Econometria

Parte Extra

Prof. Adalto Acir Althaus Junior oe

Sumário

- Tópicos
- Variáveis instrumentais
- Modelos de escolha discreta
  - ✓ Logit, Probit
- Efeito do tratamento

Variáveis Instrumentais

- Considere o seguinte modelo

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

  - ✓ Onde: $\mathrm{cov}(x_1, u) = \cdots = \mathrm{cov}(x_{k-1}, u) = 0$

    $\mathrm{cov}(x_k, u) \neq 0$

- Se estimarmos este modelo, obteremos uma estimativa viesada ou inconsistente para β_k

Variáveis Instrumentais

- Se usarmos MQO:

regressão: $y = bx + u$

$$\begin{tikzcd} \mathbf{x} \arrow[r] & \mathbf{y} \\ \mathbf{u} \arrow[ur] & \end{tikzcd}$$

Nenhuma associação entre x e u $\rightarrow$ **MQO é consistente**

Variáveis Instrumentais

- Entretanto, MQO falha na seguinte situação:

ENDOGENEIDADE → regressão: $y = bx + u$

$$
\begin{tikzcd}
\mathbf{x} \arrow[r] & \mathbf{y} \\
\mathbf{u} \arrow[u] \arrow[ur] &
\end{tikzcd}
$$

Correlação entre x e u → **MQO NÃO é consistente**

Variáveis Instrumentais

- A solução deste problema por variáveis instrumentais pode ser vista como

Regressão de variáveis instrumentais

regressão: $y = bx + u$ (*)

$$
\begin{tikzcd}
\mathbf{z} \arrow[r] & \mathbf{x} \arrow[r] & \mathbf{y} \\
& \mathbf{u} \arrow[u] \arrow[ur] &
\end{tikzcd}
$$

z **não correlacionado** com u, e sim, correlacionado com x.

O método se baseia na utilização de uma variável adicional z, não incluída em (*), que satisfaça tais condições:

(1) $Cov(z,u) = 0$

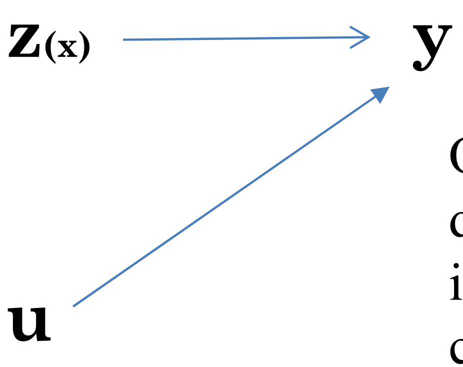
(2) $Cov(z,x) \neq 0$.

Variáveis Instrumentais

- A solução fica:

Regressão de variáveis instrumentais

regressão: $y = bz + u$ (*)



O método se baseia na utilização de uma variável adicional z, não incluída em (*), que satisfaça tais condições:

(1) $Cov(z,u) = 0$

(2) $Cov(z,x) \neq 0$.

z **não correlacionado** com u, e sim, correlacionado com x.

Variáveis Instrumentais

- A variável adicional *z* é chamada de instrumento válido para *x*, quando a variável *z* satisfaz ambas as condições.

- Em geral, temos muitas variáveis em *u*, e se mais de uma destas variáveis estiver correlacionada com *x*, neste caso, necessitamos no mínimo tantas variáveis em *z*, quantas forem as variáveis em *u* correlacionadas com *x*.

Variáveis Instrumentais

- IVs devem satisfazer duas condições
  - ✓ condição de relevância
  - ✓ Condição de exclusão
- **Condição de Relevância**
- O seguinte deve ser verdade ...
- No modelo abaixo, z satisfaz a condição de relevância se $\gamma \neq 0$

$$x_k = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1} + \gamma z + v$$

- Isso significa que z é relevante para explicar o regressor problemático, depois de "parcializar" o efeito de todos os outros regressores no modelo original

Variáveis Instrumentais

- **Condição de Exclusão**
- O seguinte deve ser verdade ...
- No modelo original, onde

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

z satisfaz a condição de exclusão se cov(z, u) = 0

- O que isso significa em palavras?
  - ✓ Resposta: z não é correlacionado com a perturbação, u ... isto é, não tem poder explicativo em relação a y após o condicionamento do outro x's
  - ✓ Podemos testar essa condição?

