3. METODOLOGIA

Neste capítulo apresentamos as metodologias comumente utilizadas para a estimação da avaliação de impacto, além da metodologia utilizada no nosso trabalho, *matching* com escore de propensão e diferenças em diferenças.

Iniciamos com as formas mais simples de avaliação do resultado de um programa. Na primeira metodologia compara-se o resultado de um grupo participante do programa antes e depois da sua implementação; na segunda, comparam-se os resultados de um grupo participante com os de um grupo não participante do programa. Outra forma de estimar o efeito da avaliação é através de um modelo de regressão. Mostramos por que tais metodologias geralmente produzem resultados viesados da estimativa do efeito do programa. Finalmente apresentamos, com mais detalhes, a metodologia de *matching*.

A exposição envolverá técnicas experimentais, como a aleatorização, e não experimentais, como as de variáveis instrumentais, *matching*, e *matching* com diferenças em diferenças.

3.1. Modelo geral para avaliação de impacto

O problema central na avaliação de qualquer programa social é o fato de que os "indivíduos" participantes do programa não podem ser observados simultaneamente no estado alternativo de não tratamento. A solução para este problema seria a estimação do resultado de uma situação contrafactual representando o que aconteceria com os participantes do programa caso essa participação não tivesse acontecido. Segundo Heckman e Smith (1995), esse problema fica um pouco mais complicado quando a participação no programa é voluntária, em vez de ocorrer por uma seleção exógena.

Considere as seguintes equações:

- $Y_1 = P(X)ß_1 + v_1$, variável de resultado do indivíduo que participou do programa;
- $Y_0 = P(X)ß_0 + v_0$,variável de resultado do indivíduo que não participou do programa;
- X: características observáveis dos indivíduos que podem influenciar na participação no programa.
- $P(X) = \begin{cases} 1, \text{ indicando a participação no programa;} \\ 0, \text{ indicando que o indivíduo não participou do programa.} \end{cases}$
- $D = Y_1 – Y_0$, o ganho para qualquer indivíduo que participou do programa.
- v_1 e v_0 = Inovações

Esse ganho não pode ser apurado tendo em vista que pelo menos um dos termos da diferença não pode ser observado. Seja qual for t (unidade de tempo), um indivíduo terá participado do programa sendo Y_1 observado e Y_0 não observado, ou não terá participado deste programa em particular, sendo Y_0 observado e Y_1 não observado. O fato de termos pelo menos Y_0 ou Y_1 não observáveis torna impossível calcularmos D. Assim, temos de desenvolver uma estimativa para uma avaliação de Y_i, utilizando uma abordagem estatística para o problema. A solução estatística identifica a média, ou qualquer outra estatística, dos efeitos causais do programa. Dado que o efeito causal para um único indivíduo não pode ser observado, esta propõe métodos para computar o efeito causal médio para a população ou para determinados subgrupos.

Seleção para o programa

Heckman (1978) desenvolveu um modelo de participação aplicado por Ashenfler e Card (1985). Este modelo apresentado a seguir foi sumarizado por Heckman, LaLonde e Smith (1999).

P(X) retrata a decisão de participação no programa. P* é a variável latente que gera P(X) de acordo com a seguinte regra:

$$P(X) = 1[P^*(X) >= 0] = 1[XT + v_p >= 0],$$

Onde 1[A] é a função indicadora que toma o valor 1 se o evento A for verdadeiro e de 0 em caso contrário. Para cada potencial realização de X, x, definimos a variável $P(X) = [XT >= v_p]$. P(X) indica se o indivíduo faria parte ou não do programa caso o seu valor de X seja fixado em x, mantendo as características não observáveis v_p constantes. Finalmente, assumimos que v_p , v_1 e v_0 são independentes de X.

Estimação

Os estudos sobre avaliação de impacto de políticas sociais devem enfocar a questão de quanto o programa é responsável pela mudança na média (ou na estatística em questão) da variável de resultado do grupo dos participantes, em comparação com o resultado que ocorreria caso eles não fizessem parte do programa.

A maneira mais intuitiva para analisarmos o efeito de um programa seria observar o mesmo grupo de participantes antes e depois de sua implementação. Porém, desta forma não estaríamos considerando as mudanças ocorridas no cenário macroeconômico que nada tiveram a ver com o programa e que poderiam ter influenciado os novos resultados do grupo analisado.

Outro método muito utilizado para apurar o impacto de programas sociais é utilizar o resultado de um grupo de não participantes do programa para estimar qual seria o resultado do grupo participante caso não tivesse participado. Entretanto, os resultados dos não participantes podem ser sistematicamente diferentes dos resultados que os participantes teriam caso não tivessem participado, gerando um viés de seleção no impacto estimado. Alguns estimadores não experimentais são capazes de ajustar esse viés sob diversas hipóteses, conforme apresentam Heckman e Robb (1985,1986), Heckman (1990ª) e Heckman e Smith (1996), os quais são apresentados mais adiante. Estimadores experimentais e não experimentais são capazes de eliminar esse viés, como veremos mais à frente.

A hipótese de independência condicional, introduzida por Lechner (1999), é uma maneira de evitar o viés de seleção. Os resultados a serem comparados devem estar condicionais à algumas variáveis X, que influenciam tanto na seleção de participação no programa quanto nos resultados potenciais. Assim, condicional à X, a probabilidade de participação no programa é estocasticamente independente dos resultados potenciais. Formalmente, teríamos:

$$Y_{0\,,}Y_1 \perp P \mid X$$

Quais seriam os fatores de interesse na avaliação de impacto de programas sociais?

O impacto causal médio do programa pode ser medido pelo impacto médio do programa, chamado na literatura de ATE (*Average Treatment Effect*):

$$E\,(Y_1 - Y_0\,) = ATE(x) = E(D \mid X = x) \qquad \textbf{(1)}$$

O termo (1) é a diferença entre o resultado esperado quando o indivíduo participa do programa e o resultado esperado de quando o indivíduo não participa do programa, para um indivíduo escolhido aleatoriamente na população.

Podemos chegar à estimativa condicional integrando (1) sob a distribuição de X:

ATE = E (D) = S ATE(X) dF(X) ˜ 1/n Soma [ATE (x_i)] , onde n é o tamanho da amostra.

