

# EAE1223: ECONOMETRIA III

## AULA 1 - INTRODUÇÃO

Luis A. F. Alvarez

26 de fevereiro de 2024

# AMOSTRAGEM E INFERÊNCIA

- Nos cursos anteriores de Econometria, boa parte das aulas foi dedicada a tópicos de **inferência estatística**, i.e. a métodos de quantificação da **incerteza** referente a uma quantidade populacional.
  - Cômputo de erros padrão, construção de testes de hipótese e intervalos de confiança.
- A incerteza, nesses cursos de Econometria, decorria fundamentalmente da **amostragem**.
  - Observávamos somente uma amostra da população de interesse, de modo que gostaríamos de quantificar o quanto poderíamos falar da população, e o quanto seria contingente à amostra.
- As propriedades dos estimadores, testes e intervalos de confiança faziam referência ao experimento mental de amostras repetidas.
  - Se repetíssemos a amostragem muitas vezes e calculássemos o estimador para cada amostra, na média das repetições o estimador acertaria parâmetro de interesse (ausência viés); ou as repetições não ficariam muito distantes de si (baixa variância amostral).

# AMOSTRAGEM ALEATÓRIA SIMPLES

- Nos cursos anteriores, as propriedades dos estimadores, testes e intervalos de confiança foram derivadas sob a hipótese de **amostragem aleatória simples**.
  - As observações na amostra eram independentes entre si, e cada uma seguia a mesma distribuição das variáveis na população.
  - Conceito serve como aproximação ao procedimento de se sortear  $n$  elementos (indivíduos, firmas, países) de uma população grande e coletar as informações deles.
  - Em dados no formato de painel, sorteamos aleatoriamente  $n$  indivíduos e acompanhamos suas variáveis por  $T$  períodos, de modo que observações de indivíduos diferentes são independentes, e a sequência de variáveis de cada indivíduo tem a mesma distribuição da população.

# SÉRIES DE TEMPO

- Neste curso, estudaremos métodos estatísticos e econométricos para dados de **séries de tempo**.
- Estes dados possuem uma estrutura especial, pois as variáveis possuem uma ordenação natural (tempo).
  - Exemplo: observamos dados de inflação e atividade econômica mensais, de janeiro de 2003 a dezembro de 2022.
- Esta estrutura nos leva a uma nova sorte de problemas.

# INCERTEZA EM SÉRIES DE TEMPO

- Primeiramente, o conceito de incerteza utilizado anteriormente necessita ser requalificado, pois é difícil ver uma série de tempo como uma amostra de uma população tal qual **naturalmente a pensamos**.
  - Em séries de tempo, a incerteza de procedimentos estatísticos decorre fundamentalmente de não sermos capazes de observar o que acontecerá em todo o futuro, nem o que aconteceu num passado distante, **nem o que poderia ter acontecido no período da amostra e não ocorreu**.
  - Essa incerteza **econômica** deve ser levada em conta, pois gostaríamos de separar o que é **contigente** ao período de análise do que é **essencial** na trajetória de uma série.
  - Propriedades dos estimadores devem ser pensadas como avaliadas em “amostras repetidas” desses cenários contrafactuais (“superpopulação”).

# O QUE MUDA

- A estrutura das séries de tempo faz com que a hipótese de dados independentes e identicamente distribuídos deixe de ser atrativa.
  - Esperamos que haja dependência do que ocorre com a inflação em  $t$  e aquilo que ocorre com ela em  $t - 1$ .
  - Além disso, em muitos cenários, é irrazoável supor que a distribuição de uma variável, que reflete o que contrafactualmente poderia ter ocorrido com ela, é a mesma em diferentes períodos.
    - A distribuição do produto Brasileiro em 1950 e 2000 não parece ser a mesma.
  - Precisamos de novos métodos para lidar com esses problemas.
- Por fim, a estrutura de séries de tempo requer cuidados adicionais na identificação e estimação de **efeitos causais**, pois devemos levar em conta as retroalimentações dos processos.

# CAUSALIDADE E ECONOMETRIA

- O conceito de causalidade, em Econometria, reflete o experimento mental associado ao qualificador *ceteris paribus*.
- Em um sistema econômico com variáveis  $(Y, X, U)$ , o efeito causal de  $X$  sobre  $Y$  é **definido** como o efeito de se perturbar  $X$  sobre  $Y$ , **mantidas as demais causas de  $Y$  constantes**.
- Primeira etapa de uma análise causal é definir o sistema econômico ou modelo causal, explicitando quais variáveis causam o quê e os efeitos causais de interesse.
  - Trata-se de **atividade puramente mental**, não dependente de amostra e envolvendo conhecimento prévio ou teoria econômica (Heckman e Pinto, 2022).