Variáveis Instrumentais

- Você encontrou uma boa IV, e agora?
  - ✓ Pode-se pensar na estimativa IV como sendo feita em duas etapas

- Primeiro estágio: regredir x_k em outros $x's$ & z

$$x_k = \partial_0 + \partial_1 x_1 + \partial_2 x_2 + \cdots + \partial_{k-1} x_{k-1} + \partial_k z + u$$

- Segundo estágio: tomar x_k previsto ($\widehat{x_k}$) do primeiro estágio e usá-lo no modelo original em vez de x_k. É por isso que também chamamos de estimativas IV mínimos quadrados de dois estágios (2SLS)

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{k-1} x_{k-1} + \beta_k \widehat{x_k} + u$$

Variáveis Instrumentais

- Os valores previstos (calculados) representam a variação em x_k que é "boa", pois é orientada apenas por fatores não correlacionados com u
- Especificamente, o valor previsto é função linear de variáveis não correlacionadas com u
- Por exemplo no Stata, use os comandos ivreg ou xtivreg (para dados em painel)

Escolha Binária: Estimação

- As 3 opções mais comuns são:
- Dist. Uniforme – dá origem ao modelo de probabilidade linear
- Dist. Logística – dá origem ao modelo logit
- Dist. Normal Padrão – dá origem ao modelo probit

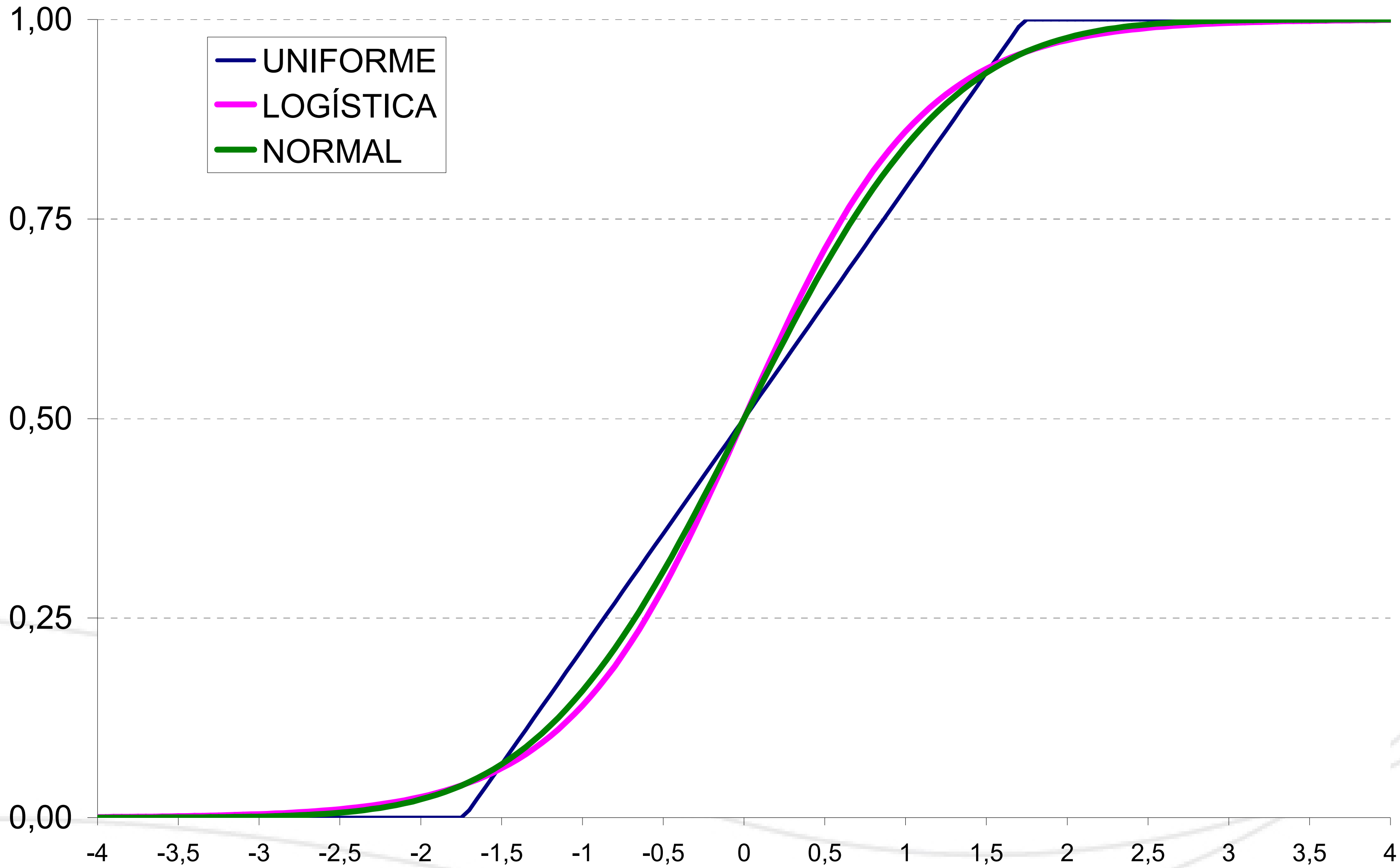
Os procedimentos descritos acima pressupõem uma certa distribuição de probabilidade para

