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

Raul Riva

FGV EPGE

October, 2025

Intro

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 β here?
- What would you expect in terms of the estimate of β?

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}:

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

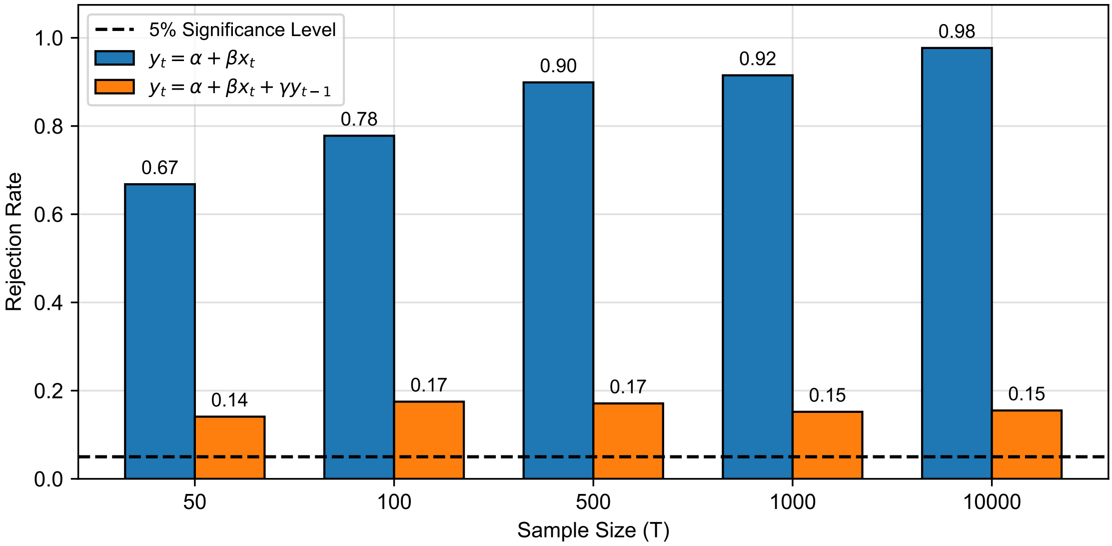
Simulation Setup

- 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 β 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 $x_0 = 0$):

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

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 $x_0 = 0$):

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

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

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

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:

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

- 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:

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

- 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 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:

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

- 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 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 $y_0 = 0$:

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

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 $y_0 = 0$:

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

- 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!

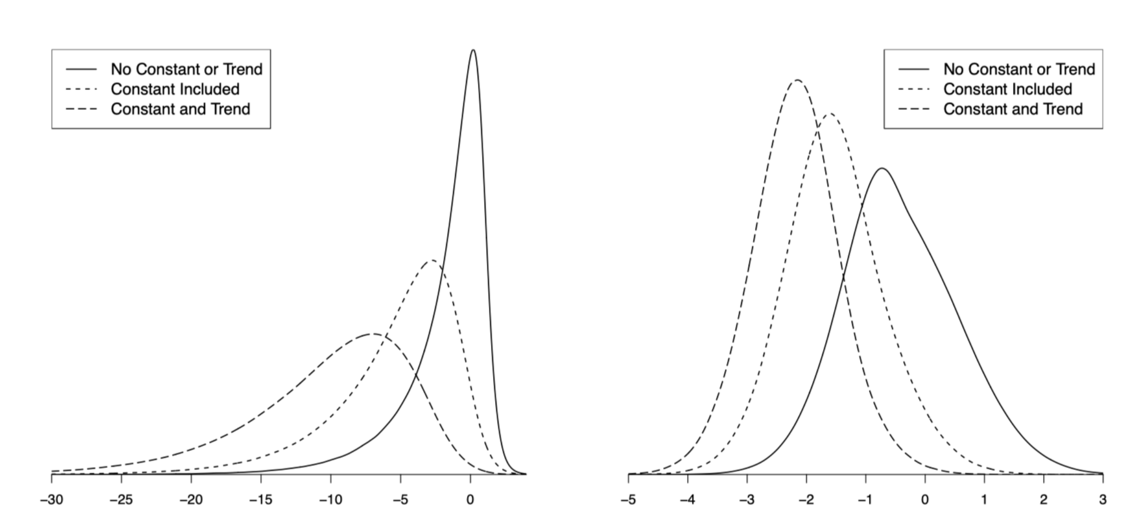


Figure 1: The Distributions for $\hat{\rho}$ and the t-statistic

AR(p) and Unit Roots

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

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

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

AR(p) and Unit Roots

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

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

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

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

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

AR(p) and Unit Roots

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

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

- It is a quick exercise to show that $\rho = \phi_1 + ... + \phi_p$;
- 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:

