

# Lecture 8: Unit Roots, AR-DL Models, and Granger Causality

---

Raul Riva

[FGV EPGE](#)

October, 2025

## Intro

---

- We have modelled time-dependence in different ways through ARMA models;
- But every time we had only one variable  $y_t$ ;
- In reality: several variables interact over time generating complicated dynamics;
- Causality in this context is super hard – no hope for experiments and so on;
- Importantly: a tool like an estimator can never ensure causality. Only careful design can.

## Quick Simulation

---

- Let  $u_t$  and  $v_t$  be two independent i.i.d.  $\sim N(0, 1)$  random variables;
- Let  $y_t = y_{t-1} + u_t$  and  $x_t = x_{t-1} + v_t$ ;
- Both  $y_t$  and  $x_t$  are random walks (non-stationary processes);
- Both are *completely independent* of each other;

## Quick Simulation

- Let  $u_t$  and  $v_t$  be two independent i.i.d.  $\sim N(0, 1)$  random variables;
- Let  $y_t = y_{t-1} + u_t$  and  $x_t = x_{t-1} + v_t$ ;
- Both  $y_t$  and  $x_t$  are random walks (non-stationary processes);
- Both are *completely independent* of each other;
- But let's say you do not know that and you try to regress  $y_t$  on  $x_t$ :

$$y_t = \alpha + \beta x_t + \epsilon_t$$

- What would be your interpretation of  $\beta$  here?
- What would you expect in terms of the estimate of  $\beta$ ?

## Simulation Setup

---

- For a sample size  $T$ , simulate many paths of length  $T$  of the pair  $(y_t, x_t)$ ;
- For each path, estimate the regression of  $y_t$  on  $x_t$  and collect  $\hat{\beta}$ ;

## Simulation Setup

---

- For a sample size  $T$ , simulate many paths of length  $T$  of the pair  $(y_t, x_t)$ ;
- For each path, estimate the regression of  $y_t$  on  $x_t$  and collect  $\hat{\beta}$ ;
- Additionally, I will also estimate the regression of  $y_t$  on  $x_t$  and  $y_{t-1}$ :