$$\tilde{\varepsilon}_i$$

Escolha Binária: Estimação

- As FDA dessas distribuições são:
  - Uniforme: $F(a) = a$
  - Logística: $F(a) = \dfrac{e^a}{1+e^a}$
  - Normal padrão: $F(a) = \int_{-\infty}^{a} (2\pi)^{-1/2} e^{(-t^2/2)} dt$

Escolha Binária: Estimação



Escolha Binária: Estimação

- Observações:
- O modelo de probabilidade linear é assim chamado por ser estimado por uma regressão linear:

$$y = F(\mathbf{x}'\boldsymbol{\beta}) + u$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_n x_n + u$$

- Pela sua simplicidade, tal modelo foi muito usado no passado
- Com o desenvolvimento dos processos computacionais, tornou-se possível trabalhar com modelos mais complexos, como logit e probit

Aplicação

- Exemplo Indústrias Declinantes
- Variável de escolha:

$$y = \begin{cases} 1 & \text{fechou} \\ 0 & \text{não fechou} \end{cases}$$

  - <u>Variáveis explicativas</u>:
    - SCALE = capacidade produtiva da planta (% da maior planta)
    - SHARE = capacidade produtiva da firma (% da indústria)
    - MULTPLANT = dummy (=1 se firma multi-planta)
    - SPECIALIST = dummy (=1 se firma não diversificada)
    - MAJORSITE = dummy (=1 se planta multi-produto)
    - CUT = utilização da capacidade total de produção da indústria

Aplicação

- Exemplo Indústrias Declinantes
- Variável de escolha:

$$y = \begin{cases} 1 & \text{fechou} \\ 0 & \text{não fechou} \end{cases}$$

$$y = \beta_0 + \beta_1 SCALE + \beta_2 SHARE + \beta_3 MULTPLANT + \beta_4 SPECIALIST + \beta_5 MAJORSITE + \beta_6 CUT + u$$

  - reg -> probabilidade linear
  - logit -> logit
  - Probit -> probit

Aplicação

Exemplo Indústrias Declinantes

TABLE 8 Logit Analysis of Plant Closures[a]

