Técnicas Econométricas para Avaliação de Impacto

Introdução aos Métodos Quase-Experimentais

Rafael Perez Ribas
Centro Internacional de Pobreza

Brasília, 23 de abril de 2008

Introdução

- Breve descrição de métodos para estimar o **efeito de um tratamento**, quando este:
    - É derivado de uma **auto-escolha**. Exemplo:
        - Trabalhador que escolhe o setor de atividade;
    - É derivado da **escolha de quem está provendo** o tratamento. Exemplos:
        - Bolsas de estudos destinados aos melhores alunos;
        - Investimentos na infra-estrutura de áreas pobres.
    - Ou existem **fatores omitidos** por trás da relação. Exemplo:
        - Retornos da escolaridade sobre o salário;
        - Discriminação no mercado de trabalho.

- A escolha pela metodologia não-experimental depende de **três fatores**:
    - O tipo de informação disponível;
    - O modelo causal;
    - O parâmetro de interesse.
- É necessário ainda saber sobre quais condições o parâmetro de interesse é identificado.
- A distinção entre resultados de uma mesma avaliação pode decorrer da imposição de **falsas suposições sobre a distribuição** das variáveis.
    - Vantagem dos modelos não-paramétricos (Heckman, 1990).

Modelo Causal

- Um modelo de efeito do tratamento pode ser descrito como um *switching model regression*:

$$D_i^*(s) = Z_i\gamma + v_i, \qquad E[v] = 0$$

$$D_i^*(s) \geq 0 \rightarrow D_i(s) = 1$$

$$D_i^*(s) < 0 \rightarrow D_i(s) = 0$$

$$s \in S$$

$$Y_i(s) = X_i(s)\beta(s) + u_i(s), \qquad E[u(s)] = 0$$

- Para facilitar, suponha que existem só dois estados (com e sem tratamento),

$$s \in \{0,1\}.$$

- Assim:

$$D^* = Z\gamma + v, \qquad E[v] = 0$$

$$D^* \geq 0 \rightarrow D = 1$$

$$D^* < 0 \rightarrow D = 0$$

$$Y_1 = X_1\beta_1 + u_1, \qquad E[u_1] = 0$$

$$Y_0 = X_0\beta_0 + u_0, \qquad E[u_0] = 0$$

- Além disso, suponha que exista um conjunto de **características observáveis comum** a todas as equações:

$$D^* = (X_c, Z)\gamma + v, \qquad E[v] = 0$$

$$Y_1 = (X_c, X_1)\beta_1 + u_1, \qquad E[u_1] = 0$$

$$Y_0 = (X_c, X_0)\beta_0 + u_0, \qquad E[u_0] = 0$$

- Relembrando, o problema da avaliação de impacto é não observamos a mesma unidade em ambos os estados. **Só observamos**:

  $$Y = Y_1 D + Y_0 (1 - D),$$

  ou ainda:

  $$Y = \left[(X_c, X_1)\beta_1\right] D + \left[(X_c, X_0)\beta_0\right](1 - D) + u_1 D + u_0 (1 - D)$$

- Para facilitar, assuma que ***X* não varia com *D***:

  $(X_c, X_1)\beta_1 = \alpha_1 + X\beta$ e $(X_c, X_0) = \alpha_0 + X\beta$

- Assim:

  $$Y = \alpha_0 + D(\alpha_1 - \alpha_0) + X\beta + u_1 D + u_0 (1 - D)$$

  $$Y = \alpha_0 + D\tau + X\beta + (u_1 - u_0) D + u_0, \quad u = (u_1 - u_0) D + u_0$$