$$y_t = \alpha + \beta x_t + \gamma y_{t-1} + \epsilon_t$$

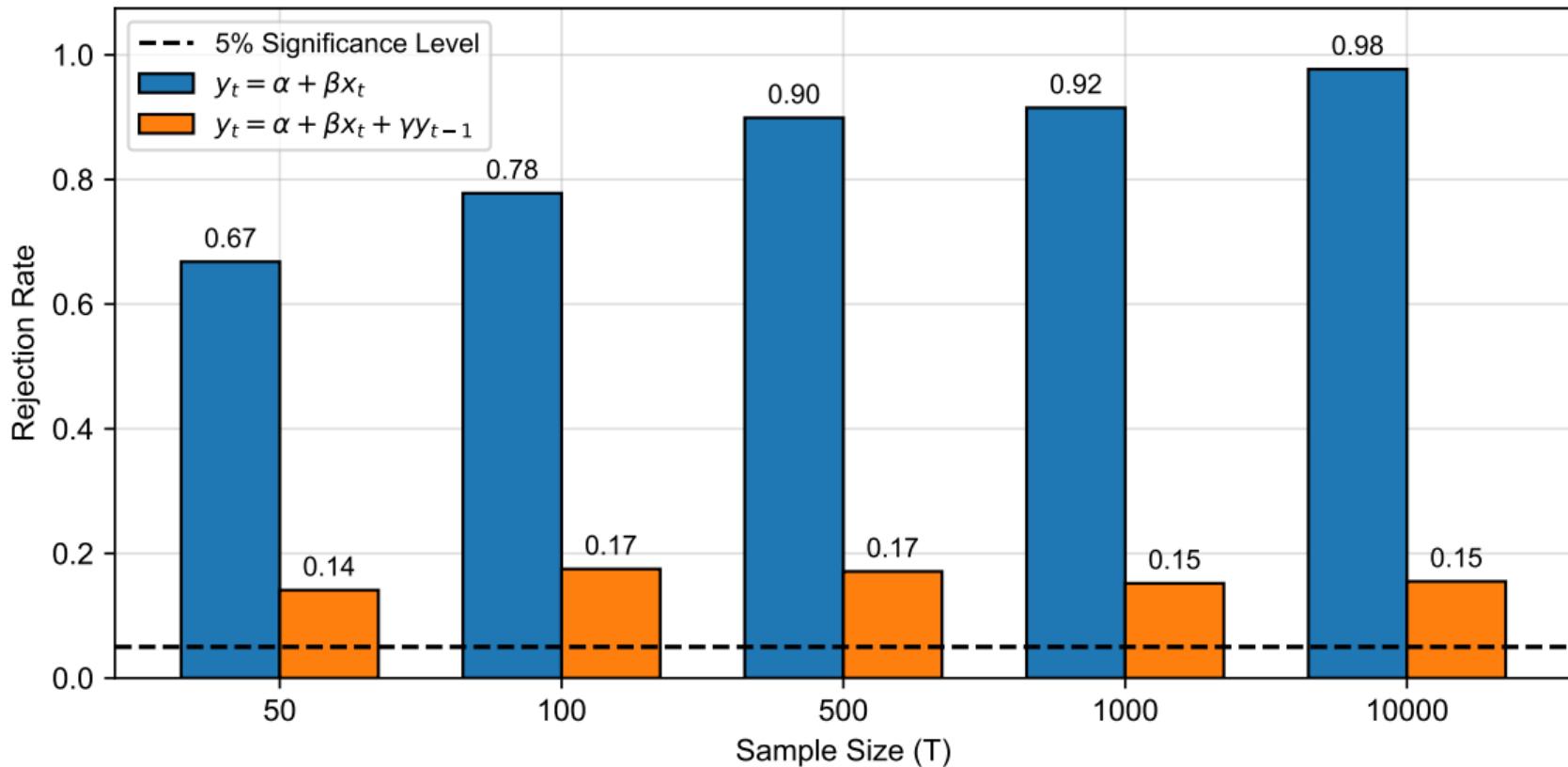
## Simulation Setup

- For a sample size  $T$ , simulate many paths of length  $T$  of the pair  $(y_t, x_t)$ ;
- For each path, estimate the regression of  $y_t$  on  $x_t$  and collect  $\hat{\beta}$ ;
- Additionally, I will also estimate the regression of  $y_t$  on  $x_t$  and  $y_{t-1}$ :

$$y_t = \alpha + \beta x_t + \gamma y_{t-1} + \epsilon_t$$

- For each run, I implement a t-test of  $H_0 : \beta = 0$  at 5% significance level;
- For each run, I record whether we would find starts in a regression table or not...
- Repeat this for many different  $T$ ;

## Rejection Rates for $\beta$ at 5% Significance Level



## What is going on?

---

- Even with a correctly specified model, we rejected the null too often!
- The reason is the lack of stationarity of the data generating process!
- Recall that (imposing that I start the simulation at  $x_0 = 0$ ):

$$Var(x_t) = Var\left(\sum_{s=1}^t v_s\right) = t;$$

## What is going on?

---

- Even with a correctly specified model, we rejected the null too often!
- The reason is the lack of stationarity of the data generating process!
- Recall that (imposing that I start the simulation at  $x_0 = 0$ ):

$$Var(x_t) = Var\left(\sum_{s=1}^t v_s\right) = t;$$

- This will imply  $\frac{1}{T} \sum_{t=1}^T x_t^2$  does *not* converge to a constant as  $T \rightarrow \infty$ ;
- This is easy to see:

$$E\left[\frac{1}{T} \sum_{t=1}^T x_t^2\right] = \frac{1}{T} \sum_{t=1}^T E[x_t^2] = \frac{1}{T} \sum_{t=1}^T t = \frac{T+1}{2} \rightarrow \infty;$$

## Empirical Takeaway

---

- Correlation does not imply causation – **especially** in time series;
- Regressions with non-stationary data can be very misleading;
- Regressions with non-stationary data only make sense in a specific context (Cointegration – more on that later);
- Also important: the asymptotic theory we saw so far **does not apply** to non-stationary data;
- We need completely new theorems – and these are one order of magnitude more sophisticated;
- Good stuff for a second-year class, right? 😊