Porém, o fator de medida que mais retrata o impacto de um programa é "o impacto médio do programa nos participantes" ou ATT (*Treatment on the Treated*). Esta medida é a que melhor responde à pergunta: como o programa mudaria o resultados dos participantes caso eles não tivessem participado do programa? Formalmente, pode ser representada como:

$$ATT(x, P(X)=1) = E(D \mid X = x, P(X) = 1) = E(Y_1 - Y_0 \mid X = x, P(X) = 1) \quad \textbf{(2)}$$

O valor de ATT é a diferença do resultado esperado de um indivíduo selecionado aleatoriamente da subpopulação de indivíduos que participaram do programa e o resultado que este indivíduo teria caso não tivesse participado do programa.

Como no caso do fator de medida ATE, podemos chegar a uma estimativa incondicional integrando ATT sob a distribuição conjunta de X e Y para aqueles que realmente participaram do programa. Sendo n_t aqueles que fizeram parte do programa, ou seja, o número de observações onde P(X)=1, ATT pode ser aproximado como:

$$ATT = E(D \mid P(X) = 1) = E(Y_1 - Y_0 \mid X, P(X)=1)$$

$$= S\ ATT(X, P(X)=1)dF(X \mid P(X) = 1)\ \tilde{}\ 1/\ n_t \text{ Soma}_{i=1} P_i\ ATT(x_i, P(x_i) = 1).$$

Os fatores de medida ATT e ATE serão iguais quando, condicional à X = x, as médias dos erros da equação de participação $?_p$ e das equações de resultado (v_1, v_0) forem independentes.

3.2. Estimação através dos termos contrafactuais

Conforme vimos anteriormente, o impacto médio de um programa nos seus participantes, ATT, será $E(Y_1 - Y_0 \mid X = x, P(X) = 1)$.

O termo $E(Y_1|X, P(X)=1)$, que descreve a distribuição condicional de Y_1 dado um vetor de variáveis independentes X e $P(X) = 1$, pode ser observado com a experiência obtida pelo grupo que participou do programa. O que está faltando é a média do termo contrafactual $E(Y_0| X, P(X) = 1)$, que expressa o que aconteceria com o grupo dos participantes caso eles não tivessem participado do programa. Formalmente, o termo contrafactual pode ser definido como:

$$E(Y_0| X, P(X) = 1) = E[E(Y_0|X, P(X)=0) \mid P(X) = 1] =$$

$$\int E(Y| X = x, P(X)=0).\, f_{X|P(X)=1}(x)\, dx$$

$E(Y_0|X, P(X)=1)$ é a média do resultado contrafactual dos indivíduos participantes caso eles não tivessem feito parte do programa, que pode ser obtida com o ajuste dos resultados observados dos indivíduos que não participaram à distribuição das variáveis X na população dos participantes.

A questão que se coloca é: como construir os termos contrafactuais?

A forma mais simples de construção do termo contrafactual é assumir que, condicionado à X, os resultados dos não participantes reflete o que teria acontecido com os participantes caso eles não tivessem participado do programa:

$$E(Y_0| X, P(X) = 0) \sim E(Y_0| X, P(X) = 1)$$

O viés de seleção associado ao impacto do programa E(D |X, P(X) = 1) aparece quando essa aproximação não é verdadeira.

$B(X) = E\ (Y_0|X, P(X) = 1) – E\ (Y_0|X, P(X) = 0)$

Estimação por regressão

Consideremos um modelo onde o resultado Y depende de um conjunto de variáveis explicativas X.

Sendo Y_0 resultado daqueles que não participaram do programa, e Y_1 resultado daqueles que participaram do programa, em um modelo de regressão tradicional temos:

$Y_0 = Xß_0 + v_0$ e

$Y_1 = Xß_1 + v_1$ com,

$E(v_0|X) = 0$ e

$E(v_1|X) = 0$.

Para melhor ilustrar os modelos, adotamos as seguintes notações:

$\mu_0\ (X) = E(Y_0|X)$ e

$\mu_1\ (X) = E(Y_1|X)$.

Assim, teremos:

$Y_0 = \mu_0 (X) + v_0$ **(3a)**

$Y_1 = \mu_1 (X) + v_1$ **(3b)** com

$E(v_0|X) = 0$ e

$E(v_1|X) = 0$.

Como apresentou Heckman (1995), o resultado observado será Y_0 ou Y_1, e pode ser escrito como:

$Y = PY_1 + (1 – P)Y_0$.

Neste caso, P representa a probabilidade que um indivíduo tem de fazer parte do programa a ser analisado.

Se substituirmos 3a e 3b na expressão anterior teremos:

$$Y= P[\mu_1(X) + v_1] + (1 – P)[\mu_0(X) + v_0] = \mu_0(X) + v_0 – P[\mu_0 (X)] – P(v_0) + P[\mu_1(X)] +P(v_1)$$

$$Y = \mu_0 (X) + P[\mu_1(X) – \mu_0(X) + v_1 – v_0] + v_0 \quad \textbf{(4)}$$

O termo que multiplica P será o ganho com o programa. Esse ganho tem duas componentes:

a) $\mu_1(X) – \mu_0(X)$ = o ganho médio para toda a população;
b) $v_0 – v_1$ = o ganho particular de cada indivíduo, participante do programa, que se difere da média.

Nessa notação o ganho médio, ATE, é: $E(D| X) = \mu_1(X) – \mu_0(X)$.

