

# Modelagem Econométrica Estrutural e Aplicações

## Aula Introdução – Parte 01

Claudio R. Lucinda

FEA/USP

2022



# Agenda

## 1 Abordagens à Econometria



# Agenda

- 1 Abordagens à Econometria
- 2 Retornos à Publicidade



# Agenda

- ① Abordagens à Econometria
- ② Retornos à Publicidade
- ③ O “Zoológico da Identificação”
  - Modelagem Econométrica Estrutural



# Modelagem Econométrica

- Eu consigo ver duas formas pelas quais você pode utilizar as técnicas econométricas (não só eu, Haile (2019) também):
  - **Descritiva:** Utilizar os métodos como formas de caracterizar a distribuição conjunta – Não se enganem, existe muito trabalho bom, valioso na academia e fora dela, que pode ser feito usando isso. E aqui dentro eu incluo os aspectos de previsão e classificação.
  - **Causal:** Utilizar os métodos como formas de descobrir relações causais. Hoje em dia este assunto é cada vez mais importante porque é a forma pela qual você pode avaliar os efeitos de uma política, e se você quer justificar esta política em termos de custo-benefício, esse é o caminho.



# Modelagem Econométrica Causal

- Dentro dessa modelagem causal, a prática criou duas vertentes:
  - **“Forma Reduzida”**: Usualmente entendida como modelos lineares estimados por RCT/IV/DID/RDD:
  - Modelos que tentam explorar a randomização (ou quasi-randomização)
  - O nome do jogo: “estratégia de identificação”
  - **Modelos Estruturais**: Modelos, usualmente não lineares nos parâmetros, que são mais difíceis de estimar:
  - Esses modelos deixam claras quais são as (usualmente numerosas) hipóteses de identificação
  - Métodos que se focam em situações em que RCT são inviáveis



# “Forma Reduzida” X Modelos Estruturais

- Existe muita animosidade entre os fãs de cada uma dessas pseudo-vertentes.
- Na prática, nós precisamos das duas abordagens
- Mas a profissão investe muito menos na segunda do que na primeira, porque:
  - A implementação de um paper estrutural é muito mais trabalhosa do que um de forma reduzida
  - Um paper estrutural exige um conhecimento prévio de outras áreas que nem sempre o pesquisador possui ou ele existe de fato.
- Isso tem implicações sobre a carreira das pessoas, o que se reflete na próxima geração de economistas.
- Só que todo modelo causal é um modelo estrutural baseado em premissas!



# Retornos à Publicidade

- Vou agora falar de um exemplo de como todo modelo causal é um modelo estrutural baseado em premissas e que, se você se esquecer disso, vai se ver com problemas.
- O exemplo é sobre uma pergunta: **Como sabemos se publicidade paga em sites de busca funciona?**
- A fonte é o *Substack* do Scott Cunningham
- [https://causalinf.substack.com/p/how-can-we-know-if-paid-search-advertising?utm\\_source=twitter&s=r](https://causalinf.substack.com/p/how-can-we-know-if-paid-search-advertising?utm_source=twitter&s=r)





## Retornos à Publicidade – Background

- O que a Google vende e o que a Google faz são coisas diferentes. Ela “faz” uma redução nos custos de busca das pessoas por meio do seu mecanismo de busca. Mas o que ela vende é publicidade. Uma proporção gigantesca da receita dela vem de vender anúncios.
- Quando vc entra na caixa de busca e coloca “relógios eBay”, vão aparecer dois tipos de resultados:
  - Busca Orgânica: Links para os sites mais populares
  - Busca Patrocinada: Links para os sites de quem pagou
- O ponto é que eles são mais ou menos substitutos perfeitos – e, se são, porque uma empresa deveria pagar pela Busca Patrocinada?
- Ou seja, temos uma questão que é CAUSAL.



## Questões Causais, RCT e Quasi-Experimentos

- Março de 2012 - o eBay realizou um experimento para saber se publicidade paga por palavra chave aumentava a receita.
- Isso virou um paper na Econometrica (“Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment”)
- Resultado: sem o pagamento, os cliques nos links pagos foram para zero, mas houve um aumento equivalente no número de cliques que não são pagos. Ou seja, no final não mudaram os volumes de cliques.