**Questions?**

## Persistence vs Unit Roots

---

## Persistence vs Unit Roots

- Let's say you have  $y_t$  and you are interested in knowing whether it is stationary or not;
- First, to determine the level of dependence you estimate an  $AR(1)$  model:

$$y_t = \rho y_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{iid } N(0, \sigma^2), \quad y_0 = 0;$$

- The OLS estimator is:

$$\hat{\rho} = \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} = \rho + \frac{\sum_{t=1}^T y_{t-1} \epsilon_t}{\sum_{t=1}^T y_{t-1}^2};$$

## Persistence vs Unit Roots

- Let's say you have  $y_t$  and you are interested in knowing whether it is stationary or not;
- First, to determine the level of dependence you estimate an  $AR(1)$  model:

$$y_t = \rho y_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{iid } N(0, \sigma^2), \quad y_0 = 0;$$

- The OLS estimator is:

$$\hat{\rho} = \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} = \rho + \frac{\sum_{t=1}^T y_{t-1} \epsilon_t}{\sum_{t=1}^T y_{t-1}^2};$$

- Let's say you get  $\hat{\rho} = 0.96$ ;
- Is this just a very persistent series or is it non-stationary?

## Persistence vs Unit Roots

- Let's say you have  $y_t$  and you are interested in knowing whether it is stationary or not;
- First, to determine the level of dependence you estimate an  $AR(1)$  model:

$$y_t = \rho y_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{iid } N(0, \sigma^2), \quad y_0 = 0;$$

- The OLS estimator is:

$$\hat{\rho} = \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} = \rho + \frac{\sum_{t=1}^T y_{t-1} \epsilon_t}{\sum_{t=1}^T y_{t-1}^2};$$

- Let's say you get  $\hat{\rho} = 0.96$ ;
- Is this just a very persistent series or is it non-stationary?
- If  $\rho < 1$ , we already know that  $\sqrt{T}(\hat{\rho} - \rho) \xrightarrow{d} N(0, \sigma^2(1 - \rho^2))$ .

## What happens if $\rho = 1$ ?

---

- If  $\rho = 1$ , this distribution vanishes – the variance goes to zero;
- This is already hinting at the need of a different asymptotic theory;

## What happens if $\rho = 1$ ?

- If  $\rho = 1$ , this distribution vanishes – the variance goes to zero;
- This is already hinting at the need of a different asymptotic theory;
- If  $\rho = 1$ , we have that  $y_t \sim N(0, \sigma^2 \cdot t)$ ;
- Notice that  $\mathbb{E}[u_t y_{t-1}] = 0$ ;
- Moreover:  $y_t^2 = y_{t-1}^2 + 2y_{t-1}u_t + u_t^2$ ;
- Using a telescoping sum and that  $y_0 = 0$ :

$$y_{t-1}u_t = \frac{1}{2} (y_t^2 - y_{t-1}^2 - u_t^2) \implies \frac{1}{T \cdot \sigma^2} \sum_{t=1}^T y_{t-1}u_t = \frac{1}{2} \left( \frac{y_T^2}{\sigma^2 T} - \sum_{t=1}^T \frac{u_t^2}{\sigma^2 T} \right);$$

## What happens if $\rho = 1$ ?

- If  $\rho = 1$ , this distribution vanishes – the variance goes to zero;
- This is already hinting at the need of a different asymptotic theory;
- If  $\rho = 1$ , we have that  $y_t \sim N(0, \sigma^2 \cdot t)$ ;
- Notice that  $\mathbb{E}[u_t y_{t-1}] = 0$ ;
- Moreover:  $y_t^2 = y_{t-1}^2 + 2y_{t-1}u_t + u_t^2$ ;
- Using a telescoping sum and that  $y_0 = 0$ :

$$y_{t-1}u_t = \frac{1}{2} (y_t^2 - y_{t-1}^2 - u_t^2) \implies \frac{1}{T \cdot \sigma^2} \sum_{t=1}^T y_{t-1}u_t = \frac{1}{2} \left( \frac{y_T^2}{\sigma^2 T} - \sum_{t=1}^T \frac{u_t^2}{\sigma^2 T} \right);$$

- What's the distribution of  $\frac{y_T^2}{\sigma^2 T}$ ?
- What's the probability limit of  $\sum_{t=1}^T \frac{u_t^2}{\sigma^2 T}$ ?

