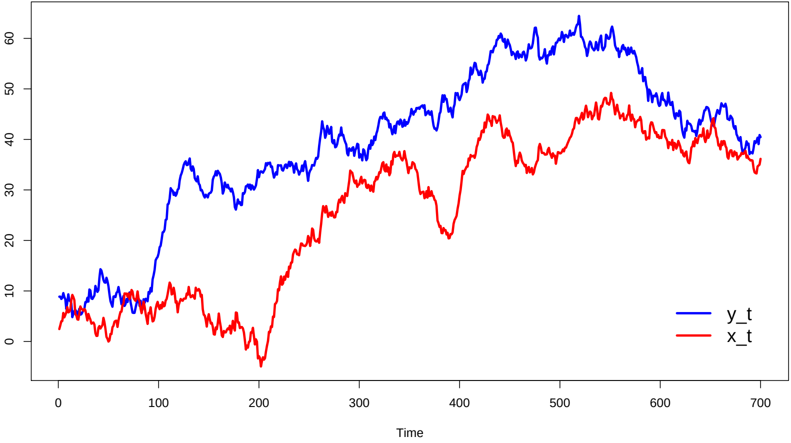
- An investigation on stationary is necessary because there is a danger of obtaining the **significant estimation** results from **unrelated data** when series exhibit unit root. Such case is said to be **spurious regression**.
- To illustrate we simulate two random walk series (y_t and x_t):

$$\begin{array}{lrcl} \text{DGP}_1: & y_t & = & y_{t-1} + \varepsilon_t \\ \text{DGP}_2: & x_t & = & x_{t-1} + \eta_t \end{array} \tag{35}$$

where $\eta_t \sim \mathcal{N}\left(0, \sigma_\eta^2\right)$ and $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$.

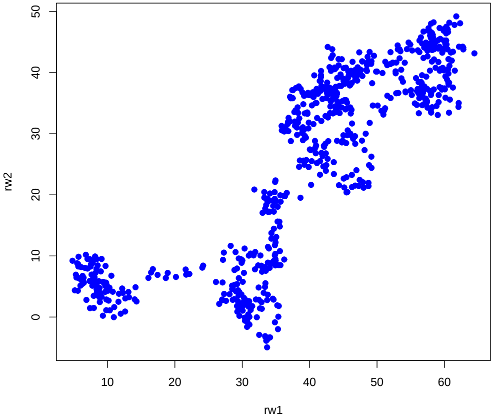
- The series y_t and x_t are simulated independently so there is no relation between these variables.

Figure: Simulated random walk series y_t and x_t



Despite no relation between series there is upward trend in both series.

Figure: Scatter plot of simulated random walk series y_t and x_t



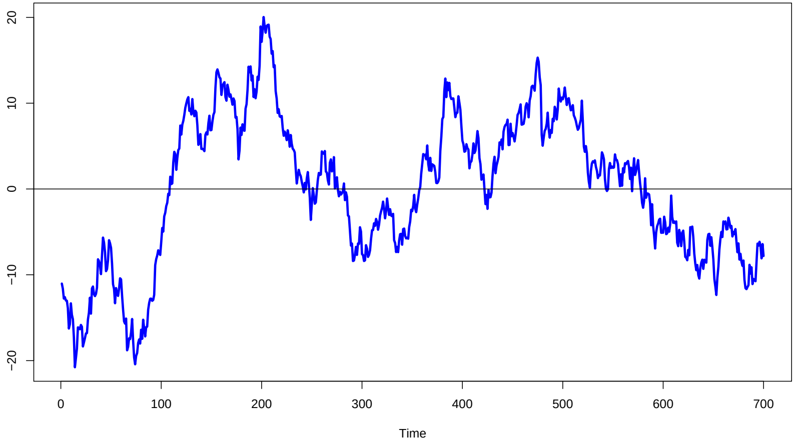
- Simple regression of y_t on x_t (standard errors are in round brackets):

$$y_t = \underset{(0.62048)}{17.818} + \underset{(0.02062)}{0.842}x_t \tag{36}$$

- t-statistic for x_t : 40.82
- R^2 is about 0.705
- But we know that series y_t on x_t were generated independently and there is **no true relationship** between these variables. Thus, the above estimates are totally meaningless or **spurious**.
- When nonstationary time series are used in a simple regression, the least squares estimator doesn't have its usual properties. As a result, t-statistics are not reliable.
- The residuals from such (spurious) regression are autocorrelated.

▸ See residuals

Figure: RESIDUALS FROM REGRESSION y_t ON x_t



DW statistics: 0.22 LM statistic (the autocorrelation of first order): 682.958[0.0000]

Back to example

Cointegration

- **Cointegration** is a special case of relationship between non-stationary variables.
- Key assumption: y_t and x_t are **integrated of order one** and the residuals e_t, i.e,:
$$e_t = y_t - \beta_0 - \beta_1 x_t \tag{37}$$
are **stationary**. Then, variables x_t and y_t are **cointegrated**.
- Intuition(1): if variables are cointegrated then they share common stochastic trend.
- Intuition(2): if variables are cointegrated then there exists long-run relationship (equilibrium) between variables.

- **First step:** testing stationarity. If the variables x_t and y_t are integrated of order one then go to the next step.
- **Second step:** estimate parameters for the long-run regression, obtain residuals (e_t) and test whether residuals are stationary:

$$e_t = y_t - \beta_0 - \beta_1 x_t \tag{38}$$

- The null and altenrative:

$$\begin{array}{lllll} \mathcal{H}_0: & e_t \sim \mathrm{I}(1) & \iff & \mathcal{H}_0: & x_t \text{ and } y_t \text{ are } \textbf{not} \text{ cointegrated} \\ \mathcal{H}_1: & e_t \sim \mathrm{I}(0) & \iff & \mathcal{H}_1: & x_t \text{ and } y_t \text{ are cointegrated} \end{array} \tag{39}$$

- But use the critical values for the cointegration test (which are different from the ADF test for a series):

Table: Critical values

Long-run regression	1%	5%	10%
$y_t = \beta_1 x_t + e_t$	-3.39	-2.76	-2.45
$y_t = \beta_0 + \beta_1 x_t + e_t$	-3.96	-3.37	-3.07
$y_t = \beta_0 + \delta t + \beta_1 x_t + e_t$	-3.98	-3.42	-3.13

Notes: critical values are taken from Hamilton (1994)

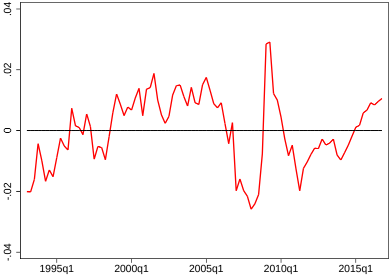
Both c_t and y_t are integrated in order order 1.

Long-run relationship:

$$\hat{c}_t = \underset{(0.174)}{3.984} + \underset{(0.013)}{0.657} y_t \qquad (40)$$

What is the long-run elasticity?

Residuals from the long-run regression



The ADF statistics: -3.52.