- **Problema:** apesar de

$$E[u] = E[u_1]D + E[u_0](1 - D) = 0 \text{ ,}$$

$$u \not\perp D$$

- Ou seja, ***D* é endógeno**.
- Neste caso, uma **diferença de médias** ou uma simples **regressão** gera o seguinte resultado **enviesado**:

$$E[Y_1 \mid D = 1] - E[Y_0 \mid D = 0] = \quad (\alpha_1 - \alpha_0)$$
$$+ \left[E(u_1 \mid D = 1) - E(u_0 \mid D = 0)\right]$$

os termos entre colchetes **não necessariamente se anulam**

- Mais especificamente, D é endógeno porque

$$v \not\perp u \text{ ,}$$

$$B = \left[E\left(u_1 \mid v \geq -(X_c, Z)\gamma\right) - E\left(u_0 \mid v < -(X_c, Z)\gamma\right)\right]$$

- Fatores não-observados, *v*, que determinaram a seleção influenciam o resultado, *Y*.
- **Solução:** fazer com que

  $v \perp u$
- **Como?**
  - **Incorporando mais variáveis**, o suficiente para que os termos não-observados sejam ortogonais.
- **Três formas** de fazer isso:
  - Mais variáveis em X_c para eliminar o viés de variáveis omitidas (seleção sobre observáveis);
  - Incluir uma variável $Z \perp u$ na equação de seleção (seleção sobre não-observáveis);
  - Incluir uma variável $Z \not\perp u$ na equação de seleção (seleção sobre observáveis).

- **Importante:** Mesmo quando *D* é exógeno, se há heterogeneidade no impacto,

  $$\tau_i = \bar{\tau} + \xi_i \,,$$

  os coeficientes de uma regressão podem não representar os parâmetros de interesse:

  $$E[\tau \mid D=1] = \bar{\tau} + E[\xi \mid D=1] \neq E[\tau \mid D=0] = \bar{\tau} + E[\xi \mid D=0]$$

- **Exemplos:**
  - Diferenças salariais entre homens e mulheres;
  - Ou entre negros e brancos;
  - Ou entre os anos de 2006 e 2007;
  - Não existe uma escolha por trás destas variáveis, mas é impossível tratá-las como aleatórias.
- Por maior que seja o controle sobre observáveis, **a composição dos grupos é distinta**.

Parâmetros de Interesse

- O parâmetro de interesse mais comum é o ATE:

  $$Y_1 = \alpha_1 + X\beta + u_1 \rightarrow E[Y_1] = \alpha_1 + E[X]\beta$$

  $$Y_0 = \alpha_0 + X\beta + u_0 \rightarrow E[Y_0] = \alpha_0 + E[X]\beta$$

  $$\tau = E[Y_1 - Y_0] = (\alpha_1 - \alpha_0)$$

  que representa o impacto médio do tratamento na **população como um todo** (independente de quem foram os tratados).

- Outro parâmetro de interesse é o ATT ou ATET:

  $$E[Y_1 \mid D=1] = \alpha_1 + E[X \mid D=1]\beta + E[u_1 \mid D=1]$$

  $$E[Y_0 \mid D=1] = \alpha_0 + E[X \mid D=1]\beta + E[u_0 \mid D=1]$$

  $$\tau_T = E[Y_1 - Y_0 \mid D=1] = (\alpha_1 - \alpha_0) + E[u_1 - u_0 \mid D=1]$$

  o ATT mostra como o **grupo que recebeu** (ou escolheu) o tratamento foi afetado.

- De maneira equivalente, o ATU é definido como:

$$E[Y_1 \mid D=0] = \alpha_1 + E[X \mid D=0]\beta + E[u_1 \mid D=0]$$

$$E[Y_0 \mid D=0] = \alpha_0 + E[X \mid D=0]\beta + E[u_0 \mid D=0]$$

$$\tau_U = E[Y_1 - Y_0 \mid D=0] = (\alpha_1 - \alpha_0) + E[u_1 - u_0 \mid D=0]$$

  isso representa o quanto o **grupo que não está sendo tratado** seria afetado caso fosse tratado.

  - Não confundir com externalidades do programa

- O ATU é de interesse, por exemplo:
  - Estimação do impacto da expansão de um programa para além do grupo já tratado.

- Um caso onde ATU é geralmente estimado, porém ele não é de interesse:
  - Decomposições à la Oaxaca-Blinder.

