Avaliação de Políticas Públicas II: Métodos não experimentais

Paula Pereda (USP)

October 30, 2020

Aula 3

voltar

Resumo da Aula 2

1. Métodos para Seleção nas Observáveis
   1.1 Análise de Regressão:
      - CIA é a hipótese de média condicional zero (elimina viés de seleção)
      - Vantagens: efeito médio, simples, heterogeneidades
      - Desvantagens: forma funcional, falta de suporte comum

   1.2 Modelo de Seleção de Heckman*:
      - Primeiro a estimar equação de seleção
      - instrumentaliza a seleção e considera correlação entre eqs

   1.3 Análise usando Propensity Score
      - Eliminar desbalanceamento dos grupos
      - reduz dimensionalidade das covariadas
      - seleção nas observáveis $\equiv$ Seleção no propensity
2. Testes de Balanceamento

Teste de Balanceamento

- Teste Bivariado
- Teste de Distribuição (Kolmogorov-Smirnov)
- **Teste Ranksum (Wilcoxon-Mann-Whitney)**

Lousa

Teste de Balanceamento e Análise de PS

- Teste de Balanceamento com amostra inicial
  - Se rej Ho, há desbalanceamento na variável, potencial inclusão no propensity score
- Realizar a Análise usando PS
- Repetir o teste de balanceamento após o uso do PS
  - Se não rej. Ho, PS resolveu a diferença entre os grupos
  - Se rej. Ho, reestruturar a análise até eliminar desbalanceamento (veremos os procedimentos de “bias correction" também)

Estimando Propensity Scores

- Especificação do PS é crucial para identificar ATT/ATE
- Heckman 97 e 98: Relevância teórica das covariadas
- Para a estimação dos PS, podemos usar modelos de probabilidade:
  - Probit
  - Logit
  - Log da Odds Ratio
- Nos modelos de probabilidade: w_i é a dummy de tratamento; X_i o vetor de covariadas; e β o vetor de parâmetros.
- Estimação por Máxima Verossimilhança

Estimando Propensity Scores por Odds Ratio

Como $E(w_i) = P(w_i/X_i = x_i) = \frac{e^{x_i\beta}}{(1+e^{x_i\beta})}$, podemos linearizar o modelo:

$$ln\left(\frac{p}{1-p}\right) = x_i\beta + \varepsilon_i$$

t.q. $p = P(w_i = 1)$.

Estima-se por MLE (Newton-Raphson Method) ou OLS

Estratégias para Especificação Correta do PS

- Melhor P(x): Bom balanceamento entre os 2 grupos nas covariadas (X_1, X_0).
- Usar muitos X's: piora o problema de suporte comum (McCaffrey et al 2004)
- Estratégias para escolha de X e forma funcional: data-driven vs estruturais
  - data-driven: stepwise com cutoff para t (Rosenbaum e Rubin, 1984; Rosenbaum e Rubin, 1985; Dehejia Wahba, 1999, Hirano e Imbens, 2001).
  - estruturais: Heckman (literatura econômica + objetivos da política)

Após estimação de PS

1. Parear observações usando e(x): PS Matching
2. Estratificar amostra usando e(x): PS Estratificado
3. Analisar Y usando pesos em função de e(x): PS Weighting
4. Usar métodos combinados

PS Matching

- Encontrar contrafactuais para tratados: Observações com probabilidades similares
- Método mais comum: Matching local (greedy)
- Várias métricas de pareamento

PS Matching: Tipos de Pareamento

1) Métrica de Mahalanobis

- Ordena aleatoriamente participantes
- Calcula distância Mahalanobis entre as covariadas do 1o tratado e demais controles (j): $d_{ij} = (X_1 - X_0)' V_{nxn}^{-1}(X_1 - X_0)$, em que X_1 e X_0 são as covariadas dos grupos de trat e controle, respectivamente, e V é a matriz de var-cov amostral das variáveis para toda a amostra
- Escolhe par de i com base em j de menor distância
- Retira par escolhido da amostra
- Repete o procedimento até que todos i's estejam pareados
- Matching 1-1

