

# Aula 00 - Parte 2

## Fluxo de Trabalho em um Modelo Estrutural

Claudio R. Lucinda

FEA/USP

2022



# Agenda

## 1 Fluxo de Trabalho em Modelagem Estrutural



# Agenda

- 1 Fluxo de Trabalho em Modelagem Estrutural
- 2 Um Exemplo Simplificado
  - Uma visão estrutural da Equação Minceriana



# Agenda

- O que vamos fazer hoje
  - Que passos são necessários para estimar um modelo estrutural?
  - Seguir cada etapa em um modelo simplificado
- O que não vamos fazer hoje
  - O material que discutimos hoje já assumirá que você tem dados
  - E que você tenha conhecimento suficiente de seus dados
  - Também pressupõe que você entenda sua linguagem de codificação preferida
  - Estas são todas as etapas não triviais, mas normalmente são abordadas em outras disciplinas
  - Vou (em certa medida) tentar ajudá-lo a desenvolver essas habilidades ao longo do curso



# Etapas para a Modelagem Estrutural

- Mike Keane deu uma PALESTRA na Universidade de Chicago em 2015 e listou estas etapas:
  - ① Desenvolvimento do Modelo Teórico
  - ② Questões Práticas de Especificação
  - ③ Resolvendo o Modelo
  - ④ Entendendo como o modelo funciona
  - ⑤ Estimativa
  - ⑥ Validação
  - ⑦ Experimentos de política



# Um Exemplo Simplificado

- Para ajudar a corrigir ideias, vamos revisitar um modelo comumente usado em econometria introdutória:

$$\log(w_i) = \beta_0 + \beta_1 s_i + \beta_2 x_i + \beta_3 x_i^2 + \varepsilon_i \quad (1)$$

- Com dados transversais e em que
- $i$  indexa indivíduos
- $w_i$  é a renda do emprego
- $s_i$  são anos de escolaridade
- $x_i$  são anos de experiência de trabalho (ou, mais comumente, *experiência de trabalho potencial*)
- $\varepsilon_i$  é qualquer outra coisa que determina a renda
- Queremos estimar  $(\beta_1, \beta_2, \beta_3)$ , que são *retornos ao investimento em capital humano*



# Vieses

- É quase certo que a equação (1) sofre de viés de variável omitida
- Ou seja, existem muitos fatores em  $\varepsilon_i$  que estão correlacionados com  $s_i$  e  $w_i$
- Assim, nossas estimativas de  $(\beta_1, \beta_2, \beta_3)$  não serão causais
- Poderíamos tentar obter estimativas causais usando uma variedade de estratégias de identificação:
- Encontre um instrumento válido para  $s_i$  (Angrist e Krueger, 1995, Card 1995)
- Explorar uma descontinuidade em  $s_i$  (Ost, Pan e Webber 2018)
- Randomize  $s_i$  (Attanasio, Meghir e Santiago, 2011)
- etc



# Uma visão estrutural da Equação Minceriana

- - Sabemos que (1) produzirá estimativas tendenciosas, mas *por quê?* Algumas possibilidades:
  - **Viés de habilidade**  $s_i$  e  $w_i$  estão ambos positivamente correlacionados com a capacidade cognitiva não observável
  - **vantagem comparativa** habilidade multidimensional não observável  $\Rightarrow$  auto-seleção na escolaridade
  - **restrições de crédito**  $s_i$  é um investimento caro; algumas pessoas podem não conseguir emprestar o suficiente
  - **heterogeneidade de preferência** (diferentes gostos por  $s_i$ , diferentes taxas de desconto)