Neste mesmo estudo, Heckman (1995) apresenta o impacto do programa nos participantes ATT como:

$$ATT = E(D|\, X, P(X) = 1) = \mu_1(X) - \mu_0(X) + E(v_1 - v_0|\, X, P(X) = 1).$$

A diferença do ganho médio (ATE) e do impacto do programa nos participantes (ATT) é o termo $E(v_1 - v_0|\, X, P(X)=1)$. Esse termo representa quanto o ganho dos participantes difere do ganho médio que toda a população teria se tivesse participado do programa. Apesar de, na média, os indivíduos ganharem $\mu_1(X) - \mu_0(X)$, para cada participante esse ganho será diferente. A equação (4) pode então ser reescrita em função desses parâmetros:

Em função de $ATE = [E(D|\, X)]$:

$$Y = \mu_0(X) + P[E(D|\, X)] + \{ v_0 + P(v_1 - v_0)\} \quad \textbf{(5)} \quad \text{e}$$

Em função de $ATT = [E(D|\, X, P(X) = 1)]$:

$$Y = \mu_0(X) + P[E(D|\, X, P(X) = 1)] + [v_0 + P[v_1 - v_0 - E(v_1 - v_0|X, P(X) = 1)] \textbf{(6).}$$

Definindo as seguintes notações:

$a(X) = \mu_0(X)$

$ß(X) = \mu_1(X) - \mu_0(X) + (v_1 - v_0)$

$E[ß(X)|\, X] = \underline{\mathbf{ß}}(X)$

$e = v_1 - v_0$

$v_0 = v$

A equação **(4)** pode ser reescrita como:

$$Y = a + P\beta + v \qquad \textbf{(4')}$$

Mesmo estando condicionado à X, β irá variar na população refletindo a heterogeneidade dos indivíduos quando se trata da resposta ao programa aplicado.

A equação **(5)** passa a ser:

$$Y = a + P\underline{\boldsymbol{\beta}} + \{v + Pe\} \qquad \textbf{(5')}$$

Onde $\underline{\boldsymbol{\beta}}$ será o impacto médio do programa na população como um todo. O impacto do programa apenas nos participantes será:

$$E(D|\ X, P(X) = 1) = \underline{\boldsymbol{\beta}}(X) + E(e|\ X, P(X) = 1) = \beta^*(X)$$

Podemos chamar $\underline{\boldsymbol{\beta}}(X)$ de $\underline{\boldsymbol{\beta}}$ e $\beta^*(X)$ de β^* deixando a condição à X implícita.

A equação análoga à **(6)** será:

$$Y = a + P\beta^* + \{v + P[e - E(e|\ X, P(X) = 1)]\} \qquad \textbf{(6').}$$

A regressão para o coeficiente de P pode ser escrita como a diferença de duas médias. Sendo ***ß+*** o coeficiente estimado para a regressão de Y em P, este pode ser representado de três maneiras:

$\boldsymbol{\beta+} = E(Y_1 | X, P(X) = 1) - E(Y_1 | X, P(X) = 0)$ **(7a)**

$\boldsymbol{\beta+} = \underline{\boldsymbol{\beta}} + E(e | X, P(X) = 1) + [E(v | X, P(X) = 1)] - E[v | X, P(X) = 0)]$ **(7b)**

$\boldsymbol{\beta+} = \beta^* + E(v | X, P(X) = 1) - E(v | X, P(X) = 0)$ **(7c)**

ß* será o ATT, impacto médio do programa nos participantes, e o termo $E(v | X, P(X) = 1) - E(v | X, P(X) = 0)$ é o viés de seleção. O viés de seleção representa como o resultado dos participantes e dos não participantes se diferencia no estado original, antes de o programa ser aplicado.

Geralmente, $E(Y_0|X, P(X) = 1) \neq E(Y_0|X, P(X) = 0)$, caso a seleção para participação em um programa não seja feita de maneira aleatória. Por exemplo, se trabalhadores mais empenhados têm uma maior propensão a participarem de programas de treinamento, o salário deles seria mais elevado mesmo se o treinamento não existisse, assim $E(Y_0 |P(X) = 1) > E(Y_0| P(X) = 0)$.

A diferença entre ß* e **<u>ß</u>** é a diferença entre o ganho médio não observável de toda a população e o ganho médio não observável daqueles que participaram do programa:

$$E(e| X, P(X) = 1) = E(v_1 - v_0 | X, P(X) = 1)$$

Ou seja, a diferença entre os estimadores está nas componentes não observáveis dos dois grupos (P(X) = 0 e P(X) = 1).

Esses estimadores irão coincidir quando:

$E(e| X, P(X) = 1) = 0 = E(v_1 – v_0 | X, P(X) = 1)$.

Casos especiais onde $ß^* = \underline{ß}$:

i) Quando $v_1 = v_0$

Neste caso, não existem componentes não observáveis no ganho. O modelo assume que, condicional à X, o efeito da participação no tratamento/programa é o mesmo para todos os indivíduos.

ii) Quando v_1 diferente de v_0 e essa diferença entre v_1 e v_0 não influencia na seleção para participação no programa. No período inicial de engajamento ao programa não conhecemos e e sua melhor previsão é e = 0 ou outra constante qualquer. Assim, sendo E(e | X, P(X) = 1) = 0 e então $ß^*= \underline{\mathbf{ß}}$ tendo em vista que e $= v_1 – v_0$.

Vale ressaltar que em ambos os casos, i e ii, a estimação de ß* ou de **ß** utilizando a diferença entre os resultados de participantes e não participantes se defronta com o problema de P ser correlacionado com v, ou seja, de a seleção para o tratamento estar correlacionada com os elementos exógenos não observáveis que influenciam no resultado.

3.3.Métodos para uma estimação não viesada do impacto de um programa

3.3.1. Aleatorização

Existem duas formas para avaliarmos o impacto de um determinado programa ou política. Uma é a maneira não experimental ou econométrica, e a outra é a forma experimental. A forma experimental consiste em separar aleatoriamente os indivíduos em um grupo de controle (indivíduos não participantes do programa) e um grupo de tratamento (participantes do programa). Podemos obter o impacto do programa comparando os resultados do grupo dos participantes com os resultados do grupo de controle, em alguns anos depois da implementação do programa.

Se a aleatorização for feita adequadamente, estatisticamente não haverá nenhuma diferença entre os grupos de controle e de tratamento, a não ser a participação no programa, não existindo assim viés de seleção.

O argumento mais forte em favor do método experimental é que, sob certas condições, o método resolve o problema da avaliação de impacto na impossibilidade de observarmos o mesmo indivíduo no estado de tratamento (participação no programa) e no estado de não tratamento (não participação no programa). Se o mesmo indivíduo pudesse ser observado nos dois estágios, o impacto do programa nesta pessoa poderia ser calculado através da comparação da variável de resultado desse indivíduo nos dois estágios, conforme dito no capítulo anterior.