Ilustração: Lousa

PS Matching: Tipos de Pareamento

2) Métrica de Mahalanobis usando e(x)

- Mesmo procedimento usando e(x) no lugar de X

3) Pareamento por vizinho mais próximo (NN)

- Sejam e_i e e_j as medidas de propensity score de tratados e controles, respectivamente, e I_1 e I_o os conjuntos dos tratados e controles
- Uma **vizinhança** $c(e_i)$ contém um participante i tratado ($i \in I_1$) e um não tratado j ($j \in I_0$) como par de i se a diferença absoluta dos e's é a menor entre todos os possíveis pares de i e j:

$$c(e_i) = Min_j|e_i - e_j|$$

- Uma vez encontrado o par, as observações são retiradas da amostra e repete-se o processo

PS Matching: Tipos de Pareamento

4) Caliper Matching: Pareamento com threshold

- NN pode gerar vizinhos distantes
- Podemos estipular uma tolerância (ϵ) para a vizinhança
- j é vizinho de i se $|e_i - e_j| < \epsilon$
- RR85 sugerem $\epsilon < 0,25\hat{\sigma}_e$, em que $\hat{\sigma}_e$ é o erro-padrão amostral de e

5) Vizinho mais próximo com threshold

- 3) + 4)

PS Matching: Tipos de Pareamento

6) Vizinho mais próximo com threshold usando métrica de Mahalanobis

- Ordena aleatoriamente participantes
- Seleciona o primeiro tratado
- Encontra j tratados (dentro do threshold)
- Calcula distância Mahalanobis entre as e(x) dos participantes
- Escolhe par de i com base em j de menor distância
- Remove-se o par e repete-se o procedimento até que todos i's estejam pareados
- Matching 1-1

Segundo RR85, 6) produz o melhor balanceamento entre as covariadas

PS Matching

Observações

- Todas as métricas e exemplos que vimos são de **matching local** (decisões ótimas localmente): Veremos exemplo de matching global
- Robustez: Testar diferentes thresholds e diferentes X's
- Pode ainda haver problema de suporte comum: TRIMMING

Procedimento de Trimming

- Selecionar subamostra com maior sobreposição entre grupos
- Validade externa X Validade interna
- Ideia: quando $e(x) \to 1$ ou $e(x) \to 0$, não há contrafactual
- Desafio: encontrar cutoff que determine o que é $e(x) \to 1$ ou $e(x) \to 0$

Procedimento de Trimming

Caso 1: X binária (mais simples) Sejam

$$X_i = \text{f, se i é mulher}$$
$$\text{m, se i é homem,}$$

- População: $N = N_f + N_m$
- q é o share da população em que $X_i = m$: $q = E(N_m/N)$
- Efeitos médios: $ATE|_{X_i=x} = \tau(x)$ e, na população, $ATE = (1-q)\tau(f) + q\tau(m)$.
- Números de controle e tratados com $X_i = x$: $N_{cx} = \sum_{i:X_i=x}(1-w_i)$ e $N_{tx} = \sum_{i:X_i=x}(w_i)$
- Então o PS é: $e(x) = N_{tx}/N_x$, ou $e(f) = N_{tf}/N_f$ e $e(m) = N_{tm}/N_m$

Procedimento de Trimming
Caso 1: X binária (mais simples), cont.

- A média amostral das duas subpopulações é

$$\bar{Y}_{cx} = 1/N_{cx} \sum_{i:X_i=x} Y_i(1 - w_i)$$
$$\bar{Y}_{tx} = 1/N_{tx} \sum_{i:X_i=x} Y_i(w_i)$$

- Sendo $V(Y_i(w)|X_i = x) = \sigma^2 \ \forall x, w$

$$\hat{\tau}(f) = (\bar{Y}_{tf} - \bar{Y}_{cf}) \text{ e } NV(\hat{\tau}(f)) \rightarrow \left(\frac{\sigma^2}{1-q} \frac{1}{ef(1-ef)} \right)$$
$$\hat{\tau}(m) = (\bar{Y}_{tm} - \bar{Y}_{cm}) \text{ e } NV(\hat{\tau}(m)) \rightarrow \left(\frac{\sigma^2}{q} \frac{1}{em(1-em)} \right)$$

Procedimento de Trimming

Caso 1: X binária (mais simples), cont.

