Técnicas econométricas para avaliação de impacto

O uso de algoritmos de emparelhamento - *matching*

Bruno César Araújo

Instituto de Pesquisa Econômica Aplicada – IPEA

Brasília, 7 de maio de 2008

Pergunta da aula de hoje

Como melhorar a imputação contrafactual e, por conseqüência, a estimativa dos efeitos de um tratamento, a partir do uso de técnicas de emparelhamento?

Problema básico de inferência causal

- Seja $D_{it}(0,1)$ um indicador se um indivíduo *i* recebeu um tratamento ou não no período *t*.
- Seja $Y^1_{i,t+s}$ (D=1) o valor da variável supostamente impactada pelo tratamento para os indivíduos que receberam o tratamento em t+s, com $s \geq 0$.
- Seja $Y^0_{i,t+s}$ (D=1) o valor desta mesma variável caso este indivíduo que recebeu o tratamento não o tivesse recebido.
- Assim, o impacto do tratamento sobre os tratados é dado por:

$$ATT = E(Y^1_{i,t+s}\,(D=1)) - E(Y^0_{i,t+s}\,(D=1))$$

Problema básico de inferência causal

- Às vezes, estamos também interessados em saber o impacto do tratamento sobre os não-tratados (a fim de estimar possíveis efeitos da expansão de um programa):

$$ATU = E(Y^1_{i,t+s}\,(D=0)) - E(Y^0_{i,t+s}\,(D=0))$$

- Ou podemos estar interessados no efeito médio do tratamento, em sua forma incondicional:

$$ATE = E(Y^1_{i,t+s}) - E(Y^0_{i,t+s})$$

O problema é que não se observa $Y^0_{i,t+s}$ e $Y^1_{i,t+s}$ para um mesmo indivíduo.

Problema básico de inferência causal

- Suponha 2 indivíduos, e ambos podem receber o tratamento. Temos, rigorosamente, 8 possibilidades para a variável de interesse:

D	D_1, D_2 $(0,0)$	D_1, D_2 $(0,1)$	D_1, D_2 $(1,0)$	D_1, D_2 $(1,1)$
$Y_1(.,.)$	$Y_1(0,0)$	$Y_1(0,1)$	$Y_1(1,0)$	$Y_1(1,1)$
$Y_2(.,.)$	$Y_2(0,0)$	$Y_2(0,1)$	$Y_2(1,0)$	$Y_2(1,1)$

A hipótese SUTVA

- A hipótese SUTVA (Stable Unit Treatment Value Assumption) desconsidera efeitos de equilíbrio geral. Assim, ao invés de oito valores possíveis para *Y*, chegamos a apenas quatro.

D	D_1, D_2 $(0,0)$	D_1, D_2 $(0,1)$	D_1, D_2 $(1,0)$	D_1, D_2 $(1,1)$
$Y_1(.,.)$	$Y_1(0)$	$Y_1(0)$	$Y_1(1)$	$Y_1(1)$
$Y_2(.,.)$	$Y_2(0)$	$Y_2(1)$	$Y_2(0)$	$Y_2(1)$

Isto é razoável?

Benchmark: experimento aleatório

- Já vimos que, quando os experimentos são aleatórios, uma boa imputação do contrafactual do grupo de casos é o próprio valor observado da variável de interesse no grupo de controle:

$$ATE = E(y_{t=1}) - E(y_{t=0})$$

O qual pode ser calculado a partir da regressão:

$$Y_i = \alpha + \beta d_i + \varepsilon_i$$

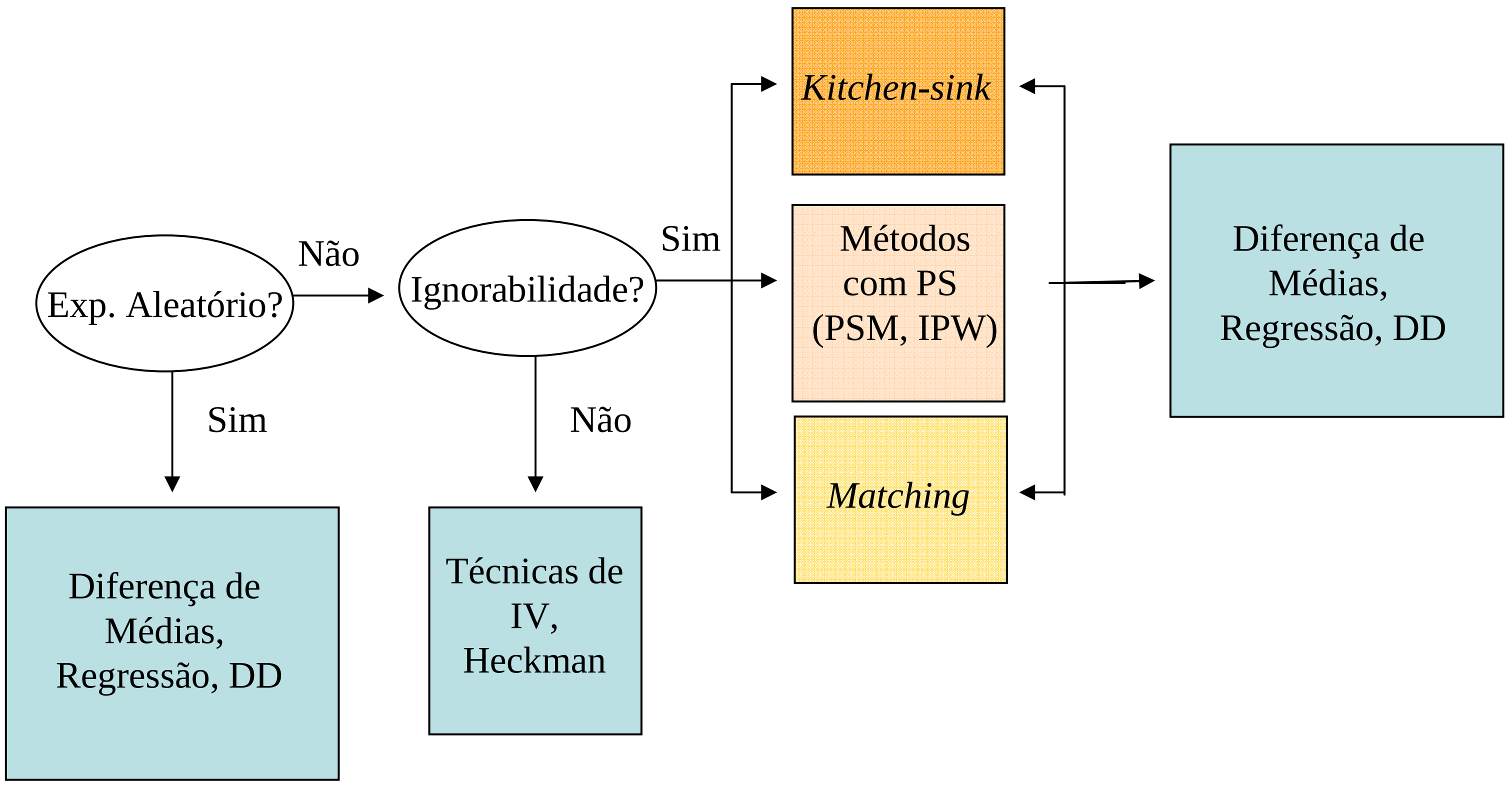
$d_i = 1$ quando o indivíduo recebe o tratamento

Assim, $\beta = \Delta y = E(y_{t=1}) - E(y_{t=0})$

E quando o experimento não for aleatório?

- Em verdade, se a participação no tratamento depende, por exemplo, da expectativa com respeito a seus benefícios, a seleção deixa de ser aleatória.
- Nestes casos, é razoável supor que a simples diferença das médias não será uma boa estimativa do ATE.
- Mesmo a estimativa DD não será razoável, uma vez que ela apenas lida com efeitos fixos no tempo.
- O que fazer?
  - Caso se saiba as variáveis determinantes da auto-seleção, pode-se utilizar técnicas relacionadas a seleção em observáveis.
  - Caso não sejam sabidas estes determinantes, utiliza-se técnicas de seleção em não-observáveis.

Esquematicamente...



Seleção em observáveis

- Quando as fontes da auto-seleção são conhecidas na forma de uma matriz de características (X), podemos usar a hipótese de "ignorabilidade" do tratamento.
- *Hipótese 1* (Abadie e Imbens, 2002)*:* Condicional às características X, D e (y_1, y_0) são independentes:

$$E(y_1|X,D)= E(y_1|X) \text{ e } E(y_0|X,D)=E(y_0|X)$$

$$ATT(X)=E(y_1 - y_0|X,D=1) \text{ e } ATE(X) = E(y_1 - y_0|X)$$

Seleção em observáveis

- Outra hipótese importante se refere à identificação: é preciso haver indivíduos que não recebam o tratamento mesmo que tenham características muito semelhantes aos que o recebem:
- *Hipótese 2:* $c < Pr(D=1|X=x) < 1-c$, para algum $c > 0$.

Como realizar a seleção em observáveis?