Suponhamos que um indivíduo possa estar no estágio “1”, de participação, ou no estágio “0”, de não participação no programa, e que seus resultados são Y_0 e Y_1, dependendo do estado. O ganho/perda com o programa seria a diferença entre os resultados nos dois estados ($Y_0 - Y_1$). Porém, como o impacto de um programa não pode ser determinado para um indivíduo em particular, devemos focar a avaliação na distribuição do impacto entre os indivíduos, $F(Y_0 - Y_1)$.

Para programas direcionados a um determinado grupo de indivíduos, a análise de impacto deve ser feita em torno daqueles que realmente participaram (P=1). Assim, como apresentado acima, o fator de interesse será $E(D \mid P(X) = 1) = E(Y_1 - Y_0 \mid X, P(X)=1)$.

A aleatorização resolve o problema de viés de seleção para as médias, criando um grupo de controle experimental formado por indivíduos que teriam participado do programa (P=1), mas que aleatoriamente foram vetados da sua participação. Esta metodologia é apresentada por Heckman e Smith (1995), e sob as condições de que a aleatorização não altera o grupo de participantes nem o seu comportamento e de que programas substitutos ao programa avaliado não estariam disponíveis, o resultado médio esperado para o grupo de controle experimental estima corretamente o resultado contrafactual $E(Y_0| X, P(X) = 1)$.

Suponhamos que $P^*=1$ para indivíduos que participariam de um programa onde a seleção é aleatória e $P^*=0$ para todos os outros casos. A aleatorização é aplicada à população com $P^*=1$. Dento dessa população determinamos aleatoriamente os indivíduos com $r=1$ para formarem o grupo de tratamento e com $r=0$ para o grupo de controle, àqueles a quem a participação no programa foi vetada. Sob as condições descritas, teremos que o resultado para o grupo de tratamento experimental mede o resultado normal para os participantes, tal que $E(Y_1|X,P(X)=1) = E(Y_1| X,r=1,P^*=1)$ e que o resultado para o grupo de controle experimental mede qual seria o resultado que os participantes teriam caso eles não fizessem parte do programa, tal que $E(Y_0| X, P(X) = 1) = E(Y_0| X, r=0, P^* = 1)$.

Assim podemos descrever o impacto médio em um programa nos participantes como a diferença de duas médias:

$$E(Y_1 - Y_0| X, P(X) = 1) = E(Y_1| X, r=0, P^* = 1) - E(Y_0| X, r=0, P^*=1)$$

Os resultados médios dos grupos de controle e de tratamento experimentais irão gerar as estimativas do lado direito da equação. Assim, a aleatorização irá funcionar como uma variável instrumental criando uma variabilidade na participação no programa entre os selecionados a participar.

Para outras estatísticas, tal como o impacto mediano de um determinado programa, o qual depende da distribuição conjunta dos resultados de participantes e não participantes, essa relação não pode ser aplicada.

Uma situação bastante comum na literatura de avaliação de impacto de programas ou tratamentos é o caso do “efeito comum”. Este se refere à situação na qual todos os indivíduos têm o mesmo ganho/perda com o programa, ou seja, $E(D|P(X)=1) = E(Y_1 - Y_0|X,P(X)=1)$ é o mesmo para todos. Essa situação favorece a análise experimental já que, mesmo se a aleatorização modificar o conjunto de participantes, o fator estimado permanece o mesmo. Neste caso, a aleatorização será capaz de gerar a distribuição completa dos resultados, de maneira que outras estatísticas possam ser estimadas, tal qual o impacto mediano e o impacto médio para toda a população.

Neste caso em especial, $Y_1 - Y_0 = D$ é constante para todos os indivíduos e pode ser identificado através de maneira experimental. Dado D, conhecendo Y_0 ou Y_1 podemos determinar o outro resultado, não observado. Se F_1 é a distribuição acumulada de Y_1 (CDF) e F_0 de Y_0 , então $F_1(Y_0 + \Delta) = F_0(Y_0)$, e através dessa relação podemos encontrar outras características da distribuição, além da média, do impacto de um programa ou tratamento.

A aleatorização não vai eliminar o viés de seleção e sim balancear o viés entre as amostras de participantes e de não participantes. Considerando um modelo simples de regressão:

$Y = a + P\beta + v$, onde:

Y = resultado de interesse;

a = resultado médio quando ninguém participa;

ß = efeito comum da participação;

v = choque aleatório observado pelo indivíduo, mas não pelo avaliador.

Neste modelo D= ß e é constante. O problema de viés de seleção aparece quando a participação, indicada por P=1, é dependente dos choques aleatórios não observados “v”, tal que E(v|P) diferente de 0. Os ganhos médios nos grupos de tratamento e de controle experimentais serão:

$$E(Y|\ r=1, P=1) = a + ß + E(v|\ P=1) \text{ e}$$

$$E(Y|\ r=0, P=1) = a + E(v|\ P=1) \text{ , respectivamente.}$$

Subtraindo as duas médias teremos ß, o parâmetro de interesse. A aleatorização não irá garantir que E(v| P*=1)=0 ou que E(v| P=1)=0. Entretanto, irá balancear o viés entre as amostras de participantes e de não participantes de tal maneira que este seja cancelado no cálculo da estimativa do impacto médio.

A principal vantagem em adotar um método experimental é que este nos leva a um grupo de controle similar ao grupo de tratamento no que diz respeito tanto às características observáveis quanto às características não observáveis. Desse modo, as diferenças encontradas entre os grupos após a implementação do programa podem ser atribuídas exclusivamente ao próprio programa.

Tanto a forma experimental como a forma não experimental de avaliação de impacto têm o mesmo problema a resolver, já que nenhum indivíduo pode ser observado simultaneamente nos estados de tratamento e de não tratamento. Na metodologia experimental, a situação contrafactual é representada pelo resultado do grupo de controle formado por indivíduos selecionados a participar do programa, mas que, através de uma seleção aleatória, foram vetados dessa participação. Na metodologia não experimental, a situação contrafactual é estimada através de métodos econométricos, que utilizam modelos de participação no programa e de resultados. É muito difícil viabilizar experimentos aleatórios em avaliação de políticas públicas, tendo em vista que alguns componentes da estratégia experimental mencionada acima podem ser inaceitáveis do ponto de vista ético ou político.