- Portanto,

$$\hat{\tau} = (N_f/N)\hat{\tau}(f) + (N_m/N)\hat{\tau}(m) \text{ e}$$
$$NV(\hat{\tau}) \to \sigma^2 \left(\frac{q}{em(1-em)} \frac{1-q}{ef(1-ef)} \right)$$

- Quando $e(f) \to 1$ ou $e(f) \to 0$, estima-se $\hat{\tau}(f)$ com baixa precisão
  - Solução para estimação mais eficiente: Dropar $x_i = f$
- Quando $e(f) = 1$ ou $e(f) = 0$, $NV(\hat{\tau})$ não existe

Procedimento de Trimming

Caso 1: X binária (mais simples), cont.

- Casos particulares:

$$(i)\ \left(\frac{em(1-em)}{ef(1-ef)} \le \frac{1-q}{1-2q}\right),\ V(\hat{\tau}(f)) \le V(\hat{\tau}) \le V(\hat{\tau}(m))$$

$$(ii)\ \left(\frac{1+q}{q} \le \frac{em(1-em)}{ef(1-ef)}\right),\ V(\hat{\tau}(m)) \le V(\hat{\tau}) \le V(\hat{\tau}(f))$$

$$(iii)\ \left(\frac{1-q}{1-2q} < \frac{em(1-em)}{ef(1-ef)} < \frac{1+q}{q}\right),\ V(\hat{\tau}) < Min[V(\hat{\tau}(m)), V(\hat{\tau}(f))]$$

- Trimming: Escolher subamostra $A \in X$ tal que efeito é estimado com menor variância, neste exemplo:

$$A = f, \text{ se } \left(\frac{em(1-em)}{ef(1-ef)} \le \frac{1-q}{1-2q}\right)$$

$$A = m, \text{ se } \left(\frac{1+q}{q} \le \frac{em(1-em)}{ef(1-ef)}\right)$$

$$A = m, f, c.c.$$

Procedimento de Trimming

Excluímos i's com $X_i \notin A$, portanto: $\tau_A = E(\tau(x_i)|X_i \in A)$

Caso 2: Generalizando para matriz X de variáveis contínuas

$$V(\hat{\tau}) = E\left(\frac{\sigma_t^2(X_i)}{e(X_i)} + \frac{\sigma_c^2(X_i)}{1 - e(X_i)}\right)$$

- Quando $e(X_i) \rightarrow 1$ ou $e(X_i) \rightarrow 0$, o limite da variáncia será alto
- Quando $e(X_i) = 0.5$, variância é mínima

Procedimento de Trimming

Caso 2: Generalizando para matriz X de variáveis contínuas, cont.

$$\tau_A = E(\tau(x_i)|X_i \in A)$$

$$V(\tau_A) = \frac{1}{q_A} E\left(\frac{\sigma^2_{t(X_i)}}{e(X_i)} + \frac{\sigma^2_{c(X_i)}}{1 - e(X_i)}|X \in A\right)$$

- em que $q_A = P(x \in A)$
- Supondo homocedasticidade:
  - $V^*(\tau_{A^*}) = \frac{\sigma^2}{q_A} E\left(\frac{1}{e(X_i)(1-e(X_i))}|X \in A\right)$
  - $A^* = \{x \in X|\alpha \leq e(x) \leq 1 - \alpha\}$, em que $\alpha = 1/2 - \sqrt{1/4 - 1/\gamma}$ e $\gamma = 2E\left(\frac{1}{e(X_i)(1-e(X_i))}|\frac{1}{e(X_i)(1-e(X_i))} < \gamma\right)$
  - Crump et al 2009 derivam $\tilde{\alpha} = 0,1$

Aula 4

voltar