Independent Variable	Mean Value	8.1	8.2	8.3
Constant	1.0	−.136‡	−.131‡	−.093‡
		(.017)	(.019)	(.031)
SCALE	.46	−.183‡	−.170‡	−.150‡
		(.034)	(.037)	(.039)
SHARE	.24	.138‡		
		(.061)		
SHARE* [multiplant firm]	.35[b]		.148‡	.140†
			(.062)	(.068)
SHARE* [single-plant firm]	.16[c]		.055	.060
			(.120)	(.153)
MULTPLANT	.45	−.017	−.030	−.035
		(.019)	(.025)	(.026)
SPECIALIST	.10	−.055†	−.054†	−.043
		(.032)	(.032)	(.033)
MAJORSITE	.60	.021	.019	.026
		(.016)	(.016)	(.016)
CU	.54			−.055
				(.040)
Number of observations		1646	1646	1480
Number of closures		124	124	104
Probability of closure		7.5%	7.5%	7.0%
Log likelihood		−417.8	−417.4	−353.4

Aplicação

- Exemplo Indústrias Declinantes
- Principais Conclusões:
- A probabilidade de fechamento de uma planta de produção:
  - diminui com o tamanho da planta
  - aumenta com o tamanho da firma
- Resultados sugerem complementaridade entre teorias de "shakeout" e "stakeout"
  - Shake out = firmas pequenas seriam expulsas de indústrias declinantes mais facilmente devido às vantagens de custo das grandes (economias de escala)
  - Stake out = firmas pequenas conseguiriam sobreviver mais tempo em indústrias declinantes pois podem manter-se lucrativas com menor produção

Treatment Effects

- Efeitos do tratamento
- Seja d igual a um indicador de tratamento do experimento que estudaremos
  - ✓ d = 0 não tratado pela experiência (ou seja, grupo controle)
  - ✓ d = 1 tratado por experiência (isto é, grupo tratado)
  - ✓ Seja y o resultado potencial de interesse
  - ✓ $y = y$ (0) para grupo não tratado
  - ✓ $y = y$ (1) para o grupo tratado
- Fácil mostrar que $y = y(0) + d\,[y(1) - y(0)]$

Treatment Effects

- Ex. # 1 - O tratamento pode ser que o programa mais médicos
  - ✓ d = 1 para municípios que receberam ações do programa
  - ✓ y pode ser um número de coisas, por exemplo nr. de atendimentos
- Ex. # 2 - O tratamento é que as empresas emitem debentures
  - ✓ d = 1 se as empresas emitem
  - ✓ y poderia ser uma série de coisas, como ROA, Lucratividade

Treatment Effects

- Agora podemos definir algumas coisas
- Efeito Médio de Tratamento (ATE) é dado por

$$\mathbf{E}[\boldsymbol{y}(\mathbf{1}) - \boldsymbol{y}(\mathbf{0})]$$

- O que isso significa em palavras?
- Resposta: É mudança esperada em y quando o indivíduo recebe o tratamento pelo programa ou experimento; este é o efeito causal que normalmente estamos interessados em descobrir!

Treatment Effects

$$\mathbf{E}[\boldsymbol{y}(\mathbf{1}) - \boldsymbol{y}(\mathbf{0})]$$

- Por que não podemos observar diretamente a ATE?
- Resposta = Só observamos um resultado...
  - ✓ Se os elementos são do grupo dos tratados, observamos y (1); se não for tratado, observamos y (0). Nós nunca observamos os dois.
  - ✓ Não conseguimos observar um determinado elemento do grupo dos tratados caso não tivesse recebido o tratamento
  - ✓ isto é nós não podemos observar o contrafactual do y (1), que seria seu resultado na ausência do tratamento, pois o indivíduo efetivamente recebeu o tratamento e não há como observar diretamente o que aconteceria caso não tivesse recebido

Treatment Effects

- Efeito médio do tratamento sobre os tratados (ATT) é dado por

$$\boldsymbol{E}\,[\boldsymbol{y}\,(\mathbf{1}) - \boldsymbol{y}\,(\mathbf{0}) \mid \boldsymbol{d} = \mathbf{1}]$$