- Existem basicamente quatro formas de se realizar ATE sob seleção em observáveis:
  - Modelos de regressão com variáveis ***X*** e interações (aulas 3 e 4);
  - Uso de técnicas de emparelhamento (*matching*) para a construção de grupos de casos e controles (aula de hoje);
  - *Matching* baseado em modelo probabilístico (PSM – próxima aula);
  - Modelos de regressão que utilizam o IPW.

Uma palavra de atenção

- Ao estimar ATE e suas variações a partir de seleção em observáveis, é preciso estar atento a:
  - A seleção em observáveis e a aleatoriedade condicional do tratamento (hipótese 1) são **hipóteses** que não estão sujeitas a teste;
  - O conjunto de observáveis deve ser suficiente, i.e, deve conter todas as variáveis determinantes da participação no tratamento (sob pena do viés de variáveis omitidas);
  - Deve haver empresas com $\boldsymbol{X=x}$ que recebam e outras que não recebam o tratamento (caso contrário, não é possível imputar o contrafactual- hipótese 2).

Modelos de regressão com variáveis *X* e interações

- A inclusão de controles pelas características individuais pode seguir a estratégia de Rubin (1977):

$$Y_{it} = \alpha + \alpha_1 d_t + \alpha_2 d_j + \beta d_{jt} + \boldsymbol{X\gamma} + (\boldsymbol{X}\text{-}\boldsymbol{E}(\boldsymbol{X}))d_{jt}\boldsymbol{\omega} + \varepsilon_{ijt}$$

Métodos de *Matching*

- Como vimos, não podemos estimar diretamente ATE a partir da diferença de médias, pois o tratamento não é aleatório.
- Contudo, supondo que a seleção se baseia em características observáveis, e que existem indivíduos de mesmos determinantes da seleção (ou suficientemente próximos) tais que um recebe o tratamento e o outro não, é razoável supor que estes indivíduos formem um par (ou conjunto) caso-controle.
- Assim, o que é feito na prática é selecionar, a partir da amostra original, um conjunto de casos e controles tais que eles tenham determinantes de seleção muito semelhantes, ao ponto de supormos que, dentro deste grupo, a seleção é aleatória. Posteriormente, aplicam-se as técnicas vistas nas aulas anteriores (teste de médias, regressão, DD, com BS etc).

Métodos de *Matching*

- Partem do pressuposto da imputação: para quem se observa y^1 supõe-se um y^0 de quem não recebeu o tratamento, mas cujas características lhe sejam semelhantes
- Em outras palavras, para y_i em que $D = 1$ encontram-se y_j (e vice-versa) tais que:

$$||X_i - X_j|| < ||X_i - X_k||, \; q.q. \; k \mid D=0$$

- Então, atribui-se $y_i(0) = y_j(0) = y_j$, e procede-se analogamente para $D = 0$

$$ATE = (1/\mathrm{N})\ \Sigma_N[D_i y_i + (1-D_i)\ y_i(1)] - (1/\mathrm{N})\ \Sigma_N[\ (1-D_i) y_i + D_i y_i(0)]$$

Técnicas de *Matching*: um balanço

- A principal vantagem do *matching* (alguns o chamam de *hard matching*) é que ele é intuitivo.

Técnicas de *Matching*: um balanço

- Contudo, suas limitações e questões em aberto são:
  1. Como toda técnica que usa seleção em observáveis, depende criticamente desta hipótese.
  2. Como proceder em casos de “empate”?
  3. Deve-se fazer *matching* com ou sem reposição?
  4. Que noção de distância ||.|| deve-se utilizar?

Técnicas de *Matching*: um balanço

5. $||X_i - X_j||$ em geral não é zero, e pode surgir um viés assintótico. Além do mais, este viés aumenta com a dimensão de *X* (ponto para o uso de PSM!)
6. Problemas com as variâncias: Não é correto dizer que *Var(ATE)* é calculada por procedimentos tradicionais. O problema não desaparece com *bootstrap* de número finito de pares, nem assintoticamente.

Felizmente...

- Abadie et al. (2004) implementaram no Stata uma rotina que permite corrigir boa parte do viés no ATE a partir da incorporação de informação de um modelo probabilístico de participação no tratamento.
- Basicamente, estima-se: $u_D = E(Y(D)|X=x)$ para os grupos de casos ($D=1$) e controles ($D=0$) por meio de um OLS e incorpora-se esta informação na imputação contrafactual (para a fórmula exata, veja artigo de Abadie et al., 2004 e Abadie e Imbens, 2002).
- Analogamente, o algoritmo dos autores estima consistentemente a variância do ATE, permitindo inclusive a opção homoscedasticidade/heteroscedasticidade entre os grupos.