Contudo, para resolver o problema de inferência causal em políticas de governo devemos utilizar estimadores “não experimentais” capazes de ajustar o viés de seleção. Apresentamos mais adiante os métodos não experimentais de estimação de impacto de programas sociais mais eficazes a saber: variáveis instrumentais, metodologia de *matching* e *matching* com diferenças em diferenças.

No caso da municipalização, a aleatorização não é possível, tendo em vista que foi um processo observado sem que tivesse sido programado, ou seja, na maioria dos estados não existiu uma política educacional direcionada para a mudança de rede por parte das escolas estaduais.

3.3.2. Método não experimental – Variáveis instrumentais

Nem sempre é possível obter dados referentes aos participantes e não participantes antes e depois da implementação do programa. Assim, quando não temos uma linha de base, a metodologia de variáveis instrumentais pode ser utilizada para medir o impacto de um determinado programa.

Essa metodologia é aplicada quando temos problema de endogeneidade dos regressores. No caso da análise de impacto, o problema existe pois, normalmente, as variáveis que influenciam na escolha da participação no programa também são correlacionadas com a variável de resultado. As variáveis instrumentais são aquelas correlacionadas com a participação no programa (P), mas que não estão na regressão da variável de resultado e nem são correlacionadas com o erro dessa equação.

Este método de avaliação de impacto foi apresentado por Heckman (1995) e para compreendê-lo devemos recorrer às seguintes definições citadas anteriormente:

$Y = a + P\beta + v$, onde:

Y = resultado de interesse;

a = resultado médio quando ninguém participa;

ß = efeito comum da participação;

v = choque aleatório observado pelo indivíduo, mas não pelo avaliador, proveniente de variáveis que influenciam no resultado Y, mas que não estão presentes na equação, com $E(v) = 0$.

P(X) = 1, se o indivíduo participa do programa/tratamento;

0, se o indivíduo não participa.

Devemos prestar muita atenção à possibilidade de P estar correlacionada com v. Indivíduos que têm um valor alto de v (ou Y) sem o programa, muitas vezes podem ser aqueles selecionados a participarem de um determinado programa. Formalmente essa relação pode ser definida como:

$E(v \mid P(X) = 1)$? 0, e neste caso a estimação simples de regressão geraria um ß viesado.

Sendo o conjunto de variáveis instrumentais Z, este tem duas propriedades:

a) É descorrelatada com v : $E(v \mid Z) = 0$;
b) Tem correlação com P : $E(P \mid Z) = Pr(P(X) = 1 \mid Z)$.

Aplicando essas duas propriedades à equação de resultado, apuramos a média condicional à Z:

$$E(Y \mid Z) = E(a + Pß + v \mid Z) = a + ß\, E(P \mid Z) + 0 = a + ß\, Pr(P(X) = 1 \mid Z).$$

Partindo da hipótese de que existem dois valores para Z, Z_1 e Z_2, com probabilidades distintas, teremos que:

$ß = [E(Y \mid Z_1) – E(Y \mid Z_2)] / [Pr(P(X) = 1 \mid Z_1) – Pr(P(X) = 1 \mid Z_2)]$ sendo que,

$E(Y \mid Z_1) = a + ß\ Pr(P(X) = 1 \mid Z_1)$ e

$E(Y \mid Z_2) = a + ß\ Pr(P(X) = 1 \mid Z_2)$

Esse método exige um instrumento Z. Porém, surgem as seguintes questões:

- Qual a credibilidade desses instrumentos?
- Sob que condições o método funciona?
- O método pode ser generalizado para casos onde ß difere entre os indivíduos, ou seja, no caso do impacto do programa não ser comum a todos?

Neste último caso, onde o impacto do programa vai variar de indivíduo para indivíduo, o argumento de variáveis instrumentais não irá funcionar, a não ser que o ganho particular de cada um não influencie na sua decisão de participação no programa. Isto é, a metodologia só irá chegar a um estimador não viesado do impacto do programa nos casos em que o ganho individual – que não pode ser previsto através das variáveis observadas no modelo – não influencia na seleção para participação no programa.

Essa metodologia deve satisfazer a duas condições básicas e a uma terceira. As variáveis instrumentais devem ser independentes dos erros das equações (5’) e (6’), e para que **ß** seja identificável:

$E[v + P e \mid X, Z] = 0$ **(A .1.a)**

E para que ß* seja identificável:

$$E[v + P(e - E(e| P(X) = 1) | X, Z] = 0 \qquad \textbf{(A .1.b)}$$

A segunda condição é que P dependa de Z da seguinte maneira:

$$E(P| X, Z) = Pr(P(X) = 1| X, Z) \qquad \textbf{(A.2)}.$$

Z não é completamente explicado por X. Nesta condição está implícita a idéia de que a probabilidade é definida para dois ou mais valores de Z, de tal forma que a probabilidade de participação no programa é uma função não trivial de Z, e que os valores dessa probabilidade irão variar de acordo com os valores de Z para pelo menos algum X. A principal idéia desta condição é que a probabilidade de participação depende não só de Z como de X também. A terceira condição é que Y irá depender de Z apenas através de P. Esta condição é uma conseqüência das duas primeiras, mas seu papel é tão importante que merece uma atenção maior. Ela pode ser dividida em duas condições separadas:

$$E(Y| X,Z) = a + \underline{\boldsymbol{ß}}\, E(P| X,Z) = a + \underline{\boldsymbol{ß}}\, Pr(P(X) = 1| X,Z) \qquad \textbf{(A. 3.a)}$$

Para ß* a hipótese é:

$E(Y\ |X,Z) = a + ß^* E(P|\ X,Z) = a + ß^* Pr(\ P(X) = 1|\ X,Z)$ **(A.3.b)**.