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

- It is a quick exercise to show that $\rho = \phi_1 + \ldots + \phi_p$;
- 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_α is the α-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 ϵ_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 ϵ_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;

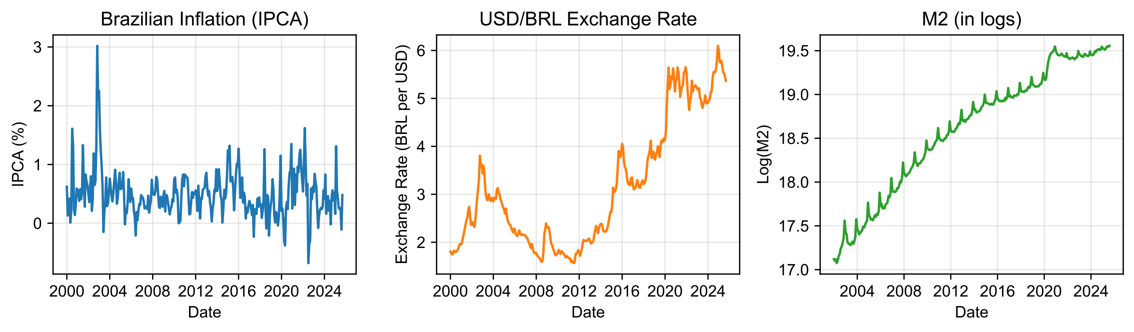
Questions?

Empirical Illustration

Empirical Illustration

- Let's apply the ADF test to some real data;
- Let's do some unit root testing on three diferent 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

An Example of Monetary Policy Adjustment

- Let i_t be the nominal interest rate set by the central bank at time t;
- Let π_t^e be the inflation expectation at t for $t+1$;
- Let r_t^* be an equilibrium real interest rate;
- Assume that $r_t^* = \alpha + \mathbf{x}_t'\beta$;
- $\mathbf{x}_t$ is a vector of other determinants of the equilibrium real rate;
- Examples: productivity growth, unemployment, foreign interest rates, risk premia, etc;
- Let $i_t^* = r_t^* + \pi_t^e$ be the "desired" nominal interest rate;

An Example of Monetary Policy Adjustment

- Let i_t be the nominal interest rate set by the central bank at time t;
- Let π_t^e be the inflation expectation at t for $t+1$;
- Let r_t^* be an equilibrium real interest rate;
- Assume that $r_t^* = \alpha + \mathbf{x}_t'\beta$;
- $\mathbf{x}_t$ is a vector of other determinants of the equilibrium real rate;
- Examples: productivity growth, unemployment, foreign interest rates, risk premia, etc;
- Let $i_t^* = r_t^* + \pi_t^e$ be the "desired" nominal interest rate;

At time t, assume that the central bank solves

$$\min_{i_t} \left((i_t - i_t^*)^2 + \theta\,(i_t - i_{t-1})^2 \right)$$

Can We Estimate This Rule?

- The solution is

$$i_t = \frac{1}{1+\theta} i_t^* + \frac{\theta}{1+\theta} i_{t-1} = \frac{1}{1+\theta} \left(\alpha + \mathbf{x}_t' \beta + \pi_t^e \right) + \frac{\theta}{1+\theta} i_{t-1}$$

- How would you recover θ from time series regressions?

Can We Estimate This Rule?

- The solution is

$$i_t = \frac{1}{1+\theta} i_t^* + \frac{\theta}{1+\theta} i_{t-1} = \frac{1}{1+\theta} \left(\alpha + \mathbf{x}_t' \beta + \pi_t^e\right) + \frac{\theta}{1+\theta} i_{t-1}$$

- How would you recover θ from time series regressions?
- Let's say you run this regression:

$$i_t = \alpha_0 + \alpha_1 i_{t-1} + \mathbf{x}_t' \delta + \gamma \pi_t^e + \varepsilon_t$$

- Under what conditions involving $(\mathbf{x}_t, \pi_t^e)$ and ε_t can we consistently estimate this rule using OLS?

Can We Estimate This Rule?

- The solution is

$$i_t = \frac{1}{1+\theta} i_t^* + \frac{\theta}{1+\theta} i_{t-1} = \frac{1}{1+\theta} \left(\alpha + \mathbf{x}_t' \beta + \pi_t^e\right) + \frac{\theta}{1+\theta} i_{t-1}$$

- How would you recover θ from time series regressions?
- Let's say you run this regression:

$$i_t = \alpha_0 + \alpha_1 i_{t-1} + \mathbf{x}_t' \delta + \gamma \pi_t^e + \varepsilon_t$$

- Under what conditions involving $(\mathbf{x}_t, \pi_t^e)$ and ε_t can we consistently estimate this rule using OLS?
- What if i_t also matters for $\mathbf{x_t}$? What about π_t^e? How to interpret coefficients then?
- What would be interesting examples here?

