

Técnicas econométricas para avaliação de impacto

O uso de métodos de regressão e introdução aos métodos de diferenças- das-diferenças

Bruno César Araújo

Instituto de Pesquisa Econômica Aplicada – IPEA

Brasília, 30 de abril de 2008

Pergunta da aula de hoje

Como os métodos de regressão se relacionam com as estratégias apresentadas nas aulas anteriores?

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 é dado por:

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

Problema básico de inferência causal

O problema é que não observamos $Y_{i,t+s}^0$ ($D=1$).

Assim, o problema de avaliação de impacto pode ser visto como um problema de *missing data* que necessita de imputação. Como fazê-la?

Tipologia das técnicas de avaliação de impacto

- 5 grandes categorias, que podem ser combinadas (Blundell e Dias, 2000):
 - Análise contrafactual a partir de experimentos aleatorizados (aula 2):
 - Problemas: são muito raros e caros; validade externa é difícil por exigirem inexistência de efeitos de equilíbrio geral.

Tipologia das técnicas de avaliação de impacto

- Experimentos naturais (muito relacionados a DD - aulas 3 e 4):
 - Duas hipóteses centrais: efeitos temporais comuns ao grupo de tratamento e de controle e ausência de mudanças de composição dentro do grupo (efeito Hendry ou Hawthorne não podem acontecer) – difícil encontrar um grupo de comparação sem o auxílio de outras técnicas.

Tipologia das técnicas de avaliação de impacto

- Análise baseada em emparelhamento – *matching* (aulas 4 e 5):
 - Se um tratamento não é aleatório, mas sabem-se os determinantes da participação, nada mais natural que encontrar um grupo de comparação composto por indivíduos com estes mesmos determinantes do grupo dos tratados. O problema é garantir que a seleção é feita apenas sobre observáveis.

Tipologia das técnicas de avaliação de impacto

- Seleção sobre não-observáveis (aulas 4 e 8):
 - Algumas possibilidades mais conhecidas são o método de IV, modelo de seleção de Heckman, modelo de efeitos fixos e diferenças-das-diferenças.
- Modelos de simulação estrutural
 - São mais utilizados quando há avaliação de políticas que envolvem mudanças de regras, levam em consideração modelos comportamentais de escolha dos indivíduos.

O desenho da pesquisa

- Blundell e Dias (2000) argumentam que a escolha do método de avaliação deve levar em consideração:
 - A natureza do programa;
 - A natureza da questão a ser respondida: se desejamos fazer uma avaliação geral, o impacto do programa nos tratados (ATT) ou a extrapolação dos impactos em caso de extensão do programa (ATU);
 - A natureza dos dados.

O desenho da pesquisa

- Sem embargo, Meyer (1994) avalia que um bom desenho de avaliação deve encontrar:
 - Variações exógenas nas variáveis explicativas;
 - Grupos de comparação que sejam comparáveis;
 - As implicações das hipóteses sob teste.

Desenhos de pesquisa e modelos de regressão

1. Abordagem de um grupo antes e depois do tratamento (à la Granger):

- Esta abordagem se refere ao acompanhamento de um mesmo grupo de indivíduos antes e depois de um determinado tratamento.

$$Y_{it} = \alpha + \beta d_t + \varepsilon_{it}$$

$$d_t = 1 \text{ quando } t = 1, \text{ e } 0 \text{ c.c.}$$

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

A hipótese básica (H0) do desenho de pesquisa é que, caso não houvesse tratamento, $\beta = 0$ e não haveria variação no Y_{it}

Desenhos de pesquisa e modelos de regressão

- Podemos também calcular diretamente $\Delta y = E(y_{t=1}) - E(y_{t=0})$ e fazemos inferência estatística, por exemplo, por bootstrapping.
- Ao invés de focarmos na dimensão temporal (i.e., ao invés de mantermos o grupo fixo e o acompanharmos no tempo) podemos considerar que t é o D , isto é, uma variável binária indicadora se o indivíduo recebeu o tratamento ou não. Neste caso, a análise é em *cross-section* e *ex-post* ao tratamento (imputação contrafactual).
- Podemos estar interessados não apenas na média dos grupos, mas na distribuição (testes Kolmogorov-Smirnoff).
- Uma vantagem desta abordagem é que ela não necessariamente precisa de painel, ela pode ter *cross-sections* consecutivos (desde que a amostragem tenha identificação de quem recebeu o não o tratamento).

Desenhos de pesquisa e modelos de regressão

- **Esta abordagem é confiável?**
 - Em geral, esta abordagem não costuma fornecer inferências válidas.
- Razões:
 - Tendência;
 - Efeitos fixos;
 - Introdução de um terceiro fator contemporâneo ao tratamento.