## eBay e Cliques

## PAID SEARCH EFFECTIVENESS

161

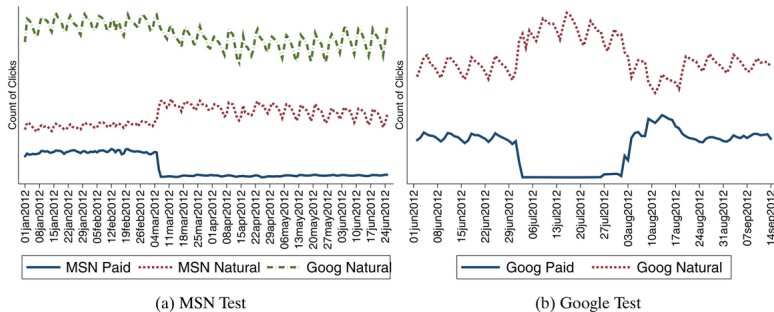


FIGURE 2.—Brand keyword click substitution. MSN and Google click-traffic counts to eBay on searches for ‘ebay’ terms are shown for two experiments where paid search was suspended (panel (a)) and suspended and resumed (panel (b)).



# eBay e Cliques (2)

TABLE I  
RETURN ON INVESTMENT<sup>a</sup>

|                                       | OLS      |          | IV       |          | DnD      |                           |
|---------------------------------------|----------|----------|----------|----------|----------|---------------------------|
|                                       | (1)      | (2)      | (3)      | (4)      | (5)      |                           |
| Estimated Coefficient                 | 0.88500  | 0.12600  | 0.00401  | 0.00188  | 0.00659  | A                         |
| (Std Err)                             | (0.0143) | (0.0404) | (0.0410) | (0.0016) | (0.0056) |                           |
| DMA Fixed Effects                     |          | Yes      |          | Yes      | Yes      |                           |
| Date Fixed Effects                    |          | Yes      |          | Yes      | Yes      |                           |
| N                                     | 10,500   | 10,500   | 23,730   | 23,730   | 23,730   |                           |
| $\Delta \ln(\text{Spend})$ Adjustment | 3.51     | 3.51     | 3.51     | 3.51     | 1        | B                         |
| $\Delta \ln(\text{Rev})$ ( $\beta$ )  | 3.10635  | 0.44226  | 0.01408  | 0.00660  | 0.00659  | C = A * B                 |
| Spend (Millions of \$)                | \$51.00  | \$51.00  | \$51.00  | \$51.00  | \$51.00  | D                         |
| Gross Revenue (R')                    | 2,880.64 | 2,880.64 | 2,880.64 | 2,880.64 | 2,880.64 | E                         |
| ROI                                   | 4,173%   | 1,632%   | -22%     | -63%     | -63%     | F = A/(1 + A) * (E/D) - 1 |
| ROI Lower Bound                       | 4,139%   | 697%     | -2,168%  | -124%    | -124%    |                           |
| ROI Upper Bound                       | 4,205%   | 2,265%   | 1,191%   | -3%      | -3%      |                           |

<sup>a</sup>The upper panel presents regression estimates of SEM's effect on sales. Columns (1) and (2) naively regress sales on spending in the pre-experiment period. Columns (3) and (4) show estimates of spending's effect on revenue using the difference-in-differences indicators as excluded instruments. Column (5) shows the reduced form difference-in-differences interaction coefficient. The lower panel translates these estimates into a return on investment (ROI) as discussed in Section 4 and shows its 95% confidence interval.