## What's the right rate?

- Putting it together:

$$\frac{1}{\sigma^2 T} \sum_{t=1}^T y_{t-1} u_t \xrightarrow{d} \frac{1}{2} (\chi^2 - 1);$$

- The numerator is  $O_p(T)$  and the limiting distribution is non-normal;

## What's the right rate?

- Putting it together:

$$\frac{1}{\sigma^2 T} \sum_{t=1}^T y_{t-1} u_t \xrightarrow{d} \frac{1}{2} (\chi^2 - 1);$$

- The numerator is  $O_p(T)$  and the limiting distribution is non-normal;
- Now we shall study the denominator. Notice that

$$\mathbb{E} \left[ \sum_{t=1}^T y_{t-1}^2 \right] = \sigma^2 \sum_{t=1}^T (t-1) = \sigma^2 \frac{(T-1)T}{2} \approx \sigma^2 \frac{T^2}{2};$$

- The denominator is  $O_p(T^2)$  we need to scale it by  $T^2$  to get meaningful limits;
- We can show that  $\frac{1}{\sigma^2 T^2} \sum_{t=1}^T y_{t-1}^2 \xrightarrow{d}$  (some complicated distribution based on functionals of a Brownian Motion);

What's the right rate?

---

## What's the right rate?

- The important thing to notice here is that the right rate is  $T$  and not  $\sqrt{T}$ :

$$T(\hat{\rho} - 1) = \frac{\frac{1}{T} \sum_{t=1}^T y_{t-1} u_t}{\frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2} \xrightarrow{d} \text{a complicated distribution}$$

- The critical values for this distribution are very different from the normal ones;
- They are fully known with any precision and can be computed through numerical simulation;
- Now we can build a test for  $H_0 : \rho = 1$  vs  $H_a : \rho < 1$ !
- The  $t$ -statistic is given by  $t = \frac{\hat{\rho}-1}{\hat{s}_T}$ , where  $\hat{s}_T = \frac{\frac{1}{T} \sum \hat{\epsilon}_t^2}{\sqrt{\sum_{t=1}^T y_t^2}}$ , which is the usual one;
- The  $t$ -statistic has a complicated and known distribution as well;

## What's the right distribution?

- The distribution of  $\hat{\rho}$  is actually more complicated than you think;
- It depends, even under the null, on two things:
  1. Whether the true process has a drift or not:  $y_t = \mu + \rho y_{t-1} + \epsilon_t$  VS  $y_t = \rho y_{t-1} + \epsilon_t$ ;
  2. Whether you included or not a constant and/or a time trend in the regression;

