

Experimentos Naturais 'Naturais'

Raphael Corbi

Universidade de São Paulo

Agosto 2017

Experimentos Naturais 'Naturais'

- ▶ a identificação de parâmetros de modelos econômicos é difícil
- ▶ variação de interesse não é ortogonal a fatores não-observáveis
- ▶ experimentos aleatórios são o melhor caminho, mas...
 - ▶ difíceis de implementar
 - ▶ difíceis de generalizar
 - ▶ caros, politicamente incorretos
- ▶ qual o second-best?
 - ▶ quão crível é a exogeneidade? experimentos naturais 'naturais'...
 - ▶ natureza = exogeneidade quase perfeita *wrt* preferência, habilidade...
 - ▶ ex.: gêmeos, data nascimento, sexo, eventos climáticos

Experimentos Naturais

- ▶ **econometria aplicada:** recuperar parâmetro de um modelo
- ▶ *estimação de modelos estruturais*
 - ▶ *estratégias complexas para estimar modelos super-parametrizados*
 - ▶ *hipóteses explícitas, análise de contrafactual*
 - ▶ *hipóteses fortes e simplistas de formas funcionais e distribuições*
- ▶ *experimentos naturais: forma reduzidas*
 - ▶ *mudança ou variação espacial em regras que afetam comportamento*
 - ▶ *interpretação simples e ortogonalidade quase garantida*
 - ▶ *hipóteses implícitas necessárias para interpretação*
 - ▶ *pequenas mudanças no modelo podem alterar a interpretação*

Aula de hoje

- ▶ (i) papers (metodologia e resultados) de experimentos naturais
- ▶ (ii) construir modelos simples específico ao problema
- ▶ (iii) hipóteses (além da exogeneidade) por traz da interpretação
- ▶ (iv) mostrar o que acontece quando relaxamos tais hipóteses
- ▶ (v) discutir quais evidências (usando o próprio experimento) podem dar suporte as hipóteses

IV e Experimentos Naturais

- ▶ $Y_i = \alpha + \beta X_i + \epsilon$
- ▶ **problema:** $Cov(X_i, \epsilon_i) \neq 0$
- ▶ **solução:** usar instrumento Z_i tal qual $Cov(Z_i, \epsilon_i) = 0$

$$\beta_{IV} = \frac{Cov(Y_i, Z_i)}{Cov(X_i, Z_i)} = \frac{Cov(X_i, Z_i)\beta + Cov(\epsilon_i, Z_i)}{Cov(X_i, Z_i)}$$

- ▶ mas o que está contido no termo do erro?

O que está no termo do erro?

- ▶ **fundamental**: descrever do que está contido no erro
- ▶ **especificação**: necessidade de critério específico ou modelo
 - ▶ quais outras variáveis além de X_i ?
 - ▶ qual a relação entre estas variáveis extras e X_i , Z_i e ϵ_i ?
- ▶ respostas podem invalidar aparente exogeneidade wrt ϵ
- ▶ o que é omitido da regressão é crítico para a identificação
- ▶ modelos comportamentais mostram quais hipóteses (tecnologia, preferências) são necessárias para justificar tais restrições de exclusão

Identificação com Experimentos Naturais: 3 casos

- 1) Investimento em capital humano e renda (escolaridade e experiência)
 - ▶ **(i)** Z =nascimento, sexo, loteria, clonagem humana (gemêos monoz.)
 - ▶ **(ii)** caso somente habilidade seja omitida, estimativas válidas para grupos diferentes
 - ▶ **(iii)** caso experiência capture melhor capital humano não-escolar, ambos viesados para lados diferentes
 - ▶ **(iv)** evidência que experiência é endógena e corr. educação
 - ▶ **(v)** controle idade prejudica identificação mesmo se $Cov(Z, A) = 0$
 - ▶ **(vi)** twin studies também podem ter viés negativo

Identificação com Experimentos Naturais: 3 casos

2) Renda Transitória e Permanente em consumo/trabalho

- ▶ **(i)** Z =eventos climáticos para preferências
- ▶ **(ii)** estimativas dependem estrutura mercado de trabalho e $U(\cdot)$