Para dois valores distintos de Z, digamos Z_1 e Z_2, tal que $Pr(P(X) = 1|X,\ Z_1)$? $Pr(P(X) = 1|X,\ Z_2)$, podemos identificar **ß** ou ß* através da diferença entre esses valores:

$$E(Y|X,\ Z_1) – E(Y|X,\ Z_2) = \mathbf{ß}\ [Pr(P(X) = 1|X,\ Z_1) – Pr(P(X) = 1|X,\ Z_2)]$$

Assim,

$$\mathbf{ß} = [E(Y|X,\ Z_1) – E(Y|X,\ Z_2)] / [Pr(P(X) = 1|X,\ Z_1) – Pr(P(X) = 1|X,\ Z_2)]$$

Substituindo a média populacional pela média amostral chegamos ao estimador de variáveis instrumentais que, sob certas condições, converge para ***ß+(X)***, estimador não viesado.

Utilizando (A.3.b), (A.2) e (A.1.b) podemos chegar também a:

$$ß^*= [E(Y|X,\ Z_1) – E(Y|X,\ Z_2)] / [Pr(P(X) = 1|X,\ Z_1) – Pr(P(X) = 1|X,\ Z_2)].$$

De maneira geral, as variáveis instrumentais são variáveis que não pertencem à equação de resultados, mas que irão influenciar a participação ou não no programa.

No modelo onde o efeito do tratamento é o mesmo para todos os participantes, ou seja, e =0 ou $?_1$ – $?_0$=0, Heckman demonstrou que o método das variáveis instrumentais pode identificar um ß não viesado, ou seja, ß* = **ß**. O ponto principal é que devemos encontrar algumas variáveis, ou variáveis Z, descorrelatadas com ? que afetem "P" e que não estão presentes na equação de resultado.

Em um modelo onde e não é um determinante de P, isto é,

$$Pr(P(X) = 1|X, Z, e) = Pr(P(X) = 1|X,Z), \text{ ou}$$

$$Pr(P(X) = 1|X,Z, Y_1 – Y_0) = Pr(P(X) = 1|X,Z),$$

As mesmas condições devem ser satisfeitas para Z. Ou seja, as variáveis Z devem ser descorrelatadas com ? e não podem estar presentes na equação de resultado. Podemos ignorar a componente, pois:

$$E(P e | X,Z) = E(e |X, Z, P(X) = 1)Pr(P(X) = 1|X,Z) = 0, \text{ tendo em vista que}$$

$$E(e |X,Z,P(X) = 1) = 0.$$

Assim, nesses dois casos particulares onde **ß** = ß* (ATE=ATT) podemos utilizar a metodologia convencional de variáveis instrumentais.

O que devemos fazer em casos mais genéricos? Consideremos a equação (6') associada a ß*. Tendo em vista que a única fonte de dependência entre o termo de erro e P seja através de ? e não de P(e – E(e |X,P(X) = 1)), o método de variáveis instrumentais parece funcionar bem. Se as condições (A–1.b), (A–2) e (A–3.b) forem satisfeitas, este pode perfeitamente estimar ß*. O que precisamos é de uma variável Z que afete a participação no programa, mas que não afete os parâmetros da equação de interesse. Isso exige que:

$E(ß + e\,|X,Z, P(X) = 1) = \mu_1(X) - \mu_0(X) + E(e\,|X,Z, P(X) = 1) = \mu_1(X) - \mu_0(X) + E(v_1 - v_0|X,Z, P(X) = 1)$ independente de Z.

Esta condição pode ser satisfeita nos dois primeiros casos onde $?_1 - ?_0 = 0$ ou $E(e \mid X, Z, P(X) = 1) = 0$ (e não pode ser previsto por Z e X). Na verdade, também pode ser satisfeita em um terceiro caso, onde e não pode ser prevista através de Z mas sim através de X. Porém, em geral, essa condição é muito difícil de ser satisfeita.

Nos casos em que o impacto do programa varia de participante para participante, qualquer aplicação da metodologia de variáveis instrumentais requer que façamos algumas deduções a respeito de como os indivíduos tomam suas decisões sobre sua participação no programa. Neste caso, não é possível uma analise estatística.

Se todas as variáveis de controle X da regressão estiverem incluídas nas variáveis observáveis Z que influenciaram P, e se "v" for descorrelatado com "e" tal que as variáveis não observáveis que afetam na colocação no programa não influenciam os resultados condicionais à X, a possibilidade de P ser correlatado com "e" é eliminada e assim P será exógeno à regressão de resultados.

Na análise do impacto da municipalização o problema de endogeneidade das variáveis pode ser observado com bastante freqüência. Porém, o fato de estarmos trabalhando com dados em painel, ou seja, o fato de termos disponíveis os dados de uma série de anos, nos permite utilizar o método de diferença em diferença, que consegue apurar o impacto de um programa mesmo na existência de viés. Esse método é especificado mais à frente.

3.3.3. Método não experimental de *matching*

A metodologia de *matching* consiste na escolha de um grupo de controle ideal a partir de uma amostra maior que a amostra do grupo de tratamento. O grupo de controle é “casado” com o grupo de tratamento a partir de um conjunto de características observadas, ou utilizando um escore de propensão (probabilidade de participar dadas certas características). Quanto mais próximos esses escores, melhor o *matching*. Conforme Ravallion (2000), um bom grupo de controle vem do mesmo ambiente econômico, social e cultural que o grupo de tratamento.