- Este é o efeito do tratamento naqueles que são tratados; ou seja, a alteração esperada em y tratado se a amostra da população de observações que são tratadas for aleatória
- O que não observamos aqui?
  - ✓ Resposta = $y\,(0) \mid d = 1$

Treatment Effects

- O efeito médio do tratamento sobre os não tratados (ATU) é dado por $\boldsymbol{E}\,[\boldsymbol{y}\,(\mathbf{1}) - \boldsymbol{y}\,(\mathbf{0}) \mid \boldsymbol{d} = \mathbf{0}]$

- É o que o efeito do tratamento teria sobre aqueles que não são tratados pelo experimento
- Não observamos $y\,(1) \mid d = 0$

Treatment Effects

- Como estimamos ATE, $E\ [y\ (1) - y\ (0)]$?

  - ✓ Resposta = Nós dependemos de $\boldsymbol{E\ [y\ (1)\ |\ d = 1] - E\ [(y\ (0)\ |\ d = 0]}$ como uma forma de inferir o ATE
  - ✓ Se interpretarmos isto como ATE, estamos assumindo que, na ausência do tratamento, o grupo tratado teria, em média, o mesmo resultado que o grupo não tratado. Mas será verdade?

Treatment Effects

$$\underbrace{\{E[y(1)|d = 1] - E[y(0)|d = 1]\}}_{\textbf{ATT}} + \underbrace{\{E[y(0)|d = 1] - E[y(0)|d = 0]\}}_{\textbf{Viés de seleção}}$$

Basta somar e diminuir o mesmo termo

- A comparação pelas simples diferença de médias não nos fornece o ATE!!

- Qual é o "viés de seleção" em palavras?
  - ✓ Resposta: Se o tratamento não for atribuído aleatoriamente, há um viés!

Treatment Effects

- O viés de seleção: $E[y(0)|d = 1] - E[\,y(0)|d = 0]$

- Definição = Qual seria a diferença em média para o grupos das observações tratadas e das não tratadas sem qualquer tratamento
  - ✓ Não observamos este contrafactual!
  - ✓ Agora vamos ver porque a aleatoriedade é a chave

Treatment Effects

- Um tratamento aleatório, d, implica que d é independente dos resultados potenciais; isto é

$$E\,[y\,(0)\mid d=1]=E\,[(y\,(0)\mid d=0]=E\,(y\,(0))$$

o valor esperado de y é o mesmo para os tratados e não tratados na ausência de qualquer tratamento

E

$$E\,[y\,(1)\mid d=1]=E\,[(y\,(1)\mid d=0]=E\,(y\,(1))$$

- Com isso, é fácil ver que o viés de seleção fica igual a zero (= 0)
- E, o ATT restante é igual a ATE!

Treatment Effects

- Expressando isso em formato de um modelo de regressão:

$$y = \beta_0 + \beta_1 d + u$$

Esta regressão só dará uma estimativa consistente de β1 Se $\mathbf{cov}\ (\boldsymbol{d}, \boldsymbol{u}) = \mathbf{0}$; isto é, o tratamento, $\boldsymbol{d}$, é aleatório e, portanto, não correlacionado com $\boldsymbol{y}\ (\mathbf{0})$

- O formato de regressão também nos permite inserir facilmente controles adicionais, X

$$y = \beta_0 + \beta_1 d + \beta_{s'} X + u$$

Treatment Effects

- Algumas técnicas e métodos para se obter uma aproximação mais precisa do ATT - ATE
  - ✓ Experimentos – Aleatorização
  - ✓ Experimentos não aleatórios
  - ✓ Matching e uso de propensity scores
  - ✓ Diff-in-diff (difference in diferences)
  - ✓ Rdd (regression discontinuity design)
  - ✓ Técnicas não paramétricas
  - ✓ Outras técnicas e desenhos de experimentos
- Modernamente, os estudos envolvem a preocupação com métodos mais robustos e bem desenhados tanto na amostragem e coleta de dados com nas estimações, visando possibilitar inferências causais