Técnicas Econométricas para Avaliação de Impacto

Outras Técnicas que Utilizam o Escore de Propensão

Rafael Perez Ribas
Centro Internacional de Pobreza

Brasília, 28 de maio de 2008

Introdução

- O Escore de Propensão (ou *Propensity Score*) é a probabilidade de uma observação receber o tratamento que está sendo investigado, condicionada às suas características observadas.

  $$\Pr\left(D_i = 1 \mid X_i\right)$$

- De uma forma mais geral, podemos tratar o PS como a probabilidade estimada de uma observação ser selecionada para uma amostra ou grupo específico da amostra.
- Portanto, a questão geral de interesse é como corrigir o viés causado pela seleção amostral não-aleatória utilizando a própria probabilidade de seleção.

- Este problema é muito comum, por exemplo, quando utilizamos amostras estratificadas, e.g. PNAD.
  - Solução: atribui-se um peso amostral para cada observação:

    $$w_{ij} = N_j / n_j, \quad i \in j$$

  - Teorema Horvitz-Thompson (1952):

    $$\hat{T}_y = \sum_{i \in S} y_i \cdot w_{ij} \text{ ou } \hat{T}_y = \sum_{i \in S} y_i / p_{ij}, \quad p_{ij} = 1/w_{ij}$$

    é o estimador consistente para o total da população

    $$T_y = \sum_{i \in U} y_i$$

- A lógica é dar um maior peso na análise para os indivíduos menos representados na amostra.
  - Seguindo isso, deriva-se uma literatura de correção do viés de seleção amostral só baseado em ponderações.
  - Por exemplo, um estimador calculado utilizando uma amostra de tratados é representativo do grupo de tratados.
  - Para tornar este estimador representativo de toda a população, precisamos dar mais peso aos tratados com menor PS, pois eles representam aquelas pessoas ausentes no grupo de tratados.
- O IPW tem a mesma função do estimador de seleção do Heckman;
  - Contudo, assumindo outras condições de seleção.

- Dois tipos de PS:
    - Conhecido: quando a seleção é aleatória dentro de estratos conhecidos;
    - Desconhecido: quando a seleção não é aleatória.
        - Ou seja, observam-se apenas alguns critérios que determinaram a probabilidade de seleção.

Modelo Básico

- ATE:

$$\tau \equiv E\left[Y_1 - Y_0\right]$$

- Efeito do tratamento sobre um grupo específico da população:

$$\tau_w \equiv \frac{\int E\left[Y_1 - Y_0 \mid X = x\right] g(x) dF(x)}{\int g(x) dF(x)}$$

onde $g(.)$ é uma função conhecida de observáveis que definem o grupo de interesse.

- Quando o grupo de interesse são os tratados (ATT),

  $$g(x) = p(x) = \Pr(D = 1 \mid X = x)$$

  e

  $$\tau_T \equiv E[Y_1 - Y_0 \mid D = 1]$$

- Problema é que $p(x)$ não é observado em muitos casos.
  - Estimável, mas não observado.
- Além disso, existe o problema fundamental da avaliação de impacto:
  - Não observamos ambos os resultados ao mesmo tempo.

- Contudo, se assumirmos a SUTVA condicional

  $D \perp Y_0 \mid X$,

  podemos estimar o efeito do tratamento condicionado a *x*:

  $$\tau_x \equiv E\left[Y_1 - Y_0 \mid X = x\right]$$
  $$= E\left[Y \mid T = 1, X = x\right] - E\left[Y \mid T = 0, X = x\right]$$

  e o ATE pode ser obtido através de uma média de τ_x sobre toda a distribuição *X*:

  $\tau = E\left[\tau_x\right]$.

- Para evitar o problema de comparar observações com o mesmo conjunto *X*, quando este é muito amplo,
  - Rosenbaum e Rubin (1983) propõem o uso de um indicador síntese: o PS.
- Teorema de Rosenbaum e Rubin:

  $$D \perp Y_0 \mid x \rightarrow D \perp Y_0 \mid p(x)$$

  Se o tratamento e o resultado são ortogonais quando condicionados a todos os *x*, eles também são ortogonais quando condicionados a *p*(*x*).
  - Isso implica que o ajuste pelo PS é suficiente para remover todos os vieses associados com as diferenças nas covariáveis.
- Outra condição é que *p*(*x*) seja uma função balanceada:

  $$x \perp D \mid p(x)$$

Estimação

- O parâmetro ATE pode ser caracterizado de acordo com a seguinte equação de momento:

  $$E\left[\psi\left(Y, D, X, \tau^{*}, p^{*}(X)\right)\right] = 0,$$

  onde

  $$\psi\left(y, d, x, \tau, p(x)\right) = \frac{y \cdot d}{p(x)} - \frac{y \cdot (1-d)}{1-p(x)} - \tau.$$

- Dado um estimador $\hat{p}(x)$ para o PS, é possível estimar τ^{*} calculando o momento médio igualado a zero como função de τ:

  $$(1/N)\sum_{i=1}^{N}\psi\left(Y, D, X, \tau^{*}, p^{*}(X)\right) = 0$$

- Isso nos leva ao seguinte estimador:

$$\hat{\tau} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{Y \cdot D}{\hat{p}(X)} - \frac{Y \cdot (1-D)}{1-\hat{p}(X)}\right).$$

- Este é o estimador semi-paramétrico do ATE.
    - Uma simples diferença ponderada pelo PS.
- Este estimador possui uma variância “conservadora”, caso $\hat{p}(x)$ for estimado parametricamente.
    - Porém, métodos de bootstrap podem ser utilizados para se chegar a resultados mais eficientes.