# MODELO ECONOMETRICO CAUSAL LINEAR

- Vamos definir o seguinte modelo causal para uma variável  $Y_t$  no período  $t$ .

$$Y_t = X_t' \beta + U_t,$$

onde  $X_t$  é um vetor  $k \times 1$  de **causas observadas** de  $Y_t$ ,  $\beta$  é um vetor  $k \times 1$  **definido** como o efeito causal de se perturbar cada um dos elementos de  $X_t$  sobre  $Y_t$ , e  $U_t$  são as demais causas **não observadas** de  $Y_t$ .

- **Exemplo:** definimos a função resposta de um banco central como:

$$i_t = (i^{\text{neutro}} + \pi^{\text{meta}}) + \gamma(\pi_{t-1} - \pi^{\text{meta}}) + u_t,$$

onde  $u_t$  são os determinantes não observados da regra de juros (choques monetários), e  $\gamma$  é definido como o coeficiente (causal) de resposta da política monetária à inflação passada.



## ESTIMAÇÃO DO MODELO CAUSAL LINEAR

- Suponha que tenhamos acesso a séries de tempo  $\{(Y_s, X_s)\}_{s=1}^T$ .
- Podemos tentar estimar o parâmetro causal  $\beta$  por MQO:

$$\hat{\beta} = \left( \sum_{t=1}^T X_t X_t' \right)^{-1} \left( \sum_{t=1}^T X_t Y_t \right)$$

- Mas:  $\hat{\beta} = \left( \sum_{t=1}^T X_t X_t' \right)^{-1} \left( \sum_{t=1}^T X_t (X_t' \beta + U_t) \right) = \beta + \left( \frac{1}{T} \sum_{t=1}^T X_t X_t' \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^T X_t U_t \right)$
- De Econometria I, sabemos que  $\hat{\beta}$  será não viciado se  $\mathbb{E}[U_s | X_1, X_2, \dots, X_T] = 0$  para  $s = 1, \dots, T$ .
  - Em Econometria 1, como as observações eram independentes,  $\mathbb{E}[U_s | X_1, X_2, \dots, X_T] = \mathbb{E}[U_s | X_s]$ , e a condição colapsava para  $\mathbb{E}[U_s | X_s] = 0$ .
  - Em séries de tempo, dependência implica que precisamos de  $\mathbb{E}[U_s | X_1, X_2, \dots, X_T] = 0 \implies \mathbb{E}[X_s U_t] = 0 \forall s, t$ .

# ESTIMAÇÃO DO MODELO CAUSAL LINEAR (CONT.)

- A condição que impomos para estimar  $\beta$  **sem viés** requer que  $\mathbb{E}[X_s U_t] = 0, \quad \forall s, t.$ 
  - Condição requer que os demais determinantes de  $Y$  não estejam sistematicamente associados com  $X$ , contemporânea ou extemporaneamente.
  - Não pode haver retroalimentação entre  $X$  e  $U$ .
  - A essa condição damos nome de **exogeneidade estrita**.
- Condição faz sentido no exemplo da regra monetária?

## ESTIMAÇÃO DO MODELO CAUSAL LINEAR (CONT.)

- Para estimar  $\hat{\beta}$  **consistentemente** quando  $T \rightarrow \infty$ , precisamos das seguinte condições:
  1.  $\left(\frac{1}{T} \sum_{t=1}^T X_t X_t'\right)$  e  $\left(\frac{1}{T} \sum_{t=1}^T X_t U_t\right)$  possuem limites em probabilidade.
  2. O limite em probabilidade de  $\left(\frac{1}{T} \sum_{t=1}^T X_t X_t'\right)$  tem posto cheio.
  3.  $\mathbb{E}[X_t U_t] = 0$  para todo  $t$ .
- Condição (1) é de natureza **estatística**: processos não podem explodir nem pode ser extremamente dependentes, de modo que as médias converjam (deve valer uma lei dos grandes números).
- Condição (2) versa sobre a **especificação** do modelo: não pode haver colinearidade perfeita em amostras grandes.
- Condição (3) é de natureza **econômica**: **exogeneidade contemporânea**.
  - Condição faz sentido no exemplo da regra monetária?

# MELHOR PREDITOR LINEAR

- E se não temos um modelo causal na cabeça? Ou se a hipótese de exogeneidade para consistência não vale? Será que o estimador de MQO estima consistentemente alguma coisa?

## DEFINIÇÃO: MELHOR PREDITOR LINEAR

Seja  $Y$  uma variável aleatória escalar e  $X$  um vetor  $k \times 1$  de preditores. O **coeficiente do melhor preditor linear** de  $Y$  como função de  $X$  é **definido** por:

$$\gamma \in \operatorname{argmin}_{c \in \mathbb{R}^k} \mathbb{E}[(Y - c'X)^2]$$

- $\gamma$  é o coeficiente que minimiza erro quadrático esperado de se projetar  $Y$  como função linear de  $X$ .