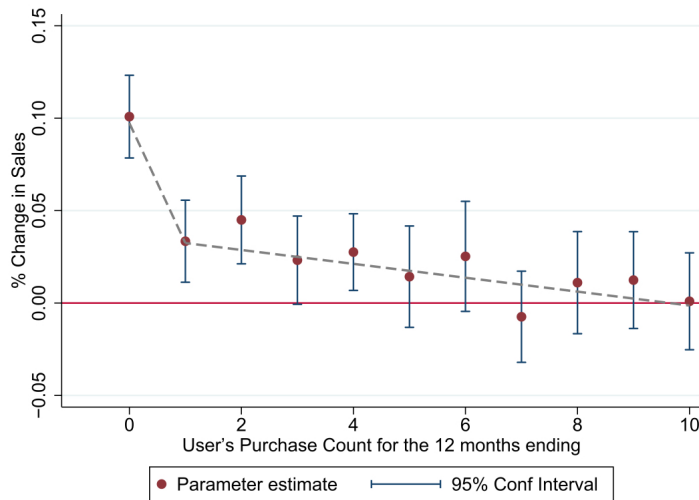


# ATT, LATE, Heterogeneidade

- Na última coluna a gente tem o ATT (Efeito Médio nos Tratados), enquanto nas com IV a gente tem o LATE (Local Average Treatment Effect)
- Isso pode significar que existe uma heterogeneidade nos efeitos, que eles documentam.
- Eles conseguem relacionar os links aos perfis de usuários – quem fez compra ou não.
- Para os usuários experientes, a publicidade não teria efeitos. Mas para os usuários novos, a publicidade era importante por informar a eles onde as coisas estavam.



# Heterogeneidade de Efeitos



# Heterogeneidade de Efeitos Importa!

- Este problema é importante não apenas no eBay.
- Estudo de 2019, “A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook” apresenta o resultado de 15 RCTs, e encontram resultados positivos.
- Tais efeitos variam dependendo de onde os anúncios aconteciam na plataforma.
- O mais legal do paper, para mim, é que ele mostra que métodos que não envolvem a randomização podem gerar resultados fortemente viesados (os exemplos dele são *matching*, estratificação e ajustamento por covariadas)