- Por definição:

$$ATE = \Pr(D=1)\cdot ATT + \Pr(D=0)\cdot ATU$$

$$\tau = \tau + \Pr(D=1)E(u_1 - u_0 \mid D=1) + \Pr(D=0)E(u_1 - u_0 \mid D=0)$$

Se $u_1, u_0 \perp D$

ou

$E[u_1 - u_0 \mid D=1] = 0$,

ATT=ATU=ATE

  - Idéia implícita de homogeneidade do impacto.

- **Importante:** para que o ATU e, por conseqüência, o ATE sejam de interesse, é relevante que a amostra de não-tratados represente uma **população de interesse**.

- O estimador pode ainda ser representativo de um efeito local (LATE):

$$LATE = \frac{E[Y \mid Z'] - E[Y \mid Z]}{E[D(Z') = 1] - E[D(Z) = 1]}, \text{ onde } Z' = Z + \varepsilon$$

- Se o impacto do tratamento é **homogêneo**, LATE=ATE.
- **Contudo**, é muito provável que a escolha de *Z* influencie o estimador.
- Correção para **erros de medida** pode ser um exemplo de quando o LATE é o parâmetro de interesse:
    - Renda de uma transferência predita pela composição domiciliar;
    - Renda permanente predita por bens duráveis.

- Um exemplo de LATE
    - Efeito de Tratamento para Pessoas na Margem da Indiferença (ou Eligibilidade), EOTM:

$$EOTM(s,s') = E\left(Y(s) - Y(s') \middle| \begin{array}{l} R(Y(s)) = R(Y(s')) \\ \left. \begin{array}{l} R(Y(s)) \\ R(Y(s')) \end{array} \right\} \geq R(Y(l)) \\ l \neq s, s' \end{array} \right)$$

Modelo de Seleção sobre Observáveis

- Assume que os fatores não-observados, *u* e *v*, não estão correlacionados.
- **Suposição de independência na média condicional**:
  $E[Y_0 \mid D=1, X_c, Z] = E[Y_0 \mid D=0, X_c, Z] = E[Y_0 \mid X_c, Z]$,
  - Se o parâmetro de interesse é de **efeito médio**.

- Se existem outros parâmetros de interesse (mediana, desvio-padrão, etc.):
  - Devemos assumir uma condição mais restritiva (versão condicional da SUTVA):

    $Y_0 \perp D \mid X_c, Z$.

- **Modelo mais comum** de seleção sobre observáveis:
  - **Regressão multivariada**.
  - Se u não é correlacionado com v, é porque X_c explica a correlação de D com Y_0.
    - Exemplo típico: seleção amostral.

- **Problema** é quando X_c explica o efeito de D sobre Y. Exemplos:
  - Resultados de testes de proficiência explicam escolaridade, mas também explicam a diferença entre salários das pessoas mais e menos escolaridade;
  - Renda domiciliar determina participação em programa social que determina gastos com alimentação, que também é determinada pela renda domiciliar.

- **Solução**, incorporar interações para melhorar o controle (Rubin, 1977):

$$Y = \alpha + D\tau + X_c\beta + \left(X_c \cdot D\right)\omega + u$$

- Contudo, se impacto é heterogêneo,

$$\hat{\tau} = E\left[Y_1 - Y_0 \mid X_c = 0\right]$$

- **Solução**:

$$Y = \alpha + D\tau + X\beta + \left(X - \overline{X}\right)D\omega + u$$

ou

$$Y = \alpha + D\tau_T + X\beta + \left(X - \overline{X}_1\right)D\omega_T + u$$

- **Problema**: *v* não ser ortogonal a *u* e a condição de **independência ser violada**, mesmo incluindo o controle de X_c.
  - Ou então, ter que incluir uma série de outras variáveis junto com as respectivas interações (**perda dos graus de liberdade**).
- **Solução:** incluir uma variável $Z \not\perp u$.
  - Tal que esta variável explique em grande parte a correlação de *v* com *u*.
  - Ou seja, retira de *v* a parte que está correlacionada com *u*.
- **Problema**: não há como incluir *Z* diretamente na equação de interesse.
  - Pois é uma variável (fracamente) endógena e correlacionada com *D*.