# MELHOR PREDITOR LINEAR E MODELO PREDITIVO

## PROPOSIÇÃO

1. Se  $\mathbb{E}[XX']$  tem posto completo, temos que  $\gamma$  é único e pode ser escrito como:

$$\gamma = \mathbb{E}[XX']^{-1}\mathbb{E}[XY]$$

2. Além disso, definindo o erro de projeção  $\epsilon = Y - X'\gamma$ , podemos escrever:

$$Y = X'\gamma + \epsilon,$$

onde,  $\mathbb{E}[X\epsilon] = 0$  **por construção**.

# ESTIMAÇÃO DE MODELO PREDITIVO

- A fórmula de  $\gamma$  é bastante parecida com a do estimador de MQO  $\hat{\beta}$ .
  - Consequentemente,  $\hat{\beta}$  estimará consistentemente o coeficiente  $\gamma$  do melhor preditor linear de  $Y$  em  $X$  se, quando  $T \rightarrow \infty$ ,  
 $\frac{1}{T} \sum_{t=1}^T X_t X_t' \xrightarrow{P} \mathbb{E}[XX']$  e  $\frac{1}{T} \sum_{t=1}^T X_t Y_t \xrightarrow{P} \mathbb{E}[XY]$ .
- Posto de outra forma, o problema de melhor prever  $Y$  como função linear de  $X$  define um **modelo preditivo linear**:

$$Y = X'\gamma + \epsilon = 0$$

cujo erro satisfaz a condição de exogeneidade  $\mathbb{E}[X\epsilon] = 0$  **por construção**. Para consistência, bastará então uma LGN.

- Condições suficientes para que a lei dos grandes números que garante a consistência do estimador de MQO  $\hat{\beta}$  para  $\gamma$  valha são:
  - (A)  $\mathbb{E}[X_t X_t'] = \mathbb{E}[XX']$  e  $\mathbb{E}[X_t Y_t] = \mathbb{E}[XY]$ .
  - (B) **As observações apresentam dependência fraca no tempo**, de modo que observações mais distantes no tempo se comportam cada vez mais como próximas de independentes.

# ESTACIONARIEDADE E DEPENDÊNCIA FRACA

- As condições (a) e (b) garantem que uma lei dos grandes números valha, de modo que  $\frac{1}{T} \sum_{t=1}^T X_t X_t' \xrightarrow{P} \mathbb{E}[XX']$  e  $\frac{1}{T} \sum_{t=1}^T X_t Y_t \xrightarrow{P} \mathbb{E}[XY]$ .
  - Condição (a) é uma versão de condição de **estacionariedade fraca**. Esta condição impõe estabilidade dos momentos de  $X$  e  $Y$ , garantindo que os potenciais limites das médias amostrais existam.
  - Condição (b) garante que haja informação suficiente no problema para que haja convergência em probabilidade para os limites.
- Uma boa parte do curso consistirá em prover ferramentas para testar se séries de tempo são estacionárias ou não, e a entender o que se deve fazer em caso negativo.
- Quanto ao item (b), não testaremos essa condição, embora seja importante saber que ela vale para processos bastante flexíveis (Carrasco e Chen, 2002).

# PROCEDIMENTO PARA AVALIAR CONSISTÊNCIA DE ESTIMADOR DE MQO DE MODELO LINEAR EM SÉRIES DE TEMPO

1. Especificar a natureza da pergunta (preditiva ou causal).
2. Postular e escrever o modelo correspondente.
3. Se o modelo é causal, avaliar a plausibilidade de exogeneidade contemporânea (teoria econômica). Se modelo é preditivo, exogeneidade contemporânea do erro do modelo vale por construção.
4. Se exogeneidade valer, e séries forem estacionárias (e fracamente dependentes), sabemos que estimador de MQO de  $\hat{\beta}$  será consistente para o parâmetro do modelo postulado.

**Obs 1:** Se a série não for estacionária, veremos no curso o que podemos dizer/devemos fazer.

**Obs 2:** Além disso, veremos como fazer inferência sobre os parâmetros, levando em conta a dependência temporal na quantificação de incerteza.



# BIBLIOGRAFIA I



Carrasco, Marine e Xiaohong Chen (2002). “Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models”. Em: *Econometric Theory* 18.1, pp. 17–39. ISSN: 02664666, 14694360. URL: <http://www.jstor.org/stable/3533024> (acesso em 23/02/2024).



Heckman, James J e Rodrigo Pinto (fev. de 2022). *Causality and Econometrics*. Working Paper 29787. National Bureau of Economic Research. DOI: 10.3386/w29787. URL: <http://www.nber.org/papers/w29787>.