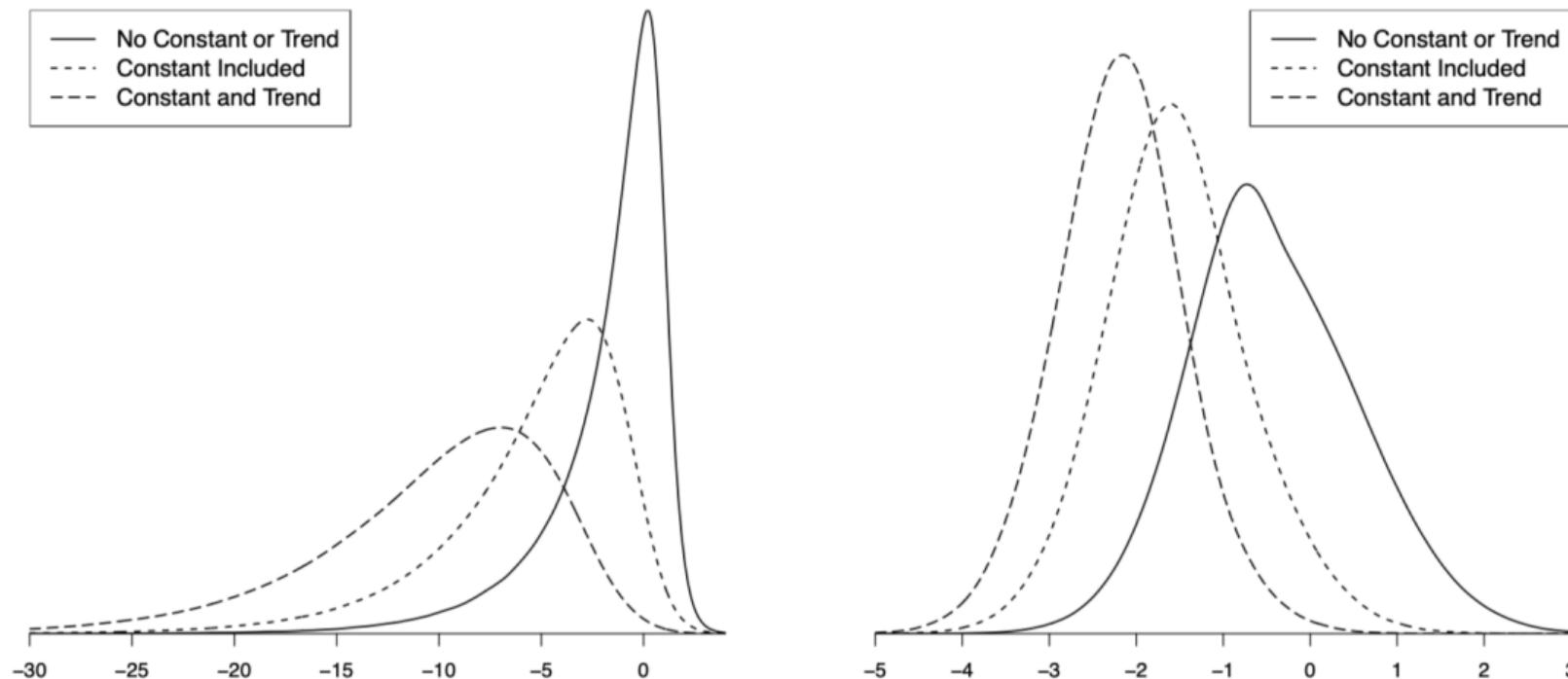
## What's the right distribution?

- The distribution of  $\hat{\rho}$  is actually more complicated than you think;
- It depends, even under the null, on two things:
  1. Whether the true process has a drift or not:  $y_t = \mu + \rho y_{t-1} + \epsilon_t$  VS  $y_t = \rho y_{t-1} + \epsilon_t$ ;
  2. Whether you included or not a constant and/or a time trend in the regression;
- A time trend is reasonable to consider since the drift implies a trend:

$$y_t = \mu \cdot t + \sum_{s=1}^t \epsilon_s;$$

- In any case, the distributions are known and available in statistical packages;
- The distributions for the  $t$ -statistic also change;

# These Distributions Are Skewed!



**Figure 1:** The Distributions for  $\hat{\rho}$  and the  $t$ -statistic

## AR( $p$ ) and Unit Roots

---

- What if the true process is an  $AR(p)$  with an unit root?

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t;$$

- Using our previous notation, a unit root is equivalent to  $\phi_1 + \phi_2 + \dots + \phi_p = 1$ ;

## AR( $p$ ) and Unit Roots

- What if the true process is an  $AR(p)$  with an unit root?

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t;$$

- Using our previous notation, a unit root is equivalent to  $\phi_1 + \phi_2 + \dots + \phi_p = 1$ ;
- Consider the vector  $\mathbf{y}_{t-1} = (y_{t-1}, \dots, y_{t-p})'$  and the  $p \times p$  matrix:

$$B = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & -1 \end{bmatrix}$$

- It maps  $\mathbf{y}_{t-1}$  to  $(y_{t-1}, \Delta y_{t-1}, \dots, \Delta y_{t-p+1})$ ;

## AR( $p$ ) and Unit Roots

- Let  $\rho \in \mathbb{R}$  and  $\beta \in \mathbb{R}^{p-1}$  be such that:

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}_{p \times 1} = B'_{p \times p} \begin{bmatrix} \rho \\ \beta \end{bmatrix}_{p \times 1} \implies y_t = \rho y_{t-1} + \beta_1 \Delta y_{t-1} + \dots + \beta_{p-1} \Delta y_{t-p+1} + \epsilon_t;$$