- Para estimar o efeito do tratamento sobre um grupo específico da população, é utilizada a seguinte função de momento:

$$\psi\left(y, d, x, \tau_w, p(x)\right) = g(x)\left(\frac{y \cdot d}{p(x)} - \frac{y \cdot (1-d)}{1-p(x)} - \tau_w\right)$$

que nos leva ao estimador:

$$\hat{\tau}_w = \sum_{i=1}^{N} g(X)\left(\frac{Y \cdot D}{\hat{p}(X)} - \frac{Y \cdot (1-D)}{1-\hat{p}(X)}\right) \Big/ \sum_{i=1}^{N} g(X)$$

- Para o caso do ATT:

$$\hat{\tau}_T = \sum_{i=1}^{N} \hat{p}(X)\left(\frac{Y \cdot D}{\hat{p}(X)} - \frac{Y \cdot (1-D)}{1-\hat{p}(X)}\right) \Big/ \sum_{i=1}^{N} \hat{p}(X)$$

- Na prática, podemos estimar:

$$\hat{\tau} = D \cdot E[Y \mid D = 1] - (1 - D) \cdot E[Y \mid D = 0],$$

ponderado por

$$\hat{w} = \frac{D}{\hat{p}(X)} + \frac{1 - D}{1 - \hat{p}(X)}$$

- Para o caso do ATT, é só utilizar o seguinte peso:

$$\hat{w}_T = D + (1 - D) \cdot \frac{\hat{p}(X)}{1 - \hat{p}(X)}$$

- E do ATU:

$$\hat{w}_U = D \cdot \frac{1 - \hat{p}(X)}{\hat{p}(X)} + (1 - D)$$

Combinação de métodos paramétricos e semi-paramétricos

- O estimador IPW é, na realidade, uma versão semi-paramétrica para controlar seleção sobre observáveis.
- A versão paramétrica seria o ajuste de regressão (Rubin, 1977):

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

- Se tanto a regressão quanto o PS são corretamente especificados, ambos estimadores são consistentes.

- A diferença é que o estimador semi-paramétrico controla as relações de X com D,
- Enquanto o estimador paramétrico controla as relações de X com Y.
- Para assumir uma relação de Y com D da forma mais ortogonal possível, Hirano e Imbens (2002) propõem combinar ambos estimadores.
  - Ou seja, calcular o estimador do ATE usando o ajuste de regressão ponderado pela probabilidade inversa de seleção.
- A regra de inclusão de covariáveis, tanto na regressão como no propensity score varia.
  - Hirano e Imbens (2002) recomendam incluir aquelas com correlação significativa (critério *ad hoc*).

- Para o caso de estimarmos um ATT:

$$Y = \alpha + D\tau + X\beta + \left(X - \overline{X}_1\right)D\omega + u,$$

ponderado por

$$\hat{w}_T = D + \left(1 - D\right)\cdot \frac{\hat{p}\left(Z\right)}{1 - \hat{p}\left(Z\right)}$$

- Importante:
    - O método do IPW não se limita somente à ponderação de regressões lineares, mas a qualquer regressão estimada por Máxima Verossimilhança.
    - Sempre gerando intervalos de confiança conservadores.

O Estimador Semi-Paramétrico Dif-Dif

- Os estimadores convencionais de Diferença-nas-Diferenças, exigem a seguinte condição:
  $$E\left(u_{i,1} - u_{i,0} \mid D = 1\right) = E\left(u_{i,1} - u_{i,0} \mid D = 0\right).$$
  - As variações devem ser ortogonais à seleção para o tratamento.
- Esta condição é, contudo, frequentemente violada quando não se tem um experimento.
- E somente incluir variáveis de controle na especificação da regressão Dif-Dif não resolve o problema se o efeito do tratamento varia com *X* (aulas anteriores).

- Solução de Abadie:
  - Estimar o modelo Dif-Dif com ponderação pelo PS.
- Além disso, é possível ainda incluir variáveis de interação na própria regressão:

$$Y_t = \alpha + D \cdot t \cdot \tau + D\gamma + \delta t + X_t \beta + \left(X_t - \overline{X}\right) D \cdot t \cdot \omega + u_t \text{ ,}$$

ponderado por

$$\hat{w} = \frac{D}{\hat{p}(Z)} + \frac{1 - D}{1 - \hat{p}(Z)}$$

Uma aplicação mais geral

- Estimação do efeito de tratamento sobre a relação de Y com X.

$$Y_0 = \alpha_0 + X\beta_0 + u_0 \quad \text{se } D = 0$$

$$Y_1 = \alpha_1 + X\beta_1 + u_1 \quad \text{se } D = 1$$

- Nas regressões para grupos específicos, os coeficientes representam não só as mudanças impostas pelo tratamento,
    - Mas também a relação entre Y e X intrínseca à composição de cada grupo.
- Exemplo:
    - Retorno da educação sobre os salários de quem estudou no sistema público ou privado.