Escore de propensão

A estimativa dos termos contrafactuais que representam o resultado que um grupo de participantes de um determinado programa teria caso não tivesse feito parte deste programa, pode ser bastante complicada na prática, tendo em vista que ela exige o uso de técnicas não paramétricas para a estimação de E (Y| X, P). Muitas vezes o número de variáveis X é muito grande, tornando a regressão não paramétrica de E (Y| X, P) complexa. Foi para facilitar essa estimativa que Rosenbaum e Rubin (1983) introduziram o conceito de escore de propensão. Esta seria uma maneira de reduzir a dimensionalidade do problema de estimação. Eles mostraram que se (Y_1 e Y_0) e P forem independentes dados X, então eles também seriam independentes dado um escore de propensão unidimensional *p(x)*.

p(x) = Pr(P(X) = 1|X), ou seja, a probabilidade de um indivíduo participar do programa dadas as suas características X . **(B.1)**

Como resultado, o impacto médio de um programa no resultado dos participantes, ATT, poderá ser estimado da seguinte maneira:

$$ATT = E(D \mid P(X) = 1) = E(Y_1 - Y_0 \mid X, P(X)=1) =$$

$= E(Y_1 – Y_0 \mid P(X)=1, p(x)) =$

$$E\{E(Y_1 \mid P(X)=1, p(x)) – E(Y_0 \mid P(X)=0, p(x)) \mid P(X) = 1\} \quad \textbf{(B.2)}$$

Para ser possível essa estimativa são necessárias as seguintes hipóteses:

Lema 1: Propriedade de balanço das variáveis pré-tratamento (de controle) dado o escore de propensão. Se *p(x)* e o escore de propensão, então:

$$P \perp X| \, p(x).$$

Se a hipótese de balanço, Lema 1, for satisfeita, a distribuição das características observáveis (e não observáveis) das observações com o mesmo escore de propensão devem ser iguais independentemente do *status* de tratamento (P = 1 ou P = 0). Em outras palavras, dado um determinado escore de propensão, a exposição ao tratamento é aleatória e, assim, observações caracterizadas como de controle ou de tratamento serão, em média, idênticas.

Qualquer modelo-padrão pode ser utilizado na estimativa do escore de propensão.

Por exemplo, Pr{P=1|X} = F(h(X)), onde F(.) é a distribuição acumulada e h(X) é a função de covariâncias de termos cuja escolha é determinada pela necessidade de se satisfazer a hipótese de balanço (Lema 1) na estimativa do escore de propensão.

Lema 2:

$$Y_1,Y_0 \perp P| \, X \Rightarrow Y_1,Y_0 \perp P| \, p(x)$$

A prova de $Y_1,Y_0 \perp P|\ X$ é imediata tendo em vista que P é uma variável dicotômica, só pode assumir os valores 0 ou 1, e a sua distribuição condicional pode ser completamente caracterizada pelo seu valor esperado E(P|.).

O lado direito do Lema 2 pode ser escrito como:

$E\,(P|\ Y_0\ ,Y_1,\ p(x)) = E[E\,(P|Y_0\ ,Y_1,\ X\ )\ |\ Y_0\ ,Y_1,\ p(x)]$, que se iguala à:

$E[E(P|X]\ |\ Y_0\ ,Y_1,p(x)) = E[p(x)|Y_0\ ,Y_1,\ p(x)] = p(x)$ pela definição do escore de propensão pois $E(P|X) = p(x)$.

Assim, a distribuição de P dado $Y_0\ ,Y_{1,}\ p(x)$ depende apenas de $p(x)$.

Como resultado, a média contrafactual $E\,(Y_0\ |\ P(X) = 1)$ também pode ser identificada através da estimativa do resultado esperado condicional ao escore de propensão $p(x)$ que deve ser ponderada pela densidade $f_{\ p|\ P\ =\ 1}$ do escore de propensão entre os participantes:

$E\,(Y_0\,|\ P(X) = 1) = E[E\,(Y_0|\ \ p(x),\ P(X) = 0)\ |\ P(X) = 1] =$

$\int E\,(Y|\ p(x)\ = ?,\ P(X) = 0).\ f_{p\ |P(X) = 1}\,(r)\ d\ r.$

Esse método requer uma regressão não paramétrica de Y em *p* unidimensional apenas conhecida como o *matching* de escore de propensão.

O método de *matching* está baseado na hipótese de que condicionado a um conjunto de variáveis X, Y_0 é independente de P, ou seja, o Lema 2, exposto acima e demonstrado por Ichino (2000).

Sob essa hipótese podemos formar um grupo de controle que reflete o grupo de tratamento. Desde que condicionada à X, a distribuição de $Y_0\ |\ P(X) = 1$ é a mesma de $Y_0\,|\ P(X) = 0$.

Em particular, quando a média existe,

$E(Y_0\ |X,\ P(X) = 1) = E(Y_0\ |X,\ P(X) = 0)$, tal que em X, $B(X) = 0$.

Esse método só elimina o viés se a independência condicional realmente existir. Ou seja, se determinantes não observáveis do resultado dos indivíduos, não incluídas em X, forem descorrelatadas com a participação no programa.

Mesmo que os suportes dos grupos de controle e tratamento sejam iguais, a distribuição das características observáveis pode ser diferente e, assim, o impacto do programa é subestimado. Isto acontece porque, dado os valores de X, existe uma relação sistemática entre a participação no programa e os resultados dos participantes. Existem variáveis observáveis que influenciam conjuntamente o resultado dos tratados e a participação no programa condicional a eles.

Os grupos de controle construídos assumindo que o Lema 2 é verdadeiro diferem dos grupos formados de maneira aleatória. A aleatorização equaliza a distribuição das características dos grupos de controle e de tratamento. Quando a construção desses grupos não é aleatória, a distribuição de suas características não é necessariamente equalizada, mesmo que o Lema 2 seja satisfeito. O suporte das distribuições de X pode ser diferente nos dois grupos e a forma da distribuição pode ser diferente mesmo em regiões de suporte comum.

Muitas vezes não é possível encontrarmos os indivíduos na amostra total para representarem todos os participantes, por isso, estimadores baseados no Lema 2 não necessariamente identificam o impacto do tratamento para todos os valores de X entre os participantes do programa, a não ser que esse impacto não dependa de X.

Existem varias metodologias de aplicação do *matching*, as mais utilizadas na literatura e apresentadas por Becker e Ichino (2002) são:

- (1) vizinho mais próximo,
- (2) *radius matching* ,
- (3) kernel *matching* e
- (4) estratificação.

O método 4 consiste na divisão das observações intervalos de acordo com o escore de propensão. Dentro de cada intervalo a média do escore de propensão das observações de controle e de tratamento devem ser iguais. Para facilitar a prática do processo, podem ser utilizados os mesmos blocos identificados pelo algoritmo da estimação do escore de propensão. Em cada intervalo computamos a diferença entre as médias da variável de resultado das observações de controle e das observações de tratamento. O ATT final será a média ponderada do ATT de cada bloco pelos pesos obtidos através da distribuição das observações tratadas entre os blocos.