3) Fertilidade e Oferta de Trabalho

- ▶ **(i)** Z =twin first e sex composition
- ▶ **(ii)** estimativas dependem se pais tem pref. sobre sexo dos filhos

Retornos do Capital Humano

- ▶ clássico problema de variável omitida (habilidade)
 - ▶ primeiras soluções: proxies (early test scores, irmãos)
 - ▶ data nascimento (Angrist e Krueger, 1991), sexo (Butcher e Case, 1994), gêmeos monozigóticos, loteria militar (Angrist, 1990))
- ▶ **problemas:**
 - ▶ range grande (0 – 18%)
 - ▶ VVO negativo (4) ou positivo (6)
 - ▶ Z problemático?
 - ▶ problema de heterogeneidade no tratamento?
 - ▶ modelo padrão de escolha educacional - como experimentos naturais são usados, a essência do problema, heterogeneidade do tratamento, hipóteses implícitas

Modelando os Retornos do Capital Humano...

- ▶ $\ln Y_a = f(S, \mu) + g(X_a, \mu)$
 - ▶ onde $\ln Y_a$ rendimento com idade a , μ habilidade, X_a experiência
- ▶ a_e = início da escola, a_k = fim da escola (min. obrigatório)
- ▶ $S_0 = a_k - a_e$ escolaridade na idade a_k
- ▶ estudantes podem aumentar um ano de escolaridade no final
- ▶ $S_1 = S_0$ ou $S_1 = S_0 + 1$
- ▶ indivíduos trabalham sempre após a escola, e nunca durante
- ▶ custo explícito de estudar c e implícito (forgone earnings)

Modelando os Retornos do Capital Humano...

- ▶ indivíduo maximiza valor presente das alternativas:
 - ▶ $V_1(s_1 = 1|S_0) = \exp[f(S_0 + 1, \mu)] \sum_{a=0}^{A-1} \beta^{a+1} \exp[g(a, \mu)] - c$
 - ▶ $V_1(s_1 = 0|S_0) = \exp[f(S_0, \mu)] \sum_{a=0}^{A-1} \beta^a \exp[g(a, \mu)]$
- ▶ regra de decisão: mais escola se $V_1(s_1 = 1|S_0) \geq V_1(s_1 = 0|S_0)$
- ▶ $s_1 = 1$ se $f(S_0 + 1, \mu) - f(S_0, \mu) \geq r + \ln\left[\frac{c}{V_1(s_1=0|S_0)} + 1\right]$
- ▶ caso retorno marginal aumente com habilidade, existe um equilíbrio separador μ^*

Butcher and Case: Child Gender Instrument

- ▶ sexo dos irmãos como instrumento para educação
- ▶ pais decidem investimento em capital humano dos filhos:
- ▶ **(i) retorno da educação diferente para meninos e meninas:**
 - ▶ meninos recebem mais educação (borrowing constraints)
 - ▶ existência de um irmão diminui a educação esperada da irmã
- ▶ **(ii) custo diferente para criar meninos e meninas**
- ▶ **(iii) existência de irmão modifica as preferências da irmã**
- ▶ evidência várias bases de dados: meninas que têm irmã acumulam menos educação (sexo é aleatório)
- ▶ Z =existência de irmã? independente tamanho da família?

Butcher and Case: Child Gender Instrument

- ▶ Problema: existência de irmã depende do tamanho da família
 - ▶ por sua vez, pode estar correlacionado com habilidade
- ▶ Modelos de joint fertility-child investment decision (Becker)
 - ▶ pais ganham utilidade do número e qualidade dos filhos
 - ▶ qualidade vem do investimento e habilidade dos pais
 - ▶ pais com maior habilidade têm menos filhos com mais qualidade
 - ▶ se habilidade hereditária, tamanho e habilidade serão correlacionados

Butcher and Case: Child Gender Instrument

- ▶ podemos incluir tamanho da família na análise econométrica?
- ▶ pais tem preferência sobre a composição de sexo dos filhos?
- ▶ 1) Incluir tamanho no FS somente
- ▶ 2) Incluir tamanho no FS e SS
- ▶ Qual a hipótese necessária para que (1) seja consistente?
- ▶ Qual a hipótese necessária para que (2) seja consistente?
- ▶ o que da pra testar?