- It is a quick exercise to show that  $\rho = \phi_1 + \dots + \phi_p$ ;
- Testing for an unit root is equivalent to testing  $H_0 : \rho = 1$  in this regression;

## AR( $p$ ) and Unit Roots

- Let  $\rho \in \mathbb{R}$  and  $\beta \in \mathbb{R}^{p-1}$  be such that:

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}_{p \times 1} = B'_{p \times p} \begin{bmatrix} \rho \\ \beta \end{bmatrix}_{p \times 1} \implies y_t = \rho y_{t-1} + \beta_1 \Delta y_{t-1} + \dots + \beta_{p-1} \Delta y_{t-p+1} + \epsilon_t;$$

- It is a quick exercise to show that  $\rho = \phi_1 + \dots + \phi_p$ ;
- Testing for an unit root is equivalent to testing  $H_0 : \rho = 1$  in this regression;
- $T(\hat{\rho} - 1) \xrightarrow{d}$  (a complicated distribution) and  $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V)$  for some matrix  $V$ ;
- Notice that only the distribution of  $\hat{\rho}$  is non-standard;

## The Augmented Dickey-Fuller (ADF) Test

- The **Augmented Dickey-Fuller (ADF) Test** consists of testing whether  $\rho = 1$ ;
- The test statistic is the usual  $t$ -statistic:

$$t_{ADF} \equiv \frac{\hat{\rho} - 1}{\hat{s}_T}$$

where  $\hat{s}_T$  is the usual standard error of  $\hat{\rho}$  in the regression;

- The test **rejects the hypothesis of a unit root** for large negative values of  $t_{ADF}$ ;
- Reject an unit root if  $t_{ADF} < c_\alpha$  where  $c_\alpha$  is the  $\alpha$ -quantile of the asymptotic distribution;
- This asymptotic distribution depends on whether you include a constant and/or a time trend in the regression;

## Other Tests for Unit Roots

---

- This is not the only test for unit roots – there are other options;
- ADF is easy to implement and very popular, but it is very sensitive to the choice of  $p$ ;
- One reasonable way of choosing  $p$  is using an IC like AIC;

## Other Tests for Unit Roots

---

- This is not the only test for unit roots – there are other options;
- ADF is easy to implement and very popular, but it is very sensitive to the choice of  $p$ ;
- One reasonable way of choosing  $p$  is using an IC like AIC;
- The **Phillips-Perron (PP)** test has the same null and alternative hypotheses;
- The PP test is based on a non-parametric correction for the variance of  $\epsilon_t$ ;
- The PP test does not require choosing a lag length  $p$ ;

## Other Tests for Unit Roots

---

- This is not the only test for unit roots – there are other options;
- ADF is easy to implement and very popular, but it is very sensitive to the choice of  $p$ ;
- One reasonable way of choosing  $p$  is using an IC like AIC;
- The **Phillips-Perron (PP)** test has the same null and alternative hypotheses;
- The PP test is based on a non-parametric correction for the variance of  $\epsilon_t$ ;
- The PP test does not require choosing a lag length  $p$ ;
- Another popular guy: **KPSS** test, based on the residuals of a regression of  $y_t$  on a constant and/or a time trend;
- The KPSS test has the opposite null and alternative hypotheses:  $H_0$ : Stationary vs  $H_a$ : Unit Root;

## The Bayesian Critique

---

- All these tests are frequentist in nature – they rely on long-run frequency properties;
- There is another way of looking at this problem: Bayesian Inference;
- Start with a prior  $p(\rho)$ , use the likelihood  $p(y_1, \dots, y_T | \rho)$  to get the posterior  $p(\rho | y_1, \dots, y_T)$ ;
- You don't have to decide between  $|\rho| = 1$  vs  $|\rho| \neq 1$ ;
- In that sense the Bayesian approach is *unified*;
- Very cool paper about it: *Understanding Unit Rooters: A Helicopter Tour* by Sims and Uhlig (1991);
- Bayesian methods are *super popular* among the empirical macro and DSGE crowd!
- Very good stuff for another second-year class! Also, huge payoffs if you are good at coding;