# Parte 1 - Desenvolvimento de Modelo Teórico

- Como a escolaridade tem um custo inicial e benefício a longo prazo, precisa de um modelo dinâmico
- Período 1: decidir quanto de escolaridade obter
- Período 2: escolha se quer ou não trabalhar; se estiver trabalhando, receba renda por (1)
- Os indivíduos escolhem o nível de escolaridade para maximizar a utilidade da vida
- Preferências (indica utilidade no período  $t$  por  $u_t$ , com  $s, x$  e  $w$  definidos anteriormente)

$$u_1(z, c, \eta_1) = f(z, c, \eta_1)$$

$$u_2(w(s, x), k, \eta_2) = g(w(s, x), k, \eta_2)$$



# Desenvolvimento do Modelo Teórico

- Com o fator de desconto  $\delta \in [0, 1]$ , a função de utilidade vitalícia descontada é então

$$V = u_1(z, c, \eta_1) + \delta u_2(w(s, x), k, \eta_2) \quad (4)$$

- As equações (1)–(4) definem nosso modelo – mas não necessariamente serão levadas aos dados!
- Este modelo ainda é **ridiculamente irreal**, mas pelo menos temos algo
- Surgem várias questões importantes (mas vamos ignorá-las por hoje)
  - Onde está a capacidade cognitiva? O que exatamente  $c$  representa? Onde estão os empréstimos?
  - Talvez as pessoas devam se preocupar com o *consumo* no período 2, não com a renda
  - O histórico familiar realmente só entra  $u_1$  e não  $\log(w)$ ?
  - $x$  em (1) deve ser uma função de  $s$ ? (Menor  $s \implies$  maior vida útil)
  - Quais são as crenças das pessoas sobre o futuro  $k$  ao decidir  $s$ ?



# Visão Geral do Modelo Teórico

- Como você pode ver, é preciso muito conhecimento para escrever até mesmo um modelo simples
- Requer conhecimento sobre o assunto e sobre matemática/economia de forma mais geral

- **Variáveis Exógenas**

- antecedentes familiares ( $z$ )
- custos de escolaridade ( $c$ )
- crianças na casa ( $k$ )

- **Variáveis Endógenas**

- escolaridade ( $s$ )
- decisão de trabalho do período 2

- **Variável de Resultado**

- renda do trabalho ( $w$ )

- **Não Observáveis**

- renda ( $\varepsilon$ )
- preferências ( $\eta_t$ )

- **Parâmetros do modelo**

- retornos ao capital humano ( $\beta$ )
- fator de desconto ( $\delta$ )
- outros parâmetros implícitos em  $f(\cdot)$  e  $g(\cdot)$



## 2. Questões Práticas de Especificação

- Agora que temos um modelo, precisamos descobrir como levá-lo aos dados
- É aqui que aplicamos o conhecimento sobre **nossos dados** e **stats/econometrics**
- Principais perguntas sobre dados:
  - Podemos observar as variáveis do modelo em nosso conjunto de dados?
  - Em caso afirmativo, eles são medidos de forma confiável?
- Principais perguntas de especificação:
  - Como modelar  $\eta_t$  e  $\varepsilon$ ? (Necessidade de fazer hipóteses distribucionais)
  - Formas funcionais de  $f(\cdot)$  e  $g(\cdot)$
  - $s$  deve ser contínuo (anos de escolaridade) ou discreto (faculdade/não)?



## 2. Questões Práticas de Especificação

- Não vamos entrar em muitos detalhes sobre isso hoje, mas a especificação é importante!
- O que determina a especificação é muitas vezes:
  - - o que é medido de forma confiável nos dados
  - o que é computacionalmente viável para estimar
- Os parâmetros do modelo precisam ser *estimado* ou *calibrado*
- Por exemplo, muitas vezes não temos dados confiáveis que nos permitam estimar  $\delta$ ; devemos calibrá-lo
- A viabilidade computacional geralmente governa como especificamos as diferentes funções
- Por exemplo, **linear nos parâmetros com não observáveis aditivamente separáveis**, como (1).