Butcher and Case: Child Gender Instrument

- ▶ como testar correlação entre habilidade e ter irmãs?
- ▶ proxy por test scores na infância
- ▶ dados de IQ test para as mães
- ▶ dados de Picture Vocabulary test para crianças
- ▶ podemos testar se sexo é aleatório e se ter irmãs é correlacionado

TABLE 2
 GENDER, SIBLING GENDER, AND "ABILITY" TEST SCORES OF CHILDREN AND THEIR MOTHERS
 (MOTHERS AGED 29+ IN 1992 FROM THE NLSY79)

	Firstborn's PPVT Score		Mother's AFQT Score		
	All mothers	Mothers with firstborn girl	All mothers	Mothers with firstborn girl	All mothers
Firstborn is a girl	-.462 (0.580) ^a	—	-6.44 (8.26)	—	—
Firstborn has a sister	—	-2.83 (1.11)	—	-31.4 (11.8)	—
Children ever born to mother	—	—	—	—	-25.9 (3.89)
Mother is Black	-16.3 (0.953)	-16.7 (1.33)	-175 (9.90)	-171 (14.1)	-172 (9.82)
Mother is Hispanic	-13.2 (1.11)	-13.8 (1.51)	-152 (11.5)	-145 (16.1)	-147 (11.5)

Butcher and Case: Child Gender Instrument

- ▶ mães com ao menos uma filha e mais filhos tem menos IQ
- ▶ isso é repassado para a próxima geração
- ▶ instrumento "irmã" é correlacionado com habilidade via fertilidade
- ▶ reconhecemos o problema (instrumento incorpora info sobre nascimentos subsequentes - fertilidade)
- ▶ porém sabemos que o sexo do primeiro filho é aleatório

- 1) condicional ao tamanho da família (decisão fertilidade é sequencial)
- 2) "sexo do primogênito" como instrumento para educação do segundo

Butcher and Case: Child Gender Instrument

- ▶ voltando ao modelo...
- ▶ custo de educação da segunda filha $c_b > c_g$
- ▶ suponha que caso primogênito homem, nunca é ótimo para as meninas acumular mais educação (logo S_0)
- ▶ primogênito mulher, meninas com mais habilidade escolhem $S_0 + 1$
- ▶ Wald Estimator: $f(S_0 + 1, \mu_H) - f(S_0, \mu_H)$
- ▶ **o efeito estimado vale para qual grupo?**

Comparando com o Angrist & Kruger (1992)...

- ▶ instrumento: trimestre de nascimento
- ▶ começa escola se 6 anos completos até 1 de Setembro
- ▶ não sai antes de completar 16 anos

- ▶ pensando no nosso modelo...
 - ▶ qual o grupo afetado pelo instrumento?
 - ▶ suponha S_0 (S_1) ótimo para μ_L (μ_H)
 - ▶ Wald Estimator: $f(S_0 + 1, \mu_L) - f(S_0, \mu_L)$
 - ▶ os efeitos estimados destes dois papers são comparáveis?

Especificação importa: idade e experiência

- ▶ a estimativa de cada instrumento pode ser diferente, pois afetam sub-populações diferentes
- ▶ caso retorno a educação seja independente de habilidade, as estimativas do modelo seriam iguais (efeitos heterogêneos)
- ▶ mas AK e BC estimam o modelo? $\ln Y_a = f(S, \mu) + g(X_a, \mu)$
- ▶ eles controlam por idade, não por experiência...
- ▶ isso pode levar a viés de sinais opostos em cada um deles
- ▶ mesmo o instrumento sendo natural, especificação importa!