**Questions?**

## Empirical Illustration

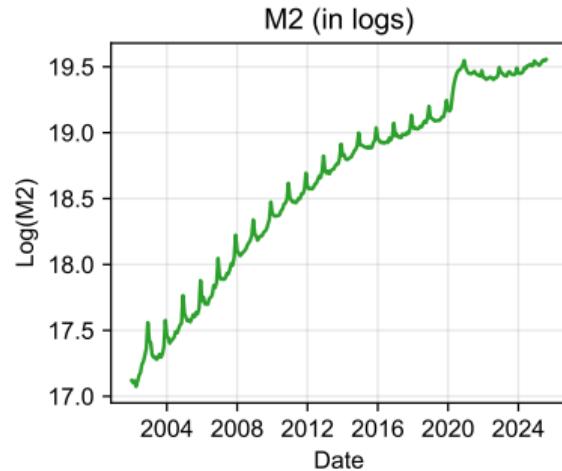
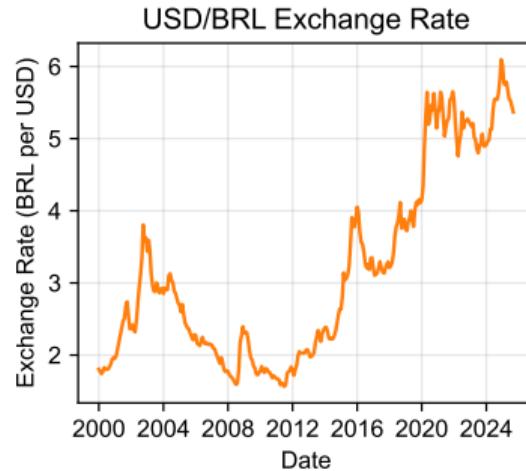
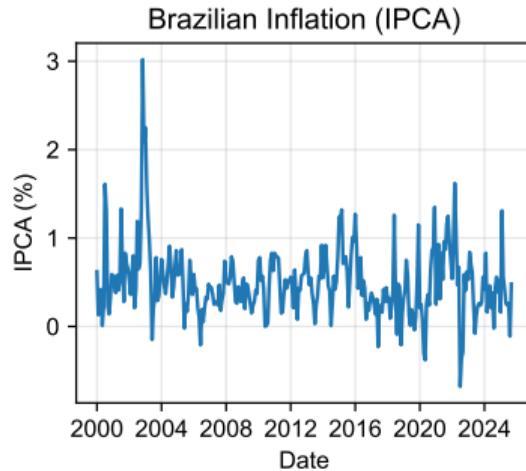
---

## Empirical Illustration

---

- Let's apply the ADF test to some real data;
- Let's do some unit root testing on three different series:
  1. Brazilian inflation (IPCA percentage increase, in levels);
  2. USD/BRL Exchange Rate (monthly average, in levels);
  3. The monetary base in Brazil (M2, in log);
- Our null hypothesis is that these series have a unit root (or are **integrated**);
- What kind of specification should we use? Constant only? Constant + Trend?

## Take a look at the data



- What kind of economic theories justify the presence (or lack thereof) of units roots?
- All series are at the monthly frequency here;
- Why does the monetary base have spikes? Would that violate stationarity?

## Running the ADF test

Series	Specification	ADF	p-value	5% Crit	Lags	N
IPCA	Constant	-8.7	0	-2.87	0	308
IPCA	Const+Trend	-8.8	0	-3.42	0	308
Exchange Rate	Constant	-0.76	0.83	-2.87	1	307
Exchange Rate	Const+Trend	-1.77	0.72	-3.42	1	307
M2 (log)	Constant	-2.84	0.05	-2.87	15	268
M2 (log)	Const+Trend	-0.88	0.96	-3.43	15	268

- Inflation: we reject the null across the board;
- For the exchange rate: we cannot reject the unit root;
- For M2: we only reject when **excluding** the trend;
- When we add the trend, we cannot reject the unit root;

## AR-DL Models

---

## Granger Causality

---

**Questions?**

The End

## References

- Chapter 11 on Hamilton's book discusses a bit of Granger Causality;
- Sections 14.40 and 14.44 on Hansen's book discuss AR-DL models and Granger Causality;
- See the Chapter 16 on Hansen's book about unit roots;
- Chapter 17 on Hamilton's book discusses non-stationary time series and unit roots;