## Vieses

Figure 10: Summary of lift results

| Campaign | Outcome      | (A)      | (C)         | (D)                       | (E)                                       | (F)   | (G)  | (H)                                       | (I)  | (J)   | (K)                      | (L)                      | (M)                                       | (N)   | (O)  |
|----------|--------------|----------|-------------|---------------------------|---|---|--|---|--|---|--------------------------|--------------------------|---|---|--|
|          |              | RCT Lift | EM          | Propensity Score Matching |   |   |  |   |  | Regression Adjustment   |                          |                          |   | Stratified Regression                       |  |
|          |              |          | Age, Gender | Age, Gender<br>+ FB Vars  | Age, Gender<br>+ FB Vars<br>+ Census Vars | Age, Gender<br>+ FB Vars<br>+ Activity Vars | Age, Gender<br>+ FB Vars<br>+ Activity Vars<br>+ FB Match Vars | Age, Gender<br>+ FB Vars<br>+ Census Vars | Age, Gender<br>+ FB Vars<br>+ Census Vars<br>+ Activity Vars | Age, Gender<br>+ FB Vars<br>+ Census Vars<br>+ Activity Vars<br>+ FB Match Vars | Age, Gender<br>+ FB Vars | Age, Gender<br>+ FB Vars | Age, Gender<br>+ FB Vars<br>+ Census Vars | Age, Gender<br>+ FB Vars<br>+ Activity Vars | Age, Gender<br>+ FB Vars<br>+ Activity Vars<br>+ FB Match Vars |
| 1        | Checkout     | 90%      | 116%        | 109%                      | 107%                                      | 85%   | 93%  | 104%                                      | 99%  | 88%   | 76%                      | 101%                     | 94%                                       | 65%   | 51%  |
| 2        | Checkout     | 1.3%     | 432%        | 161%                      | 140%                                      | 37%   | 36%  | 149%                                      | 140%   | 43%   | 33%                      | 97%                      | 90%                                       | 54%   | 40%  |
| 3        | Checkout     | 8.8%     | 65%         | 20%                       | 24%                                       | 41%   | 17%  | 21%                                       | 23%  | 38%   | 5%                       | 18%                      | 19%                                       | 30%   | 2%   |
| 4        | Checkout     | 73%      | 222%        | 145%                      | 131%                                      | 143%  | 95%  | 126%                                      | 122%   | 134%  | 100%                     | 98%                      | 87%                                       | 96%   | 74%  |
| 5        | Checkout     | 450%     | 511%        | 418%                      | 443%                                      | 463%  | 316%   | 428%                                      | 432%   | 437%  | 305%                     | 447%                     | 431%                                      | 435%  | 301%   |
| 7        | Checkout     | 2.7%     | 37%         | 30%                       | 18%                                       | 33%   | 36%  | 19%                                       | 20%  | 33%   | 35%                      | 19%                      | 19%                                       | 31%   | 33%  |
| 8        | Checkout     | -2.9%    | 48%         | 31%                       | 36%                                       | 30%   | 27%  | 36%                                       | 41%  | 54%   | 29%                      | 32%                      | 37%                                       | 52%   | 28%  |
| 9        | Checkout     | 2.4%     | 1414%       | 2062%                     | 1970%                                     | 2314%                                       | 1710%  | 1994%                                     | 1999%  | 2319%   | 1716%                    | 1962%                    | 1962%                                     | 2210%                                       | 1616%  |
| 10       | Checkout     | 2.0%     | 38%         | 23%                       | 16%                                       | 43%   | -7%  | 30%                                       | 20%  | 34%   | -13%                     | 21%                      | 21%                                       | 35%   | -11%   |
| 11       | Checkout     | 9%       | 275%        | 29%                       | 31%                                       | 38%   | 7%   | 90%                                       | 31%  | 35%   | 3%                       | 30%                      | 31%                                       | 34%   | 2%   |
| 12       | Checkout     | 1%       | 128%        | 113%                      | 130%                                      | 82%   | 82%  | 112%                                      | 113%   | 82%   | 81%                      | 112%                     | 113%                                      | 86%   | 82%  |
| 13       | Checkout     | -15%     | -39%        | -35%                      | -36%                                      | -50%  | -31%   | -35%                                      | -35%   | -31%  | -30%                     | -35%                     | -35%                                      | -31%  | -30%   |
| 14       | Checkout     | 62%      | 119%        | 80%                       | 85%                                       | 95%   | 101%   | 80%                                       | 83%  | 92%   | 90%                      | 74%                      | 77%                                       | 82%   | 84%  |
| 15       | Checkout     | 2%       | 26%         | -10%                      | -9%                                       | -10%  | -13%   | -9%                                       | -9%  | -11%  | -14%                     | -9%                      | -9%                                       | -12%  | -14%   |
| 1        | Registration | 781%     | 1024%       | 978%                      | 944%                                      | 1060%                                       | 977%   | 968%                                      | 1087%  | 983%  | 824%                     | 800%                     | 432%                                      | 348%  |  |
| 5        | Registration | 893%     | 1270%       | 1071%                     | 1055%                                     | 1070%                                       | 765%   | 1067%                                     | 1067%  | 1063%   | 728%                     | 1112%                    | 1104%                                     | 1081%                                       | 772%   |
| 8        | Registration | 63%      | 180%        | 162%                      | 159%                                      | 173%  | 167%   | 150%                                      | 153%   | 158%  | 114%                     | 157%                     | 161%                                      | 160%  | 125%   |
| 10       | Registration | 9%       | 33%         | 19%                       | 18%                                       | 33%   | -3%  | 18%                                       | 18%  | 31%   | 0%                       | 19%                      | 18%                                       | 31%   | 2%   |
| 14       | Registration | 158.1%   | 275%        | 215%                      | 219%                                      | 244%  | 241%   | 219%                                      | 219%   | 238%  | 234%                     | 219%                     | 218%                                      | 240%  | 239%   |
| 2        | Page View    | 1517%    | 4261%       | 2493%                     | 2416%                                     | 1150%                                       | 1177%  | 2408%                                     | 2422%  | 1175%   | 1187%                    | 1162%                    | 1181%                                     | 1722%                                       | 1268%  |
| 5        | Page View    | 609%     | 846%        | 771%                      | 731%                                      | 719%  | 484%   | 751%                                      | 748%   | 710%  | 477%                     | 776%                     | 769%                                      | 717%  | 498%   |
| 6        | Page View    | 14%      | 227%        | 163%                      | 169%                                      | 263%  | 255%   | 163%                                      | 160%   | 250%  | 246%                     | 111%                     | 113%                                      | 253%  | 278%   |