- **Soluções:**
  - Já conhecidas há muito tempo, porém não para corrigir o viés de seleção, mas para estimar um parâmetro de interesse (ATT ou ATE) sobre uma variável exógena.

  - ***Matching*** (método não-paramétrico),

$$\tau = E\left[Y_{1i} - \sum_{j \in \{D=0\}} W\left(X_{c,i}, X_{c,j}, Z_i, Z_j \right) Y_{0j} \right];$$

- o ***Inverse Probability Weighting*** (método semi-paramétrico),

$$W\left(X_c, Z\right) = \frac{D}{\Pr\left(D = 1 \mid X_c, Z\right)} + \frac{1 - D}{\Pr\left(D = 0 \mid X_c, Z\right)}$$

ou

$$W\left(X_c, Z\right) = \frac{D \cdot \Pr\left(D = 1 \mid X_c\right)}{\Pr\left(D = 1 \mid X_c, Z\right)} + \frac{\left(1 - D\right) \cdot \Pr\left(D = 0 \mid X_c\right)}{\Pr\left(D = 0 \mid X_c, Z\right)}$$

- Estas técnicas relaxam a imposição de formas funcionais.
- E possibilitam a estimação de parâmetros de interesse para sub-amostras. Exemplo:
  - o No caso de diferenças salariais, um estimador de *Matching* ou *IPW* pode fornecer resultados distintos de uma microssimulação contrafactual à la Oaxaca-Blinder (DiNardo; Firpo e Ñopo possuem trabalhos sobre isso).

- **Importante:**
    - Se, mesmo usando *Matching* ou *IPW*, não for incluído no modelo $Z \not\perp u$, os resultados tendem a não se distinguir da regressão de Rubin (1977).
        - A não ser devido à mudança na formal funcional.
    - Portanto, se na regressão já havia uma problema de **viés de seleção**, *matching* ou *IPW* não corrigirão este problema utilizando as mesmas variáveis X_c.
        - Estimador de *matching* **não é uma panacéia!**

Regressão Descontínua

- Um método de seleção sobre observáveis por definição.
- Ou seja, assume a mesma condição de independência.
- Porém, só gera um estimador LATE.
- Existe uma variável $Z \not\perp u$ que determina o tratamento, tal que

$$Z > z \rightarrow D = 1$$

$$Z \leq z \rightarrow D = 0$$

ou ainda

$E\left[D \mid Z = z + \varepsilon\right] > E\left[D \mid Z = z - \varepsilon\right]$, mas

$$E\left[Y_0 \mid Z = z + \varepsilon\right] = E\left[Y_0 \mid Z = z - \varepsilon\right]$$

- **Exemplos:**
  - Bolsas de estudos distribuídas de acordo com testes de proficiência;
  - Transferência de renda para pessoas abaixo da linha de pobreza.
- **Dois tipos de estimador**:
  - *Sharp*

$$E[Y \mid Z = z + \varepsilon] - E[Y \mid Z = z - \varepsilon],$$

$$\text{se } E[D \mid Z > z] = 1 \text{ e } E[D \mid Z \leq z] = 0$$

  - *Fuzzy*

$$\frac{E[Y \mid Z = z + \varepsilon] - E[Y \mid Z = z - \varepsilon]}{E[D \mid Z = z + \varepsilon] - E[D \mid Z = z - \varepsilon]}$$

- Formas de estimação:
  - Regressão paramétrica (*naïve*);
  - Regressão não-paramétrica (consistente);

Modelo de Seleção sobre Não-Observáveis

- **Condição assumida**:

  $Z \perp u$ e *Z* determina *D*.

- **Idéia**: retirar de *u* (ou *D*) o componente que está correlacionado com *v*.