Especificação importa: idade e experiência

- ▶ experiência no trabalho = capital humano não-escolar (proxy)
- ▶ controlar por idade ao invés de experiência é problemático
- ▶ estimador de Wald para BC:
 - ▶ ano extra de escola diminui experiência (sub-estima retorno)
 - ▶ identificado = retorno da educação menos perda de experiência
 - ▶ $[f(S_0 + 1, \mu_H) - f(S_0, \mu_H)] - [g(a - S_0, \mu_H) - g(a - S_0 - 1, \mu_H)]$

Especificação importa: idade e experiência

- ▶ estimador de Wald para AK:
 - ▶ super-estima o retorno da educação
 - ▶ tratamento adianta a entrada na escola (afeta educação) e também experiência depois da escola (não há trabalho antes)
 - ▶ tipo H (unconstrained) = não aumentam educação, mas experiência
 - ▶ tipo L (constrained) = aumentam educação, não experiência
 - ▶ identificado = retorno da educação mais ganho de experiência
 - ▶ $[f(S_0 + 1, \mu_L) - f(S_0, \mu_L)] + \frac{\pi_H}{1 - \pi_H} [g(a - a_k, \mu_H) - g(a - a_k - 1, \mu_H)]$

Retornos a Educação caso Experiência for uma escolha...

- ▶ hipótese até agora: indivíduos trabalham full-time após escola
- ▶ tanto AK quanto BC resolvem o problema do VVO neste caso, mantendo experiência constante X_a
- ▶ mas e esta hipótese é boa? oferta trabalho = escolha?
- ▶ experiência acumulada entre 25 e 34 de idade, por nível escolaridade
- ▶ quais as consequências para os papers?
- ▶ primeiros evidência descritiva

TABLE 3
 MEAN ACCUMULATED HOURS OF WORK
 BETWEEN THE AGES OF 25 AND 34, BY
 SCHOOL ATTAINMENT AND SEX^a
 (WEIGHTED NLSY79)

	Males	Females
HS dropout	15,741 (11,645)	6,793 (9,446)
HS graduate	18,971 (10,993)	11,892 (10,202)
Some college	18,164 (10,797)	13,929 (8,789)
College graduate	19,908 (7,662)	15,958 (7,127)

Retornos a Educação caso Experiência for uma escolha...

- ▶ basta controlar pela experiência ao invés de idade?
- ▶ não, caso participação dependa da habilidade
- ▶ mas podemos testar usando AK (Hausmann modificado)!

Retornos a Educação caso Experiência for uma escolha...

$$\ln y_a = \alpha_1 S + \alpha_2 X_a + \mu + \epsilon_a$$

$$S = \beta_1 a_e(q) + \beta_2 \mu + \epsilon_s$$

$$X_a = \gamma_1 S + \gamma_2 \mu + \epsilon_x$$

- ▶ 1) rodar o FS e salvar \hat{S}
- ▶ 2) rodar a eq. de X_a e incluir $S - \hat{S}$

TABLE 4
TESTS OF THE ENDOGENEITY OF WORK EXPERIENCE USING NATURAL NATURAL INSTRUMENTS
(NLSY79)

	Males			Females		
	AK ^a	BC ^b	AK + BC ^c	AK ^a	BC ^b	AK + BC ^c
P-value of schooling residual coefficient	.008	.031	.277	.666	.733	.097
No. of observations	2,592	1,250	1,245	2,793	1,341	1,336
First-stage R ²	.161	.172	.271	.120	.110	.187

Retornos a Educação caso Experiência for uma escolha...

- ▶ resultados não são inconsistentes com oferta de trabalho correlacionada com habilidade
- ▶ se verdade, AK e BC tem identificação falhas
- ▶ mesmo na mesma amostra, os instrumentos dariam origem a resultados diferentes caso retornos variem com habilidade
- ▶ qual a solução? 2 instrumentos? que mais?