\* Red: RCT Lift is statistically different from 0 at 5% significance level

Red: Observational method overestimates lift

Blue: Observational method underestimates lift

Color proportional to overestimation factor; darkest color reached at 3-times over- or underestimation





## Porquê dos Vieses?

- Dá errado porque eles dependem de duas hipóteses, uma delas mais importante aqui. A de **independência condicional**
- Em outras palavras, depois de condicionarmos em um conjunto de variáveis  $\mathbb{X}$ , toda variação no tratamento é independente do resultado – ou seja, o tratamento é “como se fosse randomizado”.
- Essa hipótese é ESTRUTURAL.
- Palavras do Scott:

*But believing in conditional independence is somewhat akin to believing in Big-foot. Without extremely good evidence for conditional independence, extreme skepticism is warranted because the behavioral process that generates data of this kind is based on intentional human choices that may not be random at any step, or at least not easily identifiable randomness.*



## Mas RCT é estrutural também!

- Vocês poderiam me dizer: “mas o RCT não tem esse problema, como a tabela mostra”.
- Só que p RCT **também** é um modelo estrutural – só que os parâmetros de interesse podem ser estimados com estatísticas descritivas! ( $\bar{y}_T - \bar{y}_C$ ).
- As hipóteses estruturais vieram antes – na etapa de randomização
- A randomização foi feita pensando em um modelo estrutural (não existe um experimento ideal sem um modelo por trás)



# Limites dos RCT

- Tá, mesmo que RCT seja um modelo estrutural, porque não podemos ficar só com os RCT – se quiser chamando de estrutural?
- Algumas razões:
  - ① Algumas questões causais não podem ser respondidas com um RCT
  - ② Por várias razões a experimentação pode não ser viável (razões éticas, de custos, etc e tal)
- Ou seja, vamos ter que lidar com dados observacionais
- E não vamos escapar de especificar como os não observáveis do modelo se relacionam com as outras partes.



# “Parâmetros Profundos” e Modelos Econômicos

- A forma mais fácil para o economista de especificar essa relação é por meio de um modelo econômico.
- Em alguns campos na Economia isso é mais fácil (Organização Industrial é um Exemplo), em outros é mais difícil.
- Esse modelo vai ser base para o Processo Gerador dos Dados, vai depender de Parâmetros e eles serão invariantes à política analisada, além de relacionados aos **fundamentos econômicos**, como por exemplo:
  - Preferências
  - Bens
  - Características Demográficas
  - Tecnologia de Produção
  - Informação e Expectativas
  - Espaço (Interações Sociais, localização)



# Identificação

- Uma questão importante em econometria é a de **identificação** – definida aqui como os parâmetros do modelo poderem ser determinados de forma única a partir da população observável que gerou os dados.
- Isso não é uma questão da **amostra** que temos, e sim do **modelo** e da **população** que temos.
- Identificação é um passo anterior à análise.
- Para que um modelo tenha parâmetros identificados, isso significa que se os valores deles fossem diferentes nós teríamos diferentes distribuições dos dados observáveis.
- Se isso não ocorre (a mesma distribuição dos dados observáveis pode acontecer com diferentes distribuições dos parâmetros), não temos como achar um estimador pra esses parâmetros.



# Identificação em “Forma Reduzida” e em Modelos Estruturais

- Em “forma reduzida”:
  - Ênfase na “estratégia de identificação” (processo de randomização ou porquê seu contexto é “tão bom quanto randomizado”)
  - O foco é nos efeitos de tratamento, não nos parâmetros do modelo do DGP
- Em Modelos Estruturais:
  - Ênfase em garantir que vc tenha variabilidade nos dados para identificar os parâmetros
  - Detalhar quais hipóteses vc consegue garantir a identificação como acima mencionado.