Uma das desvantagens dessa metodologia é que não são levadas em consideração observações que pertençam a um bloco onde só existam observações de controle ou apenas observações de tratamento. Uma maneira de resolver isto seria utilizando a metodologia de *matching*-vizinho mais próximo, pois esta procura para cada observação tratada a observação de controle com o escore de propensão mais próximo, que por sua vez pode ser feita com ou sem reposição; neste caso, uma observação de controle pode ser o melhor *matching* para mais de uma observação tratada. Uma vez encontrados todos os *matchings*, a diferença dos resultados de cada grupo é computada e o ATT final será a média dessas diferenças. O único problema dessa metodologia é que algumas vezes a qualidade do *matching* pode não ser muito boa tendo em vista que o controle mais próximo de uma observação tratada pode estar bem distante em termos de escore de propensão. As metodologias *radius matching* e kernel *matching* têm uma solução para este problema.

No método de *radius matching* cada unidade tratada é “casada” com um controle pertencente a uma vizinhança predefinida com base no escore de propensão do tratado. Se definirmos uma vizinhança restrita, existe uma grande chance de não encontrarmos controle dentro desta vizinhança para todas as observações tratadas. Por outro lado, quanto menor a vizinhança, melhor será a qualidade do *matching*, e mais próximos estarão os grupos de tratamento e de controle. O método de kernel leva em consideração todas as observações de tratamento e de controle, e essas são “casadas” de maneira ponderada. Todos os controles são aproveitados e o peso utilizado para cada um é inversamente proporcional à distância do seu escore de propensão e o escore da observação tratada.

Vale ressaltar que, em todos os métodos, a qualidade do *matching* pode ser melhorada quando impomos uma região de suporte comum.

A escolha do método a ser utilizado vai depender do tipo de dados que temos disponível. É clara a existência de um *trade-off* entre qualidade do *matching* e quantidade de observações “casadas”. Tendo em vista que a quantidade de escolas estaduais não é muito grande, quando fazemos uma análise por estado, utilizaremos a metodologia de kernel. Com essa metodologia, quase todas as escolas do Censo Educacional são aproveitadas, sendo retiradas da análise apenas aquelas escolas que estiverem fora da região de suporte comum.

O estimador de kernel conforme foi apresentado por Becker e Ichino (2002) será:

$$T^k = 1/ N^t \sum \{ Y^t - [\sum_{j\ e\ C} Y_j^C \ G(p_j - p_i / h_n) / \sum_{k\ e\ C} G(p_k - p_i / h_n)$$

Onde G(.) é a função do kernel e h_n é a largura da janela, e, sobre condições básicas a respeito dos dois parâmetros,

$[\sum_{j\ e\ C} Y_j^C \ G(p_j - p_i / h_n) / \sum_{k\ e\ C} G(p_k - p_i / h_n)]$ será um estimador consistente do resultado contrafactual de Y_0.

Uma grande vantagem do método experimental (aleatorização) é que este funciona para qualquer conjunto X. No método de *matching* a incerteza quanto à escolha de X que existe no método de variáveis instrumentais permanece. Mesmo que um conjunto de X satisfaça o Lema 2, uma versão reduzida deste pode não o satisfazer.

O método de diferença em diferença é capaz de lidar com esse tipo de problema.

3.3.4. Método não experimental de *matching* com diferença em diferença

Depois da aleatorização, esse método é o mais robusto e confiável para chegarmos ao verdadeiro impacto de um programa. Porém, para ser utilizado, deve-se ter uma linha de base.

A idéia é que tenhamos dados sobre os resultados e seus determinantes antes e depois da implementação do programa, tanto para o grupo que participou quanto para o grupo que não participou. A metodologia compara os resultados dos participantes depois do programa com os seus resultados antes do programa e diminui o resultado da diferença antes/depois dos não participantes. Com esse estimador as tendências comuns de tempo ou idade são eliminadas, além disso, sob a hipótese de que as características não observáveis dos indivíduos não variam no tempo, o problema de viés de seleção é resolvido.

O método foi introduzido por Heckman, Ichimura, Smith e Todd (1998) e Ichimura e Todd (1997, 1998), e identifica o estimador do impacto de um determinado programa naqueles que participaram (ATT) de forma não paramétrica. Vale ressaltar que as variáveis determinantes na participação no programa não podem ser influenciadas pelo tratamento.

Para identificar o ATT, algumas hipóteses são necessárias e serão apresentadas a seguir:

a) Tendência comum: ambos os grupos, de controle e de tratamento, devem seguir uma mesma tendência.

$$E(Y_{0,t}|X,P(X)=1)-E(Y_{0,t`}|X,P(X)=1)= E(Y_{0,t}|X,P(X)=0) - -E(Y_{0,t`}|X,P(X)=0)$$

Onde t representa um período no tempo após a implementação do programa e t` representa um período antes do programa ser implementado.

b) Não existe diferença entre os grupos antes da implementação do programa:

$$E(Y_{1,\,t`} - Y_{0,\,t`} | \; X, P(X)=1) = 0$$

c) Região de suporte comum:

Assim como na metodologia de *matching*, Heckman, Ichimura e Todd (1997) e Heckman, Ichimura, Smith e Todd (1997,1998) destacaram que a semelhança das características de ambos os grupos é essencial para o cálculo do resultado da situação contrafactual de interesse. Ou seja, os indivíduos não tratados devem ter as mesmas características X dos indivíduos que participaram do programa. Vale ressaltar que o procedimento de definição do grupo de controle já leva em consideração essa hipótese.

O estimador do impacto será o DID (*Difference in Difference*):

$$E\ (Y_1|\ X,\ P(X)=1) - E\{E\ (Y_{1t`}|\ X= x_{t`},\ P(X)=1)\} - E\{E\ (Y_{1t}|X= x_t,\ P(X)=0)\} - E\{E\ (Y_{1t`}|X= x_{t`},\ P(X)=0)\}$$

Esse método é uma extensão do método de *matching* porque não requer que não exista viés, apenas que este seja o mesmo ao longo do tempo condicional à X.