# Configurando a Especificação

- Parece que podemos estimar alguma forma do nosso modelo
- Suponham que tenhamos antecedentes familiares, custo da faculdade (esta é a variável *efc*)
- Igualmente suponham que temos situação de emprego, salário e número de filhos
- Suponha que temos *s* como binário (variável *collgrad*)
- Supondo  $x = idade - 18$  se não for graduado,  $x = idade - 22$  se graduado, que nem no paper original do Mincer
- Então só precisamos adicionar algumas suposições de forma funcional e estaremos prontos
  - $\varepsilon \sim \text{Normal}$ ,  $\eta_t \sim \text{Logistic}$
  - $u_{i1} = \alpha_0 + \alpha_1 \text{parent\_college} + \alpha_2 \text{efc} + \eta_1$
  - $u_{i2} = \gamma_0 + \gamma_1 \mathbb{E} \log w_i + \gamma_2 \text{numkids} + \eta_2$



# Parâmetros do modelo empírico

- Agora podemos detalhar os parâmetros do modelo empírico
- **parâmetros salariais** ( $\beta, \sigma_\varepsilon$ ), este último é o desv. pad. dos choques de renda
- **parâmetros de escolaridade** ( $\alpha$ )
- **parâmetros de emprego** ( $\gamma, \delta$ )
- Em seguida, escreva uma função objetivo estatística como fn. de dados e parâmetros — por exemplo. maximizar a probabilidade ou minimizar a soma dos resíduos quadrados
- Aprenderemos como fazer isso em aulas posteriores, mas não hoje



## 3. Resolvendo e 4. Entendendo como o modelo funciona

- **Resolvendo o modelo:**

- resolva o problema máximo de utilidade dinâmica para determinados valores de parâmetro
- (não estamos estimando os valores dos parâmetros ainda)
- (no caso do modelo, isso vai levar a um modelo de escolha discreta - que representaria a política ótima do modelo de programação dinâmica).

- **Compreendendo o modelo:**

- Simular dados do modelo
- certifique-se de que os dados simulados são consistentes com as implicações do modelo
- olhar para as estatísticas descritivas dos dados simulados





### 3. Resolvendo e 4. Entendendo como o modelo funciona

- Comece com um modelo o mais simples possível; certifique-se de que as coisas estão funcionando
- Ao introduzir mais complexidades, faça "estática comparativa numérica"
- Certifique-se de que os parâmetros se movem nas direções corretas
- por exemplo.  $\uparrow \beta_1 \implies \uparrow$  escolaridade (ceteris paribus)
- Se não, você provavelmente tem um bug em algum lugar



## 5. Estimativa

- A maioria dos modelos estruturais requer **estimativa não linear**
- Por exemplo, MLE/GMM ou suas contrapartes simuladas
- Na otimização não linear, os valores iniciais são cruciais
- Inicializar em valores iniciais aleatórios provavelmente dará resultados ruins
- Keane recomenda calibrar o modelo manualmente
- Por exemplo, ajustar o intercepto de cada equação aos  $\bar{Y}$ 's dos dados
- Eu recomendo estimar um modelo só com interceptos (ou com muito poucos  $X$ 's)
- Mas este conselho é específico do modelo!



## 6. Validação

- Se você tiver um bom modelo, ele deve ser **válido** (ou seja, prever bem fora da amostra)
- A validação nem sempre é possível, mas é bom fazer se puder
- Por exemplo, se dados experimentais, estimar no grupo de controle, validar no grupo de tratamento
- Por exemplo, veja se o modelo pode replicar grandes mudanças de política nos dados
- Mais simplesmente, você pode descartar metade dos seus dados e tentar prever a outra metade
- Isso normalmente não é feito se a amostra completa não for enorme



## 7. Experimentos de política

- Este é o principal motivo para fazer estimativa estrutural!
- Estimativa estrutural  $\implies$  recuperando o DGP do modelo
- Uma vez que conhecemos o DGP, podemos simular dados dele e fazer experimentos de política
  - requer ter parâmetros invariáveis de política!
- Podemos prever os efeitos de:
  - políticas propostas
  - políticas hipotéticas
- Contraste com RCTs, que apenas revelam os efeitos das políticas implementadas



# No Nosso Curso

- Na maior parte do curso, vou apresentar os modelos na forma “mais próxima dos dados”
- Ou seja, a modelagem prévia já foi feita e alguns detalhes de implementação já estão resolvidos.
- Lógico, nem todos os detalhes relevantes para a sua aplicação estão resolvidos.
- Aí, existem diferentes formas de se lidar – desde ajustes de segunda ordem até começar do zero.
- Referência destes slides: - LINK