Clonagem Natural

- ▶ caso escolaridade e experiência sejam correlacionados com habilidade, precisamos de 2 instrumentos
 - ▶ mesmo que tenhamos 2 instrumentos, duração dos intervalos pode ser endógena
 - ▶ AK e BC tem somente um, então assumem que experiência seja exógena
- ▶ abordagem alternativa: comparar duas pessoas idênticas geneticamente, com educação diferente
- ▶ análise de gêmeos y é robusta a introdução de escolhas endógenas

Clonagem Natural

- ▶ algumas hipóteses também são necessárias
- ▶ estimação pode ser feita em diferenças (lousa)
- ▶ usamos experiência observada
- ▶ podemos incluir quaisquer repressores endógenos

Clonagem Natural - Problemas

- ▶ 1) erro de medida aumento em diferenças (Ashenfelter e Krueger - pergunta pro irmão) - viés negativo
- ▶ 2) erros individuais u não ortogonais a v (e.g. diff birth weight)
 - ▶ iguais na concepção, mas não no nascimento
 - ▶ experiências diferentes no útero?
- ▶ 3) choques negativos duradouros na saúde - viés positivo na estimativa

- ▶ como gêmeos não diferem na idade e no sexo, não podemos combinar com AK e BC

Renda Permanente e Transitória

- ▶ Modelo de renda permanente:
 - ▶ consumo e poupança dependem da renda esperada futuro em todos os períodos (renda permanente)
 - ▶ choques temporários não importam (consumption smoothing)
 - ▶ contraponto de Friedman ao propensão marginal de consumir keynesiana
- ▶ se consumo sensível ao choque transitório = mercado incompleto
- ▶ problemas:
 - ▶ difícil saber quando um choque é transitório
 - ▶ variação observada na renda pode ser fruto de escolhas (agricultores investem em sementes que afetam o nível e variância da produção)
 - ▶ pode estar correlacionado a preferências por risco, que afeta consumo

Renda Permanente e Transitória

- ▶ Variáveis climáticas como fonte de variação exógena:
 - ▶ afeta a rendimento agrícola
 - ▶ exógenas a ação do agricultor
 - ▶ estacionárias ao longo do tempo da análise
 - ▶ muitos dados para estimarmos os momentos das distribuições
- ▶ mesmo sendo bons instrumentos, as estimativas se baseiam implícita ou explicitamente em hipóteses de um arcabouço teórico:
 - ▶ funcionamento do mercado de trabalho rural
 - ▶ tecnologia agrícola e preferências

Consumo e Renda sem Oferta de Trabalho

▶ **Modelo simples de escolha intertemporal:**

- ▶ produção é função somente do tempo (net of inputs)
- ▶ somente consumo da utilidade (sem lazer) = sem oferta de trabalho
- ▶ renda garantida y_0 por período
- ▶ renda agrícola $y_t = f(\omega_t)$ e $\omega_t = \bar{\omega} + \epsilon_t$

▶ Max $E\{U(c_1) + \beta U(c_2)\}$

▶ s.a. $c_1 + a_2 = y_1 + y_0$

▶ s.a. $c_2 = a_2(1 + r) + y_2 + y_0$

▶ impacto em c_2 de aumento em y_t ou y_0 é igual, mas não em c_1

Consumo e Renda sem Oferta de Trabalho

$$\text{Max}_{a_2} E\{f(\omega_1) + y_0 - a_2\} + \beta E_w U\{(1+r)a_2 + f(\omega_2) + y_0\}$$

$$\text{FOC} : U'_1 = \beta(1+r)E_w U'_2$$

Consumo e Renda sem Oferta de Trabalho

- ▶ efeito verdadeiro da renda permanente y_0 sobre poupança a_2 é igual ao βIV usando $\bar{\omega}$ como instrumento para y_1 ?

$$\frac{\partial a_2}{\partial y_0} = \frac{U_1'' - \beta(1+r)E_w U_2''}{U_1'' + \beta(1+r)^2 E_w U_2''}$$

$$\frac{\partial a_2 / \partial \bar{\omega}}{\partial y_1 / \partial \bar{\omega}} = \frac{U_1'' \partial y_1 / \partial \bar{\omega} - \beta(1+r)E_w U_2'' \partial y_2 / \partial \bar{\omega}}{[U_1'' + \beta(1+r)^2 E_w U_2''] \partial y_1 / \partial \bar{\omega}}$$

Consumo e Renda sem Oferta de Trabalho

- ▶ efeito verdadeiro da renda transitória y_1 sobre a_2 é igual ao β_{IV} usando ϵ_1 como instrumento?

$$\begin{aligned}\frac{\partial a_2}{\partial y_1} &= \frac{\partial a_2 / \partial \epsilon_1}{\partial y_1 / \partial \epsilon_1} \\ &= \frac{U_1''}{U_1'' + \beta(1+r)^2 E_w U_2''}\end{aligned}$$

- ▶ qual efeito é maior sobre a poupança?
- ▶ 1 real de renda transitória ou permanente?