The AR–DL Model

- You just saw one example of an AR-DL model;
- An **AR-DL(p,q)** model posits the following dynamics for a general y_t:

$$y_t = \alpha + \sum_{i=1}^{p} \phi_i \, y_{t-i} + \sum_{j=0}^{q} \beta_j' \, \mathbf{x}_{t-j} + \varepsilon_t, \qquad E[\varepsilon_t] = 0$$

- When $p = 0$, it's called an **ADL(q)** model;
- Potentially quite useful for forecasting;
- On its own, essentially silent about causality. You typically need a model to make sense of these coefficients;
- Estimation can be done by OLS;
- Inference on coefficients should probably use Newey-West standard errors. Why?

Interpreting AR–DL Coefficients

- Consider $y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + ... + \beta_q x_{t-q} + \varepsilon_t$
- Notice that if x_t changes, it will potentially impact $y_t, y_{t+1}, ..., y_{t+q}$ directly;
- There might be indirect effects too if x_t is persistent;
- We are frequently interested in estimating $\beta_0 + \beta_1 + ... + \beta_q \implies$ a "long-run" effect;
- You can test hypotheses about this sum using linear hypothesis tests;
- The issue here is that x_t might induce change in x_{t+1}, x_{t+2}, etc;
- To be honest you should interpret β_j's as partial correlations, unless you have a full model;

Can We Ever Hope to Get to Causality?

- To get to causality, we typically need:
  - Exogenous variation in x_t;
  - An explicit link of how x_t affects y_t over time;

Can We Ever Hope to Get to Causality?

- To get to causality, we typically need:
  - Exogenous variation in x_t;
  - An explicit link of how x_t affects y_t over time;
- There is a **huge** literature on this and what people do is a VAR;
- A VAR is a system of AR-DL models where all variables are treated symmetrically:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{K,t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K \end{bmatrix} + \sum_{i=1}^{p} \begin{bmatrix} \phi_{11,i} & \phi_{12,i} & \cdots & \phi_{1K,i} \\ \phi_{21,i} & \phi_{22,i} & \cdots & \phi_{2K,i} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{K1,i} & \phi_{K2,i} & \cdots & \phi_{KK,i} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \\ \vdots \\ y_{K,t-i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{K,t} \end{bmatrix}$$

- Depending on assumptions about the shocks and the parameters, we can get to causal interpretations of the parameters;
- This is out of scope here, but very good stuff for a second-year class! 👀

Questions?

Granger Causality

Granger Causality

- There is another concept of causality in time series: **Granger Causality**;
- Let y_t and x_t be two stationary time series;
- We say that x_t does **not** Granger-cause y_t if:

$$\mathbb{E}(y_{t+1}|y_t, y_{t-1}, \dots, x_t, x_{t-1}, \dots) = \mathbb{E}(y_{t+1}|y_t, y_{t-1}, \dots)$$

Granger Causality

- There is another concept of causality in time series: **Granger Causality**;
- Let y_t and x_t be two stationary time series;
- We say that x_t does **not** Granger-cause y_t if:

$$\mathbb{E}(y_{t+1}|y_t, y_{t-1}, \ldots, x_t, x_{t-1}, \ldots) = \mathbb{E}(y_{t+1}|y_t, y_{t-1}, \ldots)$$

- If that's the case, x_t does not help predicting y_{t+1} once we know the past of y_t;
- Notice that this is a prediction-based notion of causality;
- To test for Granger causality, we can estimate an AR-DL model:

$$H_0 : \beta_0 = \beta_1 = \ldots = \beta_q = 0 \quad \text{vs} \quad H_1 : \text{at least one } \beta_j \neq 0$$

In VAR Terms

- With two variables, a VAR(p) model is given by:

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \sum_{i=1}^{p} \begin{bmatrix} \phi_{11,i} & \phi_{12,i} \\ \phi_{21,i} & \phi_{22,i} \end{bmatrix} \begin{bmatrix} y_{t-i} \\ x_{t-i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

- We would say that x_t does not Granger-cause y_t if:

$$H_0 : \phi_{12,1} = \phi_{12,2} = \ldots = \phi_{12,p} = 0 \quad \text{vs} \quad H_1 : \text{at least one } \phi_{12,i} \neq 0$$

- This is equivalent to having lower-triangular structure in the VAR;
- Very often, economic models can be tested through tests of Granger causality;
- Example: the information contained in the yield curve is a sufficient statistic for the future path of interest rates;
- Checkout *"Asymmetric Violations of the Spanning Hypothesis"*, by Freire and Riva;

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 Hansens book about unit roots;
- Chapter 17 on Hamilton's book discusses non-stationary time series and unit roots;