- **Como fazer isso**:

  - **Método IV**, consiste em utilizar uma variável $\hat{D} \perp u$;
  - **Seleção de Heckman**, identificar em *u* o componente correlacionado com *v*;
  - **Modelo de efeitos fixos**, supondo que *D* é determinado somente por atributos fixos (muitas vezes inviável);
  - **Modelo de Diferença-nas-Diferenças**; suposição idêntica a de modelos de efeitos fixos.

Método de Variáveis Instrumentais IV

- Duas formas de implementação:
  - Substituir o indicador de tratamento, D, na equação de interesse por $g(Z)$,
    - que pode ser o próprio Z.
  - Ou utilizar o valor predito

    $$\hat{D}\left(X_c, Z\right)$$

    na equação de interesse no lugar de D.

- Dessa forma, utilizamos uma ***proxy* de *D*** que é ortogonal a *u*, pois $\left(X_c, Z\right) \perp u$.
- **Problema**: essa *proxy* varia com a escolha de *Z*.
  - Ou seja, estamos estimando um **parâmetro LATE**:

$$E\left[Y_1 - Y_0 \mid X_c, Z\right] = \left(\alpha_1 - \alpha_0\right) + E\left[u_1 - u_0 \mid Z = z\right]$$

Estimador de Seleção de Heckman

- Muito comum em análise com seleção amostral.
- Consiste em estimar a esperança condicionada de *v* (não-observado em modelos não-lineares), para posteriormente incluí-la na equação de interesse:

  $$Y = \alpha_0 + D\tau + X_c\beta + \rho E\left[v \mid X_c, Z\right] + u$$

  - Utilização da famosa razão inversa de Mills.
- Assim, $D \perp u$.

- Como no método IV, é possível estimar os coeficientes:
    - Em dois estágios;
    - Ou por meio de modelos multivariados (*joint likelihood*).
- E como no método IV, há o problema de encontrar uma variável ***Z* convincente**.
    - De fato, IV da Ms. Speedy (alocação da verba) não é convincente.
    - Seria mais interessante utilizá-la em um modelo de seleção sobre observáveis.
- **Muito importante:** Se $Z \not\perp u$, os estimadores resultantes podem ser ainda mais enviesados.

Estimador de Diferença-nas-Diferenças

- Mede a variação mais que proporcional no grupo de tratamento em relação ao grupo de controle.
- Para entendê-lo, vamos definir a seguinte equação:

$$Y_{D,t}(i) = \delta_t + \alpha_{D,t} + \eta_i + u_{i,t}$$

A primeira diferença (antes-e-depois) para cada grupo (tratamento e controle) é:

$$\begin{aligned} Y_{1,1}(i) - Y_{1,0}(i) &= (\delta_1 - \delta_0) + (\alpha_{1,1} - \alpha_{1,0}) + (\eta_i - \eta_i) + (u_{i,1} - u_{i,0}) \\ &= \delta \cdot t + \alpha_1 \cdot t + (u_{i,1} - u_{i,0}) \end{aligned}$$

$$\begin{aligned} Y_{0,1}(i) - Y_{0,0}(i) &= (\delta_1 - \delta_0) + (\alpha_{0,1} - \alpha_{0,0}) + (\eta_i - \eta_i) + (u_{i,1} - u_{i,0}) \\ &= \delta \cdot t + \alpha_0 \cdot t + (u_{i,1} - u_{i,0}) \end{aligned}$$

- Assim quando estimamos **a diferença das esperanças das diferenças**:

$$E\left[Y_{1,1}(i) - Y_{1,0}(i) \mid D = 1\right] - E\left[Y_{0,1}(i) - Y_{0,0}(i) \mid D = 0\right]_{t=1}$$

$$= (\alpha_1 - \alpha_0) + E\left(u_{i,1} - u_{i,0} \mid D = 1\right) - E\left(u_{i,1} - u_{i,0} \mid D = 0\right)$$

- **Condição:** se

$$E\left(u_{i,1} - u_{i,0} \mid D = 1\right) = E\left(u_{i,1} - u_{i,0} \mid D = 0\right),$$

  o estimador é consistente.