# Métodos Estruturais

- Diferentemente dos modelos de “forma reduzida”, não existe um pacote básico de técnicas.
- Como a abordagem depende da especificação dos modelos, as coisas são meio caso-a-caso.
- Os métodos usados são geralmente não lineares nos parâmetros
- Conhecimentos Computacionais ajudam bastante nessa hora



# Modelos Estruturais em vários campos

- Labor: Keane and Wolpin (1997), decisões de investimento em educação
- Org. Industrial: Berry, Levinsohn, and Pakes (1995), estimação de demanda
- Econ. Urbana: Ahlfeldt, Redding, Sturm, and Wolf (2015), estimação de aglomeração espacial
- Meio Ambiente: Rudik (2020), quantificar a incerteza nos IAM de meio ambiente
- Setor Público: Bayer, McMillan, Murphy, and Timmins (2016), dynamic Tiebout sorting model
- Macro: Modelos DSGE
- Internacional: Jin and Shen (2020), coordenação de dívida soberana





# Validade Externa e Interna

- **Validade Interna** diz respeito a quão “causal” é o parâmetro estimado
- **Validade Externa** diz respeito à quão generalizáveis são as estimativas em outros contextos
  - Abordagens de forma reduzida são boas em validade interna e ruins em validade externa
  - Por outro lado, se os agentes se comportam de forma parecida em diferentes contextos, modelos estruturais podem ser válidos



# Tá, mas o que usar?

- A resposta sofisticada e elegante é: **DEPENDE**
- Da natureza da questão a ser respondida
- O tipo de dados disponível
- O mecanismo pelo qual os indivíduos “recebem a política” – aleatório ou escolha dos indivíduos.



# Abordagens à Modelagem Econométrica

- Podemos dividir a modelagem econométrica em duas linhas: descritiva e estrutural.
- Vamos entender a diferença entre as duas imaginando a distribuição conjunta entre as variáveis cuja relação se busca entender,  $f(x, y)$ .
- Coisas específicas que se desejam caracterizar:
  - A distribuição condicional de  $y$  dado  $x$ ,  $f(y|x)$ ;
  - A Esperança condicional de  $y$  dado  $x$ ,  $E(y|x)$ ;
  - A Correlação (ou covariância) condicional de  $y$  dado  $x$ ,  $Corr(y|x)$  ou  $Cov(y|x)$
  - Um quantil específico  $\alpha$  da distribuição de  $y$  dado  $x$   $Q_\alpha(y|x)$ ;
  - O Melhor Preditor Linear de  $y$  dado um valor para  $x$   $BLP(y|x)$



# Econometria Estrutural e Econometria Descritiva

- Nos modelos descritivos a idéia principal é caracterizar simplesmente a distribuição conjunta.
- Na abordagem econométrica estrutural buscam-se parâmetros ou primitivas econômicas da distribuição conjunta
- Note-se que **a busca destas primitivas ou parâmetros da distribuição conjunta é sempre dependente destas premissas que limitam a distribuição conjunta.**
- Os elementos essenciais de um modelo estrutural são as hipóteses econômicas e estatísticas, as quais deveriam ser, pelo menos, razoáveis econômica e estatisticamente.
- Note-se: mesmo que você não derive explicitamente um modelo estrutural, **qualquer conclusão de ordem causal ou comportamental está implicitamente se baseando em um modelo estrutural.**



# Modelagem Estrutural, “Experimentos Naturais” e “Forma Reduzida”

- Experimentos Naturais são excelentes para identificar o efeito de interesse localmente, mas:
  - Nem sempre existe um experimento para identificar o efeito que você quer.
  - A validade de um experimento natural é bem restrita para fora do local onde o efeito é identificado.
  - E não esqueçam: o “ideal” do “experimento ideal” sempre é com base em um modelo teórico.
- “Forma Reduzida” às vezes pode ser bastante útil para responder perguntas relevantes, especialmente quando:
  - Quando a teoria aponta para efeitos determinados para um determinado coeficiente ou conjunto de coeficientes.
  - Se você tem a equação da forma reduzida exatamente identificada você até consegue voltar com os parâmetros estruturais