Consumo e Renda sem Oferta de Trabalho

- ▶ Wolpin (1982) usa desvios da média da precipitação em vilas na Índia como instrumentos para renda transitória e encontra efeito zero sobre consumo
- ▶ Paxson (1992) usa surveydata e desvios da média regional de chuva para cada uma das 4 colheitas anuais e encontra efeito transitório na renda e poupança igual a 1

Endogeneizando a Oferta de Trabalho...

- ▶ Paxson e Wolpin usam renda líquida agrícola (sem insumos)
- ▶ caso exista trabalho familiar (não pago), IV será viesado
- ▶ a medida adequada de renda e o viés dependem do que assumimos sobre trabalho familiar

- ▶ aumentamos o modelo de renda agrícola permitindo as famílias alocar trabalho familiar
- ▶ foco no efeito da renda transitória
- ▶ utilidade vem do consumo C_t e lazer L_t
- ▶ dois tipos de trabalho: braçal (contratado ou familiar) e supervisão (familiar)
- ▶ $y_t = f(h_t + m_t, s_t, \omega_t)$
- ▶ trabalho externo paga w fixo e m_t, s_t são escolhidos após ω_t

Endogeneizando a Oferta de Trabalho...

$$\text{Max}_{h_2, m_2, s_2} U[(1+r)a_2 + f(h_2 + m_2, s_2, \bar{\omega} + \epsilon_2) + y_0 - wh_2, 1 - m_2 - s_2]$$

$$h_2 : f_1 = w$$

$$m_2 : \frac{\partial U / \partial h_2}{\partial U / \partial c_2} = \geq f_1$$

$$s_2 : \frac{\partial U / \partial h_2}{\partial U / \partial c_2} = f_2$$

Endogeneizando a Oferta de Trabalho...

- ▶ Caso I: taxa marginal de substituição de lazer e consumo é igual ao salário de mercado, então tanto trabalho familiar quanto de fora será usado ($m_2 > 0$)
- ▶ Caso II: taxa marginal de substituição de lazer e consumo é maior que w , então trabalho familiar é só para supervisão ($m_2 = 0$)

Endogeneizando a Oferta de Trabalho...

- ▶ Caso I: PMg da supervisão e trabalho braçal é igual a w
 - ▶ logo suas demandas são função somente de clima e salário de mercado
 - ▶ $h_2 + m_2 = hm(\omega_2, w)es_2 = s(\omega_2, w)$
 - ▶ mas $m_2 = m(\omega_2, w, y_0)$. Se lazer bem normal?

- ▶ clima tem dois efeitos no trabalho manual:
 - ▶ clima bom aumenta produção e reduz m_2 por efeito renda
 - ▶ clima bom aumenta demanda por supervisão s_2 , o que muda a $\frac{\partial U/\partial l_2}{\partial U/\partial c_2}$
 - ▶ como $MRS = w$, a mudança é 1 para 1 ($\Delta m_2 = -\Delta s_2$)

$$\frac{\partial m_2}{\partial \omega_2} = -\frac{\partial s_2}{\partial \omega_2} + f_3 \frac{\partial m_2}{\partial y_0}$$

Endogeneizando a Oferta de Trabalho...

- ▶ dado isso, um β_{IV} deve usar rendimento líquido de insumos e trabalho familiar (avaliados no salário de mercado)

$$\frac{\partial c_2}{\partial y_2} = \frac{\partial c_2 / \partial \epsilon_2}{\partial (y_2 - wh_2 - wm_2 - ws_2) / \partial \epsilon_2}$$

- ▶ Estimativa do efeito sobre poupança de Paxson e Wolpin serão visados para cima (não levam trabalho familiar em conta)
 - ▶ choque positivo no clima subestima (super) no rendimento (poupança) pois famílias vão consumir mais lazer