- Isso implica que:
  - Toda diferença entre os grupos foi eliminada na **subtração dos efeitos fixos**;
  - A condição agora é que as **variações sejam ortogonais** à seleção para o tratamento.
    - Esta condição sempre esteve implícita nos outros estimadores.

- A condição

  $$E\left(u_{i,1} - u_{i,0} \mid D = 1\right) = E\left(u_{i,1} - u_{i,0} \mid D = 0\right)$$

  pode ser **facilmente violada** dependendo do que está sendo investigado. Exemplo:
    - Efeito de cotas em universidades sobre o desempenho de estudantes no segundo grau;
        - Grupo de tratamento: alunos com direito à cota;
        - Grupo de comparação: alunos sem direito à cota.
    - Condição é violada quando a trajetória média do desempenho de alunos é distinta.
- Por esta mesma razão, modelos de efeitos fixos podem não resolver o problema.

- **Solução**, assumir uma outra condição:

$$E\left(u_{i,1} - u_{i,0} \mid X, D = 1\right) = E\left(u_{i,1} - u_{i,0} \mid X, D = 0\right)$$

  - Ou seja, assumir **independência nas trajetórias condicionadas**, por exemplo, às características dos alunos.
- Para fazer isso, **não basta incluir variáveis** de controle em uma regressão de diferença-nas-diferenças.
- **Soluções**:
  - Incluir interações, tais como em Rubin (1977);
  - Combinar método DD com outros métodos de seleção sobre observáveis (*Matching*, *IPW*, *RDD*);
  - Combinar método DD com outros métodos de seleção sobre não-observáveis (*IV* ou Seleção de Heckman);
  - Estimar um modelo de diferenças com efeito fixo (inviável).

- Para os casos de *Matching* e *RDD*,
  a solução é mais simples do que parece:
  - Basta estimar o parâmetro **antes e depois** da implementação do tratamento e eliminar as diferenças identificadas previamente das diferenças posteriores.
- **Importante é não confundir**
  o papel do estimador de Diferença-nas-Diferenças
  com o papel dos estimadores quase-experimentais:
  - O estimador de DD garante uma forma mais consistente de se estimar um impacto, podendo ou não ser combinado com técnicas experimentais ou quase-experimentais de avaliação.

Coleta de Dados

- Métodos quase-experimentais exigem **amostras maiores** que métodos experimentais.
    - Pois somente parte da variação nas variáveis de interesse é utilizada na estimação.
    - Geralmente, o grupo de tratados tende a ser mais homogêneo, do ponto de vista da investigação, que o grupo de controle;
    - Portanto, a regra 50%-50% utilizada na amostra de experimentos não é válida nos quase-experimentos.
- Base de dados secundárias:
    - Pesquisas para cobrir uma série de outros propósitos;
    - Informações de relevância para avaliação podem estar ausentes;
        - Exemplo: Avaliação de impacto do PBF utilizando a PNAD.

- Por isso, muitas pesquisas coletam seus próprios dados:
    - Problema é que estas pesquisas podem não ter representatividade.
- Registro administrativo:
    - Contém uma grande amostra de tratados e de controle;
    - É possível observar de fato quem foi, ou está sendo, tratado e quem não foi, além de muitas das condições que determinaram esta escolha (baseline);
    - Problema é que estas informações só são representativas da população registrada.

- A busca pela representatividade não é o mais importante;
    - E pode incorrer em aumento desnecessário dos custos;
    - O foco é, primeiro, como estimar o ATT da maneira mais consistente;
    - Validade interna vem antes da externa, apesar das duas serem importantes.
- **Muito importante:**
    - Em qualquer avaliação quase-experimental (X_c, Z) deve ser completamente exógeno a D.
        - A única forma de garantir isso é que (X_c, Z) seja coletado antes do tratamento (informação na linha de base);
        - Se (X_c, Z) e D são observados simultaneamente, existe uma grande risco de inconsistência interna na avaliação.