Desenhos de pesquisa e modelos de regressão

2. Abordagem caso-controle antes e depois do tratamento (diferenças-das-diferenças):

- Muitas vezes, existe informação sobre um grupo de indivíduos que não recebeu o tratamento que sofre a mesma influência ambiental do grupo dos tratados. Este grupo geralmente é chamado de grupo de controle.
- Nestes casos, é possível estimar o que se chama comumente de diferenças-das-diferenças. A abordagem recebe este nome porque estamos interessados em um estimador do tipo:

$$\Delta y_{(j=1)} - \Delta y_{(j=0)} = (E(y_{1(t=1)}) - E(y_{1(t=0)})) - (E(y_{0(t=1)}) - E(y_{0(t=0)}))$$

Desenhos de pesquisa e modelos de regressão

- **Por que este estimador é tão popular?**
 - Porque ele expurga do efeito atribuído ao tratamento efeitos temporais que são comuns a ambos os grupos e efeitos fixos aos grupos de tratamento e de controle (i.e, diferenças intrínsecas entre os grupo de tratamento e de controle, que podem ser até mesmo não-observáveis, desde que sejam invariáveis no tempo – veja aula 3).

Desenhos de pesquisa e modelos de regressão

- **Como estimar diferenças-das-diferenças?**
- Podemos tentar estimar $\Delta y_{(j=1)} - \Delta y_{(j=0)}$ diretamente ou por meio de procedimentos de *bootstrapping*. Contudo, podemos também buscar estimá-lo com um modelo de regressão:

$$Y_{it} = \alpha + \alpha_1 d_t + \alpha_2 d_j + \beta d_{jt} + \varepsilon_{ijt}$$

$$d_t = 1 \text{ se } t=1, \text{ e } 0 \text{ c.c.};$$

$$d_j = 1 \text{ se } j = 1, \text{ e } 0 \text{ c.c.};$$

$$d_{jt} = 1 \text{ se } t=1 \text{ e } j = 1, \text{ e } 0 \text{ c.c.}$$

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

d_t é um efeito temporal comum, d_j é um efeito fixo do grupo dos casos, e d_{jt} é o efeito a ser estimado

Desenhos de pesquisa e modelos de regressão

- **Limitações da abordagem diferenças-das-diferenças**
 - Mesmo considerando que a abordagem elimina diferenças fixas (e mesmo não observáveis) entre os grupos e efeitos temporais comuns, a modelagem DD perde a validade se ainda existir interação entre $t=1$ e $j=1$ (endogeneidade, as variações não são mais ortogonais ao tratamento).
 - Como comentado na aula 3, esta condição pode ser facilmente violada (ex. efeito das cotas nos alunos de 2º grau). A solução para este problema pode ser considerar trajetórias condicionadas.

Desenhos de pesquisa e modelos de regressão

- **Limitações da abordagem diferenças-das-diferenças**
 - Outro problema ocorre quando as distribuições das variáveis Y são muito diferentes entre os grupos de casos e controles. Isto faz com que, ao fazermos transformações como logs nestas variáveis, os impactos marginais β sejam condicionais aos níveis de Y , dificultando a comparabilidade.
 - É interessante prestar atenção aos parâmetros α_1 e α_2 , se eles forem muito grandes, pode indicar presença de efeitos temporais muito fortes ou variáveis omitidas.
 - Bertrand *et al.* (2002) argumentam que a inferência estatística sobre β pode estar contaminada pela correlação serial da variável de tratamento, sobretudo em painéis longos.

Desenhos de pesquisa e modelos de regressão

3. Extensões à abordagem de diferenças-das-diferenças:

- **A inclusão de controles pelas características individuais:** a hipótese por trás deste método é a independência nas trajetórias condicionadas.
- Como já visto na aula 3, podemos fazer isto via algumas formas:
 - Incluir interações com características individuais
 - Combinar métodos DD com outros métodos de seleção sobre observáveis (*Matching*, IPW, RDD).
 - Combinar métodos DD com outros métodos de seleção sobre não-observáveis (IV, Heckman).

Desenhos de pesquisa e modelos de regressão

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

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

Desenhos de pesquisa e modelos de regressão

3. Extensões à abordagem de diferenças-das-diferenças:

- **Interações de ordem superior:** às vezes, um pesquisador não está apenas interessado no impacto do tratamento sobre o grupo dos tratados como um todo, ele pode estar interessado também no impacto do tratamento sobre um determinado subgrupo dos tratados (ex.: política de emprego para desempregados negros).
- Até agora, os modelos focalizaram a interação entre duas dummies, isto é, uma dummy indicadora se o indivíduo estava no grupo de tratamento e se ele estava no período posterior ao tratamento. Agora, temos que considerar três dummies e suas interações.

Desenhos de pesquisa e modelos de regressão

- Estas interações podem ser representadas pelo seguinte modelo econométrico:

$$Y_{it} = \alpha + \alpha_1 d_t + \alpha_2 d_j + \gamma_1 e_k + \alpha_3 d_{jt} + \gamma_2 e_{kt} + \eta d_{jk} + \beta d_{jkt} + \varepsilon_{ijt}$$

d_t , d_j , e d_{jt} são definidos como anteriormente,

$$e_k = 1 \text{ quando } k = 1, 0 \text{ c.c.}$$

$$e_{kt} = 1 \text{ quando } k = 1 \text{ e } t = 1, 0 \text{ c.c.}$$

$$d_{jk} = 1 \text{ quando } k = 1 \text{ e } j = 1, 0 \text{ c.c.}$$

$$d_{jkt} = 1 \text{ quando } k = 1, j = 1 \text{ e } t = 1, 0 \text{ c.c.}$$